

Métodos cuantitativos computarizados para biología pesquera

Isaías Hazarmabeth Salgado Ugarte
Verónica Mitsui Saito Quezada



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA



Métodos cuantitativos computarizados para biología pesquera

**Isaías Hazarmabeth Salgado Ugarte
Verónica Mitsui Saito Quezada**



Universidad Nacional Autónoma de México
Facultad de Estudios Superiores Zaragoza



Datos para catalogación bibliográfica

Autores: Isaías Hazarmabeth Salgado Ugarte, Verónica Mitsui Saito Quezada.

Métodos cuantitativos computarizados para biología pesquera.

UNAM, FES Zaragoza, noviembre de 2020.

Peso: 25.9 MB

ISBN: 978-607-30-3715-0.

Diseño de portada: Carlos Raziel Leños.

Diseño y formación de interiores: Claudia Ahumada Ballesteros.

Esta obra se realizó con apoyo de la UNAM a través de la DGAPA, PAPIME, clave PE207417 y la FES Zaragoza.

DERECHOS RESERVADOS

Queda prohibida la reproducción o transmisión total o parcial del texto o las ilustraciones de la presente obra bajo cualesquiera formas, electrónicas o mecánicas, incluyendo fotocopiado, almacenamiento en algún sistema de recuperación de información, dispositivo de memoria digital o grabado sin el consentimiento previo y por escrito del editor.

Métodos cuantitativos computarizados para biología pesquera.

D.R. © Universidad Nacional Autónoma de México

Av. Universidad # 3000, Col. Universidad Nacional Autónoma de México, C.U.,
Alcaldía de Coyoacán, C.P. 04510, Ciudad de México, México.

Facultad de Estudios Superiores Zaragoza

Av. Guelatao # 66, Col. Ejército de Oriente,
Alcaldía de Iztapalapa, C.P. 09230, Ciudad de México, México.



Prólogo	9
Prefacio	13
Agradecimientos	17
Capítulo 1. Análisis de frecuencia de tallas	21
1.1 Histogramas	21
1.2 Polígonos de frecuencia	29
1.3 Estimadores de densidad por kernel (EDKs)	33
1.4 Estimadores de densidad por kernel de amplitud variable	38
1.5 Elección de amplitud de banda	41
1.6 Cálculo eficiente de EDKs (discretizado y ASH-WARP)	42
1.7 Algunas consideraciones	45
1.8 Reglas prácticas para selección de intervalo/banda	48
1.8.1. Introducción	48
1.8.2. Reglas para elección de número de intervalos en histogramas	48
1.8.3. Reglas para elección de amplitud de intervalo en histogramas	49
1.8.4. Reglas sobreesuavizadas	51
1.8.5. Reglas para elección de número y amplitud de intervalo en polígonos de frecuencia	52
1.8.6. Reglas para elección de amplitud de banda en EDKs	54
1.8.7. Reglas sobreesuavizadas para EDKs	55
1.9 Métodos de Cómputo Intensivo:	56
1.9.1. Validación cruzada por mínimos cuadrados	56
1.9.2. Validación cruzada sesgada	58
1.9.3. Ejemplos de Aplicación	59



1.9.4. Algunos comentarios sobre la elección del parámetro de suavización	76
1.9.5 Evaluación no paramétrica de la multimodalidad	76
1.9.6 Prueba de Multimodalidad de Silverman	77
1.10 Ejemplos de aplicación	81
1.10.1. Uso de reglas de amplitud de banda	81
1.10.2. Aplicación de la prueba de multimodalidad de Silverman	85
1.10.3. Algunos comentarios adicionales	92
1.11. Caracterización de componentes gaussianos en distribuciones multimodales	100
1.11.1. Introducción	100
1.11.2. Método de Bhattacharya	102
1.12. Análisis de las frecuencias de longitud del “kidai” (<i>Dentex tumifrons</i>)	110
1.13. Análisis de frecuencia de tallas de <i>Cathorops melanopus</i> por EDKs	120
1.14. Notas acerca de los cálculos	125
1.15. Mezcla finita de distribuciones	126
1.16. Comparación de Dos Distribuciones de Frecuencia (Kolmogorov-Smirnov)	141

Capítulo 2. Alometría **149**

2.1 Introducción	149
2.2 Relaciones entre dimensiones lineales	156
2.2.1. Otros ejemplos: Equinoideos	159
2.2.2. Gasterópodo <i>Leucozonia cerata</i>	164
2.2.3. Pelecípodo <i>Pinctada mazatlanica</i>	169
2.3 Relaciones en Dimensiones mixtas	172
2.3.1. Otros ejemplos de relaciones P-L	179
2.3.2. Gasterópodo <i>Hexaplex princeps</i>	180
2.3.3. Pelecípodo <i>Hyotissa hiotis</i>	183
2.4. Relaciones entre dimensiones de volumen	184

Capítulo 3 Estimación de la edad **189**

3.1. Introducción	189
3.2. Escamas. Relación Cuerpo-Estructura	191
3.3. Otolitos	197



3.4. Relación Cuerpo-Otolito	201
3.5. Validación de la periodicidad de las marcas	202
3.6. Opérculos del gasterópodo <i>Astraea undosa</i>	203
3.7 Otolitos de la tilapia <i>Oreochromis niloticus</i> . Validación de periodicidad	214
3.7.1 Análisis de bordes	216
3.7.2. Análisis del incremento marginal	218
3.7.3. Razón del incremento marginal	220

Capítulo 4 Modelación del crecimiento individual **223**

4.1 Introducción	223
4.2 Estimación de la edad	224
4.3 Estimación de los parámetros de la función de crecimiento de von Bertalanffy (FCvB)	226
4.3.1 Gráfico de Ford-Walford	227
4.3.2 Gráfico de Gulland	228
4.3.3 Gráfico de Gulland-Holt	229
4.3.4 Método de Beverton y Holt	230
4.3.5 Regresión no lineal	230
4.4 Caso 1. Análisis de frecuencia de tallas	231
4.5 Caso 2. Longitud-edad a partir de lectura de estructuras duras	245
4.6 Algunas notas adicionales	251
4.7 Otras funciones de crecimiento	251
4.8 Caso 3. Crecimiento del caracol <i>Hexaplex princeps</i>	253
4.9 Caso 4. Crecimiento del holoturoideo (pepino de mar) <i>Holothuria inornata</i>	262
4.10 Ajuste de otras funciones con R	266
4.11 Crecimiento estacional. Ajuste para la FCvB	267
4.12 Ajuste de la FCvB por maxima verosimilitud	277

Capítulo 5 Comparación de funciones de crecimiento **285**

5.1 Introducción	285
5.2 Prueba de la T^2 de Hotelling	285



5.3 Suposiciones de la prueba T^2	288
5.4 Pruebas de razón de verosimilitud	289
5.5 Prueba global de curvas de crecimiento	291
5.6 Pruebas de máxima verosimilitud	291
5.7 Programas	293
5.8 Consideraciones sobre comparación de curvas de crecimiento	302
5.9 Aplicación a funciones Logística y Gompertz	303

Capítulo 6 Índices Morfofisiológicos **313**

6.1 Introducción	313
6.2 El caso del pez “Suzuki” (<i>Lateolabrax japonicus</i>)	316
6.2.1 Material y métodos	318
6.2.2. Análisis multivariado de covarianza (ANMUCOVA)	321
6.3 El caso del gasterópodo marino <i>Hexaplex princeps</i>	333
6.3.1. Material y método	334
6.3.2. Índices morfofisiológicos	337
6.3.3. Análisis multivariado de covarianza (ANMUCOVA)	343
6.3.4. Valores suavizados	347
6.3.5. Comparación de índices con Temperatura y Clorofila	350
6.3.6. Consideraciones acerca de los resultados	352
6.4 Comandos en Stata	354

Capítulo 7 Estadios de Madurez Gonádica **363**

7.1 Introducción	363
7.2 Escalas macroscópicas de madurez	363
7.3 Relación madurez talla	365
7.4 Escalas histológicas de madurez	372
7.5 Ciclo reproductivo del gasterópodo <i>Hexaplex princeps</i>	373
7.5.1. Gráficos de Área	377
7.5.2. Suavización no lineal resistente	383
7.5.3. Correlación cruzada	386



7.6 Talla de primera madurez	392
7.7 Comparación con curva logística de dos parámetros	396
Bibliografía	401
Apéndices	419
Apéndice 1: Uso del programa AED2020	421
Apéndice 2: Listado de Programas para Stata	448
Apéndice 3: Listado de Archivos de Datos	451
Apéndice 4: Listado de Objetos para R	454
Apéndice 5: Comandos de Stata para Análisis Multivariado de Covarianza	460
Apéndice 6: Hojas de Datos en Excel	468
Apéndice 7: Programa PAC (Programas para el Análisis del Crecimiento)	469
Apéndice 8: Ajuste de distribuciones Gaussianas por Máxima Verosimilitud de la distribución Multinomial	472
Índice Temático	477



El estudio de los recursos bióticos que permita una mejor comprensión de sus ciclos de vida, así como una mejor administración para su aprovechamiento es sin duda una tarea permanente y continua, debido a las diversas condiciones y necesidades que la humanidad impone y enfrenta. Para el caso particular de recursos pesqueros el reto es grande dada las dificultades que el ambiente establece, así como las características de movilidad de la mayoría de las especies y el marco de acceso abierto para múltiples usuarios.

La sociedad humana se beneficia de muchas maneras de los servicios a los ecosistemas generados por los recursos pesqueros, además de los beneficios económicos directos. Sin embargo, el incremento en la presión de pesca, contaminación, destrucción del hábitat, introducción de especies exóticas y otros factores de estrés continúan ejerciendo una fuerte presión sobre dichos recursos y sus hábitats.

La captura pesquera se apropia de una parte sustancial (8%) de la producción primaria global en el mar y requiere entre 24 y 35% de la producción de surgencias y plataforma continental. Además, el objetivo de manejadores pesqueros por altos rendimientos de unas pocas especies de los límites superiores de las cadenas tróficas ha movido a las comunidades bióticas hacia una composición dominada por niveles tróficos inferiores los cuales, obligadamente son capturados por las pesquerías. Todavía es necesario desarrollar criterios y métodos de valoración de los distintos bienes y servicios que los diversos recursos pesqueros nos proporcionan además de su uso directo.

Los peces constituyen más de la mitad de las 54,711 especies de vertebrados existentes con aproximadamente 29,400 especies vivientes, constituidas en 515 familias y 62 órdenes. Dentro de los invertebrados, grupos importantes por su explotación desde tiempos remotos son los moluscos (con más de 100,000 especies vivientes) que incluyen principalmente a los gasterópodos, pelecípodos y cefalópodos; en ciertas regiones del mundo son apreciados otros grupos como los equinodermos (equinoideos y holoturoideos). Los aspectos básicos importantes para la sobrevivencia de estos organismos están relacionados con la hidrología (niveles de oxígeno disuelto, nutrientes, salinidad, penetración de la luz, temperatura y contaminantes), así como los parámetros bióticos y ecología del sistema (recursos alimenticios, depredación, heterogeneidad de hábitat), todo lo cual solo se menciona como ejemplo de la magnitud del reto en investigación que se requiere y más cuando se considera que en ambientes tropicales se desarrollan las pesquerías multiespecíficas.



La biología pesquera se ha desarrollado a lo largo del tiempo como la disciplina científica que aborda el estudio de los recursos pesqueros con un enfoque holístico, buscando desarrollar métodos de evaluación que generen criterios y argumentos sólidos para un manejador de recursos. Métodos de muestreo, de integración de bases de datos, de análisis estadísticos, de valoración económica, de evaluación socio-ambiental, de gobernanza, entre otros, han sido desarrollados y aplicados para el manejo de diversos recursos pesqueros con resultados que muchas veces son solo aproximaciones teóricas.

Actualmente el uso de la tecnología facilita las diversas tareas involucradas en el desarrollo de la biología pesquera, particularmente la integración de bases de datos robustas, así como su análisis y es justo aquí donde la obra "*Métodos cuantitativos computarizados para biología pesquera*" representa una aportación única ya que hace un recorrido desde el tradicional análisis de frecuencia de tallas, las relaciones de alometría, la estimación de edad y su modelación mediante diversas funciones, hasta los índices morfofisiológicos y los estados de madurez gonádica, temas que muchas veces se encuentran disgregados en la literatura científica especializada.

Es una obra de lectura ágil y dinámica que sobre todo esta dirigida a profesionales en formación, pero que deberán contar con suficientes antecedentes en dinámica poblacional, así como en el uso de programas de cómputo especializados en estadística. Una característica sobresaliente es la aportación de ejemplos con datos reales que los autores han recabado en múltiples experiencias de investigación y que sin duda redondean la explicación teórica. También es importante destacar que se aportan herramientas conceptuales y metodológicas que manejadores de recursos apreciarán para generar opiniones argumentadas.

En el capítulo uno hay un énfasis particular de los autores en el análisis de la frecuencia de tallas como la base para la descripción de la dinámica poblacional de un recurso. El uso de los estimadores de densidad por kernel, las reglas para la selección de intervalo (número y amplitud), las pruebas de multimodalidad y sus componentes Gaussianos, constituyen una propuesta metodológica integradora y que se convierte en una herramienta valiosa con un gran peso estadístico para el análisis de datos no solo de recursos pesqueros, sino también de especies de relevancia ecológica.

Se coincide en señalar que la adecuada determinación de los parámetros de la función de crecimiento es el eje central en la determinación del mejor momento para el aprovechamiento y conservación del recurso. Y ya sea que tales parámetros se estimen mediante el análisis de la frecuencia de tallas o mediante la determinación de líneas de crecimiento en estructuras duras y el análisis de las fases de maduración gonádica, los procedimientos requieren de amplia experiencia y una fuerte inversión de tiempo de dedicación. Por tanto, los profesionistas dedicados a la biología pesquera, tanto para el desarrollo de investigación



científica como para generar opiniones en la toma de decisiones, invariablemente deberán contar con bases de datos robustas y adecuadamente estructuradas para su análisis.

Es importante destacar que la presente obra actualiza metodología tradicionalmente utilizada en la biología pesquera con el uso de herramientas de cómputo y de programas estadísticos actualizados lo cual facilita enormemente el manejo y análisis de las bases de datos mencionadas.

La gran experiencia de los autores tanto en la investigación pesquera como en la formación de recursos humanos, particularmente de posgrado, se ve reflejada en este libro, el cual, sin duda, se convertirá en un referente metodológico tanto para académicos con experiencia en el tema como para aquellos interesados en incursionar profesionalmente en la evaluación pesquera.

Luis Amado Ayala Pérez
Universidad Autónoma Metropolitana
Unidad Xochimilco



Para la sobrevivencia de la humanidad a largo plazo es inevitable conseguir una relación armoniosa con los demás componentes (bióticos y abióticos) de la Biosfera. Para ello se requiere de la comprensión profunda de las interacciones entre los participantes de los ecosistemas. Los humanos, como organismos consumidores, para subsistir necesitan de otros seres vivos. En esta relación de dependencia para la subsistencia se encuentra la pesca. Desde la antigüedad más remota la captura de organismos acuáticos ha permitido la existencia y el desarrollo de los grupos humanos y sus parientes. Es conocido que el uso excesivo de los recursos naturales puede llevar al colapso a toda una civilización. En tiempos recientes se ha tomado conciencia de que para evitar la extinción de los recursos es necesario conocerlos a profundidad con el fin de conocer el mejor momento y la cantidad adecuada de unidades (individuos en la mayoría de los casos) a capturar. Dentro de los conocimientos necesarios para conseguir una explotación que no agote el recurso biótico están, el crecimiento, la reproducción y la alimentación.

Por otra parte, el conocimiento acerca de los procesos biológicos ha ido haciéndose cada vez más cuantitativo y el manejo de la información simbólica y numérica es en la actualidad de suma complejidad por las intrincadas relaciones y su enorme cantidad. De esta forma, el uso de auxiliares para el cómputo, en la actualidad es indispensable. No sólo disponer de computadoras permite avanzar en el conocimiento, es necesario también contar con los algoritmos en adición a saber cómo manejarlos.

Con lo anterior como base se integró la presente obra, la cual pretende introducir a los interesados en el uso de herramientas computarizadas para el uso de algoritmos cuantitativos enfocados a la obtención de conocimiento de la información recabada en las investigaciones de la Biología Pesquera. Se ha incluido en diferentes instancias, el uso de Excel (hoja de cálculo), de un ejecutable escrito en Visual Basic, varias rutinas en lenguajes de Stata y R, complementados con algoritmos traducidos de Turbo-Pascal al lenguaje actual Python.

El primer capítulo se ha dedicado a un tema que se ha vuelto sumamente importante, sobre todo en el estudio de los recursos pesqueros de zonas cálidas del planeta, el análisis de la frecuencia de tallas. Se repasan histogramas (H), polígonos de frecuencia (PF) y se llega a los estimadores de densidad por kernel (EDKs), que son versiones mejoradas de las gráficas tradicionales de la información contenida en tablas de frecuencia (es decir, Hs y FPs). Se presentan, asimismo, algoritmos eficaces (ASH y WARP) de cálculo que resuelven



el problema de la cantidad enorme de cálculos que los EDKs requieren. La elección de la amplitud de clase (intervalo, banda en los EDKs) es fundamental para obtener una estimación adecuada. Para ello se incluye el uso de reglas prácticas y se presentan dos enfoques de cómputo intensivo (sólo factibles por la disponibilidad de ordenadores y programas ex profeso): la validación cruzada (mínimos cuadrados y sesgada) y bootstrap (muestreo repetitivo con reemplazamiento) suavizado para probar la modalidad de una distribución.

En el caso de las distribuciones multimodales, puede ser razonable atribuirles a la mezcla de componentes individuales mezclados en la población (cohortes). Para caracterizar a estos grupos se incluye el clásico método semigráfico de Bhattacharya, el ajuste iterativo de componentes por máxima verosimilitud (distribuciones mezcladas) e inclusive el uso de la distribución multinomial.

En el segundo capítulo se presentan las relaciones entre las dimensiones de los organismos (Alometría) no sólo enfocada a la relación entre el peso y la longitud, sino también se consideran relaciones entre longitudes y pesos entre sí con ejemplos de equinoideos y moluscos (pelecípodos y gasterópodos).

La interpretación de marcas periódicas en estructuras duras de los organismos para estimar la edad se presenta en el capítulo tercero en el que se consideran los métodos para validar primero si la estructura es de utilidad para estudiar el crecimiento (relación cuerpo-estructura) y posteriormente la periodicidad (por el análisis del tipo de borde y su incremento marginal) con ejemplos en pez y caracol marino.

Ya sea por la identificación de cohortes (distribuciones de tamaño multimodales) o la validación de la periodicidad de marcas en estructuras duras (opérculos) y su relación con el tamaño corporal es posible estimar la forma en que los organismos crecen individualmente. Tema que se incluye en el capítulo cuatro. Se repasan los tradicionales métodos basados en la linealización del modelo matemático más usado en Biología Pesquera (von Bertalanffy) y el uso del método iterativo de la regresión no lineal que permite el ajuste de otras funciones como la Logística y la de Gompertz. Además de ejemplificarlo en un caracol marino se aplica la regresión no lineal a cohortes posibles de un pepino de mar. Este capítulo finaliza con métodos iterativos para el ajuste de formas de crecimiento variables a lo largo de las estaciones del año mediante funciones de crecimiento modificadas del modelo de von Bertalanffy por la adición de uno o dos componentes cíclicos (trigonométricos).

La información sobre el crecimiento por lo general proviene de varios lugares o bien los organismos presentan dimorfismo sexual, por lo que es de interés verificar si existen



diferencias entre las maneras en que crecen grupos diferentes. En el Capítulo 5 se han incluido varias pruebas para llevar a cabo esta comparación: La T^2 multivariada de Hotelling, la prueba de razón de verosimilitud y la de máxima verosimilitud, aplicada al crecimiento de dos grupos, ajustados por funciones de von Bertalanffy, Logística y Gompertz, así como una prueba para más de dos funciones.

La reproducción y la condición de los organismos son aspectos fundamentales para el conocimiento biológico de los organismos. En el Capítulo 6 se presentan dos enfoques para su estudio: el uso de los índices morfofisiológicos tradicionales (factor de condición, índices gonadosomático y hepatosomático) comparados con el uso de medias ajustadas por tamaño corporal provenientes del Análisis Multivariado de Covarianza (ANMUCOVA). Se ejemplifica su uso con la aplicación a los datos de un pez de zona templada ("suzuki") y un caracol marino del sur de México ("caracol chino").

En el Capítulo 7 se introducen otros temas de importancia para el estudio de la reproducción como son las escalas de madurez gonádica tanto las basadas en las características macroscópicas de las gónadas, como aquellas derivadas de su estudio histológico. También se consideran funciones logísticas para utilizar la información de las escalas en la estimación del tamaño de los primeros individuos capaces de reproducirse.

Cabe señalar que (como el título del libro lo indica) en todos los Capítulos y sus secciones se presentan los comandos para llevar a cabo los procedimientos y obtener los cuadros y gráficas respectivos, no con un solo lenguaje sino por lo general en Stata y en R. La validación cruzada se ejemplifica con el uso del programa AED2020 ejecutable de VisualBasic y el ajuste por máxima verosimilitud de la distribución multinomial para caracterizar componentes gaussianos hace uso de la herramienta "Solver" de Excel.

Finalmente, en los Apéndices incluyen una guía para el uso del AED2020, listas de programas para Stata, de archivos con los datos, de objetos de R, los comandos de Stata para el ANMUCOVA, hojas de Excel, una guía para el Programa PAC (Programas para el Análisis del Crecimiento) y el ajuste de componentes gaussianos por máxima verosimilitud de la distribución Multinomial (con "Solver" de Excel).

Esta obra puede considerarse como complementaria, revisada, actualizada y aumentada de los libros publicados anteriormente sobre Suavización no paramétrica y Métodos actualizados para análisis de datos biológico-pesqueros.

Como los escritos anteriores, los autores esperamos que este texto sea de utilidad a estudiosos de la Biología Pesquera, ya sea académicos, administradores de recursos,



estudiantes de licenciatura y posgrado involucrados con la explotación de los recursos naturales renovables de México.

Información adicional sobre la obra (programas, datos, fe de erratas) está disponible en el blog “Biometría y Biología Pesquera” en la siguiente liga:

<https://blogceta.zaragoza.unam.mx/biombiolpesca/>

Isaías Hazarmabeth Salgado Ugarte
Verónica Mitsui Saito Quezada
Septiembre de 2020

Agradecimientos



Este texto no hubiera sido posible sin la participación de numerosos colaboradores, profesores, colegas y estudiantes durante numerosas etapas desarrolladas a lo largo de varios años.

Al M. en C. Ernesto Mendoza Vallejo y al Dr. José Luis Gómez Márquez por su participación en nuestras primeras etapas de formación en la Universidad Nacional Autónoma de México (UNAM) en la Escuela Nacional de Estudios Profesionales (ENEP) ahora Facultad de Estudios Superiores (FES) Zaragoza). Durante la siguiente etapa recibimos invaluable asesoría y apoyo del Dr. Armando Ortega Salas, Dr. Alejandro Yáñez Arancibia† Instituto de Ciencias del Mar y Limnología (ICMyL, UNAM), Dr Jaime Curts García (University of Texas). Asimismo de la Dra. Georgina Gluyas Millán del Centro Regional de Investigación Pesquera, La Paz, Instituto Nacional de la Pesca.

Posteriormente, durante nuestra estancia en la Universidad de Tokio, Laboratorio de Pesquerías, fuimos apoyados ampliamente por el Dr. Makoto Shimizu y el Dr. Toru Taniuchi; para la obtención de muestras por el Dr. Katsumi Matsushita y en asesoría ecológica por el Dr. Mitsuhiko Sano. En México recibimos el auxilio y apoyo de la Dra. Laura Elena Sanvicente Añorve (ICMyL, UNAM) y la Dra. Ana Bertha Villaseñor Martínez (Instituto de Geología, UNAM).

Sin la colaboración de la profesora Esther A. Uría Galicia (Departamento de Histología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional) la histología de los gasterópodos marinos no hubiera sido posible. En esta fase reciente de formación tuvimos el apoyo del Dr. José Luis Gómez Márquez (FES Zaragoza) y la Dra. Ma. de Lourdes Jiménez Badillo (Laboratorio de Ciencias Marinas y Pesquerías, Universidad Veracruzana).

Especialistas y académicos de otros países nos proporcionaron asesoría técnica y científica: Dr. Joseph Hilbe (Arizona State University), Dr. Wolfgang Härdle (Humboldt-Universität zu Berlin) y Dr. David W. Scott (Rice University), quien en su última visita a México (2016), nos obsequió la última edición de su (fundamental) libro sobre estimación de densidad.

El Ministerio de Educación, Ciencia y Cultura de Japón, el Consejo Nacional de Ciencia y Tecnología de México, la UNAM a través de la Dirección General de Asuntos del Personal Académico (DGAPA) y la FES Zaragoza a través del Programa de Apoyo a Proyectos de Investigación e Innovación (PAPIIT) IN2175596, IG201215 en codirección



con Dr. Felipe Amezcua Martínez (ICMyL Mazatlán) y Dr. Píndaro Díaz Jaimes (ICMyL Ciudad Universitaria, CdMx) y del Programa de Apoyo a Proyectos de Innovación y Mejoramiento de la Enseñanza (PAPIME) 192031, EN221403, PE205407, PE206213, RL200316, PE207417, además de la Universidad Autónoma Metropolitana Iztapalapa a través de la Cátedra Divisional “Dr Ramón Riba y Nava Esparza” (con el apoyo del Biól. Marco A. Pérez Hernández) brindaron sustento económico en algún momento para llevar a cabo alguno de los contenidos que en esta obra se incluyen actualizados.

Durante la integración de las rutinas para comparación de funciones de crecimiento, recibimos ayuda de la Dra. Ana Laura Ibañez Aguirre (UAM Iztapalapa), Dr. Xavier Chiappa Carrara (Escuela Nacional de Estudios Superiores Unidad Mérida, UNAM), Dr. Manuel Gallardo Cabello† (ICMyL, UNAM) y la Dra. Alicia Cruz Martínez (Facultad de Ciencias, UNAM).

Manifestamos nuestro agradecimiento por su tiempo y cuidadosa revisión a la Dra. Gabriela Galindo Cortes (Instituto de Ciencias Marina y Pesquerías, Universidad Veracruzana), al Dr. Luis Amado Ayala Pérez (Universidad Autónoma Metropolitana, Xochimilco), al M en E Eddie René Zambrano Román (Facultad de Ciencias Naturales, Universidad de Guayaquil, Ecuador), quien además contribuyó con versiones nuevas de varias figuras, al Dr. Gorgonio Ruiz Campos (Facultad de Ciencias, Universidad Autónoma de Baja California) y Dr. José Luis Gómez Márquez (Laboratorio de Limnología, FES Zaragoza, UNAM). Sus anotaciones nos llevaron a corregir numerosas imprecisiones y omisiones en el manuscrito preliminar. Si bien consideramos sus anotaciones y correcciones, cualquier error adicional es nuestra responsabilidad. Agradeceremos de nuestros lectores notificarnos al respecto. El Dr. Ayala Pérez además, amablemente escribió las palabras del Prólogo.

En relación con los programas, queremos hacer patente nuestro agradecimiento: en las rutinas de Visual Basic recibimos la ayuda y colaboración del M. en C. José Ricardo Rodríguez Rojas. Para la actualización de algunas rutinas de Stata la Dra. Isabel Canette (Stata Corporation) nos brindó atinadas instrucciones de programación y el Dr Christopher F. Baum puso a disposición en la internet el módulo “silvermtest” para la prueba bootstrap de multimodalidad. Joel Cervantes de Multion Consulting nos auxilió para la obtención de la última versión de Stata (16).

En esta obra se han sustituido rutinas escritas en TurboPascal por algoritmos traducidos al lenguaje actual Python. En esta labor recibimos la ayuda del Licenciado en Física Noel Isaías Plascencia Díaz y la Licenciada en Ciencias de la Tierra Mitsui Myrna Salgado Saito. Estas rutinas permitieron recuperar el uso de comandos de rutinas que ya no funcionaban con las nuevas versiones de Stata y las cuales se integraron a los guiones nuevos de R.



Se agradece al Dr. David N. Espinosa Organista, Jefe de la Unidad de Investigación y Estudios de Posgrado y su equipo de colaboradores, Mtro. José Luis Trejo, M. en C. Genaro Montaña Arias por su apoyo para disponer de espacio físico y recursos computarizados utilizados para integrar esta obra. Asimismo a la estructura administrativa de la Carrera de Biología de la FES Zaragoza.

Reconocemos la participación del Lic. Carlos Raziél Leños Castillo, Jefe del Departamento de Publicaciones y Comunicación Gráfica por coordinar la revisión del manuscrito y su auxilio para culminar con la publicación del mismo. Gracias al equipo de los sitios Blog FES Zaragoza por las facilidades otorgadas en la creación del blog de nuestro grupo académico: "<https://blogceta.zaragoza.unam.mx/biombiolpesca/>".

Por la administración de los recursos del proyecto PE207417 agradecemos al Mtro. Jorge Enrique Carbajal López (Srio. Administrativo) y a Mtra. Wendy Castro Díaz (Responsable de Proyectos).

Sin el auxilio de los buzos Manolo Jarquín, Primitivo Herrera Ordóñez y el Capitán "Beto" (Abraham Reyes López), así como de los pescadores artesanales de Puerto Ángel, la recolección de invertebrados y peces considerados en varios capítulos de esta obra no hubiera sido posible.

La integración de gran parte de los materiales de este libro no hubiera sido posible sin la participación de los alumnos, tanto de licenciatura, como de maestría y doctorado, así como de académicos posdoctorantes del Laboratorio de Biometría y Biología Pesquera. La lista completa sería muy extensa, sin embargo no podemos dejar de mencionar a Biól. Edgar Omar Guzmán Urrieta, Biól. Mariana Evelyn Meléndez Contreras, Ana Gabriela González Castillo, María Alicia Rivera Barrera, Cynthia Carapia Moreno, Viviana Axel Rodríguez Galicia, MVZ Samantha Vázquez Servín, Gabino Cordero Montoya, Ma. Del Pilar Cifuentes Salazar, Diana Karina Morales Ramírez, M. en C. Iztzel Pérez Olivares, M. en C. Candelario Jiménez Olivares, M. en C. Mario Antonio Gatica Martínez, M. en C. Mixtli Crisóstomo Pérez, Dra. Zamira A. Ávila Valle, Dra. Marcia M. Ramírez Sánchez, Dr. Manuel Feria Ortíz.

Dedicamos este libro a nuestras hijas Mitsui Myrna y Mayumi Ninfa así como a la memoria de nuestros padres Taro Saito† (VMSQ), Myrna Ugarte Mejía† y José Marte Isaías Salgado Salgado (IHSU).

Capítulo 1

Análisis de frecuencia de tallas



1.1 Histograma

La estructura de tallas de una población de peces (o de otros organismos) representa una respuesta sintética a factores ecológicos propios de las poblaciones o las comunidades (Ogle, 2016). El análisis de la frecuencia de tallas es una de las herramientas más utilizadas en la evaluación de las poblaciones explotadas por pesca. La distribución de los tamaños de los individuos ofrece una imagen instantánea que muestra las interacciones dinámicas del reclutamiento, el crecimiento y la mortalidad, por lo que permite reconocer problemas como inconsistencia en clases anuales, crecimiento lento o alta mortalidad (Neumann y Allen 2007).

Tradicionalmente, para analizar datos de frecuencia de tallas se han empleado histogramas y polígonos de frecuencia. De acuerdo con Härdle (1991) y Härdle y Simar (2015) un histograma se obtiene dividiendo la línea de los números reales en intervalos (B_j):

$$B_j(x_0, h) = [x_0 + (j-1)h, x_0 + jh], \quad j \in \mathbb{Z}$$

con amplitud $h > 0$ y origen en x_0 . Notar que debido a la especificación de un intervalo $[a, b)$ (cerrado-abierto), una observación en particular en el límite de un intervalo se coloca en el intervalo superior.

Entonces el histograma como estimador de densidad de la variable x que depende de una amplitud de intervalo h , $\hat{f}_h(x)$ es:

$$\hat{f}_h(x) = n^{-1} h^{-1} \sum_{j \in \mathbb{Z}} \sum_{i=1}^n I\{x_i \in B_j(x_0, h)\} I\{x \in B_j(x_0, h)\}$$

La primera función indicadora $I\{x_i \in B_j(x_0, h)\}$ cuenta el número de observaciones que caen dentro del intervalo $B_j(x_0, h)$, es decir, esta función $I\{ \}$ asigna un valor de 1 a los elementos pertenecientes al conjunto (especificado por el intervalo) y 0 a los demás. La segunda función indicadora localiza los conteos alrededor del valor “ x ”. En esta definición,



las frecuencias han sido normalizadas (frecuencia del intervalo dividida por el producto del número total de observaciones “ n ” con la amplitud de clase “ h ”) para que el área bajo las barras del histograma sea igual a la unidad (escala de densidad), si bien, en el uso común de los histogramas, el eje vertical (“ y ”) representa el número (frecuencia) de las observaciones que pertenecen a un intervalo (clase) aunque pueden utilizarse escalas de porcentaje ó fracción.

Los histogramas y los polígonos de frecuencia son también estimadores de la distribución de densidad del conjunto de datos. A pesar de su amplio uso, estos estimadores de densidad pueden ser muy generales y faltos de definición para propósitos del análisis detallado de la distribución (Tarter y Kronmal, 1976). Existen cuatro problemas cuando se usan histogramas (Fox, 1990):

1. Dependencia del origen. El investigador debe escoger la posición del origen de los intervalos (a menudo usando números redondeados convenientemente). Esta subjetividad puede llevar a estimaciones confusas debido a que un cambio en el origen puede cambiar las características de la distribución, como la simetría, el sesgo, así como el número y posición de las modas en la estimación de densidad (Silverman, 1986; Fox, 1990; Scott, 1992; 2015).

Para ejemplificar este problema se utilizarán los datos de longitud estándar (mm) del bagre estuarino *Cathorops melanopus* (Ariidae) (83 hembras (código numérico = 2) y 100 organismos de sexo indeterminado (código = 3) recolectados en noviembre de 1980) reportados por Salgado-Ugarte (1985). Los 100 organismos indeterminados sexualmente son una submuestra aleatoria del total de los 449 en total del archivo y codificados en la variable “*sexa*” (breve de **sexo** ajustado). Lo anterior con el fin de equilibrar la proporción de individuos. Con Stata (StataCorp, 2019) es posible construir cinco histogramas, utilizando la misma amplitud de intervalo ($h = 30$; la elección de este valor se discutirá posteriormente) pero con origen diferente (24, 30, 36, 42 y 48, respectivamente) mediante el siguiente código:

Inicialmente se carga el archivo de datos (en formato Stata)

```
. use "c:/data/bagretallas"
```

Posteriormente se construyen los histogramas y se graban en archivos.

```
. histogram bodlen if sexa==2 | sexa==3, start(24) width(30)
yttitle(" ") xttitle(" ") xlabel(24(30)300, angle(45) )
xscale(range(0 300)) saving(fig1,replace)
```



```
. histogram bodlen if sexa==2 | sexa==3, start(30) width(30)
ytittle(" ") xtittle(" ") xlab(30(30)300, angle(45) )
xscale(range(0 300)) saving(fig2,replace)

. histogram bodlen if sexa==2 | sexa==3, start(36) width(30)
ytittle(" ") xtittle(" ") xlab(36(30)300, angle(45) )
xscale(range(0 300)) saving(fig3,replace)

. histogram bodlen if sexa==2 | sexa==3, start(42) width(30)
ytittle(" ") xtittle(" ") xlab(42(30)300,angle(45) ) ylab(0(.005).015, grid)
yscale(range(0 .015)) xscale(range(0 300)) saving(fig4, replace)

. histogram bodlen if sexa==2 | sexa==3, start(48) width(30)
ytittle(" ") xtittle(" ") xlab(48(30)300, angle(45) ) ylab(0(.005).015)
yscale(range(0 .015)) xscale(range(0 300)) saving(fig5, replace)
```

El último gráfico es el promedio de los cinco anteriores y es un estimador de densidad por kernel. Estos estimadores se explicarán más adelante. Para obtenerlo se escriben los siguientes comandos:

```
. warpdenml bodlen if sexa==2 | sexa==3, b(30) m(5) k(2) gen(den30 mid30)

scatter den30 mid30, ytittle(" ") xtittle(" ") xlab(0(50)300, angle(45))
ms(i) c(J) saving(fig6, replace)
```

Finalmente, los archivos gráficos individuales se combinan en uno sólo el cual se graba para uso posterior:

```
. graph combine fig1.gph fig2.gph fig3.gph fig4.gph fig5.gph fig6.gph, cols(2)
ysize(7) l1(Densidad) b1("Longitud estándar (mm)") saving(fig1plt)
```

Como puede observarse en el gráfico combinado (Figura 1.1), existen histogramas bimodales y trimodales. Sería arbitrario escoger alguno de estos histogramas para representar la distribución de longitudes. Infortunadamente, este ejercicio de dibujar varios histogramas con el origen desplazado puede conducir al analista a seleccionar (intencionalmente o no) aquél que mejor se ajuste a sus propósitos.



Para reproducir la Figura 1.1 con R se deben de usar los siguientes comandos (R Core Team, 2018):

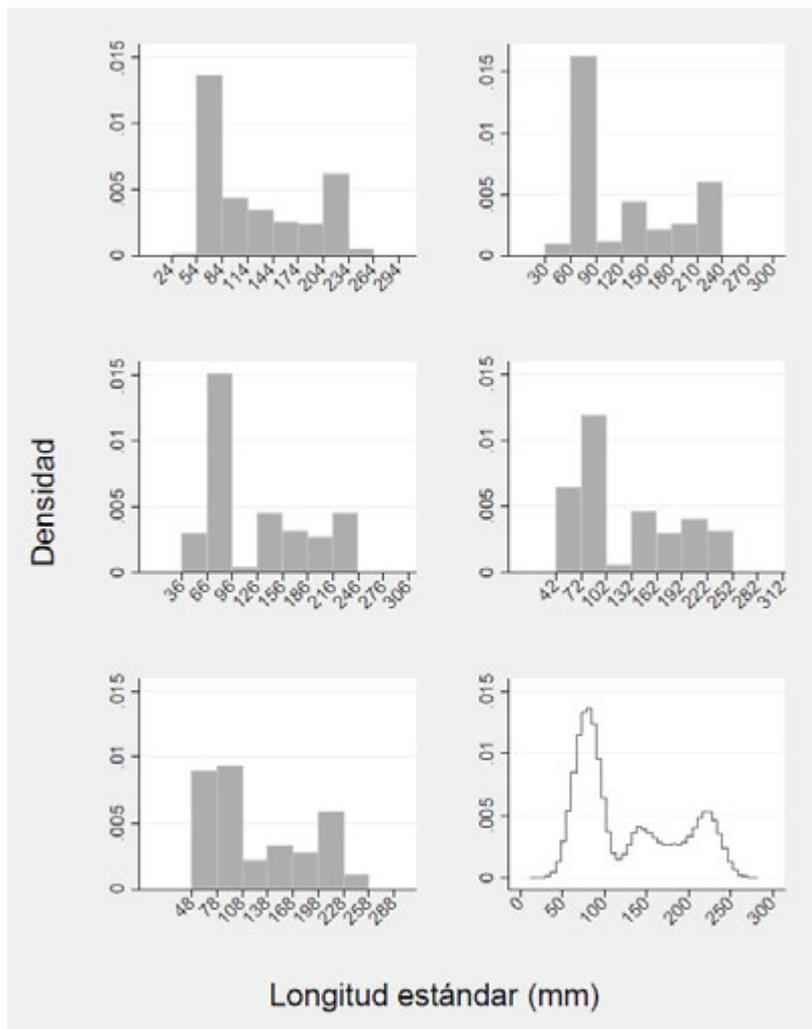


Figura 1.1. Cinco histogramas con orígenes diferentes y el Histograma Desplazado Promedio (HDP) correspondiente para los datos de longitud de hembras e indeterminados del bagre *Cathorops melanopus* (Versión Stata).



```

> setwd("c:/data")
> library(foreign)
> bagtalla <- read.dta("bagretallaso.dta")
> names(bagtalla)
[1] "bodlen" "sex" "nalea" "sexa"
> blhemmin <- bagtalla$bodlen[bagtalla$sexa==2 | bagtalla$sexa==3]
> library(ash)
Warning message:
package 'ash' was built under R version 3.5.2
> op<-par(no.readonly=TRUE)
> par(op)
> par(oma=c(3,3,0,0), mar= c(3,3,2,2))
> layout(matrix(c(1,2,3,4,5,6), byrow=TRUE, ncol=2))
> hist(blhemmin,breaks=seq(24,294,30), main=" ", freq=FALSE, ylab=" ",
xlab=" ", xlim=c(0,300))
> hist(blhemmin,breaks=seq(30,294,30), main=" ", freq=FALSE, ylab=" ",
xlab=" ", xlim=c(0,300))
> hist(blhemmin,breaks=seq(36,294,30), main=" ", freq=FALSE, ylab=" ",
xlab=" ", xlim=c(0,300))
> hist(blhemmin,breaks=seq(42,294,30), main=" ", freq=FALSE, ylab=" ",
xlab=" ", xlim=c(0,300))
> hist(blhemmin,breaks=seq(48,294,30), main=" ", freq=FALSE, ylab=" ",
xlab=" ", xlim=c(0,300))

```

El resultado se muestra en la Figura 1.2

```

> triangle <- c(1,1) #esta especificación está corregida
> ab <- c(9,279)
> f <- ash1(bin1(na.omit(blhemmin),ab,45),5,kopt=triangle)
> plot(f, type="n", xlab=" ", ylab=" ", xlim=c(0,300))
> lines(f, type="s")
> mtext(text="Longitud estándar (mm)", side=1, line=0,outer=TRUE)
> mtext(text="Densidad", side=2, line=0,outer=TRUE)

```

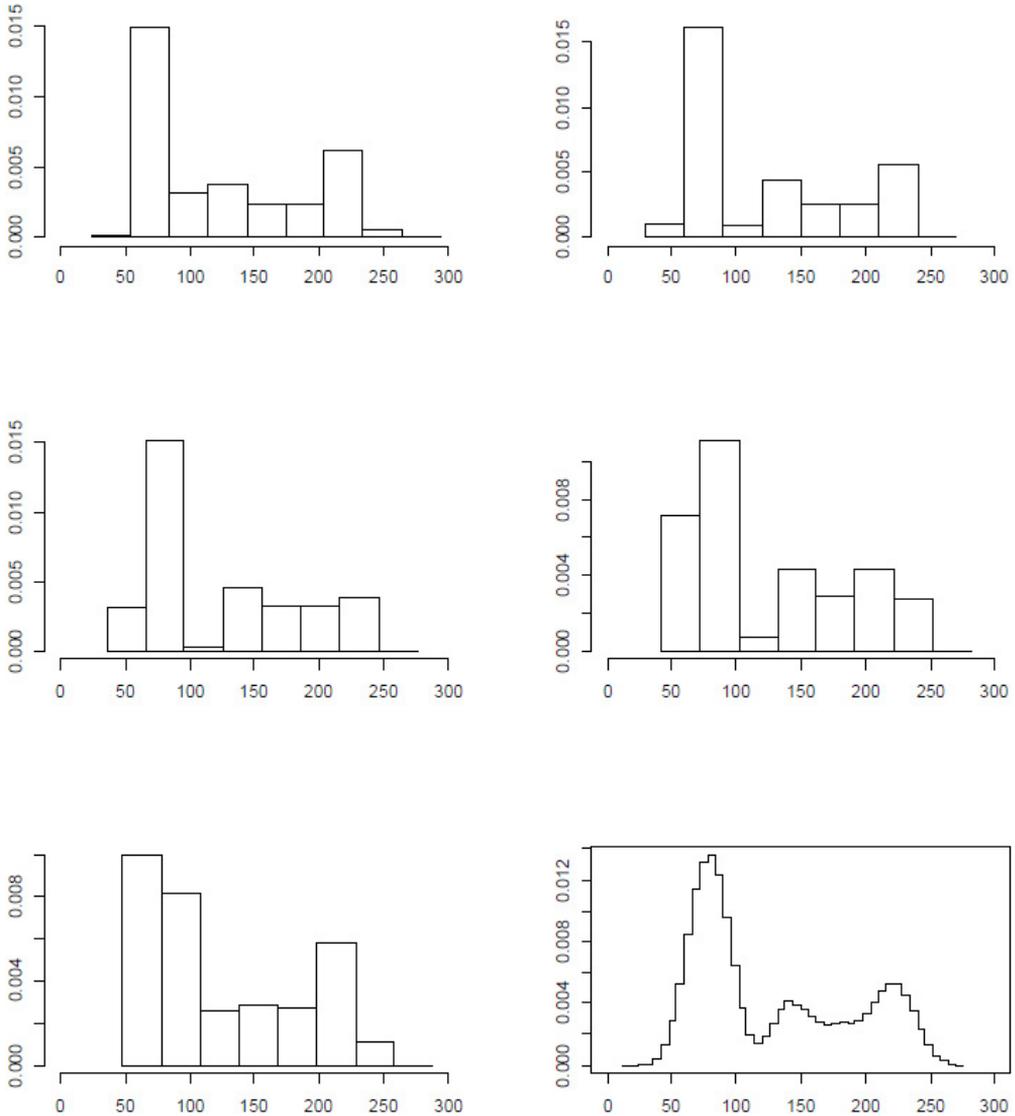


Figura 1.2. Cinco histogramas con origen diferente y el HDP correspondiente para los datos de longitud de hembras e indeterminados del bagre *Cathorops melanopus* (Versión R).



2. Dependencia en la amplitud y número de intervalos. Estos valores relacionados entre sí, se pueden considerar como parámetros que determinan la suavidad de la distribución de frecuencia (densidad). Al utilizar pocos intervalos se eliminan detalles de la distribución, mientras que numerosas clases resultan en estimaciones ruidosas. Es frecuente que el número y la amplitud de los intervalos se determine arbitrariamente a pesar de lo importante que es esta elección (valores de amplitud comúnmente utilizados son 1 y 5, cm o mm). Como ejemplo, considerar las Figuras 1.3 y 1.4 para los datos del bagre estuarino construidas mediante los comandos siguientes:

```
. histogram bodlen if sexa==2 | sexa==3, bin(5) ytitle("Densidad")
xtitle("Longitud estándar (mm)")
```

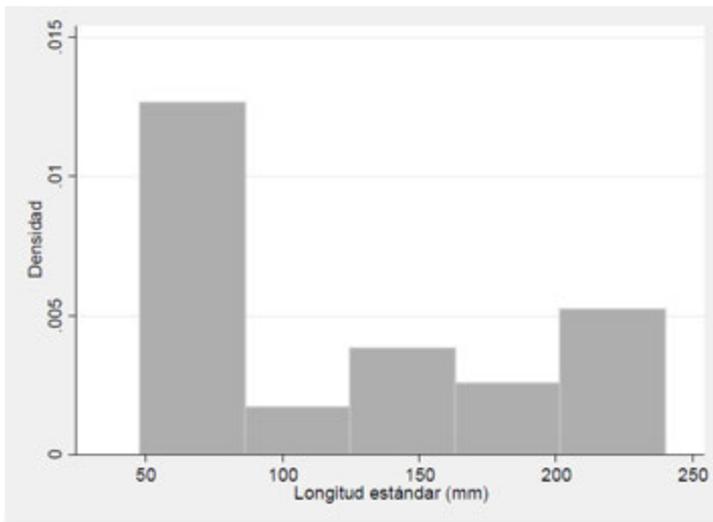


Figura 1.3. Histograma con cinco intervalos para los datos de longitud del bagre estuarino *Cathorops melanopus* (hembras e indeterminados).

```
. histogram bodlen if sexa==2 | sexa==3, bin(20) ytitle(Densidad)
xtitle("Longitud est      ándar (mm)")
```

El primer histograma con cinco intervalos muestra una distribución trimodal muy simplificada. El segundo, con 20 clases, despliega una distribución claramente multimodal con 4 modas.



3. Discontinuidad. Las discontinuidades del histograma son una función de la localización arbitraria de los intervalos y la discretización de los datos más bien que de la población que está siendo muestreada. La densidad local sólo se calcula en el punto medio (marca de clase) de cada intervalo y entonces se dibujan las barras suponiendo una densidad constante en cada clase (Chambers *et al.*, 1983).
4. Amplitud fija de intervalo. Si las clases son lo suficientemente angostas como para capturar el detalle donde la densidad es alta, entonces pueden ser demasiado angostos para evitar ruido cuando la densidad es baja. Este problema se ataca a menudo variando la amplitud de clase, pero entonces, la altura de la barra no es proporcional a su área, lo que conduce interpretaciones incorrectas.

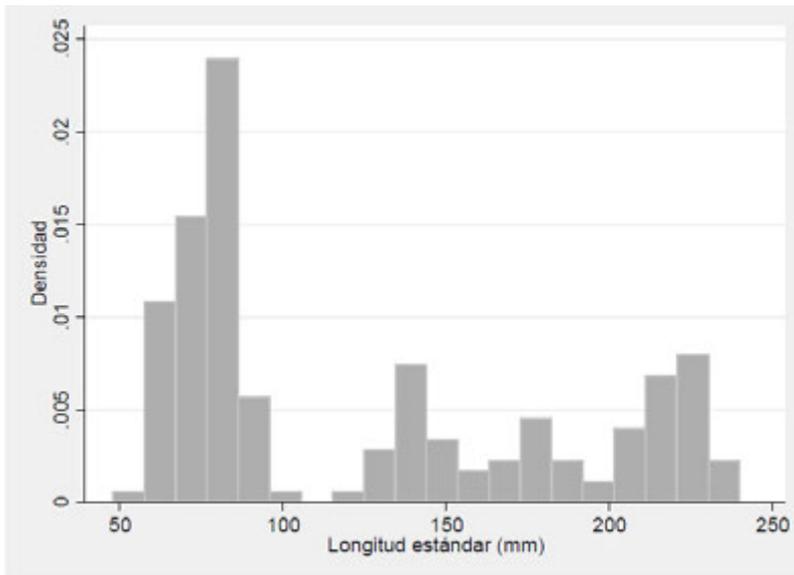


Figura 1.4. Histograma con 20 intervalos para los datos de longitud estándar del bagre estuarino *Cathorops melanopus*.



1.2 Polígonos de frecuencia

Como se señaló anteriormente, las discontinuidades del histograma limitan su utilidad como un estimador de densidad. El polígono de frecuencia (PF) es un estimador de densidad continuo derivado de la interpolación lineal de los centros de clase del histograma. Scott (1985a) examinó las propiedades teóricas de polígonos de frecuencia univariados y bivariados y encontró que poseen mejoras considerables sobre los histogramas.

Vale la pena mencionar que Fisher (1932, 1958) desaprobó a los PFs por razones gráficas al considerar que el PF conduce a una curva engañosa. De acuerdo con Él, se debe tener cuidado de distinguir a la población hipotética continua de tamaño infinito de la cual se ha sacado la muestra, de las observaciones finitas y discontinuas contenidas en dicha muestra. El trabajo de Scott ha probado que la objeción de Fisher hacia el uso de un estimador de densidad continuo no es válida, aunque debe remarcar su preocupación sobre el uso de técnicas que oscurecen el ruido estadístico con sofisticación matemática (Scott, 1992; 2015).

El polígono de frecuencia (PF) univariado es el interpolante lineal de los centros de clase de un histograma con intervalos de la misma amplitud. El PF se extiende más allá del histograma hacia un intervalo vacío a cada extremo y se ha verificado que es una función de densidad *bona fide* (es positivo con integral igual a la unidad). Considerando que el PF conecta dos valores adyacentes del histograma \hat{f}_0 y \hat{f}_1 , entre los centros de clase, el PF es descrito por la ecuación:

$$\hat{f}(x) = \left(\frac{1}{2} - \frac{x}{h}\right)\hat{f}_0 + \left(\frac{1}{2} + \frac{x}{h}\right)\hat{f}_1, \quad -\frac{h}{2} \leq x < \frac{h}{2}$$

La aleatoriedad en el polígono de frecuencia deriva enteramente de los intervalos del histograma.

Del trabajo de Scott (1985a) puede afirmarse que el PF en contraste con el histograma:

- aproxima mejor a densidades continuas por medio de interpolación lineal de intervalos más amplios.
- es menos eficiente cuando la densidad subyacente es discontinua.



- es más sensitiva con respecto a errores en la elección de la amplitud de intervalo, particularmente cuando $h < h_0$.
- requiere un error mayor en la amplitud de intervalo antes de que su error cuadrado medio integrado (ECMI) sea peor que el mejor ECMI de histograma.
- es más eficiente para los datos en relación al histograma al incrececentar el tamaño de muestra.

Como ejemplo se incluyen dos polígonos de frecuencia para los datos de longitud patrón de bagres. La Figura 1.5 muestra el *PF* con amplitud de intervalo de 9.6 y origen en 48 como en el histograma de la Figura 1.3.

Primero se generan los pares de densidad-marca de clase

```
. twoway__histogram_gen bodlen if sexa==2 | sexa==3, bin(20) gen(h9p6 m9p6)
```

Luego se extienden las marcas de clase para incluir la densidad igual a cero

```
. di 52.8-9.6  
43.2
```

```
. di 235.2+9.6  
244.8
```

```
. replace m9p6 = 43.2 in 20  
(1 real change made)
```

```
. replace h9p6 = 0 in 20  
(1 real change made)
```

```
. replace m9p6 = 244.8 in 21  
(1 real change made)
```

```
. replace h9p6 = 0 in 21  
(1 real change made)
```



Y se repite la gráfica pero con rótulos adecuados

```
. scatter h9p6 m9p6, c(1) sort ytitle(Densidad) xtitle("Longitud estándar (mm) ")
```

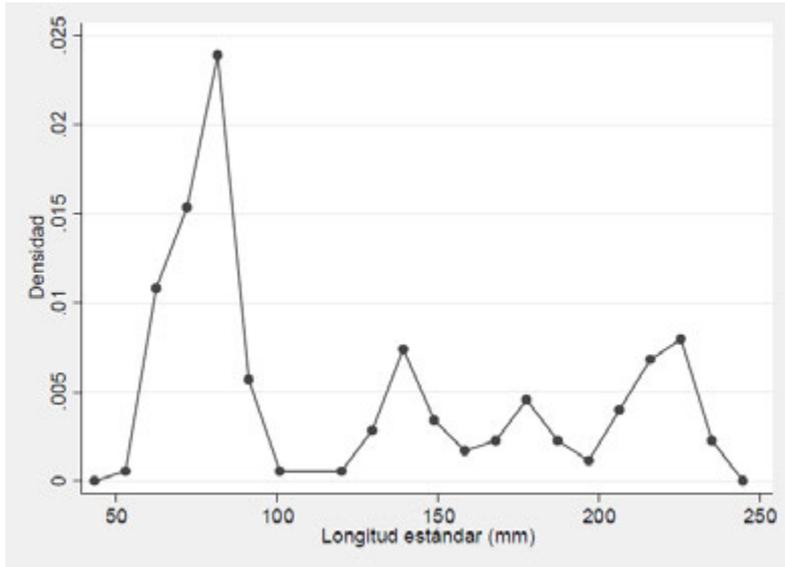


Figura 1.5. Polígono de frecuencia para los datos de longitud patrón de bagres. La amplitud de intervalo es 9.6 (como el histograma de la Figura 1.3).

El PF de la Figura 1.6 tiene una amplitud de intervalo de 45 mm con un origen en 48 mm, semejante al del histograma de la Figura 1.2.



```
. twoway__histogram_gen bodlen if sexa==2 | sexa==3, width(45) gen(h45 m45)

. l h45 m45

      +-----+
      |      h45      m45 |
      |-----|
1.    | .011779      70.5 |
2.    | .0018215    115.5 |
3.    | .0034001    160.5 |
4.    | .004493     205.5 |
5.    | .0007286    250.5 |

. di 70.5-45
25.5

. di 250.5+45
295.5

. replace m45 = 25.5 in 6
(1 real change made)

. replace h45 = 0 in 6
(1 real change made)

. replace m45 = 295.5 in 7
(1 real change made)

. replace h45 = 0 in 7
(1 real change made)

. scatter h45 m45, c(1) sort ytitle(Densidad) xtitle("Longitud estándar
(mm)") xlab(0(50)300)
```

La suavidad de estas estimaciones en comparación con los histogramas de las Figuras 1.3 – 1.4 es evidente. Más ejemplos se describirán en secciones posteriores.

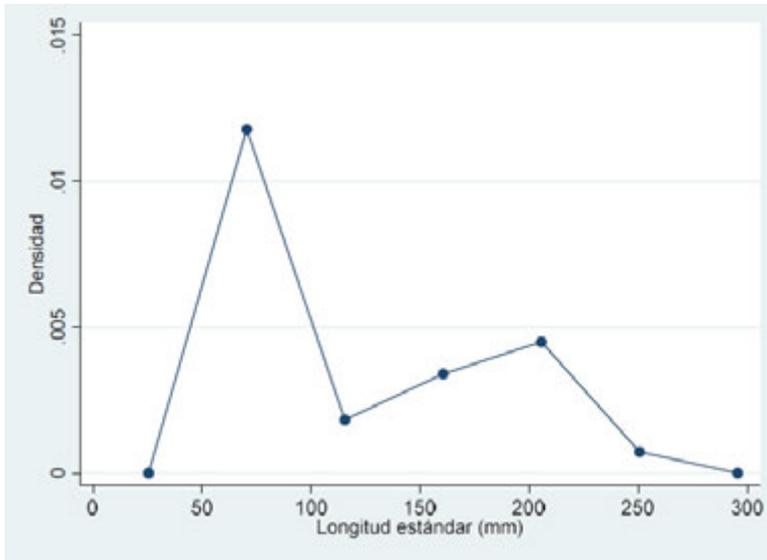


Figura 1.6. Polígono de frecuencia para los datos de longitud patrón de bagres. La amplitud de intervalo es 45 (como el histograma de la Figura 1.2)

1.3 Estimadores de densidad por kernel (EDKs)

Se han propuesto varios procedimientos para superar los inconvenientes de los histogramas. Los problemas del origen y la discontinuidad son atacados al calcular la densidad local en cada punto de los datos. Esto se logra, en esencia, por la construcción de intervalos de amplitud fija alrededor de cada punto y no en las marcas de clase.

Desde un punto de vista formal, este intervalo (llamado también ancho de ventana o amplitud de banda) puede considerarse como una función ponderal que asigna un peso positivo a cada observación dentro del intervalo y un peso nulo para los datos que caen fuera de él.

En el histograma tradicional, por ejemplo, el peso es un valor constante (función uniforme) en la forma de frecuencia, fracción, porcentaje o densidad, asignado a cada punto de los datos incluido en el intervalo. Los valores individuales para la clase se suman y el resultado puede ser representado geoméricamente por un rectángulo (la barra clásica) centrado en cada marca de clase.



En el procedimiento alternativo, para estimar la densidad, la discontinuidad se ataca al considerar funciones ponderales de variación gradual (como por ejemplo una curva Gaussiana). De esta forma es posible emplear una figura en forma de campana centrada en cada valor de los datos y sumar estas curvas individuales para obtener el resultado final (para detalles consultar a Chambers, *et al.*, 1983; Fox, 1990; Härdle, 1991; Salgado-Ugarte, 2002; Everitt y Hothorn, 2006).

Estas nociones conducen a los estimadores de densidad por kernel (EDKs), atribuidos a Rosenblat (1956) definidos por la siguiente expresión:

$$\hat{f}(x) = \frac{1}{hn} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (1.1)$$

donde $\hat{f}(x)$ es la estimación de densidad de la variable x , X_i indica una observación particular de x , n es el número de observaciones, h es la amplitud de banda (intervalo) o parámetro de suavización y $K(\bullet)$ es la función kernel, suave, simétrica con integral igual a la unidad.

El Cuadro 1.1, adaptado de Härdle (1991), enlista algunas funciones kernel más comunes. Con excepción del kernel uniforme, todas estas funciones varían gradualmente con una altura máxima en el centro y valores mínimos hacia los extremos (figura 1.7). Notar que el sostén de las funciones es de -1 a 1 con excepción del kernel gaussiano que teóricamente se calcula desde $-\infty$ a $+\infty$, aunque en la práctica los algoritmos se limiten a un intervalo menor (como de -4 a 4).

Cuadro 1.1. Algunas funciones kernel comunes.

Kernel	K(z)
Uniforme	$\frac{1}{2} I(z \leq 1)$
Triangular (ASH)	$(1 - z) I(z \leq 1)$
Epanechnikov	$\frac{3}{4}(1 - z^2) I(z \leq 1)$
Biponderado (Cuártico)	$(15/16)(1 - z^2)^2 I(z \leq 1)$
Triponderado	$(35/32)(1 - z^2)^3 I(z \leq 1)$
Coseno	$(\pi/4)\cos((\pi/2)z) I(z \leq 1)$
Gaussiano	$(1/\sqrt{2\pi})\exp(-1/2)z^2$

En todos los casos $z = (x - X_i) / h$.

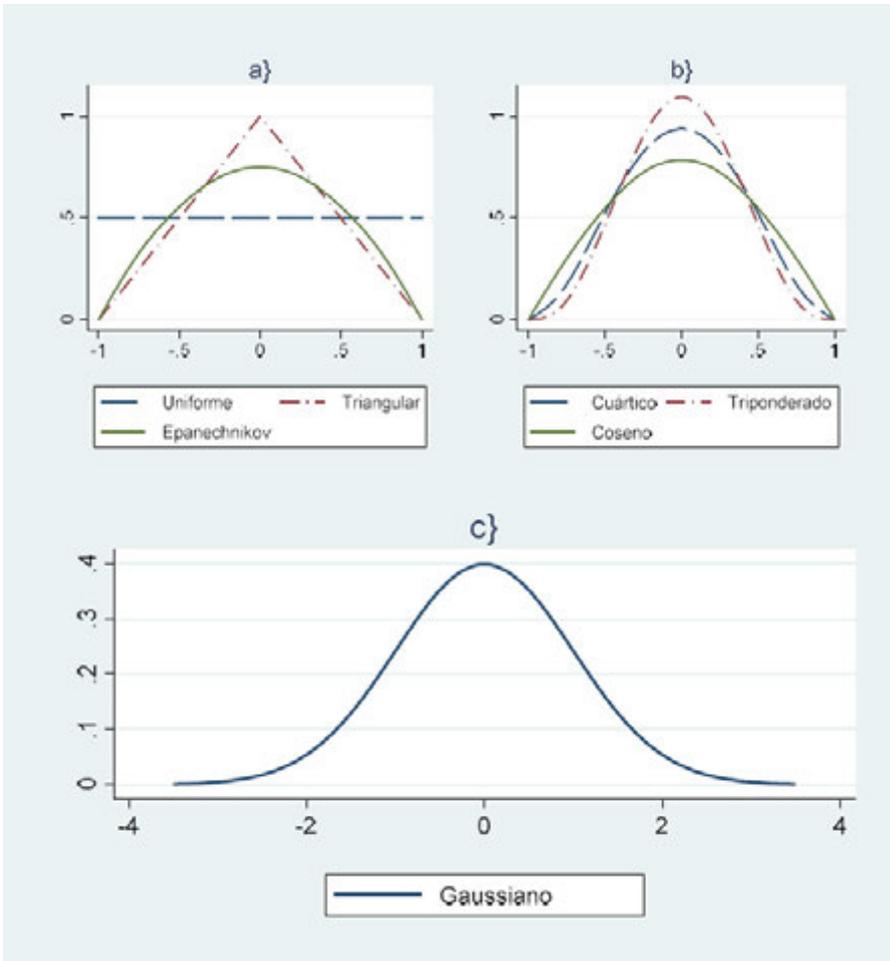


Figura 1.7. Funciones ponderales (kerneles) comunes (Salgado-Ugarte, 2013; 2017).

El resultado de aplicar un EDK gaussiano con ancho de banda de 9.5 (este valor se discutirá más adelante), a los datos de longitud de bagres se presenta en la Figura 1.8.



El código en Stata es:

```
. warpdenml bodlen if sexa==2 | sexa==3, b(9.5) m(10) k(6) t1title("Densidad  
WARP (poligonal), bw = 9.5, M = 10, ker = 6") ytitle(Densidad) xtitle("Longitud  
patrón (mm)")
```

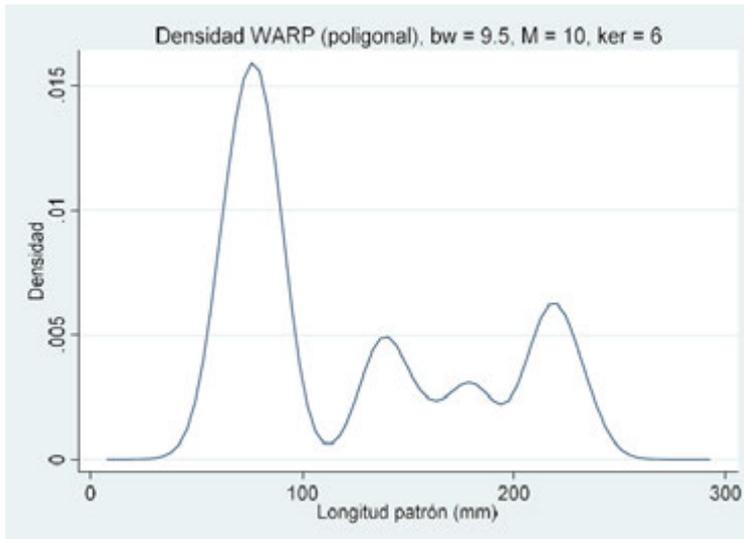


Figura 1.8. Estimador de densidad por kernel (EDK) para los datos de longitud patrón de bagres (hembras e indeterminados). El ancho de banda es 9.5 (semejante al ancho de intervalo del histograma de la Figura 1.2).

Para hacerlo en R primero se importan los datos desde el archivo en formato de datos de Stata:

```
> setwd("c:/data")  
> library(foreign)  
> bagtalla <- read.dta("bagretallaso.dta")  
> names(bagtalla)  
[1] "bodlen" "sex" "nalea" "sexa"  
> length(bagtalla$bodlen)  
[1] 641
```



Posteriormente se genera un objeto con los datos de longitud de hembras y la muestra (aleatoria) de 100 individuos indeterminados. Estos valores se han especificado por la variable “sexa” (breve de “sexo ajustado”, es decir solo tiene valor para los 100 ejemplares indeterminados seleccionados al azar):

```
> blhemin <- bagtalla$bodlen[bagtalla$sexa==2 | bagtalla$sexa==3]
```

y se construye el EDK con la función “density” incluida en R. Es necesario aplicar la función “na.omit” ya que la selección de hembras e indeterminados produce valores faltantes (“NA”):

```
> plot(density(na.omit(blhemin),bw=9.5),main="EDK, bw=9.5", ylab="Densidad",
xlab="Longitud patrón (mm)")
```

El resultado se muestra en la Figura 1.9.

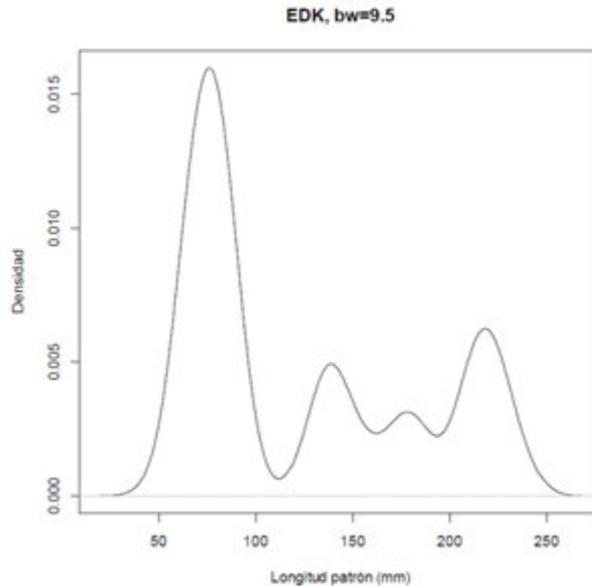


Figura 1.9. Estimador de densidad por kernel (EDK) para los datos de longitud patrón de bagres (hembras e indeterminados) obtenido con la función “density” de R. El ancho de banda es 9.5 (semejante al ancho de intervalo del histograma de la Figura 1.2).



Estos EDK emplean amplitudes fijas de banda. Esta característica hace que las estimaciones sean sensibles a ruido en las colas o en cualquier otro intervalo de la distribución con frecuencia baja.

Para atacar este inconveniente, existen procedimientos que reducen la amplitud de banda en las regiones con alta concentración de datos y la incrementa donde la concentración es baja. Es así que estos EDKs de amplitud de banda variable (Jones, 1990) proporcionan detalle donde se concentran las observaciones y eliminan fluctuaciones ruidosas en donde los datos son escasos.

1.4 Estimadores de densidad por kernel de amplitud variable

Los estimadores de la densidad por kernel descritos en las secciones anteriores utilizan amplitud fija de ventana. Esta característica provoca que las estimaciones sean vulnerables al ruido (variaciones aleatorias) en las colas o en cualquier intervalo con cuenta baja de observaciones en la distribución.

Para resolver esta dificultad se han propuesto algunos métodos que utilizan una amplitud de banda variable. La descrita a continuación se tomó de Fox (1990) quien a su vez la adaptó de Silverman (1986). La idea general es el ajustar la amplitud de ventana de tal forma que sea más angosta a densidades altas y más amplia donde existen bajas densidades. El resultado de este procedimiento de amplitud variable es retener detalle donde las observaciones se concentran y eliminar fluctuaciones ruidosas donde los datos escasean.

El algoritmo para calcular tales estimadores por kernel de amplitud variable se explica a continuación:

1. Calcular una estimación de densidad preliminar, por ejemplo la proporcionada por cualquiera de las funciones kernel con amplitud fija de ventana $f(x)$.
2. Con esta estimación inicial para cada observación de X , calcular factores locales para las ventanas w_i relacionados inversamente con la densidad:

$$w = \left[\frac{\tilde{f}_g}{\hat{f}_k(X_i)} \right]^{1/2}$$



donde

$$\tilde{f}_g = \left[\prod_{(i=1)}^n \hat{f}_K(X_i) \right]^{1/n}$$

es la media geométrica de $f(X_i)$, y por tanto los pesos tienen un producto y media geométrica de uno.

- Utilizando los pesos, calcular el estimador de densidad por kernel de amplitud variable.

$$\tilde{f}_A(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{w_i} K \left[\frac{x - X_i}{w_i h} \right]$$

Notar que la aplicación de pesos a h produce ventanas de amplitud variable que resultan en una amplitud de banda media geométrica igual a h . El factor $1/w$ se requiere para asegurar que el área total bajo la estimación de densidad sea la unidad.

- Iterar (repetir) los pasos 2 y 3, utilizando \tilde{f}_A en lugar de \tilde{f}_K para recalcular los pesos. En la práctica, la iteración produce poco cambio en las estimaciones de densidad (Fox, 1990).

La aplicación de este procedimiento se presenta en la Figura 1.7. Se puede notar que la moda dominante (con mayor frecuencia) está mejor definida y que la zona intermedia con pocos valores se encuentra algo aplanada.

Para Stata se escribieron originalmente dos programas: “varwiker.ado” y “varwike2.ado”, los cuales se han actualizado a los programas “varwiker1.ado” y “varwike21.ado”. El primero lleva a cabo el algoritmo de estimación para todos los datos en el primer y segundo pasos del método de cálculo, por lo que, si el número de datos es grande, la obtención de los resultados finales llevará más tiempo. El segundo programa, estima la densidad para cada dato en el primer paso, pero en el segundo sólo utiliza una serie de puntos uniformemente espaciados; el valor preestablecido es de 50, pero puede aumentarse a 100 o más puntos de cálculo. Para el ejemplo se utilizó éste último (“varwike21.ado”):

```
. varwike21 bodlen if sexa==2 | sexa==3, b(9.5) np(100) ytitle(Densidad)
xtitle("Longitud patrón (mm)") tlttitle("EDK de banda variable, bw(G) = 9.5, np
= 100")
```

(Figura 1.10)

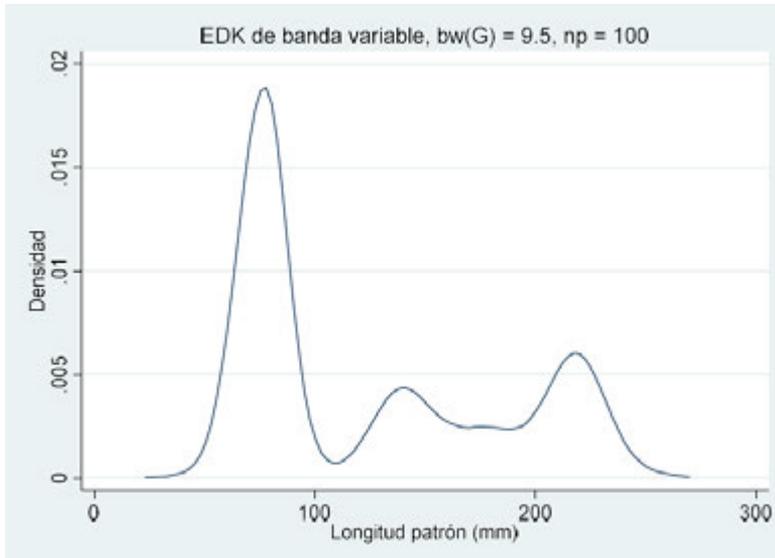


Figura 1.10. Estimador de densidad por kernel (EDK) con amplitud de banda variable para los datos de longitud patrón de bagres (hembras e indeterminados) obtenido con el comando “varwike21.ado” para Stata. El ancho de banda es 9.5 (semejante al ancho de intervalo del histograma de la Figura 1.2).

Debemos recordar que, si la amplitud de banda varía de acuerdo con la frecuencia de los datos, entonces, para el cálculo de la densidad se usan diferentes valores. Por esto, el ancho de banda inicial, como se indicó en la explicación de los pasos de cálculo, representa la media geométrica (G) de todos los valores utilizados para la estimación.

De acuerdo con Simonoff (1996), este método está cerca de ser lo mejor para la estimación de la distribución de frecuencia de datos univariados. Detalles y programas para calcular estos importantes estimadores se incluyen en Salgado-Ugarte et al. (1993), Salgado-Ugarte (2002) y Salgado-Ugarte y Pérez-Hernández (2003).



1.5 Elección de amplitud de banda

Si bien arriba se citan alternativas para resolver dos de los inconvenientes del histograma, el problema de la elección de la amplitud de intervalo (banda, h) permanece. Un enfoque para la elección del ancho de banda sugerido por Tarter y Kronmal (1976) es el variar h hasta obtener un resultado suave “satisfactorio”. Este procedimiento depende de la evaluación subjetiva del investigador, pero puede ser adecuado para fines exploratorios (Silverman, 1986) puesto que las características de la densidad “aparecen” y “desaparecen” al cambiar la amplitud de banda (Silverman, 1981a).

La teoría estadística proporciona alguna guía en la selección de una amplitud de banda óptima. Desafortunadamente, por lo general no es posible optimizar la amplitud de banda sin el conocimiento previo de la forma de la verdadera distribución de densidad. Siguiendo a Tukey (1977), Scott (1979) y Silverman (1978, 1986), la distribución Gaussiana puede emplearse como una referencia estándar al escoger h . Aplicando un kernel Gaussiano y minimizando el error cuadrado integrado medio (ECIM), la siguiente estimación de escala (dispersión ó variabilidad) puede calcularse:

$$s = \min \left[\left(\frac{\sum (x_i - \bar{x})^2}{n-1} \right)^{1/2}, \frac{\text{Dispersión de cuartos}}{1.349} \right] \quad (1.2)$$

donde la Dispersión de cuartos se refiere al cuarto superior menos el cuarto inferior, una medida resistente de dispersión aproximadamente equivalente al recorrido intercuartílico (Tukey, 1977). Entonces h puede escogerse como:

$$h = \frac{0.9s}{n^{1/5}} \quad (1.3)$$

donde s es la menor de dos estimaciones del parámetro de dispersión de la distribución Gaussiana σ : la desviación estándar y la F-pseudosigma (Dispersión de cuartos/1.349) (Hoaglin, 1983; Iglewics, 1983, Fox, 1990; Salgado-Ugarte, 2016). Este ajuste proporciona resistencia a colas densas y trabajará bien para una amplia gama de densidades pero, tiende a sobresuavizar distribuciones muy sesgadas o multimodales (Silverman, 1986). En este caso, la amplitud “óptima” de banda puede considerarse como un punto de partida para un ajuste fino posterior.



1.6 Cálculo eficiente de EDKs (discretizado y ASH-WARP)

A pesar de que los EDKs eliminan algunos de los problemas del tradicional histograma, estos estimadores presentan un inconveniente: para su determinación se requiere un gran número de cálculos.

Este inconveniente es atenuado por la disponibilidad de las computadoras, las cuales, precisamente son capaces de realizar numerosas operaciones algebraicas sin errores en un tiempo muy corto.

No obstante, la aplicación directa de la fórmula de los EDKs es ineficiente, ya que en el caso de observaciones con el mismo valor se llevan a cabo, de manera repetida, los numerosos cálculos para llegar al mismo resultado. Para evitar el hacer estos cálculos repetitivos al trabajar con lotes numéricos grandes, se han sugerido los procedimientos “discretizados” que involucran la especificación de un conjunto de valores uniformemente espaciados (por ejemplo 50) a lo largo del recorrido de los números y utilizarlos para el cálculo de las densidades (Chambers, *et al.* 1983; Fox, 1990; Salgado-Ugarte, 2002).

Scott (1985) sugirió un procedimiento alternativo para superar este problema y a su vez eliminar la influencia de la elección del origen: él propuso promediar varios histogramas con orígenes sistemáticamente diferentes con la misma amplitud de clase en lugar de escoger alguno de ellos.

Estos estimadores son los histogramas desplazados promedio (ASH por sus siglas en inglés: **A**veraged **S**hifted **H**istograms). Posteriormente, Härdle y Scott (1988), desarrollaron el concepto más general llamado promedio ponderado de puntos redondeados, que por sus siglas en inglés se conocen como **WARP** (**W**eighted **A**veraging of **R**ounded **P**oints).

El cálculo de una estimación **ASH-WARP** requiere de tres pasos:

- (1) Agrupamiento de los datos;
- (2) Cálculo de los pesos y
- (3) Ponderación de los intervalos.

La última gráfica de la Figura 1.1 muestra el resultado de promediar los cinco histogramas incluidos; es evidente la trimodalidad de los datos como resultado de una mejora significativa en la relación señal – ruido obtenida al promediar el origen.



El método ASH-WARP puede usarse para aproximar a un EDK particular utilizando la función ponderal adecuada. El WARP aproxima a la función kernel al aumentar el número de histogramas que se promedian (Härdle, 1991).

El siguiente código permite el cálculo de una estimación de densidad por kernel gaussiano discretizada. Se usa la mitad de la amplitud de banda “óptima” dada por la expresión 1.3): $19/2 = 9.5$ (lo que se explicará más adelante), la función ponderal gaussiana y un número de puntos uniformemente espaciados (100) para el cálculo. El resultado se muestra en la Figura 1.11 en la cual los 100 puntos usados para la estimación se han indicado explícitamente (símbolos circulares).

```
. kerneld bodlen if sexa==2 | sexa==3, b(9.5) k(6) np(100) ylab xlab t1("EDK
discretizado, bw = 9.5, k = 6, no. puntos = 100") s(o) l1("Densidad")
b2("Longitud patrón (mm)")
```

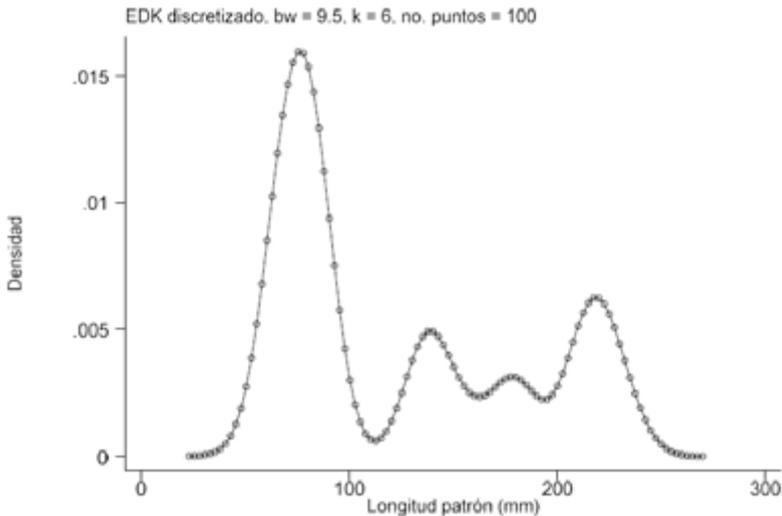


Figura 1.11. EDK discretizado para los datos de longitud patrón de bagres (hembras e indeterminados) con los 100 puntos de cálculo indicados.



La estimación ASH-WARP para estos datos se presenta en las Figuras 1.12 y 1.13. En la primera se presenta el histograma desplazado promedio con la mitad de la banda óptima. Debe señalarse que al promediar a los histogramas, la función ponderal uniforme de los histogramas se hace equivalente a una función triangular. Además, para la unión de los puntos resultantes se hace uso de la opción “step” que permite dibujar la versión discontinua para presentar el resultado de promediar cinco histogramas. Detalles adicionales del cálculo de este valor óptimo para el kernel triangular y opciones adicionales se presentarán en secciones posteriores (1.8 y 1.9).

En la segunda (Figura 1.13) por otro lado, se muestra el resultado de aumentar el número de histogramas que se promedian. Utilizando la amplitud de banda de 9.5 (mitad de la “óptima” de Silverman dada por la expresión 1.3) = 19) y resolviendo para 20 histogramas obtenemos una estimación prácticamente igual a la versión discretizada de la Figura 1.12). Teóricamente, cuando el número de histogramas a promediar (representados por “M”) es infinito, la estimación ASH (WARP) es equivalente al EDK correspondiente.

```
. warpdenml bodlen if sexa==2 | sexa==3, b(23.11) m(5) k(2) st
ytitle("Densidad") xtitle("Longitud patrón (mm)") t1title("Densidad ASH, bw =
23.11, M = 5, Ker = 2)
```

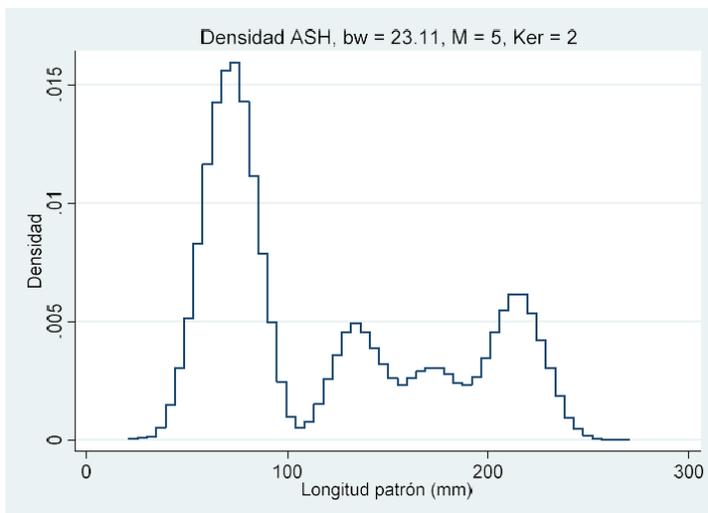


Figura 1.12. Histograma desplazado promedio (ASH) para los datos de bagre. El ancho de banda se ha transformado para ser la mitad del óptimo (46.22/2) para kernel Triangular.



1.7 Algunas consideraciones

El uso de la frecuencia de tallas para investigar el crecimiento de los organismos capturados por pesca, data desde finales del siglo antepasado cuando Petersen (1892) aplicó el (en ese entonces recientemente aparecido) histograma, a las medidas de longitud de peces y encontró que, debido a la reproducción anual de las especies de zonas templadas, la distribución resultante era multimodal y que cada una de las modas representaba un grupo de organismos con tamaño (y por tanto edad) semejante, es decir una “cohorte”.

Las cohortes pueden seguirse claramente durante la primera parte de la vida del organismo conectando las modas correspondientes en los histogramas de muestras sucesivas. Sin embargo, cuando los organismos se aproximan a su talla máxima el seguimiento de las modas se hace difícil (si no imposible) porque para entonces peces de diferentes edades han alcanzado aproximadamente el mismo tamaño (Sparre y Venema, 1992).

```
. warpdnml bodlen if sexa==2 | sexa==3, b(9.5) m(20) k(6) st
yttitle("Densidad") xttitle("Longitud patrón (mm)") tlttitle("Densidad ASH, bw =
9.5, M = 20, Ker = 6)
```

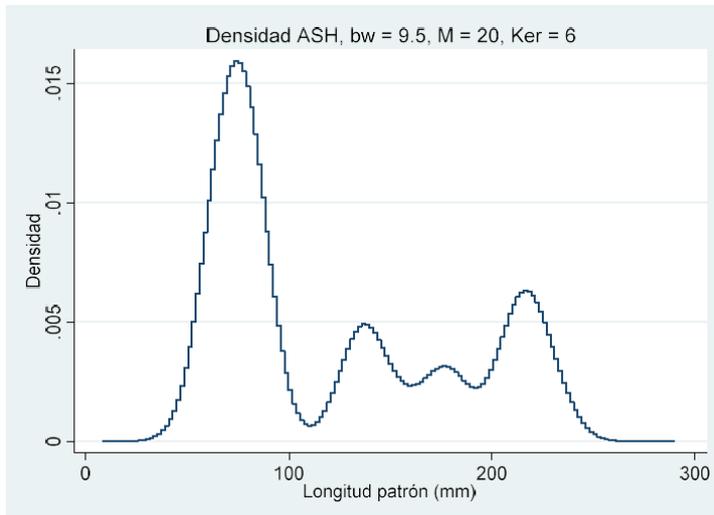


Figura 1.13. Histograma desplazado promedio (ASH) para los datos de bagre. El ancho de banda es la mitad del óptimo gaussiano y se han promediado 20 histogramas con la función ponderal Gaussiana.



El procedimiento estadístico para mostrar la distribución de las longitudes (histograma) es un suavizador de datos con la amplitud del intervalo como el parámetro de suavización (Härdle, 1991). El número de modas es una característica de los datos pero depende de la amplitud del intervalo.

De esta forma, en la investigación pesquera una pregunta fundamental es “¿Cuál es la amplitud de intervalo que manifiesta claramente la estructura modal de los datos?” A este respecto, podemos recordar varias sugerencias de la literatura.

Se usan en general intervalos de 0.5 cm para especies pequeñas (< 30 cm) y 1.0 ó 2.0 cm para especies mayores. Wolff (1989) propuso una formula derivada empíricamente para la determinación del ancho óptimo de intervalo con base en el tamaño máximo observado y el número estimado de clases de edad en la muestra.

Se ha notado que en el estudio de pesquerías se tiene que hacer a menudo un compromiso entre medir lenta y cuidadosamente un número reducido de individuos o agrupar números mayores de medidas en intervalos mayores; en este sentido se ha sugerido que la amplitud del intervalo debería ser lo suficientemente angosto para que los picos de frecuencia sucesivos queden separados por cinco o seis clases (Caddy, 1986).

Erzini (1990) argumenta que el tamaño óptimo de intervalo para el agrupamiento de datos de longitud es una función del tamaño de muestra y que las características biológicas tales como la variabilidad de la longitud a cierta edad, el patrón de reclutamiento, la tasa de crecimiento y el tamaño máximo, afectan a la claridad con la que se pueden distinguir las modas en la distribución de tallas. El artículo de Erzini apoya la sugerencia de Caddy citada arriba y además encontró que los métodos con base empírica para determinar la amplitud de intervalo pueden servir sólo para proporcionar estimaciones gruesas.

Es sorprendente que el efecto del origen de los intervalos en los histogramas resultantes no haya recibido atención en la literatura de investigación pesquera.

El ASH da una versión más detallada de la distribución de frecuencia (Figura 1.13). Como se señaló arriba, este estimador, sobre todo en su versión interpolante, converge a un EDK al incrementarse el número de histogramas desplazados que se promedian y con el uso de la función ponderal adecuada. En la práctica, es suficiente promediar entre 10 y 15 histogramas desplazados para llegar a los mismos resultados que un EDK (Härdle, 1991). Esta propiedad se utiliza para ahorrar un número considerable de operaciones algebraicas individuales en el cálculo de los estimadores de densidad por kernel.



Los EDKs resuelven algunos de los problemas del histograma y son adecuados para el análisis de datos de frecuencia de longitudes. Todas las funciones kernel incluidas en el Cuadro 1.1 tienen una eficiencia cercana a la máxima del kernel de Epanechnikov (1969) (Silverman, 1986; Salgado-Ugarte, 2002). Como consecuencia de esto, la función kernel puede escogerse con base en otras razones, como por ejemplo, el esfuerzo de cómputo requerido para su cálculo (Silverman, 1986).

Al trabajar con estimadores de densidad por kernel, el efecto del origen en la estimación desaparece, aunque el problema de la elección del parámetro de suavización (amplitud de banda) permanece. No obstante, se dispone de varias guías estadísticas al respecto. La regla para la selección de amplitud de banda presentada anteriormente se ha desarrollado para el caso de una sola distribución subyacente.

En distribuciones mezcladas, existen varios componentes (Gaussianos o de otro tipo), cada uno con parámetros diferentes (tales como la media y la desviación estándar en el caso de distribuciones Gaussianas). El número o amplitud de intervalo óptimos puede ser diferente para cada componente. Los grupos dominantes (con muchos individuos) permiten el uso de un número grande de intervalos angostos; componentes con menos individuos pueden soportar solamente unos pocos intervalos amplios.

El histograma clásico utiliza intervalos de amplitud fija, de aquí que en algunas instancias no puede presentar adecuadamente componentes dominantes y menores al mismo tiempo. Bajo estas circunstancias parece particularmente apropiado el uso de los EDKs de amplitud de banda variable, los cuales ajustan el intervalo para proporcionar detalle (decreciendo h en regiones de alta concentración de datos) y eliminar el ruido (incrementando h donde la cantidad de observaciones es escasa).

Sin embargo, como con los histogramas, es importante medir las longitudes con la precisión más alta posible para poder utilizar una gama amplia de amplitudes de intervalo.

Los EDKs permiten la aplicación de métodos para la determinación de componentes individuales en distribuciones mezcladas (Cassie, 1954; Tanaka, 1962; Morrison Cassie, 1962; Bhattacharya, 1963) de manera más clara y sencilla (ver por ejemplo, Salgado-Ugarte, *et al.* 1994; Salgado-Ugarte, *et al.* 2005). Asimismo, proporcionan valores iniciales adecuados para los métodos iterativos de ajuste de componentes por medio de criterios de máxima verosimilitud (Akamine, 1985; Macdonald y Green, 1988; Fournier, *et al.*, 1990).

Finalmente, los EDKs proporcionan varias formas de probar y evaluar la multimodalidad (para detalles ver Silverman, 1981b, 1983, 1986). Un ejemplo utilizando el método de



bootstrap suavizado de Silverman (1981b) se presenta en Salgado-Ugarte *et al.* (1997) y más recientemente en Salgado-Ugarte (2002). El uso de este método se especificará en la sección 1.9 más adelante. Otros enfoques se han sugerido por Cox (1966), Good y Gaskins (1980) y Wong (1985). Los programas para el cálculo de los EDKs se han incluido en los programas EDK2000, adjunto con el texto de Salgado-Ugarte (2002), con el programa AED2010, incluido con Salgado-Ugarte (2013) y la versión actualizada AED2020 incluido en la presente obra.

1.8 Reglas prácticas para selección de intervalo/banda

1.8.1. Introducción

La elección de la amplitud del intervalo/banda (parámetro de suavización) es uno de los problemas más relevantes de la estimación de la densidad. Como se discutió en las secciones anteriores, existen varios procedimientos para establecer un valor adecuado para este parámetro en histogramas, polígonos de frecuencia (PF), histogramas desplazados promedio (estimadores **ASH-WARP**) y estimadores de densidad por kernel (EDKs).

Algunos de estos métodos de selección se enfocan al número “óptimo” de intervalos, mientras que otros producen aproximaciones a la amplitud “óptima” de intervalo al minimizar alguna medida del error estadístico bajo ciertas condiciones y suposiciones.

En esta sección se presentan varias reglas que proporcionan algunos valores de referencia útiles para la elección del parámetro de suavización en el análisis de la densidad por histogramas, polígonos de frecuencia y estimadores de densidad por kernel, incluyendo a los histogramas desplazados promedio (**ASH**) y su generalización en el promedio ponderado de puntos redondeados (**WARP**).

1.8.2. Reglas para elección del número de intervalos en histogramas

Probablemente la regla más famosa para determinar el número de intervalos para la estimación de densidad del histograma fue propuesta por Sturges (1926). Al aplicar esta regla se espera que una variable normalmente distribuida pueda ser dividida apropiadamente de tal forma que las clases de frecuencia abarquen una serie binomial para todos los valores de N los cuales son potencias pares de dos (Doane, 1976).



Es interesante notar que no obstante en su contribución original Sturges se refiere explícitamente a la elección del intervalo de clase, técnicamente su regla es mas bien un procedimiento para escoger el número de intervalos (Scott, 1992; 2015).

De acuerdo a la sugerencia de Sturges, el número de intervalos puede calcularse como:

$$k = 1 + \log^2 n \quad (1.4)$$

donde k es el número de intervalos, y n se refiere al número de observaciones.

La formulación de Sturges es ampliamente recomendada en textos introductorios de estadística. Se ha vuelto una guía para los investigadores y a menudo es utilizada como valor pre-establecido en programas estadísticos aún cuando en varios casos no es apropiada: esta regla no es aplicable a datos con distribución asimétrica, no Gaussiana o multimodal, los cuales ocurren la mayor parte del tiempo (Doane, 1976; Scott, 1992; 2015).

Una forma de ajustar por sesgo es agregar intervalos en (1.4) de acuerdo a $1 + \hat{\gamma} \sqrt{(n/6)}$, donde $\hat{\gamma}$ es un estimador del coeficiente de sesgo estándar (Doane, 1976). Para fines exploratorios, Emerson y Hoaglin (1983) consideran que este ajuste adicional involucra muchos cálculos que podrían ser problemáticos sin una computadora. Una dificultad más seria es la no resistencia del coeficiente de sesgo.

1.8.3. Reglas para elección de amplitud de intervalo en histogramas

Como se ha citado en otros trabajos (Salgado-Ugarte, *et al.* 1993), Scott (1979) derivó una fórmula para calcular la amplitud óptima asintótica resultante en un error cuadrado integrado medio mínimo (ECIM) para histogramas. La fórmula de Scott involucra el conocimiento previo de la verdadera función de densidad, un evento más bien raro en el análisis de datos en la realidad. Por lo tanto, adoptando a la densidad Gaussiana como referencia, él propuso:

$$\hat{h} = 3.5 \hat{\sigma} n^{-1/3} \quad (1.5)$$

donde \hat{h} es la amplitud de banda estimada y $\hat{\sigma}$ es una estimación de la desviación estándar de los datos.



En su contribución original, Scott, además analizó el desempeño de esta regla cuando se aplica a tres distribuciones no Gaussianas adoptadas como modelos de referencia.

- Sesgada (densidad log-normal).
- Colas potentes (t de Student).
- Bimodal (mezcla de dos distribuciones Gaussianas).

A partir de las simulaciones realizadas, este autor concluyó que la regla de referencia Gaussiana produce un parámetro de suavización que:

- Sobresuaviza a una distribución log-normal, pero para índices de sesgo tan grandes como uno, la diferencia con la amplitud de intervalo óptima verdadera es menos del 30 %.
- Es insensible a curtosis moderada.
- Sobresuaviza datos bimodales cuando la distancia entre las modas es mayor que 2. Con datos claramente bimodales, esta regla no es adecuada (Scott, 1979).

Recientemente, este autor ha proporcionado factores de corrección para h tomando en cuenta al sesgo y la curtosis (Scott, 1992).

Una regla más robusta ha sido propuesta por Freedman y Diaconis (1981a,b), en la que se reemplaza la estimación de la desviación estándar por un múltiplo del recorrido intercuartílico (**RIC**). La regla F-D se define como:

$$\hat{h} = 2 (\text{RIC}) n^{-1/3} \quad (1.6)$$

Varios autores han comparado el desempeño de las reglas (1.4), (1.5) y (1.6) (Emerson y Hoaglin, 1983; Scott, 1992; 2015; ver también la nota técnica en el Manual de Referencia del paquete estadístico Stata, 1995; 2019) llegando a las siguientes generalizaciones:

- Las reglas de Scott y Sturges concuerdan cercanamente para tamaños de muestra entre 50 y 500.
- Para tamaños de muestra mayores, la regla de Sturges da muy pocos intervalos (produciendo una sobresuavización que desperdicia mucha de la información de los datos).



- En general, las densidades no Gaussianas requieren de más intervalos.
- La regla F-D resulta en intervalos más angostos (35% más intervalos que la regla de Scott).
- Desde un punto de vista exploratorio, la característica más interesante de las reglas de Scott y F-D es que ambas dependen principalmente de $n^{-1/3}$. Considerando el correspondiente número de intervalos, se comportaría como $n^{1/3}$, una forma funcional intermedia entre $\log(n)$ y $\sqrt{(n)}$.

1.8.4. Reglas sobreesuavizadas

Las reglas anteriormente descritas proporcionan un punto de partida simple y útil para su afinación posterior, si bien no son la respuesta definitiva al problema. Recientemente la búsqueda de procedimientos basados en los datos para minimizar el *ECIM* (o cantidades relacionadas como el error cuadrado integrado medio asintótico *ECIMA*) es el tema de considerables estudios. Los procedimientos descritos a continuación son algunos ejemplos de esta investigación (una explicación más amplia es dada por Scott, 1992; 2015).

Terrell y Scott (1985) mostraron que dependiendo del conocimiento de la escala de la densidad desconocida con base en los datos, existen límites superiores para la amplitud de intervalo de los histogramas. Por otra parte, en principio, no existe un límite inferior para h ya que la densidad desconocida puede ser arbitrariamente ruidosa. Considerando este límite superior estos autores arribaron a la siguiente expresión:

$$h_{os} \equiv \frac{b-a}{\sqrt[3]{2n}} \geq h_o \tag{1.7}$$

Donde h_o es la amplitud de intervalo óptima, h_{os} es el límite superior para h (la amplitud sobreesuavizada de intervalo) y $b-a$ es el recorrido de la muestra. Re-arreglando (1.7) es posible obtener una regla para número sobreesuavizado de intervalos:

$$\text{Número sobreesuavizado de intervalos} = \frac{b-a}{h_o} \geq \frac{b-a}{h_{os}} = \sqrt[3]{2n} \tag{1.8}$$



Eligiendo una amplitud de intervalo igual o mayor que h_{os} producirá una estimación sobreesuavizada así como utilizando un número de intervalos menor o igual al dado por la expresión (1.8).

Como concluyen Terrell y Scott (1985), las expresiones (1.7) y (1.8) dan resultados casi óptimos para una amplia variedad de densidades y proporcionan buenas estimaciones de densidad.

Estas reglas son muy simples y tienen valor mnemónico. Están encaminadas al hallazgo de un límite inferior para el número de intervalos. Sin embargo, en la estimación de la densidad el objetivo es encontrar al parámetro de suavización h_o más bien que el número de intervalos. Una solución en este contexto fue propuesta por Terrell (1990) en su *Principio de Suavización Máxima*. Restringiendo a la varianza como fija, su solución es:

$$h_{os} \text{ homoscedástica} \equiv 3.729 \sigma n^{-1/3} \geq h_o \quad (1.9)$$

La versión que utiliza al **RIC** es particularmente robusta, y por tanto especialmente útil en el caso de datos sesgados:

$$h_{os} \text{ homoscedástica robusta} \equiv 2.603 (RIC)n^{-1/3} \quad (1.10)$$

Las estimaciones conservativas sobreesuavizadas proporcionadas por estas reglas pueden conducir a evitar aseveraciones prematuras sobre la estructura en las muestras. Por el contrario, la ocurrencia de características estructurales en estimaciones conservadoras representa una evidencia sólida de su existencia. Si tales estructuras no ocurren en estimaciones sobreesuavizadas, entonces se requieren pruebas más especializadas o muestras adicionales para establecer su ocurrencia (Terrell, 1990).

1.8.5. Reglas para elección de número y amplitud de intervalos en polígonos de frecuencia

A pesar de comentarios iniciales (Fisher, 1932, 1958) desfavorables para el interpolante lineal de las marcas de clase de un histograma con amplitud uniforme de intervalo (es decir el polígono de frecuencia), el trabajo de Scott (1985a) acerca de las propiedades teóricas de polígonos de frecuencia (**PF**) uni y bivariados ha demostrado que produce estimaciones notablemente mejores en comparación con el histograma (Scott, 1992).



Los resultados mejorados producidos por el *PF* facilitan sustancialmente la localización del mínimo en el *ECIMA*, y por tanto en la estimación del número y amplitud de intervalo óptimas.

En contraste con el histograma puede afirmarse que el *PF*:

- aproxima mejor a densidades continuas mediante interpolación lineal con intervalos más amplios.
- es menos eficiente cuando la densidad subyacente es discontinua.
- es más sensitivo con respecto a los errores en la elección de amplitud de intervalo, particularmente cuando $h > h_0$; por otra parte, se requiere un error considerablemente grande en la amplitud de intervalo del *PF* antes de que su *ECIM* sea peor que el del mejor histograma (Scott, 1992).

Estas diferencias se reflejan en la resultante regla de referencia Gaussiana para el *PF*:

$$\text{Regla de referencia Gaussiana para } PF: \hat{h} = 2.15 \hat{\sigma} n^{-1/5} \quad (1.11)$$

La estimación de la desviación estándar puede ser una robusta, como *RIC*/1.349 (la Pseudosigma). Esta regla puede modificarse también tomando en cuenta factores de ajuste por sesgo y curtosis (Scott, 1992).

Como en el caso de los histogramas es posible definir límites inferiores para la amplitud de intervalo o límites superiores para el número de clases. La expresión para número sobresuavizado de intervalos para *PF* equivalente a la (1.8) de histogramas es:

Número sobresuavizado de intervalos para *PF*:

$$\frac{b-a}{h_0} \geq \left(\frac{147}{2} n \right)^{1/5} \quad (1.12)$$

Una versión diferente para el problema de sobresuavización conduce a la correspondiente regla para amplitud de intervalo:

$$h_{os} = 2.33 \hat{\sigma} n^{-1/5} \quad (1.13)$$



La diferencia entre la regla sobresuavizada (1.13) y la regla óptima Gaussiana (1.11) para PF es aproximadamente 8% lo que sugiere que la regla de referencia gaussiana sobresuaviza en la práctica (Scott, 1992; 2015).

1.8.6. Reglas para elección de amplitud de banda en EDKs

En su monografía sobre estimación de densidad, Silverman (1986) discute varias reglas para elegir h al trabajar con estimadores de densidad por kernel (EDK). Uno es el *método gráfico de prueba* (Silverman, 1978) el cual consiste en dibujar la segunda derivada de la estimación de densidad (\hat{f}) para varios valores de h y después elegir la amplitud de banda correspondiente a la gráfica con “fluctuaciones rápidas bien definidas que no ocultan por completo la variación sistemática”. Aunque algo de subjetividad está involucrada, en la práctica parece que las gráficas de prueba amplifican la variación de la estimación de densidad y de tal forma la elección de una amplitud de banda apropiada no es muy difícil. No obstante, el mismo autor reconoce que la subjetividad de este procedimiento origina que su uso principal sea una forma de verificar los resultados con otros métodos.

Además del método gráfico bosquejado arriba, Silverman (1986) propuso utilizar una distribución estándar como referencia (de forma semejante a como hizo Scott, 1979 para los histogramas). Si se emplea un kernel Gaussiano, la amplitud óptima de banda es estimada por medio de:

$$\hat{h} = 1.06 \hat{\sigma} n^{-1/5} \quad (1.14)$$

Este autor compara el resultado de su expresión con distribuciones no Gaussianas arribando a conclusiones semejantes a las de Scott (1979): la amplitud de banda resultante de (1.14) sobresuaviza datos marcadamente sesgados, es notablemente insensible a la curtosis (usando a distribuciones log-normal, y de familia t como modelos) y sobresuaviza cada vez más al volverse la distribución más bimodal. Él sugirió usar una medida robusta de la dispersión tal como el recorrido intercuartílico (RIC):

$$\hat{h} = 0.79 R n^{-1/5} \quad (1.15)$$

Esta expresión da mejores resultados en distribuciones sesgadas y de cola larga pero para el caso bimodal sobresuaviza en forma adicional. Más adelante Silverman sugiere utilizar una estimación adaptativa para la dispersión en lugar de $\hat{\sigma}$ en la ecuación (1.14):



$$A = \min \left(\hat{\sigma}, \frac{RIC}{1.349} \right) \quad (1.16)$$

(Notar que esta expresión es la misma que la (1.2) de la Sección 1.5.) lo que resulta en:

$$\hat{h} = 1.06 A n^{-1/5} \quad (1.17)$$

Härdle prefiere esta regla óptima adaptativa (su “regla mejorada del pulgar” en Härdle, 1991).

Puede valer la pena recordar que el $RIC \approx$ *Dispersión de los cuartos* (definida como en Tukey, 1977 y Hoaglin, 1983) dependiendo del método para el cálculo de los cuartiles (ver detalles en Frigge, *et al.*, 1989 o Hamilton, 1992). En Stata, la diferencia entre el RIC y la *dispersión de los cuartos* disminuye gradualmente hasta desaparecer al aumentar el tamaño de muestra (Hamilton, 1992).

Silverman sugiere la aplicación de un ajuste adicional al reducir el factor 1.06 en la ecuación (1.17) hasta 0.9 de la forma siguiente:

$$\hat{h} = 0.9 A n^{-1/5} \quad (1.18)$$

De acuerdo a sus simulaciones, para un kernel Gaussiano, esta regla proporcionará un $ECIM$ dentro del 10% del valor óptimo para las distribuciones de cola larga, asimétricas y bimodales que usó como referencia (Silverman, 1976).

1.8.7. Reglas sobresuavizadas para EDKs

Scott (1992), basado en trabajos previos (Scott y Terrell, 1987; Terrell, 1990), y considerando a la varianza como medidor de escala, derivó la siguiente regla de sobresuavización para los estimadores de densidad por kernel:

$$h_{OS} = 3 \left[\frac{R(K)}{35\sigma_k^4} \right]^{1/5} \sigma n^{-1/5} \quad (1.19)$$

Donde $R(K)$ es la “rugosidad del kernel” y σ_k^4 es la varianza cuadrada del kernel los cuales son constantes y característicos de cada kernel. El cuadro 1.5.1 contiene los valores de rugosidad y varianza para algunos kerneles comunes.



Utilizando este cuadro es posible calcular la regla de sobresuavización para los kernels incluidos. De esta forma, con un kernel bponderado (cuártico) la amplitud de banda sobresuavizada es exactamente $h_{os} = 3 \sigma n^{-1/5}$; para el kernel Gaussiano, $h_{os} = 1.144 \sigma n^{-1/5}$. Esta regla es ligeramente más amplia (8%) que la regla de referencia óptima Gaussiana, (1.14) citada anteriormente la cual usa un factor de 1.06.

Cuadro 1.2. Valores de Rugosidad y varianza para kernels comunes (adaptado de Scott, 1992; 2015).

Kernel	$R(K)$	σ_K^2
Uniforme	1/2	1/3
Triangular	2/3	1/6
Epanechnikov	3/5	1/5
Bponderado	5/7	1/7
Triponderado	350/429	1/9
Gaussiano	$0.5/\sqrt{\pi}$	1
Coseno	$\pi^2/16$	$1 - \frac{8}{\pi^2}$

Nota: los kernels están soportados en [-1,1] excepto el Gaussiano, de acuerdo a las ecuaciones de Härdle, 1991 y Scott, 1992; 2015.

1.9 Métodos de cómputo intensivo

1.9.1. Validación cruzada por mínimos cuadrados

La validación cruzada (VC) es un procedimiento bien conocido para la elección automática del parámetro de suavización. Existen dos tipos de VC: VC por máxima verosimilitud y VC por mínimos cuadrados (VCMC). Fue sugerida por Rudemo (1982) y Bowman (1984) y está basada en una idea muy simple. Considerando al **Error Cuadrado Integrado (ECI)** como una medida de distancia (d) entre el estimador \hat{f} de la densidad f puede definirse como:



$$\begin{aligned}
 d_T(h) &= \int (\hat{f}_h - f)^2(x) dx \\
 &= \int \hat{f}_h^2(x) dx - 2 \int (\hat{f}_h f)(x) dx + \int f^2(x) dx
 \end{aligned} \tag{1.20}$$

Observando la expresión (1.20), podemos notar que el primer término puede calcularse de los datos, el último término no depende de la estimación ni del parámetro de suavización (h), y sólo el término intermedio tiene que ser estimado.

El principio de la **VCMC** es minimizar el primero y segundo términos con respecto a h . Para lograr esto, la **VCMC** toma ventaja de la esperanza calculada con respecto a una observación X adicional e independiente:

$$\int (\hat{f}_h f)(x) dx = E_X[\hat{f}_h(X)]$$

Como por lo general un conjunto de datos adicional no está disponible, se define a la *estimación dejando uno fuera* como:

$$E_X[\widehat{\hat{f}}_h(X)] = n^{-1} \sum_{i=1}^n \hat{f}_{h,i}(X_i)$$

Utilizando las expresiones equivalentes anteriores se obtiene la expresión para la validación cruzada por mínimos cuadrados:

$$VC(h) = \int \hat{f}_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{h,i}(X_i) \tag{1.21}$$

Analizando las propiedades de (1.21) Scott y Terrell (1987) mostraron que es un criterio insesgado de validación cruzada. Härdle (1991) da un algoritmo de computación directa para la **VCMC**, notando que el número de iteraciones es cuadrático en el número de observaciones, un inconveniente que motiva la búsqueda de otro método más eficiente de cálculo.

A este respecto, Silverman (1986) propuso el uso de un algoritmo basado en la transformación rápida de Fourier. Scott y Terrell (1987) utilizaron un procedimiento modificado de **HDP**. Basado en la generalización del **HDP**, o sea el promedio ponderado de puntos redondeados (**PPPR**), método originalmente propuesto por Härdle y Scott (1988), Härdle (1991) presenta un algoritmo eficiente para cálculo que requiere un número lineal de iteraciones en el número de observaciones.



1.9.2. Validación cruzada sesgada

Tomando un enfoque diferente, Scott y Terrell (1987) consideraron estimar directamente el error cuadrado integrado medio asintótico (*ECIMA*). Estos autores llegaron a encontrar un sesgo en la estimación utilizando la norma L_2 y por lo tanto lo llamaron *validación cruzada sesgada* (*VCS*). La expresión general para la *VCS* es:

$$VCS(h) = \frac{1}{nh} \|K\|_2^2 + \frac{h^4}{4} \mu_2^2(K) \left[\|\hat{f}_h''\|_2^2 - \frac{1}{nh^5} \|K''\|_2^2 \right] \quad (1.22)$$

Härdle (1991) da un conjunto completo de expresiones y un algoritmo para su cálculo.

Scott y Terrell (1987) compararon a la *VCMC* (un estimador insesgado) y a la *VCS* aplicadas a datos simulados encontrando que:

- ♦ Para muestras pequeñas ($n = 25$), aproximadamente la mitad de las funciones *VCS* estimadas no tuvieron mínimos locales, aunque para $n > 40$ todas las estimaciones los tuvieron.
- ♦ La *VCS* tuvo menor desviación estándar conduciendo por tanto a una estimación adecuada del *ECIMA*.
- ♦ Si la densidad subyacente es asimétrica o con colas gruesas (Cauchy, lognormal o mezclas Gaussianas), la *VCS* tiende a sobresuavizar mientras que la *VCMC* a pesar de su mayor dispersión, en promedio, produce mejores estimaciones.

Estas generalizaciones dan alguna guía en la elección del procedimiento: si la naturaleza de la verdadera densidad no es simétrica, entonces se recomienda el empleo de la *VCMC*; de otra forma, se aconseja el uso de la *VCS*.

Scott (1992; 2015) concluye que, las reglas de sobresuavización, la *VCS* y la *VCMC* son un poderoso conjunto de herramientas para la elección de la amplitud de intervalo de histogramas y *PF*'s así como para la selección de amplitud de banda en estimadores de densidad por kernel. Este autor recomienda el uso de un gráfico con escala logarítmica $\log(h)$ y sugiere evaluar los resultados buscando fallas (falta de mínimos locales para la *VCS* o $h = 0$ degenerada para la *VCMC*); si las amplitudes de banda resultantes difieren, entonces escoger aquella que conduce a la estimación con la menor cantidad de ruido local, especialmente cerca de los picos de densidad. Los tres procedimientos deberán examinarse simultáneamente aún con muestras de tamaño muy grande.



1.9.3. Ejemplos de aplicación

Como puede verse en las ecuaciones para las reglas arriba incluidas, excepto para los selectores de amplitud de banda por validación cruzada, es sencillo estructurar algoritmos computarizados para su cálculo ya sea utilizando lenguajes de programación de alto nivel o lenguajes de macros de paquetes estadísticos (como Stata). En esta obra se incluyen dos programas: “bandw1.ado” para Stata y “bandw.R para R) los cuales calculan las reglas listadas anteriormente y las presentan en un cuadro con (número de ecuación incluido):

- Número óptimo de intervalos en histogramas (1.4)
- Número sobreesuavizado de intervalos en histogramas (1.8)
- Número sobreesuavizado de intervalos en *PF*'s (1.12)
- Amplitud óptima de intervalo en histogramas (1.5), (1.6),
- Amplitud sobreesuavizada en histogramas (1.7), (1.9) y (1.10)
- Amplitud de intervalo en *PF*'s (1.11) y (1.13)
- Amplitud óptima de banda para kernel Gaussiano (1.18) y (1.17)
- Regla para amplitud sobreesuavizada de banda para kernel Gaussiano (1.19)

Para mostrar la aplicación de este programa se utilizan los datos de longitud: a) 316 individuos de la trucha coralina (*Plectropomus leopardus*, Serranidae), adaptado del reporte original de Goeden (1978) presentados por Sparre y Venema (1992) y Salgado-Ugarte et al. (2005), y b) longitud patrón de 183 individuos (100 indeterminados y 83 hembras) presentados en la sección 1.1 de la presente obra (Salgado-Ugarte, 1985).

Los resultados para los datos longitud de trucha coralina que se presentan en el Cuadro 1.3 pueden obtenerse por

Stata:

```
. use trocolen
. bandw1 length
```



por R:

```
> setwd("c:/data")
> library(foreign)
> trocolen <- read.dta("trocoleno.dta")
> names(trocolen)
[1] "length"
> length(trocolen$length)
[1] 316
> source("c:/Rprogs/bandw.R")
> bandw(trocolen$length, "Gaussian")
```

Cuadro 1.3. Algunas reglas prácticas para elección de número y amplitud de intervalo/banda para estimación de densidad por histogramas, polígonos de frecuencia y estimadores por kernel. Datos de trucha coralina.

No. de intervalos de Sturges	9.3038
No. sobreesuavizado de intervalos	8.5817
No. sobreesuavizado de intervalos en <i>PF</i>	7.4678
Amplitud óptima Gaussiana de Scott	38.0575
Amplitud óptima robusta de Freedman-Diaconis	27.9683
Amplitud sobreesuavizada de Terrell y Scott	26.1331
Amplitud sobreesuavizada homoscedástica	40.5475
Amplitud sobreesuavizada robusta	36.4007
Amplitud óptima Gaussiana en <i>PF</i>	50.3619
Amplitud sobreesuavizada en <i>PF</i>	54.5782
Amplitud de banda óptima Gaussiana de Silverman	20.0982
Amplitud de banda óptima mejorada de Härdle	23.6713
Amplitud de banda sobreesuavizada para kernel Gaussiano de Scott	26.7972

En la Figura 1.14 se incluyen el histograma y el polígono de frecuencia con la regla de amplitud óptima de intervalo para distribución Gaussiana.



En Stata, tanto el histograma como el polígono de frecuencia (*PF*) se calcularon con el programa actualizado “warpdenm1.ado” para estimación de densidad por **ASH-WARP**. Debido a que en la estimación de densidad por **ASH-WARP** el origen de la trama utilizada para el cálculo no es de importancia al incrementarse el parámetro *M* (número de histogramas desplazados), estos programas no permiten su especificación. Es posible probar otros histogramas con diferentes orígenes utilizando otros procedimientos de cálculo. Observando a los resultados con la amplitud de banda óptima de referencia Gaussiana, hay evidencia inequívoca que sugiere la multimodalidad de la distribución de longitudes. El histograma muestra cuatro modas, mientras que el *PF* tiene dos y un “hombro” (inflexión siempre de horizontal a negativa) hacia tallas mayores.

Como se señaló en la sección 1.2, el programa “twoway__histogram_gen”, un comando interno de Stata, permite la generación de variables con los valores de densidad y de puntos medios de cálculo que pueden considerarse como centros o marcas de clase de histogramas. Primeramente se generan las variables (“d” y “m”) para graficar al histograma con la amplitud óptima (38); posteriormente se generan las variables “d50” y “m50” para graficar el polígono de frecuencia con la amplitud óptima (50):

```
. twoway__histogram_gen length, density start(220) w(38) gen(d m)
. twoway__histogram_gen length, density start(220) w(50) gen(d50 m50)
```

Luego se extienden las marcas de clase para incluir la densidad igual a cero:

```
. replace m50 = 645 in 10
(1 real change made)

. replace d50 = 0 in 10
(1 real change made)

. replace d50 = 0 in 9
(1 real change made)
```

Con estas variables con los resultados puede construirse el gráfico:

```
. twoway bar d m if inrange(m,200,600), barwidth(38) bstyle(histogram) || scatter
d50 m50, c(1) ms(i) sort ytitle("Densidad") xtitle("Longitud") legend(off)
```

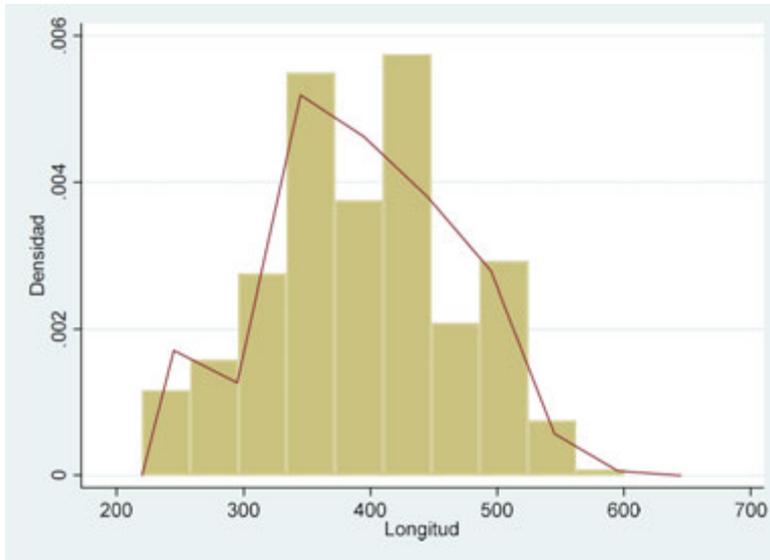


Figura 1.14. Histograma y polígono de frecuencia con la regla de referencia Gaussiana para histogramas (38) y polígonos de frecuencia (50), datos de longitud de trucha coralina.

A continuación se generan los datos (“d41” y “m41”) para el histograma sobre-suavizado y “d54” “m54” para el polígono de frecuencia sobre-suavizado.

```
. twoway__histogram_gen length, density start(220) w(41) gen(d41 m41)  
. twoway__histogram_gen length, density start(220) w(54) gen(d54 m54)
```

Con estos valores se construyen los gráficos sobre-suavizados:

```
. twoway bar d m if inrange(m,200,600), barwidth(41) bstyle(histogram) ||  
scatter d54 m54, c(l) ms(i) sort ytitle("Densidad") xtitle("Longitud")  
legend(off)
```

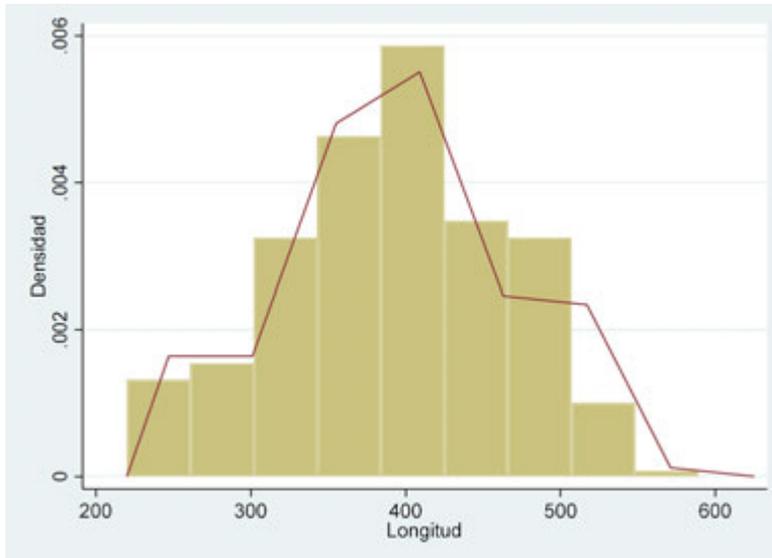


Figura 1.15. Histograma y polígono de frecuencia con la regla de amplitud sobresuavizada para histogramas (41) y polígonos de frecuencia (50), datos de longitud de trucha coralina.

El histograma y el *PF* correspondientes a las amplitudes sobresuavizadas de banda se presentan en la Figura 1.15. Para el histograma se utilizó la mayor amplitud de banda (sobresuavizada de Terrell-Scott). Como era de esperar, estas versiones muestran densidades más suaves, pero en el *PF* sobresuavizado hay un indicio de la presencia de prominencias a cada lado de la moda principal central. Por lo tanto, valdrá la pena emplear otros métodos (por ejemplo el de Silverman, 1981b; 1983) para validar una estructura más compleja.

Para uso posterior de los valores generados se recomienda guardar el archivo modificado:

```
. save trocolen2
```

El cuadro 1.4 contiene los parámetros de suavización sugeridos por las reglas prácticas para los datos de longitud de bagres ($n = 183$).



Cuadro 1.4. Algunas reglas prácticas para elección de número y amplitud de intervalo/ banda para estimación de densidad por histogramas, polígonos de frecuencia y estimadores por kernel. Datos de longitud patrón de bagres.

No. de intervalos de Sturges	8.5157
No. sobresuavizado de intervalos	7.1531
No. sobresuavizado de intervalos en PF	6.6950
Amplitud óptima Gaussiana de Scott	36.9037
Amplitud óptima robusta de Freedman-Diaconis	36.8126
Amplitud sobresuavizada de Terrell y Scott	26.8415
Amplitud sobresuavizada homocedástica	39.3183
Amplitud sobresuavizada robusta	47.9115
Amplitud óptima Gaussiana en PF	45.4047
Amplitud sobresuavizada en PF	49.2060
Amplitud de banda óptima Gaussiana de Silverman	19.0066
Amplitud de banda óptima mejorada de Härdle	22.3856
Amplitud de banda sobresuavizada para kernel Gaussiano de Scott	24.1595

Se muestran a continuación el histograma y el polígono de frecuencia con la amplitud óptima de intervalo (36 y 45 respectivamente) y los comandos para obtenerlos:

```
. twoway_histogram_gen bodlen, density start(40) w(36) gen(d36 m36)
. twoway_histogram_gen bodlen, density start(40) w(45) gen(d45 m45)

. replace d45 = 0 in 6
(1 real change made)

. replace m45 = 287.5 in 7
(1 real change made)

. replace d45 = 0 in 7
(1 real change made)

. twoway bar d36 m36 if inrange(m36,0,300), barwidth(36) bstyle(histogram) ||
scatter d45 m45, c(l) ms(i) sort ytitle("Densidad") xtitle("Longitud patrón
(mm)") legend(off)
```

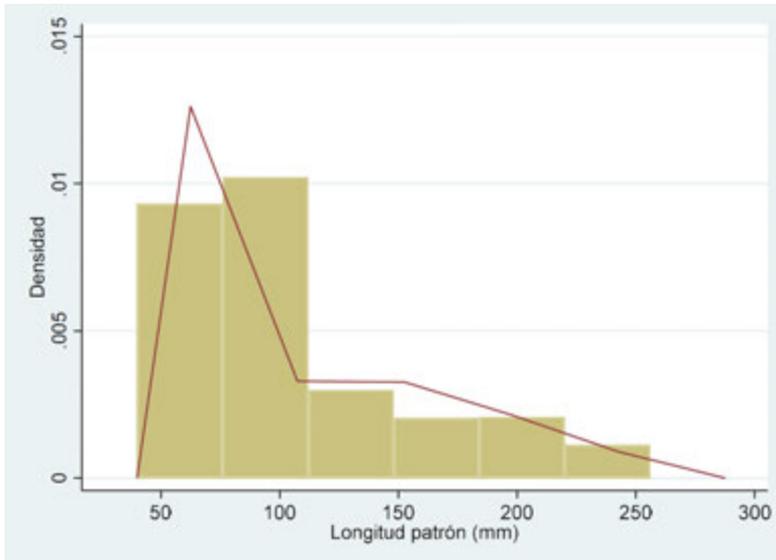


Figura 1.16. Histograma con amplitud óptima Gaussiana ($h = 36$) y polígono de frecuencia ($h = 45$) para los datos de longitud de bagres.

```
. twoway_histogram_gen bodlen, density start(40) w(47) gen(d47 m47)
. twoway_histogram_gen bodlen, density start(40) w(49) gen(d49 m49)

. replace d49 = 0 in 6
(1 real change made)

. replace m49 = 309.5 in 7
(1 real change made)

. replace d49 = 0 in 7
(1 real change made)

. twoway bar d47 m47 if inrange(m47,0,300), barwidth(47) bstyle(histogram) ||
scatter d49 m49, c(1) ms(i) sort ytitle("Densidad") xtitle("Longitud patrón
(mm)") legend(off)
```



Tanto la regla óptima Gaussiana como la sobreesuavizada robusta dan indicación de una estructura multimodal más compleja (Figuras 1.16 y 1.17 respectivamente). Al menos tres modas pueden identificarse fácilmente y los resultados sobreesuavizados proporcionan un fuerte respaldo a su existencia. Por tanto, debe realizarse un análisis adicional enfocado a la caracterización de esta estructura multimodal (ver Silverman, 1981b; Izenman y Sommer, 1988; Salgado-Ugarte, et al. 1997; Salgado-Ugarte, 2002) para un ejemplo de la estrategia a seguir; otras contribuciones para la evaluación de la multimodalidad son Comparini y Gori, 1986, Roeder, 1990, Müller y Sawitzki, 1991 y Ameijeiras-Alonso et al. 2018).

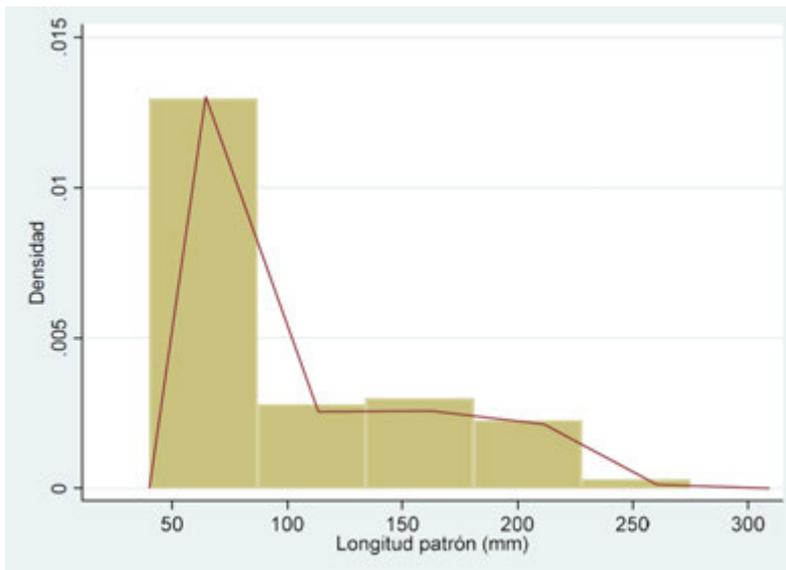


Figura 1.17. Amplitudes sobreesuavizadas de intervalo para histograma ($h = 47$) y polígono de frecuencia ($h = 49$) calculados por el método WARP.

Como lo sugiere Terrell (1990) se pueden utilizar las amplitudes sobreesuavizadas de intervalo al buscar estimaciones conservativas tanto en histogramas o polígonos de frecuencia o podemos explorar las amplitudes de banda sugeridas para estimadores por kernel. El programa “bandw1.ado” para Stata y “bandw.R” para R, como se mostró anteriormente, permiten el cálculo de dos reglas óptimas y una sobreesuavizada para los estimadores de



densidad por kernel. Estos programas permiten utilizar diferentes funciones ponderales (kerneles) de acuerdo al cuadro de conversiones de kerneles equivalentes incluido en Scott (1992; 2015) y Salgado-Ugarte (2002).

Para ejemplificar el uso de la validación cruzada, se utilizarán las rutinas incluidas en el programa “AED2020” acompañante de la presente obra. Un breve tutorial para el uso de este programa se ha publicado en Salgado-Ugarte, et al. (2005) y en el Apéndice 1 de esta obra. El programa, como las versiones “l2cvwarp.ado” y “bcvwarp.ado” y sus versiones actualizadas “l2cvwarp2.ado” y “bcvwarp2.ado” escritas para Stata, emplea a los EDKs calculados por el método eficiente WARP que hace posible el cálculo de numerosos estimadores en poco tiempo.

A continuación se describen los pasos para los datos de trucha coralina.

- Se crea un archivo de texto con una sola variable para que sea leído por AED2020. Para ello puede hacerse a partir del archivo en formato de datos en Stata:

```
. outfile length using trocolen.raw
```

- Posteriormente se lee en AED2020: Archivo-Abrir (una variable)- Archivos RAW(*.raw) – trocolen.raw
- Aparece un resumen estadístico de los datos leídos.
- Se elige del menú de caída la opción Validación Cruzada y en primer lugar se escoge “Mínimos cuadrados”.
- Se abre un cuadro de diálogo en donde se deben especificar los valores para la estimación:
 - Delta: la precisión para estimar la banda óptima. De manera pre-establecida está la unidad.
 - m inicial: el procedimiento escala la amplitud de banda con el número de histogramas a promediar con el valor de delta (amplitud de banda = delta*m). En esta opción se elige el valor inicial del intervalo de búsqueda (valor pre-establecido de 1)
 - m final: el valor final del intervalo de búsqueda (valor pre-establecido de 30).



- Kernel (función ponderal). Es posible emplear varias funciones (cuadrada, triangular, Epanechnikov, Biponderada (Cuártica), Triponderada y Gaussiana.
- Una vez establecidos estas opciones se oprime el botón “Calcular”

El resultado de esta rutina se presenta en dos ventanas. En la parte superior derecha aparece un cuadro con los cinco valores mínimos del puntaje de la validación cruzada con los valores de m y de amplitud de banda resultante del escalamiento arriba citado. El primer número presenta el ancho de banda recomendado por el método. Para los datos de la trucha coralina estos valores se incluyen en la Figura 1.18. En ella puede apreciarse que se han modificado los valores de m inicial y final para obtener una curva más simétrica del puntaje de validación cruzada. La curva muestra claramente un mínimo, cuyo valor numérico se especifica en la ventana superior derecha. El resultado sugiere una amplitud de banda de 22 para un kernel Cuártico (biponderado).

El EDK correspondiente se puede calcular también. Se escoge Estimar-WARP y en el cuadro de diálogo que se abre, se marca “Biponderada” y se escribe “22” en “Amplitud de banda”, se cambia el valor de 5 a 10 en “Número de histogramas” y se oprime el botón “Calcular Densidad”. El resultado se presenta en la Figura 1.19.

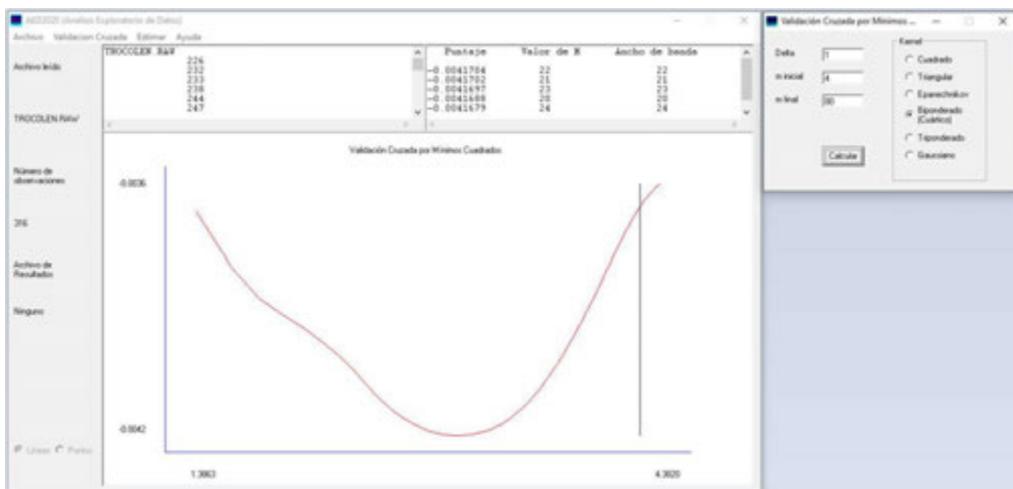


Figura 1.18. Captura de pantalla con el resultado de la rutina de validación cruzada por mínimos cuadrados del AED2020. La línea vertical indica el valor de ancho de banda sobreesuavizado.

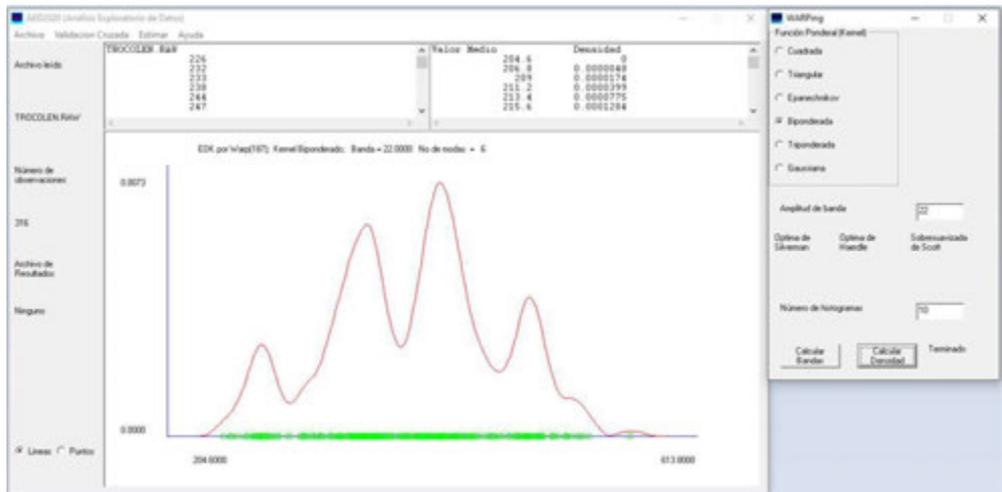


Figura 1.19. EDK con el ancho de banda indicado por la VC por mínimos cuadrados, 22 para kernel bponderado.

Los algoritmos incluidos en la rutina de Validación Cruzada (VC) del AED2020 siguen los procedimientos de Scott y Terrell (1987) basados en los programas de Härdle (1992). Estos algoritmos hacen uso del WARP como método eficiente para calcular a los EDKs

Considerando una de las recomendaciones explícitas de Scott (1992; 2015), y como se sugiere por las gráficas de Härdle (1991) se usa una escala logarítmica para los ejes de M y h . Además, como es sugerido por Scott se incluye una línea de referencia que representa la amplitud de intervalo sobreesuavizada. El cuadro con los cinco valores menores del puntaje (score) de la VC permite obtener la amplitud “óptima” de intervalo. Si un valor mínimo no aparece en el intervalo de M utilizado, pero se sugiere por una tendencia monótona decreciente hacia alguno de los extremos, es posible localizarlo mediante un proceso iterativo de ensayo y error, desplazando el intervalo al especificar valores inicial y final diferentes para la estimación.

De esta manera, por medio de las rutinas del AED2020, es posible estimar el parámetro de suavización “óptimo” para una amplia gama de estimadores de densidad por kernel. Al especificar un kernel triangular es posible obtener la amplitud óptima de intervalo para el *HDP*, y al aumentar el número de histogramas desplazados en la estimación de densidad *WARP* (M) la del estimador de densidad por kernel triangular correspondiente. Las otras funciones ponderales se aplican de acuerdo a sus respectivos estimadores de densidad.



El procedimiento para utilizar la validación cruzada sesgada (VCS) en la estimación del parámetro de suavización es semejante al descrito anteriormente para la VC por mínimos cuadrados (MC). En el menú “Validación Cruzada” se elige “Sesgada”; al aparecer el cuadro de diálogo se especifica uno de los dos kernels soportados (cuártico o triponderado). Como en la VCMC es necesario establecer el valor de “delta” (pre-establecido de uno) así como los valores inicial y final de “m” para la búsqueda. En la Figura 1.20 se muestra el resultado de una búsqueda preliminar utilizando el mismo kernel que con la VCMC (cuártico). Es interesante notar la existencia de dos regiones con valores mínimos de puntaje. El valor más bajo de la curva ocurre a la derecha de la línea que marca la amplitud de banda sobresuavizada. Siendo este valor sobresuavizado el máximo que nos da información sobre la distribución de los datos, el mínimo a su derecha nos sugiere una amplitud de banda que no es de interés. Por tanto nos enfocamos en el mínimo hacia la izquierda.

El resultado de la búsqueda iterativa con el fin de aislar el mínimo nos conduce a la Figura 1.21. El ancho de banda sugerido es de 31 para kernel bponderado, es decir, un poco mayor que el anterior. El resultado de utilizar este ancho de banda se muestra en la Figura 1.22.

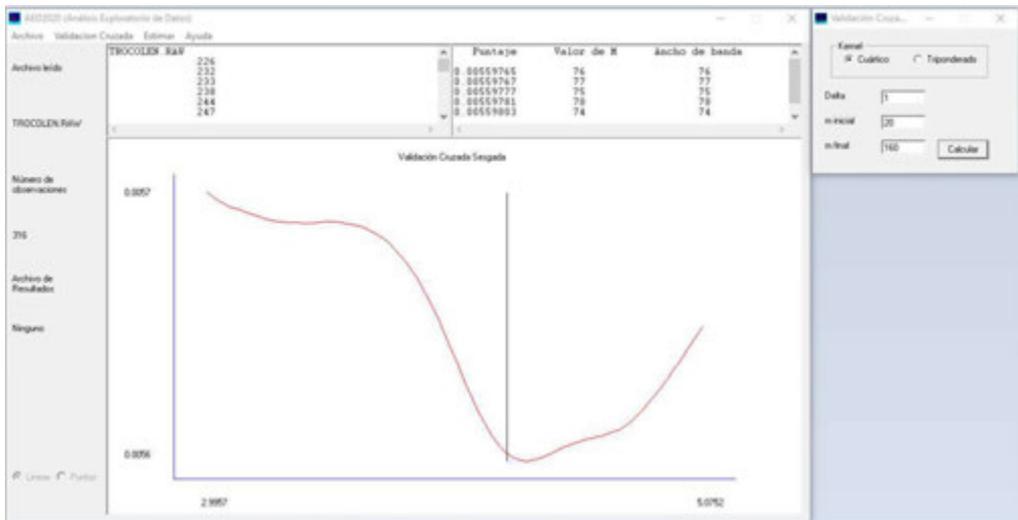


Figura 1.20. Captura de pantalla con el resultado de la rutina de validación cruzada sesgada del AED2020. La línea vertical indica el valor de ancho de banda sobresuavizado. El mínimo hacia la derecha no es de interés.

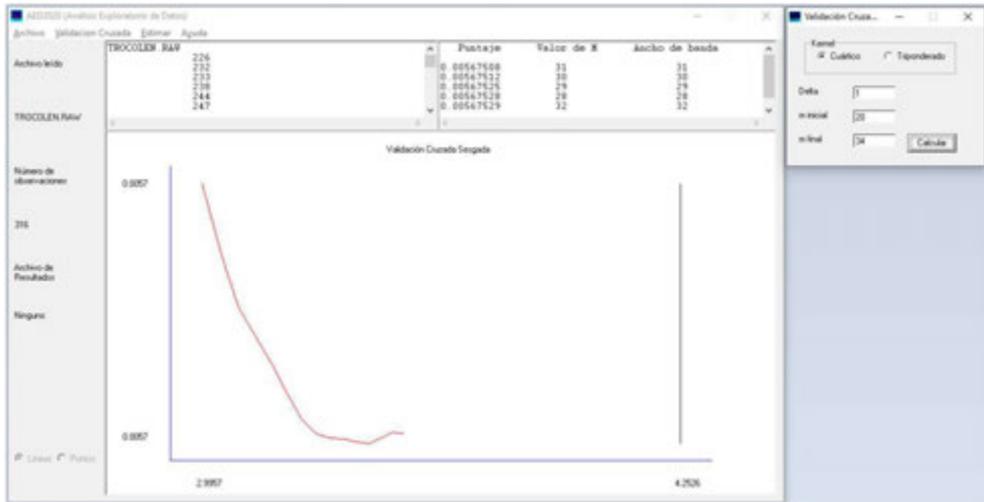


Figura 1.21. Captura de pantalla con el resultado de la rutina de validación cruzada sesgada del AED200. Se ha hecho enfoque en el mínimo local hacia la izquierda de la banda sobreesuavizada.

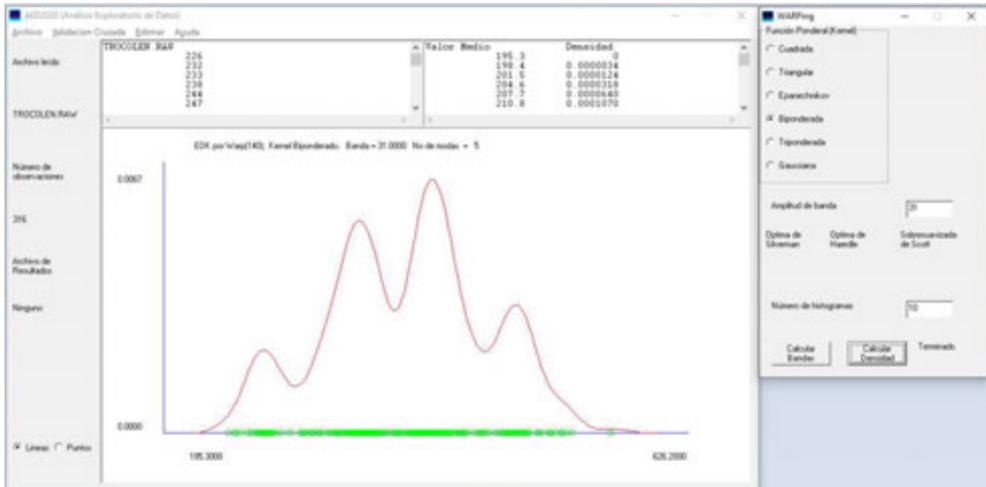


Figura 1.22. EDK con el ancho de banda indicado por la VC sesgada, 31 para kernel bponderado.



Resulta indudable la naturaleza multimodal de la distribución de longitudes de la trucha coralina con al menos cuatro modas.

Como se mencionó anteriormente, por medio del cuadro con factores de conversión para suavización equivalente incluido en Härdle (1991), Scott (1992; 2015), Salgado-Ugarte (2002) y modificado en el Cuadro 1.5 abajo, es posible transformar la amplitud de banda óptima obtenida con cualquiera de las dos funciones sostenidas por la VCS hacia cualquiera de los kernels de dicho Cuadro para calcular la correspondiente estimación de densidad si fuera necesario.

Cuadro 1.5. Factores de conversión entre kernels comunes (Adaptado de Härdle, 1991; Scott, 1992, 2015; Salgado-Ugarte, 2002).

A / De	Unif.	Trian.	Epan.	Bipon.	Tripon.	Cos.	Gauss.
Uniforme	1.000	0.715	0.786	0.663	0.584	0.761	1.740
Triangular	1.398	1.000	1.099	0.927	0.817	1.063	2.432
Epanech.	1.272	0.910	1.000	0.844	0.743	0.968	2.214
Cuártico	1.507	1.078	1.185	1.000	0.881	1.146	2.623
Triponder.	1.711	1.225	1.345	1.136	1.000	1.302	2.978
Coseno	1.315	0.941	1.033	0.872	0.768	1.000	2.288
Gaussiano	0.575	0.411	0.452	0.381	0.336	0.437	1.000

Para el segundo conjunto de datos (longitud de bagres, $n = 138$) es necesario preparar un archivo de texto para poder leerlo en AED2020:

```
. use bagretallas4

. keep bodlen sexa

. drop if sexa==.
(349 observations deleted)

. drop if sexa==1
(109 observations deleted)
```



- . drop sexa
- . saveold baghsmi, version(11)
- . outfile bodlen using baghsmi

En AED2020 se usa Abrir (Una Variable), cambiar en tipo de archivo a “Archivos RAW(*.raw)” y escoger el archivo “baghsmi.raw”. Se abre “Validación Cruzada” – “Mínimos cuadrados” y se prueba con los valores pre-establecidos en el cuadro de diálogo (delta = 1; m inicial = 1, m final = 50) cambiando el kernel a “Biponderado (Cuártico)” para comparar con la amplitud de banda sesgada. El resultado (Figura 1.23) sugiere una amplitud de banda de 10 (valor del eje horizontal en el mínimo de la curva y especificado en el cuadro superior derecho con los resultados).

El EDK con amplitud de banda = 10 con kernel biponderado, es claramente multimodal (Figura 1.24).

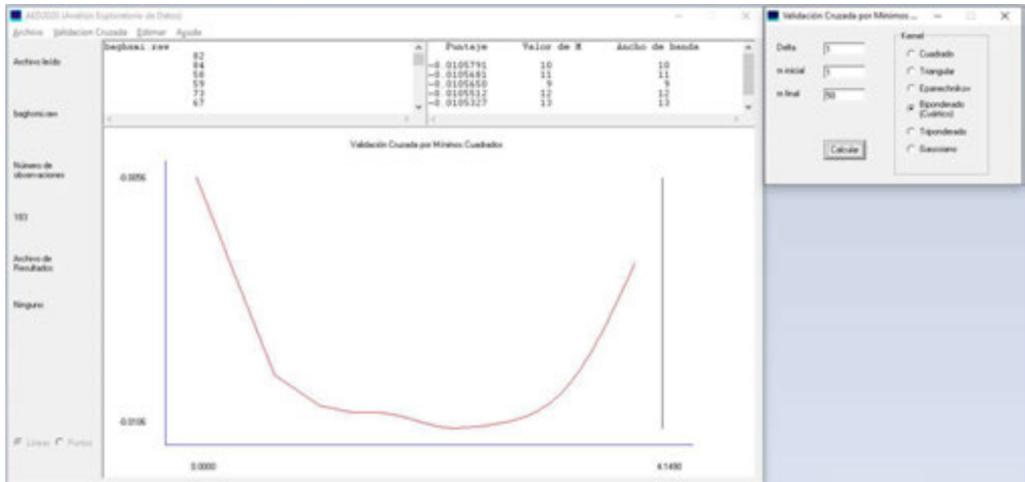


Figura 1.23. Captura de pantalla con el resultado de la rutina de validación cruzada por mínimos cuadrados del AED2020. La línea vertical indica el valor de ancho de banda sobresuavizado.



La búsqueda por Validación Cruzada Sesgada (*VCS*) conduce a una curva de puntajes que, además de un mínimo local de interés por estar a la izquierda de la línea que marca el valor de la banda sobre-suavizada, la existencia de otro mínimo que por parecer estar a la derecha de la banda sobre-suavizada, no es de interés (Figura 1.25). El mínimo a la izquierda sugiere una amplitud de banda de 15, mayor que la indicada por la Validación Cruzada por Mínimos Cuadrados (*VCMC*). El EDK con amplitud de banda = 15 con kernel bponderado se presenta en la Figura 1.26.

Tanto la *VCMC* como la *VCS* producen estimaciones de densidad multimodales (Figuras 1.24 y 1.26). En la práctica es común que la *VCS* resulte en valores mayores que la de mínimos cuadrados. No obstante, la cercanía de los valores y el resultado multimodal de las distribuciones son resultados que deben tomarse en cuenta seriamente. En conclusión, existe una fuerte evidencia para una distribución multimodal de la longitud corporal estándar de los bagres.

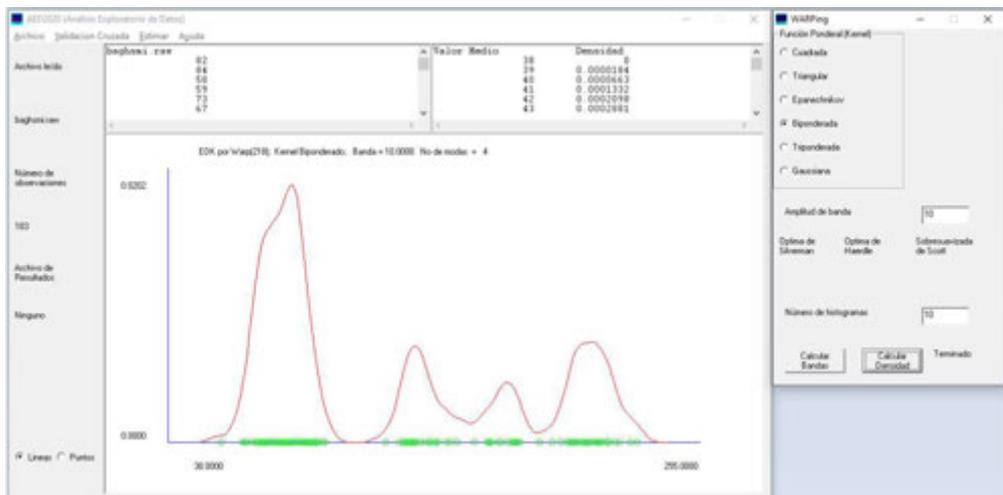


Figura 1.24. EDK con el ancho de banda indicado por la VC por mínimos cuadrados (10) para kernel bponderado.

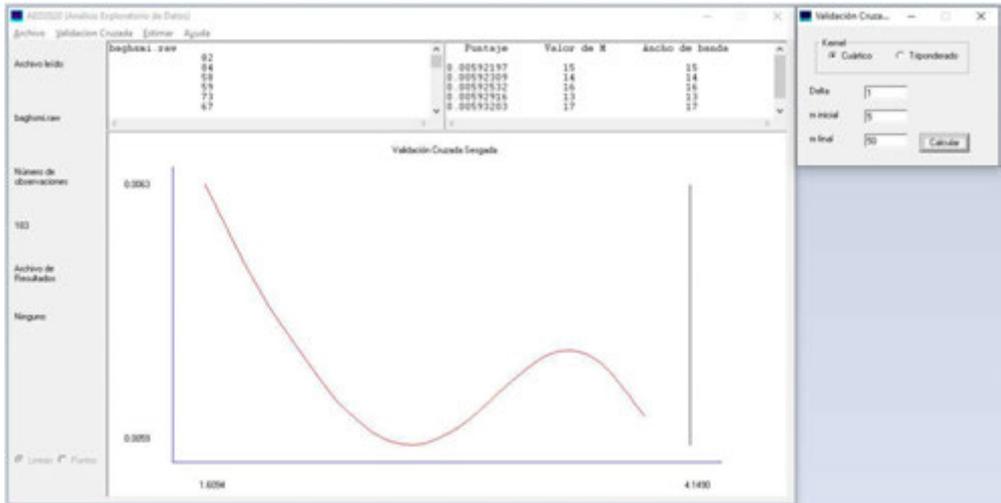


Figura 1.25. Captura de pantalla con el resultado de la rutina de validación cruzada sesgada del AED2020. La línea vertical indica el valor de ancho de banda sobreesuavizado.

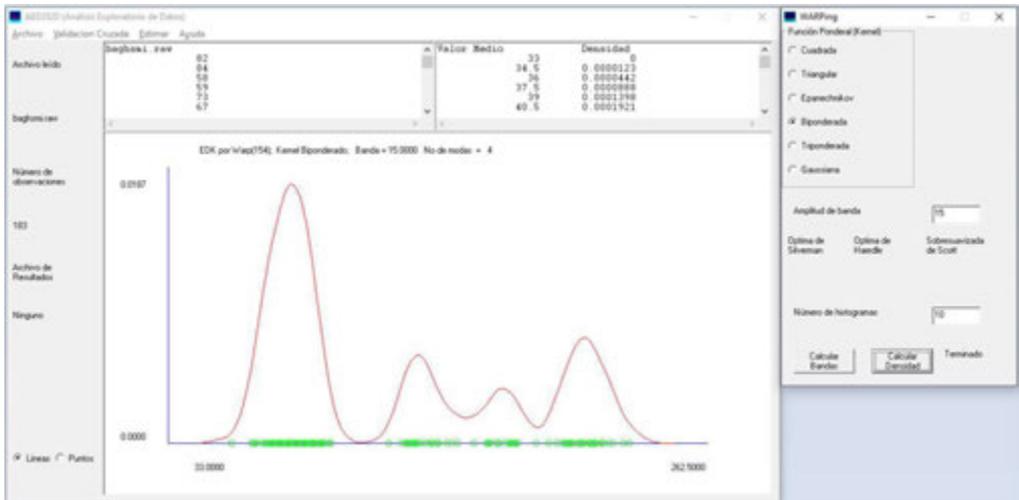


Figura 1.26. EDK con el ancho de banda indicado por la VC sesgada (15) para kernel bponderado.



1.9.4. Algunos comentarios sobre la elección del parámetro de suavización

Se han propuesto algunos otros procedimientos no discutidos en la presente obra. Como ejemplos se pueden citar:

- Validación cruzada por muestreo repetitivo bootstrap (Taylor, 1989).
- Métodos Plug-In (Sheather y Jones, 1991; Hall, *et al.*, 1991).

Esperamos que esta obra motive el interés sobre la aplicación de estos útiles métodos para elegir el parámetro de suavización en la estimación de densidad. Mientras tanto, consideramos que la colección de reglas y métodos presentados proporcionan una guía práctica para la selección de una estimación de densidad adecuada.

Es necesario reconocer que todos los procedimientos introducidos en este Capítulo tienen limitaciones, y como se ha afirmado por varios autores (Marron, 1986; Scott, 1992; 2015) la práctica analítica de dibujar varias estimaciones con diferentes parámetros de suavización no será reemplazada completamente por métodos automáticos de suavización. Desde un punto de vista exploratorio, cualquier elección de amplitud de banda produce estimaciones de utilidad; valores grandes de h permiten reconocer características estructurales generales tales como simetría, casos extraordinarios, modas y localización, mientras que valores pequeños de h revelan estructuras locales que pueden ser reales o artefacticos no presentes en la densidad “verdadera”.

No obstante, la búsqueda de un procedimiento para la elección completamente automática de la amplitud de banda o intervalo ha motivado el desarrollo de algoritmos novedosos como aquellos que se han mencionado en este apartado. En la actualidad este es un campo activo de investigación en Estadística (Scott, 2015).

1.9.5. Evaluación no paramétrica de la multimodalidad

Introducción

Los investigadores que trabajan con formas complejas de distribución han vuelto su atención en años recientes hacia las técnicas no-paramétricas tales como la estimación de densidad por kernel (Silverman, 1986) donde los componentes individuales mezclados pueden detectarse por la identificación de modas (máximos locales) en la distribución subyacente (Izenman y Sommer, 1988).



No hay garantía de que una mezcla de densidades unimodales resulte en una densidad multimodal con modas correspondiendo al número y localización de cada componente individual. Existe una dependencia sobre tanto el espaciamiento de las modas como de las formas relativas de las distribuciones de los componentes.

Sin embargo, en muchas instancias prácticas, la existencia de más de una sola moda sugiere la evidencia de una mezcla. Varias pruebas han sido propuestas para detectar la multimodalidad en una distribución (Hartigan y Hartigan, 1985). Por ejemplo, Good y Gaskins (1980) usaron el método de verosimilitud penalizada para estimación de densidad junto con ciertas técnicas “quirúrgicas” para “búsqueda de protuberancias”, mientras que el método propuesto por Silverman (1981b) combina la estimación de densidad por kernel con un procedimiento de prueba jerárquico de muestreo repetitivo (bootstrap). Ambos métodos son no-paramétricos, basados en los datos y de cómputo intensivo.

Los contextos específicos desempeñan a menudo un papel prominente en la relación de las modas empíricas con la existencia de componentes mezclados. Así, Silverman (1978a, 1981a) en un estudio de mortalidad infantil súbita o “cot death” sugirió que una densidad de mortalidad esencialmente bimodal pudiera explicarse como una mezcla de la densidad de muerte en hospital (infantes que mueren en el hospital por causas conocidas) con una densidad contaminada y desplazada. De manera similar, Good y Gaskins (1980) atribuyeron a un ajuste trimodal de un conjunto de datos sobre condritas carbonosas a la existencia de (al menos) tres tipos diferentes de condritas. En el caso de mezclas filatélicas analizadas por Izenman y Sommer (1988), el contexto histórico desempeña el papel principal en la identificación de los componentes en la mezcla. La frecuencia de tamaños en peces puede representar diferentes grupos de edad.

En esta Sección se presentará el procedimiento de muestreo repetitivo suavizado sugerido por Silverman (1981b, 1983) el cual es descrito en más detalle ya que utiliza estimadores de densidad por kernel, procedimientos revisados anteriormente.

1.9.6 Prueba de multimodalidad de Silverman

Descripción de la prueba

Esta prueba propuesta por Silverman (1981b) usa técnicas no paramétricas de estimación de densidad por kernel para determinar el número más probable de modas en la densidad subyacente.



Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de una distribución con densidad desconocida f . Definamos a \hat{F}_n como la función empírica de distribución acumulada de la muestra obtenida al asignar masa de $1/n$ a cada X_i . Como una clase general de estimadores no paramétricos de densidad de f , Rosenblatt (1956) propuso

$$\hat{f}(x) = \int_{-\infty}^{+\infty} w_n(x, u) d\hat{F}_n(u) = n^{-1} \sum_{j=1}^n w_n(x, X_j), \quad x \in \mathfrak{R} \quad (1.23)$$

donde la función ponderal positiva $w_n(x, u)$ satisface $\int_{-\infty}^{+\infty} w_n(x, u) = 1$. El estimador de densidad por kernel para f se obtiene estableciendo

$$w_n(x, u) = h^{-1} K((x - u)/h), \quad x \in \mathfrak{R} \quad (1.24)$$

donde el kernel K es una función de densidad de probabilidad, y la amplitud de banda o amplitud de ventana $h = h(n) > 0$ del kernel converge a 0 al $n \rightarrow \infty$. Por tanto, el estimador de densidad por kernel se escribe como

$$\hat{f}(x) = (nh)^{-1} \sum_{j=1}^n K((x - X_j)/h), \quad x \in \mathfrak{R} \quad (1.25)$$

Como se ha mostrado previamente, la elección de h en 1.25 es un problema estadístico de importancia: Un valor demasiado pequeño de h produce una estimación de densidad demasiado dependiente en los valores de la muestra, mientras que un valor demasiado grande para h produce el efecto opuesto y sobreesuaviza la estimación de densidad removiendo peculiaridades interesantes. En la Sección anterior se han presentado varios criterios óptimos para la elección de la amplitud de banda. Debe señalarse sin embargo, que la estimación de h óptima no es el interés principal aquí, sino es el conteo de modas, no en la suavización óptima (aunque ambos temas estén relacionados) (Izenman y Sommer, 1988).

La prueba de multimodalidad de Silverman puede describirse en los siguientes términos. La hipótesis nula H_0^k , establece que f posee cuando más k modas, mientras que la hipótesis alternativa, H_1^k , establece que f tiene más de k modas ($k = 1, 2, \dots$). Si establecemos que

$$N(f) = \#\{x: f'(x) = 0, f''(x) < 0\} \quad (1.26)$$

sea el número de modas en f , entonces $H_0^k: N(f) \leq k$ and $H_1^k: N(f) > k$. Si \hat{f}_h es el estimador de densidad por kernel de f con amplitud de banda h , entonces un estadístico de interés es $N(\hat{f}_h)$. Definiendo la k ésima amplitud crítica de banda como

$$h_{k, \text{crit}} = \inf \{h: N(\hat{f}_h) \leq k\}, \quad (1.27)$$



la amplitud de banda menor que aún es compatible con H_0^k .

El método de Silverman es altamente dependiente de las propiedades del kernel Gaussiano, por lo que es necesario establecer

$$K(t) = (2\pi)^{-1/2} \exp(-t^2/2), \quad t \in \mathfrak{R}, \quad (1.28)$$

como la función kernel en 1.25. Silverman mostró que para el kernel 1.28, $N(\hat{f}_h)$ es una función de h decreciente continua hacia la derecha y por tanto $N(\hat{f}_h) > k$ si y sólo si $h = h_{k,crit}$. Por lo tanto, para encontrar $h_{k,crit}$ se cuentan las modas en cada estimación de densidad \hat{f}_h para diferentes valores de h . Cuando $h = h_{k,crit}$, \hat{f}_h desplegará k modas más un “hombro” notable en su gráfica, y si h se reduce un poco más, brotará una moda adicional ($k + 1$) en el lugar de ese “hombro”. Se denomina **densidad crítica** a cualquier densidad con un hombro, tal como $\hat{f}_{h_{k,crit}}$. De lo anterior se deriva

$$\Pr\{h_{k,crit} > h\} = \Pr\{N(\hat{f}_h) > k \mid X_1, X_2, \dots, X_n \text{ es una muestra extraída de } f\} \quad (1.29)$$

Puesto que no es difícil sacar muestras de una densidad crítica, Silverman combinó la propiedad de monotonicidad de $N(\hat{f}_h)$ con el procedimiento bootstrap de muestreo repetitivo con reemplazamiento para construir una prueba funcional para la multimodalidad.

La significancia de $h_{k,crit}$ se evalúa a través del siguiente algoritmo:

1. Extraer n veces con reemplazamiento de la muestra aleatorio original X_1, X_2, \dots, X_n , para obtener una muestra bootstrap $X_{1^*}, X_{2^*}, \dots, X_{n^*}$.
2. Se obtiene una muestra bootstrap suavizada, $Y_1^*, Y_2^*, \dots, Y_n^*$ por medio de calcular

$$Y_j^* = c_k \{X_j^* + h_{k,crit} Z_j\} \quad j = 1, 2, \dots, n, \quad (1.30)$$

donde Z_j es una desviación Gaussiana estándar independiente, y

$$c_k = (1 + [h_{k,crit}/s]^2)^{1/2} \quad (1.31)$$

se utiliza para restablecer la escala de tal forma que la varianza de Y_j^* sea igual a la varianza de la muestra s^2 de los datos originales (ver Efron, 1982). En otros términos, muestreamos de la convolución $c_k \{\hat{F}_n * F_{norm}\}$ donde F_{norm} es la función distribución $N(0, h_{k,crit}^2)$. Esta convolución es más suave que \hat{F}_n por sí sola.



3. Usar (1.30), (1.25) y (1.28) para formar una estimación de $\hat{f}_{h_{k,crit}}^*$ de f .
4. Repetir los pasos 1-3 un gran número B de veces. Definamos a $\hat{f}_{h_{k,crit}}^{*b}$ para denotar la estimación de densidad para la b ava muestra bootstrap suavizada.
5. Establecer

$$I_{k,b} = 1, \text{ si } N(\hat{f}_{h_{k,crit}}^{*b}) > k$$

$$= 0, \text{ en cualquier otro caso} \quad (1.32)$$

Entonces,

$$P_k = B^{-1} \sum_{b=1}^B I_{k,b} \quad (1.33)$$

es el nivel de significancia estimado (o valor p) de $h_{k,crit}$ y detenerse hasta obtener un valor lo suficientemente “grande”.

Silverman (1981) no ofreció sugerencias de que tan “grande” pueda ser para seguir la regla de alto arriba citada. En su artículo aplicó el procedimiento a los datos de meteoritos condritas de Good y Gaskins (1980) ($n = 22$) y mostró que las amplitudes de banda críticas tuvieron valores p :

$P_1 = 0.08$, $P_2 = 0.05$, $P_3 = 0.79$ y $P_4 = 0.93$ parando a una $k = 3$ para una densidad trimodal.

En un artículo subsiguiente, Silverman (1983) mostró teóricamente que esta prueba bootstrap puede ser conservativa. Puesto que no se han publicado estudios de simulación de esta prueba, los investigadores la ven como una técnica analítica exploratoria de datos. Izenman y Sommer (1988) sugieren que debe tomarse una actitud flexible en la aplicación de la prueba a datos con estructuras de distribución complejas y con colas largas. Estos autores comentan que no hay razón para esperar que la secuencia de valores p (1.33) sea monótonamente creciente; por cierto que el propio estudio de Silverman sobre las condritas ilustra este punto. Además, es posible, dependiendo de la posición de las modas, que se observen en la práctica grandes fluctuaciones en los valores de p . Basados en esta experiencia y en las anotaciones previas en relación con la naturaleza conservativa de la prueba bootstrap, los citados autores sugieren aplicar una regla de paro flexible con un valor nominal de p de 0.40 hasta que estudios detallados se lleven a cabo. Izenman y Sommer, no obstante, recomiendan fuertemente el estudio de las gráficas de las estimaciones de densidad cerca de cada banda crítica y las gráficas de las estimaciones de densidad de las propias bandas críticas durante el desarrollo de la prueba de Silverman.



1.10. Ejemplos de aplicación

Se presenta a continuación un ejemplo de la aplicación de los métodos presentados arriba. De julio, 1980 a agosto 1981 se recolectó un total de 2436 individuos del bagre estuarino *Cathorops melanopus* en la Laguna de Tampamachoco localizada en la costa Noreste de México. Debido a que se encontró diferencia estadística significativa entre sexos se consideró conveniente analizar los datos por sexos separados. Los organismos de sexo indeterminado a pesar de aparecer en unos pocos meses, fueron tan abundantes que predominaron en la recolecta total. Con el fin de utilizar los datos como ejemplo de análisis de frecuencia de tallas para estimar el crecimiento, se consideraron sólo los especímenes capturados en noviembre de 1980 y se extrajo aleatoriamente una submuestra de tamaño 100 de organismos indeterminados con el fin de equilibrar las abundancias. Para este ejemplo se incluyeron las 83 hembras capturadas y la submuestra de 100 indeterminados para evaluar su multimodalidad.

Para calcular las estimaciones de densidad por kernel se utilizaron las versiones actualizadas de los programas para Stata escritos por Salgado-Ugarte et al. (1993); Salgado-Ugarte (2002); Salgado-Ugarte *et al.* (2005).

1.10.1. Uso de reglas de amplitud de banda

De la Sección 1.8.10 se tiene el cuadro 1.4 con los resultados de las reglas para amplitud de intervalo/banda para el conjunto de datos aquí considerado.

En la investigación de la distribución de los datos por medio de estimadores de densidad por kernel se utilizará aquí, en primera instancia, la estrategia de tres etapas:

- i) Aplicación de la banda sobreesuavizada
- ii) Aplicación de la banda “óptima”
- iii) En su caso, utilizar la mitad de la banda “óptima”.

En la etapa i), cualquier característica de la distribución que se aparte de la gaussianidad (sesgo, multimodalidad, “hombros”) es un carácter significativo de los datos. Es así, que si el resultado se aparta notablemente de la forma de campana de Gauss, el valor menor



(banda óptima de Silverman) de los incluidos en el cuadro no es el “óptimo”. El ancho de banda que muestra de mejor manera la distribución de los datos es un valor menor. No obstante, se aplica en el paso ii) y si los datos no son gaussianos, las características de sesgo y multimodalidad serán más evidentes. Como este valor “óptimo” no lo es en este caso, se debe usar un valor menor. Como hasta este punto no se sabe que tanto menos debe valer, se utiliza un criterio mediano y se toma entonces para representar a la distribución de los datos, la mitad de la banda “óptima”.

Se presenta el EDK con la amplitud de banda sobresuavizada de Scott (24.15) en la Figura 1.28. Es evidente por lo menos la bimodalidad y el sesgo positivo de la distribución.

```
. use bagretallas
```

```
. warpdenml bodlen, b(24.15) m(10) k(6) ytitle("Densidad") xtitle("Longitud patrón (mm)") t1("Densidad WARP (poligonal), bw = 24.15, m = 10, k = 6")
```

Con el conocimiento de que la banda “óptima” de Silverman ($h_0 = 19$) no es tal, la aplicación de este valor más angosto resulta en una distribución claramente trimodal (Figura 1.28).

```
. warpdenml bodlen, b(19) m(10) k(6) ytitle("Densidad") xtitle("Longitud patrón (mm)") t1("Densidad WARP (poligonal), bw = 19, m = 10, k = 6")
```

Aplicando el paso siguiente entonces, el EDK con la mitad del óptimo de Silverman ($19/2 = 9.5$) se presenta en la Figura 1.29. En esta Figura los tamaños de las hembras se han separado en tres grupos para que junto con los indeterminados se obtenga una distribución tetramodal.

```
. warpdenml bodlen, b(9.5) m(10) k(6) ytitle("Densidad") xtitle("Longitud patrón (mm)") t1("Densidad WARP (poligonal), bw = 9.5, m = 10, k = 6")
```

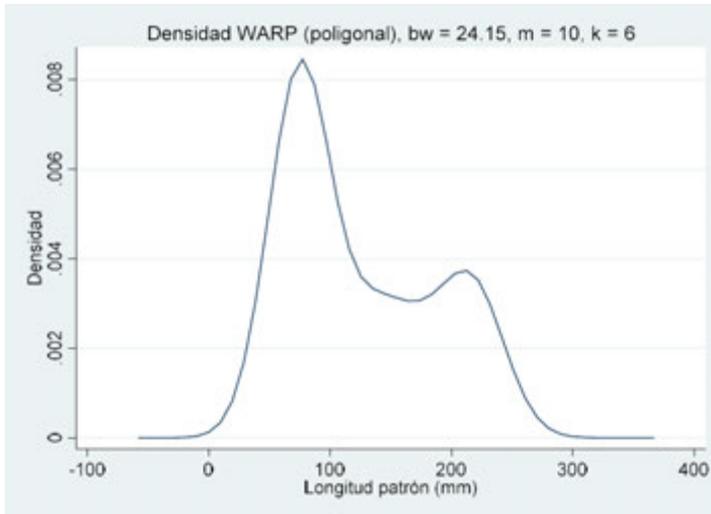


Figura 1.27. EDK con la banda sobre-suavizada (24.15) de Scott para los datos de bagre estuarino *Cathorops melanopus* (hembras e indeterminados).

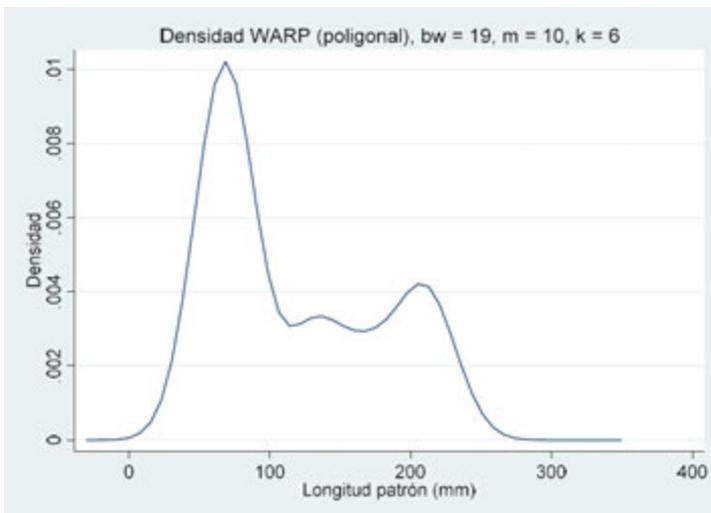


Figura 1.28. EDK con la banda “óptima” de Silverman (19) para los datos de bagre estuarino *Cathorops melanopus*.

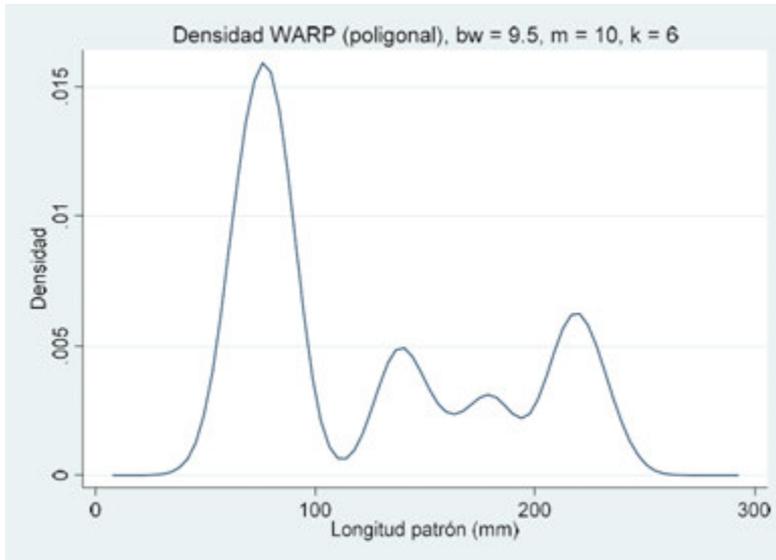


Figura 1.29. EDK con la mitad de la banda “optima” de Silverman ($19/2 = 9.5$) para los datos de bage estuarino *Cathorops melanopus*.

El valor de ancho de banda recomendado por la Validación Cruzada por mínimos cuadrados fue de 10 para kernel bponderado, el cual transformado por los factores del Cuadro 1.5 equivale a $10 * 0.381 = 3.81$, un valor menor a 9.5. La amplitud de banda cuártica (bponderada) recomendada por la Validación Cruzada Ssegada fue de 15, que transformada a Gaussiana es $(15 * 0.381) 5.72$, valor también menor que la mitad de la amplitud “optima” de Silverman.

La naturaleza multimodal de estos datos es innegable, pero, para verificarla de manera formal, se someterán estos datos a la prueba de multimodalidad de Silverman por Bootstrap suavizado.



1.10.2. Aplicación de la prueba de multimodalidad de Silverman

Lo primero que se debe hacer es identificar a las “amplitudes de banda críticas”, es decir, aquel valor último de ancho de banda compatible con un cierto número de modas.

En Stata esto se lleva a cabo con los programas actualizados de Salgado-Ugarte et al. (1997), Salgado-Ugarte (2002) y Salgado-Ugarte *et al.* (2005). Rutinas equivalentes están incluidas en el programa AED2020.

La identificación de las amplitudes críticas se lleva a cabo con el programa “critiband.ado”, el cual requiere la especificación como opciones de comando de un valor de amplitud de ventana alto (bwh siglas de las palabras en inglés bandwidth high) y uno bajo (bwl, bandwidth low) para fijar los límites de búsqueda. Es necesario también especificar la precisión con la cual el programa disminuirá del valor alto hasta el bajo con la opción “step”. Este programa hace uso de la rutina “numodes.ado” que cuenta el número de modas por medio del registro del cambio de signo de las primeras diferencias de las densidades producidas por “warpdenm1.ado”. Es por esto que se necesita una alta precisión en las estimaciones, lo cual se logra aumentando el número de histogramas a promediar (la opción “m”). En nuestra experiencia una opción que proporciona estimaciones adecuadas es fijar la opción a m(40).

Si la precisión es alta step(0.01) se obtiene un gran número de estimaciones, por lo que será necesario buscar las bandas críticas por etapas. Para los datos del ejemplo considerado, se muestran los fragmentos representativos de la búsqueda:

```
. critiband bodlen, bwh(40) bwl(20) st(.01) m(40)
Estimation number = 1      Bandwidth = 40      Number of modes = 1
Estimation number = 2      Bandwidth = 39.99  Number of modes = 1
Estimation number = 3      Bandwidth = 39.98  Number of modes = 1

Estimation number = 445    Bandwidth = 35.56  Number of modes = 1
Estimation number = 446    Bandwidth = 35.55  Number of modes = 1
Estimation number = 447    Bandwidth = 35.54  Number of modes = 1
Estimation number = 448    Bandwidth = 35.53  Number of modes = 2
Estimation number = 449    Bandwidth = 35.52  Number of modes = 2

Estimation number = 1750   Bandwidth = 22.51  Number of modes = 2
Estimation number = 1751   Bandwidth = 22.5   Number of modes = 2
Estimation number = 1752   Bandwidth = 22.49  Number of modes = 2
```



```
Estimation number = 1753      Bandwidth = 22.48      Number of modes = 3
Estimation number = 1754      Bandwidth = 22.47      Number of modes = 3
```

El primer comando permitió encontrar las bandas críticas para una y dos modas. Se continuó con intervalos menores para las siguientes.

```
. critiband bodlen, bwh(20) bwl(10) st(.01) m(40)
Estimation number = 1      Bandwidth = 20      Number of modes = 3
Estimation number = 2      Bandwidth = 19.99   Number of modes = 3
Estimation number = 3      Bandwidth = 19.98   Number of modes = 3
Estimation number = 4      Bandwidth = 19.97   Number of modes = 3

Estimation number = 600    Bandwidth = 14.01   Number of modes = 4
Estimation number = 601    Bandwidth = 14      Number of modes = 4
Estimation number = 602    Bandwidth = 13.99   Number of modes = 3
Estimation number = 603    Bandwidth = 13.98   Number of modes = 4
Estimation number = 604    Bandwidth = 13.97   Number of modes = 4
```

Este intervalo de bandas hizo posible encontrar la banda crítica para tres modas. La búsqueda con un intervalo menor permitió identificar un número mayor de bandas críticas.

```
. critiband bodlen, bwh(5) bwl(1) st(.01) m(40)
Estimation number = 1      Bandwidth = 5      Number of modes = 4
Estimation number = 2      Bandwidth = 4.99   Number of modes = 4
Estimation number = 3      Bandwidth = 4.98   Number of modes = 4
Estimation number = 4      Bandwidth = 4.97   Number of modes = 4

Estimation number = 89     Bandwidth = 4.12   Number of modes = 4
Estimation number = 90     Bandwidth = 4.11   Number of modes = 5
Estimation number = 91     Bandwidth = 4.1    Number of modes = 4
Estimation number = 92     Bandwidth = 4.09   Number of modes = 5
Estimation number = 93     Bandwidth = 4.08   Number of modes = 5

Estimation number = 107    Bandwidth = 3.94   Number of modes = 5
Estimation number = 108    Bandwidth = 3.93   Number of modes = 5
Estimation number = 109    Bandwidth = 3.92   Number of modes = 5
Estimation number = 110    Bandwidth = 3.91   Number of modes = 6
Estimation number = 111    Bandwidth = 3.9    Number of modes = 6
```



Estimation number = 150	Bandwidth = 3.51	Number of modes = 6
Estimation number = 151	Bandwidth = 3.5	Number of modes = 7
Estimation number = 152	Bandwidth = 3.49	Number of modes = 6
Estimation number = 153	Bandwidth = 3.48	Number of modes = 7
Estimation number = 154	Bandwidth = 3.47	Number of modes = 7
Estimation number = 175	Bandwidth = 3.26	Number of modes = 7
Estimation number = 176	Bandwidth = 3.25	Number of modes = 8
Estimation number = 177	Bandwidth = 3.24	Number of modes = 7
Estimation number = 178	Bandwidth = 3.23	Number of modes = 8
Estimation number = 179	Bandwidth = 3.22	Number of modes = 8
Estimation number = 204	Bandwidth = 2.97	Number of modes = 9
Estimation number = 205	Bandwidth = 2.96	Number of modes = 9
Estimation number = 206	Bandwidth = 2.95	Number of modes = 8
Estimation number = 207	Bandwidth = 2.94	Number of modes = 9
Estimation number = 208	Bandwidth = 2.93	Number of modes = 9
Estimation number = 213	Bandwidth = 2.88	Number of modes = 9
Estimation number = 214	Bandwidth = 2.87	Number of modes = 9
Estimation number = 215	Bandwidth = 2.86	Number of modes = 9
Estimation number = 216	Bandwidth = 2.85	Number of modes = 10
Estimation number = 217	Bandwidth = 2.84	Number of modes = 10

Con este intervalo menor se identificaron las bandas críticas de 5, 6, 7, 8 y hasta 9 modas. Es interesante notar que los primeros intervalos de búsqueda son muy amplios y que a partir de la amplitud de banda crítica para cuatro modas, la separación de las siguientes es muy pequeña. Esto parece ser una indicación del número estadísticamente significativo de modas en la distribución.

Una vez identificados los valores críticos, se lleva a cabo la prueba de Bootstrap suavizado para multimodalidad de Silverman. Para ello, se requieren varios pasos. Primero se generan las muestras por Bootstrap (muestreo con repetición); se establece un valor inicial (semilla) y se usa la rutina “bootsamb.ad” la cual requiere como argumentos el nombre de la variable (“bodlen”), el valor de la banda crítica (35.54 para una moda) y la desviación estándar de los datos (59.86215); además se establece el número (B) de muestras bootstrap con la opción “iterations”. En este ejemplo se usará un número mayor ($B = 600$) al que usó Silverman en la presentación del método ($B = 100$). Con estos programas es posible realizar un número mayor de repeticiones. El programa produce los valores de las muestras con el ajuste de la desviación estándar (bootstrap suavizado) y repite los datos originales.



Para evitar tener en la memoria estas repeticiones de los datos originales se borran de la memoria y se conservan los valores requeridos por la prueba de Silverman: “ysm” (valores ajustados) y “_rep” (identificador de la muestra).

Con estas muestras bootstrap se utiliza el programa “silvtest1.ado”, el cual, para cada muestra calcula el EDK por medio del método WARP y cuenta sus modas. Cada vez que la estimación de densidad tenga un número mayor de modas que las especificadas, se cuenta para obtener la estimación del valor de P dividiendo este número por el total de repeticiones. Para usar “silvtest.ado” se requiere especificar las variables “ysm” y “_rep” y como opciones **critbw** (amplitud crítica), **mval** (no. de histogramas a promediar, se usa 40), **nurini** (valor inicial del intervalo de búsqueda), **nurfin** (valor final del intervalo de búsqueda), **cnmodes** (número de modas para la amplitud crítica probada).

A continuación, se incluye la transcripción de la aplicación de estos pasos para obtener el valor de P para una moda:

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
bodlen	183	125.0273	59.86215	48	240

Para obtener la desviación estándar de los datos de longitud estándar

```
. set seed 12345

. boot bootsamb, arg(bodlen 35.54 59.86215) i(600)
warning: data in memory will be lost.
Press enter to continue, Ctrl-Break to abort.
```

Bootstrap:

```
Program: bootsamb
Arguments: bodlen 35.54 59.86215
Options:

Replications: 600
Data set size: 183
Sample size: _N
```

Contains data

obs: 109,800

bootsamb bootstrap



```
vars:                4
-----
      storage   display   value
variable name  type     format   label     variable label
-----
_rep           long     %12.0g
bodlen        float    %9.0g
ysm           float    %9.0g
_obs          long     %12.0g
              observations
-----
Sorted by:
    Note: Dataset has changed since last saved.

. keep ysm _rep
```

Al usar “bootsamb.ado” se presenta la advertencia de que los datos en la memoria serán sustituidos por los valores calculados por el programa. Se oprime la barra espaciadora y al aparecer los resultados y para ahorrar espacio en la memoria del programa se conservan sólo las variables utilizadas por el siguiente comando (“ysm” y “_rep”).

El programa “silvtest.ado” calcula en este caso 600 EDKs, cuenta las modas de cada uno y calcula el valor de P . El listado completo de los resultados presentados depende del número de muestras. De manera pre-establecida, el programa grafica los resultados. Es muy recomendable en primera instancia poner atención a la variación gradual de los EDKs obtenidos con un valor cada vez menor de amplitud de banda. Esta graficación continua es una animación que permite observar como el número de modas va aumentando al disminuir la amplitud de ventana.

```
. silvtest ysm _rep, cr(35.54) m(40) nuri(1) nurf(600) cnm(1)

bs sample    1           Number of modes = 1
bs sample    2           Number of modes = 1
bs sample    3           Number of modes = 1

bs sample    598         Number of modes = 1
bs sample    599         Number of modes = 1
bs sample    600         Number of modes = 1
```

Critical number of modes = 1

Pvalue = 1 / 600 = 0.0017



En el desplegado de resultados se ha omitido el despliegue del número de modas y tan sólo se incluyeron los iniciales y los finales. El programa presenta el número de estimaciones que tuvieron más de las modas probadas, en este caso no hubo ninguna muestra bootstrap que su EDK tuviera más de una moda, por lo que el valor estimado de P es igual a $1/600 = 0.0017$.

Como primer paso para probar ahora dos modas es recuperar los valores originales ya que para la prueba de Silverman estos se borraron de la memoria. Ya que se recuperaron, se cambia el valor inicial para los números pseudoaleatorios (semilla) y se repiten los pasos: `bootsamb.ado`, `keep` y `silvtest.ado` pero ahora con la banda crítica para dos modas.

```
. use baghsmi, clear

. set seed 12347

. boot bootsamb, arg(bodlen 22.49 59.86215) i(600)
```

Se omite desplegado de resultados

```
. keep ysm _rep

. silvtest ysm _rep, cr(22.49) m(40) nuri(1) nurf(600) cnm(2)
bs sample      1                Number of modes = 2
bs sample      2                Number of modes = 2
bs sample      3                Number of modes = 1

bs sample      599              Number of modes = 2
bs sample      600              Number of modes = 2
```

```
Critical number of modes =      2
```

```
Pvalue =          5 / 600 = 0.0083
```

Se repiten estos pasos (desde la recuperación de los datos originales) por cada número de modas a probar y los resultados se presentan como en el Cuadro 1.6.



Cuadro 1.6. Amplitudes críticas de banda y niveles estimados de significancia para los datos de longitud de bagres (hembras e indeterminados) $n = 183$.

Número de modas	Amplitudes críticas	Valores de p
1	35.54	0.0017 (1)
2	22.49	0.0083 (5)
3	13.99	0.0250 (15)
4	4.10	0.7950 (477)
5	3.92	0.5650 (339)
6	3.49	0.5800 (348)
7	3.24	0.4667 (280)
8	2.95	0.4967 (298)
9	2.86	0.3133 (188)

NOTA: Los valores de p se obtuvieron de $(\#)/B$, donde $B = 600$ muestras con reemplazo bootstrap de tamaño 183.

El Cuadro 1.6 muestra las amplitudes críticas de banda y sus correspondientes valores p para los datos de longitud de bagres. Estos resultados indican que la distribución de las longitudes es consistente con la ocurrencia de cuatro modas. De este cuadro puede observarse que la secuencia de valores de p no es estrictamente monótonamente creciente, pero después del valor de p para cuatro modas (0.795) no se obtienen valores significativamente menores al valor de referencia de 0.40. Esto proporciona una evidencia sólida para apoyar la multimodalidad de los datos además de asegurar el número de las modas (4) de la distribución. Por otra parte, como se mencionó anteriormente, esta prueba se sabe que es más bien conservativa (Silverman, 1983) y por tanto puede subestimar el número de modas. Sin embargo, proporciona un valor mínimo confiable para el número de modas de un conjunto de datos (Roeder, 1990).

En los datos de longitud de bagres cuatro modas ocurren para h entre 13.98 y 4.10, lo que sugiere una amplitud de $(13.98+4.10)/2 = 9.04$. El EDK con amplitud 9 se presenta en la Figura 1.30. La naturaleza tetramodal de estos datos se confirma. Es interesante notar que este valor es cercano al empíricamente obtenido al dividir la banda “óptima” de Silverman a la mitad. La amplitud por la prueba de Silverman fue mayor que las obtenidas por validación cruzada (aproximadamente 4 y 6), las cuales, sin embargo, quedan en el intervalo de cuatro modas determinado por el valor de las bandas críticas.

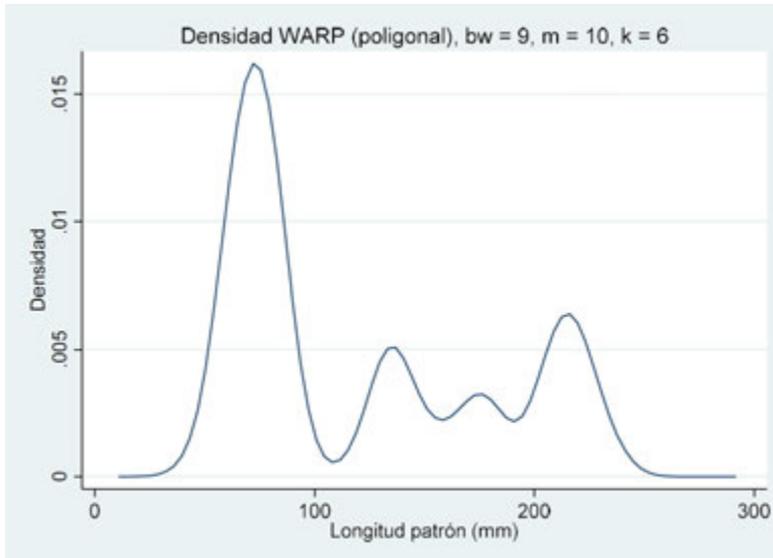


Figura 1.30. EDK con la amplitud de banda recomendada por la prueba de Silverman $(13.98 + 4.10)/2 \approx 9$ para los datos de bagre estuarino *Cathorops melanopus*.

1.10.3 Algunos comentarios adicionales

Como complemento y validación de los programas para esta sección, se incluye la aplicación de la prueba de multimodalidad a tres conjuntos de datos reportados en la literatura para evaluar su grado de modalidad con el procedimiento de Silverman:

- los datos de meteoritos condritas (tomados de Scott, 1992), Cuadro 1.7, tres modas.
- los datos de grosor de timbres postales mexicanos antiguos (tomados de Izenman y Sommer, 1988) Cuadro 1.8, siete modas y
- los datos de velocidad de galaxias (de Roeder, 1990) Cuadro 1.9, tres modas.

Salvo algunas diferencias menores en los valores de amplitudes críticas de banda (debidas a diferencia en el procedimiento para el cálculo de la estimación de densidad por kernel) se obtuvieron las mismas conclusiones que las de los reportes originales.



También se analizó la prueba de multimodalidad para los datos del géiser “Old Faithful” del parque nacional de Yellowstone en los E.U.A. (adaptados de Härdle, 1991). Cuadro 1.10, dos modas.

Hay que notar que para estos conjuntos de datos el número de repeticiones bootstrap para cada banda crítica es de 600, lo que representa el valor más alto reportado para la prueba de Silverman hasta el presente. Para los datos del géiser, el resultado sugiere que la duración de los períodos eruptivos del géiser se distribuye en forma bimodal.

Al disponer las rutinas para esta prueba en AED2020 y haber escrito las rutinas equivalentes en R, a continuación se presentan los resultados comparativos para cada conjunto de datos (Cuadros 1.7 a 1.11). En todos los casos, se llega a la misma conclusión sobre la modalidad de los datos.

Cuadro 1.7. Amplitudes críticas de banda y niveles de significancia estimados para los datos de velocidad de galaxias (Roeder, 1990) $n = 82$.

Semilla	Número de modas	Amplitudes críticas	Valor de p		
			AED2020	Stata	R
12345	1	3007	0.0067 (4)	0.0117 (7)	0.0750 (459)
12347	2	2439	0.0100 (6)	0.0067 (4)	0.0250 (15)
12349	3	924	0.5233 (314)	0.5700 (342)	0.7183 (431)
12351	4	867	0.2167 (130)	0.2467 (148)	0.4083 (245)
12353	5	717	0.2383 (143)	0.2867 (172)	0.4050 (243)
12355	6	659	0.1500 (90)	0.1700 (102)	0.2283 (137)
12357	7	444	0.3800 (228)	0.3533 (212)	0.5117 (307)

NOTA: Los valores de p se obtuvieron de $(\#)/B$, donde $B = 600$ muestras con reemplazo bootstrap de tamaño 82.



Cuadro 1.8. Amplitudes críticas de banda y niveles de significancia estimados para los datos de condritas (Scott, 1992) $n = 22$.

Semilla	Número de modas	Amplitudes críticas	Valor de p		
			AED2020	Stata	R
12345	1	2.40	0.1850 (111)	0.1617 (97)	0.2950 (177)
12347	2	1.83	0.0567 (34)	0.0533 (32)	0.1167 (70)
12349	3	0.69	0.6450 (387)	0.6650 (399)	0.7650 (459)
12351	4	0.47	0.7150 (429)	0.7317 (439)	0.7917 (475)

NOTA: Los valores de p se obtuvieron de $(\#)/B$, donde $B = 600$ muestras con reemplazo bootstrap de tamaño 22.

Cuadro 1.9. Amplitudes críticas de banda y niveles de significancia estimados para los datos de grosor de timbres postales (Izenman & Sommer, 1988) $n = 485$.

Semilla	Número de modas	Amplitudes críticas	Valor de p		
			AED2020	Stata	R
12345	1	0.00669	0.0000 (0)	0.0000 (0)	0.0067 (4)
12347	2	0.00321	0.3367 (202)	0.3400 (204)	0.5900 (354)
12349	3	0.00298	0.0717 (43)	0.0633 (38)	0.2250 (135)
12351	4	0.00281	0.0100 (6)	0.0083 (4)	0.0617 (37)
12353	5	0.00261	0.0033 (2)	0.00033 (2)	0.0233 (14)
12355	6	0.00238	0.0000 (0)	0.0000 (0)	0.0050 (3)
12357	7	0.00149	0.4900 (294)	0.5133 (308)	0.6867 (412)
12359	8	0.00138	0.2833 (170)	0.2583 (155)	0.4933 (296)
12361	9	0.00106	0.6333 (380)	0.6050 (363)	0.8417 (505)

NOTA: Los valores de p se obtuvieron de $(\#)/B$, donde $B = 600$ muestras con reemplazo bootstrap de tamaño 485.



Cuadro 1.10. Amplitudes críticas de banda y niveles de significancia estimados para los datos de duración de erupciones del géyser “Old Faithful” (duración en minutos; Härdle, 1992) $n = 272$.

Semilla	Número de modas	Amplitudes críticas	Valor de p		
			AED2020	Stata	R
12345	1	0.828	0.0000 (0)	0.0000 (0)	0.0000 (0)
12347	2	0.127	0.5467 (328)	0.4983 (299)	0.7367 (442)
12349	3	0.087	0.9283 (557)	0.8967 (538)	0.9850 (591)

NOTA: Los valores de p se obtuvieron de $(\#)/B$, donde $B = 600$ muestras con reemplazo bootstrap de tamaño 272.

Cuadro 1.11. Amplitudes críticas de banda y niveles estimados de significancia para los datos de longitud de bagres (hembras e indeterminados) $n = 183$.

Semilla	Número de modas	Amplitudes críticas	Valor de p		
			AED2020	Stata	R
12345	1	35.54	0.0017 (1)	0.0017 (1)	0.0067 (4)
12347	2	22.49	0.0000 (0)	0.0083 (5)	0.0133 (8)
12349	3	13.99	0.0233 (14)	0.0250 (15)	0.0217 (13)
12351	4	4.10	0.8350 (501)	0.7950 (477)	0.9483 (569)
12353	5	3.92	0.5767 (346)	0.5650 (339)	0.8283 (497)
12355	6	3.49	0.5300 (318)	0.5800 (348)	0.7983 (479)
12357	7	3.24	0.4867 (292)	0.4667 (280)	0.7950 (477)
12359	8	2.95	0.4883 (293)	0.4967 (298)	0.7167 (430)
12361	9	2.86	0.2783 (167)	0.3133 (188)	0.6117 (367)

NOTA: Los valores de p se obtuvieron de $(\#)/B$, donde $B = 600$ muestras con reemplazo bootstrap de tamaño 183.



Se incluyen los comandos para llevar a cabo la prueba de Silverman con R (datos de bagres):

```
> setwd("c:/data")
> library(foreign)
> library(reticulate)
Warning message:
package 'reticulate' was built under R version 3.5.3
> baghsmi<-read.dta("baghsmi.dta")
> source("c:/Rprogs/bandw.R")
> bandw(baghsmi$bodlen)
```

Some practical number of bins and binwidth-bandwidth rules
for univariate density estimation using histograms
frequency polygons (FP) and kernel density estimators

```
=====
Sturges' number of bins                = 8.5157
Oversmoothed number of bins            <= 7.1531
-----
FP oversmoothed number of bins         <= 6.695
=====
Scott's optimal Gaussian binwidth      = 36.9037
Freedman-Diaconis optimal robust binwidth = 36.8126
Terrell-Scott's oversmoothed binwidth  >= 26.8415
Oversmoothed homoscedastic binwidth    >= 39.3183
Oversmoothed robust binwidth           >= 47.9115
-----
FP optimal Gaussian binwidth           = 45.4047
FP oversmoothed binwidth                >= 49.206
=====
Gaussian kernel
=====
Silverman's optimal bandwidth           = 19.0066
Haerdle's better optimal bandwidth     = 22.3856
Scott's oversmoothed bandwidth          = 24.1595
=====
```

```
> source("c:/Rprogs/critiband.R")
> source("c:/Rprogs/warpcdenpy.R")
> critiband(baghsmi$bodlen, bwh=36, bwl=2, st=.01, m=40)
Estimation number = 1   Bandwidth = 36   Number of modes = 1
Estimation number = 2   Bandwidth = 35.99   Number of modes = 1
```



```

...
Estimation number = 47   Bandwidth = 35.54   Number of modes = 1
Estimation number = 48   Bandwidth = 35.53   Number of modes = 2
Estimation number = 49   Bandwidth = 35.52   Number of modes = 2
...
Estimation number = 1352 Bandwidth = 22.49   Number of modes = 2
Estimation number = 1353 Bandwidth = 22.48   Number of modes = 3
Estimation number = 1354 Bandwidth = 22.47   Number of modes = 3
...
Estimation number = 2202 Bandwidth = 13.99   Number of modes = 3
Estimation number = 2203 Bandwidth = 13.98   Number of modes = 4
Estimation number = 2204 Bandwidth = 13.97   Number of modes = 4
...
Estimation number = 3191 Bandwidth = 4.1     Number of modes = 4
Estimation number = 3192 Bandwidth = 4.09   Number of modes = 5
Estimation number = 3193 Bandwidth = 4.08   Number of modes = 5
...
Estimation number = 3209 Bandwidth = 3.92   Number of modes = 5
Estimation number = 3210 Bandwidth = 3.91   Number of modes = 6
Estimation number = 3211 Bandwidth = 3.9    Number of modes = 6
...
Estimation number = 3252 Bandwidth = 3.49   Number of modes = 6
Estimation number = 3253 Bandwidth = 3.48   Number of modes = 7
Estimation number = 3254 Bandwidth = 3.47   Number of modes = 7
...
Estimation number = 3277 Bandwidth = 3.24   Number of modes = 7
Estimation number = 3278 Bandwidth = 3.23   Number of modes = 8
Estimation number = 3279 Bandwidth = 3.22   Number of modes = 8
...
Estimation number = 3306 Bandwidth = 2.95   Number of modes = 8
Estimation number = 3307 Bandwidth = 2.94   Number of modes = 9
Estimation number = 3308 Bandwidth = 2.93   Number of modes = 9
...
Estimation number = 3315 Bandwidth = 2.86   Number of modes = 9
Estimation number = 3316 Bandwidth = 2.85   Number of modes = 10
Estimation number = 3317 Bandwidth = 2.84   Number of modes = 10

```

```

> sd(baghsmi$bodlen)
[1] 59.86215
> source("c:/Rprogs/bootsam.R")

```

Métodos cuantitativos computarizados para biología pesquera



```
> set.seed(12345)
> bsam <- bootsam(baghsmi$bodlen, 35.54, 59.86215, 600)
```

```
> silvtest(bsam, 35.54, 40, 1, 600, 1)
Estimation number = 1   Number of modes = 1
Estimation number = 2   Number of modes = 1
...
Estimation number = 600   Number of modes = 1
```

```
Critical number of modes = 1
```

```
Pvalue = 4 / 600 = 0.006666667
```

```
> set.seed(12347)
> bsam <- bootsam(baghsmi$bodlen, 22.49, 59.86215, 600)
> silvtest(bsam, 22.49, 40, 1, 600, 2)
Estimation number = 1   Number of modes = 2
```

```
...
Estimation number = 600   Number of modes = 2
```

```
Critical number of modes = 2
```

```
Pvalue = 8 / 600 = 0.013333333
```

```
> set.seed(12349)
> bsam <- bootsam(baghsmi$bodlen, 13.99, 59.86215, 600)
> silvtest(bsam, 13.99, 40, 1, 600, 3)
Estimation number = 1   Number of modes = 3
Estimation number = 600   Number of modes = 3
```

```
Critical number of modes = 3
```

```
Pvalue = 13 / 600 = 0.021666667
```

```
> set.seed(12351)
> bsam <- bootsam(baghsmi$bodlen, 4.1, 59.86215, 600)
> silvtest(bsam, 4.1, 40, 1, 600, 4)
Estimation number = 1   Number of modes = 7
...
Estimation number = 600   Number of modes = 6
```



Critical number of modes = 4

Pvalue = 569 / 600 = 0.9483333

```
> set.seed(12353)
> bsam <- bootsam(baghsmi$bodlen,3.92,59.86215,600)
> silvtest(bsam,3.92,40,1,600,5)
Estimation number = 1   Number of modes = 6
...
Estimation number = 600   Number of modes = 7
```

Critical number of modes = 5

Pvalue = 497 / 600 = 0.8283333

```
> set.seed(12355)
> bsam <- bootsam(baghsmi$bodlen,3.49,59.86215,600)
> silvtest(bsam,3.49,40,1,600,6)
Estimation number = 1   Number of modes = 7
...
Estimation number = 600   Number of modes = 7
```

Critical number of modes = 6

Pvalue = 479 / 600 = 0.7983333

```
> set.seed(12357)
> bsam <- bootsam(baghsmi$bodlen,3.24,59.86215,600)
> silvtest(bsam,3.24,40,1,600,7)
Estimation number = 1   Number of modes = 8
...
Estimation number = 600   Number of modes = 9
```

Critical number of modes = 7

Pvalue = 477 / 600 = 0.795

```
> set.seed(12359)
> bsam <- bootsam(baghsmi$bodlen,2.95,59.86215,600)
> silvtest(bsam,2.95,40,1,600,8)
```



```
Estimation number = 1    Number of modes = 8
...
Estimation number = 600  Number of modes = 9

Critical number of modes = 8

Pvalue = 430 / 600 = 0.7166667
> set.seed(12361)
> bsam <- bootsam(baghsmi$bodlen,2.86,59.86215,600)
> silvtest(bsam,2.86,40,1,600,9)
Estimation number = 1    Number of modes = 6
...
Estimation number = 600  Number of modes = 6

Critical number of modes = 9

Pvalue = 367 / 600 = 0.6116667
```

Como puede observarse, los comandos se escribieron siguiendo la misma lógica que las rutinas escritas para Stata y AED2020. Nótese también que al establecer y anotar el valor inicial (semilla) antes de la generación de cada muestra bootstrap es posible replicar exactamente los resultados de la prueba. La conclusión es la misma que la obtenida por las rutinas de Stata y AED2020: los datos de longitud corporal estándar de los bagres en esta muestra tienen una distribución con cuatro modas. Asumiendo periodicidad anual y considerando a los grupos modales como cohortes (Sección 1.11), es posible estimar la función de crecimiento, tema que se verá en el Capítulo 4.

1.11. Caracterización de componentes gaussianos en distribuciones multimodales

1.11.1 Introducción

Las distribuciones mezcladas (aquéllas producto de la mezcla de varias distribuciones Gaussianas o de otro tipo) frecuentemente surgen en datos biológicos y ecológicos. En las ciencias biológico-pesqueras, el análisis de estas mezclas de datos es un tópico de particular importancia, sobre todo al estudiar la biología de los organismos en áreas tropicales o



subtropicales. Una buena parte de las muestras de longitudes de peces en estas regiones exhiben distribuciones con modas múltiples. Desde Petersen (1892) se ha reconocido que estas modas son evidencia de que la muestra contiene una mezcla de individuos de varias generaciones. Algunas especies tienen un período reproductivo corto que ocurre una vez al año. En estos casos, las modas pueden identificarse con una generación en particular (cohorte). La correspondencia entre las modas y los grupos de edad es particularmente clara para los grupos de edades jóvenes aunque se hace menos marcada en los grupos mayores debido al solapamiento de las distribuciones al irse aproximando las generaciones al tamaño máximo de la especie.

El análisis de los datos de frecuencia de talla provee información sobre la edad y el crecimiento que son necesarios para la evaluación del stock. La dificultad de estimar las edades por métodos directos tales como la lectura de estructuras duras o estudios de marca-recaptura explica un poco la importancia del análisis de la distribución de los tamaños. Aun cuando es posible realizar estudios directos para estimar la edad, el análisis de la frecuencia de tallas es un método útil para la validación (Pauly y Morgan, 1987).

Se han propuesto una variedad de métodos, tanto gráficos como analíticos para el análisis de distribuciones mezcladas de frecuencia y su separación en componentes individuales. Estos métodos incluyen el uso de papel en escala probabilística (Harding, 1949; Cassie, 1954), ajuste gráfico a parábola por ensayo y error (Tanaka, 1962), diferencias logarítmicas (Bhattacharya, 1967; Matsumiya, *et al.*, 1991) y máxima verosimilitud (Hasselblad, 1966). Otros procedimientos más sofisticados se han publicado para el caso donde las modas de grupos de edad describen una curva de crecimiento (Macdonald y Pitcher, 1979; Schnute y Fournier, 1980; Liu, *et al.*, 1989). El problema general, cuando el número de modas no se conoce de antemano, permanece como un tópico activo de investigación.

Se dispone actualmente con varios programas computarizados que implementan los métodos arriba citados y que se aplican a una o varias muestras: ELEFAN I (Gayaniilo *et al.* 1989; 1994), LFSA (Sparre, 1987b; Sparre *et al.* 1989), MIX (Macdonald y Green, 1988; 1992) y MULTIFAN (Fournier *et al.* 1990).

Este capítulo presenta programas para el paquete estadístico Stata que permiten la realización del método gráfico-analítico de Bhattacharya (1967) para estimar los componentes Gaussianos en una distribución de frecuencia. En el procedimiento de Bhattacharya, el analista no necesita conocer previamente el número de componentes. Los programas de este capítulo incorporan algunas de las modificaciones y sugerencias de Pauly y Caddy (1985), Sparre y Venema (1992) y Erzini (1990).



La siguiente sección del capítulo da una introducción general al método de Bhattacharya. Posteriormente el método (por medio de los programas) se aplica a los datos de frecuencia de longitud del porgy (Tanaka, 1962). Estos datos son los que originalmente usó el autor para dar el ejemplo práctico y se notará que son adecuados para su aplicación directa.

En la sección que sigue se muestra la aplicación del método a los datos de longitud de hembras e indeterminados del bagre estuarino *Cathorops melanopus* presentado en secciones anteriores.

1.11.2 Método de Bhattacharya

El método de Bhattacharya está diseñado para estimar componentes Gaussianos en distribuciones mezcladas. Su aplicación es directa cuando los componentes están bien separados, pero aún en casos de solapamiento es posible separarlos (Bhattacharya, 1967).

Para entender el método se presenta el caso más simple de un solo componente Gaussiano (Figura 1.31). Este gráfico es un polígono de frecuencia que se produjo extrayendo 10,000 números pseudo-aleatorios con un valor inicial (semilla = 10032020) de una población Gaussiana con media de cero y desviación estándar de 2.0, los cuales se agruparon en clases de amplitud 0.5 en el intervalo de -8.5 a 8 (ver Scott, 1992, Salgado-Ugarte, 2002, Secciones 1.1 y 1.2 de esta obra para detalles acerca del histograma y polígonos de frecuencia). En Stata esto se consigue:

```
. drop _all

. set obs 10000
number of observations (_N) was 0, now 10,000

. set seed 10032020

. gen gausdat=rnormal(0,2)
```

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
gausdat	10,000	.0041417	2.00011	-8.000182	7.534985



```
. twoway__histogram_gen gausdat, start(-8.75) width(.5) gen(h x)

. scatter h x, ms(oh) c(1) xlab(-8(1)8) xtitle("Marcas de clase")
ytitle("Frecuencia")
```

Para obtener las marcas de clase y sus respectivas frecuencias se hizo uso del comando “twoway__histogram_gen” de Stata (Harrison, 2014). Una vez obtenidas estas variables, debe borrarse de la memoria la variable que se agrupó (en este caso “gausdat”) y los valores faltantes y conservar sólo las frecuencias (“h”) y las marcas de clase (“x”):

```
. keep h x

. drop if h==.
(9,968 observations deleted)

. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
x	32	-.25	4.690416	-8	7.5
h	32	312.5	357.0633	1	1083

La Figura 1.32 muestra el gráfico de Bhattacharya para los datos simulados construido con el programa actualizado “bhatapl1.ado”.

```
. bhatapl1 h x, yline(0) ytitle("Diferencias logarítmicas") xtitle("Marcas de
clase")
```

Puede notarse que a excepción de los extremos de la secuencia, el resto de los puntos de la gráfica siguen una tendencia rectilínea con inclinación negativa.

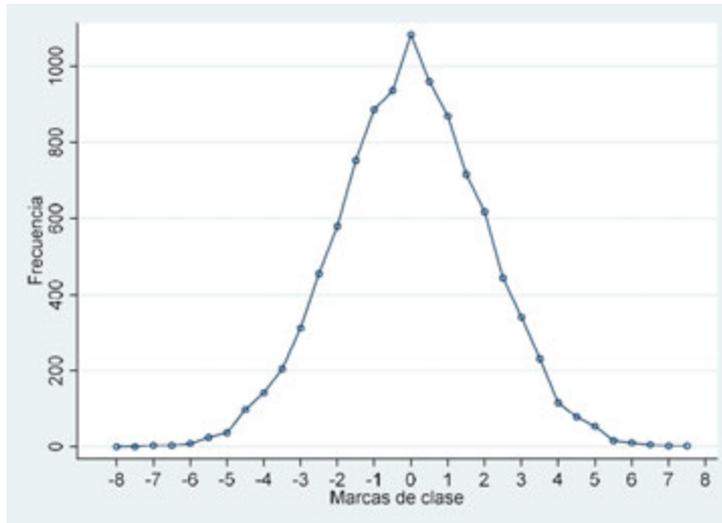


Figura 1.31. Polígono de frecuencia para datos Gaussianos simulados ($\mu = 0$ y $\sigma = 2$) con origen en -9.75 y ancho de intervalo de 0.5 .

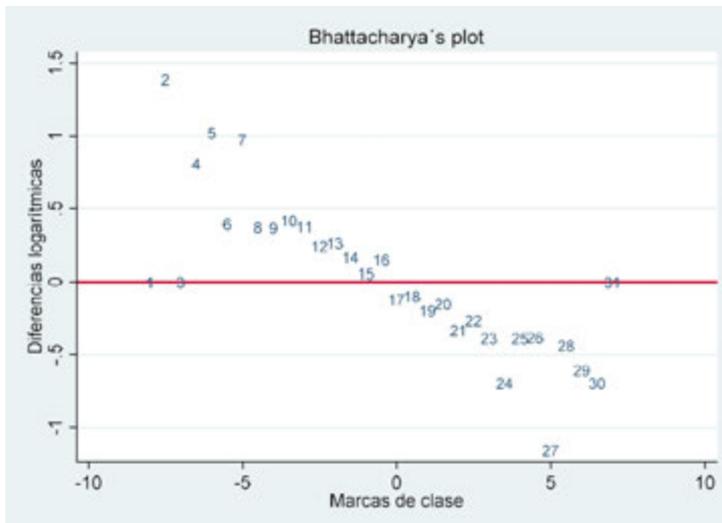


Figura 1.32. Gráfico de Bhattacharya para los datos simulados con un solo componente Gaussiano.



La Figura 1.33 muestra el ajuste por el método para los datos de la observación 11 a la 17 (de las 32 que se originan con los intervalos especificados arriba) por medio del programa actualizado “bhatgauc1.ado”:

```
. bhatgauc1 h x in 11/17, ytitle("Frecuencia") xtitle("Marcas de clase")
R-square = 0.8301          Adj R-square = 0.7962
Mean = -0.0628
s.d. = 1.9228
component size = 9689
```

Además del resultado gráfico, la información numérica muestra que el componente ajustado es muy cercano a la población de la cual la muestra de 10,000 observaciones se extrajo (media de 0.063 y desviación estándar de 1.92) conteniendo 9,689 elementos.

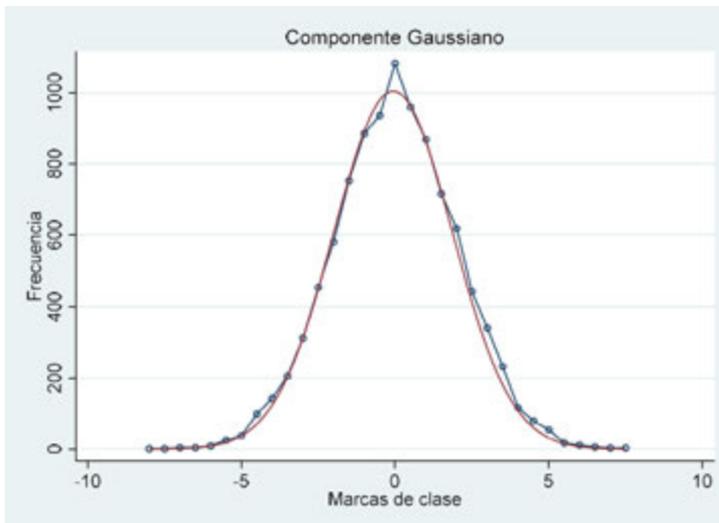


Figura 1.33. Componente Gaussiano ajustado (continua) por el método de Bhattacharya a las frecuencias simuladas observadas (círculos).

Un enfoque más moderno se presenta en la Figura 1.34, que presenta el estimador de densidad por kernel con la amplitud óptima Gaussiana de Silverman.

Métodos cuantitativos computarizados para biología pesquera



```
. drop _all

. set obs 10000
number of observations (_N) was 0, now 10,000

. set seed 10032020

. gen gausdat=rnormal(0,2)

. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
gausdat	10,000	.0041417	2.00011	-8.000182	7.534985

```
. bandwl gausdat
```

Some practical number of bins and binwidth-bandwidth rules for univariate density estimation using histograms, frequency polygons (FP) and kernel density estimators

```
=====
Sturges' number of bins = 14.2877
Oversmoothed number of bins <= 27.1442
```

```
-----
FP oversmoothed number of bins <= 14.9024
=====
```

```
Scott's optimal Gaussian binwidth = 0.3249
Freedman-Diaconis optimal robust binwidth = 0.2473
Terrell-Scott's oversmoothed binwidth >= 0.5723
Oversmoothed homoscedastic binwidth >= 0.3462
Oversmoothed robust binwidth >= 0.3218
```

```
-----
FP optimal Gaussian binwidth = 0.6815
FP oversmoothed binwidth >= 0.7386
=====
```

```
Gaussian kernel (6)
```

```
=====
Silverman's optimal bandwidth = 0.2816
```



```
Haerdle's 'better' optimal bandwidth = 0.3317
Scott's oversmoothed bandwidth = 0.3626
```

```
. warpdenml gau, b(.28) m(10) k(6) ytitle("Densidad") xtitle("Puntos medios")
t1("Densidad WARP (poligonal), bw=0.28, m=10, k=6") gen(den mid)
```

La curva resultante es prácticamente indistinguible de la función Gaussiana poblacional.

Para poder utilizar los programas de caracterización de componentes gaussianos, es necesario transformar los valores de densidad ("den" en la opción "gen") a frecuencia. Esto se logra con una regla de tres. Los puntos medios pueden considerarse como marcas de clase. Como se aclaró antes, se conservan sólo las variables de interés ("frec" y "mid").

```
. gen frec=den*10000/sumden[_N]
(9,836 missing values generated)

. keep frec mid
```

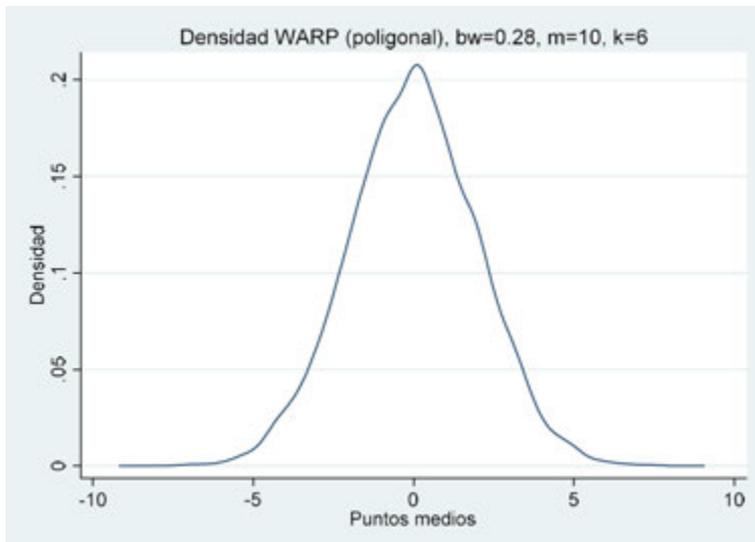


Figura 1.34. Estimador de densidad por kernel con amplitud óptima de banda (Silverman) $h = 0.28$ para datos simulados.

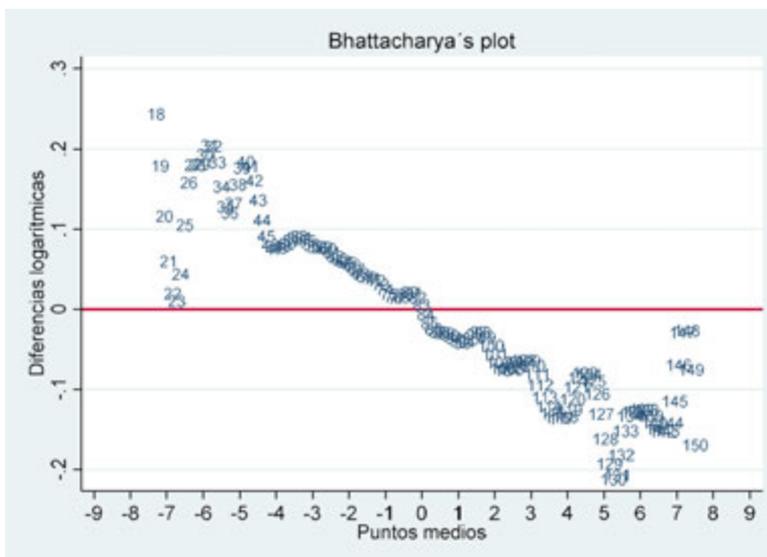


Figura 1.35. Gráfico de Bhattacharya para los valores de frecuencia suavizada del EDK con banda óptima de Silverman, datos simulados con un solo componente Gaussiano.

El gráfico de Bhattacharya para los valores de frecuencia suavizada obtenidos a partir de la densidad (Figura 1.35) muestra una tendencia rectilínea con pendiente negativa muy clara. Cabe aclarar que por medio del selector “in” del comando “scatter” de Stata se graficaron los valores sin tomar en cuenta las colas de la distribución:

```
. bhatapl1 freq mid in 18/150, yline(0) ytitle("Diferencias logarítmicas")
xtitle("Puntos medios") xlab(-9(1)9)
```

Con este gráfico se ajustó un componente Gaussiano que recupera los valores originales de manera muy cercana (Figura 1.36)

```
. bhatgauc1 freq mid in 54/115, ytitle("Frecuencia") xtitle("Puntos medios")
t1("Componente gaussiano")
R-square = 0.9716          Adj R-square = 0.9711
Mean = 0.0230
s.d. = 1.9908
component size = 9950
```

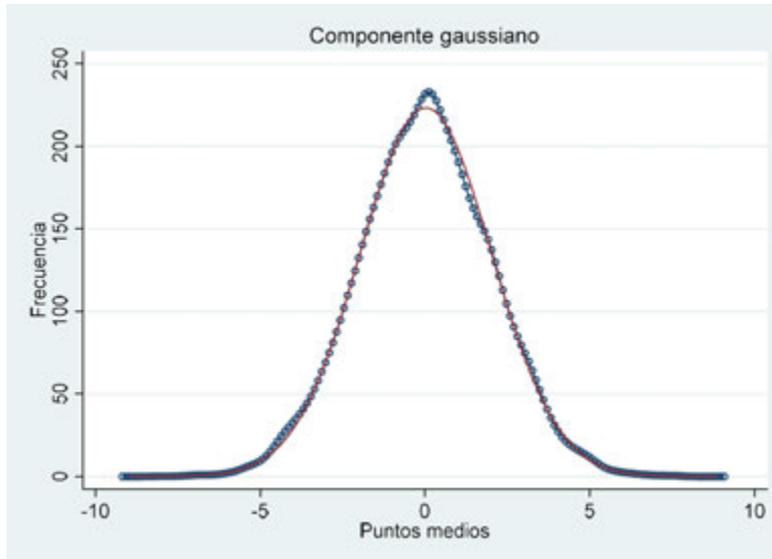


Figura 1.36. Componente Gaussiano ajustado (línea continua) por el método de Bhattacharya a las frecuencias simuladas suavizadas (círculos).

Considérese la curva suave en forma de campana que describe la distribución normal ideal de la cual estos datos se han extraído. Tomemos en cuenta en particular, la derivada de esta curva normal respecto al eje x . La pendiente de la curva es positiva a la izquierda de la moda, cero en la moda y negativa a la derecha. Como Bhattacharya muestra, la derivada logarítmica disminuye aproximadamente de manera lineal. Los puntos en las figuras 1.32 y 1.35 que describen aproximadamente una línea recta con pendiente negativa son las diferencias logarítmicas de frecuencias adyacentes a lo largo de la curva empírica simulada.

El método de Bhattacharya empieza por graficar estas diferencias logarítmicas. Si los componentes Gaussianos están bien separados, el gráfico exhibirá segmentos lineales con pendiente negativa, un segmento por cada componente. Si los componentes se sobrelapan fuertemente, puede ser difícil de detectarlos en su totalidad. En este caso, los componentes dominantes pueden ser estimados y sustraídos de los datos y entonces, los componentes restantes pueden ser reconocidos y estimados por diferencia.



Existen cuatro pasos en el método de Bhattacharya:

- a) Se calculan los logaritmos de valores sucesivos de frecuencia;
- b) Se grafican estas diferencias y los componentes se determinan visualmente. Este paso es la parte gráfica del método.
- c) Se estima por regresión lineal por mínimos cuadrados el segmento lineal correspondiente a cada componente.
- d) Se recuperan los parámetros de cada componente: media, desviación estándar y frecuencia (o proporción) a partir de los ángulos e interceptos descritos por los segmentos lineales de pendiente negativa (por medio de las expresiones de Bhattacharya (1967).

Los pasos a) y b) se llevan a cabo con el programa “bhatapl1.ado”; los dos siguientes c) y d) se realizan con el programa “bhatgauc1.ado”. Cabe mencionar que este programa usa versiones modificadas de las fórmulas originales para recuperar los parámetros.

El método puede aplicarse repetidamente a los residuales para detectar componentes Gaussianos oscurecidos por los componentes dominantes que los sobrelapan.

1.12 Análisis de las frecuencias de longitud del “kidai” (*Dentex tumifrons*)

Se han escrito dos programas actualizados de los originalmente presentados por (Salgado-Ugarte, *et al.* 1994; 2000). Para demostrar estos programas se utilizarán los bien conocidos datos de longitud furcal del pez “kidai” (en japonés) o “yellowback seabream” (en inglés), conocido actualmente con el nombre científico de *Dentex tumifrons* y tomados de Tanaka (1962), que son los datos considerados por Bhattacharya (en parte) y por Tanaka en sus propuestas originales. Las medidas se agruparon en intervalos de 1 mm. El cuadro 1.12 lista los datos y la figura 1.37 despliega el correspondiente polígono de frecuencia.

```
. cd c:/data
c:\data

. use kidai

. scatter freq midpoi

. scatter freq midpoi, ms(oh) c(1) ytitle("Frecuencia (no. de individuos)")
xtitle("Longitud furcal (cm)") xlab(6(6)36)
```



Cuadro 1.12. Datos de frecuencia de longitud furcal “kidai”
Dentex tumifrons (tomados de Tanaka, 1962).

Centro de clase	Frecuencia	Centro de clase	Frecuencia	Centro de clase	Frecuencia
7.5	7	17.5	448	27.5	114
8.5	79	18.5	512	28.5	64
9.5	509	19.5	719	29.5	22
10.5	2240	20.5	673	30.5	0
11.5	2341	21.5	445	31.5	2
12.5	623	22.5	341	32.5	2
13.5	476	23.5	310	33.5	0
14.5	1230	24.5	228	34.5	0
15.5	1439	25.5	168	35.5	1
16.5	921	26.5	140	36.5	0

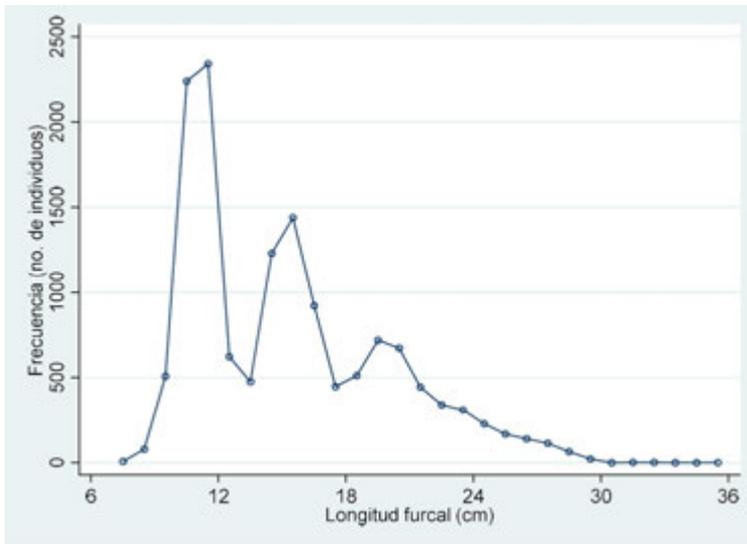


Figura 1.37. Polígono de frecuencia para datos de longitud furcal del “kidai” (*Dentex tumifrons*). Datos tomados de Tanaka (1962).



El polígono de frecuencia exhibe claramente tres modas y una cola derecha que va disminuyendo paulatinamente hacia la derecha. Las modas para las tallas menores son muy pronunciadas. La cola suave hacia tallas mayores puede ser el resultado del solapamiento de componentes mayores de talla.

El comando actualizado `bhataplt1.ado` lleva a cabo los dos primeros pasos del método de Bhattacharya (cálculo de diferencias logarítmicas y graficación contra las marcas de clase). El resultado se presenta en la Figura 1.38. Es posible utilizar el selector “in” del lenguaje de comandos del Stata para limitar la gráfica hacia alguna región dentro de los datos. En este gráfico se distinguen al menos cuatro componentes Gaussianos: uno dominante de las observaciones 3 a 5, uno más de las observaciones 7 a 9, uno de 12 a 14 y otro en el intervalo de 20 a 22. Pudiera haber un componente adicional solapado por sus vecinos en el intervalo de 15 a 19.

```
. bhataplt1 freq midpoi, yline(0) ylab(-2(1)2) xlab(6(6)36)  
ytitle("Diferencias logarítmicas de frecuencia") xtitle("Longitud furcal  
(cm)") ttitle(".")
```

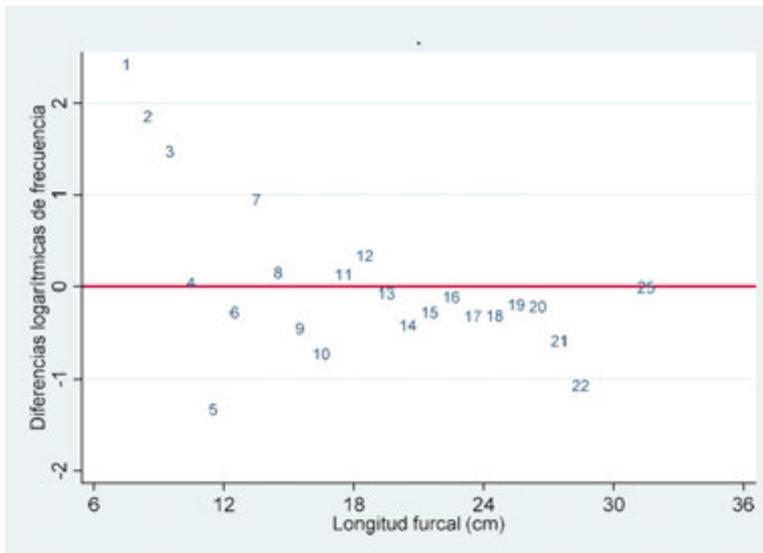


Figura 1.38. Gráfico de Bhattacharya para los datos de longitud furcal del “kidai” (*Dentex tumifrons*).



El tercer paso (que se ha modificado para dejar de ser gráfico y volverse analítico) es estimar las pendientes e interceptos de los segmentos lineales correspondientes a cada componente y derivar los parámetros de los componentes Gaussianos de estas estimaciones. El programa (actualizado) `bhatgauc1.ado` hace estos cálculos para determinar uno por uno cada componente. El selector “in” puede utilizarse para restringir el intervalo de estimación donde la tendencia lineal sea lo más clara posible.

Aplicado al ejemplo del “kidai” y seleccionando de la marca de clase 3 a la 5, se obtiene la Figura 1.39 en la que se distingue un ajuste adecuado de un componente Gaussiano a la primera moda observada en el polígono de frecuencia. Además de la gráfica, el comando proporciona datos numéricos que se presentan en el Cuadro 1.13. Con los valores presentados en este cuadro podemos evaluar el componente ajustado. En general debe buscarse el coeficiente de determinación más cercano a la unidad y la desviación estándar lo menor posible (para evitar el solapamiento con otros componentes).

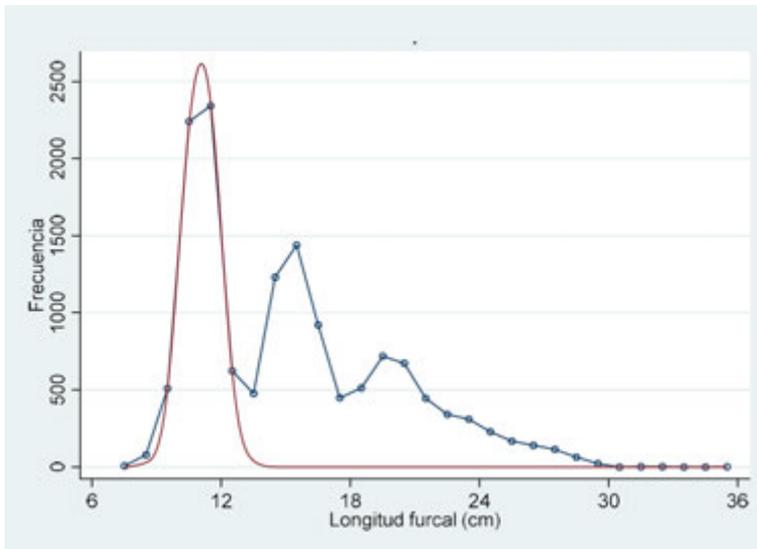


Figura 1.39. Ajuste del primer componente Gaussiano para los datos de longitud furcal del “kidai” (*Dentex tumifrons*).



Cuadro 1.13. Resultados analíticos del primer componente Gaussiano.

$r^2 =$	0.9998	r^2 ajustada =	0.9996
Media =	11.0480	Desv. Estándar =	0.8443
Tamaño =	5783		

```
. bhatgauc1 freq midpoi in 3/5, xlab(6(6)36) ytitle("Frecuencia")
xtitle("Longitud furcal (cm)") t1title(".")
```

Si el ajuste es satisfactorio, es posible usar la opción “generate” del programa para crear una nueva variable con el primer componente estimado. Los demás componentes evidentes se ajustan de manera equivalente y un resumen de los puntos utilizados se presenta en el Cuadro 1.14:

Cuadro 1.14. Resumen de puntos considerados para determinación de componentes Gaussianos.

Número de componente	Intervalo de datos utilizado
1	3 a 5
2	7 a 9
3	12 a 14
5	20 a 22

Deliberadamente se ha numerado al último componente como “5” ya que en la región anterior de longitud y después del “3”, la disposición ruidosa de las diferencias logarítmicas sugiere la existencia de un componente enmascarado por sus vecinos. Para calcular los parámetros de esta curva Gaussiana menor entre dos cercanas, se suman los componentes generados en los pasos anteriores (1 hasta 3 más el 5), se restan estos valores ajustados a la frecuencia original condicionando la diferencia a valor mínimo de 0 y se repite el gráfico de Bhattacharya para estos residuos poniendo énfasis en el intervalo de solapamiento.



```

. bhatgauc1 freq midpoi in 3/5, xlab(6(6)36) ytitle("Frecuencia")
xtitle("Longitud furcal (cm)") t1title(".") gen(gaucol)
R-square = 0.9998          Adj R-square = 0.9996
Mean = 11.0480
s.d. = 0.8443
component size = 5783

. bhatgauc1 freq midpoi in 7/9, xlab(6(6)36) ytitle("Frecuencia")
xtitle("Longitud furcal (cm)") t1title(".") gen(gauc2)
R-square = 0.9939          Adj R-square = 0.9878
Mean = 15.3153
s.d. = 1.1971
component size = 4498

. bhatgauc1 freq midpoi in 12/14, xlab(6(6)36) ytitle("Frecuencia")
xtitle("Longitud furcal (cm)") t1title(".") gen(gauc3)
R-square = 0.9980          Adj R-square = 0.9960
Mean = 19.8759
s.d. = 1.6295
component size = 2988

. bhatgauc1 freq midpoi in 20/22, xlab(6(6)36) ytitle("Frecuencia")
xtitle("Longitud furcal (cm)") t1title(".") gen(gauc5)
R-square = 0.9937          Adj R-square = 0.9875
Mean = 26.5694
s.d. = 1.5229
component size = 533

. gen sumgl235 = gaucol+gauco2+gauco3+gauco5

. gen freca = freq-sumgl235

. replace freca=0 if freca<0
(13 real changes made)

. bhataplt1 freca midpoi, yline(0) ylab(-2(1)2) xlab(6(6)36)
ytitle("Diferencias logarítmicas de frecuencia") xtitle("Longitud furcal
(cm)") t1title(".")

```



El gráfico de Bhattacharya para estas frecuencias residuales ajustadas muestra claramente la existencia de un componente Gaussiano de los centros de clase 16 a 18 (Figura 1.40). Se utiliza de nuevo el programa bhatgauc1.ado para ajustar componente a estos valores. Los resultados finales se presentan en el Cuadro 1.15.

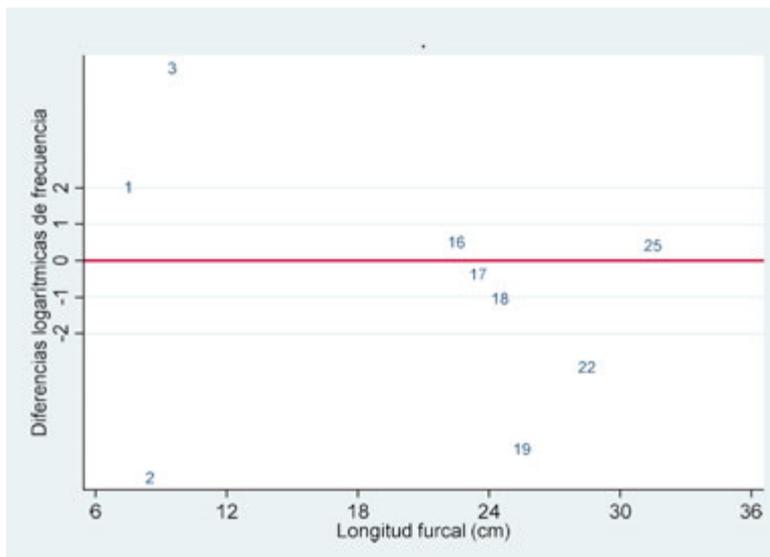


Figura 1.40. Gráfico de Bhattacharya para los valores residuales de frecuencia.

```
. bhatgauc1 freca midpoi in 16/18, xlab(6(6)36) ytitle("Frecuencia")
xtitle("Longitud furcal (cm)") tlttitle(".") gen(gauco4)
R-square = 0.9932          Adj R-square = 0.9865
Mean = 23.6219
s.d. = 1.1370
component size = 638
```



Cuadro 1.15. Comparación de resultados con otros autores.

Parámetro	Método	Componentes				
		1	2	3	4	5
Medias	A	11.048	15.315	19.876	23.622	26.569
	B	11.05	15.32	19.85	23.58	26.82
	C	11.02	15.33	19.85	23.46	26.92
	D	10.99	15.26	19.84	23.50	26.82
	E	11.03	15.28	19.86	23.62	26.62
	F	11.0	15.3	19.7	23.5	27.2
	G	11.00	15.30	19.70	23.45	27.26
Desviaciones	A	0.844	1.197	1.629	1.137	1.523
Estándar	B	0.844	1.161	1.412	1.212	1.443
	C	0.76	1.15	1.32	1.29	1.54
	D	0.8	1.2	1.4	1.2	1.4
	E	0.81	1.13	1.60	1.07	1.47
	F	0.87	1.14	1.43	1.55	1.19
	G	0.83	1.10	1.39	1.62	1.12
	Proporciones	A	0.4005	0.3115	0.2069	0.0442
B		0.4072	0.3110	0.1860	0.0642	0.0316
C		0.4049	0.3164	0.1788	0.0693	0.0307
D		0.4007	0.3194	0.1873	0.0598	0.0328
E		0.4065	0.3067	0.2087	0.0420	0.0361
F		0.411	0.305	0.183	0.077	0.024
G		0.4106	0.3056	0.1787	0.0827	0.0224

A = Salgado-Ugarte, *et al.* (1994; 2000,2020); B = Buchanan-Wollaston (1929); C = Cassie (1954); D = Tanaka (1962); E = Bhattacharya (1964); F = Akamine (1985); G = Macdonald y Green (1988).

La gráfica con todos los componentes generados y caracterizados se presenta en la Figura 1.41.

```
. scatter freq midpoi, ms(oh) c(1) lpattern(dash) xlab(6(6)36) xscale(range(6 36)) || mspline gaucol midpoi, bands(40) xaxis(2) xscale(range(6 36) axis(2)) xlab(11.05 15.32 19.88 23.62 26.57, axis(2)) xtitle(" ", axis(2)) || mspline
```



```
gauco2 midpoi, bands(40) || mspline gauco3 midpoi, bands(40) || mspline gauco4  
midpoi, bands(40) || mspline gauco5 midpoi, bands(40) ytitle("Frecuencia")  
xtitle("Longitud furcal (cm)") legend(label(1 "FO") label(2 "CG1") label(3  
"CG2") label(4 "CG3") label(5 "CG4") label(6 "CG5") ring(0) pos(1) cols(1))  
xline(11.05 15.32 19.88 23.62 26.57)
```

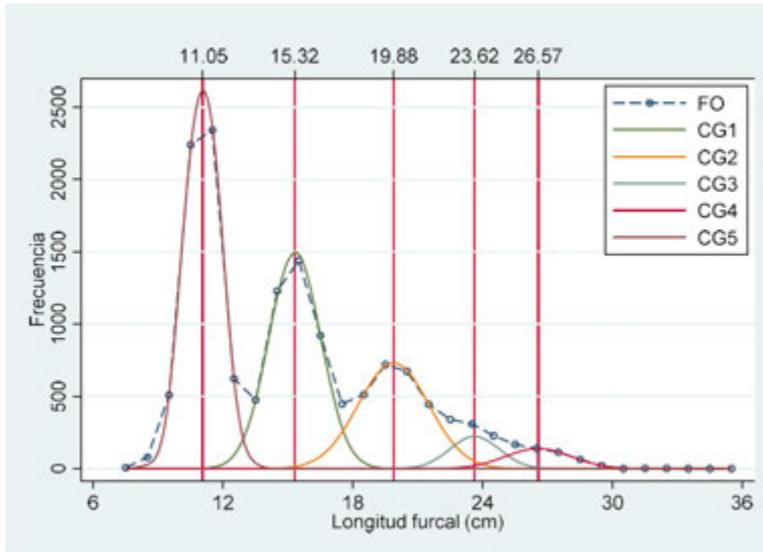


Figura 1.41. Frecuencia observada y componentes Gaussianos estimados.
Se indican los valores medios de cada componente (FO: Frecuencia
observada; CG(*n*) Componente Gaussiano *n*).

Como una medida de la semejanza de la suma de los componentes Gaussianos con la frecuencia mezclada original se llevó a cabo un análisis de correlación. La Figura 1.42 presenta la gráfica de la suma de Gaussianos contra la frecuencia original incluyendo (tan sólo como referencia) la recta de regresión. El valor para el índice de correlación es muy cercano a 1 (0.9986, $P < 0.001$) lo que indica que no puede rechazarse que ambas distribuciones sean iguales.



```
. pwcorr sumgauc freq, sig
```

	sumgauc	freq
sumgauc	1.0000	
freq	0.9986	1.0000

```
. scatter sumgauc freq, ms(Oh) || lfit sumgauc freq , legend(label(1 "Observados")
label(2 "Calculados") ring(0) pos(11) cols(1)) ytitle("Suma de componentes
Gaussianos") xtitle("Frecuencia Observada Suavizada")
```

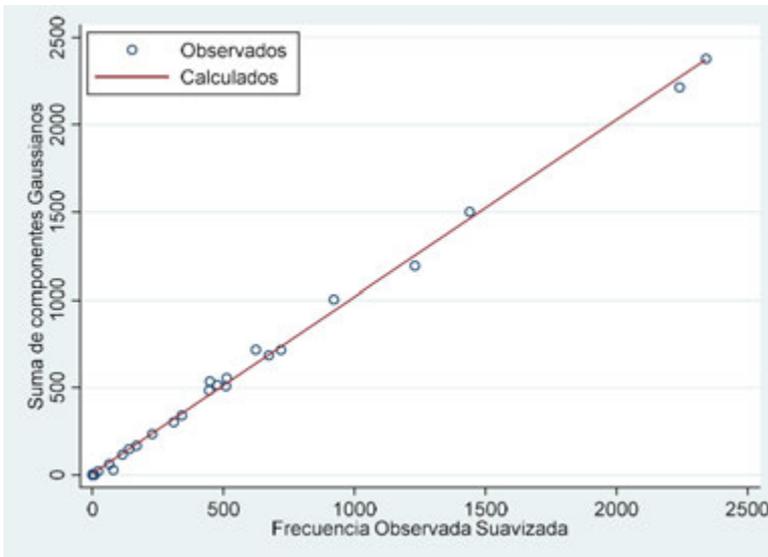


Figura 1.42. Correlación entre suma de componentes Gaussianos y frecuencia observada suavizada.



El cuadro 1.15 contiene una comparación de los valores obtenidos mediante los programas presentados en este capítulo con aquellos reportados para el mismo conjunto de datos por otros autores. Las diferencias son mínimas y se pueden explicar por la diferente implementación de métodos ó programas.

1.13 Análisis de frecuencia de tallas de *Cathorops melanopus* por EDKs

En esta sección presentaremos la caracterización de componentes gaussianos para representar a la distribución multimodal de las tallas de las hembras e indeterminados del bagre estuarino *Cathorops melanopus* capturados en noviembre de 1990 utilizados anteriormente.

El Cuadro 1.16 presenta un resumen de todas las amplitudes de banda obtenidas para este conjunto de datos por medio de los métodos y reglas discutidos en secciones anteriores.

Cuadro 1.16. Amplitudes de banda por los métodos revisados (biponderada transformada a Gaussiana por factor 0.381 del Cuadro 1.5).

Critero	Kernel Gaussiano	Kernel Biponderado
Regla sobresuavizada	24.15	
Regla óptima Silverman	19.01	
Regla ½ de óptima	9.5	
VC Mínimos Cuadrados	3.36	10
VC Sesgada	5.04	15
Prueba de Silverman	9	

Si consideramos que la Prueba Bootstrap de Multimodalidad de Silverman es un procedimiento robusto ($b = 9.0$) y que coincide aproximadamente con la regla empírica de la mitad del valor óptimo de Silverman ($b = 19.01/2 \approx 9.5$), la utilizaremos para identificar a los componentes gaussianos individuales que juntos dan como resultado la distribución multimodal observada. Para esto utilizamos los siguientes comandos. Para evitar interferencia con los datos originales únicamente conservaremos en la memoria de Stata a



las variables de densidad (den9) y los puntos de cálculo (mid9). Como el archivo original contiene mayor número de datos que los resultantes de la opción “generate” (en breve “gen”) se eliminan los valores faltantes generados (con el comando “drop” y la condición “if den==.” Donde el punto (“.”) representa a los valores faltantes (“missing values”). Para mostrar la opción “at” del comando de Stata “kdensity” utilizaremos a este comando para (re) calcular los valores de densidad.

```
. warpdenm1 bodlen if sexa==2 | sexa==3, b(9) m(10) k(6) gen(den9 mid9)
. kdensity bodlen if sexa==2 | sexa==3, bw(9) at(mid9) gauss (gen md9 kd9)
. keep den9 mid9
. drop if den9==.
(562 observations deleted)
```

Ya que los programas “warpdenm1.ado” y “kdensity” producen resultados de frecuencia en escala de densidad, antes de aplicar el método de Bhattacharya (“bhataplt1.ado”) debemos de transformarlos a frecuencia, la cual, al provenir de un estimador de densidad por kernel, podemos considerar como “suavizada” por no dar valores enteros. Para esto se hace una regla de tres, en donde la suma acumulada de las densidades obtenidas equivale al tamaño de la muestra ($n = 183$) de este conjunto de datos y cada valor de densidad es una frecuencia. En Stata esto se obtiene con los comandos (Figura 1.43):

```
. gen sumden=sum(kd9)
. gen frec = 183*kd9/sumden[_N]
. bhataplt1 frec mid, yline(0)
. bhataplt1 frec mid9, yline(0) ytitle("Diferencias logaritmicas de
frecuencia") xtitle("Puntos medios de Longitud patrón (mm)") t1title(".")
```

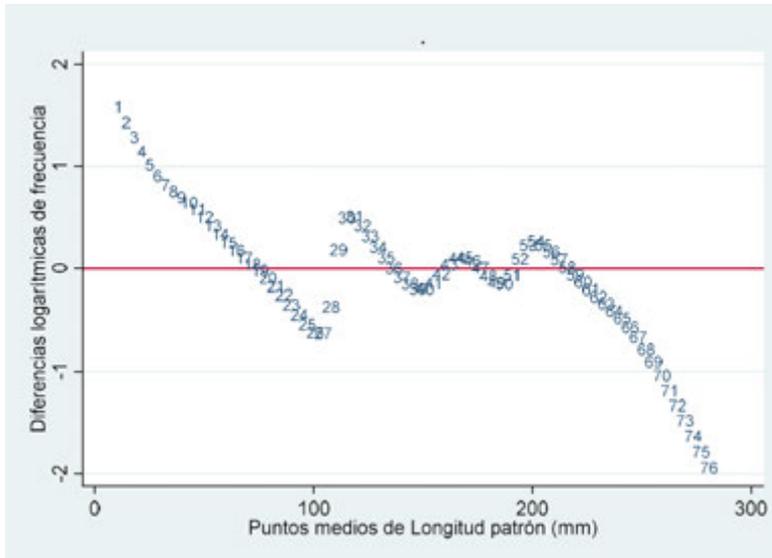


Figura 1.43. Gráfico de Bhattacharya para longitud patrón (mm) de hembras e indeterminados de *Cathorops melanopus*

En la Figura 1.43 claramente se aprecian varios segmentos lineales con pendiente negativa que indican la presencia de un componente gaussiano. Los comandos siguientes listan el procedimiento a seguir para la caracterización de cada uno de los componentes.

```
. bhatgaucl freq mid9 in 10/18, ytitle("Frecuencia") xtitle("Puntos medios de
Longitud patrón (mm)") t1("Componente Gaussiano") gen(gauco1)
R-square = 0.9983          Adj R-square = 0.9981
Mean = 75.4190
s.d. = 12.9791
component size = 97
```

```
. bhatgaucl freq mid9 in 31/36, ytitle("Frecuencia") xtitle("Puntos medios de
Longitud patrón (mm)") t1("Componente Gaussiano") gen(gauco2)
R-square = 0.9989          Adj R-square = 0.9986
Mean = 138.8662
s.d. = 11.2625
```



```

component size = 26

. bhatgauc1 freq mid9 in 47/48, ytitle("Frecuencia") xtitle("Puntos medios de
Longitud patrón (mm)") t1("Componente Gaussiano") gen(gauco3)
R-square = 1.0000          Adj R-square =      .
Mean = 178.5755
s.d. = 12.9284
component size = 19

. bhatgauc1 freq mid9 in 57/69, ytitle("Frecuencia") xtitle("Puntos medios de
Longitud patrón (mm)") t1("Componente Gaussiano") gen(gauco4)
R-square = 0.9893          Adj R-square = 0.9883
Mean = 219.4601
s.d. = 12.8259
component size = 38
    
```

El intervalo de observaciones especificado mediante el selector “in” fue determinado después de hacer algunos ajustes considerando que los puntos centrales de cada tendencia rectilínea con pendiente negativa representan de mejor manera al componente gaussiano y de preferencia se escoge el intervalo con menor desviación estándar para evitar un exceso de solapamiento de los componentes individuales. Si no se toma en cuenta lo anterior, la suma de los componentes será mayor que la frecuencia observada.

```
. gen sumgauc=gauco1+gauco2+gauco3+gauco4
```

La Figura 1.44 presenta el resumen gráfico de los pasos realizados antes. Esta gráfica presenta la frecuencia suavizada del EDK con amplitud de banda de 9, cada uno de los componentes gaussianos y su suma.

```

. scatter freq mid9, ms(oh) c(1) xaxis(2) xlabel(0(50)300, axis(2))
xscale(range(0 300) axis(2)) xtitle(" ", axis(1)) || mspline sumgauc mid9,
bands(40) || mspline gauco1 mid9, bands(40) xaxis(1) xlabel(75.4 138.6 178.5
219.4) xscale(range(0 300) axis(1)) || mspline gauco2 mid9, bands(40)
|| mspline gauco3 mid9, bands(40) || mspline gauco4 mid9, bands(40)
legend(cols(1) pos(2) ring(0) label(1 "FO") label(2 "SCG") label(3 "CG1")
label(4 "CG2") label(5 "CG3") label(6 "CG4")) ytitle("Frecuencia suavizada" ")
xtitle("Longitud patrón (mm)", axis(2)) xline(75.4 138.8 178.5 219.4)
    
```

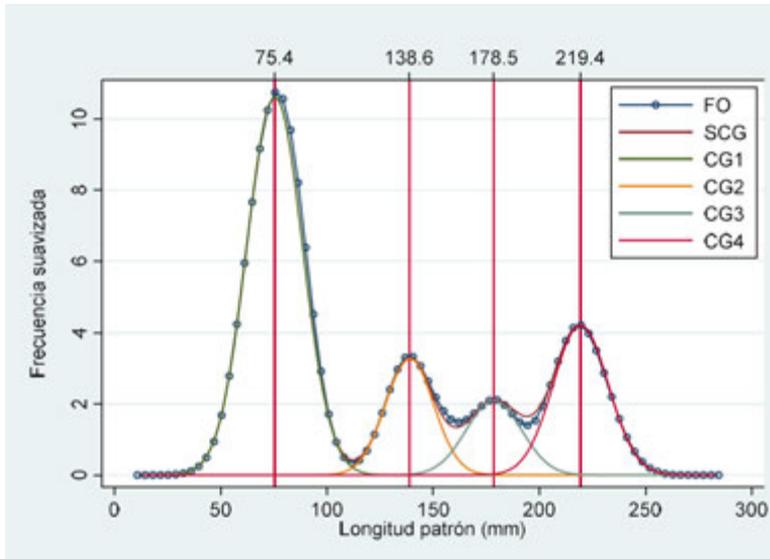


Figura 1.44. Frecuencia observada, suma de componentes Gaussianos (CGs) y CGs estimados. Se indican los valores medios de cada componente (FO: Frecuencia observada; SCG: suma de componentes gaussianos; CG(*n*) Componente Gaussiano *n*).

Como en el ejemplo anterior la correlación entre la suma de los componentes Gaussianos con la frecuencia mezclada original se muestra gráficamente en la Figura 1.45 que incluye como referencia la recta de regresión. El valor para el índice de correlación es muy cercano a 1 (0.9994, $P < 0.001$) lo que indica que, como en el caso anterior, no puede rechazarse que ambas distribuciones sean iguales.

```
. pwcorr sumgauc frec, sig
```

	sumgauc	frec
sumgauc	1.0000	
frec	0.9984	1.0000
	0.0000	



```
. scatter sumgauc frec, ms(Oh) || lfit sumgauc frec , legend(label(1
"Observados") label(2 "Calculados") ring(0) pos(11) cols(1)) ytitle(Suma de
Componentes Gaussianos) xtitle("Frecuencia Observada Suavizada")
```

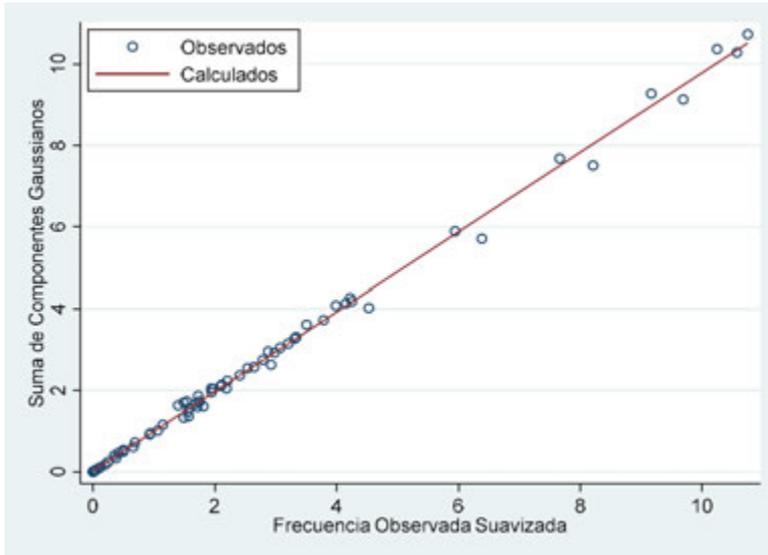


Figura 1.45. Correlación entre suma de componentes Gaussianos y frecuencia observada suavizada.

1.14 Notas acerca de los cálculos

- Debido al uso de los logaritmos, cuando las frecuencias son cero, se generan valores faltantes. Pudiera agregarse un valor constante pequeño (por ejemplo 1/6) a las frecuencias con muchos ceros.
- El programa `bhatgauc.ado` utiliza el comando `"regress"` de Stata para estimar la línea de las diferencias logarítmicas para un componente Gaussiano en particular, sea el componente k . La media (μ_k) y la desviación estándar (σ_k) del k componente son estimados de acuerdo a las siguientes expresiones (Pauly y Caddy, 1985; Sparre *et al.* 1989, respectivamente):



$$\mu_k = 0.5d - a_k / b_k$$

y

$$\hat{\sigma}_k = \sqrt{-d/b_k}$$

Donde d es la amplitud de intervalo, a y b son el intercepto y la pendiente, respectivamente de la línea estimada por mínimos cuadrados.

- El tamaño del componente k es estimado mediante la expresión (Bhattacharya, 1967):

$$\hat{N}_k = \sum y / \sum \hat{P}$$

Donde y son los valores de frecuencia y P son los valores ajustados correspondientes de probabilidad Gaussiana. La sumatoria se efectúa sobre el intervalo especificado por el selector "in". En contraste, Bhattacharya (1967) sólo consideró dos puntos en la estimación del tamaño del componente. Si la tendencia lineal es clara, el componente Gaussiano tiene poco solapamiento con otros y este procedimiento produce estimaciones confiables de N_k . Si se incluyen puntos que se desvían de la línea, los componentes estimados no se ajustarán adecuadamente a las frecuencias observadas.

1.15 Mezcla finita de distribuciones

Como se ha señalado, las distribuciones multimodales pueden considerarse como una mezcla de componentes (por ejemplo Gaussianos) individuales que pueden ajustarse a un modelo paramétrico de mezclas gaussianas (normales). Estos modelos pertenecen a la clase de distribuciones finitas mezcladas descritas en detalle por McLachlan y Peel (2000), (Everitt y Hothorn, 2006).

El interés por ajustar un modelo paramétrico a una distribución mezclada se inició por el trabajo de Karl Pearson a fines del siglo antepasado (1890s). Motivado por la sugerencia de la existencia de dos componentes gaussianos en la distribución de tamaño relativo del frente con la longitud del cuerpo de 1000 cangrejos colectados en la Bahía de Nápoles por parte de su colega Weldon (1893). Pearson enfrentó la tarea por medio del método de los momentos estadísticos, lo que lo llevó a resolver un polinomio de grado 9, una tarea "heroica" por la cantidad inmensa de labor de cálculo (Everitt y Hothorn, 2006).



Desde el punto de vista de la Biología Pesquera, y como se ha señalado antes, el análisis de las distribuciones multimodales de los tamaños de los organismos capturados ha sido de interés primordial desde que apareció el histograma (Pearson, 1891; 1894; Riaño-Rufilanchas, 2017) y lo aplicó Petersen (1892) al estudio de los tamaños de las capturas. Para caracterizar los componentes Gaussianos considerados como parte de una mezcla por máxima verosimilitud, puede citarse la propuesta de Hasselblad (1966) (NORMSEP), el método computarizado del programa MIX (McDonald y Green, 1988) y su reciente actualización, el paquete “mixdist” para R (Macdonald y Du, 2018).

En la actualidad, con la disponibilidad de computadoras veloces y programas adecuados, esta tarea se lleva a cabo con el algoritmo EM (Expectation-Maximization) con la finalidad de optimizar una función de máxima verosimilitud para modelos de mezclas finitas como lo demostraron McLachlan y Basford en 1988 (McLachlan y Peel, 2000).

Siguiendo a Everitt y Hothorn (2006), la distribución tetramodal de las tallas de bagre está dada explícitamente por:

$$f(x) = p_1\phi(x, \mu_1, \sigma_1^2) + p_2\phi(x, \mu_2, \sigma_2^2) + p_3\phi(x, \mu_3, \sigma_3^2) + p_4\phi(x, \mu_4, \sigma_4^2)$$

Esta distribución tiene 15 parámetros por estimar, pero se intentará simplificar haciendo que se estimen las proporciones de cada componente en función de la primera. En R, la función de verosimilitud se puede escribir:

```
> logL <- function(param, x) {
+ d1 <- dnorm(x, mean = param[2], sd = param[3])
+ d2 <- dnorm(x, mean = param[4], sd = param[5])
+ d3 <- dnorm(x, mean = param[6], sd = param[7])
+ d4 <- dnorm(x, mean = param[8], sd = param[9])
+ -sum(log( param[1]*d1 + (1-param[1])*0.31*d2 + (1-param[1])*0.24*d3 +
(1-param[1])*0.44*d4)) }
```

Los valores de proporción en la función “logL” (estimados en referencia al tamaño del componente 1) y los iniciales para la búsqueda iterativa (“startparam”) fueron tomados de los resultados del método de Bhattacharya ajustados. Posteriormente se usa el optimizador “optim” para encontrar los valores de los 15 parámetros que minimicen verosimilitud logarítmica negativa.



```
> startparam <- c(p = 0.5, mu1=72, sd1=13, mu2=135, sd2=11, mu3= 175, sd3=13,
mu4=216, sd4=12)
> opp <- optim(startparam, logL, x = baghsmi$bodlen, method = "L-BFGS-B",
lower = c(0.01, rep(1,8)), upper = c(0.99, rep(300, 8)))
> opp
$`par`
p      mu1      sd1      mu2      sd2      mu3      sd3      mu4      sd4
0.5464392 75.7649186 9.3721949 139.6922568 7.6476720 177.1524240
8.7660986 219.4507007 9.2381461
```

Además de este ajuste directo, es posible utilizar alguno de los paquetes disponibles para la estimación de componentes en distribuciones mezcladas. Everitt y Hothorn (2016) presentan dos “mclust” y “flexmix”.

A continuación se presenta el uso del paquete “mclust” (Fraley, *et al.* 2005; Scrucca, *et al.* 2016) para R. En la versión 5 del paquete se usa la rutina para ajustar la densidad (“densityMclust”):

```
> library(mclust)
> mclres <- densityMclust(baghsmi$bodlen)
fitting ...

=====| 100%

> summary(mclres, parameters=TRUE)
-----
Density estimation via Gaussian finite mixture modeling
-----

Mclust E (univariate, equal variance) model with 6 components:

log-likelihood  n df      BIC      ICL
-872.1099 183 12 -1806.734 -1946.794

Mixing probabilities:
      1      2      3      4      5      6
0.14984714 0.16875554 0.22784542 0.15149201 0.09397897 0.20808091

Means:
      1      2      3      4      5      6
```



69.30554 76.64205 79.35161 140.09746 177.96255 219.53012

Variances:

	1	2	3	4	5	6
	72.13399	72.13399	72.13399	72.13399	72.13399	72.13399

De manera un poco inesperada, el método ajusta a la distribución tetramodal con seis componentes con la misma varianza. Estos resultados se discutirán más adelante.

El otro paquete para R es “flexmix”, cuya funcionalidad se describe en Leisch (2004). La mezcla de cuatro distribuciones gaussianas puede ajustarse usando:

```
> library("flexmix")
Loading required package: lattice
Warning message:
package 'flexmix' was built under R version 3.5.3

> fl <- flexmix(bodlen~1, data=baghsmi, k=4)

> parameters(fl)
              Comp.1      Comp.2      Comp.3      Comp.4
coef.(Intercept) 140.172843 219.295732 177.73997 75.760021
sigma             8.098604   9.430247   7.47533  9.396054

> summary(fl)

Call:
flexmix(formula = bodlen ~ 1, data = baghsmi, k = 4)

      prior size post>0 ratio
Comp.1 0.152  28     33 0.848
Comp.2 0.210  38     47 0.809
Comp.3 0.091  17     28 0.607
Comp.4 0.546 100    100 1.000

'log Lik.' -871.8918 (df=11)
AIC: 1765.784   BIC: 1801.088
```

Como en el caso anterior, estos resultados se compararán más adelante.



En las versiones recientes de Stata, se dispone también de un comando para ajustar distribuciones de frecuencia por mezcla finita (“fmm”). A continuación su aplicación para la distribución tetramodal del bagre estuarino. Primeramente se hace un ajuste preliminar

```
. fmm 4: regress bodlen  
(se omiten resultados)
```

se copia la matriz e(b) con los parámetros estimados y se modifican los valores de la matriz b, copia de e(b), cambiándolos por los obtenidos por Bhattacharya, medias y varianzas (desviaciones estándar al cuadrado):

```
. matrix b = e(b)  
  
. matrix b[1,5]=76  
  
. matrix b[1,6]=139  
  
. matrix b[1,7]=179  
  
. matrix b[1,8]=220  
  
. matrix b[1,9]=13^2  
  
. matrix b[1,10]=11^2  
  
. matrix b[1,11]=13^2  
  
. matrix b[1,12]=12^2
```

Una vez modificados los valores de la matriz, se repite el comando de ajuste, pero especificando que se usen los valores modificados de la matriz b:

```
. fmm 4, from(b): regress bodlen  
(se omiten líneas)
```

```
Finite mixture model                Number of obs      =  
183  
Log likelihood = -871.88967  
-----
```



	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----					
1.Class	(base outcome)				
-----+-----					
2.Class					
_cons	-1.276224	.2168303	-5.89	0.000	-1.701203
	-1.8512441				
-----+-----					
3.Class					
_cons	-1.795766	.2792237	-6.43	0.000	-2.343034
	-1.248498				
-----+-----					
4.Class					
_cons	-.9547463	.19087	-5.00	0.000	-1.328845
	-.580648				
-----+-----					

Class : 1
 Response : bodlen
 Model : regress

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----					
bodlen					
_cons	75.76002	.9370392	80.85	0.000	73.92346
	77.59658				
-----+-----					
var(e.bodlen)	87.80332	12.41811			66.54642
	115.8503				
-----+-----					

Class : 2
 Response : bodlen
 Model : regress

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
--	-------	-----------	---	------	----------------------

Métodos cuantitativos computarizados para biología pesquera



Interval]

```
-----+-----
bodlen      |
      _cons |   140.1979   1.671382   83.88   0.000   136.9221
143.4738
-----+-----
var(e.bodlen) |    65.60777   21.06048                      34.97153
123.0824
-----+-----
```

```
Class      : 3
Response   : bodlen
Model      : regress
```

```
-----+-----
            |      Coef.   Std. Err.      z    P>|z|      [95% Conf.
Interval]
-----+-----
bodlen      |
      _cons |   177.7604    2.12119   83.80   0.000   173.603
181.9179
-----+-----
var(e.bodlen) |    54.36773   32.52708                      16.83004
175.6294
-----+-----
```

```
Class      : 4
Response   : bodlen
Model      : regress
```

```
-----+-----
            |      Coef.   Std. Err.      z    P>|z|      [95% Conf.
Interval]
-----+-----
bodlen      |
      _cons |   219.2822    1.602307  136.85   0.000   216.1418
222.4227
-----+-----
var(e.bodlen) |    88.70535   22.97251                      53.39609
147.3636
-----+-----
```



Para recuperar el tamaño de los componentes se recuperan los valores de proporción para cada componente. Esto se logra con:

```
. estat lcprob
```

Latent class marginal probabilities Number of obs = 183

```
-----
```

Class	Delta-method			
	Margin	Std. Err.	[95% Conf. Interval]	
1	.5464483	.0368013	.4738495	.6171239
2	.1525079	.0270893	.1066093	.213446
3	.0907106	.0226779	.0549991	.1460266
4	.2103333	.0304561	.1567896	.2761733

```
-----
```

El comando despliega información de los pasos iterativos del ajuste (omitidos) y finalmente un resumen estadístico de cada componente, incluyendo errores estándar, pruebas de significancia de los estimadores e intervalos de confianza.

Para graficar la distribución ajustada se utiliza el comando posterior “predict”

```
. predict den2, density marginal
```

Y con la variable generada se construye la gráfica de frecuencia observada con la calculada (Figura 1.46)

```
. kdensity bodlen , w(9) gauss lpattern(dash) addplot(line den2 bodlen, sort)
ytitle("Densidad") xtitle("Longitud patrón (mm)") legend(label(1 "Densidad
Observada") label(2 "Densidad Estimada por 'fmm'") pos(1) ring(0)) title("
") note(" ")
```

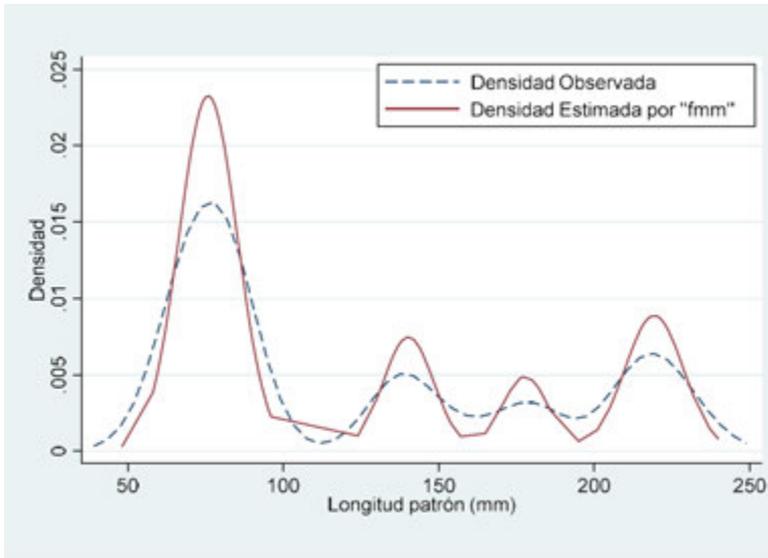


Figura 1.46. Comparación de la densidad observada (línea punteada) con la densidad estimada por el comando “fmm” (Finite Mixture Model).

En lenguaje de programación de Stata se escribió el comando actualizado “gaussgen1.ado” que permite generar un componente gaussiano especificando los puntos medios (marcas de clase), la media, la desviación estándar y el número de elementos que contiene el componente. De esta forma, para los componentes obtenidos al especificar la función de verosimilitud y optimizarla por R se generan los componentes y se hace su gráfica:

```
. gaussgen1 mid9 75.76 9.37 100 cg1logL

. gaussgen1 mid9 139.69 7.65 26 cg2logL

. gaussgen1 mid9 177.15 8.77 20 cg3logL

. gaussgen1 mid9 219.45 9.24 37 cg4logL

. scatter frec mid9, ms(oh) c(1) lpattern(dash) tltitle("Función log-
verosimilitud") || mspline cg1logL mid, bands(40) || mspline cg2logL mid,
```



```
bands(40) || mspline cg3logL mid, bands(40) || mspline cg4logL mid, bands(40)
legend(label(1 "FO") label(2 "CG1") label(3 "CG2") label(4 "CG3") label(5
"CG4") cols(1) pos(1) ring(0) ytitle("Frecuencia") xtitle("Puntos medios de
Longitud patrón (mm)"))
```

Cuadro 1.17. Comparación de parámetros de componentes gaussianos obtenidos por los métodos arriba presentados para datos de *Cathorops melanopus* (n = 183).

Estimador	Bhatta.	McLust	logL	flexmix	fmm
CG1					
Media	75.42	75.10	75.76	75.76	75.76
Des. Est.	12.98	8.29	9.37	9.40	9.37
n	97.00	100.00	100.00	100.00	100.00
P	0.53	0.55	0.55	0.55	0.55
CG2					
Media	138.87	140.10	139.69	140.17	140.20
Des. Est.	11.26	8.29	7.65	8.10	8.10
n	26.00	28.00	26.00	28.00	28.00
P	0.14	0.15	0.14	0.15	0.15
CG3					
Media	178.57	177.96	177.15	177.74	177.76
Des. Est.	12.93	8.29	8.77	7.48	7.37
n	19.00	17.00	20.00	17.00	17.00
P	0.10	0.09	0.11	0.09	0.09
CG4					
Media	219.46	219.53	219.45	219.30	219.28
Des. Est.	12.82	8.29	9.24	9.43	9.42
n	38.00	38.00	37.00	38.00	38.00
P	0.21	0.21	0.20	0.21	0.21

Como se puede apreciar, el resultado no se ajusta muy bien a la frecuencia observada, pero puede irse modificando para mejorar la estimación (Figura 1.47).



Lo mismo se aplica para los resultados del paquete “mclust”.

```
. gaussgen1 mid9 69.3 72.13^.5 27 cg1mcl
. gaussgen1 mid9 76.64 72.13^.5 31 cg2mcl
. gaussgen1 mid9 79.35 72.13^.5 42 cg3mcl
. gaussgen1 mid9 140.1 72.13^.5 28 cg4mcl
. gaussgen1 mid9 178 72.13^.5 17 cg5mcl
. gaussgen1 mid9 219.5 72.13^.5 38 cg6mcl

. scatter frec mid9, ms(oh) c(1) lpattern(dash) tlttitle("Mclust") || mspline
cg1mcl mid, bands(40) || mspline cg2mcl mid, bands(40) || mspline cg3mcl mid,
bands(40) || mspline
> cg4mcl mid, bands(40) || mspline cg5mcl mid, bands(40) || mspline cg6mcl
mid, bands(40) legend(label(1 "FO") label(2 "CG1") label(3 "CG2") label(4
"CG3") label(5 "CG4") label
> (6 "CG5") label(7 "CG6") cols(1) pos(1) ring(0)) ytitle("Frecuencia")
xtitle("Puntos medios de Longitud patrón (mm)")
```

Puede apreciarse que el procedimiento, basado en el criterio de información Bayesiano (BIC) produce una descomposición algo extraña de la frecuencia de longitudes de los organismos indeterminados (Figura 1.48). Para mejorar la estimación, se promediaron las medias y se sumaron los tamaños de cada componente para unificarlos. Los comandos y la gráfica resultante se presentan a continuación:

```
. di (69.3+76.64+79.35)/3
75.096667

. di 27+31+42
100

. gaussgen1 mid9 75.09 72.13^.5 100 cg1amcl

. scatter frec mid9, ms(oh) c(1) lpattern(dash) tlttitle("Mclust ajustado") ||
mspline cg1amcl mid, bands(40) || mspline cg4mcl mid, bands(40) || mspline
cg5mcl mid, bands(40) |
> | mspline cg6mcl mid, bands(40) legend(label(1 "FO") label(2 "CG1")
label(3 "CG2") label(4 "CG3") label(5 "CG4") cols(1) pos(1) ring(0))
ytitle("Frecuencia") xtitle("Puntos m
> edios de Longitud patrón (mm)")
```

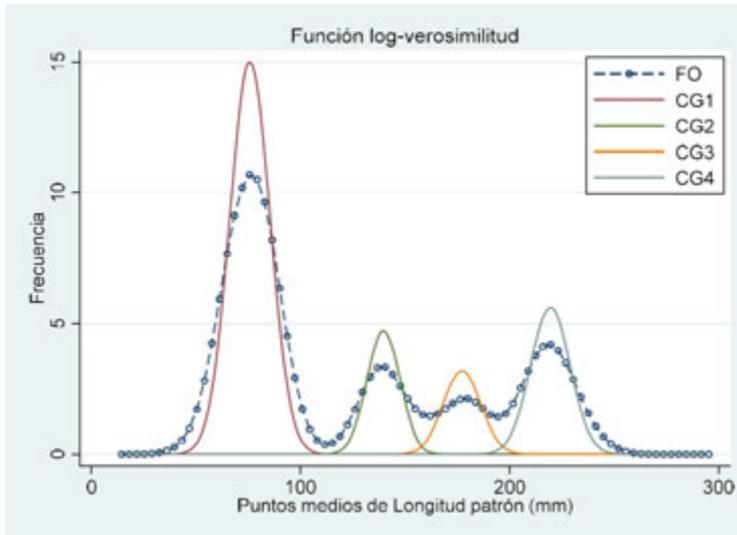


Figura 1.47. Componentes gaussianos ajustados por especificación de la función logarítmica negativa de verosimilitud.

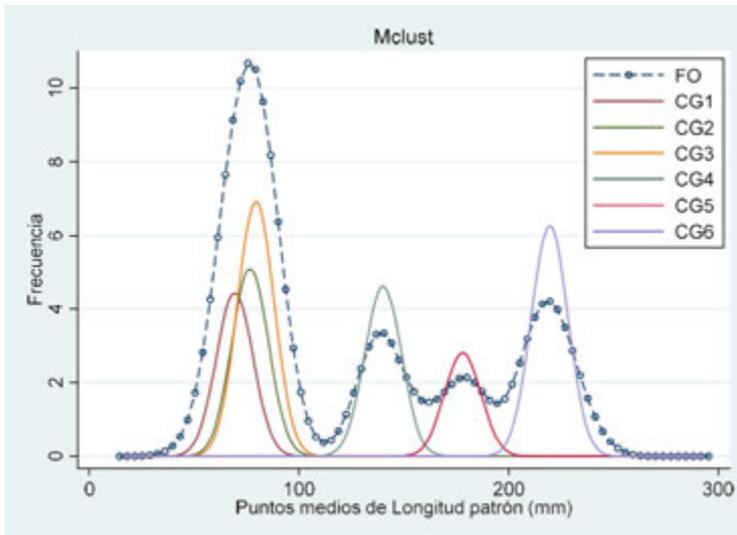


Figura 1.48. Componentes gaussianos ajustados por el paquete “mclust” para R.

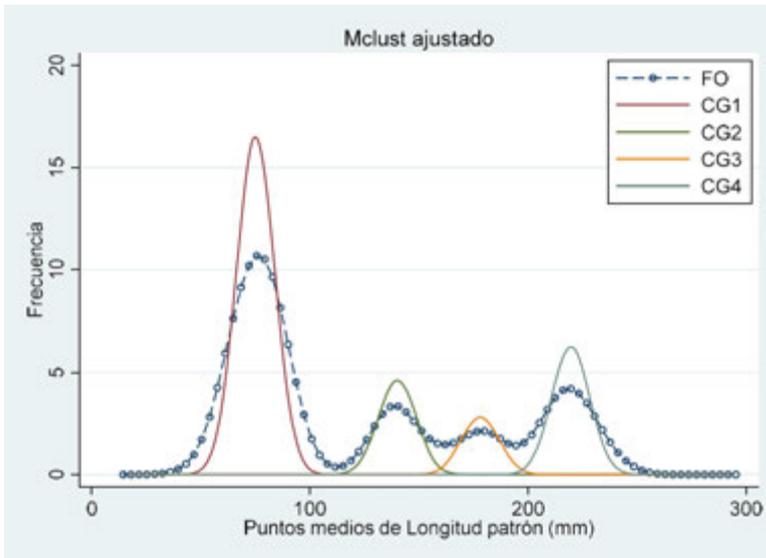


Figura 1.49. Componentes gaussianos caracterizados por el paquete “mclust” para R con los tres primeros componentes unificados.

El resultado parece ahora más adecuado (Figura 1.49).

Para los componentes estimados por el paquete “flexmix” para R (Figura 1.50):

```
. gaussgen1 mid9 75.76 9.4 100 cg1FM
. gaussgen1 mid9 140.17 8.1 28 cg2FM
. gaussgen1 mid9 177.74 7.48 17 cg3FM
. gaussgen1 mid9 219.3 9.43 38 cg4FM

. scatter frec mid9, ms(oh) c(1) lpattern(dash) tlttitle("Flex Mix") ||
mspline cg1FM mid, bands(40) || mspline cg2FM mid, bands(40) || mspline cg3FM
mid, bands(40) || mspline cg4FM mid, bands(40) legend(label(1 "FO") label(2
"CG1") label(3 "CG2") label(4 "CG3") label(5 "CG4") cols(1) pos(1) ring(0))
ytittle("Frecuencia") xtittle("Puntos medios de Longitud patrón (mm)")
```

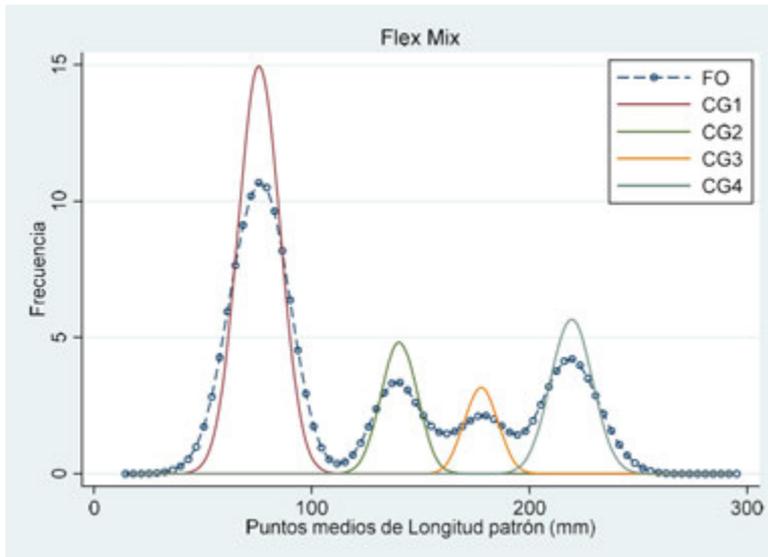


Figura 1.50. Componentes gaussianos ajustados por el paquete “flexmix” para R.

Finalmente, para los componentes estimados por el comando “fmm” de Stata (Figura 1.51):

```
. gaussgen1 mid9 75.76 9.37 100 cg1fmm

. gaussgen1 mid9 140.2 8.1 28 cg2fmm

. gaussgen1 mid9 177.76 7.37 17 cg3fmm

. gaussgen1 mid9 219.28 9.42 38 cg4fmm

. scatter frec mid9, ms(oh) c(1) lpattern(dash) ttitle("Stata fmm") ||
mspline cg1fmm mid, bands(40) || mspline cg2fmm mid, bands(40) || mspline
cg3fmm mid, bands(40) || mspline cg4fmm mid, bands(40) legend(label(1 "FO")
label(2 "CG1") label(3 "CG2") label(4 "CG3") label(5 "CG4") cols(1) pos(1)
ring(0)) ytitle("Frecuencia") xtitle("Puntos medios de Longitud patrón (mm)")
```

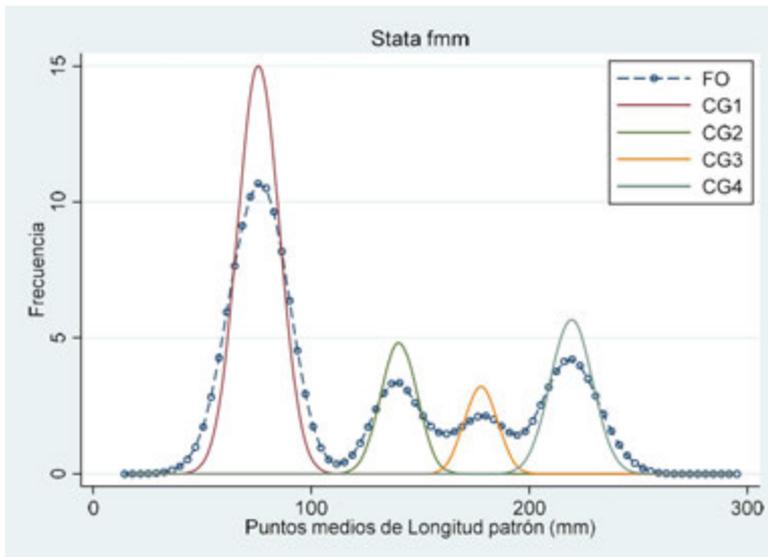


Figura 1.51. Componentes gaussianos ajustados por el comando “fmm” de Stata.

Como puede apreciarse del Cuadro 1.17 y las Figuras, los valores medios son muy semejantes. La diferencia se presenta en los valores de la desviación estándar de los componentes. Todas las desviaciones estimadas por los procedimientos de máxima verosimilitud son semejantes entre sí pero menores a las obtenidas por Bhattacharya.

En esta sección se decidió presentar el uso básico de los métodos y funciones basados en la máxima verosimilitud para el análisis de distribuciones mezcladas. Seguramente estas diferencias pueden hacerse menores y mejorar el ajuste de los componentes gaussianos a la frecuencia observada mediante especificaciones más precisas a los comandos.

Ajustes adicionales a estos procedimientos y el uso del paquete “mixdist” (Macdonald y Du, 2018) se presentarán en una monografía adicional posterior.

En el Apéndice 8 se presenta el procedimiento de ajuste de distribuciones gaussianas por medio de la máxima verosimilitud de la distribución multinomial (Haddon, 2001).



1.16 Comparación de dos distribuciones de frecuencia (Kolmogorov-Smirnov)

La prueba de Kolmogorov (1933)-Smirnov (1933) es un método dentro de la Estadística No Paramétrica (Siegel y Castelan, 1995) que sirve para comparar dos distribuciones de frecuencia de variables continuas y es sensible a diferencias en la localización (nivel), dispersión, asimetría, entre otras propiedades de la distribución de variables cuantitativas (Sokal y Rohlf, 1981).

En general, esta prueba no es muy poderosa para diferencias en las colas de las distribuciones pero es poderosa para hipótesis alternativas que involucran multimodalidad o acumulaciones en los datos (StataCorp, 2019), lo que la hace particularmente importante para el análisis de distribuciones de frecuencia multimodales comunes en los datos biológico-pesqueros.

El método tradicional considera los cuadros de frecuencia de las dos distribuciones. La prueba se basa en las diferencias absolutas de frecuencia acumulada de las dos muestras. La diferencia absoluta máxima se compara con los valores críticos calculados de acuerdo a las siguientes expresiones:

$$D_{\alpha} = k_{\alpha} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

donde

$$k_{\alpha} = \sqrt{\frac{1}{2} \left[-\ln\left(\frac{\alpha}{2}\right) \right]}$$

Estos valores críticos pueden ser consultados también en Tablas (Sokal y Rohlf, 1981). Como en otras pruebas estadísticas, si el valor calculado excede al valor crítico de Tablas o calculado, se rechaza la hipótesis nula de que las muestras provienen de distribuciones con la misma distribución.

Sokal y Rohlf (1981) presentan la comparación de dos cuadros de frecuencia de las longitudes de salmones juveniles (*Oncorhynchus kisutch*) registrados en trampas río abajo (Shapovalov y Taft, 1954). Este ejemplo se reproduce aquí con la información del Cuadro 1.18.



Cuadro 1.18. Prueba aproximada para dos distribuciones de frecuencia (tamaño de muestra grande) Longitud de salmón plateado a dos periodos temporales (tomados de Sokal y Rohlf, 1981).

(1) Longitud	(2) LSC	(3) f_1	(4) F_1	(5) f_2	(6) F_2	(7) F_1/n_1	(8) F_2/n_2	(9) D
60	62.5	1	1		0	0.002	0.000	0.002
65	67.5	4	5		0	0.012	0.000	0.012
70	72.5	33	38		0	0.090	0.000	0.090
75	77.5	94	132		0	0.314	0.000	0.314
80	82.5	125	257	1	1	0.610	0.003	0.608
85	87.5	98	355	5	6	0.843	0.017	0.826
90	92.5	44	399	6	12	0.948	0.034	0.913
95	97.5	18	417	5	17	0.990	0.049	0.942
100	102.5	3	420	16	33	0.998	0.095	0.903
105	107.5	1	421	25	58	1.000	0.166	0.834
110	112.5		421	45	103	1.000	0.295	0.705
115	117.5		421	57	160	1.000	0.458	0.542
120	122.5		421	87	247	1.000	0.708	0.292
125	127.5		421	65	312	1.000	0.894	0.106
130	132.5		421	20	332	1.000	0.951	0.049
135	137.5		421	5	337	1.000	0.966	0.034
140	142.5		421	5	342	1.000	0.980	0.020
145	147.5		421	5	347	1.000	0.994	0.006
150	152.5		421	1	348	1.000	0.997	0.003
155	157.5		421		348	1.000	0.997	0.003
160	162.5		421	1	349	1.000	1.000	0.000

En este cuadro se han incluido las marcas de clase, los límites superiores de clase, las frecuencias absolutas y las frecuencias acumuladas de muestras de individuos juveniles capturados en dos eventos de captura diferentes. Las diferencias absolutas se realizan con la frecuencia acumulada como proporción del tamaño de cada muestra (valor absoluto de



columna 7 – columna 8). Se ha indicado la diferencia absoluta máxima (0.942) la cual es contrastada con los valores críticos para dos valores de significancia ($\alpha = 0.05$ y $\alpha = 0.01$) los cuales se calculan:

$$k_{.05} = \sqrt{\frac{1}{2} \left[-\ln\left(\frac{0.05}{2}\right) \right]} = 1.3581 \quad \text{y} \quad k_{.01} = \sqrt{\frac{1}{2} \left[-\ln\left(\frac{0.01}{2}\right) \right]} = 1.6276$$

$$D_{.05} = 1.3581 \sqrt{\frac{421 + 349}{421(349)}} = 0.0983 \quad \text{y} \quad D_{.01} = 1.6276 \sqrt{\frac{421 + 349}{421(349)}} = 0.1178$$

Como 0.942 es \gg 0.1178, no se acepta la hipótesis nula y se concluye que las dos muestras no provienen de poblaciones con la misma distribución.

Como ejemplo adicional se incluye a continuación la comparación de las distribuciones de frecuencia de talla de dos muestras del bagre estuarino *Cathorops melanopus* obtenidas en dos fechas distintas (marzo y noviembre de 1980) (Salgado-Ugarte, 1985):

Cuadro 1.19. Prueba aproximada para dos distribuciones de frecuencia (tamaño de muestra grande) Longitud del bagre estuarino *Cathorops melanopus* en dos periodos diferentes, marzo y noviembre (Salgado-Ugarte, 1985).

(1) Longitud	(2) <i>L</i> _{sc}	(3) <i>f</i> ₁	(4) <i>F</i> ₁	(5) <i>f</i> ₂	(6) <i>F</i> ₂	(7) <i>F</i> ₁ / <i>n</i> ₁	(8) <i>F</i> ₂ / <i>n</i> ₂	(9) <i>d</i>
35	40	0	0	0	0	0	0	0.000
45	50	0	0	1	1	0	0.0034	0.003
55	60	0	0	11	12	0	0.0411	0.041
65	70	0	0	33	45	0	0.1541	0.154
75	80	0	0	38	83	0	0.2842	0.284
85	90	0	0	16	99	0	0.339	0.339
95	100	12	12	2	101	0.0632	0.3459	0.283
105	110	27	39	1	102	0.2053	0.3493	0.144
115	120	33	72	3	105	0.3789	0.3596	0.019
125	130	12	84	21	126	0.4421	0.4315	0.011
135	140	5	89	36	162	0.4684	0.5548	0.086



Cuadro 1.19. Prueba aproximada para dos distribuciones de frecuencia (tamaño de muestra grande) Longitud del bagre estuarino *Cathorops melanopus* en dos periodos diferentes, marzo y noviembre (Salgado-Ugarte, 1985) (continuación).

(1) Longitud	(2) L_{sc}	(3) f_1	(4) F_1	(5) f_2	(6) F_2	(7) F_1/n_1	(8) F_2/n_2	(9) d
145	150	8	97	18	180	0.5105	0.6164	0.106
155	160	17	114	11	191	0.6	0.6541	0.054
165	170	12	126	13	204	0.6632	0.6986	0.035
175	180	5	131	15	219	0.6895	0.75	0.061
185	190	5	136	5	224	0.7158	0.7671	0.051
195	200	9	145	8	232	0.7632	0.7945	0.031
205	210	8	153	20	252	0.8053	0.863	0.058
215	220	11	164	26	278	0.8632	0.9521	0.089
225	230	16	180	10	288	0.9474	0.9863	0.039
235	240	4	184	4	292	0.9684	1	0.032
245	250	5	189	0	292	0.9947	1	0.005
255	260	1	190	0	292	1	1	0.000
265	270	0	190	0	292	1	1	0.000

La diferencia máxima absoluta de 0.339 es mayor que los valores críticos ($D_{.05} = 0.127$ y $D_{.01} = 0.157$) por lo que, como en el caso anterior, no se acepta la hipótesis nula y se concluye que ambas muestras no provienen de poblaciones con la misma distribución.

Actualmente, los programas estadísticos pueden llevar a cabo esta comparación sin necesidad de construir un cuadro de frecuencia, sino con los datos directamente. En Stata, el procedimiento es:

```
. use ksbagretallas
```

```
. tab sample
```

```

sample |      Freq.      Percent      Cum.
-----+-----
      1 |         190         39.42         39.42

```



6	292	60.58	100.00
-----+-----			
Total	482	100.00	

```
. ksmirnov bodlen, by(sample)
```

Two-sample Kolmogorov-Smirnov test for equality of distribution functions

Smaller group	D	P-value
1:	0.0528	0.527
6:	-0.3406	0.000
Combined K-S:	0.3406	0.000

Note: Ties exist in combined dataset;
there are 169 unique values out of 482 observations.

El comando “ksmirnov” lleva a cabo la prueba de igualdad de dos distribuciones. El primer renglón del cuadro con resultados, prueba la hipótesis de que la muestra 1 (marzo) contiene valores menores que la muestra 2 (noviembre); la diferencia más grande es de 0.0528 y su valor de $P = 0.527$ es no significativo.

El renglón siguiente prueba que la muestra de marzo tiene valores mayores que la muestra de noviembre y su diferencia (-0.3406) es altamente significativa, por lo que esta hipótesis alternativa (marzo > noviembre) no se rechaza.

Finalmente, el tercer renglón prueba la igualdad de las distribuciones de las dos poblaciones de las cuales provienen las muestras y la máxima diferencia (0.3406) es altamente significativo ($P < 0.001$) por lo no se acepta la hipótesis de igualdad de distribuciones de las poblaciones. Nótese que la conclusión es la misma que la basada en el cuadro de frecuencia y los valores calculados son similares (0.339 y 0.3406).

Con R, podemos hacer la misma prueba con unos pocos pasos para definir a las variables continuas (nombradas para facilidad “x” y “y”):

```
> setwd("c:/data")
> library(foreign)
> ksbt <- read.dta("ksbagretallas.dta")
> mar <- subset(ksbt, sample==1)
> nov <- subset(ksbt, sample==6)
> x <- mar$bodlen
> y <- nov$bodlen
> ks.test(x,y)
```



Two-sample Kolmogorov-Smirnov test

```
data: x and y
D = 0.34063, p-value = 5.022e-12
alternative hypothesis: two-sided
```

Warning message:

```
In ks.test(x, y) : p-value will be approximate in the presence of ties
```

Como podemos ver, los valores de D y su significancia son los mismos, con el mensaje de que los datos tienen valores “empatados”, es decir valores con la misma magnitud repetidos.

Las razones de las diferencias pueden apreciarse claramente en los EDKs de cada muestra construidos con la mitad del valor óptimo de Silverman para kernel Gaussiano.

En la Figura 1.52 correspondiente al mes de marzo predominan organismos mayores a 100 mm de longitud patrón, mientras que los individuos de noviembre (Figura 1.53) manifiestan claramente la aparición de organismos pequeños (menores a 100 mm de longitud patrón) reclutados a la fracción poblacional capturada y resultado del evento reproductivo de los meses anteriores (Salgado-Ugarte, 1985).

```
. warpdnml bodlen if sample==1, b(7) m(10) k(6) xlab(0(100)300) ytitle("Densidad")
t1("Densidad WARP (poligonal), bw=7, m=10, k=6") xtitle("Longitud patrón (mm)")
xline(100)
```

```
. warpdnml bodlen if sample==6, b(8) m(10) k(6) xlab(0(100)300) ytitle("Densidad")
t1("Densidad WARP (poligonal), bw=7, m=10, k=6") xtitle("Longitud patrón (mm)")
xline(100)
```

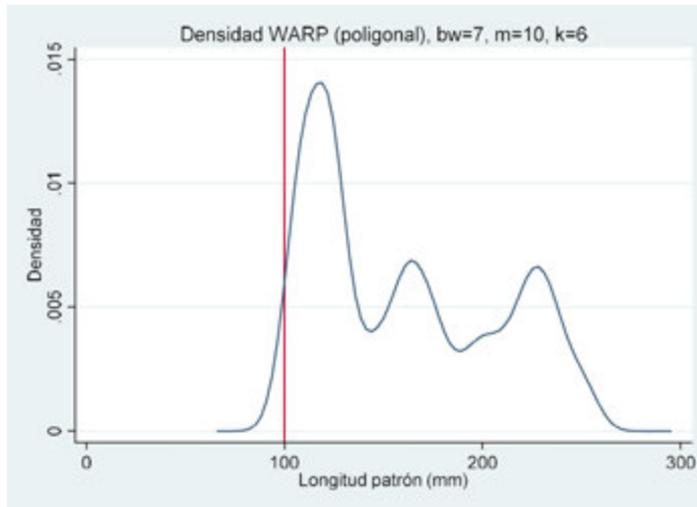


Figura 1.52. Estimador de densidad por kernel (EDK) para la muestra de marzo de *Cathorops melanopus* ($n = 190$; 75 machos, 57 hembras y 58 indeterminados).

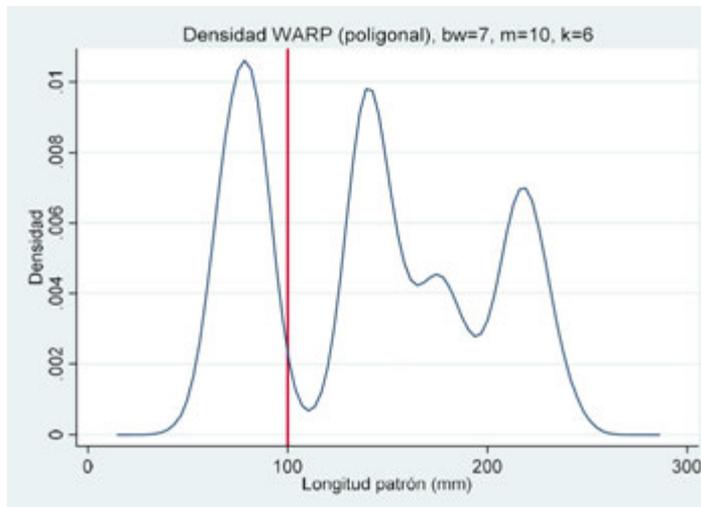


Figura 1.53. Estimador de densidad por kernel (EDK) para la muestra de noviembre de *Cathorops melanopus* ($n = 292$; submuestra de 100 indeterminados, 109 machos y 83 hembras).

Capítulo 2

Alometría: relaciones entre longitudes y pesos



2.1 Introducción

El término “alometría” en un sentido amplio se refiere a las diferencias en las proporciones correlacionadas con cambios en magnitud absoluta de un organismo completo o de las partes específicas consideradas (Gould, 1966). El estudio de las proporciones entre las partes orgánicas ha sido de interés para los estudiosos de las entidades biológicas desde tiempos antiguos.

Uno de ellos fue Georges Cuvier, quien en 1798 postuló la “Ley de la correlación de las partes” que sostiene la posibilidad de reconocer a un animal por un único hueso o incluso por una parte de ese resto por medio del conocimiento de las relaciones cuantitativas entre los componentes estructurales de los seres vivos (Cuvier 1798; 1810 citado en Caponi, 2004).

Si bien, el interés sobre el tamaño y las proporciones de las partes orgánicas viene desde la antigüedad, quien destacó el tema en el siglo pasado fue D’Arcy Wentworth Thompson con su obra “On Growth and Form” (Thompson, 1917; 1961; 2014) en la que sigue la historia del tema hasta Aristóteles quien argumentó a partir de matemáticas elementales que la superficie y el volumen de un cuerpo geométrico aumentan proporcionalmente al cuadrado y al cubo, respectivamente de las dimensiones lineales (McMahon y Bonner, 1986).

Para medir los cambios en las proporciones Thompson propuso el uso de las “transformaciones cartesianas” que consiste en una malla cartesiana rectilínea sobre un dibujo bidimensional del organismo en una cierta etapa de desarrollo y la construcción de otra sobre el mismo organismo en etapa distinta de crecimiento. La deformación de la retícula indica que partes se han contraído y cuales se han expandido. Este enfoque, aunque integral, puede considerarse como una aproximación al hacer la simplificación de un espécimen tridimensional a un dibujo en dos dimensiones (McMahon y Bonner, 1986).

Otra aproximación más sencilla es la desarrollada por J.S. Huxley (1932) y Georges Teissier (1931) ambos citados en McMahon y Bonner (1986), la cual es considerar la relación entre dos parámetros. Para comparar los tamaños relativos de dos partes de un organismo X e Y una forma de especificarlo formalmente sería mediante una relación potencial:





$$Y = aX^b \dots\dots\dots(2.1)$$

donde a y b son constantes. A esta función potencial se le llama “Ecuación Alométrica”.

Debido a consideraciones prácticas y estadísticas, se acostumbra utilizar la versión linealizada de esta expresión, es decir la versión logarítmica:

$$\log_{10} Y = \log_{10} a + b \log_{10} X \dots\dots\dots(2.2)$$

La constante b es el llamado “coeficiente alométrico” que nos indica la naturaleza de los cambios proporcionales entre X y Y . Esta constante es igual a 1 cuando las variables relacionadas son de la misma dimensión (lineales con lineales o volumétricas con volumétricas) y se dice entonces que el crecimiento es “Isométrico”. Si consideramos una dimensión volumétrica en relación con una lineal, el valor del coeficiente será idealmente igual a 3 si el crecimiento es isométrico. Valores diferentes a 1 o 3 indican un crecimiento “alométrico”; cuando la constante produce valores mayores (a 1 o 3) se tiene un crecimiento “hiperalométrico” y si son menores (a 1 o 3) el crecimiento es “hipoalométrico” (Froese, *et al.* 2011).

Suele explicarse adicionalmente que en el caso de hiperalometría con una variable volumétrica (peso) y otra lineal (longitud), el organismo crece más en peso que en longitud, mientras que la hipoalometría sugiere el crecimiento acentuado de la longitud en relación con el peso.

Para ejemplificar estas relaciones se pueden “crear” entidades geométricas sencillas (un cubo y una esfera) y variar la tasa de incremento de sus tres dimensiones (X , Y , Z).

El Cuadro 2.1 muestra los resultados obtenidos con ayuda de una hoja de cálculo (Excel) en el caso de la variación del volumen de un cubo y una esfera a lo largo de 20 intervalos (de tiempo) y con una tasa constante de incremento igual a 1.5. La entidad crece con las mismas proporciones siempre, es decir tiene crecimiento “Isométrico”.



Cuadro 2.1. Magnitudes de una entidad isométrica (tasas de incremento constantes = 1.5).

Tiempo	X	Y	Z	Volumen (cubo)	Volumen (esfera)
1	1.00	1.00	1.00	1.00	0.52
2	1.50	1.50	1.50	3.38	14.14
3	2.25	2.25	2.25	11.39	47.71
4	3.38	3.38	3.38	38.44	161.03
5	5.06	5.06	5.06	129.75	543.48
6	7.59	7.59	7.59	437.89	1834.25
7	11.39	11.39	11.39	1477.89	6190.58
8	17.09	17.09	17.09	4987.89	20893.20
9	25.63	25.63	25.63	16834.11	70514.56
10	38.44	38.44	38.44	56815.13	237986.65
11	57.67	57.67	57.67	191751.06	803204.96
12	86.50	86.50	86.50	647159.82	2710816.74
13	129.75	129.75	129.75	2184164.41	9149006.48
14	194.62	194.62	194.62	7371554.88	30877896.88
15	291.93	291.93	291.93	24878997.72	104212901.96
16	437.89	437.89	437.89	83966617.31	351718544.13
17	656.84	656.84	656.84	283387333.43	1187050086.43
18	985.26	985.26	985.26	956432250.32	4006294041.69
19	1477.89	1477.89	1477.89	3227958844.83	13521242390.69
20	2216.84	2216.84	2216.84	10894361101.31	45634193068.59

Debido al aumento considerable en las dimensiones con el paso del tiempo, es conveniente utilizar los valores logarítmicos de las dimensiones. El Cuadro 2.2 contiene los valores para una entidad con un crecimiento “Hiperalométrico” y el Cuadro 2.3 muestra el caso de una entidad con un crecimiento “Hipoalométrico”.



Cuadro 2.2. Magnitudes de una entidad hiperalométrica (tasa de incremento de $Y = 3$).

Tiempo	X (1.5)	Y (3)	Z (1.5)	Volumen (cubo)	Volumen (esfera)
1	1.00	1	1.00	1.00	0.52
2	1.50	3	1.50	6.75	3.53
3	2.25	9	2.25	45.56	23.86
4	3.38	27	3.38	307.55	161.03
5	5.06	81	5.06	2075.94	1086.96
6	7.59	243	7.59	14012.60	7336.98
7	11.39	729	11.39	94585.08	49524.63
8	17.09	2187	17.09	638449.29	334291.27
9	25.63	6561	25.63	4309532.72	2256466.06
10	38.44	19683	38.44	29089345.87	15231145.88
11	57.67	59049	57.67	196353084.65	102810234.71
12	86.50	177147	86.50	1325383321.42	693969084.29
13	129.75	531441	129.75	8946337419.57	4684291318.98
14	194.62	1594323	194.62	60387777582.09	31618966403.09
15	291.93	4782969	291.93	407617498679.13	213428023220.84
16	437.89	14348907	437.89	2751418116084.15	1440639156740.64
17	656.84	43046721	656.84	18572072283568.00	9724314307999.29
18	985.26	129140163	985.26	125361487914084.00	65639121578995.20
19	1477.89	387420489	1477.89	846190043420066.00	443064070658218.00
20	2216.84	1162261467	2216.84	5711782793085450.00	2990682476942970.00

Cuadro 2.3. Magnitudes de una entidad hipoalométrica (Tasa de incremento $X = 5$).

Tiempo	X (5)	Y (1.5)	Z (1.5)	Volumen (cubo)	Volumen (esfera)
1	1	1.00	1.00	1.00	0.52
2	5	1.50	1.50	11.25	5.89
3	25	2.25	2.25	126.56	66.27
4	125	3.38	3.38	1423.83	745.51



Cuadro 2.3. Magnitudes de una entidad hipoalométrica
(Tasa de incremento $X = 5$) (continuación).

Tiempo	X (5)	Y (1.5)	Z (1.5)	Volumen (cubo)	Volumen (esfera)
5	625	5.06	5.06	16018.07	8387.04
6	3125	7.59	7.59	180203.25	94354.20
7	15625	11.39	11.39	2027286.53	1061484.74
8	78125	17.09	17.09	22806973.46	11941703.38
9	390625	25.63	25.63	256578451.40	134344163.00
10	1953125	38.44	38.44	2886507578.19	1511371833.70
11	9765625	57.67	57.67	32473210254.68	17002933129.10
12	48828125	86.50	86.50	365323615365.20	191282997702.36
13	244140625	129.75	129.75	4109890672858.46	2151933724151.56
14	1220703125	194.62	194.62	46236270069657.60	24209254396705.00
15	6103515625	291.93	291.93	520158038283648	272354111962931.00
16	30517578125	437.89	437.89	5851777930691040	3063983759582980.00
17	152587890625	656.84	656.84	65832501720274200	34469817295308500.00
18	762939453125	985.26	985.26	740615644353085000	387785444572220000.00
19	3814697265625	1477.89	1477.89	8331925998972210000	4362586251437480000.00
20	19073486328125	2216.84	2216.84	93734167488437300000	49079095328671700000.00

El siguiente código en R permite mostrar los resultados de graficar los valores logarítmicos del volumen de la esfera con la dimensión horizontal (X) en los tres casos de crecimiento (isométrico, hiper e hipo alométrico). Los coeficientes alométricos se obtuvieron con la función “lm” (en este ejemplo realiza regresión lineal) para cada grupo de valores.

```
> setwd("c:/data")
> library(foreign)
> formdatos <- read.dta("formdatos.dta")
> names(formdatos)
[1] "tiempo" "equis" "ye" "zeta" "vcubo" "vesfera" "x2"
"y2"
[9] "z2" "vc2" "ve2" "x3" "y3" "z3" "vc3"
"ve3"
> lm(log10(vesfera)~log10(equis), data=formdatos)
```



```
Call:
lm(formula = log10(vesfera) ~ log10(equis), data = formadatos)

Coefficients:
(Intercept)  log10(equis)
  0.4538      3.0735
> lm(log10(ve2)~log10(x2), data=formadatos)

Call:
lm(formula = log10(ve2) ~ log10(x2), data = formadatos)

Coefficients:
(Intercept)  log10(x2)
 -0.2816     4.7098
> lm(log10(ve3)~log10(x3), data=formadatos)

Call:
lm(formula = log10(ve3) ~ log10(x3), data = formadatos)

Coefficients:
(Intercept)  log10(x3)
 -0.2816     1.5039

> plot(log10(ve3)~log10(x3), data=formadatos, type="p", ylab="Log10(Volumen
esfera)", xlab="Log10(X)")
> lines(log10(ve3)~log10(x3), data=formadatos)
> points(log10(ve2)~log10(x2), data=formadatos, pch=2)
> lines(log10(ve2)~log10(x2), data=formadatos)
> points(log10(vesfera)~log10(equis), data=formadatos, pch=3)
> lines(log10(vesfera)~log10(equis), data=formadatos)
> ex.b1 <- expression(Isométrico (italic(b) == 3))
> ex.b2 <- expression(Hiperalométrico (italic(b) == 4.7))
> ex.b3 <- expression(Hipoalométrico (italic(b) == 1.5))
> text(5,11, ex.b1)
> text(4,17, ex.b2)
> text(10,10, ex.b3)
```

Se obtiene la Figura 2.1.

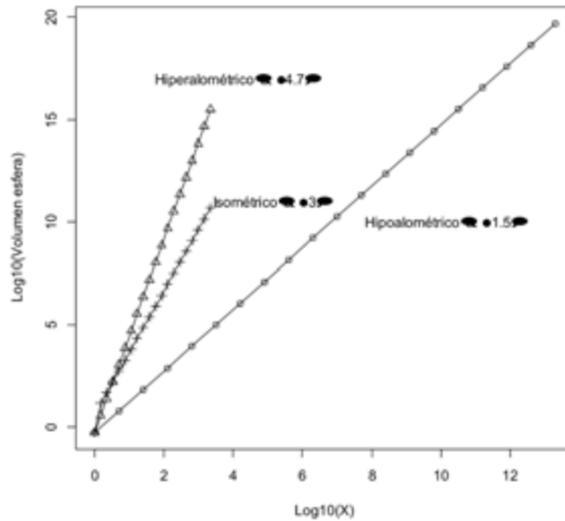


Figura 2.1. Versión gráfica con R para los tres casos de crecimiento en escala logarítmica.

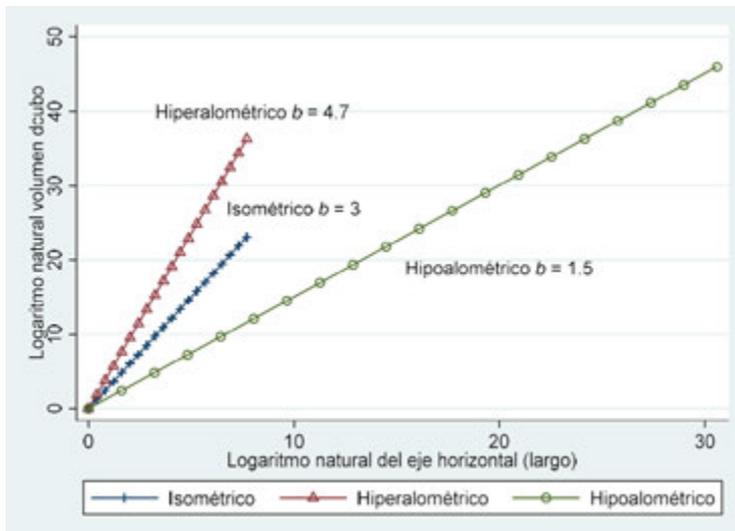


Figura 2.2. Versión gráfica con Stata para los tres casos de crecimiento en escala logarítmica.



2.2 Relaciones entre dimensiones lineales

En los organismos de importancia pesquera es práctica común medirlos con diversos fines. El motivo más importante es para verificar si los organismos capturados cumplen con las tallas mínimas establecidas (en el caso de que las haya) o bien si tienen el tamaño adecuado para que puedan comercializarse. Para la evaluación como recurso natural renovable es necesario medir sistemáticamente su tamaño.

En el caso de los peces existe una serie de medidas establecidas tanto de longitud como de peso. El manual de Ciencia Pesquera de la FAO (Holden y Rait, 1975) cita 16 medidas de longitud para los peces además de otras tantas más específicas. Lozano-Cabo (1983) define cinco longitudes principales, pero recientemente se ha hecho común el uso de cuatro medidas de longitud con variantes según la anatomía peculiar de algunas especies (Cailliet *et al.*, 1986; Espino-Barr, *et al.*, 2008; Anislado-Tolentino, *et al.*, 2016; Figura 2.3):

- Total: medida desde la sínfisis maxilar o mandibular hasta el extremo de los lóbulos de la aleta caudal ligeramente comprimidas.
- Patrón: (o Estándar): como la total pero hasta el extremo anterior del lóbulo carnoso del cuerpo.
- Furcal: como las anteriores pero hasta la horquilla caudal
- Altura: distancia corporal máxima perpendicular al eje antero-posterior.

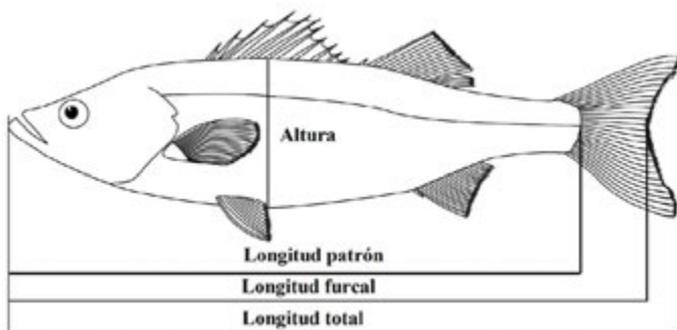


Figura 2.3. Principales medidas de longitud en peces (dibujo por E.R. Zambrano modificado).



En ocasiones, durante el manejo y traslado de los especímenes ícticos, la aleta caudal puede dañarse y en estos casos, la medición correcta de la longitud total (e incluso de la furcal) se hace imposible. Por lo tanto, en esta obra se recomienda el uso de la longitud patrón, si bien con base en las consideraciones de Howe (2002), se hace preciso medirla según la definición arriba indicada.

A pesar de lo señalado arriba, frecuentemente se lleva a cabo la relación entre la longitud total y la patrón o la furcal con fines comparativos entre reportes que utilizaron alguna de estas medidas y diferente a la utilizada en un informe particular. El caso más común es entre la longitud total y la patrón (o estándar).

Un ejemplo se presenta en la Figura 2.4 para los datos de longitud total y patrón de hembras y juveniles del bagre estuarino *Cathorops melanopus* (Salgado-Ugarte, 1985). La regresión entre estas variables permite la determinación de una de las longitudes a partir de las otras. Se ha utilizado la escala logarítmica para obtener la forma tradicional del coeficiente alométrico. En este caso, la estimación puntual (0.9676) y los límites del intervalo de confianza del 95% (0.9645, 0.9706) sugieren una relación de tipo ligeramente hipoalométrica.

```
. use bagtoto
. regress lntl lnbl if sex==2 | sex==3
```

Source	SS	df	MS	Number of obs	=	1,708
Model	262.696842	1	262.696842	F(1, 1706)	>	99999.00
Residual	1.17807205	1,706	.000690546	Prob > F	=	0.0000
Total	263.874914	1,707	.154584015	R-squared	=	0.9955
				Adj R-squared	=	0.9955
				Root MSE	=	.02628

lntl	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnbl	.9675609	.0015687	616.78	0.000	.9644841 .9706377
_cons	.3803711	.007314	52.01	0.000	.3660258 .3947165

```
. scatter lntl lnbl if sex==2 | sex==3 , ms(Oh) || lfit lntl lnbl if
sex==2 | sex == 3, lw(1) ytitle("Logaritmo natural de longitud total")
xtitle("Logaritmo natural de longitud patrón") legend(label(1 "Observados")
label(2 "Calculados"))
```

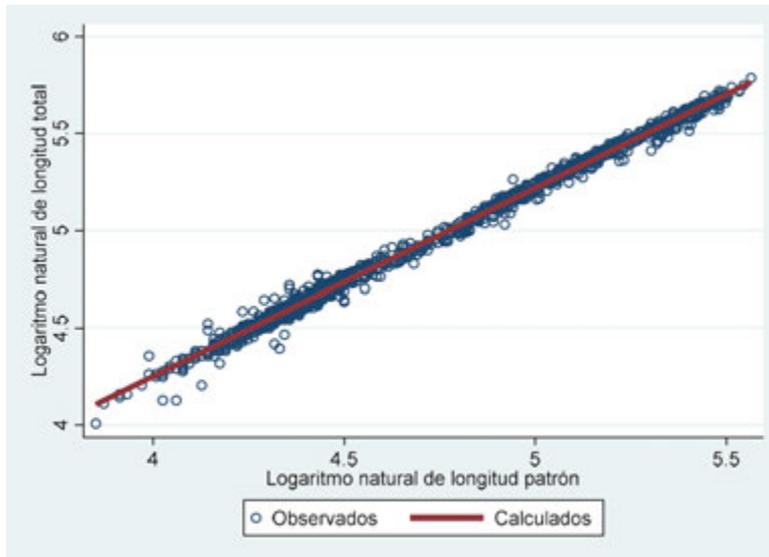


Figura 2.4. Relación entre la longitud total y la longitud patrón de hembras y juveniles de *Cathorops melanopus*. La línea muestra el ajuste lineal con un coeficiente ligeramente hipoalométrico de 0.9675.

Actualmente, con la disponibilidad de computadoras veloces y software estadístico es posible ajustar directamente la ecuación alométrica por medio de procedimientos iterativos de regresión no lineal. Para el caso de Stata, en esta obra se presentan los programas “nlpotfun.ado” y su versión actualizada “nlpfun2.ado” los cuales especifican la función no lineal potencial para ser introducida a la rutina de cálculo por regresión no lineal.

A diferencia de la lineal, la regresión no lineal hace uso de métodos numéricos iterativos para encontrar la combinación de parámetros que minimizan una función de residuos al cuadrado. Stata utiliza un método mixto “Gauss-Newton” para encontrar la combinación de parámetros que dan el valor mínimo de desviaciones al cuadrado de los datos con el modelo resultante.

A continuación su aplicación a los datos de hembras e indeterminados del bagre:



```
. use bagtoto

. nl pfun2: totlen bodlen if sex==2 | sex==3
(obs = 1,708)
```

```
Iteration 0: residual SS = 3.13e+07
Iteration 1: residual SS = 2.93e+07
.
Iteration 8: residual SS = 25799.4
Iteration 9: residual SS = 25799.4
```

Source	SS	df	MS		
Model	40941975	2	20470987.3	Number of obs =	1,708
Residual	25799.405	1706	15.122746	R-squared =	0.9994
Total	40967774	1708	23985.8162	Adj R-squared =	0.9994
Potential function totlen = a*bodlen^b				Root MSE =	3.888798
				Res. dev. =	9484.362

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/a	1.488808	.0110761	134.42	0.000	1.467084	1.510532
/b	.9639746	.0014853	649.02	0.000	.9610615	.9668878

Debemos destacar la semejanza del coeficiente alométrico entre el obtenido por mínimos cuadrados de las variables transformadas logarítmicamente ($b = 0.9676$) con el coeficiente por regresión no lineal ($b = 0.9640$). Los intervalos de confianza en ambos casos, aunque cercanos a 1, no lo abarcan por lo que estrictamente la relación entre estas variables es hipoalométrica.

2.2.1 Otros ejemplos. Equinoideos

A continuación se muestran varios ejemplos de relaciones entre dimensiones lineales de varios organismos acuáticos con formas diversas.

Como parte de las actividades del Laboratorio de Biometría y Biología Pesquera (FES Zaragoza, UNAM), se han venido analizando datos provenientes de las capturas de la pesca artesanal de diferentes especies de Puerto Ángel, Oaxaca México. Aún sin un interés pesquero importante (a diferencia de su importancia en países asiáticos), los erizos de mar (Echinodermata: Echinoidea), son abundantes en los litorales rocosos aledaños a esta



localidad y representan un recurso potencial. En la zona son abundantes dos especies: el erizo negro, *Equinometra vanbrunti* y el erizo rosado, *Toxopneustes rosseus* (Figura 2.5).



Figura 2.5. Vistas aboral del erizo negro *Equinometra vanbrunti* y lateral del erizo rosado *Toxopneustes roseus* (Fotografías por González-Castillo A.G. y Carapia-Moreno C.).

La relación alométrica entre la altura y el ancho de los caparazones (testas) de *E. vanbrunti* recolectados en un ciclo anual de abril 2017 a abril 2018 (González-Castillo y Rivera-Barrera, 2018) se lleva a cabo con los siguientes comandos de Stata. Primero con la transformación logarítmica y después con el ajuste por regresión no lineal a los datos originales:

```
. use erizosigma  
  
. gen lnaltura = log(altura)  
  
. gen lnancho = log(ancho)  
  
. regress lnaltura lnancho
```



Source	SS	df	MS	Number of obs	=	184
Model	4.97700059	1	4.97700059	F(1, 182)	=	924.49
Residual	.979800092	182	.005383517	Prob > F	=	0.0000
				R-squared	=	0.8355
				Adj R-squared	=	0.8346
Total	5.95680068	183	.032550823	Root MSE	=	.07337

lnaltura	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnancho	1.260349	.0414515	30.41	0.000	1.178562	1.342136
_cons	-1.17352	.0651769	-18.01	0.000	-1.30212	-1.044921

```
. nl pfun2: altura ancho
(obs = 184)
Iteration 0: residual SS = 8.930433
.
Iteration 5: residual SS = 5.019126
```

Source	SS	df	MS	Number of obs	=	184
Model	967.78927	2	483.894634	R-squared	=	0.9948
Residual	5.0191261	182	.027577616	Adj R-squared	=	0.9948
				Root MSE	=	.1660651
Total	972.8084	184	5.28700215	Res. dev.	=	-140.5397

Potential function altura = a*ancho^b

altura	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/a	.3141114	.0218545	14.37	0.000	.2709907	.3572321
/b	1.252211	.0431518	29.02	0.000	1.167069	1.337353

```
. predict altenol
(option yhat assumed; fitted values)
```

```
. scatter altura ancho, ms(Oh) || scatter altenol ancho, c(1) ms(i) sort
lw(1) legend(label(1 "Observados") label(2 "Calculados") pos(5) ring(0))
ytitle("Altura de la testa (cm)") xtitle("Ancho de la testa (cm)") text(3.3 3.6
"{it:H} = 0.31{it:A}{super:1.25}") title({it:Echinometra vanbrunti})
```

Tanto la regresión lineal con los datos transformados a escala logarítmica, como la regresión no lineal a la función alométrica potencial con los datos originales, indican un crecimiento hiperalométrico entre la altura y el ancho de la testa del erizo negro de mar *E. vanbrunti* (Figura 2.6).

Los ajustes para el erizo rosa *Toxopneustes roseus* se presentan enseguida y el resultado en la Figura 2.7:

Métodos cuantitativos computarizados para biología pesquera



```
. use erizotorogama

. gen lnaltura = log(altura)

. gen lnancho = log(ancho)

. scatter lnaltura lnancho, ms(Oh)

. regress lnaltura lnancho
```

Source	SS	df	MS	Number of obs		
Model	3.0296225	1	3.0296225	F(1, 186)	=	391.07
Residual	1.44092834	186	.007746927	Prob > F	=	0.0000
Total	4.47055085	187	.023906689	R-squared	=	0.6777
				Adj R-squared	=	0.6760
				Root MSE	=	.08802

lnaltura	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnancho	1.196962	.0605272	19.78	0.000	1.077553	1.31637
_cons	-1.157887	.1191831	-9.72	0.000	-1.393011	-.9227624

```
. nl potfun altura ancho
(obs = 188)
```

```
Iteration 0: residual SS = 796.746
Iteration 1: residual SS = 604.7306
Iteration 2: residual SS = 417.7995
Iteration 3: residual SS = 22.05607
Iteration 4: residual SS = 16.07339
Iteration 5: residual SS = 15.99468
Iteration 6: residual SS = 15.99464
```

Source	SS	df	MS	Number of obs		
Model	2137.37636	2	1068.68818	F(2, 186)	=	12427.66
Residual	15.9946404	186	.08599269	Prob > F	=	0.0000
Total	2153.371	188	11.4541011	R-squared	=	0.9926
				Adj R-squared	=	0.9925
				Root MSE	=	.2932451
				Res. dev.	=	70.25349

```
Potential function, altura=b0*ancho^b1
```

altura	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
b0	.2919359	.0367033	7.95	0.000	.2195275	.3643442
b1	1.235916	.0630154	19.61	0.000	1.111599	1.360233



(SEs, P values, CIs, and correlations are asymptotic approximations)

```
. predict altenol
(option yhat assumed; fitted values)

. scatter altura ancho, ms(Oh) || scatter altenol ancho, c(1) ms(i) sort lw(1)
legend(label(1 "Observados") label(2 "Calculados") pos(5) ring(0)) ytitle("Altura
de la testa (cm)") xtitle("Ancho de la testa (cm)") text(4.6 5.7 "{it:H}
= 0.29{it:A}{super:1.24}") ylab(2(1)5) xlab(5(1)10) title({it:Toxopneustes
roseus})
```

Ambas especies presentan un crecimiento de testa similar, pero son claramente diferentes en su aspecto externo. *E. vanbrunti* es oscuro (morado casi negro) con espinas largas y agudas mientras que *T. roseus* es de color rosa claro con espinas cortas y no tan agudas (Figura 2.5).

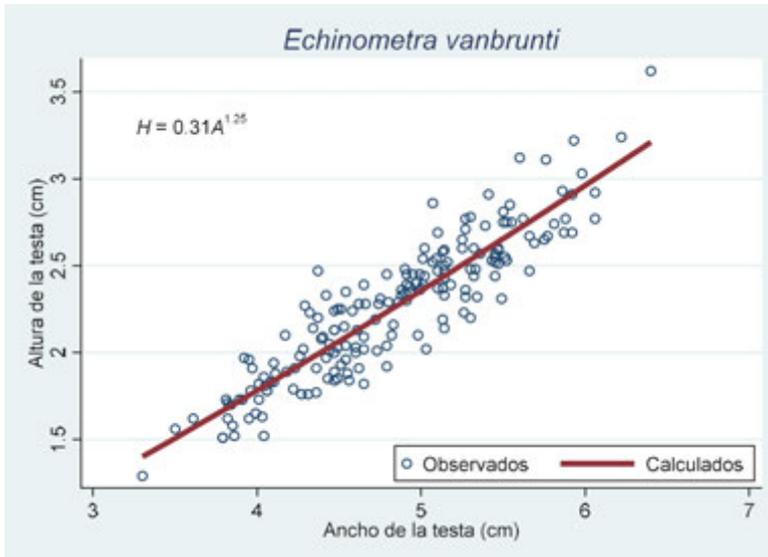


Figura 2.6. Relación alométrica entre la Altura (H) y el ancho (A) de las testas de *E. vanbrunti*. El coeficiente alométrico (1.25) así como el intervalo con 95% de confianza (1.17, 1.34) indican un crecimiento hiperalométrico. La línea muestra el ajuste directo por regresión no lineal.

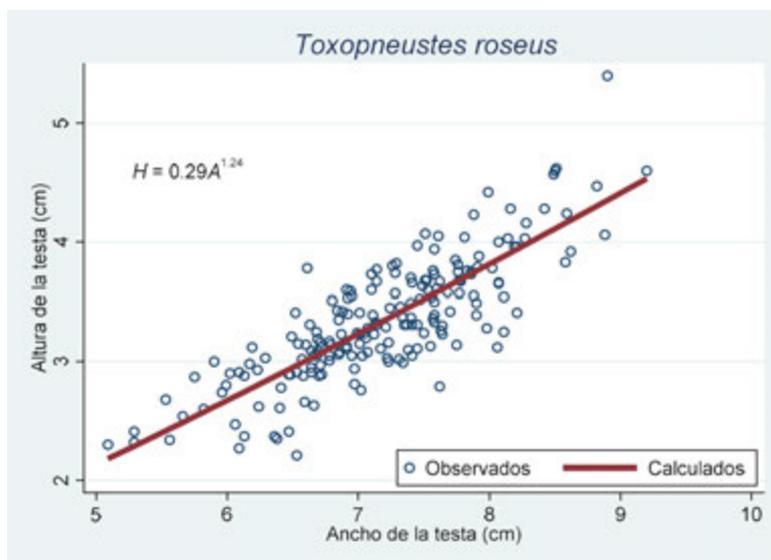


Figura 2.7. Relación alométrica entre la Altura (H) y el ancho (A) de las testas de *Toxopneustes roseus*. El coeficiente alométrico (1.24) así como el intervalo con 95% de confianza (1.11, 1.36) indican un crecimiento hiperalométrico. La línea muestra el ajuste directo por regresión no lineal.

2.2.2. Gasterópodo *Leucozonia cerata*

Si bien, esta especie de molusco no se explota intensamente en los litorales de Puerto Ángel, Oaxaca, México, tiene cierta demanda para la elaboración de artesanías. Desde el punto de vista ecológico participa en la trama trófica del ecosistema local y su concha sirve de hábitaculo para otros organismos (cangrejos ermitaños) (Rodríguez-Galicia, 2015).

Los ejemplares aquí considerados se obtuvieron de muestras recolectadas en periodos aproximadamente mensuales de febrero 2014 a junio 2015. En laboratorio los especímenes se midieron en largo y ancho de su concha (Figura 2.8).

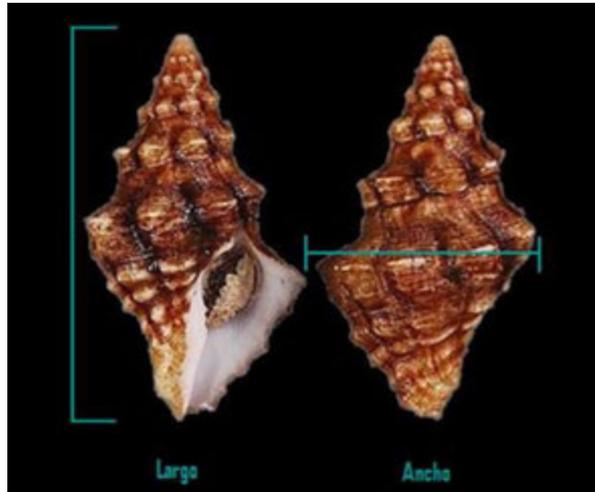


Figura 2.8. Medidas morfológicas tomadas a *Leucozonia cerata*

Los comandos para llevar el análisis alométrico, primero con la transformación logarítmica de las variables y regresión lineal y posteriormente con la estimación directa de los datos por medio de regresión no lineal con Stata son:

```
. use leuco2
. gen lnlar = log(lar)
. gen lnanc = log(anc)
. regress lnlar lnanc
```

Source	SS	df	MS	Number of obs	=	78
-----+-----				F(1, 76)	=	1802.73
Model	6.87410783	1	6.87410783	Prob > F	=	0.0000
Residual	.289800027	76	.003813158	R-squared	=	0.9595
-----+-----				Adj R-squared	=	0.9590
Total	7.16390786	77	.093037764	Root MSE	=	.06175

lnlar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
lnanc	.9876257	.0232609	42.46	0.000	.9412976 1.033954
_cons	.6358246	.0257585	24.68	0.000	.5845222 .687127
-----+-----					



```
. nl potfun lar anc
(obs = 78)
```

```
Iteration 0: residual SS = 1936.929
Iteration 1: residual SS = 493.068
Iteration 2: residual SS = 186.8252
Iteration 3: residual SS = 16.26879
Iteration 4: residual SS = 8.547736
Iteration 5: residual SS = 8.543581
Iteration 6: residual SS = 8.543581
```

Source	SS	df	MS	Number of obs =	78
Model	2654.82942	2	1327.41471	F(2, 76) =	11808.11
Residual	8.54358134	76	.112415544	Prob > F =	0.0000
				R-squared =	0.9968
				Adj R-squared =	0.9967
				Root MSE =	.3352843
				Res. dev. =	48.85518
Total	2663.373	78	34.1458077		

```
Potential function, lar=b0*anc^b1
```

lar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
b0	1.842501	.0657133	28.04	0.000	1.711622 1.973381
b1	1.011126	.0292085	34.62	0.000	.9529518 1.069299

(SEs, P values, CIs, and correlations are asymptotic approximations)

```
. scatter lar anc, ms(Oh) || scatter larenol anc, c(1) ms(i) sort lw(1)
legend(label(1 "Observados") label(2 "Calculados") pos(5) ring(0)) ytitle("Largo
de la concha (cm)") xtitle("Ancho de la concha (cm)") text(9 1.5 "{it:H} =
1.84{it:A}{sup:1.01}") title({it:Leucozonia cerata})
```

Los coeficientes alométricos obtenidos de la regresión lineal con variables transformadas a logaritmo (intervalo de confianza del 95%) = 0.99, (0.94, 1.03) y el obtenido directamente por regresión no lineal = 1.01 (0.95, 1.07) ambos indican un crecimiento isométrico (Figura 2.9).

Trabajando con R, es posible reproducir los resultados del ajuste directo por regresión no lineal por medio de los siguientes comandos (Figura 2.10):



```

> setwd("c:/data")

> library(foreign)

> leuco <- read.dta("leuco2.dta")

> potfun <- function(x, a, b) a*x^b

> res.potfun <- nls(lar~potfun(anc, a, b), start=c(a=.1, b=1), data=leuco)

> res.potfun
Nonlinear regression model
  model: lar ~ potfun(anc, a, b)
  data: leuco
    a      b
1.843 1.011
residual sum-of-squares: 8.544

Number of iterations to convergence: 5
Achieved convergence tolerance: 1.09e-07

> confint(res.potfun)
Waiting for profiling to be done...
      2.5%      97.5%
a 1.7146121 1.977081
b 0.9532057 1.069864

> ex.pf <- expression(italic(L) == italic(aA)^italic(b))

> plot(lar~anc, data=leuco, ylab="Largo de la concha (cm)", xlab="Ancho de la
concha (cm)", main=expression(italic("Leucozonia cerata")))

> curve(potfun(x, a=1.843, b=1.011), add=TRUE,lwd=2)

> legend(1.5,8,ex.vpf)

> legend(3.5,3, c("Observados", "Calculados"),lty=c(0,1), pch=c('o','.'),
lwd=c(1,2))

```

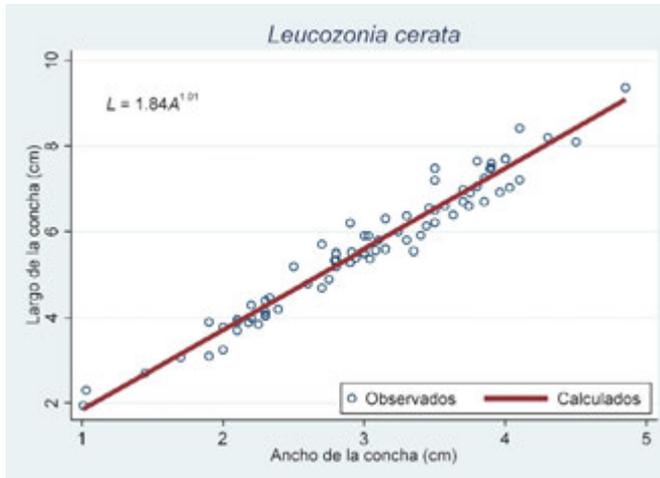


Figura 2.9. Relación alométrica entre el Largo (L) y el ancho (A) de la concha de *Leucozonia cerata*. El coeficiente alométrico (1.01) así como el intervalo con 95% de confianza (0.95, 1.07) indican un crecimiento isométrico. La línea muestra el ajuste directo por regresión no lineal.

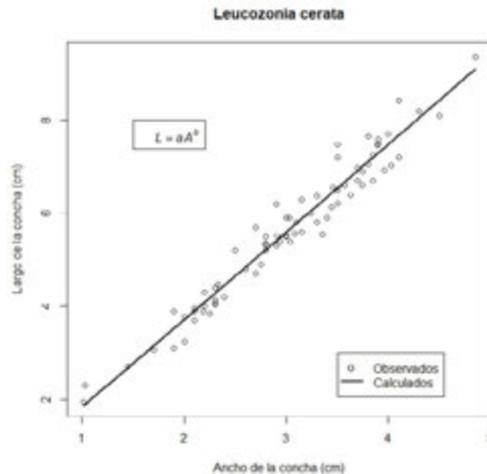


Figura 2.10. Relación alométrica entre el Largo (L) y el ancho (A) de la concha de *Leucozonia cerata* realizada con el programa R. La línea muestra el ajuste directo por regresión no lineal.



2.2.3. Pelecípodo *Pinctada mazatlanica*

La “Madreperla” (*Pinctada mazatlanica*) es un molusco bivalvo lamelibranquio que pertenece a las ostras perleras (Figura 2.11). Se caracteriza por tener capas de nácar en la superficie interna de las valvas y son productoras de perlas que han sido muy valoradas por el hombre desde la antigüedad.

Las ostras perleras representan para México un importante recurso integral ya que utiliza la concha, el nácar, el músculo (callo) como alimento y desde luego la perla. Sin embargo, la explotación poco controlada y la escasez de estudios sobre ecología y biología de las especies han impedido la recuperación de las poblaciones naturales (Torres-Zepeda y Cruz-Pérez, 2002; Meléndez-Contreras, 2015).

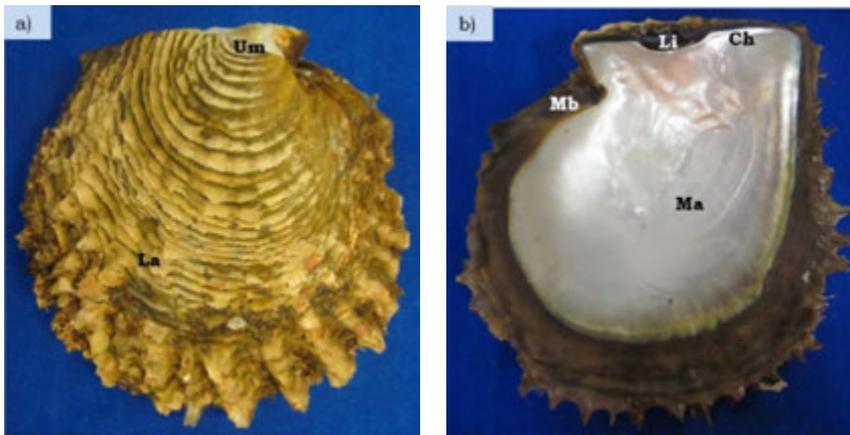


Figura 2.11. Fotografía de la valva derecha de *P. mazatlanica*. A) Vista externa que muestra el periostraco, las lamelas comarginales (La) y parte del umbo (Um); b) vista interna donde se observa la marca del músculo aductor (Ma), el ligamento (Li), la charnela (Ch) y la muesca del biso (Mb) (Tomada de Meléndez-Contreras, 2015).

El ejemplo que se presenta proviene de datos de recolectas, en lo posible, mensuales realizadas mediante buceo libre desde agosto 2013 a noviembre 2014 en el litoral rocoso aledaño a Puerto Ángel, Oax. Desde Bahías de Mazunte hasta Dominguillo. En total se obtuvieron 526 organismos.



La Figura 2.12 muestra las medidas biométricas tomadas a las conchas. Se incluyen los comandos de Stata para llevar a cabo el cálculo de la ecuación alométrica directamente mediante regresión no lineal. Los resultados se presentan en la Figura 2.12.

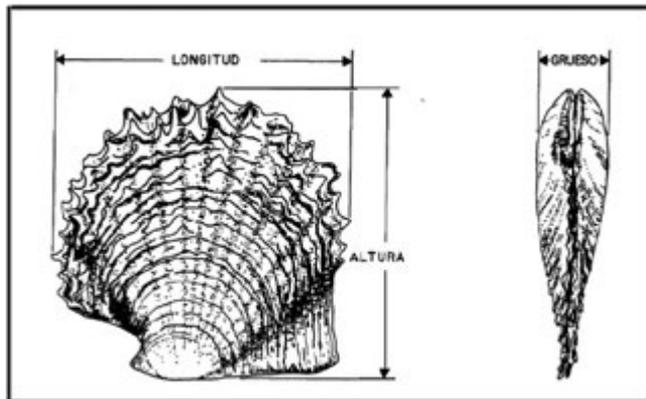


Figura 2.12. Medidas biométricas de la madreperla *Pinctada mazatlanica* (tomada de Saucedo-Lastra, 1995; Meléndez-Contreras, 2015).

```
. use pinmazadatos
```

```
. nl potfun longitud altura
(obs = 523)
```

```
Iteration 0: residual SS = 44284.38
Iteration 1: residual SS = 1125.939
Iteration 2: residual SS = 1024.657
Iteration 3: residual SS = 763.4023
Iteration 4: residual SS = 691.3748
Iteration 5: residual SS = 691.3614
```

Source	SS	df	MS
Model	60769.9331	2	30384.9666
Residual	691.361359	521	1.32698917
Total	61461.2945	523	117.516815

```
Number of obs = 523
F( 2, 521) = 22897.67
Prob > F = 0.0000
R-squared = 0.9888
Adj R-squared = 0.9887
Root MSE = 1.15195
Res. dev. = 1630.169
```

```
Potential function, longitud=b0*altura^b1
```



longitud	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
b0	2.664675	.1658951	16.06	0.000	2.33877	2.990581
b1	.5845802	.0257473	22.70	0.000	.5339989	.6351615

(SEs, P values, CIs, and correlations are asymptotic approximations)

```
. predict lonenol
(option yhat assumed; fitted values)
(3 missing values generated)
```

```
. scatter longitud altura, ms(Oh) || scatter lonenol altura, c(1) ms(i) sort
lw(1) legend(label(1 "Observados") label(2 "Calculados") pos(5) ring(0))
ytile("Longitud de la concha (cm)") xtitle("Altura de la concha (cm)") text(15
7 "{it:L} = 2.66{it:H}{super:0.58}") title({it:Pinctada mazatlanica})
```

La Figura 2.13 muestra una ligera curvatura negativa en la tendencia de la longitud con la altura de la concha. Esta relación alométrica también se destaca por un claro crecimiento hipoalométrico entre la longitud y la altura de la concha (coeficiente alométrico = 0.58).

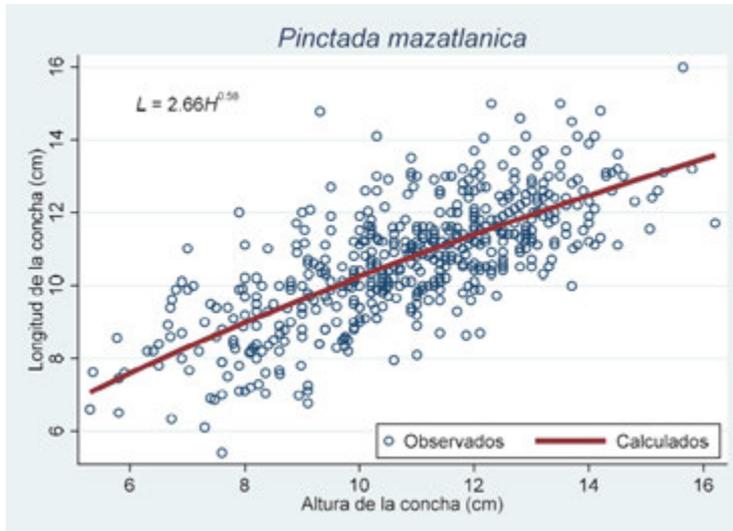


Figura 2.13. Relación alométrica entre la Longitud (L) y la altura (H) de la concha de *Pinctada mazatlanica*. El coeficiente alométrico (0.58) así como el intervalo con 95% de confianza (0.53, 0.64) indican un crecimiento hipoalométrico. La línea muestra el ajuste directo por regresión no lineal.



2.3 Relaciones entre dimensiones mixtas

De las relaciones entre dimensiones lineales y de volumen, la que destaca sobre otras es la relación entre el peso y la longitud de un organismo. En Biología Pesquera, a pesar de que algunos autores han menospreciado al modelado de la relación entre el peso y la longitud (RPL) de los peces (Hilborn y Walters, 2001) y que otros consideran este aspecto como no importante como para merecer ser publicado en la literatura científica (Froese, 2006), la revisión de este autor destaca finalmente su valor al proporcionar aspectos ecológicos importantes. Jellyman *et al.* (2013) resumen estos aspectos (Ogle, 2016):

La RPL es importante para calcular tasas de crecimiento y modelación bioenergética. Además es de utilidad para:

- Estimar el peso a partir de las longitudes.
- Convertir crecimiento longitudinal en ponderal.
- Estimar la biomasa de una comunidad a partir de datos de longitud.
- Estimar la condición de los organismos.
- Hacer comparaciones de rasgos de historias de vida.

Como relación alométrica, el cálculo de los coeficientes tradicionalmente se ha realizado por medio de regresión lineal de la expresión logarítmica (2.2) donde Y es el peso y X es la longitud. Para los datos de las hembras ($n = 556$) del bagre estuarino *Cathorops melanopus* (Salgado-Ugarte, 1985) con R se obtiene:

```
> setwd("c:/data")
> library(foreign)
> bagres <- read.dta("bagtoto.dta")
> setwd("c:/data")
> library(foreign)
> bagres <- read.dta("bagtoto.dta")
```

Se crea un subconjunto de datos sólo con las hembras (subset), se generan los valores logarítmicos del peso eviscerado: $\lnew <- \log(\text{eviwei})$, de la longitud patrón: $\lnbl <- \log(\text{bodlen})$, se lleva a cabo la regresión lineal (lm) y se despliegan los resultados. Debe notarse que los grados de libertad de 538 sugieren que el número de pares peso-longitud es de 540. Esto debido a que el peso eviscerado sólo se pudo registrar para 540 individuos.



Por otra parte, se utiliza el peso eviscerado (eviwei) para evitar la interferencia del peso de las vísceras que circunstancial (contenido estomacal) o estacionalmente (gónadas) presentan variaciones importantes y tratar de recuperar el peso somático de los individuos. Esto es importante para la estimación de la condición y del estado reproductivo de los especímenes (Salgado-Ugarte, *et al.* 2005).

```
> bagh <- subset(bagres, sex==2)
> lnew <- log(eviwei)
> lnbl <- log(bodlen)
> lnrplh <- lm(lnew~lnbl, data=bagh)
> summary(lnrplh)
```

```
Call:
lm(formula = lnew ~ lnbl, data = bagh)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.45741 -0.06564  0.00965  0.07704  0.33571
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.27182    0.12104   -93.13  <2e-16 ***
lnbl         2.99446    0.02348   127.51  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1192 on 538 degrees of freedom
(16 observations deleted due to missingness)
Multiple R-squared:  0.968,    Adjusted R-squared:  0.9679
F-statistic: 1.626e+04 on 1 and 538 DF,  p-value: < 2.2e-16
```

Posteriormente se imprimen los intervalos de confianza del 95% para los estimadores del coeficiente alométrico (pendiente) y de la ordenada al origen. Asimismo se piden los resultados del análisis de varianza aplicado a la regresión.

```
> confint(lnrplh)

            2.5 %      97.5 %
(Intercept) -11.509585 -11.034062
lnbl         2.948326   3.040592
> anova(lnrplh)
Analysis of Variance Table

Response: lnew
      Df Sum Sq Mean Sq F value    Pr(>F)
```



```
lnbl          1 231.044 231.044 16258 < 2.2e-16 ***
Residuals 538  7.646  0.014
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para obtener la Figura 2.14:

```
> plot(lnew~lnbl)
> tmp <- range(bagh$lnbl)
> xs <- seq(tmp[1],tmp[2],length.out=99)
> ys <- predict(lnrplh,data.frame(lnbl=xs))
> lines(ys~xs, lwd=2)
```

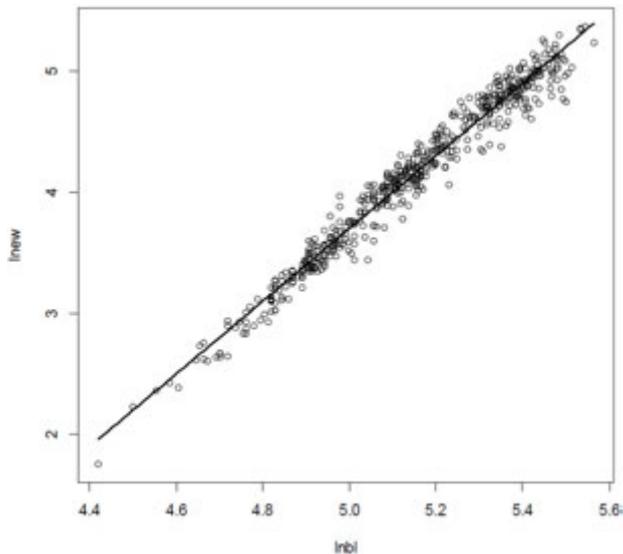


Figura 2.14. Relación logarítmica Peso – longitud de hembras de *Cathorops melanopus*. La línea muestra el ajuste lineal con un coeficiente alométrico de 2.99446 no significativamente diferente del valor isométrico de 3.

Para representar los datos en su escala original es necesario volver los valores estimados logarítmicos a gramos y milímetros en este caso, mediante la función inversa. Como se usaron logaritmos naturales, se eleva a su valor el número base e. Se sabe que los valores



re-transformados de logaritmos a escala original tienen un cierto sesgo y por tanto es necesario aplicar un factor de corrección (Sprugel, 1983; Ogle, 2016) que es igual a:

$$e^{\frac{(\log_e(\text{base})s_{y|X})^2}{2}}$$

```
> exp(.1192^2/2)
[1] 1.00713
> btys <- 1.00713*exp(ys)
```

Ya que R utiliza logaritmos naturales de manera pre-establecida, en la expresión anterior para el factor de corrección no es necesario calcular el logaritmo base e del número “ e ” ya que su resultado obviamente es la unidad que multiplica al error de la estimación ($s_{y|X}$) obtenido en la regresión (“residual standard error”). Se especifica porque es común el uso de logaritmos con base 10, en cuyo caso si se necesita calcular.

Cabe aclarar que para graficar los valores estimados (Figura 2.15), en lugar de utilizar los valores observados de la variable horizontal se crea una trama de valores uniformemente espaciados que cubren el recorrido (amplitud) de los datos de longitud (lo que se consigue con la función “range” y la generación de una secuencia (“seq”) del mínimo al máximo con un número suficiente de pasos con “length.out=99”).

```
> plot(lnew~lnbl)
> tmp <- range(bagh$lnbl)
> xs <- seq(tmp[1],tmp[2],length.out=99)
> ys <- predict(lnrplh,data.frame(lnbl=xs))
> lines(ys~xs, lwd=2)
> plot(eviwei~bodlen, data=bagh)
> btxs <- exp(xs)
> lines(btys~btxs, lwd=2)
```

De los patrones observados de las gráficas es claro el efecto de “megágono abierto” en la variabilidad de los datos peso-longitud conforme aumenta el tamaño de los organismos. La transformación logarítmica ayuda a uniformizar el comportamiento de la varianza a lo largo de los tamaños representados (Emerson y Stoto, 1983; Emerson y Hoaglin, 1983; Salgado-Ugarte, 2017).

Con la disponibilidad de computadoras y programas apropiados, es posible actualmente ajustar estos datos sin la transformación logarítmica por medio de procedimientos iterativos de regresión no lineal. De nuevo, con R esto se lleva a cabo con las siguientes funciones:

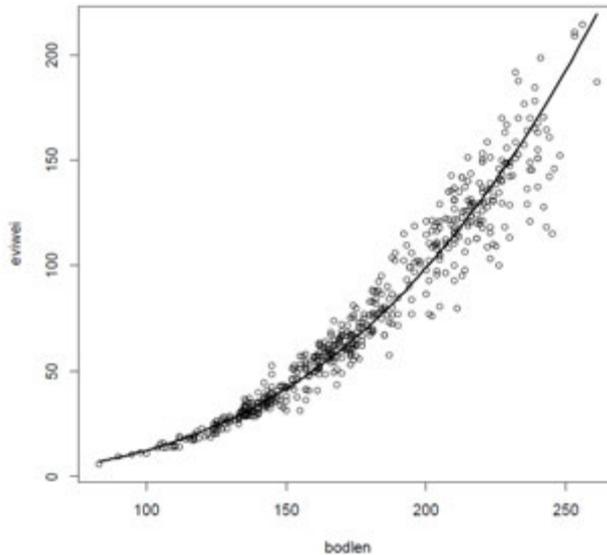


Figura 2.15. Relación Peso – longitud para hembras de *Cathorops melanopus*. La curva muestra el ajuste potencial con el exponente alométrico de 2.99446 no significativamente diferente del valor isométrico de 3.

Se genera la función con la expresión potencial (“potfun”):

```
> potfun <- function(l,a,b) a*l^b
```

Se utiliza la función “nls”, rutina que lleva a cabo la regresión no lineal por medio de algoritmos numéricos de minimización (de desviaciones al cuadrado). Los valores iniciales se probaron anteriormente, pero si se tienen mensajes de error “singular gradient”, entonces se cambian hasta obtener convergencia hacia una combinación de parámetros adecuada.

```
> rplh <- nls(eviwei~potfun(bodlen,a,b),data=bagh,start=c(a=.5,b=.5))
```

Los resultados se despliegan con “summary” y “confint”:

```
> summary(rplh)
```

```
Formula: eviwei ~ potfun(bodlen, a, b)
```



Parameters:

```

      Estimate Std. Error t value Pr(>|t|)
a 3.995e-05  7.894e-06   5.061 5.72e-07 ***
b 2.779e+00  3.697e-02  75.164 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 11.18 on 538 degrees of freedom

Number of iterations to convergence: 48

Achieved convergence tolerance: 2.038e-06

(16 observations deleted due to missingness)

```
> confint(rplh)
```

Waiting for profiling to be done...

```

      2.5%      97.5%
a 2.724644e-05 0.0000582907
b 2.707828e+00 2.8501257709

```

La gráfica de este ajuste, utiliza la función “predict” y la secuencia de valores de longitud creados anteriormente:

```

> yest <- predict(rplh,data.frame(bodlen=btxs))
> lines(yest~btxs, lwd=2)

```

Se puede comparar esta línea con el ajuste lineal logarítmico re-expresado en escala original:

```
> lines(btys~btxs, lwd=2,lty="dashed")
```

El resultado se muestra en la Figura 2.16. Es interesante notar que el ajuste por regresión no lineal produce un coeficiente de alometría de 2.78 y su intervalo de confianza no incluye al 3, lo que indica un crecimiento de tipo hipoalométrico (si bien, cercano a la isometría). La curva por ajuste no lineal pasa ligeramente por debajo de la re-transformada a tallas pequeñas, coincide a intermedias y va por encima a tallas mayores.

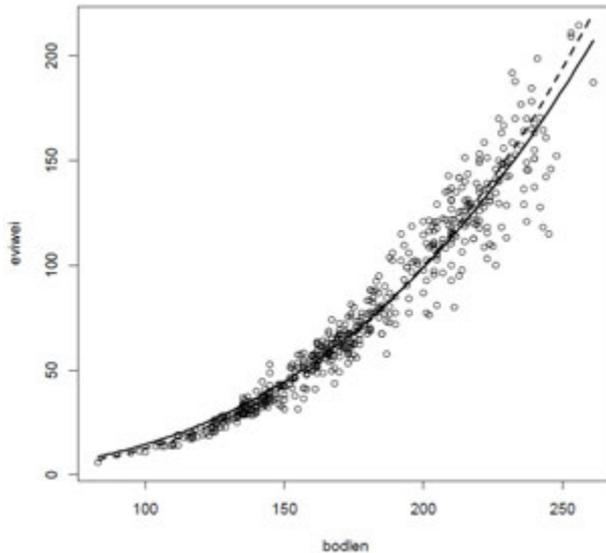


Figura 2.16. Comparación de la Relación Peso – longitud para hembras de *Cathorops melanopus*. Del ajuste lineal re-transformado (línea continua) con el ajuste por regresión no lineal de los datos originales (línea punteada).

Con Stata, los mismos resultados pueden obtenerse con el comando:

```
. use bagtoto
. nl pfun2 eviwei bodlen if sex==2
(obs = 540)
```

El comando “nl” llama a la rutina para regresión no lineal de Stata y lo aplica al programa “pfun2” que se escribió con la especificación de la ecuación alométrica potencial. Posteriormente se calculan los valores ajustados:

```
. predict ewe if sex==2
(option yhat assumed; fitted values)
(1,883 missing values generated)
```

Se hace la gráfica (Figura 2.17)



```
. scatter eviwei ewe bodlen if sex==2, ms(oh i) c(. 1) sort lw(. .6)
legend(label(1 "Observados") label(2 "Calculados")) ytitle("Peso eviscerado
(g)") xtitle("Longitud estándar (mm)")
```

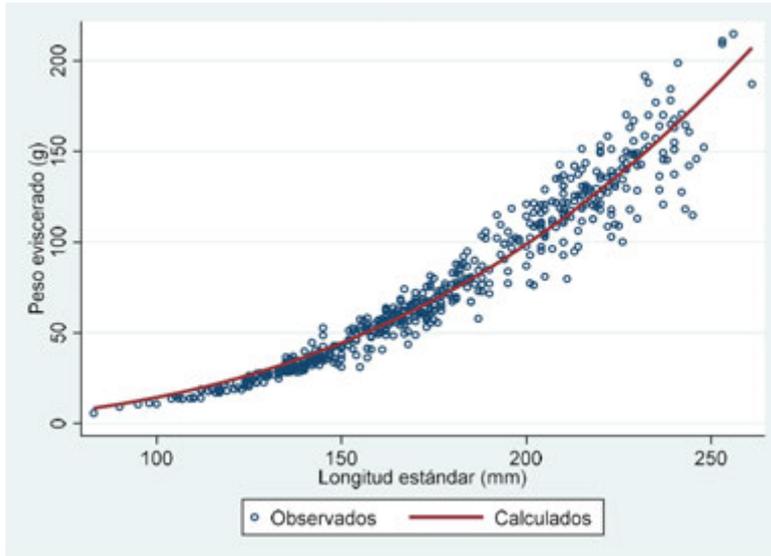


Figura 2.17. Relación Peso – Longitud para hembras de *Cathorops melanopus* con Stata.

2.3.1. Otros ejemplos de relaciones P-L

A continuación se incluye la aplicación de los métodos para establecer la relación peso longitud con diversos e3specímenes de diferentes tipos.

En primer lugar se presenta el caso del erizo de mar *Equinometra vanbrunti*, recolectado en los litorales rocosos cercanos a Puerto Ángel Oaxaca. Se presenta la relación Peso total – Ancho del caparazón de 183 individuos recolectados de agosto de 2017 a abril de 2018 como parte de las actividades de los proyectos del Laboratorio de Biometría y Biología Pesquera de la FES Zaragoza, UNAM y en este caso del proyecto del Laboratorio de Investigación Formativa de Ramírez-Gómez (2018).



El ajuste por regresión no lineal se hizo en Stata y dio como resultado un coeficiente alométrico de 2.64 significativamente diferente de 3, por lo que indica que esta especie tiene un crecimiento “hipoalométrico”.

```
. use evantros

. nl pfun2: pt ancho if especie==1
(obs = 183)

Iteration 0:  residual SS = 676660.9
Iteration 1:  residual SS = 675857.1
.
Iteration 13: residual SS = 9095.139
Iteration 14: residual SS = 9095.139
```

Source	SS	df	MS		
Model	680085.54	2	340042.771	Number of obs =	183
Residual	9095.1393	181	50.2493885	R-squared =	0.9868
				Adj R-squared =	0.9867
				Root MSE =	7.08868
Total	689180.68	183	3766.01465	Res. dev. =	1234.131

Potential function pt = a*ancho^b

pt	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
/a	.8696082	.1092717	7.96	0.000	.6539979 1.085218
/b	2.637708	.0760153	34.70	0.000	2.487718 2.787698

```
. scatter pt ancho if especie==1, ms(oh) || mspline ptespl ancho, bands(50) lw(.6)
ytitle("Peso total (g)") xtitle("Ancho del caparazón (cm)") legend(label(1
"Observados") label(2 "Calculados") pos(5) ring(0))
```

La gráfica con los datos observados y los calculados se incluye en la Figura 2.18.

2.3.2. Gasterópodo *Hexaplex princeps*

Relación Peso total – Largo de la concha del caracol marino *Hexaplex princeps* recolectado de febrero de 2012 a diciembre de 2014 en varios sitios litorales cercanos a Puerto Ángel, Oaxaca. En total la muestra fue de 774 individuos de los cuales 389 fueron machos y 385 hembras. Para los machos y trabajando con R se obtuvo:



```

> hpd <- read.dta("hexaprincepsdatos.dta")
> names(hpd)
 [1] "fecha"      "no"         "largo"      "pt"         "pgd"        "pg"         "pc"         "po"
"pv"         "sexo"
[11] "index"     "diat"       "dias"       "fc"         "igs"        "igd"        "lnpg"       "lnpc"
"lnpgd"     "lnpv"
[21] "denh"      "midh"       "denm"       "midm"       "diasa"      "equis"      "lnlargo"
> hpm <- subset(hpd, sexo==1)
> length(hpm)
 [1] 27
> plot(pt~largo, data=hpm)
> rplhpm <- nls(pt~potfun(largo,a,b),data=hpm,start=c(a=.5,b=.5))

> summary(rplhpm)
Formula: pt ~ potfun(largo, a, b)

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a  0.78889   0.13182    5.984 4.95e-09 ***
b  2.27138   0.07194   31.574 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25.34 on 387 degrees of freedom

Number of iterations to convergence: 12
Achieved convergence tolerance: 3.477e-06

> confint(rplhpm)
Waiting for profiling to be done...
      2.5%      97.5%
a 0.5734312 1.081819
b 2.1352717 2.408541
> tmp <- range(hpm$largo)
> xm <- seq(tmp[1],tmp[2],length.out=99)
> yem <- predict(rplhpm, data.frame(largo=xm))
> lines(yem~xm, lwd=2)

```

Estos resultados indicaron que *Hexaplex princeps* crece hipoalométricamente en peso total y largo de la concha (Figura 2.19).

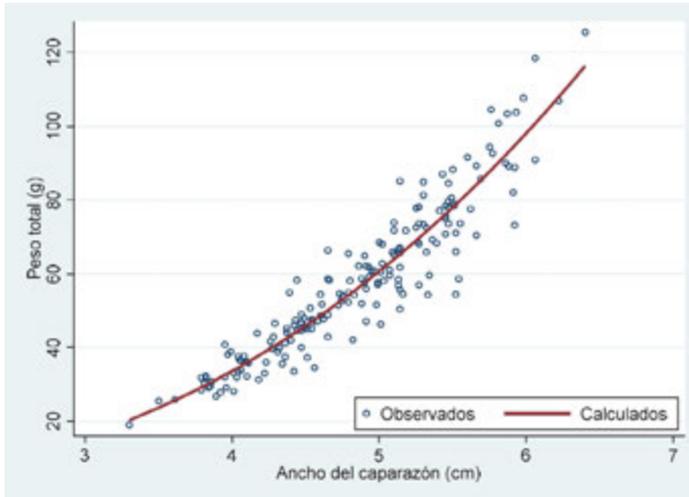


Figura 2.18. Relación Peso total – Ancho de caparazón del erizo marino *Equinometra vanbrunti* $n = 183$ (Adaptados de Ramírez-Gómez, 2018). El coeficiente alométrico 2.64 indica un crecimiento hipoalométrico estadísticamente significativo.

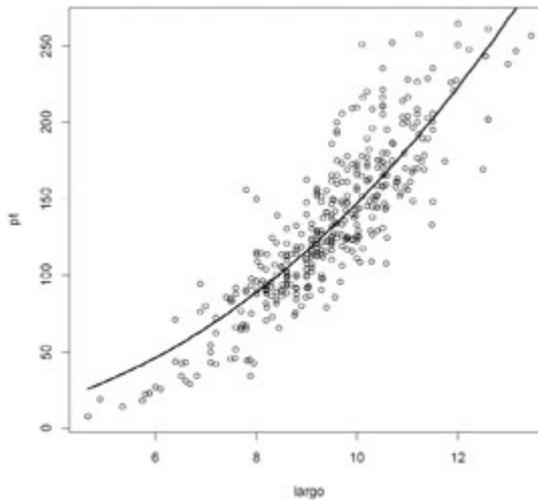


Figura 2.19. Relación Peso total – Largo de la concha del caracol marino *Hexaplex princeps*. El coeficiente alométrico de 2.27 indicó un crecimiento hipoalométrico estadísticamente significativo.



2.3.3. Pelecípodo *Hyotissa hiotis*

Para este ejemplo se consideraron los individuos del pelecípodo conocido comúnmente como “Callo Garra de León”, *Hyotissa hiotis* (Cordero-Montoya, 2016; Vázquez-Servín, 2017) recolectados ($N = 265$) de febrero 2016 a agosto 2017. Se consideró el peso de los individuos después de haber retirado a los epibiontes, (organismos que, debido al tamaño grande de esta especie, utilizan su concha como sustrato para fijarse) con el largo de la concha.

Los comandos para Stata son:

```
. use garraleon
. nl potfun psepi largo if largo>10,
(obs = 251)

Iteration 0:   residual SS = 6.36e+07
Iteration 1:   residual SS = 6.36e+07

Iteration 21:  residual SS = 8650080
Iteration 22:  residual SS = 8647743
Iteration 23:  residual SS = 8647743

-----+-----
Source |          SS      df      MS          Number of obs =      251
-----+-----+-----
Model | 56231636.2      2 28115818.1          F( 2, 249) =      809.56
Residual | 8647742.52    249 34729.8896          Prob > F      =      0.0000
-----+-----+-----
Total | 64879378.7    251 258483.581          R-squared      =      0.8667
                                          Adj R-squared  =      0.8656
                                          Root MSE      =     186.3596
                                          Res. dev.     =    3334.593

Potential function, psepi=b0*largo^b1
-----+-----
psepi |          Coef.   Std. Err.      t    P>|t|    [95% Conf. Interval]
-----+-----+-----
b0 |   .1103142     .0716924     1.54   0.125   - .0308866   .2515151
b1 |   1.707824     .1306817    13.07   0.000   1.450441   1.965206
-----+-----

(SEs, P values, CIs, and correlations are asymptotic approximations)

. predict psepie if largo > 10

. scatter psepi psepie largo if largo>10, ms(oh i) c(. 1) ytitle("Peso
sin epibiontes (g)") xtitle("Largo de la concha (mm)") sort legend(label(1
"Observados") label(2 "Calculados"))
```

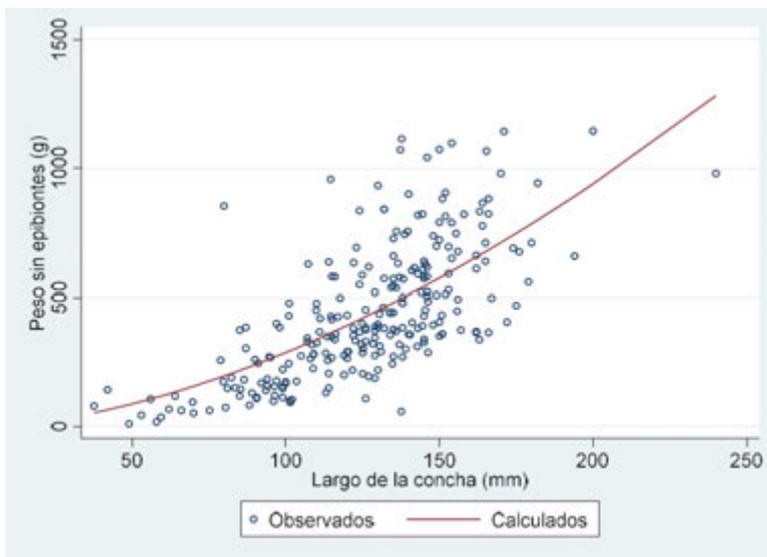


Figura 2.20. Relación Peso sin epibiontes – Largo de la concha del pelecípodo marino *Hyotissa hyotis*. El coeficiente alométrico de 1.71 indicó un crecimiento marcadamente hipoalométrico estadísticamente significativo.

Para este ajuste no se consideraron a los individuos menores a 10 mm de largo de concha (Figura 2.20). El coeficiente alométrico para esta relación resultó ser marcadamente hipoalométrico (1.71) comparado con la isometría (exponente alométrico = 3).

2.4 Relaciones entre dimensiones de volumen

Las relaciones entre variables relacionadas con el volumen de los organismos como el peso, muestran tendencias lineales. Además de permitir uso práctico para estimación de una variable conociendo la otra, la comparación de pesos del cuerpo o de las estructuras orgánicas proporciona valiosa información ecológica. Si se analizan los pesos de las gónadas, se tiene información acerca del ciclo reproductivo de las especies; si el interés es sobre estructuras tróficas (peso del contenido estomacal, del estómago, del tracto digestivo)



se obtiene valiosa información acerca de aspectos tróficos y su variación a lo largo de los ciclos estacionales de la localidad.

Ejemplos de relaciones Peso – Peso.

Relación Peso total – Peso eviscerado de bagres (Figura 2.21):

```
> bagt <- read.dta("bagtoto2.dta")
> bagtm <- subset(bagt, sex==1)
> rptpebm <- nls(totwei~potfun(eviwei,a,b),data=bagtm,start=c(a=.5,b=.5))
> summary(rptpebm)
```

Formula: totwei ~ potfun(eviwei, a, b)

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
a	1.218175	0.024334	50.06	<2e-16 ***
b	0.975897	0.004224	231.05	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.027 on 699 degrees of freedom

Number of iterations to convergence: 7

Achieved convergence tolerance: 2.802e-07

(12 observations deleted due to missingness)

```
> confint(rptpebm)
```

Waiting for profiling to be done...

	2.5%	97.5%
a	1.1710310	1.2669869
b	0.9675873	0.9842399

```
> tmp <- range(na.omit(bagtm$eviwei))
> xm <- seq(tmp[1],tmp[2],length.out=99)
> yem <- predict(rptpebm, data.frame(eviwei=xm))
> plot(totwei~eviwei, data=bagtm)
> lines(yem~xm, lwd=3, col="Red")
```

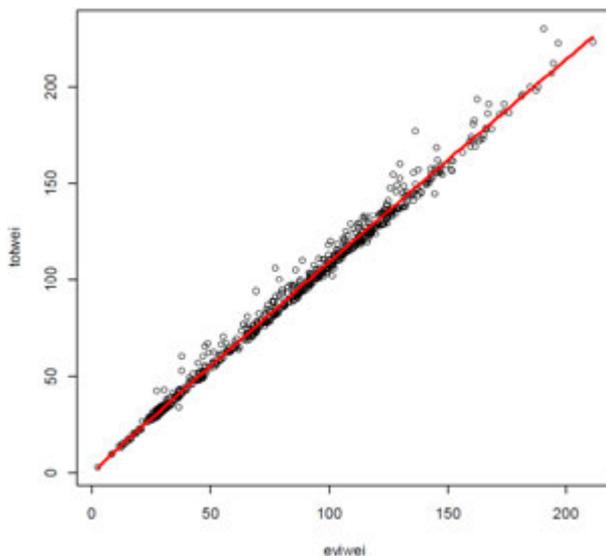


Figura 2.21. Relación Peso total – Peso eviscerado del bagre estuarino *Cathorops melanopus* (machos). El coeficiente alométrico de 0.98 indicó un crecimiento hipoalométrico estadísticamente significativo, cercano a la isometría.

Peso del tracto digestivo – Peso total de *Haemulon flaviguttatum*

Se usan los datos de pesos del tracto digestivo y del organismo completo del “blanco” (*Haemulon flaviguttatum*) capturados de agosto de 2011 a octubre de 2013 y reportados por Cifuentes-Salazar (2017).

```
. use trofipeces

. nl pfun2: tdig ptol if ncien=="Haemulon flaviguttatum" & (ptol !=. & tdig !=.)
(obs = 175)

Iteration 0: residual SS = 4608.74
Iteration 1: residual SS = 1581.38
...
Iteration 7: residual SS = 832.9761
```



Iteration 8: residual SS = 832.9761

Source	SS	df	MS			
Model	5111.3828	2	2555.69142	Number of obs =	175	
Residual	832.97611	173	4.81489078	R-squared =	0.8599	
Total	5944.359	175	33.9677654	Adj R-squared =	0.8583	
Potential function $tdig = a * ptol^b$				Root MSE =	2.194286	
				Res. dev. =	769.6668	
tdig	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/a	.0105423	.0050879	2.07	0.040	.0005	.0205845
/b	1.153536	.0866347	13.31	0.000	.9825385	1.324533

```
. predict ptdehf if ncien=="Haemulon flaviguttatum"
(option yhat assumed; fitted values)
(538 missing values generated)
```

```
. scatter tdig ptdehf ptol if ncien=="Haemulon flaviguttatum" & (ptol !=.
& tdig !=.), ms(oh i) c(. 1) sort ytitle("Peso del tracto digestivo (g)")
xtitle("Peso total (g)") || lfit tdig ptol if ncien=="Haemulon flaviguttatum"
& (ptol !=. & tdig !=.), lpattern(dash) legend(rows(1) label(1 "Observados")
label(2 "Calculados") label(3 "Ajuste lineal"))
```

El coeficiente alométrico de 1.15 indica un crecimiento hiperalométrico, pero el intervalo de confianza (.98,1.32) no rechaza a la isometría entre el peso del tracto digestivo y el peso total de los organismos. Los datos mostraron una dispersión creciente con el tamaño y una ligera tendencia curvilínea como puede apreciarse con el ajuste lineal indicado con la línea punteada en la Figura 2.22.

La marcada variabilidad alrededor de las curvas de ajuste se explica por los pesos diferentes debidos a la circunstancia de cada individuo respecto a su estado de repleción, es decir si fue capturado después de haberse alimentado o si al momento de la captura no había podido capturar presas para alimentarse. *H. flaviguttatum* es un depredador activo del ecosistema que se alimenta principalmente de crustáceos (Barquera-Rebolledo, 2018).

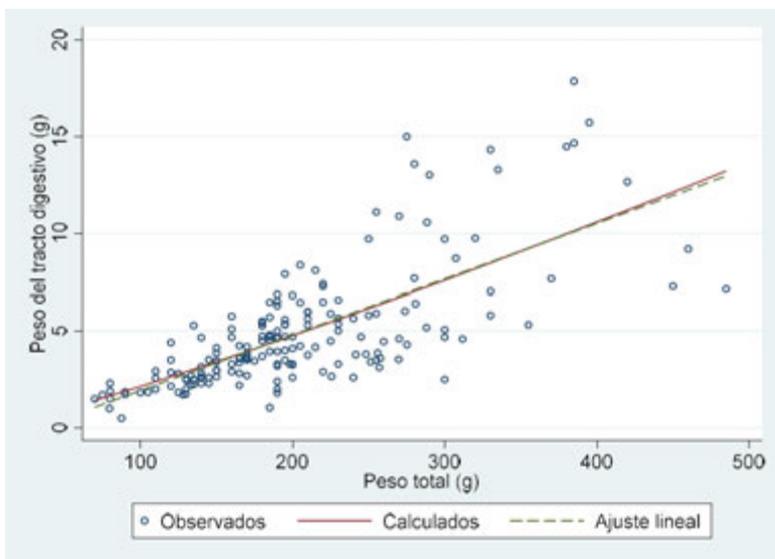


Figura 2.22. Relación (potencial) Peso tracto digestivo – Peso total del “blanco” *Haemulon flaviguttatum*. El coeficiente alométrico de 1.15 no fue significativamente diferente de la isometría. Se compara con un ajuste lineal (línea punteada).

Capítulo 3

Estimación de la edad



3.1. Introducción

Entre los aspectos biológicos más importantes que se requieren para conseguir un manejo adecuado de una población de peces está la determinación de la edad y el crecimiento. Estas características son determinantes para la propuesta de tamaños de malla de las redes que garanticen la captura de animales adultos con crecimiento lento y evitar el retirar de la población a los organismos antes de que se reproduzcan (Saito-Quezada, 2004).

El conocimiento de la composición de edades de una población de peces es esencial para un adecuado manejo de los recursos. La información correcta de la edad es necesaria para predicciones de longevidad, establecimiento de registros del índice de crecimiento, edad de madurez, edad a la cual se realizan migraciones y el conocimiento de períodos críticos en la vida de una población (Rounsefell y Everhart, 1953; Everhart y Youngs, 1981; Morales-Linares, 2006).

Para determinar el patrón de crecimiento de una especie es indispensable conocer el tamaño del organismo y su edad.

La edad de los individuos que viven en la naturaleza no siempre es fácil de conocer o estimar.

De acuerdo a Gómez-Márquez, *et al.* (2016) para estimar la edad de los peces (y de otros organismos acuáticos) existen tres tipos de métodos generales:

- Estimación directa, la cual es posible al criar a los organismos.
 - Estimación indirecta, la cual en general se refiere al análisis de la frecuencia de tallas que bajo ciertas condiciones es multimodal con cada moda representando a una cohorte, o grupo de organismos con la misma edad.
 - Estimación por lectura de marcas periódicas en estructuras óseas.
-



En lo que sigue nos enfocaremos a este último tipo de método. Las estructuras esqueléticas de los organismos, al formarse por la acumulación sucesiva de materiales orgánicos e inorgánicos contienen información sobre las condiciones ambientales y fisiológicas de los organismos que las forman. El estudio de la dimensión temporal a partir de estructuras duras en los organismos se ha denominado “Esclerocronología”. Esta disciplina cubre los problemas de la estimación de la edad tanto como la estimación del tiempo, la duración y los eventos en la vida. Los métodos esclerocronológicos están basados en el estudio de varios tipos de señales que proveen referencias temporales, ya sea estructurales, químicas y ópticas (Panfili *et al.*, 2002).

El tipo de estructura utilizado depende del tipo de organismo estudiado. En el caso de los peces las estructuras más usadas son:

- Escamas
- Otolitos
- Vértebras
- Radios o espinas
- Opérculos

Para organismos invertebrados, la estructura también depende del grupo taxonómico:

Moluscos:

- Conchas (Gasterópodos)
- Valvas (Pelecípodos)
- Opérculos (Gasterópodos)
- Estatolitos (Gasterópodos)
- Picos (Cefalópodos)

Equinoideos:

- Madreporita
- Placas del caparazón



3.2. Escamas. Relación Cuerpo-Escama

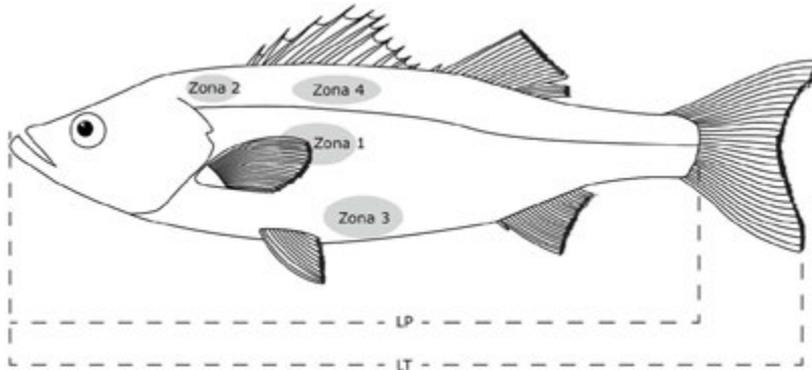


Figura 3.1. Zonas corporales del pez donde comúnmente se obtienen muestras de escamas; LP = longitud patrón; LT = longitud total (dibujo por E.R. Zambrano modificado de Salgado-Ugarte, 1995; Salgado-Ugarte, 2016).

Para ejemplificar los métodos para el procesamiento cuantitativo de los datos obtenidos del análisis de las escamas se utilizarán los datos tomados de Salgado-Ugarte (1995; 2016). Estos datos contienen la información de 2804 escamas extraídas de cuatro zonas corporales (Figura 3.1) del “Japanese sea bass” (*Lateolabrax japonicus*) proveniente de la pesca comercial de la Bahía de Tokio, Japón. Las zonas se escogieron considerando las recomendaciones generales (Bagenal y Tesch, 1978) y para comparación directa con otros trabajos previos acerca de la especie (Kuwatani, 1962; Hatanaka y Sekino, 1962b).

Las escamas fueron lavadas con agua para remover mucosidad y restos orgánicos adheridos. Ya limpias, varias de ellas en buen estado y no regeneradas, se montaron entre dos portaobjetos de vidrio para su posterior tratamiento. Las escamas fueron medidas de acuerdo al esquema de la Figura 3.2 ayuda de un proyector de Perfiles Nikon V-16e con accesorio para impresión de datos y platina micrométrica.

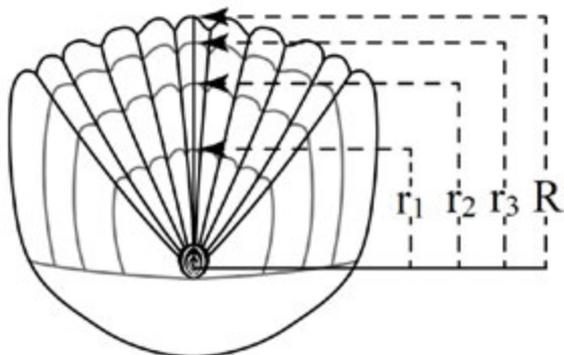


Figura 3.2. Esquema con las medidas tomadas a partir del Foco de las escamas: R , radio de la escama; r_i radio del anillo i (dibujo por E.R. Zambrano modificado de Salgado-Ugarte, 1995).

Los gráficos de la relación cuerpo-escama se presentan en la Figura 3.3 que puede realizarse por medio del comando y opciones de Stata:

```
. use scale111

. scatter bodlen rad, by(zone, note(" ") compact) ms(0h) ytitle("Longitud
estándar (mm)") xtitle("Radio de la escama (mm)") subtitle("Zona", prefix)
```

En estos gráficos es posible apreciar que la relación lineal más clara para las escamas es la de la zona 4. Lo que también se verifica con los valores del coeficiente de determinación (r^2) y del error cuadrado medio (error de estimación) para cada conjunto de datos por Zona (Cuadro 3.1).

Los comandos de Stata para construir el Cuadro 3.1 son:

```
. regress bodlen rad if zone==1
```

El cual se repite para los valores de cada zona restante (zone = 2, zone = 3, zone = 4).



Cuadro 3.1. Valores estadísticos de las regresiones de la relación cuerpo-escama para cada zona corporal de *Lateolabrax japonicus*.

Zona	ECM	r ²
1	55.602	0.6035
2	52.611	0.6294
3	47.359	0.7566
4	34.772	0.8754

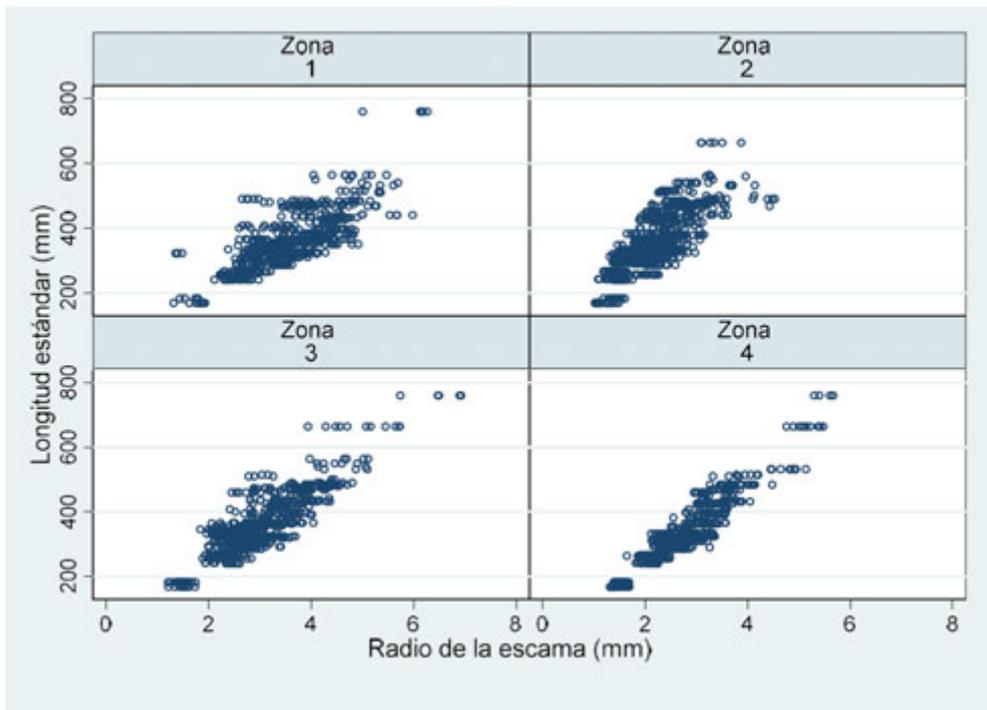


Figura 3.3. Comparación de las relaciones entre la longitud del cuerpo y el radio de la escama para las 4 zonas corporales consideradas y mostradas en la Figura 3.1



Una forma conveniente de distinguir el borde de la escama de los radios de los anillos es codificarlo con un número muy diferente a la cantidad de radios que se espera encontrar en los individuos. Para estos datos, el valor que se escogió fue “30”.

Si se asume que los anillos se forman cada año, entonces es posible asignar una edad que equivale al número máximo de anillos de cada individuo (menores que el código “30” que indica al borde). De esta forma es posible organizar la información en un cuadro estadístico. Para incluir a los organismos sin anillo y asignados a la edad “0” se tiene que cambiar su código:

```
. replace rad= 0 if rad==30 & age==0
. table age rad if rad<30 & zone==4, c(mean radlen sd radlen n radlen)
format(%9.3f) row
```

El resultado se muestra en el Cuadro 3.2, el cual permite hacer cálculos posteriores para la modelación del crecimiento.

Cuadro 3.2. Resumen estadístico (media, desviación estándar y frecuencia) de radios por edad para total de individuos del “suzuki” (*Lateolabrax japonicus*) n = 153.

Edad	Anillo											
	0	1	2	3	4	5	6	7	8	9	10	11
0	1.301											
	0.285											
	22											
1		1.532										
		0.331										
		55										
2		1.539	2.489									
		0.261	0.3									



Cuadro 3.2. Resumen estadístico (media, desviación estándar y frecuencia) de radios por edad para total de individuos del “suzuki” (*Lateolabrax japonicus*) n = 153 (continuación).

Edad	Anillo											
	0	1	2	3	4	5	6	7	8	9	10	11
		54	54									
3		1.671	2.643	3.158								
		0.281	0.264	0.331								
		15	15	15								
4		1.799	2.765	3.322	3.582							
		0	0	0	0							
		1	1	1	1							
5		2.194	3.123	3.536	3.837	4.032						
		0.211	0.001	0.272	0.517	0.699						
		2	2	2	2	2						
6		1.309	2.357	2.731	3.036	3.292	3.43					
		0.239	0.099	0.241	0.395	0.617	0.704					
		2	2	2	2	2	2					
7		1.508	2.628	3.054	4.158	4.383	4.773	5.028				
		0	0	0	0	0	0	0				
		1	1	1	1	1	1	1				
11		1.239	2.086	2.537	2.918	3.281	3.696	3.963	4.253	4.398	4.812	5.054
		0	0	0	0	0	0	0	0	0	0	0
		1	1	1	1	1	1	1	1	1	1	1
Total	1.301	1.557	2.533	3.128	3.486	3.718	3.832	4.496	4.253	4.398	4.812	5.054
	0.285	0.305	0.307	0.36	0.556	0.642	0.758	0.753				
	22	131	76	22	7	6	4	2	1	1	1	1



Con los datos del archivo scale111.dta es posible también utilizar el incremento marginal para validar la periodicidad de formación de los anillos en las escamas.

```
. sum marinc if zone==4 & age==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
marinc	55	.3607916	.1137075	.0622318	.5507569

```
. graph box marinc if zone==4 & age==1, over(cdate) yline(.36) ytitle("Incremento marginal (mm)")
```

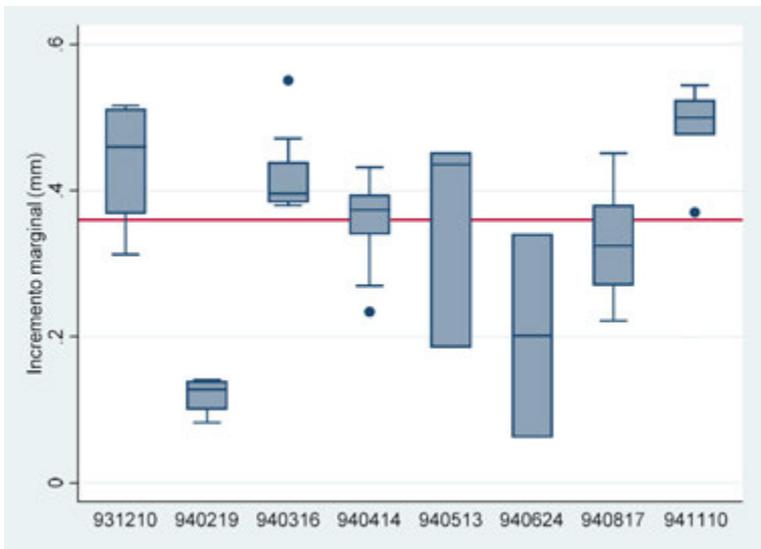


Figura 3.4. Diagramas de caja y bigotes para valores de incremento marginal de escamas de *L. japonicus* con 1 anillo de crecimiento (fecha en formato año-mes-día). La línea horizontal es el valor medio de los incrementos marginales.



En este gráfico se puede apreciar claramente un mínimo durante el mes de febrero, si bien en junio se observa un valor mediano por debajo del promedio con mucha variabilidad (caja muy amplia).

3.3. Otolitos

Los otolitos son, en la mayor parte de los casos una estructura confiable y precisa para estimar la edad de los peces. Comparadas con las escamas poseen la ventaja de no ser afectados por la reabsorción (sin embargo, al respecto consultar las obras de Mugiya y Uchimura, 1989). Para muchas especies se conoce que las escamas sistemáticamente subestiman la edad de los peces más viejos del punto del crecimiento asintótico (Beamish y Chilton, 1981; Beamish y McFarlane, 1983; Lowerre-Barbieri *et al.*, 1994). Sin embargo, el método implica la muerte del pez. Aún con especímenes capturados, la extracción de estas estructuras puede alterar su calidad como producto de consumo. No obstante, cuando se requiere de una estimación de edad precisa, esta puede obtenerse de la observación y lectura de los otolitos.

Los otolitos pueden examinarse a dos niveles principales: contar las marcas regulares o la enumeración de incrementos diarios de crecimiento (Brothers, 1987). En el primer caso, las marcas pueden observarse en el otolito completo o bien “chamuscado”, pulido, cortado o cortado y pulido, teñido o sumergido en varios agentes líquidos aclarantes. Descripciones metodológicas detalladas se presentan en Blacker (1974), Bagenal y Tesch (1978), Chilton y Beamish (1982), Izuka y Katayama (2008).

Hace algún tiempo, los estudios pioneros de Irie (1955, 1957) sobre la formación de los otolitos contienen una descripción muy detallada del desarrollo de los otolitos de *Lateolabrax japonicus*. Encontró que:

- a) Las zonas translúcidas (hialinas) se forman de verano a otoño y que las opacas de invierno a primavera.
- b) El carbonato de calcio se deposita en la superficie del otolito de verano a otoño cuando se considera que los peces están activos e ingieren alimento vigorosamente; por otro lado, se deposita una gran cantidad de material orgánico y poco carbonato de calcio de invierno a primavera (principalmente durante la primavera) cuando se considera que los peces están relativamente inactivos e ingieren poco alimento.



- c) Cada año se forma una zona translúcida y una opaca y por tanto estas zonas pueden utilizarse para la estimación de la edad de manera similar a las marcas de crecimiento de las escamas.

Este autor llevó a cabo estudios posteriores (Irie, 1960) con otras especies y confirmó lo reportado anteriormente: el calcio disuelto en el agua de mar se deposita en la superficie del otolito; en verano y otoño, cuando el pez crece rápidamente, se deposita una gran cantidad de calcio mientras que la proteína es escasa (formación de zona translúcida); en invierno y primavera, cuando el pez crece lentamente, se deposita poca cantidad de carbonato cálcico y la proteína llena los surcos entre los granos cristalinos del mineral (formación de la zona opaca). Encontró que la temperatura y la cantidad de alimento ingerida por el pez afectan la formación de granos microcristalinos y no la cantidad de mineral disuelto en el agua. Cuando el ambiente es favorable y la condición fisiológica del pez es buena, los cristales se forman abundantemente produciendo un otolito de crecimiento rápido independientemente de la concentración de calcio en el agua. Por el contrario, en un ambiente desfavorable, casi no hay formación de microcristales y el otolito crece lentamente. Un ambiente bueno implica que el pez vive en temperatura cálida y puede ingerir alimento en abundancia; las condiciones desfavorables implican que el pez vive en aguas frías y se alimenta poco. Estas condiciones ocurren durante el verano-otoño y del invierno a la primavera en las zonas templadas. Otra conclusión importante de sus estudios es que la formación de las zonas en el otolito está conectada íntimamente con la historia del pez y, una vez formadas, quedan preservadas. Es por esto que el otolito puede utilizarse como una guía efectiva para investigar la historia de vida de un individuo (Salgado-Ugarte, 1995; 2016).

El primer ejemplo presenta los datos obtenidos por Salgado-Ugarte (1995) también publicados posteriormente en Salgado-Ugarte (2016). Consisten en un total de 407 otolitos sagita (205 izquierdos y 202 derechos) que se midieron y extrajeron de un total de 216 peces. Los especímenes se midieron (longitudes total y patrón) y pesaron con una precisión de 1 mm y 0.1 g respectivamente. Los peces se disectaron para determinar el sexo y los dos otolitos sagita se extrajeron cortando la parte superior del cráneo (método “open-hatch” descrito por Secor *et al.* 1991), lavados con agua para remover membranas y otros restos de tejidos, secados con papel absorbente y almacenados en frascos de vidrio. Cada otolito se pesó con una balanza analítica (Inaba Seisakusho Ltd) con precisión de 0.1 mg.

Se llevó a cabo un análisis preliminar para medir largo y ancho de cada uno de 154 otolitos con ayuda de un proyector de perfiles (Nikon, V-16e). Posteriormente, los otolitos completos (con el lado cóncavo hacia arriba) en glicerina y examinados microscópicamente con luz reflejada contra un fondo negro para medir los radios del otolito y sus anillos con 0.01mm



de precisión con el accesorio de medición anexo al microscopio. Bajo esta iluminación, las zonas opacas se observan blanquecinas y las translúcidas oscuras. Se consideraron cinco direcciones radiales a partir del núcleo para medir los otolitos (Figura 3.5).

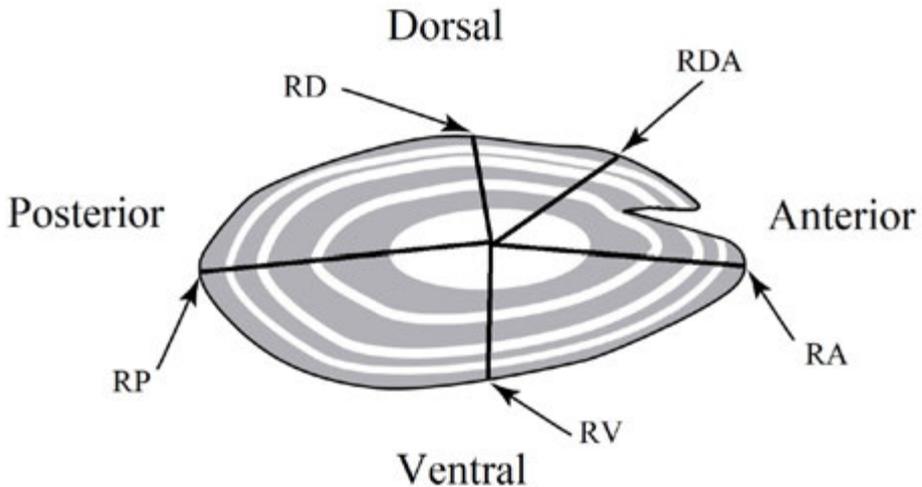


Figura 3.5. Esquema que muestra los ejes de medición (otolito izquierdo). Las zonas opacas se indican en blanco y las translúcidas están sombreadas. RA, radio anterior; RP, radio posterior; RD, radio dorsal; RV, radio ventral; RDA, radio antero-dorsal (dibujo por E.R. Zambrano modificado de Salgado-Ugarte, 1995).

El otolito derecho de cada individuo (cuando estaba dañado o cristalizado, se tomó el otolito izquierdo) se incluyó en resina poliéster y se seccionó transversalmente a través del núcleo con una máquina de corte de precisión (Exact Cutting Grinding System, BS-3000). Las secciones de aproximadamente 0.5 mm de grosor se colocaron entre dos portaobjetos y se observaron con luz transmitida en campo brillante con un proyector de perfiles (Nikon, V-16e) a 20x. Se consideraron dos ejes para medir los radios: del núcleo al margen ventral (radio ventral de sección, RVS) y del núcleo uno de los márgenes proximales (radio proximal de sección, RPXS) a lo largo de la muesca del surco (Figura 3.6)

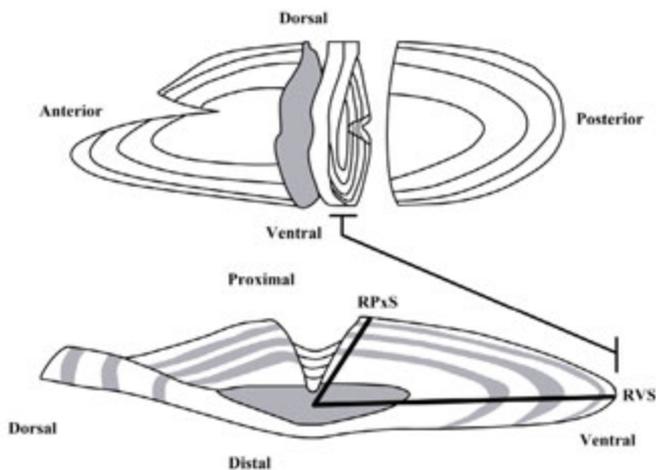


Figura 3.6a. Esquema que muestra la sección del otolito y los ejes radiales considerados para su medición (otolito derecho). Las secciones se observaron con luz transmitida por lo que las zonas translúcidas se indican en color blanco y las opacas están sombreadas. RVS, radio ventral de sección; RPxS, radio proximal de sección (dibujo por E.R. Zambrano modificado de Salgado-Ugarte, 1995).

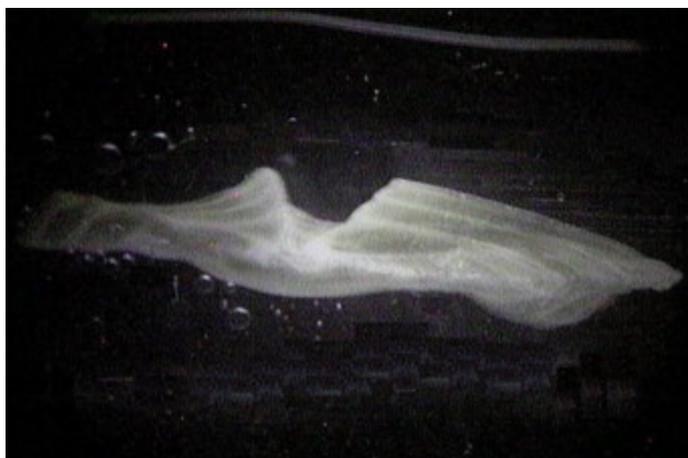


Figura 3.6b. Sección de otolito de *Lateolabrax japonicus* aproximadamente equivalente al esquema de la Figura 3.6a.



3.4. Relación Cuerpo-Otolito

La exploración preliminar de los largos, anchos y pesos de los otolitos izquierdo y derecho se muestran en la Figura 3.7. Como se esperaba, las medidas de longitud exhiben relaciones lineales. Las dimensiones lineales con los pesos muestran relaciones no lineales, pero éstos, entre sí se relacionan linealmente. La relación entre largo y ancho de otolitos muestra un ligero incremento de la variabilidad al aumentar las dimensiones.

```
. use bodoto2
. graph matrix bodlen olel owil oler owir otowel otower, half ms(+) iscale(*1.3)
```

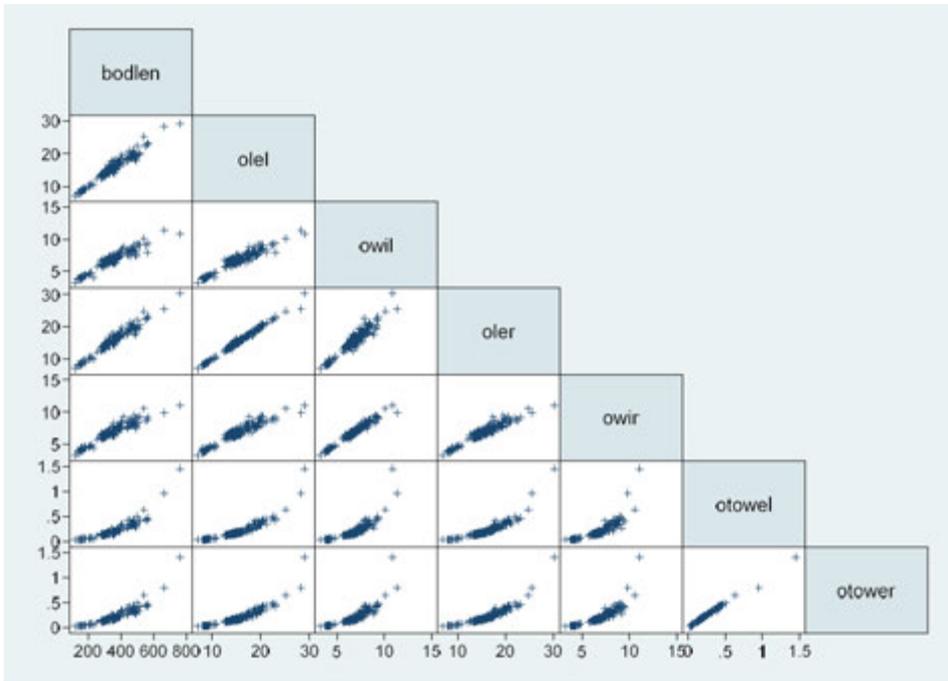


Figura 3.7. Matriz de gráficos de dispersión (diagonal inferior) para largo, ancho y peso de otolitos izquierdo y derecho; bodlen, longitud estándar; ole(l,r), longitud del otolito (izquierdo, derecho); owi(l,r), ancho del otolito (izquierdo, derecho); otowe(l,r), peso del otolito (izquierdo, derecho) *Lateolabrax japonicus*.



3.5. Validación de la periodicidad de las marcas

Para validar la periodicidad de la formación de las marcas se analizó el incremento marginal. Si bien todos los incrementos marginales tuvieron un comportamiento semejante a lo largo del estudio, el correspondiente a la dirección dorsal fue el mas claro y es el que se presenta en la Figura 3.8:

```
. use otage113
. graph box dmarinc , over(cdate, lab(alternate)) ytitle("Incremento marginal (dorsal)")
```

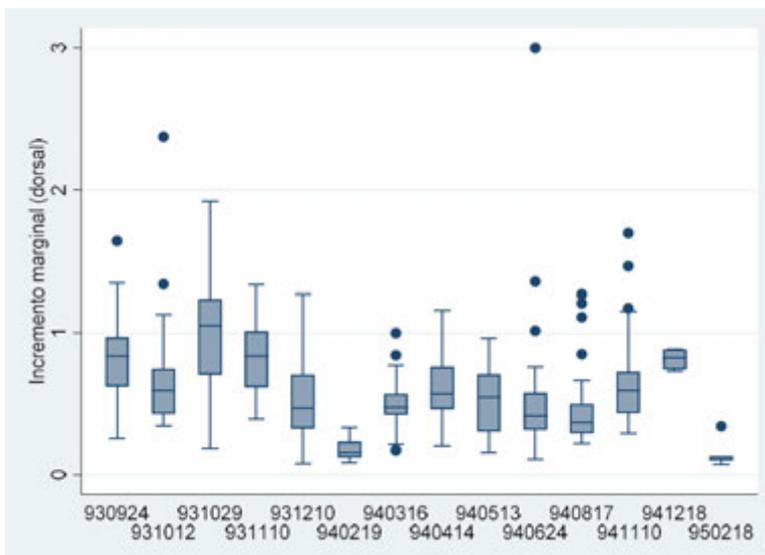


Figura 3.8. Diagramas de caja y bigotes para el incremento marginal de los otolitos de *Lateolabrax japonicus* a lo largo del eje dorsal por fecha de muestreo

Claramente puede apreciarse que los valores mínimos se tienen en la temporada fría del año (febrero) lo que indica la formación de la zona opaca del crecimiento. Esto concuerda con lo reportado en estudios previos mencionados anteriormente (Irie, 1955; 1957 y 1960), lo que respalda fuertemente la periodicidad anual de las marcas en los otolitos de *L. japonicus* de la Bahía de Tokio (Salgado-Ugarte, 1995; 2016).



Al comparar los resultados de las distintas orientaciones medidas, resalta la existencia de numerosas marcas no periódicas a lo largo de los ejes anterior y posterior. Por tanto, los ejes dorsal y antero dorsal fueron los más apropiados para estimar el número de marcas y por tanto la edad.

El examen de los otolitos completos permitió la estimación de un ejemplar hembra de 15 años y un macho de 14. Anteriormente, el espécimen de más edad reportado tuvo 8 años.

3.6. Opérculos del gasterópodo *Astraea undosa*

Astraea undosa (Wood, 1828), es uno de los grandes gasterópodos prosobranquios, conocido en inglés como “wavy turban” (Abbott 1974) y en México como “caracol panocha”. Habita en aguas moderadamente profundas, típico de zonas costeras templadas y se distribuye desde Ventura, California, EUA, hasta Punta Abreojos en la península de Baja California, México (Morris, 1969; Keen, 1971). Es estrictamente un gasterópodo herbívoro oportunista con ausencia de preferencia alimentaria diferencial por sexo. Su concha es grande, sólida y pesada, cónica, más ancha que alta, con las vueltas y la base aplanadas, alcanzando una longitud de 20 cm con un diámetro equivalente. Presenta de 6 a 8 vueltas con pendiente pronunciada, un ápice agudo. Cada vuelta está decorada con una serie de prominencias verticales onduladas y una banda nudosa prominente semejante a una cuerda enredada a lo largo de la línea de sutura. Tienen un opérculo calcáreo, engrosado (Morris, 1969), la cara externa presenta tres crestas espinosas que parten desde el borde, con periostraco córneo café-oliváceo y la cara externa blanquecina. De acuerdo con Checa y Jiménez-Jiménez (1998) se trata de un opérculo rigiclaudata espiral.

Esta especie se captura comercialmente en la península de Baja California y aunque su edad y crecimiento han sido estudiados anteriormente (Cupul-Magaña y Torres-Moye, 1996; Gluyas-Millán *et al.*, 1999; Gluyas-Millán, *et al.*, 2000), debido a su importancia como sucedáneo del abulón entre otras cosas, se hizo el estudio con ejemplares de “La Bocana”, lugar donde no existen antecedentes (Saito-Quezada, 2010). De este estudio se muestra el análisis de los opérculos.

Se obtuvo una muestra representativa de *Astraea undosa* de la captura comercial de la costa occidental de la península de Baja California (La Bocana). El muestreo se llevó a cabo en julio de 2006. Los caracoles fueron 287 ejemplares recolectados por medio de buceo semiautónomo tipo Hooka. A cada caracol se le midió el diámetro basal (*DB*), el peso total (*PT*) y el peso del callo (*PC*). Asimismo se conservó el opérculo. Para el análisis de los datos se utilizaron métodos estadísticos exploratorios (Salgado-Ugarte, 1992). Posteriormente en su caso, se aplicaron métodos estadísticos confirmatorios como la regresión lineal simple.



A los opérculos se les midió el largo y el ancho con un vernier (0.02 mm) y se pesaron en una balanza digital. En primera instancia, para determinar la edad se contaron los anillos de los opérculos sin pulir con luz reflejada.



Figura 3.9. Foto de opérculo de *Astraea undosa*; a la izquierda parte interior del opérculo con periostraco, a la derecha parte exterior del opérculo.

Para poder determinar el método esclerocronológico de corte y pulido adecuado a los opérculos de la muestra de *Astraea undosa*, como un paso preliminar se obtuvieron opérculos de *A. undosa* y *A. turbanica* en tiendas de conchas de La Paz, Baja California Sur. En laboratorio se hicieron pruebas para elegir entre varios métodos de corte y pulido, el más conveniente para la observación de los anillos de crecimiento: (a) los opérculos fueron sumergidos en agua, durante varios días hasta que facilitara retirar el periostraco; (b) tres opérculos de cada especie fueron seccionados, por medio de una cortadora Lapidary Combination Unit, Model LU6X, Lortor Inc., 2850 N.W., SEATTLE, WA 9819. Para ello se aisló la costilla central de las tres crestas presentes en el opérculo. En los opérculos de *A. turbanica* se podían obtener bien las crestas separadas, pero en *A. undosa* la cresta se quebraba cuando se querían pulir para poder hacer la observación de los anillos de crecimiento. (c) En otros opérculos de cada especie se pulieron las tres costillas del lado externo con cuidado para no romperlos y llegar a ver a trasluz las líneas de crecimiento; (d) Se observaron con luz transmitida y luz reflejada.



Figura 3.10. Foto de la cresta central aislada del opérculo de *A. undosa*.

Debido a las secciones centrales de los opérculos de *A. undosa* fueron muy quebradizas, se escogió el método de primero desbastar las crestas externas y pulir los opérculos para después observarlos con luz transmitida.



Figura 3.11. Foto del opérculo de *A. undosa* con las tres crestas cortadas y pulidas (luz reflejada).



Una vez que se determinó que los anillos se veían mejor en opérculos con las tres crestas pulidas, se optó por este método y se procedió a la lectura de los anillos de crecimiento. Se midieron las longitudes de los radios de cada anillo desde la zona considerada como focal cerca de la espira hasta el inicio de cada zona opaca situada después de cada banda translúcida y finalmente hasta el extremo del opérculo (borde) con un vernier (0.02mm).

Debido al crecimiento curvilíneo del opérculo los radios se midieron a partir de la zona focal hasta la parte media del anillo en el opérculo (Figura 3.12). La lectura se realizó en dos tiempos diferentes y de manera aleatoria, sin considerar el DB del caracol.

Cuando la primera lectura no coincidía con la segunda se solicitó la ayuda de un segundo lector (lo que sucedió pocas veces) quien definió el número de anillos considerado para posteriores estimaciones. La periodicidad de las marcas de crecimiento se consideró anual de acuerdo a los resultados de Cupul-Magaña *et al.*, (1996) y Gluyas-Millán, *et al.*, (1999). Por tanto, el número de anillos proporcionó la edad estimada para el ejemplar.

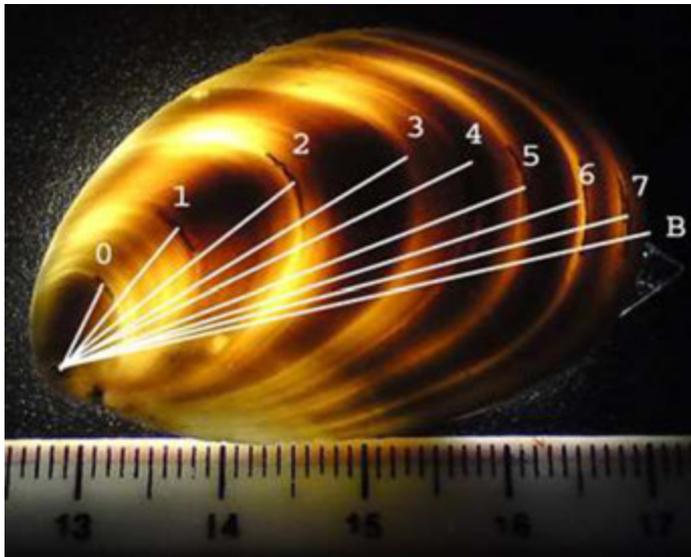


Figura 3.12. Foto del opérculo (luz transmitida) del opérculo de *A. undosa*, donde se muestran los anillos de crecimiento y la medición de cada uno de los siete anillos y el borde (B) a partir de la zona focal.



De los 153 opérculos de *A. undosa* algunos estaban fracturados y otros fragmentados por lo que fue posible analizar 147 sin pulir, en los cuales se observaron de 0 a 11 anillos, siendo los más frecuentes los opérculos con 5 (45) y 6 (41) anillos; sólo se distinguió un opérculo con dos anillos (Cuadro 3.3). La Figura 3.13 muestra la relación entre el diámetro basal y la edad estimada con los opérculos sin pulir.

```
. use sinpulir.dta
. tab asp, sum(diabas)
```

Al pulirlos, un número adicional de opérculos no pudo observarse por lo que sólo 140 pudieron considerarse.

Cuadro 3.3. Resumen de anillos en opérculos sin pulir y media del diámetro basal de *A. undosa*.

Número de anillo	Media del DB (mm)	Desviación estándar	Frecuencia
0	48	2.64	3
1	58.75	2.5	4
2	76	0	1
3	75	5.29	3
4	85	4.24	6
5	91.86	5.79	45
6	94.29	6.82	41
7	101.38	4.92	18
8	104.38	5.88	13
9	101.2	5.84	5
10	110.8	1.30	5
11	109	1	3
Total	93.59	12.44	147



Como el arreglo de los datos no es trivial, enseguida presentaremos un resumen breve de su tratamiento para obtener los resultados finales. Primeramente, la información se obtiene por individuo y se organiza en un cuadro como el siguiente (3.4). Este cuadro presenta los datos por cada individuo y se anotan los valores de los radios que cada uno tiene (aquí indicados como A1 hasta A11. El código A30 se usa para marcar el borde de cada óperculo.

. use astraeaanillos

Cuadro 3.4. Arreglo de datos de radios de óperculos por individuo (primeros 20).

No	radio1	radio2	radio3	radio4	radio5	radio6	radio7	radio8	radio9	radio10	radio11	radio30
1	0.726	1.414	2.34	2.948	3.28	3.451						3.668
2	1.98	1.92	2.692	3.118	3.426	3.642						3.746
3	0.94	1.922	2.568	3	3.152							3.452
4	0.692	1.534	2.128	2.776	3.1							3.26
5	1.952	3.106	3.334	3.682	4	4.272	4.294	4.364	4.43			4.492
6	1.41	2										2.3
7	1.1	2.646	3.03	3.452	3.678	3.952	4.05	4.2	4.374			4.528
8	0.946	1.796	2.382	2.624	2.756	2.848	3.198	3.326	3.57	3.668		3.81
9	0.9	1.606	2	2.48	2.91							3.9
10	1.202	2.212	3.108	3.668	4.128	4.422						4.892
11	1.098	1.898	2.216	2.83	3	3.192	3.442	3.658	3.8	3.85		3.9
12	1.256	2.178	2.562	2.9	3.158							3.278
13	0.808	1.642	2.342	2.832	3.368	3.744						3.984
14	1	2.02	2.42	2.94	3.1	3.356						3.5
15	1.276	1.978	2.472	2.674								2.79
16	0.9	1.562	1.95	2.574	2.774	2.862	2.98					3.2
17	1.17	1.9	2.466	2.842	3.174							3.354
18	1.318	2.9	3.246	3.752	3.932	4.156						4.282
19	1.118	1.6	2.178	2.324	2.572	2.85	3	3.21	3.442	3.698	3.852	3.974
20	0.54	1.27	1.866	2.692	3.284							3.452



Este arreglo de trabajo debe modificarse a un arreglo vertical para hacerlo coincidir con la información adicional para cada individuo (diámetro basal, pesos, fecha, etc.) para tener un cuadro como el 3.5. En Stata, existe un comando que puede reorganizar la información del Cuadro 3.4:

```
. reshape long radio, i(no) j(noradio)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 30)

Data                                wide  ->  long
-----
Number of obs.                      153  ->  1836
Number of variables                  13   ->    3
j variable (12 values)              ->  noradio
xij variables:
      radio1 radio2 ... radio30  ->  radio
```

Este comando pasa de un arreglo “amplio” (horizontal) a uno “largo” (vertical). A este archivo se le agregan los datos restantes para completar la información (Cuadro 3.5). Como se obtienen valores faltantes ya que no todos los individuos tienen el mismo número de anillos, se requiere eliminar los valores faltantes:

```
drop if radio==.
```

Cuadro 3.5. Información de los radios de opérculo reorganizada para análisis estadístico: DB = Diámetro basal; PT = Peso total; PC = Peso del callo. En número de radio = 30 indica el borde del opérculo (primeros dos individuos).

No.	DB	PT	PC	Fecha	Radio	No.radio
1	95	151.2	31.9	jul-06	0.726	1
1	95	151.2	31.9	jul-06	1.414	2
1	95	151.2	31.9	jul-06	2.34	3
1	95	151.2	31.9	jul-06	2.948	4
1	95	151.2	31.9	jul-06	3.28	5
1	95	151.2	31.9	jul-06	3.451	6
1	95	151.2	31.9	jul-06	3.668	30
2	102	301.9	42.1	jul-06	1.98	1
2	102	301.9	42.1	jul-06	1.92	2
2	102	301.9	42.1	jul-06	2.692	3



Cuadro 3.5. Información de los radios de opérculo reorganizada para análisis estadístico: DB = Diámetro basal; PT = Peso total; PC = Peso del callo. En número de radio = 30 indica el borde del opérculo (primeros dos individuos) (*continuación*).

No.	DB	PT	PC	Fecha	Radio	No.radio
2	102	301.9	42.1	jul-06	3.118	4
2	102	301.9	42.1	jul-06	3.426	5
2	102	301.9	42.1	jul-06	3.642	6
2	102	301.9	42.1	jul-06	3.746	30

Con los datos de este arreglo es posible hacer un resumen estadístico de los radios como el que se presenta en el cuadro 3.6. El número de radios fue de 1 hasta 11. Con Stata este resumen puede obtenerse de un archivo de datos:

```
. use astraadad2
. tabulate nuradio, sum(radio)
```

Cuadro 3.6. Resumen estadístico de los radios medidos en opérculos de *Astraea undosa*.

Número de Anillo	Media del radio	Desviación Estándar	Frecuencia
1	1.15	0.28	140
2	1.87	0.32	133
3	2.46	0.30	132
4	2.96	0.30	130
5	3.30	0.29	115
6	3.61	0.37	57
7	3.70	0.41	24
8	3.85	0.39	17
9	3.99	0.42	11
10	3.69	0.12	4
11	3.85	0.00	1
Total	2.67	0.95	905



Es conveniente verificar si la lectura de los radios fue adecuada. Morales-Nin (1992) sugiere hacer un polígono de frecuencia para los radios de cada anillo. Es de esperar una distribución unimodal (en teoría Gaussiana) para cada anillo. En la Figura 3.13 se muestran no los polígonos de frecuencia sino los estimadores de densidad por kernel de los radios de cada anillo en donde hubo al menos cuatro individuos (anillo 10).

```
. two kdensity radio if nuradio<30, by(nuradio, cols(1) yrescale compact
note(" ")) ysize(10) gauss xlabel(0(.5)4.5) xtitle("Longitud de radio (mm)")
yttitle("Densidad")
```

En esta figura se observa que en general las distribuciones son unimodales aunque en algunas hay indicación de dos modas, con la moda “verdadera” como dominante y las menores como valores correspondientes a otros anillos incorrectamente clasificados. El sesgo de algunas distribuciones pudiera sugerir el uso de una medida alterna a la media aritmética, en este caso, las modas dominantes de cada distribución.

Con esta información es posible estimar la edad, asumiendo la periodicidad anual propuesta por Gluyas-Millán *et al.* (1999).

```
. gen edad = nuradio[_n-1] if nuradio==30
```

Con este comando se genera una variable que sólo tiene valor cuando nuradio es igual a 30, es decir en el borde de los opérculos. En ocasiones es conveniente tener repetida la edad por cada observación, lo cual se consigue con la siguiente orden:

```
. by no: gen edad2 = nuradio[_N-1]
```

Esta orden de Stata aprovecha las capacidades de realizar operaciones por grupo (en este caso “no” que identifica a cada individuo) y el uso de subíndices para las observaciones. “_N” especifica a la última posición de cada grupo (individuo). Recordemos que cada individuo tiene varias observaciones, cada una correspondiendo a un valor de radio de opérculo.

Una vez estimada la edad, es posible ahora observar la relación entre la talla y la edad. Un resumen de esta relación se presenta en el Cuadro 3.7 y en la Figura 3.14:

```
. tab edad2 if nuradio==30, sum(lontot)

. scatter lontot edad if edad==nuradio, ms(Oh) ytitle("Diámetro basal (mm)")
xtitle("Edad (años)") xlabel(0(1)11)
```

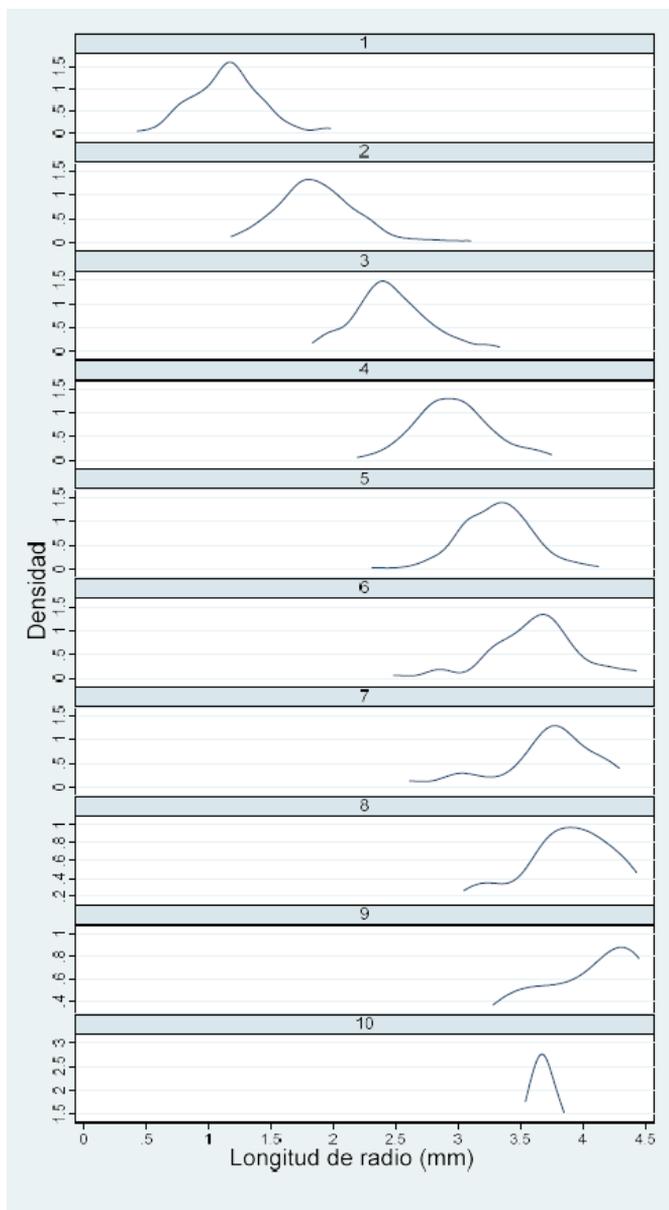


Figura 3.13. EDKs Gaussianos para los radios de los anillos de opérculos pulidos de *Astraea undosa*.



Cuadro 3.7. Resumen de Diámetro Basal (mm) y edad de individuos de *A. undosa*.

Edad	Media	Desviación estándar	Frecuencia
1	58.14	6.20	7
2	76.00	0.00	1
3	74.00	7.07	2
4	88.33	6.47	15
5	92.53	5.24	58
6	100.50	5.62	34
7	99.33	7.22	6
8	105.50	4.37	6
9	111.00	5.03	7
10	104.66	6.11	3
11	108.00	0.00	1
Total	93.85	12.5437	140

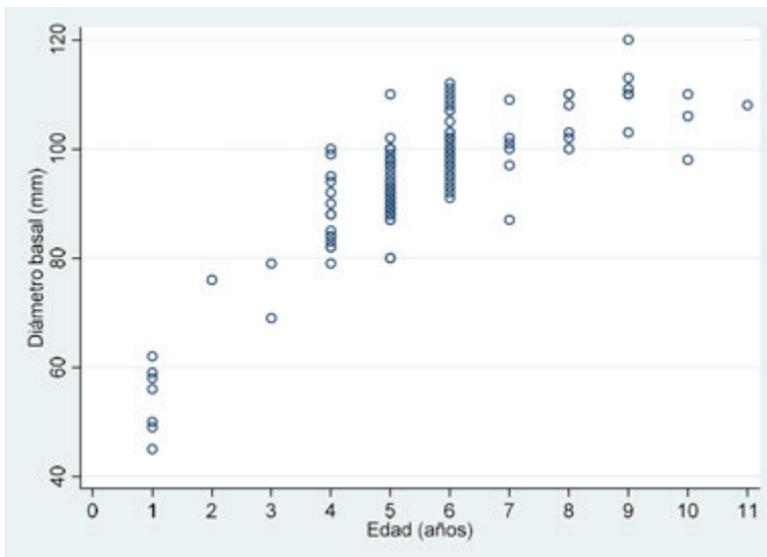


Figura 3.14. Gráfica de dispersión del diámetro basal contra la edad de *Astraea undosa*.



3.7. Otolitos de la tilapia *Oreochromis niloticus*. Validación de periodicidad

En esta sección mostraremos un repaso de tres métodos utilizados para validar la periodicidad en la formación de marcas de crecimiento, tomando como ejemplo los datos de la Tilapia *Oreochromis niloticus* de la presa “Emiliano Zapata” en el Estado de Morelos, recolectados mensualmente de enero 1998 a febrero de 1999 (Saito-Quezada, 2004). Los ejemplares fueron obtenidos de la pesca local artesanal realizada por los lugareños por medio de atarraya con 6.5 cm de luz de malla.

Cada mes se midieron y sexaron 100 individuos y de éstos, al azar se seleccionaron 30 para análisis más detallados, entre ellos la extracción, preparación y lectura de otolitos (Figuras 3.15 y 3.17).

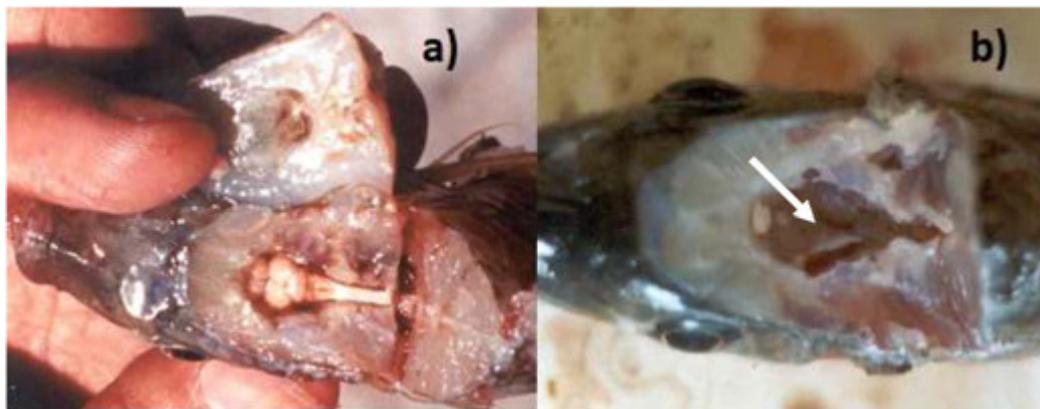


Figura 3.15. a) Corte de la cabeza de un espécimen exponiendo el cerebro y b) corte con el cerebro removido para poder observar los otolitos señalados por la flecha (abajo y arriba).

Los datos recabados fueron organizados en un arreglo vertical de las variables muestra, fecha, individuo, anillo, radio, longitud patrón, incremento marginal, hialino, opaco, número de días, entre otras. Con este arreglo general se llevaron a cabo los cálculos necesarios.

. use miotome7



Cuadro 3.8. Arreglo vertical de datos para cálculo de porcentaje de tipo de borde, incremento marginal y razón de incremento marginal: 1) No. de muestra; 2) No. de individuo; 3) Sexo; 4) Longitud patrón (cm); 5) No. de anillo (40 indica al borde; 6) Radio de anillo; 7) Incremento marginal; 8) Razón de incremento marginal; 9) Indicador de borde opaco; 10) indicador de borde hialino.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	1	1	14.7	1	1.00
1	1	1	14.7	2	1.50
1	1	1	14.7	3	1.99
1	1	1	14.7	4	2.55
1	1	1	14.7	5	3.5
1	1	1	14.7	40	3.96	0.46	0.4842	1	.
1	2	2	12.3
1	3	1	12.6	1	1.30
1	3	1	12.6	2	1.90
1	3	1	12.6	40	3.19	1.29	2.1500	.	1
1	4	2	12.5	1	1.86
1	4	2	12.5	40	3.53	1.67	0.8978	.	1
1	5	1	12.3	1	1.80
1	5	1	12.3	2	2.60
1	5	1	12.3	40	3.20	0.60	0.7500	1	.
1	6	1	13.0	1	1.20
1	6	1	13.0	2	2.10
1	6	1	13.0	40	3.40	1.30	1.4444	.	1
1	7	1	12.9	1	2.00
1	7	1	12.9	2	2.90



3.7.1. Análisis de bordes

El tipo de borde de los otolitos (opaco o hialino) es importante ya que diversos estudios pioneros (Irie, 1955; 1957; 1960) verificados recientemente (Fablet *et al.*, 2011) establecen que las zonas translúcidas se forman de la acreción por precipitación de aragonita (forma cristalina del carbonato de calcio) a temperaturas elevadas con alta disponibilidad de alimento y las zonas opacas por acumulación predominante de material orgánico (proteína otolina) durante periodos de bajas temperaturas y poco alimento (Salgado-Ugarte, 1995; 2016).

Los resultados para el tipo de borde de estos individuos se presentan en el Cuadro 3.9 y gráficamente en la Figura 3.16. En ella, puede distinguirse un periodo predominante de formación de marcas (alto porcentaje de otolitos con borde opaco y bajo con bordes hialinos) durante los meses de secas (de noviembre a febrero) y uno secundario durante los meses de lluvias, particularmente junio.

Cuadro 3.9. recuencia y porcentaje de tipos de borde de otolitos por fecha de muestreo (número de días transcurridos a partir de la primera fecha 10/01/1998).

Mes de muestreo (no. días)	Tipo de borde			
	Opaco		Hialino	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
ENE (0)	14	56.0	11	44.0
FEB (35)	22	75.9	7	24.1
ABR (98)	2	8.0	23	92.0
MAY (133)	7	28.0	18	72.0
JUN (153)	9	31.0	20	69.0
JUL (181)	6	24.0	19	76.0
AGO (215)	1	4.0	24	96.0
SEP (251)	6	23.1	20	76.9
OCT (280)	5	25.0	15	75.0
NOV (308)	12	57.1	9	42.9
DIC (335)	8	31.0	18	69.0
FEB/99 (398)	21	72.0	8	28.0
Total	113	37.0	192	63.0



Los comandos de Stata para calcular los porcentajes de borde son:

```
. gen edias=date(fecha,"DMY")
. di date("10/01/1998","DMY")
13889
. gen días=edias-13889
. bysort días: gen sumo=sum(o)
. bysort días: gen sumot = sumo[_N]
. bysort días: gen sumh=sum(h)
. bysort días: gen sumt=sumo[_N]
. gen porsot = sumot/(sumot+sumht)* 100 if anillo==40
. gen porsht = sumht/(sumot+sumht)*100 if anillo==40

. use miotome7

. scatter porsot porsht días, ms(Oh Th) c(1 1) xlab(0 "E" 35 "F" 67 "M" 98 "A"
133 "M" 153 "J" 181 "J" 215 "A" 251 "S" 280 "O" 308 "N" 335 "D" 368 "E" 398
"F") xline(351.5) legend(label(1 "Opacos") label(2 "Hialinos") pos(11) ring(0))
yscale(range(0 110)) text(105 330 "1998") text(105 370 "1999") ytitle("Porcentaje
de frecuencia") xtitle("Fecha de muestreo")
```

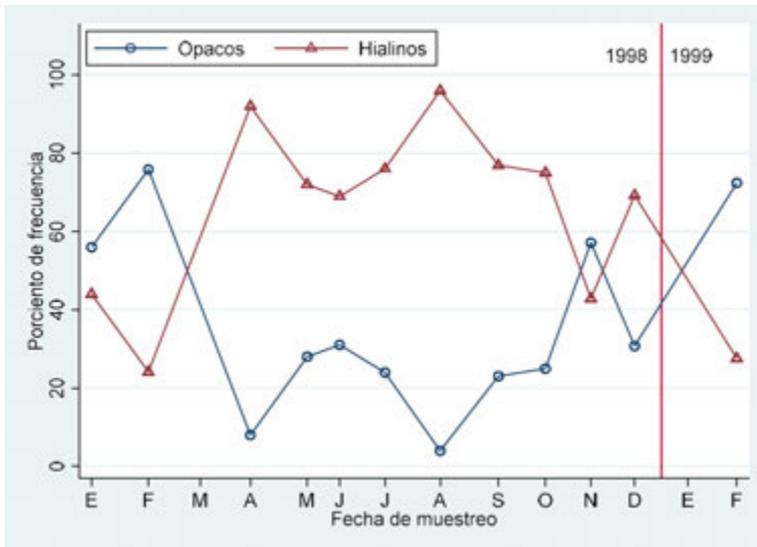


Figura 3.16. Porcentaje de otolitos con bordes hialinos y opacos por fecha de muestreo



3.7.2. Análisis del incremento marginal

Un segundo enfoque utilizado para establecer la periodicidad de la formación de las marcas encontradas en los otolitos consiste en un análisis detallado de las medidas marginales obtenidas en la medición de los radios de los otolitos (Figura 3.17). En este reporte se analizó la distribución de las medidas marginales por medio de estimadores de densidad por kernel para identificar a las modas predominantes y utilizarlas como representantes por fecha de muestreo. Estas modas por fecha de muestreo se presentan en la Figura 3.18. En ella puede apreciarse un mínimo absoluto en los meses de secas (de noviembre a febrero) y otro menos marcado durante las lluvias (julio).

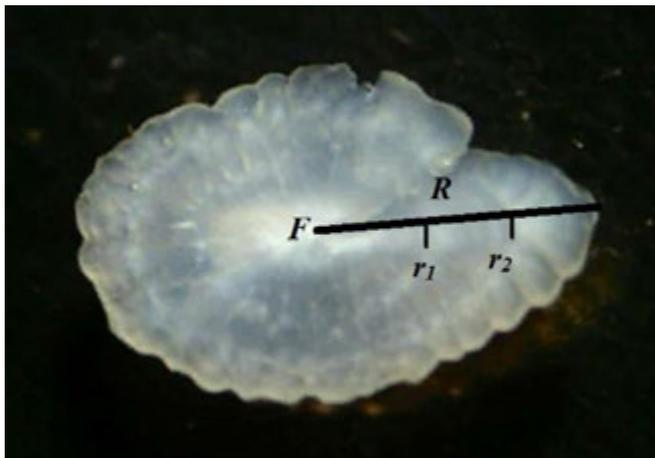


Figura 3.17. Medición de otolitos: F = foco del otolito, R = radio total (a la captura), r_1 y r_2 radios de los anillos respectivos. El incremento marginal para este otolito se calculó como $R - r_2$ y la razón de incremento marginal como $(R - r_2) / (r_2 - r_1)$.

Para obtener el incremento marginal, del arreglo de datos especificado arriba se utilizaron los siguientes comandos:

```
. gen inc = radio[_n]-radio[_n-1] if nradio==40
```



Y para obtener los valores modales se utilizaron los comandos:

```
. bandw1 inc if mue==1
. warpdenm inc if mue==1, b(.15) m(10) k(6) numo mo
```

en donde $b(.15)$ representa la amplitud de banda óptima de Silverman obtenida de “bandw1.ado”. Así para cada muestra y estos valores se almacenaron en el archivo “modainc.dta”.

```
. use modainc
. scatter mode dias, ms(Oh) c(1) xlab(0 "E" 35 "F" 67 "M" 98 "A" 133 "M" 153 "J"
181 "J" 215 "A" 251 "S" 280 "O" 308 "N" 335 "D" 368 "E" 398 "F") xline(351.5)
text(1.9 330 "1998") text(1.9 370 "1999") ytitle("Modas de incremento marginal
(mm)") xtitle("Fecha de muestreo")
```

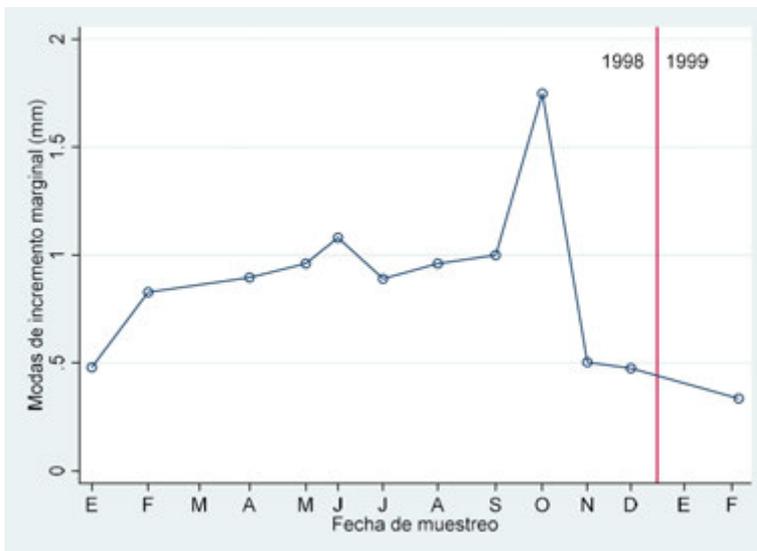


Figura 3.18. Valores modales de incremento marginal por fecha de muestreo



3.7.3. Razón del incremento marginal

Además del estudio de las medidas directas del incremento marginal para validar la periodicidad de las marcas de crecimiento, se utiliza la *Razón de Incremento Marginal*, definida como la división del incremento marginal entre la longitud de la banda de crecimiento anterior (o la distancia del foco al anillo en los individuos con un solo anillo). Como en el caso anterior, la distribución de estos valores en cada mes se analizó y se tomó la moda dominante como representante mensual. La figura 3.19 muestra de manera más clara las oscilaciones sugeridas en la Figura 3.18 y de nuevo puede apreciarse que valores bajos ocurren en los meses de secas (noviembre a febrero) y a la mitad de la temporada de lluvias (julio).

```
. use miotome7
. gen mir = inc/(radio[_n-1]-radio[_n-2])
```

Y como con los incrementos, se estimaron las modas dominantes de las razones por muestreo:

```
. bandw1 mir if mue==1
. warpdenm mir if mue==1, b(.19) m(10) k(6) numo mo
```

y como en los incrementos se guardaron en el mismo archivo “modainc.dta” para realizar la Figura

```
. scatter mir dias, ms(Oh) c(1) xlab(0 "E" 35 "F" 67 "M" 98 "A" 133 "M" 153 "J"
181 "J" 215 "A" 251 "S" 280 "O" 308 "N" 335 "D" 368 "E" 398 "F") xline(351.5)
text(1.15 330 "1998") text(1.15 370 "1999") ytitle("Modas de razón incremento
marginal") xtitle("Fecha de muestreo")
```

Con base en este análisis puede concluirse que las marcas de crecimiento se forman dos veces al año, durante la temporada seca fría (época desfavorable) y en la mitad de la época cálida del año (época reproductiva). Esta formación bianual ha sido reportada también en otros lugares del mundo (Admassu y Casselman, 2000).

Con la información talla-edad, entonces es posible modelar el crecimiento. Además, se han desarrollado varios métodos que utilizan la composición de edades para evaluar a los recursos pesqueros (Sparre y Venema, 1992).

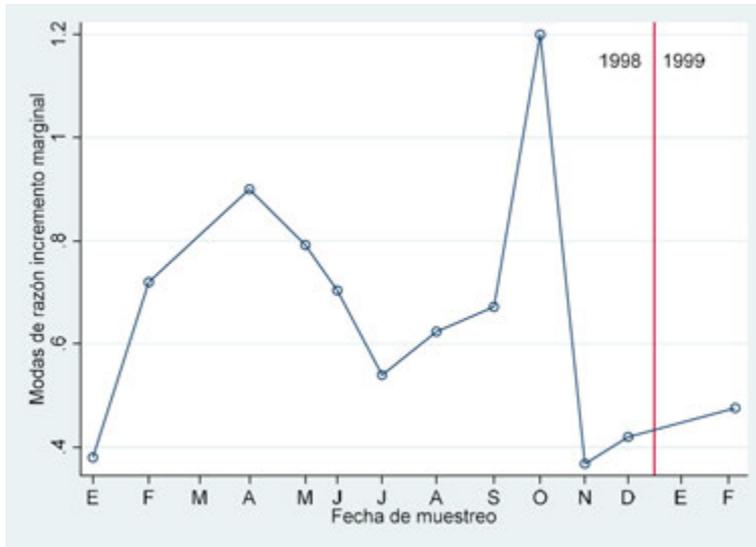


Figura 3.19. Valores modales de la razón de incremento marginal por fecha de muestreo.

Capítulo 4

Modelación del crecimiento individual



4.1. Introducción

En la actualidad, se reconoce que el propósito del manejo de pesquerías es asegurar una producción sostenible a lo largo del tiempo de las poblaciones explotadas (stocks) de peces (y otros organismos acuáticos) a través de acciones regulatorias y de mejoras sin olvidar la promoción de bienestar económico y social de los pescadores y las industrias asociadas. Para lograr este propósito, las autoridades responsables del manejo deben diseñar, justificar (políticamente) y administrar (forzar) una serie de restricciones sobre la actividad pesquera (Hilborn y Walters, 1992). Estas decisiones deben estar basadas en el conocimiento biológico de las especies explotadas para una adecuada evaluación de los stocks.

Como se revisó en el capítulo 3, el conocimiento de la edad de los peces es una característica biológica muy importante que proporciona información para la evaluación de una pesquería comercial. Una vez que se dispone de información de edad y tallas, es posible investigar el patrón de crecimiento de las especies. Como introducción hacia los modelos de crecimiento, iniciaremos con algunas consideraciones.

En una población cerrada a la migración, la biomasa de una población puede cambiar por el aumento (reproducción, reclutamiento), la disminución (mortalidad natural o por pesca) o el crecimiento de los individuos (Haddon, 2001; King, 2007; Ogle, 2016). Este crecimiento individual es el resultado de procesos bioenergéticos constructivos (anabolismo) y destructivos (catabolismo) que dependen de características relativamente estables de la historia de vida (peces longevos crecen más lentamente que especies de vida corta) y factores ambientales abióticos y bióticos relativamente variables como la disponibilidad de alimento, la cual refleja aspectos tales como la temperatura y la competencia (intra e interespecífica). Por lo tanto, el crecimiento es un proceso biótico que sintetiza una amplia variedad de factores bióticos y abióticos en un ecosistema acuático (Haddon, 2001; Ogle, 2016).



El crecimiento de los individuos puede ser definido como el incremento de tamaño, ya sea en alguna dimensión lineal (longitud, anchura) o peso (total, eviscerado) en función del tiempo (edad). En los animales marinos a menudo se registra la “longitud estándar” (a este respecto consultar a Howe, 2002) debido a la facilidad de medición precisa a bordo de las embarcaciones donde el movimiento por las olas dificulta el uso de balanzas (King, 2007).

En este capítulo presentaremos varios procedimientos con sus respectivos algoritmos computarizados diseñados para estimar los parámetros de algunas funciones de crecimiento utilizadas en Biología Pesquera: tradicionales: von Bertalanffy, Logística, Gompertz, Richards; y estacionales: Somers, Pitcher y MacDonald.

Para la función de crecimiento de von Bertalanffy (FCVB) se presentarán tres derivaciones lineales de la ecuación básica del crecimiento así como el enfoque más moderno de la regresión no lineal, la cual será aplicada al resto de los modelos revisados.

Los programas son aplicados a datos provenientes de la lectura de partes duras y el análisis de la frecuencia de tallas llevado a cabo por la aplicación de estimadores de densidad por kernel. Las distribuciones multimodales son analizadas por medio de la versión actualizada de la determinación de componentes gaussianos (Bhattacharya, 1967, Salgado-Ugarte, *et al.* 1994).

Este capítulo pretende introducir una forma alternativa de analizar datos de frecuencia de tallas (por métodos no paramétricos de cómputo intensivo) en conjunto con versiones automatizadas de los procedimientos tradicionales para obtener expresiones del crecimiento a partir de los tamaños estimados a cada edad. Los mismos algoritmos computarizados pueden aplicarse a tamaños promedio a las edades estimadas por interpretación de marcas en las estructuras duras de peces (u otros organismos acuáticos explotados).

4.2 Estimación de la edad

Podemos mencionar dos grandes tipos de métodos para estimar la edad de los peces y otros animales acuáticos: - análisis de frecuencia de tallas y - lectura de partes duras (escamas, otolitos entre otras).

Análisis de frecuencia de tallas.- Cuando es posible medir el tamaño de un gran número de individuos muestreados de un stock de peces u otros animales acuáticos (por ejemplo, crustáceos o moluscos) es posible analizar en detalle la distribución estadística de sus



tamaños. Tradicionalmente, para esta tarea se han venido utilizando el histograma o el **polígono de frecuencia**. No obstante, en años recientes se ha propuesto el uso de histogramas mejorados denominados estimadores de densidad por kernel por ser uno de los métodos estadísticos más apropiados para el análisis de las distribuciones de tallas animales (Salgado-Ugarte *et al.* 1993, 1995a, 1995b, 1997). Si los eventos reproductivos de la especie son de naturaleza discreta y periódica, será posible seguir a los grupos o clases dominantes de talla como modas en la distribución de tamaños. Si la reproducción ocurre a intervalos regulares de tiempo, puede ser posible aproximar las edades de cada uno de los grupos de tamaño (King, 1995; 2007).

El número de grupos de organismos con edad semejante puede estimarse con mucha precisión gracias a la disponibilidad de poderosos procedimientos asociados a los EDKs tales como aquellos que permiten la elección de la mejor amplitud de banda (intervalo) y que se derivan de reglas prácticas (óptimas y sobreesuavizadas), así como de métodos estadísticos de cómputo intensivo como la validación cruzada (por mínimos cuadrados y sesgada) y la prueba de bootstrap (esquema de muestreo con repetición) para multimodalidad.

Una vez que se ha obtenido una distribución multimodal es posible concebir a cada grupo como una distribución gaussiana. Los componentes individuales en esta distribución mezclada pueden determinarse utilizando algún método para caracterización de distribuciones normales, como por ejemplo el procedimiento de Bhattacharya. Estas modas representan grupos de peces con edad semejante (cohortes). Las distribuciones gaussianas quedan determinadas por dos parámetros, media aritmética y desviación estándar. Con las medias estimadas para cada cohorte y suponiendo ó conociendo el período de tiempo que las separa, es posible entonces estimar los parámetros de crecimiento de la función de von Bertalanffy, la más utilizada en evaluaciones de recursos pesqueros.

Interpretación de partes duras.- Para estimar la edad de los peces que habitan regiones templadas del planeta, es posible contar marcas (anillos) anuales formados en las partes duras (escamas, otolitos entre otros). Los anillos se forman debido a fluctuaciones estacionales del ambiente. En los lugares de zonas intertropicales, donde los cambios estacionales son pequeños, frecuentemente es muy difícil encontrar marcas periódicas en las partes duras (Sparre y Venema, 1992). No obstante, en estas áreas con ambientes uniformes, las partes duras muestran incrementos diarios. En otras instancias, se forman marcas periódicas debido a eventos biológicos tales como la reproducción. Por tanto, el mejor compromiso para el estudio de los stocks de especies tropicales es un análisis de gran número de datos de tamaño (para estudiar su frecuencia) combinado con un tamaño modesto de lecturas de marcas periódicas (anillos diarios o por reproducción) en estructuras duras (Sparre y Venema, 1992; King, 1995; 2007).



4.3 Estimación de los parámetros de la función de crecimiento de von Bertalanffy (FCvB)

Para describir cuantitativamente el crecimiento de un pez, se han desarrollado varias expresiones matemáticas. De acuerdo con Haddon (2001) la función de crecimiento desarrollada por von Bertalanffy (1938) fue introducida a la Biología Pesquera por Beverton y Holt (1957). La función de crecimiento de von Bertalanffy (FCvB) ha sido muy útil pues sigue cercanamente el crecimiento observado en un gran número de especies. Desde entonces, han aparecido en la literatura varios modelos de crecimiento, como por ejemplo los propuestos por Beverton y Holt (1957), Ursin (1968), Ricker (1975) Gulland (1983), Pauly (1984), Pauly y Morgan (1987). Además, la FCvB se ha vuelto una de las partes fundamentales en la literatura biológica porque se integra a modelos complejos que describen la dinámica de las poblaciones de peces (Sparre y Venema, 1992).

La expresión matemática de von Bertalanffy para el crecimiento, expresa la longitud (L) en función de la edad (t) del pez:

$$L_t = L_{\infty}(1 - \exp[-K(t - t_0)]) \quad (4.1)$$

El lado derecho de la ecuación contiene a la edad, t , y tres parámetros: L_{∞} , K y t_0 . Especies diferentes tendrán un conjunto particular de parámetros, los cuales pueden interpretarse biológicamente: L_{∞} , es “la longitud media de peces muy viejos (estrictamente, infinitamente viejos)”. También es llamada la “longitud asintótica”. K es un “parámetro de curvatura” el cual determina que tan rápido el pez alcanza su L_{∞} . Un valor alto para K indica a una especie de vida corta que alcanza su longitud asintótica en uno o dos años. Un valor bajo de K resulta en una curva de crecimiento aplanada para las especies que necesitan muchos años para acercarse a su L_{∞} .

El tercer parámetro, t_0 , a menudo se le llama “parámetro de condición inicial” y determina el punto hipotético en el tiempo cuando el pez tiene una longitud de cero. Por supuesto que esto no tiene sentido biológico ya que al salir del huevo (y que para nuestros fines es cuando comienza el crecimiento), la larva ya tiene una cierta longitud. Esta longitud a la eclosión puede llamarse L_0 cuando ponemos $t = 0$ como el día de nacimiento.

$$L_0 = L_{\infty}(1 - \exp[-Kt_0]) \quad (4.2)$$

Sin embargo, L_0 puede no ser una estimación realista de la longitud a la eclosión porque la larva del pez puede seguir una forma de crecimiento diferente si consideramos que el desarrollo del embrión en el huevo puede ser afectado por factores distintos. No obstante,



los peces mayores (que son sujetos a explotación pesquera) por lo general siguen bien la FCVB (Sparre y Venema, 1992).

4.3.1. Gráfico de Ford-Walford

Para estimar la FCVB podemos utilizar el “método de Petersen” que permite estimar la curva de crecimiento a partir de la posición relativa de las modas en la distribución de longitudes de una muestra. La suposición crucial es que las modas están separadas por intervalos uniformes de tiempo (un año en forma típica).

Otro método sencillo para estimar los parámetros de la ecuación de von Bertalanffy con tamaños separados por intervalos uniformes de tiempo es el llamado gráfico de Ford (1933) – Walford (1946). La derivación de este gráfico se basa en una curva de crecimiento con t_0 igual a cero:

$$L_t = L_\infty(1 - e^{-Kt}) \quad (4.3)$$

$$L_t = L_\infty - L_\infty e^{-Kt}$$

$$L_\infty - L_t = L_\infty e^{-Kt} \quad (4.4)$$

Restando L_t a L_{t+1} en la Ecuación 4.3) tenemos:

$$L_{t+1} - L_t = L_\infty(1 - e^{-K(t+1)}) - L_\infty(1 - e^{-Kt})$$

desarrollando y trasponiendo términos:

$$\begin{aligned} &= L_\infty e^{-Kt} - L_\infty e^{-K(t+1)} \\ &= L_\infty e^{-K} - L_\infty e^{-Kt} e^{-K} \\ &= L_\infty e^{-Kt} (1 - e^{-K}) \end{aligned} \quad (4.5)$$

Sustituyendo 4.4) en 4.5) resulta en:

$$\begin{aligned} L_{t+1} - L_t &= (L_\infty - L_t)(1 - e^{-K}) \\ &= L_\infty(1 - e^{-K}) - L_t + L_t e^{-K} \\ L_{t+1} &= L_\infty(1 - e^{-K}) + e^{-K} L_t \end{aligned} \quad (4.6)$$



Esta es una ecuación lineal y sugiere que la longitud a la edad t (L_t) puede graficarse contra la longitud a un año después (L_{t+1}). La línea recta que se ajusta a estos datos tendrá una pendiente de $b = e^{-K}$ y un intercepto de $a = L_{\infty}(1 - e^{-K})$. De esta forma es posible estimar K y L_{∞} como (King, 1995; 2007):

$$K = -\ln(b)$$

$$L_{\infty} = a/(1 - b)$$

4.3.2. Gráfico de Gulland

Este método es una variación del gráfico de Ford-Walford y fue propuesto por Chapman (1961) y después por Gulland (1969). Se basa en el uso de un intervalo de tiempo constante Δt . Puede demostrarse que la FCVB implica que:

$$L_{t+\Delta t} - Lt = L_{\infty}(1 - e^{-K\Delta t}) - L_t(1 - e^{-K\Delta t}) \quad (4.7)$$

Por tanto, puesto que K y L_{∞} son constantes, $(1 - e^{-K\Delta t})$ permanecerá constante y la ecuación se vuelve una función lineal $y = a + bx$, donde

$$a = L_{\infty}(1 - e^{-K\Delta t}) \text{ y } b = -(1 - e^{-K\Delta t})$$

Los parámetros de crecimiento son entonces derivados de (Sparre y Venema, 1992):

$$K = -(1/\Delta t) \ln(1 + b)$$

y

$$\begin{aligned} L_{\infty} &= -a/b \\ &= a/(1 - e^{-K\Delta t}) \end{aligned}$$



4.3.3. Gráfico de Gulland-Holt

Este procedimiento, propuesto por Gulland y Holt (1959) toma en cuenta que los peces incrementan su longitud al envejecer, pero su “tasa de crecimiento” (incremento en longitud por unidad de tiempo), disminuye con la edad, tendiendo a hacerse cero en peces muy viejos. Considerando la definición de la tasa de crecimiento:

$$\frac{\Delta L}{\Delta t} = \frac{L_{t+\Delta t} - L_t}{\Delta t} \quad (4.8)$$

y considerando que la relación matemática entre la longitud de un pez y la tasa de crecimiento a cualquier tiempo es una función lineal:

$$\frac{\Delta L}{\Delta t} = a + bL_t$$

utilizando la FCVB llegamos a:

$$\frac{\Delta L}{\Delta t} = K(L_\infty - L_t) \quad (4.9)$$

La ecuación (4.9) puede re-escribirse como:

$$\frac{\Delta L}{\Delta t} = KL_\infty - K\bar{L}_t \quad (4.10)$$

donde $\bar{L}_t = \frac{L_{t+\Delta t} + L_t}{2}$

Cuando Δt es pequeña, \bar{L}_t puede ser una aproximación razonable del tamaño medio. Una ventaja sobre otros métodos es que Δt no necesita ser constante. Usando a \bar{L}_t como la variable independiente y $\Delta L/\Delta t$ como la variable de respuesta, la ecuación 4.10) se vuelve una relación lineal que permite la estimación de los parámetros por medio de una regresión de acuerdo a la siguiente expresión:

$$\Delta L/\Delta t = a + b\bar{L}_t \quad (4.11)$$

donde $K = -b$ y $L_\infty = -a/b$ (Sparre y Venema, 1992; Gómez-Márquez, 1994).



4.3.4. Método de Beverton-Holt

Este método (Beverton y Holt, 1957) es usado para estimar K y t_0 para una L_∞ dada, la cual puede provenir de cualquiera de los métodos presentados anteriormente. De la FCVB podemos obtener:

$$L_\infty - L_t = L_\infty (\exp[-K(t - t_0)])$$

$$\ln(L_\infty - L_t) = \ln L_\infty - Kt + Kt_0$$

Arreglando algebraicamente finalmente tenemos:

$$\ln(L_\infty - L_t) = \ln L_\infty + Kt_0 - Kt \quad (4.12)$$

De nuevo, esta es una función lineal ($y = a - bx$) de t vs. $\ln(L_\infty - L_t)$ que permite la recuperación de los parámetros mediante (Gómez-Márquez, 1994):

$$K = -b \text{ y } a = \ln L_\infty + Kt_0$$

Despejando t_0 , tenemos:

$$t_0 = \frac{a - \ln L_\infty}{K}$$

4.3.5. Regresión No Lineal

De acuerdo con Sparre y Venema (1992), la regresión no lineal por mínimos cuadrados es un método superior a cualquiera de los procedimientos arriba descritos para estimar los parámetros de la FCVB (desde un punto de vista teórico y estadístico). El trabajo de cálculo es considerable y en la práctica es necesario utilizar una computadora (con los programas debidos) para poder llevar a cabo las operaciones requeridas. El método estima los parámetros de crecimiento en una forma tal que la suma de los cuadrados de las desviaciones de las observaciones respecto del modelo son minimizadas. Esto es, busca la combinación de valores para L_∞ , K y t_0 que resultan en la suma mínima de desviaciones cuadradas.



4.4 Caso 1. Análisis de frecuencia de tallas

Para ejemplificar el análisis de frecuencia de longitudes utilizaremos los datos de longitud del bagre *Cathorops melanopus* ($n = 641$) recolectados en noviembre de 1980 en el sistema lagunar de Tampamachoco, Ver. Reportados por Salgado-Ugarte (1985). Un breve resumen estadístico se presenta en el Cuadro 4.1:

Cuadro 4.1. Resumen estadístico de longitud patrón de *Cathorops melanopus* de Tampamachoco.

Sexo	Media	Desv. estándar	Frecuencia
Machos	164.9633	34.0392	109
Hembras	184.3855	36.3324	83
Indeterminado	75.7795	9.6996	449
Total	105.0078	49.5904	641

Cuadro 4.2. Amplitudes de banda críticas y niveles estimados de significancia para longitud patrón de *Cathorops melanopus* $n = 641$.

Número de modas	Amplitudes de banda críticas	Valores de P
1	23.36	0.0000
2	19.43	0.0000
3	9.64	0.1750
4	3.78	0.7417
5	3.23	0.7750
6	3.02	0.7250

Nota: Los valores de P se obtuvieron de $B = 120$ repeticiones bootstrap de tamaño 641.

Estos datos de longitud se sometieron a la prueba para multimodalidad de Silverman utilizando los programas incluidos en Salgado-Ugarte, *et al.* (1997) y Salgado-Ugarte (2002) para obtener los resultados incluidos en el Cuadro 4.2.



A continuación se presentan los comandos para llevar a cabo la prueba:

```
. use catflen
. critiband bodlen, bwhigh(25) bwlow(2) stsize(0.01) mval(40)

Estimation number = 63      Bandwidth = 23.38      Number of modes = 1
Estimation number = 64      Bandwidth = 23.37      Number of modes = 1
Estimation number = 65      Bandwidth = 23.36      Number of modes = 1
Estimation number = 66      Bandwidth = 23.35      Number of modes = 2
Estimation number = 67      Bandwidth = 23.34      Number of modes = 2
.
Estimation number = 556     Bandwidth = 19.45      Number of modes = 2
Estimation number = 557     Bandwidth = 19.44      Number of modes = 2
Estimation number = 558     Bandwidth = 19.43      Number of modes = 2
Estimation number = 559     Bandwidth = 19.42      Number of modes = 3
Estimation number = 560     Bandwidth = 19.41      Number of modes = 3
.
Estimation number = 1536    Bandwidth = 9.65       Number of modes = 3
Estimation number = 1537    Bandwidth = 9.64       Number of modes = 3
Estimation number = 1538    Bandwidth = 9.63       Number of modes = 4
Estimation number = 1539    Bandwidth = 9.62       Number of modes = 4
Estimation number = 1540    Bandwidth = 9.61       Number of modes = 4
.
Estimation number = 2122    Bandwidth = 3.79       Number of modes = 4
Estimation number = 2123    Bandwidth = 3.78       Number of modes = 4
Estimation number = 2124    Bandwidth = 3.77       Number of modes = 5
Estimation number = 2125    Bandwidth = 3.76       Number of modes = 5
Estimation number = 2126    Bandwidth = 3.75       Number of modes = 5
.
Estimation number = 2176    Bandwidth = 3.25       Number of modes = 5
Estimation number = 2177    Bandwidth = 3.24       Number of modes = 5
Estimation number = 2178    Bandwidth = 3.23       Number of modes = 5
Estimation number = 2179    Bandwidth = 3.22       Number of modes = 6
Estimation number = 2180    Bandwidth = 3.21       Number of modes = 6
.
Estimation number = 2198    Bandwidth = 3.03       Number of modes = 6
Estimation number = 2199    Bandwidth = 3.02       Number of modes = 6
Estimation number = 2200    Bandwidth = 3.01       Number of modes = 7
Estimation number = 2201    Bandwidth = 3          Number of modes = 7
Estimation number = 2202    Bandwidth = 2.99       Number of modes = 7
```

Un vez determinadas las bandas críticas se procede a generar las muestras bootstrap. Esto se hace para cada banda crítica. Sólo se presenta el caso para las muestras con la banda crítica de 4 modas. Como se hace uso del generador de números pseudoaleatorios de Stata y con fin de llevar registro preciso del procedimiento se establece un valor inicial de semilla.



```
. keep bodlen
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bodlen	641	105.0078	49.5904	48	243

```
. boot bootsamb, arg(bodlen 3.78 49.5904) i(120)
```

La combinación del comando boot de Stata con nuestro programa bootsamb, el cual requiere como argumentos la variable a muestrear, el valor de banda crítica y la desviación estándar de los datos originales permite además establecer el número de repeticiones (iteraciones) en este caso de 120. Con las muestras bootstrap se procede entonces a ejecutar la prueba de Silverman. Como se mencionó anteriormente sólo se presenta un extracto para la prueba de 4 modas:

```
. keep ysm_rep
. silvtest ysm_rep, cr(3.78) m(40) nurf(120) cnm(4)
bs sample 1          Number of modes = 5
bs sample 2          Number of modes = 4
.
bs sample 119       Number of modes = 5
bs sample 120       Number of modes = 4
```

```
Critical number of modes = 4
```

```
Pvalue = 89 / 120 = 0.7417
```

Este conjunto de datos claramente manifiesta la existencia de cuatro modas. Debido a que histogramas suavizados (EDKs) con cuatro modas se obtienen utilizando amplitudes de banda de 9.63 a 3.88, utilizamos una amplitud intermedia para la presentación final $(9.63 + 3.78)/2 \approx 6.7$. El resultado se presenta en la Figura 4.2 junto con los componentes gaussianos obtenidos a continuación.

Los EDKs utilizan la escala de densidad de manera pre-establecida para transformar la frecuencia a número de individuos es necesario aplicar una regla de proporcionalidad directa: “la suma de las densidades es igual a n, como una densidad dada es a...”

```
. warpdnml bodlen, b(6.7) m(10) k(6) gen(den6p7 mid6p7)
. gen sumden=sum(den6p7)
. sum
. gen frec = den*641/sumden[_N]
(543 missing values generated)
```



Posteriormente, aplicamos el método de Bhattacharya para descomposición gaussiana (Bhattacharya, 1967) por medio de versiones actualizadas (Salgado-Ugarte, *et al.* 2000) de los programas incluidos en Salgado-Ugarte *et al.* (1994). El primero de estos programas calcula las diferencias logarítmicas en función de los centros de clase del histograma utilizado y dibuja la gráfica con el número de par ordenado como símbolo de graficación con el fin de definir a partir de puntos consecutivos, tendencias lineales con pendiente negativa. Cada una de estas líneas representa un componente gaussiano individual en la distribución mezclada. En la Figura 4.1 se presenta esta gráfica y se aprecian claramente varios puntos con tendencias lineales con pendiente negativa.

```
. bhatapl11 frec mid, yline(0)
```

El segundo programa permite la caracterización del componente gaussiano ajustado a las frecuencias observadas reportando sus parámetros, media y desviación estándar, así como el número de individuos que comprende. El resultado se presenta en forma de cuadro y se grafica la curva gaussiana resultante sobre las frecuencias observadas.

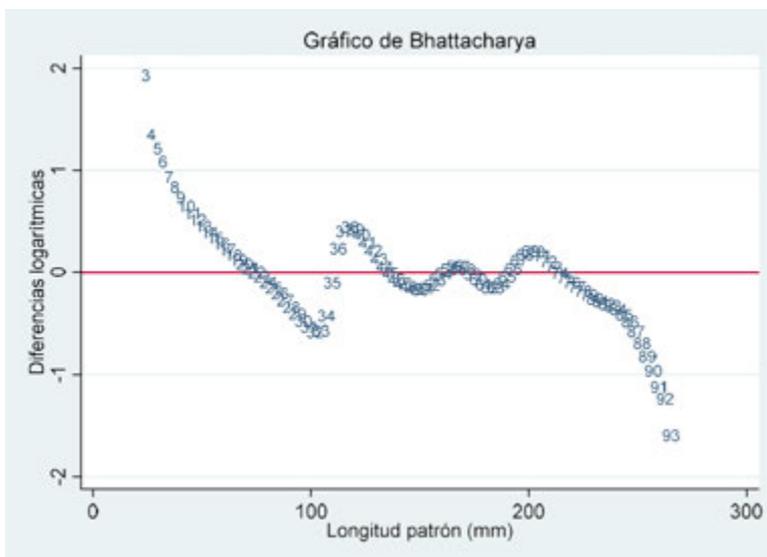


Figura 4.1. Gráfico de Bhattacharya para los datos de longitud patrón del bagre *Cathorops melanopus* de Tampamachoco, Ver.



Por ejemplo, para el primer componente, eligiendo el intervalo de observaciones 13 a 22, se obtiene un componente gaussiano con media de 75.6083, d.e. de 11.9029 con 441 individuos. El ajuste lineal tiene una r^2 ajustada de 0.9939. Esta distribución de Gauss se presenta como la primera de izquierda a derecha en la Figura 4.2.

```
. bhatgauc1 freq mid in 13/22
R-square = 0.9945          Adj R-square = 0.9939
Mean = 75.6083
s.d. = 11.9029
component size = 441
```

Se continúa con los demás componentes:

```
. bhatgauc1 freq mid in 40/45, gen(gauco2)
R-square = 0.9993          Adj R-square = 0.9991
Mean = 137.5868
s.d. = 9.7123
component size = 75
```

```
. bhatgauc1 freq mid in 58/61, gen(gauco3)
R-square = 0.9990          Adj R-square = 0.9985
Mean = 174.8263
s.d. = 11.9622
component size = 42
```

```
. bhatgauc1 freq mid in 72/76, gen(gauco4)
R-square = 0.9960          Adj R-square = 0.9947
Mean = 215.9293
s.d. = 11.8713
component size = 66
```

El cuadro 4.3 resume los parámetros de los cuatro componentes determinados los cuales se presentan también en la Figura 4.2, la cual se construye con el siguiente comando:

```
. scatter freq mid, ms(oh) || mspline gauco1 mid, bands(40) || mspline
gauco2 mid, bands(40) || mspline gauco3 mid, bands(40) || mspline gauco4
mid, bands(40) ytitle("Frecuencia") xtitle("Puntos medios de Longitud patrón
(mm)") legend(off)
```



Cuadro 4.3. Componentes gaussianos estimados y sus parámetros.

Componente	Intervalo de datos utilizados	Media	Desv. estándar	No. de individuos
1	13 a 22	75.6083	11.9029	441
2	40 a 45	137.5868	9.7123	75
3	58 a 61	174.8262	11.9622	42
4	72 a 76	215.9292	11.8713	66

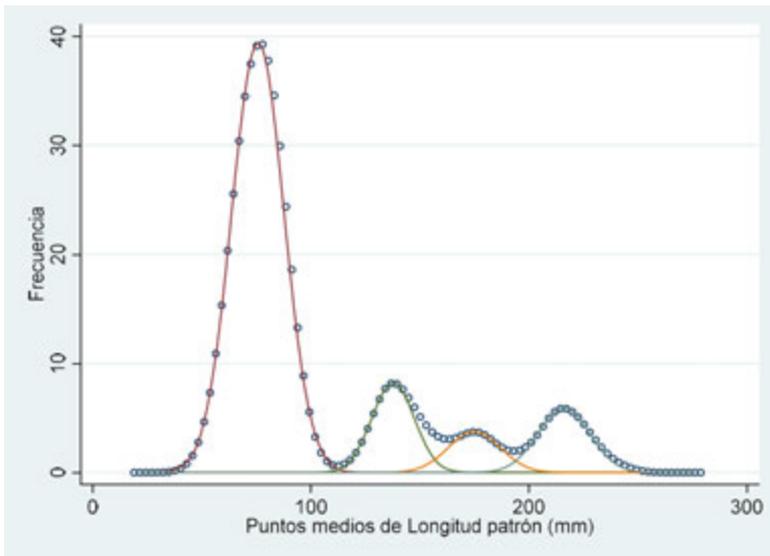


Figura 4.2. Frecuencia suavizada observada (de EDK) y componentes gaussianos estimados para los datos de longitud de bagres.

Los valores medios son entonces usados para estimar los parámetros de la FCVB. A continuación se irán presentando los procedimientos con una breve mención de los programas elaborados e incluidos en Salgado-Ugarte, *et al.* (2000) y actualizados por los autores para la presente obra.



Para estimar los parámetros de la ecuación de von Bertalanffy por los métodos linealizados, en primer lugar se incluye el gráfico de Ford-Walford, realizado con el programa “fordwaln.ado” y que estima los parámetros L_{∞} and K de la función de crecimiento de von Bertalanffy por medio de los datos de longitud media y de acuerdo al procedimiento propuesto por Ford (1933) y Walford (1946). Este programa dibuja la gráfica de l_t vs. l_{t+1} (Figura 4.3) y proporciona un cuadro con los resultados numéricos (Cuadro 4.4).

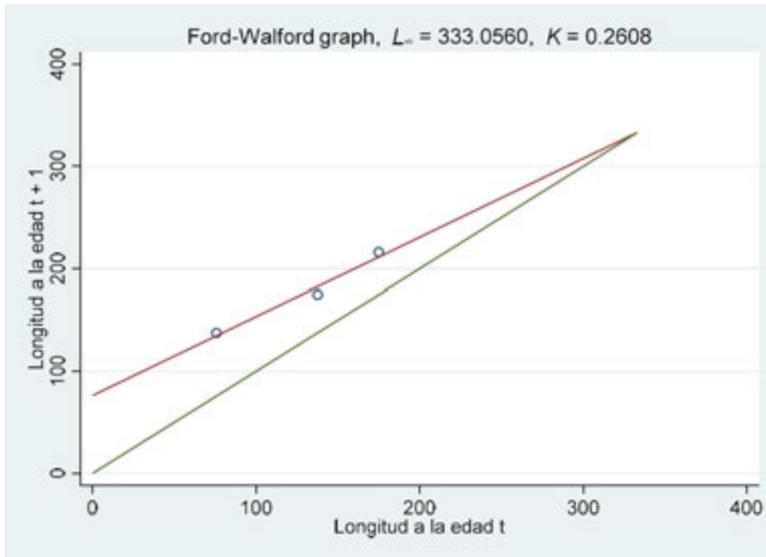


Figura 4.3. Gráfico de Ford-Walford para valores medios estimados por el método de Bhattacharya (datos de longitud patrón de bagres).

```
. use vbcatage, clear
. fordwaln mbl, ytitle("Longitud a la edad t + 1") xtitle("Longitud a la edad t")
```

Estimation of L_{∞} and K values by the Ford-Walford Method

Intercept = 76.4705	Slope = 0.7704
R-square = 0.9709	Adj R-square = 0.9418
L_{∞} = 333.0560	K = 0.2608



Cuadro 4.4. Estimación de L_{∞} y K por el método de Ford-Walford.

Intercepto =	76.4705	Pendiente =	0.7704
r^2 =	0.9709	r^2 Ajustada =	0.9418
L_{∞} =	333.0560	K =	0.2608

El mismo valor estimado para L_{∞} puede obtenerse con el programa para el método de Gulland (gullpltn.ado). Como en el programa para Ford-Walford el algoritmo para el método de Gulland estima los parámetros L_{∞} y K de la FCVB utilizando longitudes medias de acuerdo al procedimiento propuesto por Chapman (1961) y posteriormente por Gulland (1969). El programa dibuja la gráfica de l_t vs. $l_{t+1} - l_t$ (delta de longitud, Figura 4.4) y proporciona un cuadro con los resultados numéricos (Cuadro 4.5).

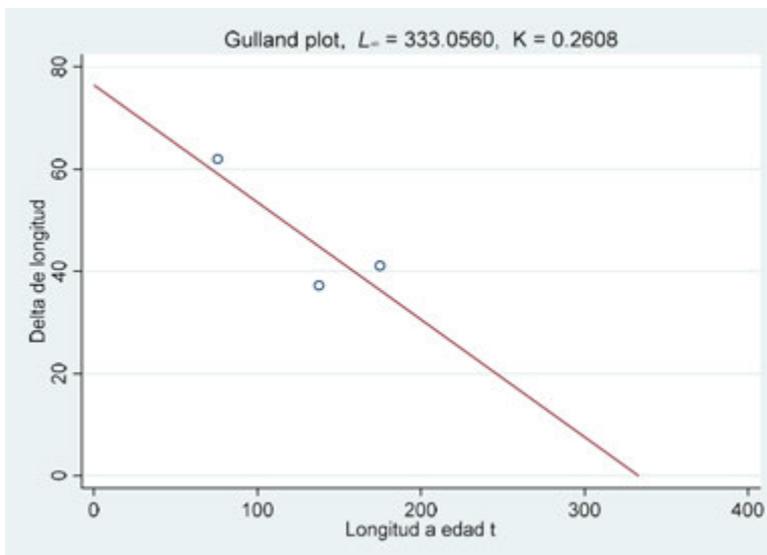


Figura 4.4. Gráfico de Gulland para valores medios de estimados por el método de Bhattacharya's para los datos de longitud de bagres.



```
. gullpltn mbl, ytitle("Delta de longitud") xtitle("Longitud a edad t")
```

Estimation of L_{inf} and K values by the Gulland Method

```
-----
Intercept = 76.4705          Slope          = -0.2296
R-square   = 0.7477          Adj R-square   = 0.4953
L_inf.     = 333.0560        K              = 0.2608
-----
```

Cuadro 4.5. Estimación de L_{∞} y K por el método Gulland.

Intercepto =	76.4705	Pendiente =	-0.2296
r^2 =	0.7477	r^2 Ajustada =	0.4953
L_{∞} =	333.0560	K =	0.2608

Cada moda puede representar un grupo de peces con edad semejante (cohorte). Suponiendo cohortes anuales y una edad inicial de 1, podemos generar una variable de edad. Con esta variable de “edad” es posible utilizar el método de Gulland-Holt para la estimación de los parámetros de la FCVB. El respectivo programa (gullholt.ado) estima los parámetros L_{∞} y K de la función de crecimiento de von Bertalanffy utilizando longitudes medias a cada edad de acuerdo al procedimiento propuesto por Gulland y Holt (1959). Este programa dibuja la relación lineal utilizada para la estimación (Figura 4.5) y proporciona un cuadro con los valores estimados de L_{∞} y K . Aplicado a los datos de bagre tenemos los resultados siguientes que también se muestran en el Cuadro 4.6:

```
. gulholtn mbl age, ytitle("Delta de longitud/Delta de edad") xtitle("Longitud
media por delta de edad") t1title("Gráfico de Gulland-Holt, {it:L{sub:{&infin}}})
= 342.3808, {it:K} = 0.2466")
```

Estimation of L_{inf} and K values by the Gulland-Holt Method

```
-----
Intercept = 84.4403          Slope          = -0.2466
R-square   = 0.6798          Adj R-square   = 0.3596
L_inf.     = 342.3808        K              = 0.2466
-----
```

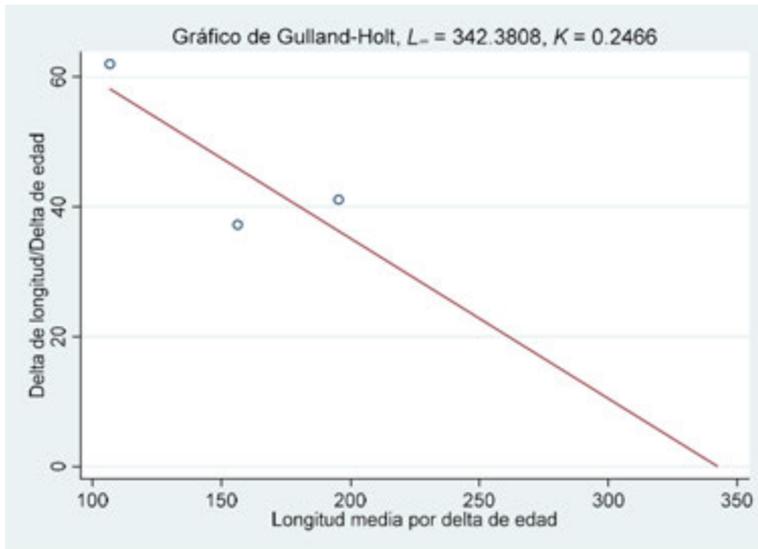


Figura 4.5. Gráfico de Gulland-Holt para estimación de los parámetros de la FCVB, datos de longitud de bagres.

Cuadro 4.6. Estimación de L_{∞} y K por el método de Gulland-Holt.

Intercepto =	84.4403	Pendiente =	-0.2466
r^2 =	0.6798	r^2 Ajustada =	0.3596
L_{∞} =	342.3808	K =	0.2466

Con la estimación de L_{∞} (333.06) obtenida por Ford-Walford o el gráfico de Gulland y suponiendo una edad anual podemos estimar t_0 con el método de Beverton y Holt (programa bevhltn.ado).

```
. bevhltn mbl age, linf(342.3808) ytitle("\{it:Ln\}(\{it:L\}_{sub:\{&infin\}} - \{it:l\}_{sub:t})")
xtitle("Edad (años)") ttitle("\Gráfico de Beverton-Holt, \{it:K\} = 0.2440, \{it:t\}_{sub:0} = - 0.0346")
```



Estimation of K and t_0 values by the Beverton-Holt Method

```
-----
Intercept = 5.8069          Slope          = -0.2574
R-square   = 0.9960          Adj R-square  = 0.9940
K          = 0.2574          t_0          = -0.0055
-----
```

Estimated vonBertalanffy Growth Function

```
-----
l_t = 333.0600 * (1 - exp( -0.2574 *(t + 0.0055 )))
-----
```

Este programa estima K y t_0 para una L_∞ dada los parámetros de la función de crecimiento de von Bertalanffy por medio de los datos del tamaño medio (longitud) a la edad de acuerdo al procedimiento propuesto por Beverton and Holt (1957). Este programa dibuja la gráfica del $\ln(L_\infty - l_t)$ - edad (Figura 4.6) y presenta un cuadro con los parámetros K y t_0 estimados además de la función de crecimiento de von Bertalanffy resultante y sus valores calculados con la opción “generate”. La Figura 4.7 muestra la curva ajustada sobre los puntos observados y la longitud asintótica.

Cuadro 4.7. Estimación de K y t_0 por el método de Beverton-Holt.

Intercepto =	5.8069	Pendiente =	- 0.2574
r² =	0.9960	r² Ajustada =	0.9940
K =	0.2574	t₀ =	-0.0055

Función de crecimiento de von Bertalanffy estimada:

$$l_t = 333.0600 (1 - \exp(-0.2574 *(t + 0.0055)))$$

```
. scatter mbl age, ms(Oh) || mspline mble age, bands(30) ytitle("Longitud
patrón (mm)") xtitle("Edad (años)") legend(label(1 "Observados") label(2
"Calculados")) yscale(range (50 350)) ylab(50(50)350) yline(333.06)
```

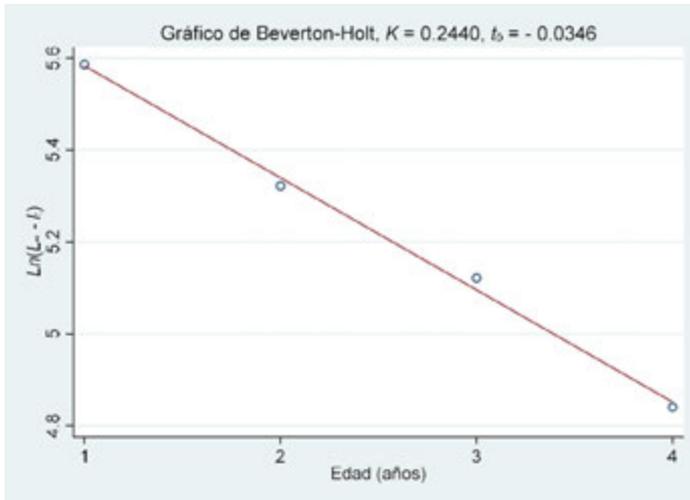


Figura 4.6. Gráfico de Beverton-Holt para estimar los parámetros K y t_0 de la FCVB para los datos de longitud de bagres.

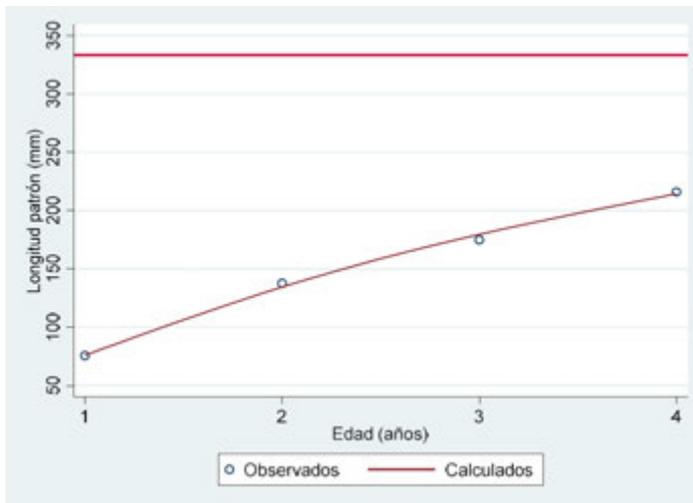


Figura 4.7. Medias observadas a cada edad y curva estimada de la FCVB para los datos de longitud patrón de bagres.



Podemos utilizar una regresión no lineal para estimar la función de crecimiento de von Bertalanffy con el programa `nlvbgf.ado`, el cual utiliza la rutina general de ajuste de modelos no lineales del paquete Stata. El programa particular establece la función objetivo y los parámetros y la rutina general hace la estimación por mínimos cuadrados y un procedimiento iterativo de Gauss-Newton (Royston, 1992; StataCorp, 1999; 2001; 2013; 2019). Los valores iniciales especificados son 500, 1, y 0.01 for L_{∞} , K and t_0 , respectivamente. La aplicación de este programa a los datos de longitud media a edad para las longitudes de bagres se incluyen en el Cuadro 4.8.

```
. nl vbgf mbl age
```

Como podemos observar, los parámetros estimados son casi idénticos pero el enfoque no lineal proporciona información estadística adicional como intervalos de confianza así como valores de t y significancia de los parámetros.

Cuadro 4.8. Regresión no lineal para datos de longitud de bagres.

Fuente	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	Número de obs =	4
				$F(3, 1) =$	998.30
Modelo	101802.37	3	33934.1242	$\text{Prob} > F =$	0.0233
Error	33.9919	1	33.9919	$r^2 =$	0.9997
				r^2 Ajustada =	0.9987
Total	101836.36	4	25459.0911	Error medio de estimación =	5.8302
				Desviación del error =	19.9108
Función de crecimiento de von Bertalanffy $L = L_{\infty}(1 - \exp(-K(\text{edad} - t_0)))$					
L	Coficiente	Error estándar	t	$P > t $	[Intervalo de Confianza del 95%]
L_{∞}	342.2142	100.766	3.40	0.182	-938.2685 1622.697
K	0.2435	0.1303	1.87	0.313	-1.4123 1.8993
t_0	-0.0407	0.2729	-0.15	0.906	-3.5080 3.4266

(Errores estándar, valores de P , intervalos de confianza y correlaciones son aproximaciones asintóticas); L = longitud patrón.



Estas estimaciones se mejoran considerablemente ponderando la regresión con el tamaño de los componentes gaussianos. En nuestro ejemplo la frecuencia está en la variable “size” (Cuadro 4.9):

```
. nl vbgf mbl age [freq = size]
(sum of wgt is 624)
```

Cuadro 4.9. Regresión no lineal ponderada para datos de longitud de bagres.

Fuente	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	Número de obs =	624	
				$F(3, 621) =$	941494	
Modelo	8299936.14	3	2766645.38	Prob > F =	0.0000	
Error	1824.8512	621	2.9385	$r^2 =$	0.9998	
				r^2 Ajustada =	0.9988	
Total	8301761	624	13304.1042	Error medio de estimación =	1.7142	
				Desviación del error =	2731.48	
Función de crecimiento de von Bertalanffy $L = L_{\infty}(1 - \exp(-K(\text{edad} - t_0)))$						
L	Coefficiente	Error estándar	t	P> t	[Intervalo de Confianza del 95%]	
L_{∞}	323.8079	2.7040	119.75	0.000	318.4977	329.1181
K	0.2734	0.0043	64.12	0.000	0.2650	0.2818
t_0	0.0256	0.0064	3.98	0.000	0.0130	0.0383

(Errores estándar, valores de P, intervalos de confianza y correlaciones son aproximaciones asintóticas); L = longitud patrón.



4.5 Caso 2. Longitud - Edad a partir de interpretación de estructuras duras.

Para ejemplificar este caso se utilizarán los datos de edad estimada para el pez japonés “suzuki” (*Lateolabrax japonicus*) reportados por Salgado-Ugarte (1995). Se presentan los resultados de las estimaciones de edad por lectura de escamas de individuos capturados durante los peces de primavera (inicio de la estación de crecimiento, Cuadro 4.10).

```
. use vbsuzuki, clear
```

Cuadro 4.10. Número de individuos por fecha de muestreo y edad estimada. La media de longitud patrón por edad se incluye en la columna de totales.

Número de muestra (1994)					
Edad	19/02	16/03	14/04	13/05	Total (Media)
0	14	3	0	0	17 (168.18)
1	4	4	2	1	11 (257.10)
2	0	5	2	7	14 (361.29)
3	0	2	0	2	4 (462.00)
11	0	0	1	0	1 (760.00)
Total	18	14	5	10	47

En este caso, al tener datos de edad, probamos primero la regresión no lineal:

```
. nl vbgf mbl age
. predict mblenl
(option yhat assumed; fitted values)

. scatter mbl age, ms(Oh) || mspline mblenl age, bands(30) ytitle("Longitud
patrón (mm)") xtitle("Edad (años)") yscale(range(200 1000)) ylab(200(200)1000)
legend(label(1 "Observados") label(2 "Calculados")) yline(862.24)
```



Cuadro 4.11. Regresión no lineal para datos de Longitud - Edad de *Lateolabrax japonicus*.

Fuente	Suma de Cuadrados	Grados de libertad	Cuadrados Medios	Número de obs =	5	
				$F(3, 2) =$	1183.25	
Modelo	1015387.3	3	338462.435	Prob > F =	0.0008	
Error	572.0887	2	286.0443	$r^2 =$	0.9994	
				r^2 Ajustada =	0.9986	
Total	1015959.4	5	203191.879	Error medio de estimación =	16.9128	
				Desviación del error =	37.8887	
Función de crecimiento de von Bertalanffy $L = L_{\infty}(1 - \exp(-K(\text{edad} - t_0)))$						
L	Coefficiente	Error estándar	t	$P > t $	[Intervalo de Confianza del 95%]	
L_{∞}	862.244	44.0435	19.577	0.003	672.7401	1051.748
K	0.1767	0.0257	6.872	0.021	.0661	0.2873
t_0	-1.1502	0.2169	-5.302	0.034	-2.0836	-0.2168

(Errores estándar, valores de P, intervalos de confianza y correlaciones son aproximaciones asintóticas); L = longitud patrón.

El intervalo de tiempo para las edades no es uniforme, por lo que es posible utilizar el método de Gulland-Holt para un ajuste alternativo (Cuadro 4.12):

```
. gulholtn mbl age, ytitle("Delta de longitud/Delta de edad") xtitle("Longitud media por delta de edad") tlttitle("Gráfico d
> e Gulland-Holt, {it:L{sub:{&infin}}} = 964.0516, {it:K} = 0.1432")
```

Estimation of L_{∞} and K values by the Gulland-Holt Method

```
-----
Intercept = 138.0650          Slope          = -0.1432
R-square   = 0.6190          Adj R-square  = 0.4285
L_inf.    = 964.0516          K              = 0.1432
-----
```

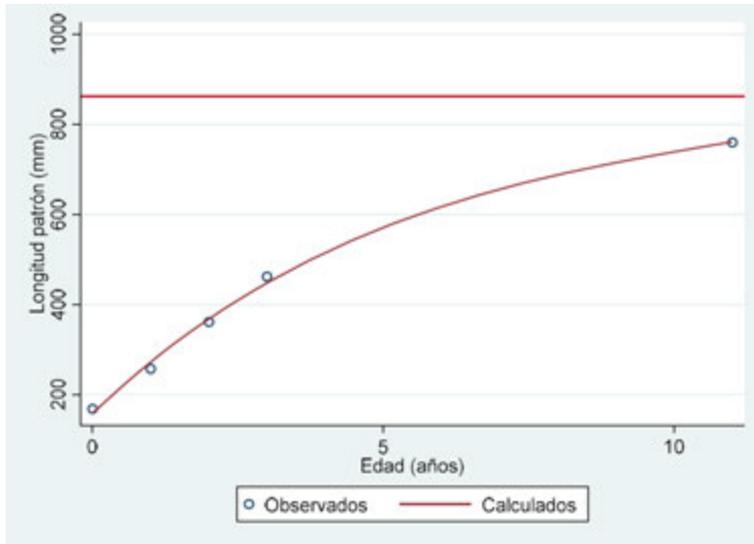


Figura 4.8. FCVB por regresión no lineal para datos de longitud media a cada edad del “suzuki” (*Lateolabrax japonicus*).

Cuadro 4.12. Estimación de L_{∞} y K por el método de Gulland-Holt para el “suzuki” *Lateolabrax japonicus*.

Intercepto =	138.0650	Pendiente =	-0.1432
r^2 =	0.6190	r^2 Ajustada =	0.4285
L_{∞} =	964.0516	K =	0.1432

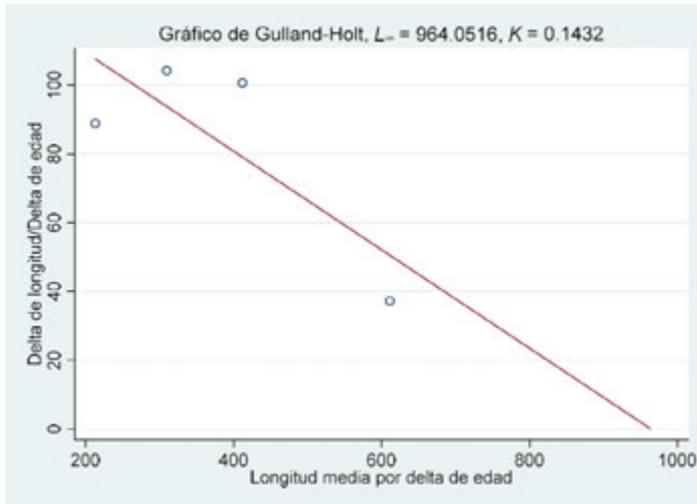


Figura 4.9. Gráfico de Gulland-Holt plot para los datos de longitud media a edad del "suzuki" (*Lateolabrax japonicus*).

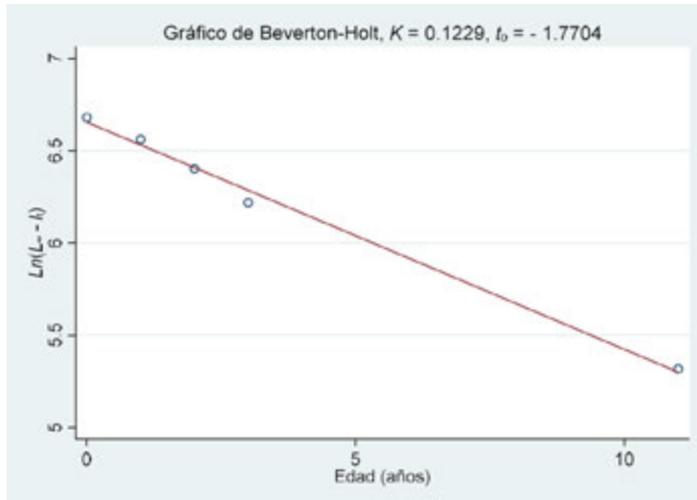


Figura 4.10. Gráfico de Beverton-Holt para estimar los parámetros K y t_0 de la FCVB de los datos de longitud media a edad del "suzuki" (*Lateolabrax japonicus*).



Usando este valor de L_{∞} en el procedimiento de Beverton-Holt llegamos a (Cuadro 4.13):

```
. bevhltm mbl age, linf(964.0516) ytitle("`it:Ln`({it:L}{sub:{&infin}}) - {it:l}{sub:t})") xtitle("Edad (años)") tlttitle("Gráfico de Beverton-Holt, {it:K} = 0.1229, {it:t}{sub:0} = - 1.7704")
```

Estimation of K and t_0 values by the Beverton-Holt Method

Intercept =	6.6536	Slope =	-0.1229
R-square =	0.9946	Adj R-square =	0.9929
K =	0.1229	t_0 =	-1.7704

Estimated von Bertalanffy Growth Function

$$l_t = 964.0516 * (1 - \exp(-0.1229 * (t + 1.7704)))$$

Cuadro 4.13. Estimación de K y t_0 por el método de Beverton-Holt para el “suzuki” *Lateolabrax japonicus*.

Intercepto =	6.6536	Pendiente =	-0.1229
r^2 =	0.9946	r^2 Ajustada =	0.9929
K =	0.1229	t_0 =	-1.7704

Función de crecimiento de von Bertalanffy estimada:

$$l_t = 964 (1 - \exp(-0.1229 * (t + 1.7704)))$$

```
. scatter mbl age, ms(Oh) || mspline mblebh age, bands(30) ytitle("Longitud patrón (mm)") xtitle("Edad (años)") yscale(range(200 1000)) ylab(200(200)1000) legend(label(1 "Observados") label(2 "Calculados")) yline(964.06) ylab(200(200)1000) xlab(0(2)12)
```

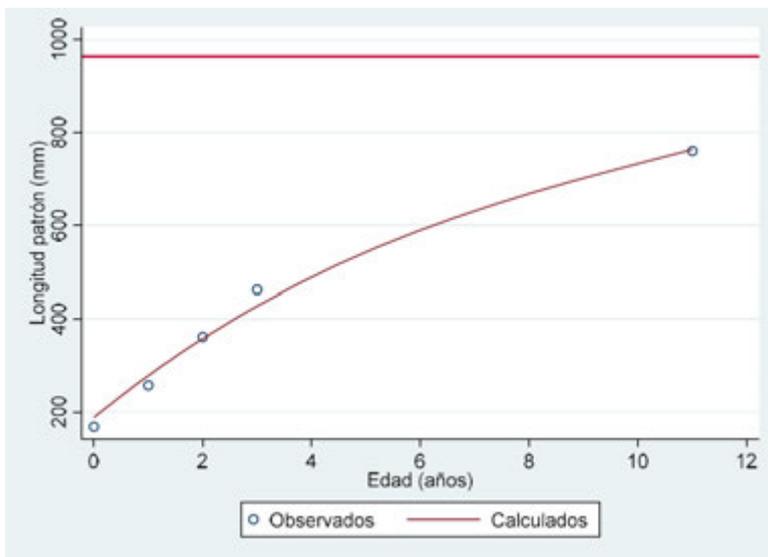


Figura 4.11. Ajuste alternativo de la FCVB para los datos de longitud media a edad del “suzuki” (*Lateolabrax japonicus*).

Aparentemente, el enfoque de regresión no lineal ajusta de manera más cercana a la FCVB a los valores de longitud media a cada edad del “suzuki”.

Hemos encontrado que los programas escritos para ejecutar las rutinas anteriormente presentadas, aunque muy simples en su mayoría, proporcionan un conjunto muy útil de herramientas para el análisis de datos biológico-pesqueros, permitiendo ahorrar tiempo y esfuerzo así como por contener enfoques tradicionales junto con los modernos. Las expresiones de crecimiento obtenidas con la ayuda de estos programas en combinación con información biológica adicional pueden introducirse en algoritmos más complejos de simulación como aquellos de Hilborn y Walters (1992), Haddon (2001) o King (1995; 2007). Se invita al lector a probar los programas y comunicarnos sus experiencias.



4.6 Algunas notas adicionales

Existen varios procedimientos computarizados adicionales para estimar el crecimiento a partir de edades estimadas por análisis de frecuencia de tallas o lectura de estructuras duras. En particular podemos mencionar aquellos incluidos en el FAO-ICLARM Stock Assessment Tools (FiSAT) (Gayanilo, *et al.* 1994).

Cada uno de los programas mencionados en el presente capítulo tienen sus archivos de ayuda (fordwal.hlp, gullplot.hlp, gullholt.hlp, bevholt.hlp and nlvbgf.hlp) y se incluyen en el disco acompañante. Además de los archivos de ayuda bhataplt.hlp y bhatgauc.hlp, se ha incluido un archivo general de ayuda: bhattado.hlp.

4.7 Otras funciones de crecimiento

Como se ha presentado en las secciones anteriores, la función de crecimiento de von Bertalanffy es muy popular para describir el crecimiento de organismos que crecen continuamente a lo largo de su vida y que es acelerado cuando son jóvenes y se va haciendo más lento en edades avanzadas.

Esta ecuación no sólo ha sido utilizada para expresar el crecimiento de organismos acuáticos (peces, moluscos, crustáceos, etc.), sino también describe adecuadamente el crecimiento de organismos terrestres como elefantes (Sukumar *et al.*, 1988; Shrader *et al.*, 2006), bóvidos (Malhado *et al.*, 2008; Ahrestani y Prins, 2011) y mamíferos en general (Zullinger *et al.*, 1984).

Otra tendencia común en los seres vivos es crecer lentamente a edades tempranas, aceleradamente en etapas intermedias y hacerse lento de nuevo al final de su periodo de vida. Estas trayectorias de desarrollo pueden ser descritas matemáticamente por otras funciones asintóticas comunes como la Logística (forma sigmoide simétrica), la de Gompertz (sigmoide asimétrica) u otras un poco más sofisticadas como la de von Bertalanffy para crecimiento en peso o la expresión generalizada de Richards (1959).

A diferencia del modelo exponencial de crecimiento sin restricciones propuesto por Malthus en 1798, aunque según Murray (2001) esta función fue propuesta primero por Euler, en la práctica, el crecimiento está limitado por los recursos disponibles (Grosjean, 2001). Es así que Verlhust (1838) (Delmas, 2004) propuso una expresión matemática con un término auto limitante que representa teóricamente un recurso limitante, lo que resulta en una ecuación asintótica conocida como la Función Logística de Crecimiento. Esta función,



originalmente propuesta para el crecimiento poblacional, proporciona una representación matemática para la ley autocatalítica de la fisiología y la química. La forma integral de su ecuación puede ser escrita como:

$$W_t = W_\infty / (1 + \exp(-g(t - t_0))) \quad (4.13)$$

Esta curva sigmoideal tiene un punto de inflexión donde W es igual a $W_\infty/2$ y delimita dos mitades antisimétricas (los dos extremos de la “S” son similares. La tasa instantánea de crecimiento cuando $W = 0$ es el parámetro g y t_0 es el tiempo al cual la tasa de crecimiento es máxima; es también la abscisa del punto de inflexión de la curva para el cual $W = W_\infty/2$. Tiene dos asíntotas el cero ($W = 0$) y el valor de tamaño asíntótico W_∞ (Grosjean, 2001).

De acuerdo con Moreau (1987), la curva logística ha sido usada en muy pocas ocasiones para describir el crecimiento individual de los peces y cita al trabajo de Miura *et al.* (1976) quienes describen la forma de crecimiento estacional anual de cuatro grupos de edad del salmón *Onchorhynchus rhodurus*. Recientemente y debido a la disponibilidad de programas de ajuste no lineal, la función logística está siendo usada cada vez más. Un ejemplo de su aplicación se encuentra en Chale-Matsau, *et al.* (2001) quienes ajustaron una curva logística de crecimiento a datos históricos de *Polysteganus undulosus* capturado en la pesquería de anzuelo y línea de KwaZulu-Natal de 1962 a 1963. No obstante, esta función frecuentemente ajusta el crecimiento de organismos invertebrados. Ejemplos son Grosjean (2001) para el erizo de mar *Paracentrotus lividus*; Katsanevakis (2007) para el mejillón *Pinna nobilis*; Aragón-Noriega (2013) para el bivalvo *Atrina maura*; Saito-Quezada (2018) para los gasterópodos marinos *Hexaplex princeps* y *Opeatostoma pseudodon* entre otros.

La función de Gompertz es otra curva en forma de “S” con asíntotas tanto superior como inferior ($W_t = 0$ y W_∞), un punto de inflexión ($t_0, W_\infty/e$) pero mitades asimétricas. Esta curva expresa la tasa decreciente de crecimiento a través de las etapas adultas de la vida del pez y puede considerarse como la expresión de la actividad de dos tipos diferentes y opuestos de factores reguladores durante el crecimiento (Ricker, 1979). Una expresión integrada para la función de Gompertz es:

$$W_t = W_\infty \exp(-\exp(-g(t - t_0))) \quad (4.14)$$

Donde t_0 es la abscisa del punto de inflexión y g es la tasa instantánea de crecimiento cuando $t = t_0$ (Fletcher 1973, en Moreau, 1987; Grosjean, 2001).



Como un ejemplo del uso de esta función, Moreau (1987) cita el artículo de Zweifel y Lasker (1976) quienes usaron dos ciclos sucesivos de Gompertz para describir el crecimiento de larvas de la anchoveta *Engraulis mordax*. Aunque existen algunos reportes de su uso, sobre todo con el empleo de las computadoras (Silliman, 1967), no se usa por lo general en problemas de biología pesquera ya que la función de elección es la de von Bertalanffy, la cual se inserta en los cálculos de rendimiento. No obstante Winsor (1932) y Ebert (1999) citan varias aplicaciones para datos de crecimiento (Grosjean, 2001).

Moreau (1987) encuentra que la función de Gompertz describe de cerca el peso a cada edad, siguiendo los incrementos cada vez más grandes durante los años tempranos y a los datos de longitud a cada edad con un punto de inflexión. Este es el patrón que presentan usualmente los peces anádromos los cuales tienen un crecimiento lento inicial en el agua dulce seguido por un crecimiento más rápido durante las etapas de vida marina (Moore y Moore 1974). La función de Gompertz es adecuada para especies tropicales de corta vida con cambios significativos en su alimentación al pasar de etapas juveniles a adultas. Este autor nota la necesidad de investigar adicionalmente el uso de esta función debido a que los cálculos de rendimiento pueden efectuarse con cualquier ecuación de crecimiento con la ayuda de las computadoras.

Citas adicionales al uso de la función de Gompertz son Chen y Soong (2002) para el crecimiento del caracol *Coralliophila violácea* del Indo-Pacífico además de los trabajos citados anteriormente para la función Logística (Katsanevakis, 2007; Aragón-Noriega, 2013; Saito-Quezada, 2018).

4.8 Caso 3. Crecimiento del caracol *Hexaplex princeps*

Para estudiar algunos aspectos de la biología pesquera del “caracol chino” *Hexaplex princeps* (Figura 4.12) de los litorales rocosos aledaños a Puerto Ángel, Oaxaca, México, se obtuvieron recolectas con periodicidad aproximadamente mensual de febrero de 2012 a noviembre de 2014 con la ayuda de los buzos locales. Para estimar la edad y modelar el crecimiento, Saito-Quezada (2018) midió las conchas desde el ápice hasta la base (largo) y se separaron los opérculos, los cuales fueron pulidos para facilitar la identificación y medición de los anillos de crecimiento de acuerdo a las recomendaciones de Morris (1969), Gluyas-Millán *et al.* (1999) y Saito-Quezada (2010).



Figura 4.12. “Caracol chino” *Hexaplex princeps*, concha y opérculo (Saito-Quezada, 2018).

De los 804 individuos obtenidos en 21 recolectas fue posible analizar 665 opérculos (buen estado y distinción de marcas de crecimiento). En conjunto, se llevaron a cabo un total de 10,808 medidas en los opérculos (radios y bordes). Un resumen de esta información se presenta en el Cuadro 4.14. y su versión gráfica en la Figura 4.13.

```
. use anillostotales
```

```
. sum a1-a30
```



Cuadro 4.14. Resumen numérico de las medidas de los anillos en los opérculos de *Hexaplex princeps* (D.E. = Desviación Estándar).

Anillo	Frecuencia	Media	D.E.	Mínimo	Máximo
a1	663	2.6708	0.2349	1.09	3.92
a2	664	4.4228	0.4443	1.68	6.14
a3	665	6.0862	0.6674	2.42	8.48
a4	665	7.6744	0.9056	3.71	11.11
a5	665	9.2106	1.1021	4.61	13.55
a6	665	10.8230	1.2787	5.74	16.15
a7	664	12.4292	1.4420	6.80	17.92
a8	663	14.1189	1.6721	7.54	20
a9	662	15.8272	1.8378	8.99	21.51
a10	658	17.5735	2.0250	10.06	24.73
a11	651	19.3803	2.2203	11.99	28.07
a12	636	21.1773	2.3185	12.85	29.73
a13	596	22.8867	2.4193	13.31	32.22
a14	538	24.4263	2.3616	16.19	32.01
a15	414	25.7385	2.3722	17.69	32.55
a16	281	26.9275	2.4503	18.91	34.61
a17	182	28.1220	2.4342	20.42	33.61
a18	110	29.2079	2.2966	22.53	33.95
a19	55	30.3131	2.3202	23.83	36.63
a20	26	31.0177	2.4115	25.63	35.05
a21	11	31.5636	2.9446	27.06	35.05
a22	4	31.6550	3.1058	28.52	35.28
a23	3	34.3433	2.5637	31.62	36.71
a24	1	32.7900	-	32.79	32.79
a25	1	34.0100	-	34.01	34.01
a26	1	35.0100	-	35.01	35.01
Borde	664	7.6348	1.1646	3.50	38.95
Total	10808	15.5318	8.2814	1.09	38.95



```
. graph box a1-a26, legend(rows(3) symxsize(5)) ytitle("Radio (mm)" " ")
```

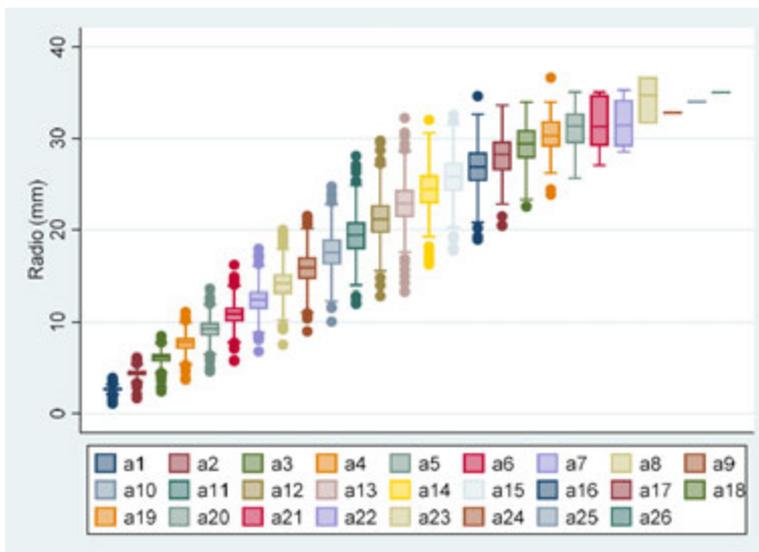


Figura 4.13. Diagramas de caja y bigotes para los radios de los anillos en óperculos de *Hexaplex princeps*.

Con las medidas de los radios se calculó la razón de incremento marginal. Los resultados se procesaron con el suavizador no lineal resistente compuesto 4253EH, doble (Velleman 1980; Salgado-Ugarte 2017) aplicado a los datos multiplicados por una constante (10^6) en escala logarítmica (García-Berthou y Moreno-Amich 1993) para evitar sesgo y trabajo con números menores que cero). En los casos de valores faltantes en la secuencia, estos se remplazaron por interpolación lineal (medias aritméticas de valores adyacentes). El resultado para las hembras se presenta en la Figura 4.14.

```
. use validahp
. scatter smlnrhm index, c(1) xlab(1 "F" 2 "M" 3 "A" 4 "M" 5 "J" 6 "J" 7 "A" 8
"S" 9 "O" 10 "N" 11 "D" 12 "E" 13 "F" 14 "M" 15 "A" 16 "M" 17 "J" 18 "J" 19 "A"
20 "S" 21 "O" 22 "N" 23 "D" 24 "E" 25 "F" 26 "M" 27 "A" 28 "M" 29 "J" 30 "J"
31 "A" 32 "S" 33 "O" 34 "N") xline(11.5 23.5) text(13.9 2 "2012") text(13.9 13
"2013") text(13.9 25 "2014") ytitle("Ln de {it:RIM} suavizado") xtitle("Tiempo
(mes)") yline(13.68) xsize(8)
```



Es claro en esta figura que valores mínimos se tuvieron a fines e inicios del año. Por esto, es bastante razonable considerar una periodicidad anual en la formación de las marcas de crecimiento, las cuales parecen originarse a fines de cada año.

Tomando como base la periodicidad anual, es posible estimar funciones de crecimiento a la información de los tamaños (largo de la concha) para cada edad estimada. Estos datos se presentan en la Figura 4.14 para las hembras de *H. princeps*.

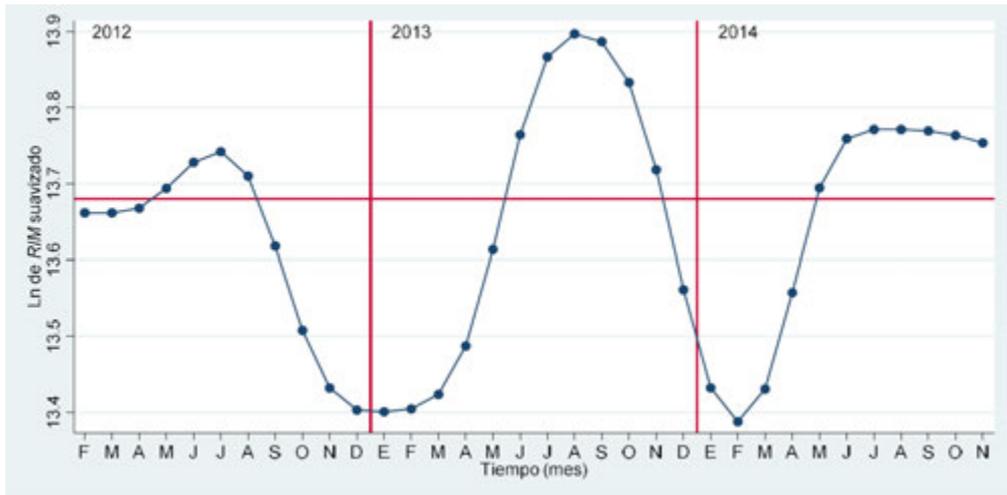


Figura 4.14. Valores medianos (por 10^6) logarítmicos de la Razón de Incremento Marginal (RIM) suavizados por 4253EH, doble para las hembras a lo largo de la temporada de estudio. Se incluye línea horizontal con la mediana de todos los valores como referencia y líneas verticales para separar años.

Los comandos para construir esta gráfica se presentan a continuación. Primeramente se usa la regresión no lineal para estimar los parámetros de cada modelo y se utiliza en cada caso el comando “predict” para la obtención de los valores estimados de longitud por cada modelo. Como se aprecia en la Figura 4.15, fue posible ajustar los tres modelos a estos datos. Un resumen de los parámetros estimados por cada función de crecimiento se incluye en el Cuadro 4.15.



```
. use anillostotales2
```

```
. nl vbgf largo edad if sexo==1
. predict levb if sexo==1
```

Salida omitida

```
. nl gom3 largo edad if sexo==1
. predict lgomeh if sexo==1
```

Salida omitida

```
. nl log3 largo edad if sexo==1
. predict llogeh if sexo==1
```

Salida omitida

Como puede apreciarse en la Figura 4.15, las curvas ajustadas son muy semejantes y una interrogante que surge es ¿qué modelo es el mejor? Desde un punto de vista práctico cualquiera de ellos conduce a estimaciones razonables. Pudiéramos elegir tomando como base el tamaño asintótico más razonable. La talla máxima (largo) medida en los ejemplares hembras de este estudio fue de 13.45 cm. y la literatura reporta un valor de 15.24. De acuerdo con este criterio el modelo que más se acerca en su asíntota a estos valores es el de von Bertalanffy.

También pudiéramos tomar en cuenta el valor del coeficiente de determinación de las regresiones. Por una pequeña diferencia, este valor parece favorecer al modelo logístico.

Cuadro 4.15. Parámetros estimados por cada modelo y coeficiente de determinación de cada regresión (no lineal).

Función	L_{∞}	K	t_0	r^2
von Bertalanffy	12.53	0.12	4.01	0.9902
Gompertz	11.74	0.19	7.23	0.9902
Logístico	11.34	0.25	8.86	0.9903



Aplicando un punto de vista estadístico mas elaborado, es posible considerar los “Criterios de Información (CI)”. Estas métricas simples combinan una medida de “falta de ajuste” de un modelo a un conjunto de datos con una “penalización” por el número de parámetros (Ogle, 2016). Actualmente dos valores son los más utilizados: el “Criterio de Información de Akaike (AIC por sus siglas en inglés)” (Akaike, 1974) y el “Criterio de Información Bayesiano (BIC)” (Schwartz, 1987). Valores pequeños de CI indican un mejor modelo, esto es, falta de ajuste y número de parámetros bajos (Burnham y Anderson, 2002; Ogle, 2016).

El cuadro 4.16 resume los resultados de la obtención de estos índices por medio de Stata y de R de acuerdo a los comandos presentados. Nótese que los comandos de Stata son una variación de los anteriormente presentados que consiste en agregar dos puntos (“:”) después de los modelos incluidos en Stata (“gom3” y “log3”). Para von Bertalanffy es necesario escribir la función de acuerdo a la sintaxis para “ecuaciones sustituibles” de Stata (la expresión algebraica entre paréntesis y los parámetros entre llaves).

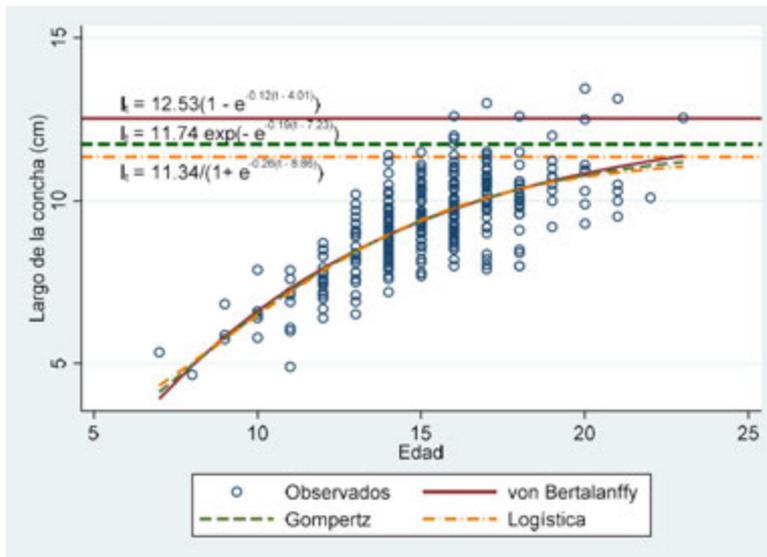


Figura 4.15. Datos de talla (longitud de la concha) – edad (años) y funciones de crecimiento (von Bertalanffy, Gompertz y Logística) para las hembras de *Hexaplex princeps*.



Debe notarse que aunque muy parecidos, los valores puntuales resultantes de la aplicación de los respectivos comandos en cada programa no son exactamente iguales, lo que puede atribuirse a su implementación particular en cada uno. No obstante, la información que proporcionan es la misma. Los valores mayores son los obtenidos a partir de la función de crecimiento de von Bertalanffy, seguida por la de Gompertz y los valores más bajos de la función Logística (si bien la diferencia entre ellos es pequeña).

```
. scatter largo lvbeh lgomeh llogeh edad if sexo==1, ms(Oh i i i) c(. 1 1
1) lpattern(. solid dash dash_dot) lw(. 5 .5 .5) sort ytitle("Largo de la
concha (cm)") yline(12.53, lcolor(maroon) lw(.5) ) text(13.9 "l{subscript:t}
= 12.53(1 - e{superscript:-0.12(t - 4.01)})", justification(left)) text(12.1
9 " l{subscript:t} = 11.74 exp(- e{superscript:-0.19(t - 7.23)})" ,
justification(left)) ylab(5(5)15) text(10.9 9 " l{subscript:t} = 11.34/(1+
e{superscript:-0.26(t - 8.86)})", justification(left)) legend(order(1 2 3 4)
lab(1 "Observados") lab(2 "von Bertalanffy") lab(3 "Gompertz") lab(4 "Logística")
rows(2)) || scatter largo edad if sexo==1, ms(i) yline(11.74, lpattern(dash)
lcolor(dkgreen) lw(.5)) || scatter largo edad if sexo==1, ms(i) yline(11.34,
lpattern(dash_dot) lcolor(orange) lw(.5))
```

```
. nl (largo = {L} * (1 - exp(-{k}*(edad - {t0}))) ) if sexo==1 & edad!=",
initial(L 50 k 1 t0 .1)
```

Salida omitida

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	336	.	-453.8641	3	913.7282	925.1796

Note: BIC uses N = number of observations. See [R] BIC note.

```
. nl gom3: largo edad if sexo==1
```

Salida omitida

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
-------	---	----------	-----------	----	-----	-----



```
-----+-----
      . |          336          . -452.7655          3          911.531          922.9823
-----+-----
```

Note: BIC uses N = number of observations. See [R] BIC note.

```
. nl log3: largo edad if sexo==1
```

Salida omitida

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

```
-----+-----
      Model |          N   ll(null)   ll(model)          df          AIC          BIC
-----+-----
      . |          336          .   -452.134          3          910.268          921.7194
-----+-----
```

Note: BIC uses N = number of observations. See [R] BIC note.

Trabajando con R

```
> setwd("c:/data")
> library(foreign)

> anitot <- read.dta("anillostold.dta")

> anitoth <- subset(anitot, sexo==1)

> vbgf <- function(t,L,k,t0) L*(1-exp(-k*(t-t0)))

> res.vb <- nls(largo~vbgf(edad, L, k, t0), start= c(L=12, k=.1, t0=1),
na.exclude, data=anitoth)

> AIC(res.vb)
[1] 915.7282
> BIC(res.vb)
[1] 930.9967

> res.gom <- nls(largo~gomf(edad, L, k, t0), start= c(L=12, k=.1, t0=1),
na.exclude, data=anitoth)
> AIC(res.gom)
[1] 913.531
> BIC(res.gom)
[1] 928.7994
```



```
> res.log <- nls(largo~logf(edad, L, k, t0), start= c(L=12, k=.1, t0=1),
na.exclude, data=anitoth)
> AIC(res.gom)
[1] 913.531
> AIC(res.log)
[1] 912.268
> BIC(res.log)
[1] 927.5365
```

Todo lo anterior nos lleva a que con los valores del cuadro 4.16 el modelo que mejor ajusta a los datos obtenidos es la función Logística.

Cuadro 4.16. Comparación de valores de los criterios de información de Akaike y Bayesiano obtenidos con los programas R y Stata.

Función de crecimiento	Resultados con R		Resultados con Stata	
	AIC	BIC	AIC	BIC
vBGF	915.7282	930.9967	913.7282	925.1796
Gompertz	913.531	928.7994	911.531	922.9823
Logística	912.268	927.5365	910.268	921.7194

4.9 Caso 4. Crecimiento del holoturoideo (pepino de mar) *Holothuria inornata*

Los pepinos de mar son un recurso marino muy apreciado en países de Asia, tanto como alimento como medicamento tradicional desde hace cientos de años lo que los ha llevado a sobre-explotación y a riesgo de extinción (Rahman, 2015). En México, el interés por este recurso ha surgido en las últimas décadas (Fajardo-León y Vélez, 1996; Herrero-Perezrul, 2004; López-Rocha, 2012).

De noviembre de 2017 a octubre de 2018 se llevaron a cabo visitas a Puerto Ángel, Oax. México y con la ayuda de los buzos locales se capturaron individuos del pepino de mar “arenero” cuyo nombre científico es *Holothuria inornata* (Figura 4.16). Los individuos se conservaron en frío para su transporte y descongelados en laboratorio se midieron y disectaron. Para investigar el crecimiento como primera aproximación se utilizó el análisis de la distribución de tallas, que en este caso fue el largo del cuerpo. En agosto de 2018 se capturaron 76 individuos (Morales-Ramírez, 2018).



Figura 4.16. Individuos capturados del pepino de mar “arenero” *Holothuria inornata*.

```
. use pepinodatos
. bandw1 largo if colecta==6
```

Con la mitad de la banda óptima de Silverman al ser datos no gaussianos con indicación multimodal se obtiene el estimador de densidad por kernel de la Figura 4.17. Con las opciones “numodes” y “modes” del comando actualizado “warpendenml.ado” se cuenta y estiman las modas de la distribución.

```
. warpendenml largo if colecta==6, b(.57) m(10) k(6) numo mo ytitle("Densidad")
xtitle("Largo del cuerpo (cm)") t1("Densidad WARP (poligonal), bw = 0.57, m =
10, k = 6")
```

```
Number of modes = 7
```

```
Modes in WARPing density estimation, bw = .57, M = 10, Ker = 6
```

```
Mode ( 1 ) = 7.9800
Mode ( 2 ) = 9.5760
Mode ( 3 ) = 11.4000
```



Mode (4) =	19.3800
Mode (5) =	22.1160
Mode (6) =	25.0800
Mode (7) =	25.9920

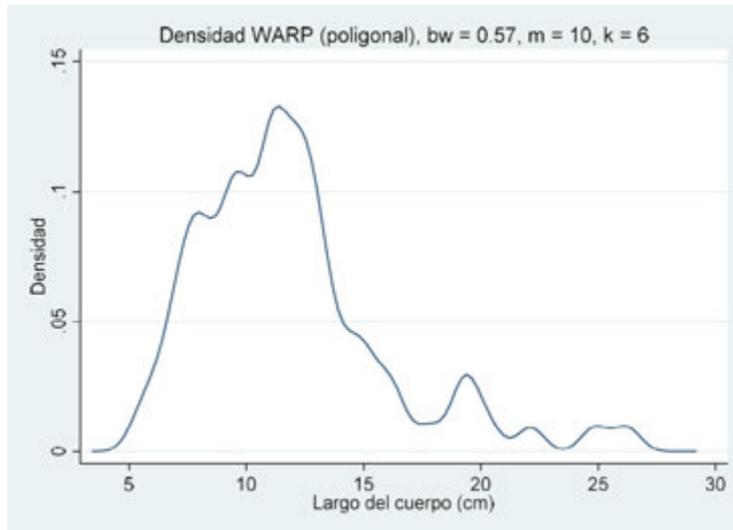


Figura 4.17. Distribución de tallas de individuos de *Holothuria inornata* colectados en agosto de 2018.

Los valores modales considerados como representantes de cohortes y asumiendo una periodicidad anual mostraron una tendencia claramente sigmoideal. Por esto se probó el ajuste mediante funciones de crecimiento asintóticas sigmoideales, Logística y de Gompertz.

Los resultados se presentan en la Figura 4.17. Sin dejar de reconocer que este es un intento muy preliminar para describir el patrón de crecimiento de *H. inornata*, los valores obtenidos para el modelo de crecimiento Logístico resultan bastante razonables, con un valor de r^2 (0.9956) muy cercano a la unidad y con todos los parámetros estimados significativamente ($P < 0.05$; 95 % de confianza). El tamaño asintótico obtenido concuerda bien con lo reportado para la especie en otros lugares de México (Prieto-Rios *et al.*, 2014) quienes reportan tallas hasta de 40 cm y miembros del mismo género (Nisperuza-Pérez *et al.*, 2018) con tallas de 30 cm.



```
. use pepinodatos
```

```
. nl log3: largo edad
(obs = 7)
```

```
Iteration 0: residual SS = 176.6955
Iteration 1: residual SS = 151.4026
Iteration 2: residual SS = 123.5133
Iteration 3: residual SS = 99.67854
Iteration 4: residual SS = 95.935
Iteration 5: residual SS = 42.31181
Iteration 6: residual SS = 12.71436
Iteration 7: residual SS = 10.90341
Iteration 8: residual SS = 10.81472
Iteration 9: residual SS = 10.81244
Iteration 10: residual SS = 10.81241
Iteration 11: residual SS = 10.8124
Iteration 12: residual SS = 10.8124
Iteration 13: residual SS = 10.8124
```

Source	SS	df	MS		
Model	2443.9697	3	814.656558	Number of obs =	7
Residual	10.812405	4	2.70310121	R-squared =	0.9956
				Adj R-squared =	0.9923
				Root MSE =	1.644111
Total	2454.7821	7	350.683154	Res. dev. =	22.90863

3-parameter logistic function, largo = $b1/(1 + \exp(-b2*(edad - b3)))$

largo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/b1	30.03391	4.049058	7.42	0.002	18.79193	41.2759
/b2	.5585666	.1450583	3.85	0.018	.1558202	.961313
/b3	3.284915	.6266665	5.24	0.006	1.54501	5.024821

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	7	.	-11.45431	3	28.90863	28.74636

Note: BIC uses N = number of observations. See [R] BIC note.



El ajuste de la ecuación de Gompertz si bien parecido al logístico tuvo un coeficiente de determinación menor (0.9904) y valores de Criterios de Información (Akaike y Bayesiano) mayores (30.48 y 30.31, respectivamente). Por lo que sólo se reporta el modelo Logístico.

```
. scatter largo edad, ms(Oh) || mspline larloge edad, bands(40) yscale(range(5  
35)) ylab(5(5)35) yline(30) ytitle("Largo de  
> 1 cuerpo (cm)") xtitle("Edad") legend(label(1 "Observados") label(2  
"Calculados") pos(5) ring(0))
```

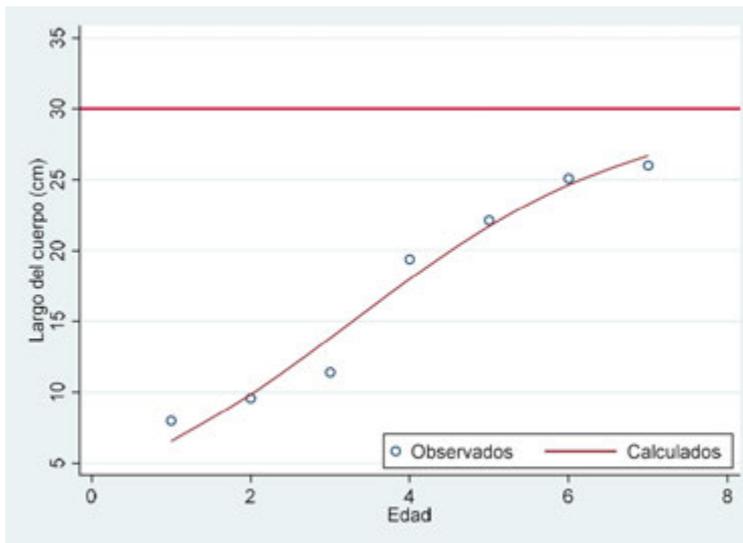


Figura 4.18. Datos talla-edad y ajuste por regresión no lineal de la Función Logística de crecimiento para individuos de *H. inornata*, agosto 2018 Puerto Ángel, Oax. México (Morales-Ramírez, 2018).

4.10 Ajuste de otras funciones con R

En el apéndice 4 se ha incluido el ajuste de los modelos de von Bertalanffy, Logístico, Gompertz y adicionalmente el de Richards a los datos de crecimiento de erizo de mar (Grosjean, 2001; Verzani, 2014) y la curva “lineal” de von Bertalanffy a datos de la corvina roja (Beckman *et al.*, 1988; Wilson *et al.*, 1993; Verzani, 2014).



4.11 Crecimiento estacional. Ajuste para la FCvB

El crecimiento de la mayor parte de las especies de plantas y animales ectotermos tales como peces, reptiles y numerosos grupos de invertebrados (crustáceos, moluscos, equinodermos) varía de manera considerable a lo largo del año por efectos de las fluctuaciones en las condiciones ambientales (temperatura, luz, cantidad de alimento, entre otras) (García-Berthou, *et al.*, 2012).

Lo anterior ocurre claramente en regiones templado-frías del planeta en donde las condiciones ambientales manifiestan oscilaciones muy marcadas. Es así que el crecimiento disminuye durante los meses fríos de invierno y se acelera durante primavera y verano (Pitcher y MacDonald, 1973; Haddon, 2001).

En latitudes tropicales durante el año se presentan temporadas de lluvia y de estío que producen cambios ambientales (por ejemplo, la entrada de nutrientes por las descargas fluviales) que modifican la tasa de crecimiento de los organismos acuáticos (King, 2007; García-Berthou *et al.*, 2012).

Las variaciones en las tasas de crecimiento pueden ser tan importantes que hacen que el modelo tradicional de crecimiento de von Bertalanffy proporcione una descripción incompleta del patrón de desarrollo (Pitcher y MacDonald, 1973).

Es por esto que se han propuesto varios modelos para describir esta variación. Quizás debido al uso predominante de la Función de Crecimiento de von Bertalanffy (FCvB) que la mayoría son modificaciones a su expresión clásica. Una de ellas es la propuesta por Pitcher y MacDonald (1973) y Pauly y Gaschütz (1979) aquí en la versión modificada por Haddon (2001):

$$L_t = L_\infty \left(1 - e^{-\left[C \sin \left(\frac{2\pi(t-s)}{52} \right) + K(t-t_0) \right]} \right) \quad (4.15)$$

donde C está relacionada con la magnitud de las oscilaciones arriba y debajo de la curva de crecimiento no estacional, s es el punto de inicio en el tiempo para la curva sinusoidal (relacionada con la fase) y las otras constantes, K , L_∞ y t_0 son los parámetros definidos en la función de crecimiento de von Bertalanffy. La variable t es la edad a la longitud L_t . El valor de 52 indica que la escala de tiempo de los eventos son semanas (por lo tanto s , t y t_0 serán medidos en semanas. En efecto, esta ecuación es la curva de von Bertalanffy con una onda sinusoidal agregada (Haddon, 2001).



Para seguir con el razonamiento de Haddon (2001), presentaremos su ejemplo de análisis a la submuestra (Cuadro 4.17) de datos de pececillos (*Phoxinus phoxinus*) reportados por Pitcher y MacDonald (1973).

Cuadro 4.17. Datos de longitud (mm) de *Phoxinus phoxinus* medida a intervalos de tiempo (semanas) tomados de Haddon (2001) quien los midió de una figura de Pitcher y MacDonald (1973).

Edad	Longitud	Edad	Longitud	Edad	Longitud
1	3	52	30	114	65
4	9	60	36	122	67
8	15	64	38	132	68
10	12	70	48	134	64
14	19	72	45	138	65
24	24	82	49	146	67
28	24	90	49	152	68
30	21	94	52	158	71
38	21	104	59	172	75
48	24	112	61	176	73

La Figura 4.19 muestra el patrón de crecimiento observado (círculos). Puede notarse una clara variación del crecimiento a lo largo de las semanas. Ajustando el modelo especificado por la ecuación 4.15 se obtiene la curva incluida en el gráfico, al cual se le ha sobrepuesto también un ajuste clásico de von Bertalanffy (línea punteada).

Presentamos el ajuste llevado a cabo con Stata. Para poder llevarlo a cabo fue necesario actualizar nuestros programas. El primero de ellos es “svbgff.ado” que utiliza la modalidad de utilizar un programa evaluador de función, el cual requiere que se le den valores iniciales en un vector fila. Este vector se crea utilizando el lenguaje para matrices de Stata. Una vez especificados los valores iniciales se llama al programa y el comando “nl” de Stata lleva a cabo la regresión no lineal para ajustar la función. Los comandos son:

```
. use seasonvdata
. matrix ivals = (100, .01, .1, .1, 4)
```



```
. nl svbgff @ obsl weeks, parameters(Li K t0 C s) initial(ivals)
(obs = 30)
```

```
Iteration 0: residual SS = 353.3595
Iteration 1: residual SS = 196.8611
Iteration 2: residual SS = 196.8576
Iteration 3: residual SS = 196.8576
Iteration 4: residual SS = 196.8576
```

Source	SS	df	MS		
Model	72677.142	5	14535.4285	Number of obs =	30
Residual	196.85759	25	7.87430354	R-squared =	0.9973
				Adj R-squared =	0.9968
				Root MSE =	2.806119
Total	72874	30	2429.13333	Res. dev. =	141.5748

obsl	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/Li	106.8611	9.559888	11.18	0.000	87.17212	126.55
/K	.0067863	.0010863	6.25	0.000	.0045491	.0090235
/t0	-5.200722	2.644609	-1.97	0.060	-10.6474	.245952
/C	.0601118	.0160822	3.74	0.001	.0269898	.0932338
/s	4.005121	1.507403	2.66	0.014	.9005674	7.109675

```
. predict obsle
```

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	30	.	-70.7874	5	151.5748	158.5808

```
. nl vbgfn: obsl weeks
```

Salida omitida

```
. predict obsle1
```

```
. estat ic
```



Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	30	.	-80.34045	3	166.6809	170.8845

```
. scatter obs1 weeks, ms(0h) || mspline obsle weeks, bands(40) legend(label(1
"Observados") label(2 "vBGF estacional") label(3 "vBGF regular") pos(5) ring(0)
cols(1)) ytitle("Longitud (mm)") xtitle("Edad (semanas)") || mspline obsle1
weeks, bands(40) lpattern(dash)
```

Para el ajuste de la FCvB regular se usó la también actualizada versión “nlvbgfn.ado” que utiliza una sintaxis diferente por ser un programa que especifica una “expresión sustituible” (StataCorp, 2019). Aunque es evidente que el modelo estacional recupera de mejor manera el comportamiento descrito por los datos (Figura 4.19), se obtuvieron los valores de los criterios de información (Akaike y Bayesiano), los cuales son menores en el modelo estacional (AIC = 151.45, BIC = 158.58) que aquellos del modelo regular de von Bertanffy (AIC = 166.68, BIC = 170.88).

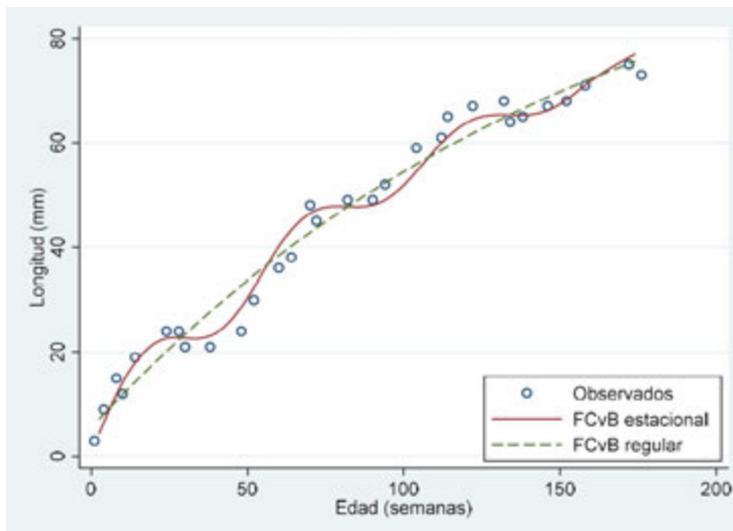


Figura 4.19. Ajuste de la FCvB estacional y la FCvB regular a los datos de *Phoxinus phoxinus* (Pitcher y MacDonald, 1973) modificados por Haddon (2001).



Haddon (2001) lleva a cabo la evaluación de este modelo por medio del análisis de los residuales (para más información sobre este procedimiento estadístico consultar a Bolen y Jackman, 1990; Salgado-Ugarte, 2017) y obtiene una gráfica que muestra un patrón en los valores que sugieren un comportamiento cíclico no considerado por el modelo anterior (Figura 4.20). Por esto, sugiere el empleo de un modelo que contenga una segunda expresión sinusoidal para explicarlo. La ecuación sería:

$$L_t = L_\infty \left(1 - e^{-\left[C_1 \text{sen} \left(\frac{2\pi(t-s_1)}{52} \right) + C_2 \text{sen} \left(\frac{2\pi(t-s_2)}{p} \right) + K(t-t_0) \right]} \right) \quad (4.16)$$

donde p es el periodo del segundo ciclo y los otros parámetros son como antes, sólo duplicados en cada ciclo.

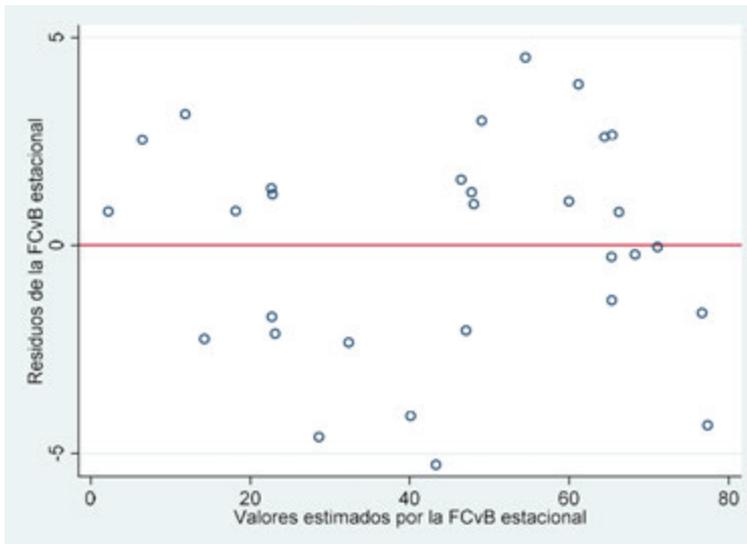


Figura 4.20. Gráfica de residuos de la FCvB estacional. Los puntos sugieren un patrón cíclico no explicado.

Este modelo se ha incluido en el programa “nlsvbf2.ado” el cual también requiere un vector fila con los valores iniciales a utilizar por el método iterativo (Gauss-Newton) de regresión no lineal incluido en el programa Stata.

Métodos cuantitativos computarizados para biología pesquera



```
. matrix ivals = (100,.01,.1,.1,5,.5,-50,100)
. nl svbfff2 @ obs1 weeks, parameters(Li K t0 C1 s1 C2 s2 p) initial(ivals)
(obs = 30)
```

Iteration 0: residual SS = 680.4992

.
.
.

Iteration 12: residual SS = 4.54e-11

Source	SS	df	MS		
Model	72677.141	8	9084.64266	Number of obs =	30
Residual	4.536e-11	22	2.0618e-12	R-squared =	1.0000
				Adj R-squared =	1.0000
				Root MSE =	1.44e-06
Total	72677.141	30	2422.57138	Res. dev. =	-731.3919

obs1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/Li	106.8611	.0000167	6.4e+06	0.000	106.861	106.8611
/K	.0067863	2.13e-09	3.2e+06	0.000	.0067863	.0067863
/t0	-5.200719	6.29e-06	-8.3e+05	0.000	-5.200732	-5.200706
/C1	.0601118	1.62e-08	3.7e+06	0.000	.0601118	.0601118
/s1	4.00512	8.71e-07	4.6e+06	0.000	4.005118	4.005122
/C2	-8.92e-09	1.76e-08	-0.51	0.618	-4.55e-08	2.76e-08
/s2	-80.93414	159.0912	-0.51	0.616	-410.869	249.0007
/p	136.2855	139.6875	0.98	0.340	-153.4087	425.9797

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	30	.	365.696	7	-717.3919	-707.5836

Note: BIC uses N = number of observations. See [R] BIC note.

```
. predict les2
```

```
. scatter obs1 weeks, ms(Oh) || mspline les2 weeks, bands(40) || mspline obsle
weeks, bands(40) lpattern(dash) legend(label(1 "Observados") label(2 "FCvB
estacional dos ciclos") label(3 "FCvB estacional un ciclo") pos(5) ring(0)
cols(1)) ytitle("Longitud (mm)") xtitle("Edad (semanas)")
```

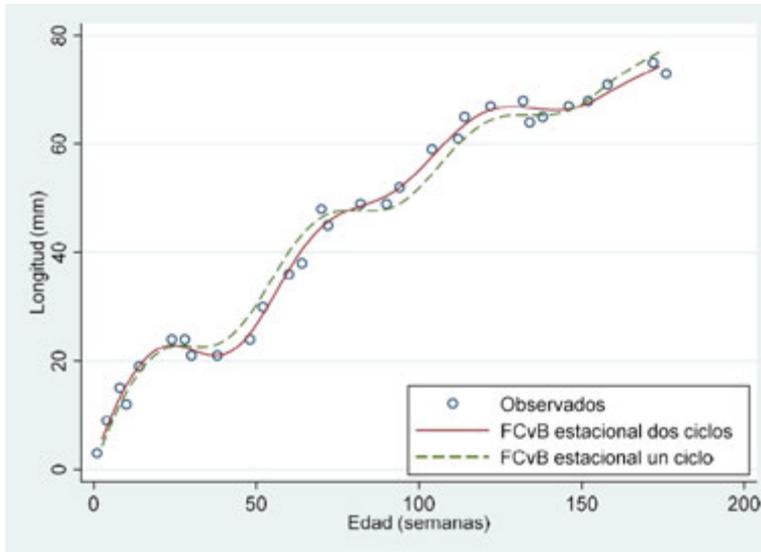


Figura 4.21. Funciones de crecimiento estacional con dos ciclos (continua) y con un ciclo (punteada) para los datos de *Phoxinus phoxinus*.

La curva resultante se presenta en la Figura 4.21. Como era de esperarse, el modelo de crecimiento estacional de von Bertalanffy con dos ciclos ajusta mejor a los puntos observados.

Haddon (2001) destaca que si bien, el agregar términos a la expresión matemática pudiera parecer arbitrario, lo que se busca es encontrar un equilibrio entre la generación de un modelo realista y un ajuste óptimo de los datos. El segundo ciclo sinusoidal podría tener un periodo mayor al estacional anual por lo que sugiere una valiosa dirección para investigación a futuro.

En su libro sobre Biología Pesquera, King (2007) presenta una versión para datos de edad en meses y lo aplica a los datos semanales de la “almeja blanca” (surf clam) *Donax deltoides* de Australia del sur:

$$L_t = L_\infty \left(1 - \exp - \left[C \operatorname{sen} \left(\frac{2\pi(t-s)}{12} \right) + K(t-t_0) \right] \right) \quad 4.17$$

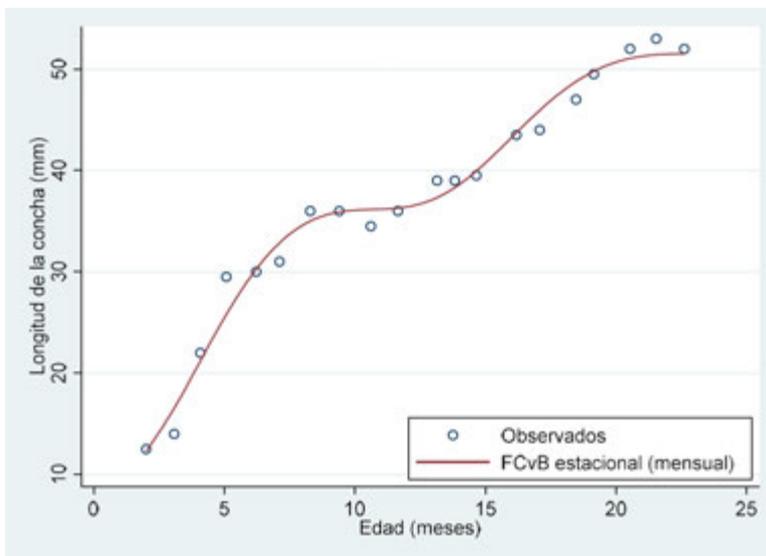


Figura 4.22. Curva de crecimiento estacional de von Bertalanffy para datos modales de frecuencia de tallas de la “almeja blanca” (*Donax deltoides*). La edad se estimó en meses.

Esta expresión se ha incluido en el programa “nlsvbffm.ado” en donde la “m” antes de la extensión nos indica que es la versión para trabajar con edades registradas mensualmente. Como los datos están registrados en semanas, para usar la expresión 4.17 incluida en el citado programa, es necesario convertir las semanas a meses. Los valores iniciales especificados en la matriz “IV” (de “Initial Values”) se escogieron observando la gráfica de los tamaños contra el tiempo y estimándolos de manera razonable con base en ajustes previos.

```
. use seasonking

. gen agemonths = ageweeks/4

. matrix IV = (60,.1,.01,1,1)

. nl svbfffm @ obs1 agemonths, parameters(Li K t0 C s) initial(IV)
```



```
. scatter obs1 agemonths, ms(Oh) || mspline sle agemonths, bands(40) legend(label(1
"Observados") label(2 "FCvB estacional (mensual)") pos(5) ring(0) cols(1))
ytittle("Longitud de la concha (mm)") xtittle("Edad (meses)")
```

Otra de las expresiones utilizadas para la descripción del crecimiento con variación estacional es la de Somers (1988). Este autor mostró que la expresión propuesta por Pitcher y MacDonald (1973)

$$L_t = L_\infty (1 - \exp[-[C \text{sen}(2\pi(t - s)) + K(t - t_0)])] \quad 4.18$$

presenta ciertos inconvenientes (cumplimiento de definiciones de parámetros sólo bajo circunstancias muy estrictas) por lo que para corregirlos propone las siguientes funciones:

$$L(t) = L_\infty (1 - \exp(-K(t - t_0) - S(t) + S(t_0))), \quad 4.19$$

$$\text{con } S(t) = (CK/2\pi) \text{ seno}(2\pi(t - t_s)),$$

$$\text{y } S(t_0) = (CK/2\pi) \text{ seno}(2\pi(t_0 - t_s))$$

y donde C modula la amplitud de las oscilaciones de crecimiento (proporción de disminución en crecimiento a la profundidad de la oscilación en el “invierno”); t_s es el intervalo entre el tiempo 0 y el inicio de la porción convexa de la primera oscilación de crecimiento (punto de inflexión); y el resto de las variables y parámetros como se han definido anteriormente.

Si $C = 0$, entonces no hay oscilación estacional y el modelo se reduce al crecimiento típico de von Bertalanffy. Si $C = 1$, entonces el crecimiento se detiene por completo en el “punto de invierno” ($WP = t_s + 0.5$), mientras que valores $0 < C < 1$ resultan en un crecimiento reducido durante el invierno. Valores de $C > 1$ ($o < 0$) permiten decrementos estacionales en la longitud-a-edad promedio y pueden no ser realistas en organismos cuyos esqueletos no pueden encogerse (Pauly *et al.* 1992) pero que podrían ser el resultado de una mortalidad de invierno dependiente del tamaño (García-Berthou *et al.* 2012).

Siendo consistentes con las expresiones anteriormente presentadas, y bajo las sugerencias de García-Berthou *et al.* (2012) se escribió el programa “nlsvbfgfs.ado” que utiliza un programa evaluador de función y que para su aplicación, requiere de la introducción de valores iniciales en un vector fila. Como ejemplo, se reproduce el análisis de los datos talla-edad de la anchoveta chilena (Cubillos, *et al.* 2001) disponibles en el paquete “FSAdata” para el lenguaje R (Ogle, 2019).

Métodos cuantitativos computarizados para biología pesquera



```
. use anchovechile

. matrix IV = (20, .5, .1, 1, .5)

. nl svbgfs @ tlcmm age, parameters(Li K t0 C ts) initial(IV)
(obs = 207)
```

Iteration 0: residual SS = 194.8895

.

Iteration 9: residual SS = 149.9283

Source	SS	df	MS		
Model	37080.863	5	7416.1727	Number of obs =	207
Residual	149.92828	202	.742219226	R-squared =	0.9960
				Adj R-squared =	0.9959
				Root MSE =	.8615215
Total	37230.792	207	179.858897	Res. dev. =	520.6703

tlcmm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/Li	21.95922	1.005531	21.84	0.000	19.97653	23.9419
/K	.3672645	.0425516	8.63	0.000	.2833622	.4511668
/t0	-.6043762	.056121	-10.77	0.000	-.7150342	-.4937182
/C	.9549112	.1572839	6.07	0.000	.6447823	1.26504
/ts	.2897906	.0261219	11.09	0.000	.2382841	.3412971

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	207	.	-260.3351	5	530.6703	547.3339

Note: BIC uses N = number of observations. See [R] BIC note.

```
. predict tles
```

```
. scatter tlcmm age, ms(Oh) || mspline tles age, bands(40) legend(label(1
"Observados") label(2 "FCvB estacional (Somers)") pos(5) ring(0) cols(1))
ytitle("Longitud total (cm)") xtitle("Edad (años)") lw(.8) yline(21.96)
yscale(range(5 22))
```

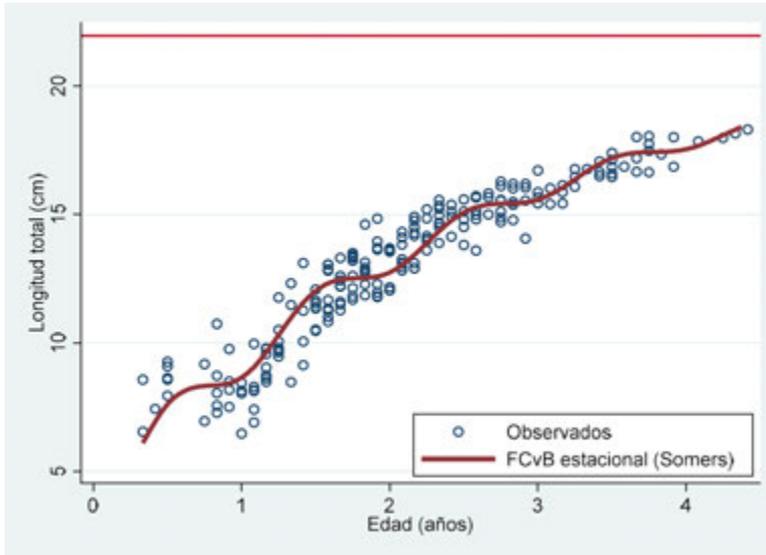


Figura 4.23. Datos talla-edad de la anchoveta chilena y ajuste de la FCvb estacional con la expresión de Somers (1988).

El uso del modelo de crecimiento estacional de von Bertalanffy de acuerdo a la expresión de Somers con el programa R pueden consultarse en la viñeta “Von Bertalanffy Growth Model – Extra” que especifica el uso del paquete “FSA” y sus datos “FSAdata” (Ogle, 2013; Ogle *et al.*, 2020).

4.12 Ajuste de la FCvB por Máxima Verosimilitud

Además de considerar minimizar la suma de las desviaciones al cuadrado respecto de un modelo, existe otro enfoque para estimar sus parámetros, el cual consiste en maximizar la verosimilitud del conjunto de datos al modelo.

El método de Máxima Verosimilitud se basa en la posibilidad de asignar una probabilidad a un conjunto de observaciones (la verosimilitud) en función de un modelo con uno o varios parámetros cuyos valores han de estimarse. Los valores de los parámetros que



maximizan la probabilidad de obtener los datos observados son las estimaciones por máxima verosimilitud (Roff, 2006).

Este enfoque se basa en la probabilidad de la intersección de eventos independientes que se obtiene por la multiplicación de sus probabilidades.

Si consideramos a cada dato como la realización de un experimento y que sus desviaciones de acuerdo a un modelo (con ciertos parámetros) tienen asociada una distribución de probabilidad Gaussiana, entonces, la probabilidad de que un conjunto dado de resultados se obtiene multiplicando las probabilidades individuales. Para la función de crecimiento de von Bertalanffy y la verosimilitud (L del inglés "Likelihood") se obtendría con la siguiente expresión:

$$L = \prod_{i=1}^n \varphi(y_i) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{l_t - \theta_1[1 - e^{-\theta_2(t - \theta_3)}]}{\sigma}\right)^2} \quad 4.20$$

Para simplificar los cálculos, se trabaja con los valores logarítmicos de la verosimilitud. Este cambio de escala no cambia el resultado ya que el punto máximo de la escala logarítmica es exactamente el mismo obtenido con los valores originales. La expresión aplicando logaritmos naturales es:

$$\ln(L) = -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (l_t - \theta_1[1 - e^{-\theta_2(t - \theta_3)}])^2 \quad 4.21$$

Esta expresión no lineal no tiene solución analítica exacta y debe de calcularse por métodos numéricos. Afortunadamente, varios paquetes estadísticos pueden llevar a cabo estos cálculos.

Para ejemplificar el uso de este método se utilizarán los datos talla-edad de hembras de la merluza del Pacífico reportados por Kimura (1980) que se incluyen en el Cuadro 4.18.

En Stata, el ajuste por máxima verosimilitud se lleva a cabo en varios pasos por medio del comando general "ml". Una de las formas de uso consiste en escribir un programa que contenga al modelo y la correspondiente expresión para el logaritmo de la verosimilitud. Este programa es "mlvbgf.ado" el cual se incluye en la declaración del modelo, línea de comando que especifica además los parámetros del modelo de acuerdo a la sintaxis del programa "ml". En la línea siguiente, que es opcional pero conveniente para incrementar la posibilidad de llegar a una solución (conseguir la convergencia) se especifican valores iniciales para cada parámetro (en el orden del comando precedente). La estimación iterativa de los parámetros se inicia con el comando "ml max".



Cuadro 4.18. Longitud media a cada edad de hembras de la Merluza del Pacífico (*Merluccius productus*).

Edad (años)	Longitud media
1	15.4
2	28.03
3.3	41.18
4.3	46.2
5.3	48.23
6.3	50.26
7.3	51.82
8.3	54.27
9.3	56.98
10.3	58.93
11.3	59
12.3	60.91
13.3	61.83

```
. use kimudat

. ml model lf mlvbgf (K: mlf = age, nocons) (b2:) (Li:) (sigma:)

. ml init 1 -.01 50 1, copy

. ml max

initial:      log likelihood = -538.5788
rescale:      log likelihood = -538.5788
rescale eq:   log likelihood = -43.298554
Iteration 0:  log likelihood = -43.298554 (not concave)
Iteration 1:  log likelihood = -34.933709 (not concave)
Iteration 2:  log likelihood = -30.554626 (not concave)
Iteration 3:  log likelihood = -27.866959 (not concave)
Iteration 4:  log likelihood = -26.66943
.
Iteration 8:  log likelihood = -23.616561
Iteration 9:  log likelihood = -23.61656
```



Log likelihood = -23.61656

Number of obs = 13
 Wald chi2(1) = 117.45
 Prob > chi2 = 0.0000

	mlf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
K	age	.2962532	.0273358	10.84	0.000	.242676	.3498305
b2	_cons	-.0169648	.0468736	-0.36	0.717	-.1088354	.0749058
Li	_cons	61.2333	1.129484	54.21	0.000	59.01955	63.44705
sigma	_cons	1.488427	.2919045	5.10	0.000	.9163047	2.060549

Como se utilizó una combinación lineal de los parámetros K y t_0 (el parámetro indicado como $b2$), el valor de t_0 individual se obtiene dividiendo el valor $b2$ entre K : $-0.0169648/0.2962532 = -0.05726453$, y con este valor se general los valores calculados:

```
. di -.0169648/.2962532
-.05726453

. gen mlfe4 = 61.2333*(1 - exp(-.2962532*(age + .05726453)))

. scatter mlf age, ms(Oh) || mspline mlfe4 age, bands(40) yline(61.23)
legend(label(1 "Observados") label(2 "Calculados") p

> os(5) ring(0)) ytitle("Longitud media (cm)") xtitle("Edad (años)") lw(.6)
```

En la Figura 4.24 se presentan los datos observados y el ajuste de la curva de von Bertalanffy. Cabe señalar que maximizar la verosimilitud es equivalente a la minimización de las desviaciones al cuadrado (Kimura, 1980; Roff, 2006). Tan sólo se requiere modificar la función objetivo:

$$\hat{\sigma}^2 = \sum_{i=1}^n (l_t - \hat{\theta}_1 [1 - e^{-\hat{\theta}_2(t - \hat{\theta}_3)}])^2 / (n - 3) \tag{4.22}$$



que pudiéramos explicar como el resultado de:

$$\text{Minimizar } \sum_{i=1}^n (\text{Valor observado} - \text{Valor predicho})^2 \quad 4.23$$

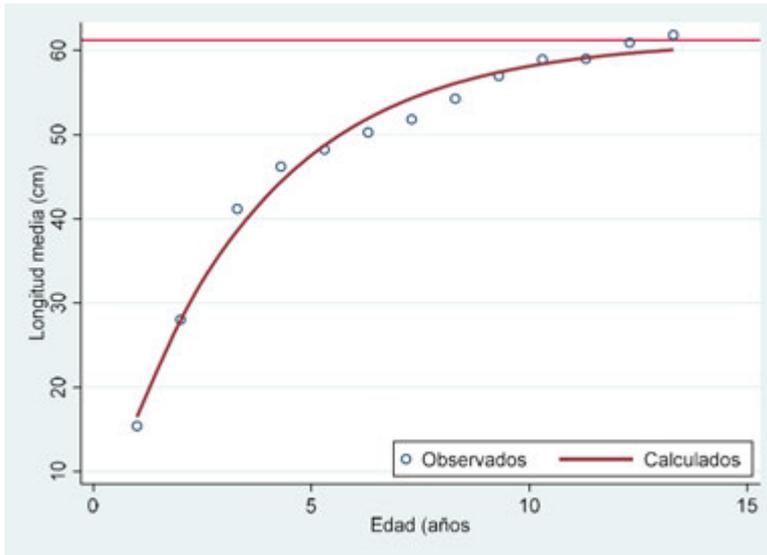


Figura 4.24. Ajuste por máxima verosimilitud de la función de crecimiento de von Bertalanffy para datos talla-edad de la Merluza del Pacífico, *Merluccius productus* (Kimura, 1980). La línea horizontal indica el valor asintótico (L_{∞}).

Este ajuste también puede realizarse con el programa alternativo “mlvbgf2.ado” que utiliza una simplificación lineal de los parámetros K y t_0 diferente a la del programa anterior sugerida por Isabel Canette de Stata Corporation en donde “xb” contiene a K y t_0 . El valor de K aparece como coeficiente de la variable independiente (edad) y t_0 se obtiene dividiendo el valor correspondiente a la constante (ordenada al origen de la combinación lineal) dividida entre la pendiente (K):

Métodos cuantitativos computarizados para biología pesquera



```
. ml model lf mlvbgf2 (xb: mlf = age) (Li:) (sigma:)

. ml init 1 -.01 50 1, copy

. ml max

initial:      log likelihood = -5.054e+14
rescale:      log likelihood = -1652213.9
rescale eq:   log likelihood = -31.456163
Iteration 0:  log likelihood = -31.456163 (not concave)
Iteration 1:  log likelihood = -19.60944 (not concave)
Iteration 2:  log likelihood = -16.702229 (not concave)
Iteration 3:  log likelihood = -12.971348 (not concave)
Iteration 4:  log likelihood = -12.483807
Iteration 5:  log likelihood = -12.334563
Iteration 6:  log likelihood = -11.705492
Iteration 7:  log likelihood = -11.672479
Iteration 8:  log likelihood = -11.67036
Iteration 9:  log likelihood = -11.670359

                                         Number of obs   =          13
                                         Wald chi2(1)     =         117.45
Log likelihood = -11.670359              Prob > chi2      =          0.0000

-----+-----
      mlf |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
xb       |
      age |  -.2962532     .0273358   -10.84  0.000   - .3498304   - .242676
      _cons | -.0169648     .0468736    -0.36  0.717   - .1088354   .0749058
-----+-----
Li       |
      _cons |      61.2333    1.129483    54.21  0.000    59.01955    63.44705
-----+-----
sigma    |
      _cons |      1.488427   .2919045     5.10  0.000    .9163047    2.060549

. di -.0169648/-.2962532
.05726453

. gen mlfe5 = 61.2333*(1 - exp(-.2962532*(age + .05726453)))
```

Los resultados son los mismos que con “mlvbgf.ado”, aclarando que los signos obtenidos para K y t0 son los que deben utilizarse directamente en la ecuación de von Bertalanffy para la generación de los valores calculados (K negativa y t0 positivo).



Estas estimaciones, al estar basadas en la máxima verosimilitud, permiten obtener los valores de los criterios de información, Akaike y Bayesiano, métricas útiles para la selección de modelos alternativos.

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll (null)	ll (model)	df	AIC	BIC
.	13	.	-11.67036	4	31.34072	33.60052

Note: BIC uses N = number of observations. See [R] BIC note.

Capítulo 5

Comparación de funciones de crecimiento



5.1. Introducción

Una vez que una función de crecimiento se ha ajustado de manera satisfactoria a los datos, puede surgir la necesidad de comparar aquellas pertenecientes a diferentes grupos de peces, definidos por sexo, por sitio de muestreo, etc. Es frecuente que machos y hembras de una especie muestren diferencias en el crecimiento o bien que el crecimiento difiera dependiendo de la región geográfica donde los peces fueron capturados. Se han propuesto varios procedimientos estadísticos para comparar los parámetros de las funciones de crecimiento. Debido a la naturaleza no lineal de las funciones de crecimiento, no es posible utilizar directamente métodos estándar como el análisis de covarianza. Algunos procedimientos por ejemplo, linealizan la FCvB para llevar a cabo análisis comparativos univariados para cada parámetro (Misra, 1980); Kimura (1980) propone comparar los parámetros por el método de razón de verosimilitud, mientras que Bernard (1981) propone una prueba multivariada para comparar los tres parámetros a la vez. Cerrato (1990) hace una evaluación de varios métodos de comparación y sus resultados indican que el método de razón de verosimilitud debiera utilizarse. Contribuciones posteriores (Roff, 2006) plantean el uso de pruebas basadas en la máxima verosimilitud.

Esta sección presenta una serie de rutinas para probar diferencias entre los parámetros de dos funciones de crecimiento del tipo de von Bertalanffy. Además, se muestran también versiones modificadas de estos algoritmos para probar la diferencia entre dos funciones Logísticas o de Gompertz.

5.2 Prueba de la T^2 de Hotelling

En Biología Pesquera, el modelo utilizado predominantemente para describir cuantitativamente el crecimiento de los peces es la función de crecimiento de von Bertalanffy (FCvB). Al investigar las poblaciones de peces uno puede terminar con varias expresiones, por ejemplo, una para machos y otra para hembras. Una pregunta interesante es si las diferencias observadas en los parámetros de la expresión son o no diferentes significativamente.



Como afirmó Bernard (1981) para responder a la pregunta anterior, se puede considerar lo siguiente:

Los parámetros L_{∞} (longitud asintótica), K (coeficiente de crecimiento), y t_0 (tiempo al cual la longitud es cero) de la FCVB están por lo general, correlacionados y no debieran ser estimados independientemente. Cuando se comparan dos grupos de peces, las pruebas estadísticas univariadas no son apropiadas debido a las covarianzas de los parámetros cuando estos últimos están correlacionados significativamente. Cuando las correlaciones entre los valores de los parámetros son significativas, puede utilizarse el procedimiento multivariado de la T^2 de Hotelling para probar las diferencias entre dos grupos. En este método, todas las diferencias entre los valores paramétricos se prueban simultáneamente considerando las covarianzas entre los parámetros.

Esta prueba multivariada se lleva a cabo calculando las expresiones incluidas a continuación. Cabe aclarar que al elaborar los programas se notaron (y corrigieron) varios errores tipográficos en las ecuaciones impresas. Estas imprecisiones fueron notadas antes por otros autores y con algunos comentarios adicionales se publicaron posteriormente (Headrick y Margraf, 1982; Misra, 1982) y respondidas por el autor mismo (Bernard, 1982).

El estadístico T^2 de Hotelling se basa en los vectores de los valores paramétricos y las matrices de varianza-covarianza de la regresión suponiendo una distribución de probabilidad multivariada conjunta. Los parámetros de la FCvB L_{∞} , K , t_0 , y las diferencias correspondientes de los valores para los dos grupos de peces se expresan como vectores columna.

$$\vec{\mathbf{P}}_1 = \begin{bmatrix} L_{\infty(1)} \\ K_{(1)} \\ t_{0(1)} \end{bmatrix}, \vec{\mathbf{P}}_2 = \begin{bmatrix} L_{\infty(2)} \\ K_{(2)} \\ t_{0(2)} \end{bmatrix}, \text{ y } [\vec{\mathbf{P}}_1 - \vec{\mathbf{P}}_2] = \begin{bmatrix} L_{\infty(1)} - L_{\infty(2)} \\ K_{(1)} - K_{(2)} \\ t_{0(1)} - t_{0(2)} \end{bmatrix} \quad (5.3a)$$

ó de manera equivalente, para la función Logística, longitud asintótica (L_{∞}), tasa instantánea de crecimiento (g) y tiempo cuando la tasa de crecimiento es máxima (t_0), y para la función de Gompertz, longitud asintótica (L_{∞}), tasa instantánea de crecimiento (g) y tiempo en el punto de inflexión (t_0):

$$\vec{\mathbf{P}}_1 = \begin{bmatrix} L_{\infty(1)} \\ g_{(1)} \\ t_{0(1)} \end{bmatrix}, \vec{\mathbf{P}}_2 = \begin{bmatrix} L_{\infty(2)} \\ g_{(2)} \\ t_{0(2)} \end{bmatrix}, \text{ y } [\vec{\mathbf{P}}_1 - \vec{\mathbf{P}}_2] = \begin{bmatrix} L_{\infty(1)} - L_{\infty(2)} \\ g_{(1)} - g_{(2)} \\ t_{0(1)} - t_{0(2)} \end{bmatrix} \quad (5.3b)$$



\vec{P}_1 y \vec{P}_2 son vectores columna de los valores paramétricos estimados a partir de medidas de edad y longitud en dos grupos de peces, y $[\vec{P}_1 - \vec{P}_2]$ es un vector columna de las diferencias entre las estimadas del mismo parámetro.

El estadístico T^2 se calcula como el producto de las diferencias en los parámetros $[\vec{P}_1 - \vec{P}_2]$ elevada al cuadrado y la inversa de la matriz de varianza-covarianza conjunta.

$$T^2 = \frac{N_1 N_2}{N_1 + N_2} [\vec{P}_1 - \vec{P}_2]' \mathbf{S}^{-1} [\vec{P}_1 - \vec{P}_2] \quad (5.4)$$

donde N_1 y N_2 son el número de peces en cada grupo,

$[\vec{P}_1 - \vec{P}_2]'$ es la traspuesta (vector fila) de $[\vec{P}_1 - \vec{P}_2]$, y

\mathbf{S}^{-1} es la inversa de la matriz conjunta de varianza-covarianza,

$$\mathbf{S} = \frac{(N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2}{N_1 + N_2 - 2} \quad (5.5)$$

donde \mathbf{S}_1 y \mathbf{S}_2 son las matrices de varianza-covarianza estimadas de L_∞ , K (ó g), y t_0 de cada uno de los dos grupos de peces.

\mathbf{S}_i y \vec{P}_i pueden estimarse juntos cuando los datos de longitud – edad se han ajustado a la FCvB utilizando procedimientos numéricos (regresión no lineal). El valor calculado del estadístico T^2 se utiliza para probar la hipótesis nula.

H_0 : $\vec{P}_1 - \vec{P}_2$ contra la hipótesis alternativa

H_A : $\vec{P}_1 \neq \vec{P}_2$

Debido a que la T^2 es la generalización multivariada de la t -Student univariada, el valor de tablas del estadístico F (con algunas modificaciones) se utiliza para probar su significancia de acuerdo a la siguiente equivalencia:

$$T_0^2 = \frac{3(N_1 + N_2 - 2)}{N_1 + N_2 - 4} F_{\alpha; 3, N_1 + N_2 - 4} \quad (5.6)$$



donde F es obtenida de una tabla estándar (o calculada con la fórmula o procedimiento correspondiente incluidas en hojas de cálculo o paquetes estadísticos) y tiene 3 con $N_1 + N_2 - 4$ grados de libertad.

Cuando el estadístico T^2 es significativo (se rechaza H_0), es posible calcular intervalos de confianza simultáneos de Roy-Bose alrededor de las diferencias de los valores de los parámetros para encontrar cuáles contribuyeron en mayor proporción para la significancia. El intervalo es simétrico alrededor de cada diferencia (Morrison, 1978).

$$(p_{1j} + p_{2j}) \pm \sqrt{\frac{N_1 + N_2}{N_1 N_2} s_{jj}^2 T_0^2} \quad (5.7)$$

Para evaluar la significancia relativa entre los parámetros se obtienen los valores de F necesarios para causar que los intervalos de confianza contengan al cero por medio de un re-arreglo de las ecuaciones 5.6 y 5.7:

$$F_0 = \frac{N_1 N_2 (N_1 + N_2 - 4) (p_{1j} - p_{2j})^2}{3 (N_1 + N_2) (N_1 + N_2 - 2) s_{jj}^2} \quad (5.8)$$

donde F_0 es el valor crítico de F . Los cambios en los valores de F_0 se deben a cambios en α ya que los grados de libertad permanecen constantes. Las diferencias en crecimiento entre los dos grupos de peces se hacen manifiestos por altos valores de F_0 para cada parámetro; mientras más alto, mayor será la significancia en la explicación de la diferencia.

5.3. Suposiciones de la prueba de T^2

Para aplicar la prueba multivariada en la comparación de funciones de crecimiento, ya sea de von Bertalanffy (L_{∞} , K and t_0), Logística ó de Gompertz (L_{∞} , g , t_0), se supone que ambos conjuntos de estimadas se obtienen de dos distribuciones conjuntas trivariadas normales (gaussianas) con una estructura de varianza común. De esta forma, es válida la estimación de una varianza conjunta (S_1 y S_2) al desconocer la verdadera matriz de varianza-covarianza. Con matrices de varianza-covarianza desiguales, la probabilidad de un error de Tipo I (α) y el poder de la prueba de T^2 se desvía de los valores en la tabla con los mismos grados de libertad. Sin embargo, cuando los tamaños de las muestras son iguales, las diferencias en las matrices de varianza-covarianza no afectan el nivel de error o el poder de la prueba.



La normalidad de una distribución multivariada puede ser probada indirectamente por las correspondientes probabilidades marginales univariadas (debido a que las distribuciones marginales normales a menudo indican una distribución conjunta también normal). Si las pruebas estándar de bondad de ajuste no indican distribuciones marginales normales, será posible re-exresar los datos para lograr distribuciones más simétricas. En ciertos casos, aún con distribuciones marginales normales, la distribución conjunta multivariada puede apartarse de la normalidad (gaussianidad). En tales circunstancias, el sesgo en la T^2 de Hotteling puede minimizarse usando números grandes de observaciones. Ya que los estimadores son medias en la distribución trivariada normal, sus distribuciones son aproximadamente normales y se hará mejor si aumenta el tamaño muestral (N_i). De manera optimista, esta robustez significa que la prueba de T^2 de Hotteling puede usarse para evaluar diferencias en crecimiento sobre una gama amplia de distribuciones (Bernard, 1981). No obstante, Misra (1982) hizo anotaciones sobre estas expectativas optimistas (las cuales fueron contestadas por Bernard en el mismo comentario) y posteriormente Cerrato (1990) llevó a cabo una revisión crítica de este procedimiento. Ejemplos adicionales de la aplicación de esta prueba se reportan en los trabajos de Ibañez-Aguirre *et al.* (1999) y Cruz-Martínez (2000).

5.4 Pruebas de razón de verosimilitud

Otro enfoque para analizar las diferencias en crecimiento son las pruebas de razón de verosimilitud (PRV). Estos procedimientos establecen el resultado de la hipótesis de dos o más curvas independientes como un caso base contra el cuál comparar todas las hipótesis alternativas. Para las funciones de crecimiento (FCvB, Gompertz ó Logística) las hipótesis alternativas serán: a) coincidencia de todas las curvas, b) valores iguales de L_∞ , c) valores iguales de K ó g , d) valores iguales de t_{0r} , e) valores iguales de L_∞ y K ó g , f) valores iguales de L_∞ y t_{0r} y g) valores iguales de K ó g y t_{0r} . Si la hipótesis a) (coincidencia de curvas) no se rechaza, no es necesario proceder con las siguientes hipótesis alternativas. Si a) es rechazada, será necesario entonces llevar a cabo comparaciones por pares para encontrar cuales son los parámetros de las curvas que difieren (Kimura, 1980; Haddon, 2001).

Los parámetros de los modelos de crecimiento pueden estimarse mejor usando métodos de máxima verosimilitud. Estas estimaciones son deseables porque son consistente, asintóticamente normales y asintóticamente alcanzan el valor menor posible de la varianza. Para la curva de von Bertalanffy (y otros modelos no lineales) sólo pueden encontrarse estimaciones por máxima verosimilitud apropiadas con algoritmos iterativos. Existen varios algoritmos especializados, pero la forma más fácil de calcularlos es por el uso



alguno de los programas disponibles para ajuste no lineal por mínimos cuadrados. Debe recordarse que las soluciones por mínimos cuadrados obtenidas iterativamente pueden ser minimizaciones de $S(L_{\infty}, K, t_0)$ locales en lugar de globales, así que será necesario proporcionar varias combinaciones de valores iniciales y de preferencia obtenidos de la mejor información disponible (Kimura, 1980).

Para comparar dos curvas de crecimiento, puede utilizarse el método de razón de verosimilitud (Kimura, 1980) que es simple. En primer lugar, encontrar la mejor curva para cada conjunto de datos. Luego, comparar la suma total de residuos de este caso base con la obtenida al considerar una sola ecuación ajustada a los dos grupos de datos. Si existe diferencia significativa, se requerirá el ajuste de funciones con una o varias restricciones, como por ejemplo, dos ecuaciones con la misma L_{∞} y diferentes K s y t_0 s.

La prueba calcula un estadístico que es comparado con la distribución χ^2 con grados de libertad iguales al número de restricciones, (i.e. el número de parámetros supuesto igual) de acuerdo a la siguiente ecuación:

$$\chi_k^2 = -N \times \ln \left(\frac{SCR_{\Omega}}{SCR_{\omega}} \right) \quad (5.9)$$

donde k son los grados de libertad, N es el número total de observaciones combinado de las dos curvas, SCR_{Ω} es la suma total del cuadrado de los residuos obtenida de ajustar ambas curvas por separado y SCR_{ω} es la suma total de residuos cuadrados derivados de ajustar a las curvas con alguna de las restricciones supuestas (K s iguales, L_{∞} s iguales, t_0 s iguales). La suma de residuales cuadrados más pequeña siempre será la provista por el caso base (SCR_{Ω}) al tener el valor más grande de parámetros independientes (supuestos).

En este capítulo se presentan cuatro programas sencillos para llevar a cabo ajustes por mínimos cuadrados de regresión lineal de la FCVB con el fin de realizar comparaciones de sus parámetros por razón de verosimilitud. Estos programas se emplearon como base para implementar los programas con las restricciones requeridas con otras funciones de crecimiento como la Logística y la de Gompertz. Los programas calculan las estimaciones de los parámetros y las sumas de residuos cuadrados para el caso base, L_{∞} s, K s (o g s) y t_0 s iguales como las presenta Kimura (1980) y Haddon (2001).



5.5 Prueba global de curvas de crecimiento

En 1992, Chen y colaboradores presentaron un método para comparar simultáneamente varias funciones de crecimiento al que denominaron Análisis de la Suma de Cuadrados Residual (ASCR). Es diferente a las pruebas de razón de verosimilitud aunque es análogo al utilizar la suma de cuadrados residual (Haddon, 2001). La prueba lleva a cabo una comparación global para verificar si dos o más curvas son estadísticamente diferentes, es decir, la hipótesis nula es que las curvas son coincidentes. Para calcular el ASCR se requieren 4 pasos:

- 1) Para cada conjunto i de datos, ajustar una curva y calcular la suma de cuadrados residuales, SCR_i y los grados de libertad asociados Gl_i .
- 2) Los SCR_i y Gl_i resultantes para cada curva se suman para dar $\sum SCR_i$ y $\sum DF_i$.
- 3) Se conjuntan los datos de todas las curvas, se ajusta una nueva curva a los datos combinados y se calculan la SCR_p y los Gl_p conjuntos.
- 4) Se calcula un valor del estadístico F de acuerdo a la siguiente expresión:

$$F = \frac{\frac{SCR_p - \sum SCR_i}{Gl_p - \sum Gl_i}}{\frac{\sum SCR_i}{\sum Gl_i}} = \frac{SCR_p - \sum SCR_i}{3(K-1)} \cdot \frac{\sum Gl_i}{\sum SCR_i} \cdot \frac{N-3K}{N-3K} \quad (5.10)$$

Donde el estadístico F tiene $3(K-1)$ y $(N-3K)$ grados de libertad, siendo K el número de curvas a compararse y N el tamaño de muestra combinado.

5.6 Pruebas de máxima verosimilitud

Con base en la máxima verosimilitud es posible llevar a cabo diversas pruebas de hipótesis (Roff, 2006) como la comparación de modelos con la misma estructura pero con diferencia en el número de parámetros. En el caso de la función de von Bertalanffy, pudiéramos querer comparar un modelo con θ_3 con otro sin θ_3 .



$$l_t = \theta_1 (1 - e^{-\theta_2(t - \theta_3)}) \text{ vs. } l_t = \theta_1 (1 - e^{-\theta_2 t}) \quad (5.11)$$

Para hacerlo se utiliza la distribución aproximada como ji-cuadrada de la desvianza. Para estos modelos las desvianzas son:

$$D_{n-3} = \frac{1}{\sigma^2} \sum_{t=1}^n (l_t - \hat{\theta}_{F,2} [1 - e^{-\hat{\theta}_{F,2}(t - \hat{\theta}_{F,3})}])^2 \sim \chi_{n-3}^2 \quad (5.12)$$

$$D_{n-2} = \frac{1}{\sigma^2} \sum_{t=1}^n (l_t - \hat{\theta}_{R,2} [1 - e^{-\hat{\theta}_{R,2} t}])^2 \sim \chi_{n-2}^2 \quad (5.13)$$

donde los subíndices F y R significan “Full” (Completo en inglés) y “Reducido” respecto a los modelos. Por la naturaleza aditiva de la ji-cuadrada, tenemos que $D_{n-2} - D_{n-3} \sim \chi_1^2$. Pero aún tenemos el problema de la inconveniente σ^2 . Para eliminar este parámetro y producir un estadístico F construimos el siguiente cociente

$$\frac{D_{n-2} - D_{n-3}}{D_{n-3} / (n - 3)} \sim F_{1, n-3} \quad (5.12)$$

Si las estimadas por máxima verosimilitud se encuentran por mínimos cuadrados, es posible utilizar la siguiente fórmula general para comparar dos modelos:

$$\frac{(SCR_R - SCR_F) / (F - R)}{SCR_F / (n - F)} \sim F_{F-R, n-F} \quad (5.12)$$

donde SCR_F es la suma de cuadrados del modelo completo (“Full”) con F parámetros, SCR_R es la suma de cuadrados del modelo reducido con R parámetros ($F > R$) y n es el tamaño total de la muestra.

Con este procedimiento se pueden comparar dos curvas de crecimiento, por ejemplo de von Bertalanffy correspondientes a machos y a hembras. Puede haber diferencia en los tres parámetros de la función o sólo en uno. La prueba general se efectúa ajustando por separado las curvas de cada población y después con los datos en conjunto. La comparación de cada parámetro se hace con pruebas específicas a posteriori. Todo lo anterior es posible hacerlo utilizando variables binarias indicadoras (dummy). Más adelante se dan instrucciones para llevar a cabo el análisis con instrucciones para el paquete estadístico Stata. Estas rutinas son equivalentes a los listados de instrucciones para el programa *S-plus* en los apéndices de la obra de Roff (2006).



5.7 Programas

El programa para efectuar la prueba de T^2 basada en Bernard (1981) usa el lenguaje de matrices de Stata y es muy sencillo. Los programas para llevar a cabo las pruebas de razón de verosimilitud usan la herramienta de regresión no lineal disponible en el Stata para las funciones Logística y Gompertz en sus versiones de tres parámetros y el programa de Salgado-Ugarte, *et al.* (2000) para la función de crecimiento de von Bertalanffy.

Utilizando la información de la Tabla 1 de Bernard (1981) con longitud a edad de machos y hembras de la perca oceánica del Pacífico *Sebastes alutus* (gallineta del Pacífico) capturados en 1977 se introducen las matrices y se utiliza el comando (hotegfun.ado):

```
. matrix Sm = (294.7427,-0.6018,-32.5704\-.6018,.0013,.073\
32.5704,.073,4.2296)

. matrix Sf = (1596.1804,-1.1895,-91.0468\ -1.1895,.0009,.0713\
91.0468,.0713,5.7613)

. matrix Pm=(441.16,.13,-3.36)

. matrix Pf=(505.97,.09,-4.57)

. hotegfun Sm Sf Pm Pf 78 78 99
```

Los resultados se presentan en el Cuadro 5.1:

Los valores numéricos obtenidos con nuestros programas difieren de los publicados por lo que consideramos como redondeo en las operaciones con matrices y errores de imprenta (que se han detectado en varias partes del impreso original). La conclusión de la prueba es que machos y hembras crecen de manera diferente ($T^2_{calculada} = 865.18 \gg T^2_{tablas} = 11.89$). Tanto L_{∞} como K son diferentes (los intervalos de confianza no incluyen al cero) lo que indica que las hembras alcanzan tallas mayores y crecen más lentamente que los machos y sólo t_0 pudiera tener el mismo valor (los intervalos de confianza incluyen al cero) en ambos sexos. Los valores críticos de F indican que el parámetro más significativo que contribuye a la diferencia es L_{∞} seguida por K .



Cuadro 5.1. Prueba de T^2 de Hotelling para comparación de parámetros de la FCVB de dos poblaciones (Bernard, 1981).

Matriz de varianza-covarianza (S)			Inversa de S		
945.46165	-0.89565	-61.8086	0.00553845	0.2810506	0.06446791
	0.0011	0.07215		17277.305	-246.06116
		4.99545			4.5517389
[-64.8100 0.0400 1.2100] = [P ₁ - P ₂]'					
$T^2 = 865.1831$ $T^2_{0.01; 3, 152} = 11.8937$ $F_{0.01; 3, 152} = 3.9131$					
				F crítica	
-81.7904	$\leq L_{\infty 1} - L_{\infty 2} \leq$	-47.8296		57.0041	
0.0217	$\leq K_1 - K_2 \leq$	0.0583		18.6635	
-0.0243	$\leq t_{01} - t_{02} \leq$	2.4443		3.7606	

Para ejemplificar la prueba de razón de verosimilitud se utilizarán los datos de la Tabla 1 de Kimura (1980) que muestra la longitud promedio a varias edades de machos (1) y hembras (2) de la merluza nortea *Merluccius productus* (en inglés Pacific hake) capturados en las aguas costeras de California, Oregon, y Washington incluidas en el Cuadro 5.2 y en la Figura 5.1. Los pasos para realizar estas pruebas son los siguientes:

- Se estima el caso base sin restricciones, esto es, por sexos separados (programa `nlvbgfno.ado`)
- Luego se usa una sola ecuación para describir el crecimiento de los peces, ignorando el sexo (rutina `nlvbgf.ado`).
- Se calcula el valor de χ^2 correspondiente a la prueba de razón de verosimilitud y se determina su significancia.
- Si la prueba indica diferencias significativas por sexo, se procede a estimar las diferencias por parámetro.
- Primero, se estiman las funciones de crecimiento con L_{∞} s iguales (programa `nlvbgfel.ado`) y los residuos cuadrados se comparan con los del caso base para obtener el valor del estadígrafo χ^2 .



- Se estima el crecimiento con K_s iguales (programa `nlvbgfek.ado`) y se contrasta con el caso base para determinar el valor de χ^2 .
- Se estima el crecimiento con t_0 s iguales (programa `nlvbgfet.ado`) y se obtiene el valor de χ^2 .

Cuadro 5.2. Datos de longitud – edad para la merluza norteña *Merluccius productus* (adaptados de Kimura, 1980).

Edad (años)	Frecuencia	Longitud media	Sexo
1	385	15.4	2
2	36	28.03	2
3.3	17	41.18	2
4.3	135	46.2	2
5.3	750	48.23	2
6.3	1073	50.26	2
7.3	1459	51.82	2
8.3	626	54.27	2
9.3	199	56.98	2
10.3	97	58.93	2
11.3	44	59	2
12.3	11	60.91	2
13.3	6	61.83	2
1	385	15.4	1
2	28	26.93	1
3.3	13	42.23	1
4.3	83	44.59	1
5.3	628	47.63	1
6.3	1134	49.67	1
7.3	1761	50.87	1
8.3	432	52.3	1
9.3	93	54.77	1
10.3	21	56.43	1
11.3	8	55.88	1



```
. use kimudat
```

```
. scatter ml age if sex==1, ms(Sh) || mspline mlem age, bands(40) || scatter ml
age if sex==2, ms(Oh) || mspline mleh age, bands(40) legend(label(1 "Machos")
label(3 "Hembras") label(2 "FCvB machos") label(4 "FCvB hembras") pos(5)
ring(0)) ytitle("Longitud del cuerpo (cm)") xtitle("Edad (años)")
```

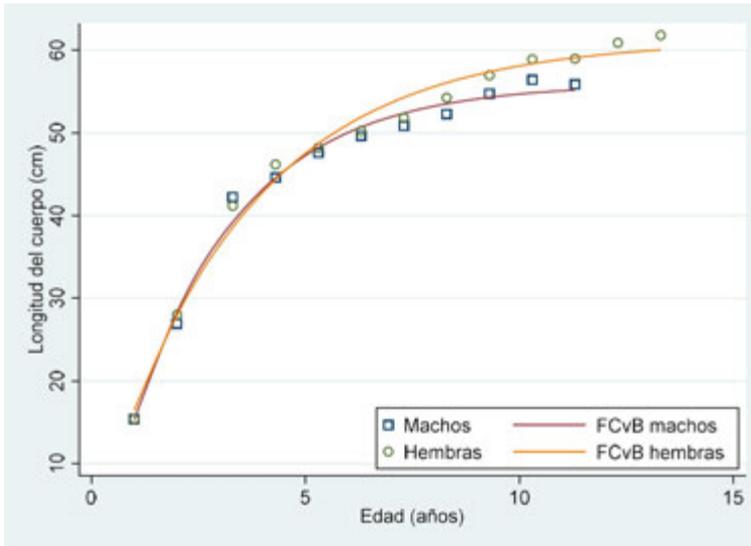


Figura 5.1. Datos de longitud-edad de la merluza del Pacífico (Kimura, 1980)

Los resultados de los pasos anteriores se resumen en el Cuadro 5.3. Como en Kimura (1980), las pruebas de razón de verosimilitud indican que existe una diferencia significativa ($P = 0.007$) en los modelos por sexo. Por tanto se procede a efectuar las pruebas siguientes de las cuales se tuvo:

- diferencia significativa ($P = 0.002$) en L_{∞} entre sexos
- diferencia en el límite de significancia ($P = 0.05$) en las Ks
- diferencias no significativas para t_0 ($P = 0.267$).



Los pasos anteriores se han incluido en un programa que integra las cinco rutinas involucradas en los cálculos de cada uno de los pasos anteriores: vblrtest.ado (por “von Bertalanffy likelihood ratio test”).

Cuadro 5.3. Pruebas de razón de verosimilitud para la FCVB para machos (1) y hembras (2). Datos usados por Kimura (1980).

Restricciones lineales	Modelos equivalentes	SCR	χ^2	gl	P
Ninguna	$l_{1j} = 55.98 [1 - \exp\{-0.386(t_{1j} - 0.171)\}]$ $l_{2j} = 61.23 [1 - \exp\{-0.296(t_{2j} + 0.057)\}]$	48.22			
$L_{\infty 1} = L_{\infty 2}$	$l_{1j} = 59.40 [1 - \exp\{-0.297(t_{1j} + 0.111)\}]$ $l_{2j} = 59.40 [1 - \exp\{-0.337(t_{2j} - 0.087)\}]$	71.60	9.49	1	0.002
$K_1 = K_2$	$l_{1j} = 57.43 [1 - \exp\{-0.330(t_{1j} + 0.021)\}]$ $l_{2j} = 60.14 [1 - \exp\{-0.330(t_{2j} - 0.095)\}]$	56.34	3.74	1	0.053
$t_{01} = t_{02}$	$l_{1j} = 56.45 [1 - \exp\{-0.361(t_{1j} - 0.057)\}]$ $l_{2j} = 60.77 [1 - \exp\{-0.313(t_{2j} - 0.057)\}]$	50.76	1.23	1	0.267
$L_{\infty 1} = L_{\infty 2}$ $K_1 = K_2$ $t_{01} = t_{02}$	$l_j = 59.29 [1 - \exp\{-0.320(t_j - 0.010)\}]$	79.76	12.08	3	0.007

```
. vblrtest ml age sex
```

Likelihood ratio test for two von Bertalanffy Growth Functions

Linear constraints	Models	RSS	PLR	DF	P
none	l_t1= 55.98[1-exp{- 0.386(t-0.171)}] l_t2= 61.23[1-exp{- 0.296(t+0.057)}]	48.22			
Li1 = Li2	l_tt= 59.29[1-exp{- 0.320(t-0.010)}]	79.76	12.08	3	0.007



K1 = K2										
t01 = t02										

Li1 = Li2		l_t1= 59.40 [1-exp{- 0.297 (t+0.111) }]		71.60		9.49		1		0.002
		l_t2= 59.40 [1-exp{- 0.337 (t-0.087) }]								

K1 = K2		l_t1= 57.43 [1-exp{- 0.330 (t+0.021) }]		56.34		3.73		1		0.053
		l_t2= 60.14 [1-exp{- 0.330 (t-0.095) }]								

t01 = t02		l_t1= 56.45 [1-exp{- 0.361 (t-0.057) }]		50.76		1.23		1		0.268
		l_t2= 60.77 [1-exp{- 0.313 (t-0.057) }]								

El cuadro resultante construido por el programa “vblrtest.ado” muestra primero la prueba de diferencia entre grupos ya que esta define si el procedimiento continúa o no con las pruebas para la comparación de cada parámetro.

Es posible aplicar a este conjunto de datos la prueba de Hotelling. Primeramente se hace el ajuste no lineal (se omiten los resultados para no ser repetitivos) para los machos y se recuperan matriz de varianza covarianza (Sm) y vector de parámetros (Pm); se repite la operación pero ahora para las hembras (Sh y Ph). Con estas matrices y vectores se utiliza el programa hotegfun.ado para obtener los valores del Cuadro 5.4:

```
. nl vbgf ml age if sex==1
(obs = 11)

. matrix Sm = e(V)

. matrix Pm = e(b)

. nl vbgf ml age if sex==2
(obs = 13)

. matrix Sh = e(V)

. matrix Ph = e(b)

. hotegfun Sm Sh Pm Ph 11 13 99
```

```
=====
Multivariate test for Growth Function comparison
from two populations (based on Bernard, 1981)
```



```

Matrix S                               Matrix S inverse
| 1.337492 -0.03171492 -0.09877241 | | 2.41400934 85.961766 -5.0373613 |
|           .00114213  .00429182 | |           5380.74578 -563.687203 |
|           .02590538 | |           112.78315 |

-----

| -5.2553  0.0893  0.2286 | = [P1 - P2]'

-----

T^2 = 142.2401  T^2_0.01: 3,20 = 16.2960  F_0.01: 3,20 = 4.9382

-----

Confidence intervals of 99 %                                Critical F
-7.1679 <= L_inf1 - L_inf2 <= -3.3427                      37.2832
0.0334 <= K1 - K2 <= 0.1452                                12.6154
-0.0376 <= t_o1 - t_o2 <= 0.4948                          3.6424

=====
    
```

Cuadro 5.4. Prueba de T^2 de Hotelling para comparación de parámetros de la FCVB de machos y hembras (Kimura, 1980).

Matriz de varianza-covarianza (S)			Inversa de S		
1.337492	-0.031715	-0.0987724	2.41400933	85.961766	-5.0373613
	0.001142	0.00429182		5380.7458	-563.68721
		0.02590539			112.783151
[-5.2553 0.0893 0.2286] = [P ₁ - P ₂]'					
$T^2 = 142.2401$ $T^2_{0.01: 3,20} = 16.2960$ $F_{0.01: 3,20} = 4.9382$					
				F crítica	
-7.1679	$\leq L_{\infty 1} - L_{\infty 2} \leq$	-3.3427		37.2832	
0.0334	$\leq K_1 - K_2 \leq$	0.1452		12.6154	
-0.0376	$\leq t_{01} - t_{02} \leq$	0.4948		3.6424	



Los resultados son muy semejantes a los obtenidos con las pruebas de razón de verosimilitud: hay una diferencia significativa entre sexos originada principalmente por los valores de L_{∞} , seguidos por el parámetro de curvatura K y sin diferencia entre las estimadas de t_0 .

El método de Chen *et al.* (1992) aplicado a los datos de Kimura (1980) involucra el ajuste de la función de von Bertalanffy para cada sexo y para la muestra combinada. Sustituyendo en la ecuación los valores de sumas de cuadrados residuales obtenidas de los programas anteriores se tiene:

$$F_c = \frac{\frac{79.7645 - (19.4234 + 28.8004)}{3(2 - 1)}}{\frac{(19.4234 + 28.8004)}{24 - 3(2)}} = 3.9243$$

con 3 y 18 grados de libertad, lo que resulta en un valor de $P = 0.0256$ que indica una diferencia entre las curvas por sexo.

Es interesante notar que el método de Chen *et al.* (1992) proporciona los mismos resultados que la prueba F de Roff (2006) cuando se comparan dos curvas. Es importante recordar que la prueba del Análisis de la Suma de Cuadrados Residual permite la comparación global de dos o más curvas. Sustituyendo los valores obtenidos para los mismos datos (Kimura, 1980) se tiene:

$$F_c = \frac{(79.76455 - 48.2238)/(6 - 3)}{48.2238/(24 - 6)} = 3.9243$$

el cual con 3 y 24-6 (18) grados de libertad resulta en un valor de $P = 0.0256$. Todo lo anterior indica que debe preferirse el ajuste de curvas de crecimiento por sexos separados. Debe notarse que esta prueba proporciona exactamente los mismos resultados que la propuesta de Chen (1992) para dos funciones.

Como se deriva del planteamiento utilizado por Roff (2006) es posible comparar la influencia de los diferentes parámetros en la diferencia general de las curvas de machos y hembras:

Longitudes asintóticas (L_{∞} s):

$$F_c = \frac{(71.60215 - 48.2238)/(6 - 5)}{48.2238/(24 - 6)} = 8.7262, P = 0.0002415$$



Coefficientes de crecimiento (Ks):

$$F_c = \frac{(56.33682 - 48.2238)/(6 - 5)}{48.2238/(24 - 6)} = 3.0283, P = 0.0372$$

Edad teórica a tamaño cero (t_0 s):

$$F_c = \frac{(50.7578 - 48.2238)/(6 - 5)}{48.2238/(24 - 6)} = 0.9458, P = 0.47566$$

Estos pasos también se han integrado en una sola rutina “vbmltest.ado” por “von Bertalanffy maximum likelihood test”:

```
. vbmltest ml age sex
```

Maximum likelihood F tests for two von Bertalanffy Growth Functions					
Linear constraints	Models	RSS	F	DF	P
none	l_t1= 55.98[1-exp{- 0.386(t-0.171)}] l_t2= 61.23[1-exp{- 0.296(t+0.057)}]	48.22			
Li1 = Li2 K1 = K2 t01 = t02	l_tt= 59.29[1-exp{- 0.320(t-0.010)}]	79.76	3.92	3 18	0.0256
Li1 = Li2	l_t1= 59.40[1-exp{- 0.297(t+0.111)}] l_t2= 59.40[1-exp{- 0.337(t-0.087)}]	71.60	8.73	1 18	0.0085
K1 = K2	l_t1= 57.43[1-exp{- 0.330(t+0.021)}] l_t2= 60.14[1-exp{- 0.330(t-0.095)}]	56.34	3.03	1 18	0.099
t01 = t02	l_t1= 56.45[1-exp{- 0.361(t-0.057)}] l_t2= 60.77[1-exp{- 0.313(t-0.057)}]	50.76	0.95	1 18	0.344

Los resultados se resumen en el Cuadro 5.5. Los valores obtenidos para la comparación de los parámetros de la función de von Bertalanffy coinciden con los resultados de las pruebas de Hotelling (Bernard, 1981) y de razón de verosimilitud (Kimura, 1980), marcando la influencia principal para la diferencia por sexos de la longitud asintótica, en menor grado del parámetro de curvatura y la no diferencia en el tiempo cero.



Cuadro 5.5. Pruebas de F de Roff (2006) de verosimilitud para la FCvB para machos (1) y hembras (2). Datos usados por Kimura (1980).

Restricciones lineales	Modelos equivalentes	SCR	F	gl	P
Ninguna	$l_{1j} = 55.98 [1 - \exp\{-0.386(t_{1j} - 0.171)\}]$ $l_{2j} = 61.23 [1 - \exp\{-0.296(t_{2j} + 0.057)\}]$	48.22			
$L_{\infty 1} = L_{\infty 2}$ $K_1 = K_2$ $t_{01} = t_{02}$	$l_j = 59.29 [1 - \exp\{-0.320(t_j - 0.010)\}]$	79.76	3.92	3, 18	0.0256
$L_{\infty 1} = L_{\infty 2}$	$l_{1j} = 59.40 [1 - \exp\{-0.297(t_{1j} + 0.111)\}]$ $l_{2j} = 59.40 [1 - \exp\{-0.337(t_{2j} - 0.087)\}]$	71.60	8.73	5, 18	0.0002
$K_1 = K_2$	$l_{1j} = 57.43 [1 - \exp\{-0.330(t_{1j} + 0.021)\}]$ $l_{2j} = 60.14 [1 - \exp\{-0.330(t_{2j} - 0.095)\}]$	56.34	3.02	5, 18	0.037
$t_{01} = t_{02}$	$l_{1j} = 56.45 [1 - \exp\{-0.361(t_{1j} - 0.057)\}]$ $l_{2j} = 60.77 [1 - \exp\{-0.313(t_{2j} - 0.057)\}]$	50.76	0.94	5, 18	0.476

5.8 Consideraciones sobre comparación de curvas de crecimiento

Cuando se derivan de datos longitud - edad, las estimaciones de L_{∞} , K (ó g), y t_0 forman una distribución de probabilidad multivariada. Al seleccionar una prueba multivariada, tal como la T^2 de Hotelling, las covarianzas entre las estimaciones de los parámetros se incorporan a la prueba y no representan más un problema; no se requiere ningún tratamiento adicional para las covarianzas.

El sesgo en el valor de la T^2 que resulta de conjuntar las matrices de varianza-covarianza estimadas de estructuras de varianza diferentes (varianzas heterogéneas) es eliminado al seleccionar $N_1 = N_2$.

Es más difícil cumplir con la suposición de normalidad (gaussianidad). Las transformaciones son complejas y a menudo no producen distribuciones conjuntas normales. No obstante, el estadístico F usado en la T^2 de Hotelling es poco afectado por datos moderadamente no-normales, especialmente con valores altos de N .



Al comparar los valores críticos de F se obtiene un ordenamiento por importancia de la influencia de los parámetros en la diferencia y se elimina cualquier confusión causada por grados de libertad altos.

Este procedimiento se ha empleado en años recientes para comparar funciones de crecimiento. Sin embargo, Cerrato (1990) ha hecho algunas anotaciones en relación con la formulación de Bernard (1981). Basado en una comparación empírica entre razón de verosimilitud, t de Student, χ^2 univariada y pruebas de T^2 , Cerrato (1990) concluye que las pruebas de razón de verosimilitud es el más preciso y debería ser el procedimiento de elección.

Las pruebas de razón de verosimilitud permiten tanto el análisis global como (en caso de diferencia) la prueba individual de los parámetros, permitiendo conocer cuál es el de mayor relevancia para la existencia de la diferencia.

El análisis de la suma de cuadrados residual (ASCR) propuesto por Chen *et al.* (1992) sólo permite llevar a cabo una comparación global de curvas. Sin embargo, puede aplicarse para comparar más de dos funciones. Cuando se comparan dos grupos, el resultado coincide con la formulación de Roff (2006).

Las pruebas de verosimilitud basadas en el estadístico F propuestas por Roff (2006), además de la comparación total de dos muestras, y a semejanza de las pruebas de Kimura (1980), hacen posible comparar cada uno de los parámetros de la función de crecimiento de von Bertalanffy.

Cabe señalar finalmente que si las regresiones se ponderan por el número de organismos de cada edad, los resultados de las pruebas anteriores, en general, llevan a indicar diferencia entre los grupos. Se requiere investigación adicional al respecto.

5.9 Aplicación a funciones Logística y Gompertz

Los programas presentados anteriormente para la ecuación de von Bertalanffy pueden modificarse para la inclusión de otras funciones de crecimiento. A continuación presentamos los programas para comparación de curvas de crecimiento Logística y Gompertz aplicados a los datos de crecimiento por retrocálculo del caracol marino *Hexaplex princeps* proveniente de la pesca artesanal de Puerto Ángel, Oaxaca, México tomados de (Saito-Quezada, 2018) incluidos en el Cuadro 5.6



Cuadro 5.6. Longitudes retrocalculadas y edades estimadas para hembras y machos del caracol *Hexaplex princeps* (tomados de Saito-Quezada, 2018).

Edad estimada (años)	Longitud de la concha	Longitud de la concha
1	3.25	3.41
2	3.55	3.73
3	3.87	4.04
4	4.17	4.35
5	4.50	4.67
6	4.83	5.01
7	5.19	5.37
8	5.57	5.76
9	5.97	6.17
10	6.42	6.59
11	6.89	7.05
12	7.38	7.51
13	7.87	7.96
14	8.33	8.38
15	8.70	8.77
16	9.03	9.15
17	9.36	9.54
18	9.65	9.91
19	9.97	10.30
20	10.23	10.53
21	10.41	10.73
22	10.59	10.55
23		11.12
24		11.00
25		11.40
26	12.24	11.74



Primeramente se aplica la prueba de T^2 de Hotelling (Bernard, 1981) para comparar las funciones de crecimiento logísticas de hembras y machos.

```
. use vbrchpcomb
. nl log3 lrc edad if sexo==1
```

Salida omitida

```
. predict lrceorgh
. matrix Sh = e(V)
. matrix Ph = e(b)
. nl log3 lrc edad if sexo==2
```

Salida omitida

```
. predict lrceogm
. matrix Sm = e(V)
. matrix Pm = e(b)
. hotegfun Sh Sm Ph Pm 23 26 99
```

Los resultados de estos comandos se presentan en el Cuadro 5.7.

Posteriormente se comparan las ecuaciones ajustadas de Gompertz, comparación que se presenta en el Cuadro 5.8.

```
. nl gom3 lrc edad if sexo==1
```

Salida omitida

```
. predict lrcegomh
. matrix Sh = e(V)
. matrix Ph = e(b)
. nl gom3 lrc edad if sexo==2
```

Salida omitida

```
Predict lrcegom
. matrix Sm = e(V)
. matrix Pm = e(b)
. hotegfun Sh Sm Ph Pm 23 26 99
```



Cuadro 5.7. Prueba de T^2 de Hotelling para comparación de parámetros de la función de crecimiento Logística (adaptada de la prueba para dos FCVB de Bernard, 1981) de hembras y machos de *Hexaplex princeps*.

Matriz de varianza-covarianza (S)			Inversa de S		
0.08751745	-0.00113154	0.1206424	354.153484	4771.80938	-203.86809
	0.00001831	-0.0015370		276881.058	-864.66910
		0.17359993			139.781785
[0.9430 -0.0101 1.7098] = [P1 - P2]'					
$T^2 = 407.4279$ $T^2_{0.01; 3,45} = 13.3142$ $F_{0.01; 3,45} = 4.2492$					
				<i>F crítica</i>	
0.6340	$\leq L_{\infty 1} - L_{\infty 2} \leq$	1.2520		39.5767	
-0.0145	$\leq K_1 - K_2 \leq$	-0.0056		21.5251	
1.2746	$\leq t_{01} - t_{02} \leq$	2.1449		65.5865	

Cuadro 5.8. Prueba de T^2 de Hotelling para comparación de parámetros de la función de crecimiento de Gompertz (adaptada de la prueba para dos FCvB de Bernard, 1981) de hembras y machos de *Hexaplex princeps*.

Matriz de varianza-covarianza (S)			Inversa de S		
0.48134467	-0.00287705	0.46361667	142.708034	5748.18675	-110.33672
	0.00001975	-0.0026921		483120.598	-2969.9057
		0.45938375			96.1255957
[2.1230 -0.0113 2.4869] = [P1 - P2]'					
$T^2 = 310.4933$ $T^2_{0.01; 3,45} = 13.3142$ $F_{0.01; 3,45} = 4.2492$					
				<i>F crítica</i>	
1.3983	$\leq L_{\infty 1} - L_{\infty 2} \leq$	2.8476		36.4689	
-0.0159	$\leq K_1 - K_2 \leq$	-0.0067		25.1889	
1.7790	$\leq t_{01} - t_{02} \leq$	3.1948		52.4376	



La prueba para las dos funciones indican diferencias entre sexos, con el parámetro t_0 , seguido por las K_s como los que mas influyen para la diferencia.

A continuación se llevan a cabo las comparaciones por las pruebas de razón de verosimilitud RV (Kimura, 1980) y de máxima verosimilitud MV (Chen *et al.*, 2002; Roff, 2006). Para ello se escribieron los programas “l3lrtest.ado” y “l3mltest.ado” los cuales efectúan la comparación de dos funciones Logísticas por RV y MV, respectivamente y “g3lrtest.ado” y “g3mltest.ado”, programas que llevan a cabo la comparación de dos funciones de Gompertz por RV y MV también respectivamente. El “3” incluido en el nombre de los programas indica que las funciones, tanto la Logística como la de Gompertz son las versiones con tres parámetros.

Para la comparación por sexos de la Función Logística, los comandos que llaman a los programas para las pruebas son:

```
. l3lrtest lrc edad sexo
```

Salida omitida

```
. l3mltest lrc edad sexo
```

Salida omitida

Los resultados, incluidos en los Cuadros 5.9 y 5.10 indican diferencias en las expresiones de crecimiento Logísticas destacando, la influencia de la edad al tamaño cero y la longitud asintótica en la diferenciación y en menor grado el del parámetro de curvatura.

Cuadro 5.9. Pruebas de razón de verosimilitud para la Función de Crecimiento Logístico para hembras (1) y machos (2) de *Hexaplex princeps* (Kimura, 1980).

Restricciones lineales	Modelos equivalentes	SCR	χ^2	g^l	P
Ninguna	$l_{1j} = 13.98/[1+ \exp\{-0.119(t_{1j} - 11.222)\}]$ $l_{2j} = 13.04/[1+ \exp\{-0.129(t_{2j} - 9.512)\}]$	0.91			
$L_{\infty 1} = L_{\infty 2}$ $K_1 = K_2$ $t_{01} = t_{02}$	$l_j = 13.37/[1+ \exp\{-0.125(t_{ij} - 10.178)\}]$	1.27	16.54	3	0.001



Cuadro 5.9. Pruebas de razón de verosimilitud para la Función de Crecimiento Logístico para hembras (1) y machos (2) de *Hexaplex princeps* (Kimura, 1980) (continuación).

Restricciones lineales	Modelos equivalentes	SCR	χ^2	gl	P
$L_{\infty 1} = L_{\infty 2}$	$l_{1j} = 13.37/[1+ \exp\{-0.126(t_{1j} - 10.376)\}]$ $l_{2j} = 13.37/[1+ \exp\{-0.123(t_{2j} - 9.967)\}]$	1.02	5.87	1	0.015
$K_1 = K_2$	$l_{1j} = 13.59/[1+ \exp\{-0.124(t_{1j} - 10.692)\}]$ $l_{2j} = 13.27/[1+ \exp\{-0.124(t_{2j} - 9.815)\}]$	0.96	3.00	1	0.083
$t_{01} = t_{02}$	$l_{1j} = 13.25/[1+ \exp\{-0.128(t_{1j} - 10.183)\}]$ $l_{2j} = 13.49/[1+ \exp\{-0.121(t_{2j} + 10.183)\}]$	1.10	9.45	1	0.002

Cuadro 5.10. Pruebas de máxima verosimilitud para la Función de Crecimiento Logístico de hembras (1) y machos (2) de *Hexaplex princeps* (Roff, 2006).

Restricciones lineales	Modelos equivalentes	SCR	F	gl	P
Ninguna	$l_{1j} = 13.98/[1+ \exp\{-0.119(t_{1j} - 11.222)\}]$ $l_{2j} = 13.04/[1+ \exp\{-0.129(t_{2j} - 9.512)\}]$	0.91			
$L_{\infty 1} = L_{\infty 2}$ $K_1 = K_2$ $t_{01} = t_{02}$	$l_j = 13.37/[1+ \exp\{-0.125(t_{ij} - 10.178)\}]$	1.27	5.75	3, 43	0.0021
$L_{\infty 1} = L_{\infty 2}$	$l_{1j} = 13.37/[1+ \exp\{-0.126(t_{1j} - 10.376)\}]$ $l_{2j} = 13.37/[1+ \exp\{-0.123(t_{2j} - 9.967)\}]$	1.02	5.47	1, 43	0.0240
$K_1 = K_2$	$l_{1j} = 13.59/[1+ \exp\{-0.124(t_{1j} - 10.692)\}]$ $l_{2j} = 13.27/[1+ \exp\{-0.124(t_{2j} - 9.815)\}]$	0.96	2.72	1, 43	0.106
$t_{01} = t_{02}$	$l_{1j} = 13.25/[1+ \exp\{-0.128(t_{1j} - 10.183)\}]$ $l_{2j} = 13.49/[1+ \exp\{-0.121(t_{2j} + 10.183)\}]$	1.10	9.14	1, 43	0.004



Para la comparación de las funciones de Gompertz, se introducen los comandos siguientes:

```
. g3lrtest lrc edad sexo
```

Salida omitida

```
. g3mlttest lrc edad sexo
```

Salida omitida

La salida omitida se ha formateado convenientemente y se incluye en los Cuadros 5.11 y 5.12.

Cuadro 5.11. Pruebas de razón de verosimilitud para la Función de Crecimiento de Gompertz para hembras (1) y machos (2) de *Hexaplex princeps* (Kimura, 1980).

Restricciones lineales	Modelos equivalentes	SCR	χ^2	gl	P
Ninguna	$l_{1j} = 17.00 \exp[-\exp\{-0.064(t_{1j} - 9.294)\}]$ $l_{2j} = 14.88 \exp[-\exp\{-0.076(t_{2j} - 6.807)\}]$	1.29			
$L_{\infty 1} = L_{\infty 2}$ $K_1 = K_2$ $t_{01} = t_{02}$	$l_j = 15.61 \exp[-\exp\{-0.071(t_{ij} - 7.727)\}]$	1.69	12.99	3	0.005
$L_{\infty 1} = L_{\infty 2}$	$l_{1j} = 15.64 \exp[-\exp\{-0.072(t_{1j} - 7.973)\}]$ $l_{2j} = 15.64 \exp[-\exp\{-0.070(t_{2j} - 7.522)\}]$	1.44	5.27	1	0.022
$K_1 = K_2$	$l_{1j} = 15.92 \exp[-\exp\{-0.070(t_{1j} - 8.255)\}]$ $l_{2j} = 15.50 \exp[-\exp\{-0.070(t_{2j} - 7.367)\}]$	1.39	3.59	1	0.058
$t_{01} = t_{02}$	$l_{1j} = 15.48 \exp[-\exp\{-0.073(t_{1j} - 7.789)\}]$ $l_{2j} = 15.88 \exp[-\exp\{-0.068(t_{2j} + 7.789)\}]$	1.51	7.56	1	0.006



Como en las comparaciones por sexo en la Función Logística, ambas pruebas indican diferencias significativas con la edad y resaltan al tamaño cero y la longitud asintótica como las responsables principales de la diferencia.

En la Figura 5.2 se presentan los cuatro ajustes indicando sus longitudes asintóticas. Puede observarse que las curvas son muy semejantes con diferencias claras sólo en los valores de longitud infinita.

```
. two mspline lrcelogh edad, bands(40) || mspline lrcelogm edad, bands(40)
lpattern(dash) || mspline lrcegomh edad, bands(40) || mspline lrcegommm edad,
bands(40) yline(13.98 13.04 17 14.88) yscale(range(2 18)) ylab(2(2)18)
legend(label(1 "Lh") label(2 "Lm") label(3 "Gh") label(4 "Gm") label(5 "lRCh")
label(6 "lRCm") pos(5) ring(0)) || scatter lrc edad if sex==1, ms(Oh) || scatter
lrc edad if sex==2, ms(Sh) text(17.6 3 "{it:L{sub:{&infin}}} = 17.0, Gh)")
text(15.5 3 "{it:L{sub:{&infin}}} = 14.88, Gm)") text(14.4 3 "{it:L{sub:{&infin}}}
= 13.98, Lh)") text(13.4 3 "{it:L{sub:{&infin}}} = 13.04, Lm)")
```

Cuadro 5.12. Pruebas de máxima verosimilitud para la Función de Crecimiento de Gompertz de hembras (1) y machos (2) *Hexaplex princeps* (Roff, 2006).

Restricciones lineales	Modelos equivalentes	SCR	F	gl	P
Ninguna	$l_{1j} = 17.00 \exp[-\exp\{-0.064(t_{1j} - 9.294)\}]$ $l_{2j} = 14.88 \exp[-\exp\{-0.076(t_{2j} - 6.807)\}]$	1.29			
$L_{\infty 1} = L_{\infty 2}$ $K_1 = K_2$ $t_{01} = t_{02}$	$l_j = 15.61 \exp[-\exp\{-0.071(t_j - 7.727)\}]$	1.69	4.35	3, 43	0.0092
$L_{\infty 1} = L_{\infty 2}$	$l_{1j} = 15.64 \exp[-\exp\{-0.072(t_{1j} - 7.973)\}]$ $l_{2j} = 15.64 \exp[-\exp\{-0.070(t_{2j} - 7.522)\}]$	1.44	4.88	1, 43	0.0325
$K_1 = K_2$	$l_{1j} = 15.92 \exp[-\exp\{-0.070(t_{1j} - 8.255)\}]$ $l_{2j} = 15.50 \exp[-\exp\{-0.070(t_{2j} - 7.367)\}]$	1.39	3.27	1, 43	0.078
$t_{01} = t_{02}$	$l_{1j} = 15.48 \exp[-\exp\{-0.073(t_{1j} - 7.789)\}]$ $l_{2j} = 15.88 \exp[-\exp\{-0.068(t_{2j} + 7.789)\}]$	1.51	7.18	1, 43	0.010

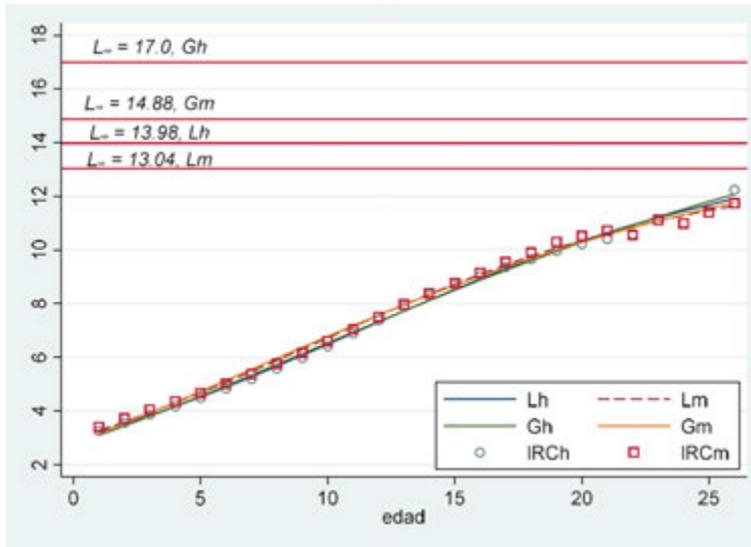


Figura 5.2. Ajuste de funciones Logística y Gompertz a datos longitud-edad retrocalculada a partir de opérculos del caracol marino *Hexaplex princeps* (Saito-Quezada, 2018).

Capítulo 6

Índices morfofisiológicos



6.1. Introducción

El estudio de la reproducción es un tópico fundamental en las ciencias pesqueras. Se han establecido varios métodos para llevar a cabo este objetivo que van desde procedimientos histológicos detallados hasta índices derivados de medidas corporales. Estos índices morfométricos están basados generalmente en medidas corporales tales como el peso total y eviscerado, de gónadas, del hígado y la longitud corporal. En general, estas medidas están correlacionadas positivamente con el tamaño y se utilizan para estimar otros procesos relacionados. De esta forma, el peso de las gónadas está relacionado con la madurez y el desove, los cuales a su vez dependen del sexo, estadio de madurez sexual, tamaño individual, temporada del año, etc. El peso total o eviscerado se usa para estimar la condición (robustez o grado de bienestar de los animales) suponiendo que los individuos más pesados a una longitud dada están en mejor condición. El peso corporal es un reflejo de los procesos fisiológicos involucrados en la reproducción y supervivencia (García-Berthou y Moreno-Amich, 1993).

Para estudiar el ciclo reproductivo se ha empleado (ver Capítulo 7) el índice gonadosomático (IGS):

$$IGS = 100GW^{-1}$$

donde W es el peso total y G es el peso de las gónadas. Además de este índice numérico, se usan otros en conjunto para describir el estado general del pez. Un ejemplo es el factor de condición de Fulton (Ricker, 1975; pero ver interesante discusión en Nash *et al.*, 2006 y Valencia-Hernández, 2003), K_1 :

$$K_1 = 100WL^{-3}$$

Factor de condición de Ricker (1975), K_2 :

$$K_2 = 100WL^{-b}$$



y el factor de condición relativa K_3 :

$$K_3 = 100Wa^{-1}L^{-b}$$

donde L es la longitud y a y b son los parámetros de la relación longitud-peso:

$$W = aL^b$$

Otros índices para describir la condición es el índice hepatosomático (IHS):

$$IHS = 100HW^{-1}$$

donde H es el peso del hígado.

En años recientes se han publicado varios artículos sobre la evaluación de los índices morfométricos existentes y se han propuesto varios nuevos (DeVlaming *et al.*, 1982; Erickson *et al.*, 1985; Bolger y Connolly, 1989; Cone, 1989; Patterson, 1992; García-Berthou y Moreno-Amich, 1993). Los problemas de los métodos tradicionales pueden resumirse como sigue:

- 1) Algunos de ellos son índices de razón: (factor de condición de Fulton, índice Gonadosomático, e índice Hepatosomático) y estos cocientes:
 - Presentan incremento de variabilidad en comparación con las variables originales usadas para el cálculo.
 - Proporcionan una estimación sesgada de la verdadera media de la razón.
 - Poseen distribuciones de razón poco comunes, no Gaussianas y a menudo intratables.
 - Tienden a hacer difícil el entendimiento de las relaciones entre las variables.
- 2) Especialmente para el índice de Fulton, IGS , y el IHS , suponen incorrectamente una relación lineal (crecimiento isométrico).
- 3) A menudo suponen incorrectamente independencia respecto de la longitud.



- 4) La mayoría de los índices suponen que la variación de la condición está indicada solamente por la variación del estimador a . Esto implica que las pendientes b de la relación longitud-peso son las mismas entre grupos.

Algunos autores han presentado nuevos enfoques que intentan superar algunas de las dificultades arriba listadas. Por ejemplo, Erickson *et al.* (1985) propuso el “índice gonadal relativo” (IGR):

$$IGR = a_i \times \frac{W}{S^b}$$

donde i indica la etapa gonadal y S es una medida del “tamaño” del pez, (como el peso corporal – peso gonadal). Este procedimiento supone errores multiplicativos lognormales y para la estimación por medio de métodos tradicionales de regresión se utiliza una transformación logarítmica. Sólo si las pendientes no difieren significativamente entre las etapas de desarrollo gonádico es correcto usar la pendiente conjunta b para calcular IRG .

Patterson (1992) propuso que la condición pudiera estudiarse mejor utilizando un solo modelo para analizar simultáneamente la respuesta del peso del pez a varios factores. Este enfoque se beneficia de la simplicidad de usar un análisis integrado y evita los problemas de las distribuciones sesgadas de los cocientes. En su método, todas las variables independientes que afectan el peso se incluyen al mismo tiempo que la longitud de los peces para obtener una solución por mínimos cuadrados. La evaluación de la significancia de las variables se hace mediante un análisis de covarianza. El coeficiente de cada parámetro se usa entonces como una medida de su efecto en la condición del pez. Su cálculo puede efectuarse por medio de programas estadísticos capaces de estimar modelos lineales generalizados.

Un paso adicional en la complejidad se incluye en el método propuesto por García-Berthou y Moreno-Amich (1993). En lugar de tener una sola variable de respuesta en el modelo lineal, todas las medidas morfométricas relacionadas con la actividad reproductiva se incluyen simultáneamente en conjunto con factores tales como el sexo y la fecha de muestreo y se ajustan por el uso de covariadas (como la longitud del cuerpo). La propuesta de estos autores emplea el análisis multivariado de covarianza (ANMUCOVA). Las medias predichas del modelo describen las variaciones estacionales y de otros factores haciendo innecesario el empleo de índices.

Un enfoque relacionado con el factor de condición relativo es el “peso relativo” (Wr) que se calcula mediante la expresión (Blackwell, et al. 2000):



$$W_r = (W/W_s) \times 100$$

donde W = peso observado del individuo y W_s es el peso estándar para un espécimen de la misma longitud. Este “peso estándar”, para ser representativo de la especie, debe calcularse con un número elevado de organismos provenientes de todas las localidades de su ámbito de distribución y durante todas las épocas del año. Si esto se consigue, los valores obtenidos permitirán (entre otras aplicaciones potenciales) la comparación de valores entre localidades con fines de evaluación y manejo. Este índice ha ido ganando aceptación para aplicarlo primeramente a especies de pesca deportiva y últimamente a otras especies de América del Norte y algunas de otras regiones (Ogle, 2016). Sin embargo consultar a Stevenson y Wood (2006) o Cade *et al.* (2011) para considerar otras medidas.

6.2. El caso del pez “suzuki” (*Lateolabrax japonicus*)

En el presente capítulo se presentará el uso del ANMUCOVA para describir el ciclo reproductivo y la condición del “suzuki” (*Lateolabrax japonicus*) de la captura comercial realizada en la Bahía de Tokio (Salgado-Ugarte, 1995). El “suzuki” es una especie de importancia económica que habita aguas japonesas. Es muy apreciado como alimento y ha sido considerado como una especie prometedora para cultivo marino en invierno (Matsumiya, *et al.*, 1982).

Varios autores han estudiado aspectos reproductivos de *L. japonicus*. Mito (1957) estudió el desarrollo de huevos y larvas de *L. japonicus*. Este autor llevó a cabo inseminación artificial para observar el desarrollo de los huevos y larvas hasta la eclosión. De acuerdo con sus observaciones, la temporada de desove parece extenderse de noviembre a marzo en las aguas marinas de Kyushu.

El huevo es flotante, incoloro y de forma esférica; mide de 1.35-1.44 mm de diámetro con un solo glóbulo graso de color amarillo claro (raramente dividido en 2 o 3) el cual tiene un diámetro de 0.34-0.38 mm. La eclosión tomó lugar de 4.5 a 5 días a temperaturas del agua de 11.0-16.2 C. Las larvas eclosionadas tienen longitud total de 4.42-4.60 mm, con el glóbulo graso situado en la parte anterior del saco vitelino, sobresaliendo parcialmente de su superficie. El número de miotomas es 19+18=37 (el número vertebral de este pez es 18+17=35). En una semana el saco vitelino y el glóbulo graso se consumió completamente y la larva alcanzó una longitud total de 5.0-5.3 mm. Durante la absorción del saco vitelino no ocurrieron cambios notables.



Las larvas, con longitud total de 5-15 mm son pelágicas y viven en la región costera. Las post-larvas más grandes ó juveniles con longitud total de aproximadamente 20 mm ó más, se encuentran en el fondo cerca de la costa, en la zona de pastos marinos de los estuarios o en las lagunas de marea, y se alimentan principalmente de copépodos y otras larvas pequeñas de crustáceos.

Kosaka (1969) encontró que *L. japonicus* de la Bahía de Sendai desova una vez al año durante un período muy corto que va de inicios de diciembre hasta inicios de enero. Comparados con localidades más al sur, los peces de la Bahía de Sendai muestran una tendencia hacia un período de desove más corto. El lugar de desove se localiza en la parte meridional de la Bahía de Sendai y en una región marina más al sur donde la profundidad del agua es de aprox. 60 m y el tipo de fondo consiste de gránulos y rocas expuestas.

El tamaño biológico mínimo es 245 mm de longitud corporal al segundo año de vida en machos y de 340 - 370 mm al tercer año de vida en las hembras.

Hayashi (1969, 1971, 1972) estudió el ducto reproductivo y el proceso de maduración gonádica de *L. japonicus* de la Bahía de Wakasa. Un breve resumen de sus hallazgos se presenta a continuación:

Como en el caso de la mayoría de los teleósteos, los testículos de esta especie consisten en dos partes: hilos de tejido conectivo (testículo mediastino) y estroma en el cual se localizan numerosos túbulos seminíferos. Después del desove, se distingue una capa de células epiteliales seminíferas a lo largo de la pared de los túbulos y se desarrolla hasta la aparición de las espermatogonias primarias. Las células epiteliales (consideradas como homólogas a las células de Sertoli) aparentemente tienen tres funciones:

- a) la descomposición y reabsorción de los restos de células espermáticas en varios estadios de maduración en los túbulos
- b) la fuente original de las espermatogonias primarias
- c) la provisión nutricional de las células germinales

Las espermatogonias primarias se hacen visibles en los túbulos de abril a agosto. Después de la mitosis cambian a espermatogonias secundarias a mediados de octubre. Posteriormente prosiguen rápidamente su proceso de maduración a través de etapas meióticas. Los espermatozoides aparecen en el túbulo por primera vez a mediados de noviembre. En la Bahía de Wakasa, con base en las observaciones microscópicas, la temporada de desove de



L. japonicus ocurre en las dos ó tres semanas entre finales de diciembre y principios de enero. El cambio estacional del índice gonadosomático refleja cercanamente el proceso histológico de maduración y es útil para evaluar la madurez. El macho de esta especie requiere de dos años para madurar y el tamaño biológico mínimo del macho es aproximadamente de 260 mm de longitud estándar.

La oogénesis de *L. japonicus* es semejante a la de otros peces óseos. Sin embargo, en esta especie, la oogénesis está marcada por el patrón característico de distribución de las vesículas de yema y el período más bien corto necesario para la maduración de los oocitos en la etapa de vesícula de yema dentro del huevo maduro (aproximadamente tres meses de septiembre a principios de enero). Este período más corto requerido para la oogénesis parece estar cercanamente relacionado con el modo de vida (la migración de peces adultos hacia las aguas marinas frente a las costas y está íntimamente relacionado con la verificación del proceso de maduración.

Las células foliculares desempeñan un papel importante en el suministro de nutrimentos para el oocito en desarrollo y la función fagocigótica del oocito atrésico. Las células foliculares parecen funcionar como las células seminíferas epiteliales del testículo lo que sugiere su homología. Los cambios estacionales de los valores del índice gonádico reflejan con cierta precisión la madurez del pez. Tanto la observación microscópica de los ovarios y los datos del índice gonádico indican que en la Bahía Wakasa, el desove ocurre de fines de diciembre a mediados de enero, principalmente a principios de enero. La talla de primera madurez para las hembras es de aproximadamente 350 mm de longitud estándar, la cual corresponde al tamaño del grupo de 2 ó tres años de edad.

Hirai (1988) estudió (mediante microscopía electrónica de barrido) los micrópilos y las estructuras a su alrededor en los huevos pelágicos de nueve especies y observó que *L. japonicus* posee un número reducido de protuberancias.

6.2.1. Material y métodos

Organismos de *Lateolabrax japonicus* (406 individuos) fueron obtenidos de la captura comercial de la Bahía de Tokio de septiembre de 1993 a junio de 1995. Los especímenes fueron preservados en hielo y transportados al laboratorio donde se congelaron. Los especímenes fueron medidos (longitud total y estándar) al mm más cercano y pesados (± 0.1 g). Todos los individuos adultos (más grandes) fueron disectados para determinar el sexo y para tomar el peso de gónadas, hígado, estómago, contenido estomacal e intestino



(± 0.01 g). Una submuestra de ejemplares juveniles (de menor tamaño) fue tomada para considerar las medidas anteriores (Cuadro 6.1); sólo se incluyeron aquellos con peso gonadal apreciable. Mayores detalles se encuentran en Salgado-Ugarte (1995).

Si se considera la recomendación de varios autores (Stearns, 1983, García-Berthou y Moreno-Amich, 1993), para el análisis se calculó un “peso eviscerado” como:

peso eviscerado = peso total - (peso de gónada + peso del hígado + peso del estómago + peso del intestino)

Para el ANMUCOVA se consideraron las siguientes variables:

Variables de respuesta: peso eviscerado, peso de la gónada y peso del hígado.

factores: sexo, fecha de muestreo, edad

covariada: longitud estándar

Los valores de peso y longitud fueron multiplicados por una constante (10^6) y transformados a logaritmo natural para evitar trabajar con valores negativos. Se probaron varios diseños de análisis debido a que un modelo completamente factorial no podía aplicarse por la no ocurrencia de individuos en algunas combinaciones de factores (por ejemplo, no hubo juveniles antes de invierno y en verano). Los datos se dividieron en siete conjuntos de acuerdo a la siguiente clave:

1. Total de individuos
2. Machos
3. Hembras
4. Machos 1 año de edad
5. Machos 2 años de edad
- 6 Hembras 1 año de edad
- 7 Hembras 2 años de edad



Cuadro 6.1. Número de individuos por sexo y fecha de muestreo considerados para el análisis de aspectos reproductivos y condición.

Fecha	Sexo			Total
	Machos	Hembras	Indeterminados	
1993				
09 - 24	2	9	0	11
10 - 12	3	6	0	9
10 - 29	4	5	0	9
11 - 10	6	8	0	14
12 - 10	7	8	3	18
1994				
02 - 19	2	4	14	20
03 - 16	11	13	1	25
04 - 14	27	4	1	32
05 - 13	6	10	0	16
06 - 24	7	7	2	16
08 - 17	6	10	4	20
11 - 10	14	8	1	23
Total	95	92	26	213

Para todos estos conjuntos de datos se consideró a la fecha como factor y a la longitud corporal como covariada. Como el primer paso se obtuvo la significancia de las interacciones de la covariada con los factores. Después de verificar la significancia de la interacción, el segundo paso en el análisis fue calcular el diseño estándar del ANMUCOVA. Como tercer paso, se llevaron a cabo y se interpretaron contrastes de medias *a posteriori*. El nivel de significancia en todos los análisis fue de 0.05.

Los cálculos se efectuaron utilizando el módulo para modelos lineales generalizados (GLM) del paquete estadístico MINITAB (MINITAB Inc., 1994) y el procedimiento MANOVA del SPSS-X (SPSS Inc., 1986). Para llevar a cabo los contrastes conjuntos univariados *a posteriori* para valores medios adyacentes por fecha, se empleó el subcomando CONTRAST = REPEATED en el MANOVA del SPSS.



6.2.2. Análisis multivariado de covarianza (ANMUCOVA)

En lo que sigue se presentan los resultados obtenidos con el procedimiento MANCOVA. La Figura 6.1 presenta las variables transformadas logarítmicamente consideradas para el análisis. A pesar de la dispersión variable alrededor de una tendencia lineal, puede notarse la linealidad en la relación entre los datos morfométricos.

Los resultados del ANMUCOVA se presentan gráficamente en las siguientes figuras de acuerdo al esquema descrito a continuación:

Número total de individuos:	Figuras 6.2 – 6.3.
Número total de machos:	Figuras 6.4 - 6.5.
Número total de hembras:	Figuras 6.6 - 6.7.
Machos edad 1:	Figuras 6.8 - 6.9.
Machos edad 2:	Figuras 6.10 - 6.11.
Hembras edad 1:	Figuras 6.12 - 6.13.
Hembras edad 2:	Figuras 6.14 - 6.15.

Estas gráficas muestran los valores medios ajustados por la covariada y clasificadas por la fecha de muestreo. La significancia de los cambios en las medias ajustadas pueden evaluarse por los códigos en la Tabla 6.2.

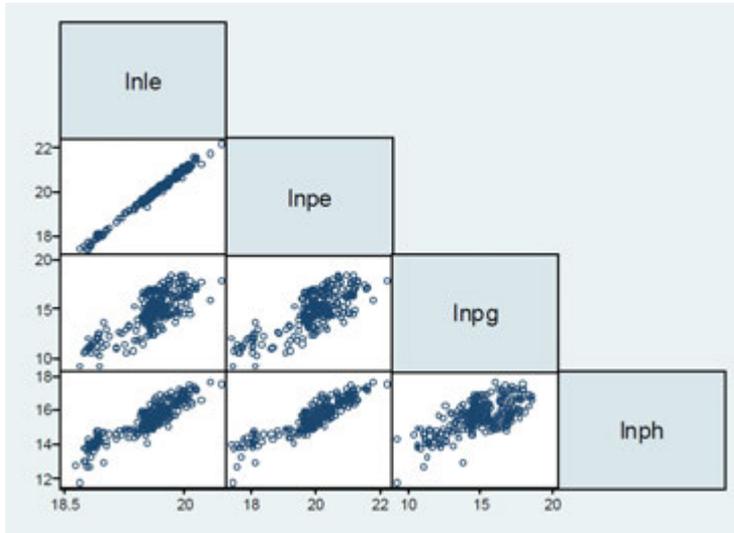


Figura 6.1. Variables transformadas (logaritmo natural) consideradas para el análisis: ln = logaritmo natural; le, longitud estándar; pe = peso eviscerado; pg = peso de gónada; ph = peso del hígado.

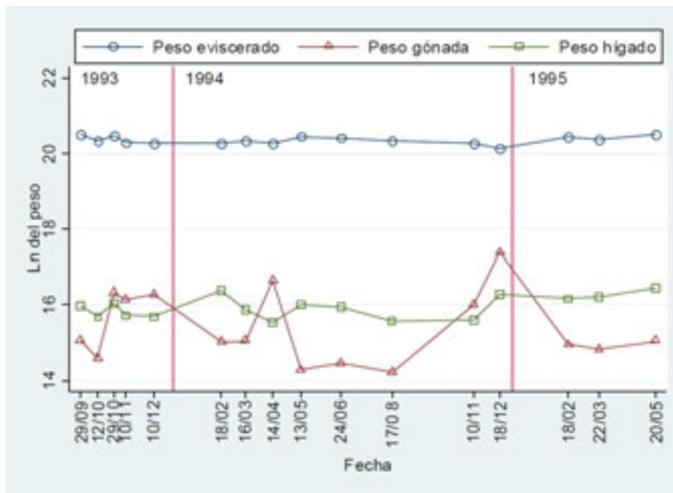


Figura 6.2. Medias ajustadas por longitud por fecha de muestreo. Datos en conjunto.

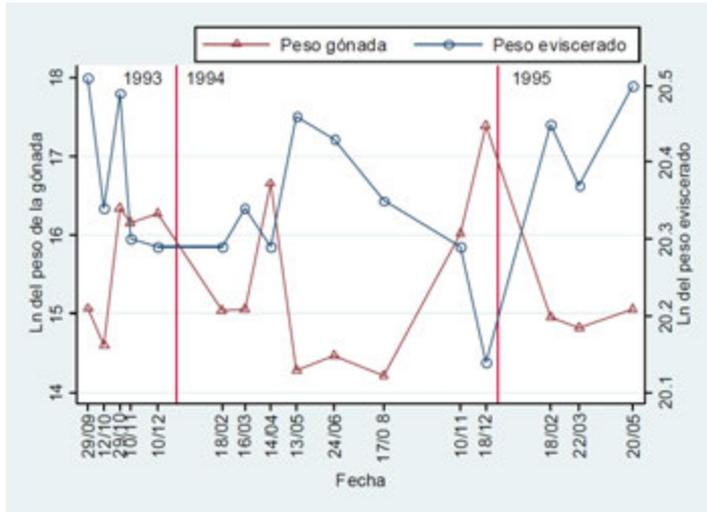


Figura 6.3. Medias ajustadas por longitud (covariada) de peso gonadal y eviscerado por fecha de muestreo. Datos en conjunto.

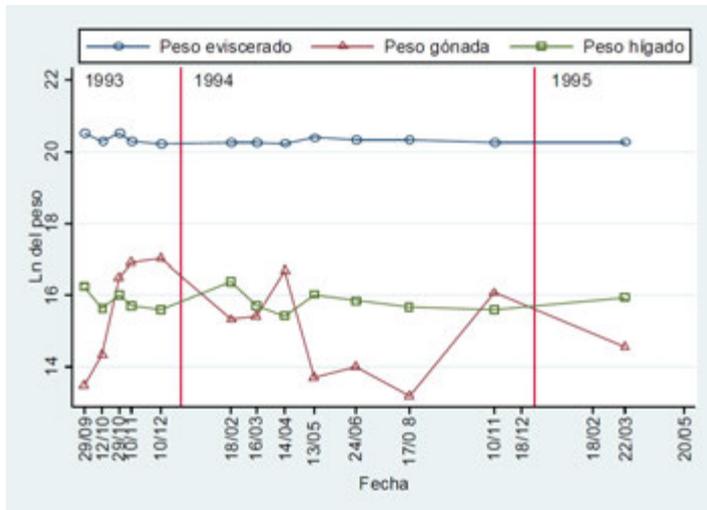


Figura 6.4. Medias ajustadas por longitud (covariada) para valores medios de peso eviscerado, gonádico y del hígado por fecha de muestreo. Machos.

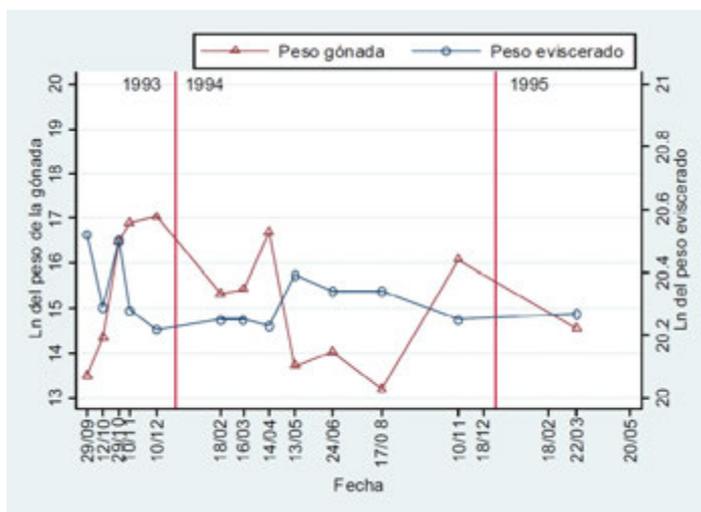


Figura 6.5. Medias ajustadas por longitud (covariada) de peso gonadal y eviscerado, por fecha de muestreo. Machos.

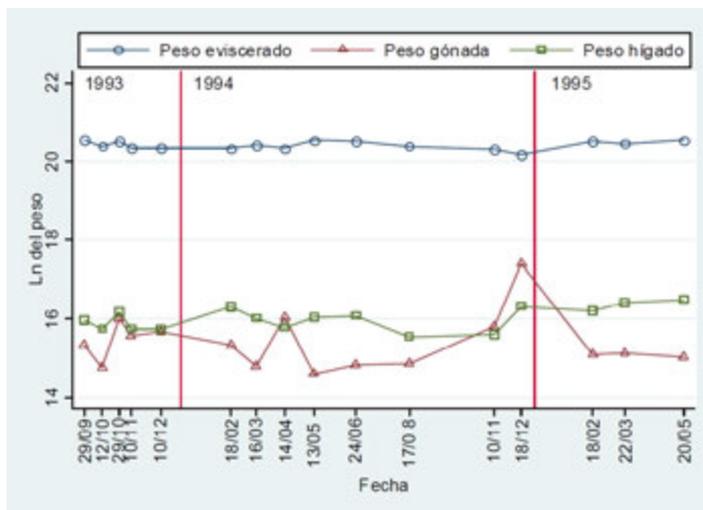


Figura 6.6. Medias ajustadas por longitud (covariada) de peso eviscerado, gonadal y del hígado por fecha de muestreo. Hembras.

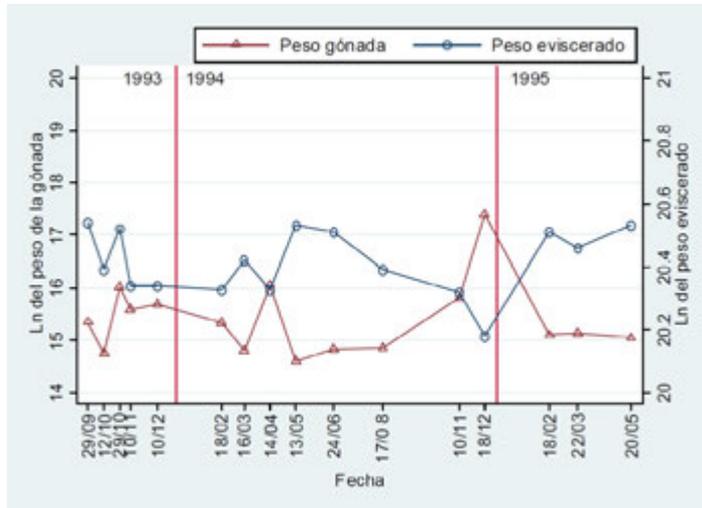


Figura 6.7. Medias ajustadas por longitud (covariada) de peso gonadal y viscerado por fecha de muestreo. Hembras.

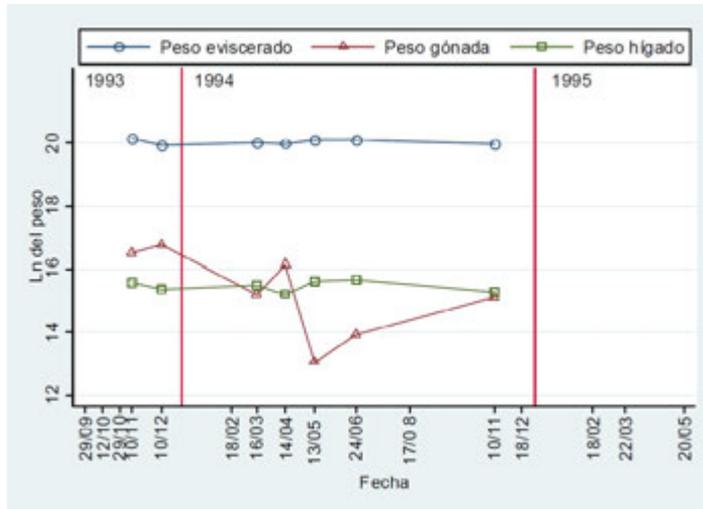


Figura 6.8. Medias ajustadas por longitud (covariada) de peso viscerado, gonadal y del hígado por fecha de muestreo. Machos edad 1.

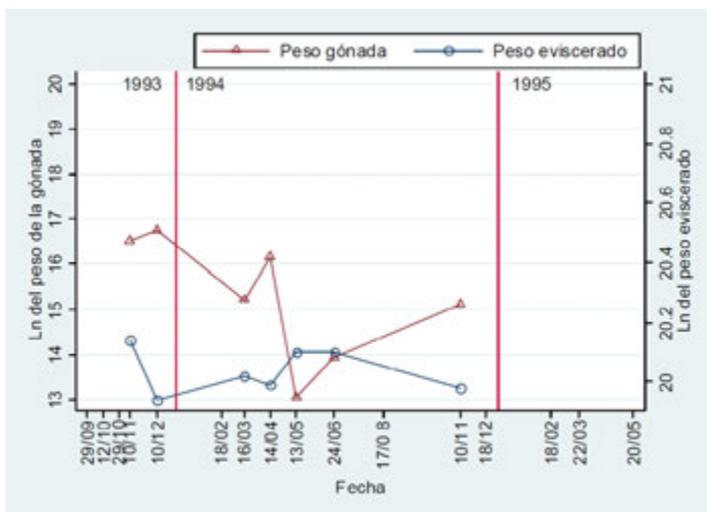


Figura 6.9. Medias ajustadas por longitud (covariada) de peso viscerado y gonadal. Machos edad 1.

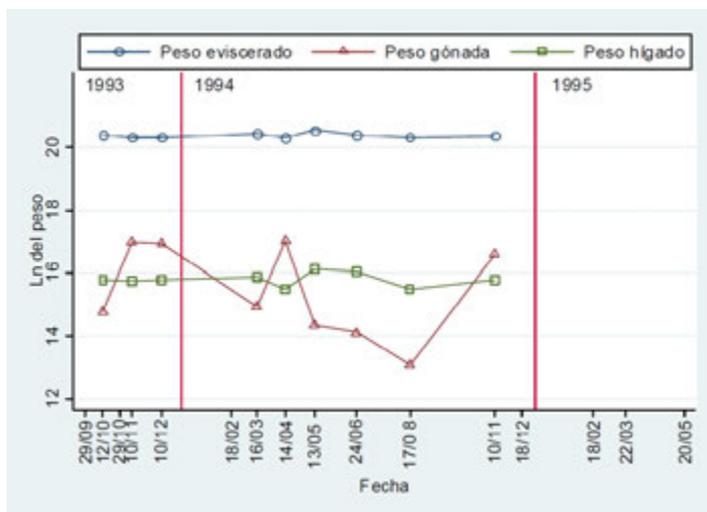


Figura 6.10. Medias ajustadas por longitud (covariada) de peso viscerado, gonadal y del hígado. Machos edad 2.

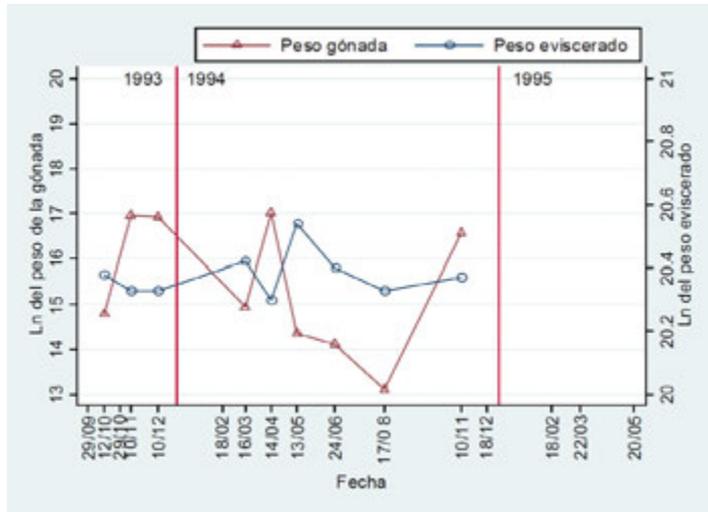


Figura 6.11. Medias ajustadas por longitud (covariada) de peso eviscerado y gonadal. Machos edad 2.

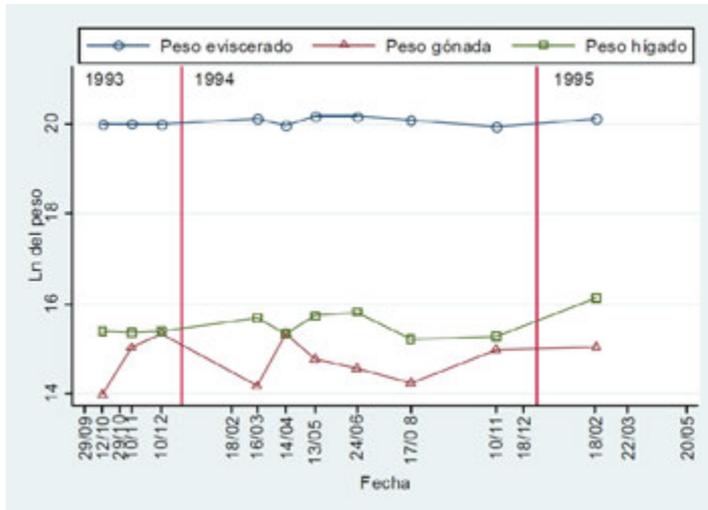


Figura 6.12. Medias ajustadas por longitud (covariada) de peso eviscerado, gonadal y del hígado. Hembras edad 1.

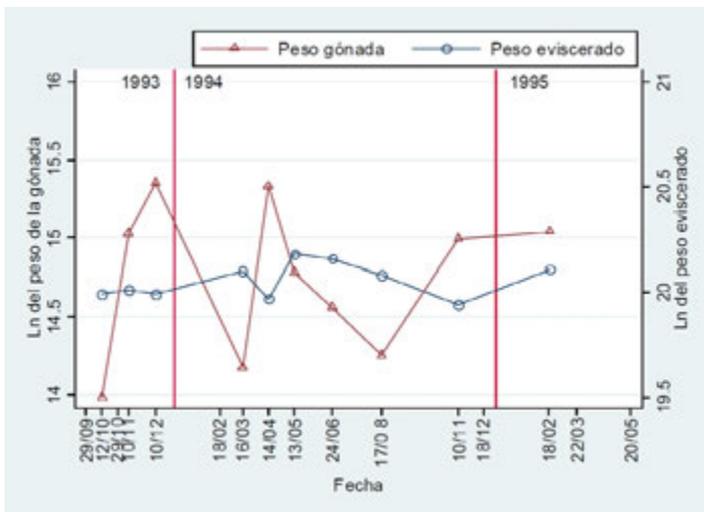


Figura 6.13. Medias ajustadas por longitud (covariada) de peso eviscerado y gonadal. Hembras edad 1.

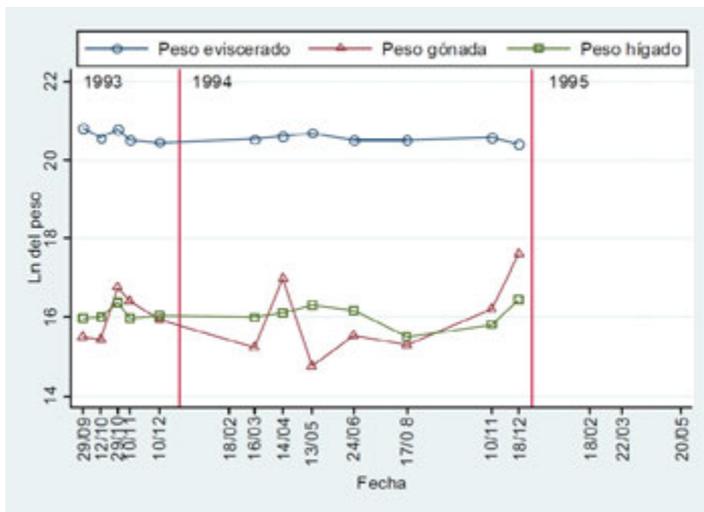


Figura 6.14. Medias ajustadas por longitud (covariada) de peso eviscerado, gonadal y del hígado. Hembras edad 2.

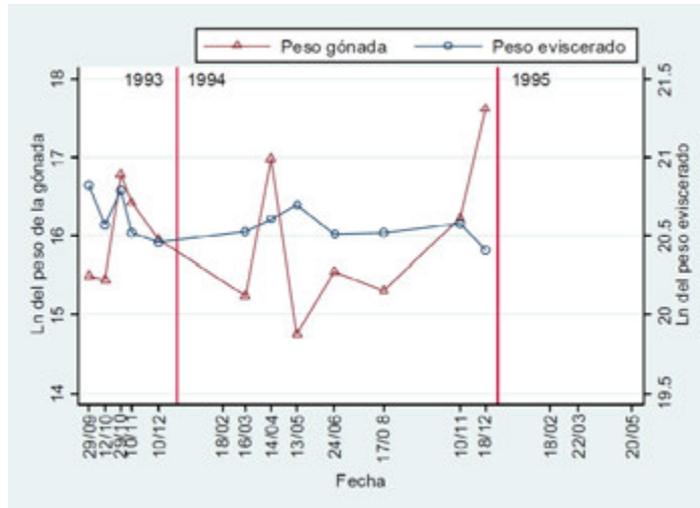


Figura 6.15. Medias ajustadas por longitud (covariada) de peso eviscerado y gonadal. Hembras edad 2.

La transformación logarítmica hace lineales las relaciones entre las variables consideradas para el análisis como se muestra en la Figura 6.1. Se observó que los individuos de menor tamaño poseen un comportamiento diferente. Después, el análisis se llevó a cabo considerando a cada sexo por separado.

Para el ANMUCOVA, se ajustó un diseño preliminar con interacción de factores con covariada. En todos los casos, los términos de la interacción no fueron significativos. Por lo tanto, fue apropiado continuar con el diseño sin las interacciones.

Los resultados del ANMUCOVA preliminar presentados en las Figuras 6.2 – 6.3, indicaron valores medios altos durante noviembre a diciembre (1993 y 1994). La significancia estadística de los cambios en las medias ajustadas se incluye en el Tabla 6.2 parte a). En abril de 1994 se presentó un valor extraordinario alto. Este valor elevado fue verificado al revisar los individuos de dicha fecha de muestreo y los valores altos fueron corroborados. No existe reporte previo sobre valores altos de los índices gonádicos durante la primavera. Todos los estudios concluyen que el período de desove es muy corto y ocurren en los meses de invierno, como se citó en la parte introductoria del capítulo. Hubo una indicación clara de una relación inversa entre las medias del peso de hígado y las medias de las gónadas:



valores medios elevados de peso hepático se presentaron en verano y bajos en invierno. Aún en los valores inusuales del mes de abril, se observó esta correlación negativa (valores medios bajos del hígado y altos de la gónada). Estas conclusiones se corroboraron estadísticamente mediante las pruebas con nivel de significancia del 0.05.

En este estudio se siguió la recomendación del uso del peso eviscerado para tener una medida más adecuada de la condición. Los resultados preliminares (Figura 6.3) indicaron claramente una relación negativa entre las medias de peso eviscerado con las medias de peso gonádico, corroboradas por las pruebas de contrastes múltiples a posteriori (Tabla 6.2).

El análisis por sexos mostró la misma tendencia. En el caso de los machos, las gónadas manifestaron valores altos en noviembre y diciembre. El valor extraordinario de primavera también estuvo presente. Se observó una relación negativa entre las medias gonádicas y las hepáticas. El peso eviscerado se comportó de manera semejante a los pesos medios del hígado. Todas las afirmaciones anteriores están respaldadas por las pruebas de valores adyacentes.

En las hembras, se observó una tendencia semejante a la de los machos y también fue corroborada por las pruebas estadísticas de comparaciones múltiples a posteriori. Se tuvieron valores medios en el peso de los ovarios de noviembre a diciembre con el pico irregular en abril también presente. Se reconoció asimismo, la correlación negativa entre las medias ajustadas del peso hepático con el gonadal.

Un análisis más detallado se llevó a cabo estudiando las medias por grupo de edad. Para los machos de 1 año de edad, a pesar de que en varios meses no hubo peces pertenecientes a esta cohorte, se apreció una indicación de valores altos de noviembre a diciembre. Además, el valor alto inusual de primavera también se detectó, así como la tendencia negativa entre gónadas e hígado, pero la significancia de ésta última fue limitada por los tamaños de muestra reducidos. El efecto fue más claro al observar las medias del peso eviscerado en relación con los valores medios del peso gonádico. Sin embargo, y de nuevo por el reducido número de ejemplares, no se tuvo significancia estadística. Los machos del grupo 2 mostraron claramente el pico invernal en las medias de peso gonádico y un patrón claramente opuesto entre los pesos hepático y eviscerado (apoyado significativamente por los contrastes de valores adyacentes). Las gónadas de los especímenes machos de edad 2 fueron siempre más pesados que los de edad 1.



Cuadro 6.2. Representación esquemática de la significancia de los contrastes conjuntos univariados para comparar fechas adyacentes para variables del ciclo reproductivo y la condición.

a) Datos en conjunto																
Fecha	1993					1994							1995			
Mes	09	10	10	11	12	02	03	04	05	06	08	11	12	02	03	05
Día	29	12	29	10	10	19	16	14	13	24	17	10	18	18	22	20
LPE	+	-	+	-	-	-	+	+	+	+	+	-	-	+	+	+
LPG	-	-	+	+	+	-	-	+	-	-	-	+	+	-	-	-
LPH	-	-	-	-	-	+	-	-	+	+	-	-	+	+	+	+
b) Machos																
LPE	+	-	+	-	-	-	-	-	+	+	+	+				+
LPG	-	-	+	+	+	-	-	+	-	-	-	+				-
LPH	+	-	-	-	-	+	-	-	+	+	+	+				-
c) Hembras																
LPE	+	-	-	-	-	-	-	-	+	+	-	-	-	+	+	+
LPG	-	-	+	+	+	+	+	+	-	-	-	+	+	-	-	-
LPH	-	-	+	-	-	+	+	+	+	+	-	-	+	+	+	+
d) Machos edad 1																
LPE					NS	NS	NS	NS	NS	NS	NS	NS				
LPG					+	+		-	+	-	-	-				
LPH					+	+		+	-	+	+	+				
e) Machos edad 2																
LPE		-		-	-		-	-	+	+	+	+				
LPG		-		+	+		-	+	-	-	-	+				
LPH*		-		-	-		-	-	+	+	+	+				
f) Hembras edad 1																
LPE*		-		-	-		+	+	+	+	+	-		-		
LPG*		-		+	+		-	+	-	-	-	+		+		
LPH		+		+	+		+	+	+	+	-	-		+		
g) Hembras edad 2																
LPE		NS	NS	NS	NS	NS		NS	NS	NS	NS	NS	NS	NS		
LPG*		-	-	+	+	+		-	+	-	-	-	-	+		
LPH		NS	NS	NS	NS	NS		NS	NS	NS	NS	NS	-	+		

El nivel de significancia fue de 0.05 excepto para las celdas marcadas con un asterisco (*) las cuales usaron un nivel de significancia de 0.1. El significado de los signos es: “+” = valores altos significativos; “-” = valores bajos significativos; NS = no significativo.



Para las hembras, se observó el mismo comportamiento general, tanto para la edad 1 (contrastos significativos) y 2 (variaciones de pesos eviscerados y hepáticos no significativos), pero con mayores diferencias en peso gonádico. Por ejemplo (y después de la re-transformación a la escala original) el peso medio gonádico más alto para las hembras de edad 1 fue de 4.64 g, mientras que para la edad 2, los tres valores medios más altos fueron 18.07, 23.91 y 44.90 g. El valor medió mínimo para las gónadas de hembras de edad 2 fue de 2.57 g, peso muy semejante al peso medio máximo de las gónadas de hembras de edad 1.

Estas observaciones sugieren que, a pesar de que el patrón de variación del peso gonádico medio exhibido por los individuos de edad 1 es semejante a los especímenes de edad 2, existe una diferencia apreciable en peso. Esto puede sugerir que sólo los individuos mayores a 2 años de edad se reproducen efectivamente. Esta observación está respaldada por los hallazgos de Hayashi (1971,1972) y Kosaka (1969) lo que indica que los machos en el segundo año de vida y las hembras de 2 (ó 3) años de edad llevan a cabo reproducción activa.

La evidencia encontrada en el estudio del presente ejemplo, corrobora que la época de desove de esta especie es el invierno. Sin embargo cabe mencionar el pico irregular en los valores medios del peso gonádico durante la primavera de 1993. Los peces con esta característica que no concuerda con el patrón conocido se obtuvieron de la captura comercial. Sus gónadas fueron revisadas y se pudo verificar que los valores no son producto de equivocaciones de medición. Sin un conocimiento más preciso de la localidad de pesca de estos peces es difícil inferir una posible causa. Los estudios histológicos llevados a cabo en la Bahía de Wakasa, indican un período de maduración muy rápido limitado a los fines de diciembre y hasta mediados de enero. En localidades más al norte se ha reconocido una tendencia al acortamiento de los periodos de desove.

La condición somática baja durante la estación de desove es una observación común en numerosas especies de peces y da indicación de que el crecimiento somático está limitado debido al desarrollo de las gónadas. Este decremento en condición somática se ha atribuido al consumo y agotamiento de las reservas corporales durante la maduración de las gónadas (Htun-Han, 1978). El decremento en el peso del hígado pudo haberse debido al paso de materiales de reserva del hígado a las gónadas como Larson (1974) encontró en la trucha *Salvelinus fontinalis*.

El uso del análisis multivariado de covarianza tienen algunas ventajas. Por ejemplo, un análisis preliminar (no incluido aquí) del índice gonadosomático clásico mostró una heterocedasticidad marcada que hace difícil las comparaciones. Al utilizar esta propuesta nueva este problema no existió en lo absoluto. Además, al considerar a la longitud estándar



como covariada, todas las medias fueron ajustadas al tamaño de los peces. El método y las comparaciones múltiples de valores adyacentes permiten evaluación de la significancia estadística a lo largo de las fechas de muestreo.

Es posible proponer las siguientes generalizaciones del análisis presentado anteriormente:

- El análisis multivariado de covarianza es un método muy valioso que permite estudiar adecuadamente el ciclo reproductivo del “suzuki” *Lateolabrax japonicus*.
- El período de desove, indicado por los valores medios ajustados más altos de gónadas y bajos de hígado y eviscerado, ocurre de noviembre a febrero como lo indican las muestras de la captura comercial de la Bahía de Tokio.
- Un pico extraordinario en el valor medio ajustado de las gónadas se detectó en abril. La causa de este valor alto no usual no pudo investigarse adicionalmente por la carencia de información adicional sobre el lugar preciso de captura.
- El análisis de las medias gonádicas por grupo de edad puede indicar que los machos, y en especial las hembras de *Lateolabrax japonicus* se reproducen después de alcanzar los dos años de edad. Lo anterior concuerda con los reportes previos por Kosaka (1969) y Hayashi (1971; 1972).

Se encontró una tendencia opuesta entre los valores medios de las gónadas y los valores medios de peso hepático y somático (eviscerado), lo que sugiere el enfoque del crecimiento hacia el desarrollo de las gónadas durante la estación reproductiva.

El ANMUCOVA puede aplicarse para el estudio de otros índices morfométricos, como por ejemplo el índice de repleción.

6.3. El caso del gasterópodo marino *Hexaplex princeps*

Este apartado se basa en la investigación doctoral de la segunda autora del presente libro (Saito-Quezada, 2018).

En el litoral del Pacífico mexicano se explotan artesanalmente varias especies de gasterópodos, entre los que se encuentra el comúnmente llamado “caracol chino” llamado así por la ornamentación de su concha. *Hexaplex princeps* (su nombre científico) es un



gasterópodo Murícido que se distribuye desde el Golfo de California hasta Perú. Su carne es apreciada para su consumo como alimento y sus conchas se utilizan para elaborar artesanías ornamentales.

Debido a la anatomía diferente y peculiar de los gasterópodos, sobre todo en relación con organismos vertebrados como los peces, el estudio de sus órganos debe hacerse adaptando los métodos tradicionales utilizados principalmente para especies ícticas. Por ejemplo, se sabe que en algunos gasterópodos relacionados con *Hexaplex princeps* la gónada está embebida en la glándula digestiva, la cual en cierta forma equivale al hígado y páncreas de los vertebrados, dando lugar al complejo espiral visceral (Vasconcelos *et al.*, 2008; Gharsallah *et al.*, 2010; Elhasni *et al.*, 2010).

6.3.1. Material y método

Los individuos se obtuvieron con periodicidad aproximadamente mensual de febrero de 2012 a noviembre de 2014 con la ayuda de buzos locales en varias localidades con litoral rocoso aledañas a Puerto Ángel, Oaxaca, México. El Cuadro 6.3 presenta un resumen del total de especímenes recolectados: en total 774, 389 hembras y 385 machos.

Para el Índice de Condición (*IC*) se consideró el peso del callo (pie) expresado como un porcentaje del peso corporal, para tener el seguimiento del grado de bienestar de los individuos en el periodo de estudio.

El Índice Gónada-Glándula (*IGG*) se calculó con el peso de la espiral visceral (gónada+glándula digestiva) expresado como un porcentaje del peso corporal (Salgado-Ugarte *et al.*, 2005; Vasconcelos *et al.*, 2012).

Como alternativa a los índices morfofisiológicos (*IC*, *IGG*) se utilizó el **Análisis Multivariado de Covarianza (ANMUCOVA)** y se analizaron las medias mensuales ajustadas por tamaño (largo de la concha como covariada) de los pesos de la espiral visceral y el peso del callo. Para asegurar la linealidad de las relaciones, la homogeneización de las varianzas y evitar el uso de valores negativos se utilizó la transformación logarítmica (neperiana) y multiplicación de los valores originales por una constante (10^6) (García-Berthou y Moreno Amich, 1993; Salgado-Ugarte, 1995; Salgado-Ugarte *et al.*, 2005).



Cuadro 6.3. Resumen del número de individuos por sexo y por fecha.

Fecha	Sexo		Total
	Hembras	Machos	
23/02/2012	11	15	26
22/03/2012	16	18	34
26/04/2012	5	6	11
18/05/2012	4	9	13
15/06/2012	14	9	23
20/07/2012	2	6	8
23/08/2012	17	8	25
28/09/2012	67	42	109
26/10/2012	12	18	30
24/11/2012	12	13	25
26/04/2013	26	20	46
23/08/2013	4	10	14
25/10/2013	5	9	14
22/11/2013	22	16	38
24/01/2014	53	40	93
21/02/2014	0	2	2
21/03/2014	10	16	26
25/04/2014	27	33	60
23/05/2014	26	27	53
15/08/2014	17	25	42
24/10/2014	9	4	13
28/11/2014	25	32	57
21/12/2014	5	7	12
Total	389	385	774



Para reducir el ruido y aclarar los patrones en las tendencias de los valores de los índices y las medias ajustadas se empleó un suavizador no lineal resistente, especificado por las siguientes siglas: 4253eh, doble, que significa el uso de medianas corredizas de amplitud 4, seguidas por medianas corredizas de amplitud 2, 5 y 3; ajuste de valor terminal (e), promedio móvil ponderado “Hanning” y aplicación del mismo tratamiento (4253eh) a los residuales para adicionarlos a los valores obtenidos (doble) según lo recomendado por Tukey (1977), Velleman (1980), Velleman y Hoaglin (1981), Goodall (1990), Salgado-Ugarte y Curts-García (1992, 1993), Salgado-Ugarte (1992, 2013, 2017). Al requerir estos procedimientos secuencias completas faltantes se utilizó un procedimiento de interpolación simple (media aritmética) entre valores adyacentes a los faltantes.

Cuadro 6.4. Resumen del número de individuos por sexo y por fecha para el cálculo del Índice de Condición (IC) y el Índice Gónada-Glándula (IGG).

Fecha	Sexo					
	Hembras		Machos		Total	
	IC	IGG	IC	IGG	IC	IGG
23/02/2012	11	-	15	-	26	-
22/03/2012	16	-	18	-	34	-
26/04/2012	-	-	-	-	-	-
18/05/2012	4	-	9	-	13	-
15/06/2012	14	-	9	-	23	-
20/07/2012	2	-	6	-	8	-
23/08/2012	17	-	8	-	25	-
28/09/2012	67	-	42	-	109	-
26/10/2012	12	-	18	-	30	-
24/11/2012	12	12	13	13	25	25
26/04/2013	26	26	20	20	46	46
23/08/2013	4	4	10	10	14	14
25/10/2013	5	5	9	9	14	14
22/11/2013	22	22	16	16	38	38
24/01/2014	53	53	40	40	93	93
21/02/2014	0	0	2	2	2	2



Cuadro 6.4. Resumen del número de individuos por sexo y por fecha para el cálculo del Índice de Condición (IC) y el Índice Gónada-Glándula (IGG) (continuación).

Fecha	Sexo					
	Hembras		Machos		Total	
	IC	IGG	IC	IGG	IC	IGG
21/03/2014	10	10	16	16	26	26
25/04/2014	27	27	33	33	60	60
23/05/2014	26	26	27	27	53	53
15/08/2014	17	17	25	25	42	42
24/10/2014	9	9	4	4	13	13
28/11/2014	25	25	32	32	57	57
21/12/2014	5	5	7	7	12	12
Total	384	241	379	254	763	495

6.3.2. Índices morfofisiológicos

Se inició con el análisis de los índices morfofisiológicos. Se presentan los índices de condición (peso callo/ peso total × 100) y el índice gónadaglándula (peso de la espiral visceral / peso total × 100). El Cuadro 6.4 muestra el número de organismos para los cuales fue posible el cálculo de los índices respectivos. Este número varió debido a que en algunas ocasiones no fue posible pesar todas las estructuras involucradas.

La Figura 6.16 muestra el comportamiento del índice de condición para el total de los individuos; la Figura 6.17 el índice gónada-glándula y la Figura 6.18 compara los valores medianos de cada fecha de ambos índices.

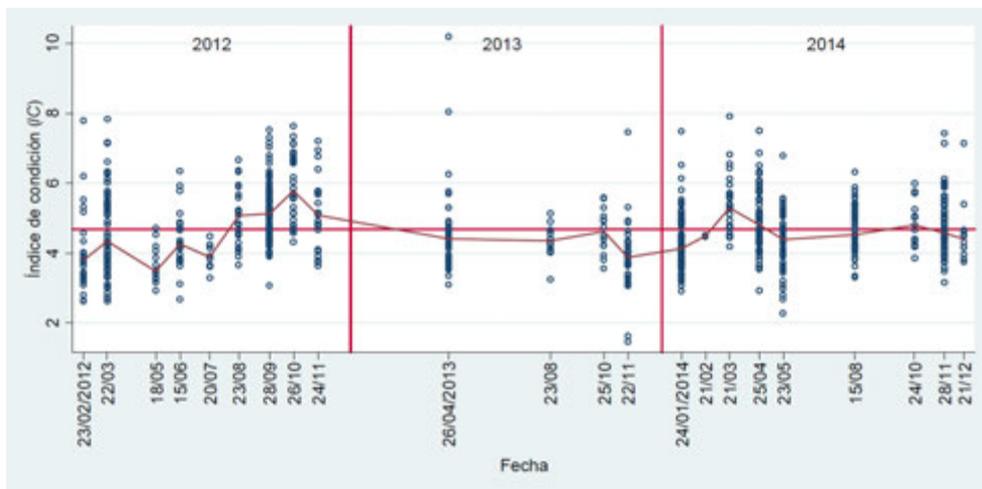


Figura 6.16. Índice de condición (IC) para el total de individuos de *Hexaplex princeps* por fecha de colecta. Los valores medianos de cada colecta están conectados linealmente. La línea horizontal es el promedio total.

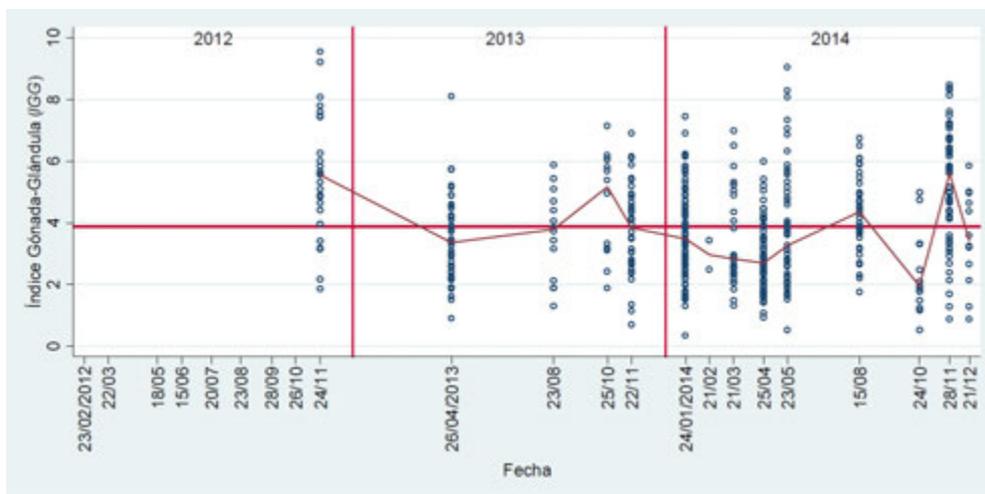


Figura 6.17. Índice Gónada-Glándula (IGG) para el total de individuos de *H. princeps* por fecha de colecta. Los valores medianos de cada colecta están unidos linealmente y la línea horizontal es el promedio total.

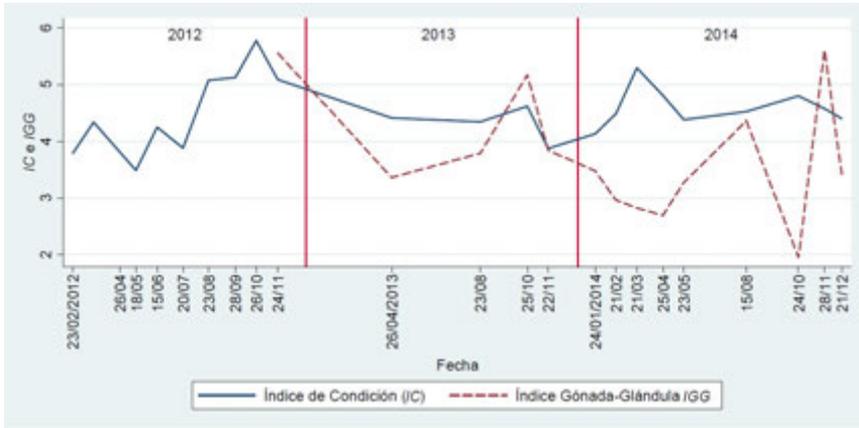


Figura 6.18. Gráfico de comparación de los valores medianos de los dos índices (IC e IGG) considerados.

El análisis de los índices se llevó a cabo también para cada sexo. Los gráficos de los valores puntuales y medianos (conectados por una línea) por fecha de colecta se presentan en las siguientes Figuras: Condición 6.19 y 6.20; Gónada-Glándula 6.21 y 6.22.

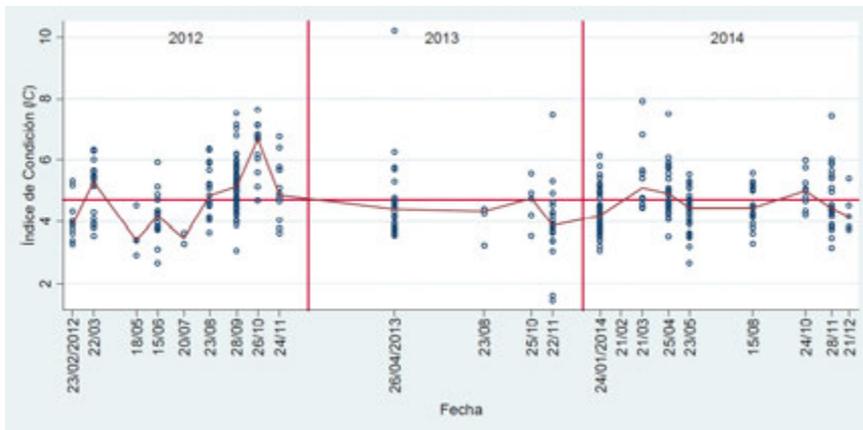


Figura 6.19. Índice de condición (IC) para las hembras de *Hexaplex princeps* por fecha de colecta. Los valores medianos de cada colecta están conectados linealmente y la línea horizontal es el promedio total.

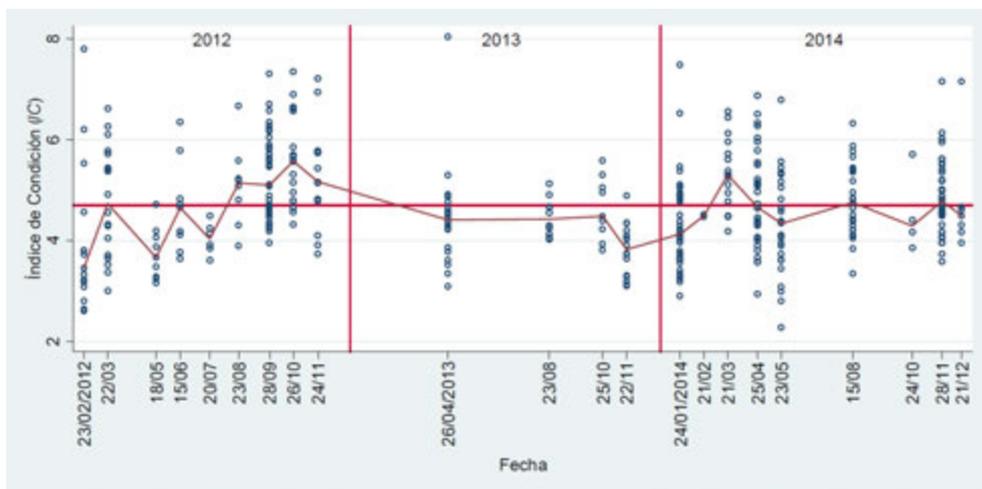


Figura 6.20. Índice de condición (IC) para los machos de *Hexaplex princeps* por fecha de colecta. La línea une los valores medianos de cada colecta.

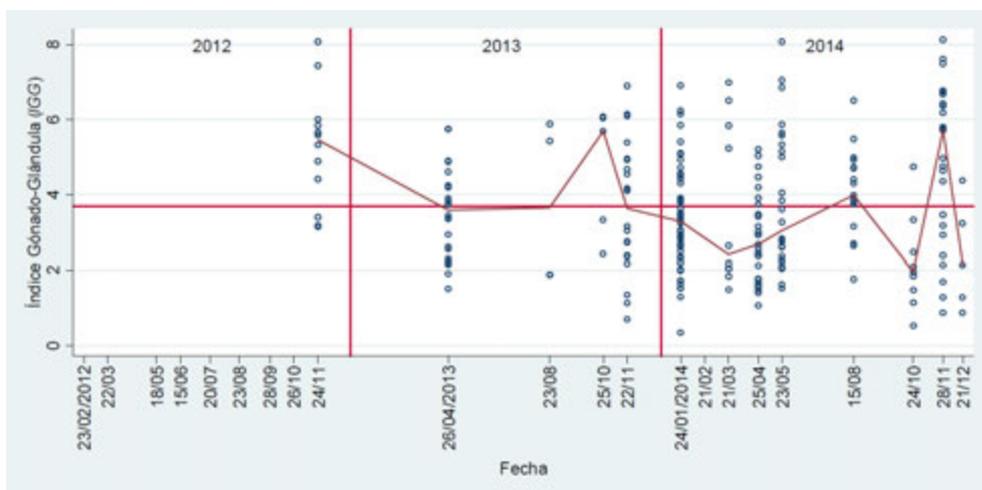


Figura 6.21. Índice Gónada-Glándula para las hembras de *Hexaplex princeps* por fecha de colecta. Los valores medianos de cada colecta están conectados linealmente y la línea horizontal es el valor promedio.

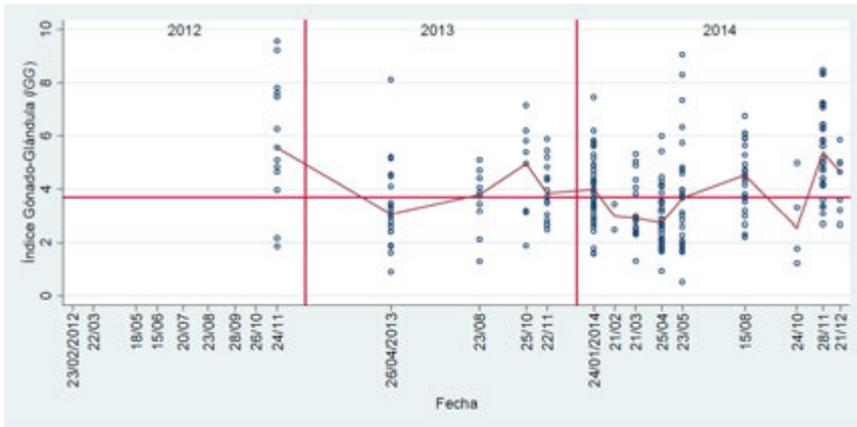


Figura 6.22. Índice Gónada-Glándula para los machos de *Hexaplex princeps* por fecha de colecta. Los valores medianos de cada colecta están conectados linealmente y la línea horizontal es el valor promedio.

A continuación se incluyen las gráficas comparativas entre los índices de condición y de la gónada - glándula en las hembras y machos de *H. princeps* (Figuras 6.23 y 6.24)

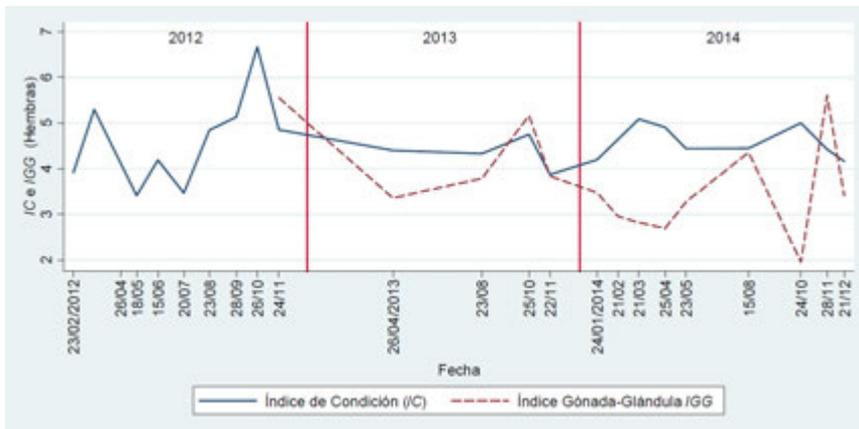


Figura 6.23. Comparación de los valores medianos del Índice de Condición (IC) y del Índice Gónada-Glándula (IGG) para las hembras de *Hexaplex princeps* por fecha de colecta.

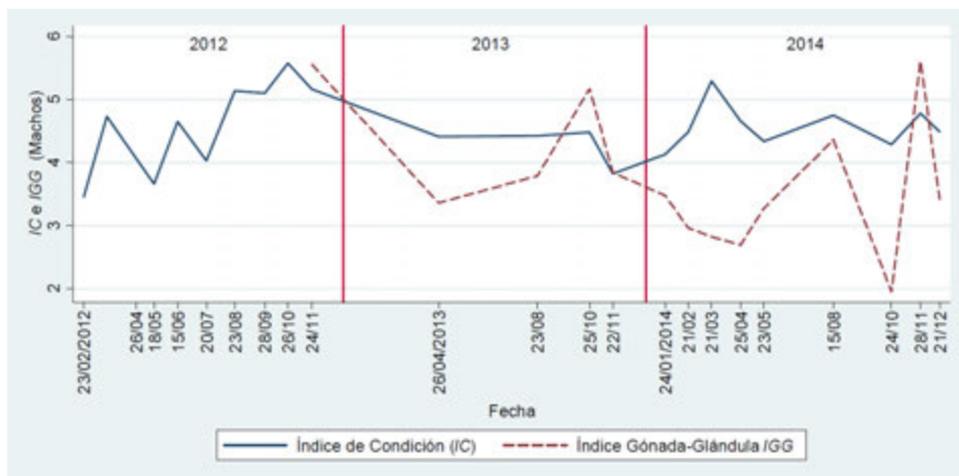


Figura 6.24. Comparación de los valores medianos del Índice de Condición (IC) y del Índice Gónada-Glándula (IGG) para los machos de *Hexaplex princeps* por fecha de colecta.

Cuadro 6.5. Número de individuos considerados para el Análisis Multivariado de Covarianza (ANMUCOVA).

Fecha	Hembras	Machos	Total
26/04/2013	26	20	46
23/08/2013	4	10	14
25/10/2013	5	9	14
22/11/2013	22	16	38
24/01/2014	53	40	93
21/02/2014	0	2	2
21/03/2014	10	16	26
25/04/2014	27	33	60
23/05/2014	26	27	53
15/08/2014	17	25	42
24/10/2014	9	4	13
28/11/2014	25	32	57
21/12/2014	5	7	12
Total	229	241	470



6.3.3. Análisis multivariado de covarianza

El número de individuos considerados para este análisis se muestran en la Cuadro 6.5. En ella se incluyen aquellos en las que fue posible medir las variables a considerar: largo de la concha, peso de la gónada-glándula, peso del callo y peso blando.

En este análisis, las variables de respuesta son el peso de la gónada-glándula, el peso del callo y el peso de las partes blandas; la covariada es el largo de la concha y el factor es la fecha de recolecta.

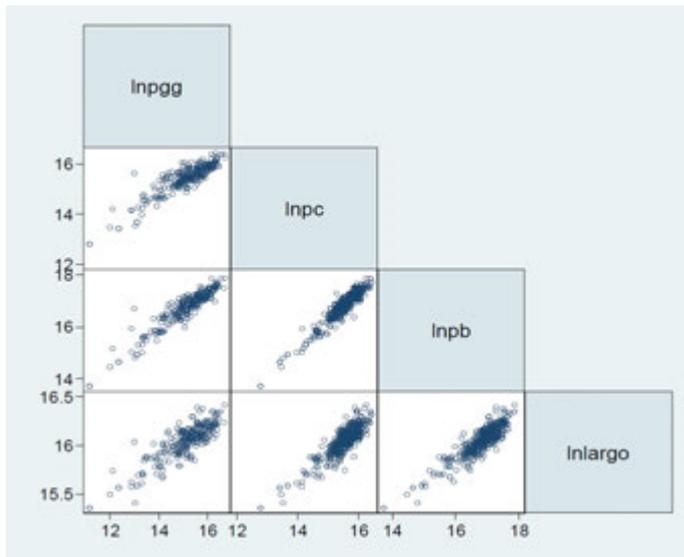


Figura 6.25. Matriz de gráficos de dispersión (mitad) para los datos de las hembras de *H. princeps*; lnpgg = Logaritmo natural del peso de gónada-glándula; lnpc = Logaritmo natural del peso de callo; lnpb = Logaritmo natural del peso blando; ln largo = Logaritmo natural del largo de la concha.

En el diseño preliminar, en el análisis de las hembras la interacción de factor con covariada resultó significativa y se incluyó en el análisis. Para los machos, esta interacción no fue significativa, por lo que en el diseño final, ésta no se incluyó.

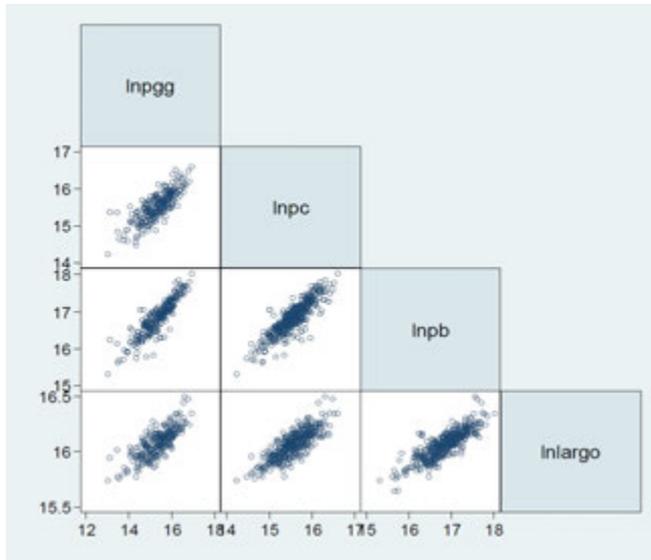


Figura 6.26. Matriz de gráficos de dispersión (mitad) para los datos de los machos de *H. princeps*; lnpg = Logaritmo natural del peso de gónada; lnpc = Logaritmo natural del peso de callo; lnpb = Logaritmo natural del peso blando; lnlar = Logaritmo natural del largo de la concha.

Posterior al comando para el análisis multivariado de covarianza (manova), se utilizaron las rutinas posteriores a la estimación: “margins” en combinación con la variable factor (en este caso nombrada “días”) para obtener los valores medios ajustados; “margins” en combinación con el prefijo “a” para el factor que produce la comparación de los valores medios sucesivos.

Estos comandos producen los valores medios ajustados para cada una de las variables de respuesta en el diseño y sus contrastes entre valores sucesivos (ver Sección 6.3 abajo). Esta información se ha resumido en las Figuras 6.27 (hembras) y 6.28 (machos).



Cuadro 6.6. Análisis Multivariado de Covarianza para datos de hembras de *H. princeps*, N = 241.

Fuente	Criterio		gl	gl(F)	F	P	
Modelo	W	0.0530	25	(75,637.6)	14.20	0.0000	a
	P	1.4306		(75,645)	7.84	0.0000	a
	L	9.9238		(75,635)	28.01	0.0000	a
	R	9.2051		(25,215)	79.16	0.0000	u
Residual			215				
Fecha	W	0.7684	12	(36,630.1)	1.63	0.0124	a
	P	0.2481		(36,645)	1.62	0.0140	a
	L	0.2803		(36,635)	1.65	0.0110	a
	R	0.1776		(12,215)	3.18	0.0003	u
LnLargo	W	0.3684	1	(3,213)	121.73	0.0000	e
	P	0.6316		(3,213)	121.73	0.0000	e
	L	1.7146		(3,213)	121.73	0.0000	e
	R	1.7146		(3,213)	121.73	0.0000	e
Interacción	W	0.7707	12	(36,630.1)	1.61	0.0143	a
	P	0.2455		(36,630.1)	1.60	0.0160	a
	L	0.2769		(36,630.1)	1.63	0.0127	a
	R	0.1751		(12,215)	3.14	0.0004	u
Residual			215				
Total			240				

e = exacto, a = aproximado, u = límite superior en F.

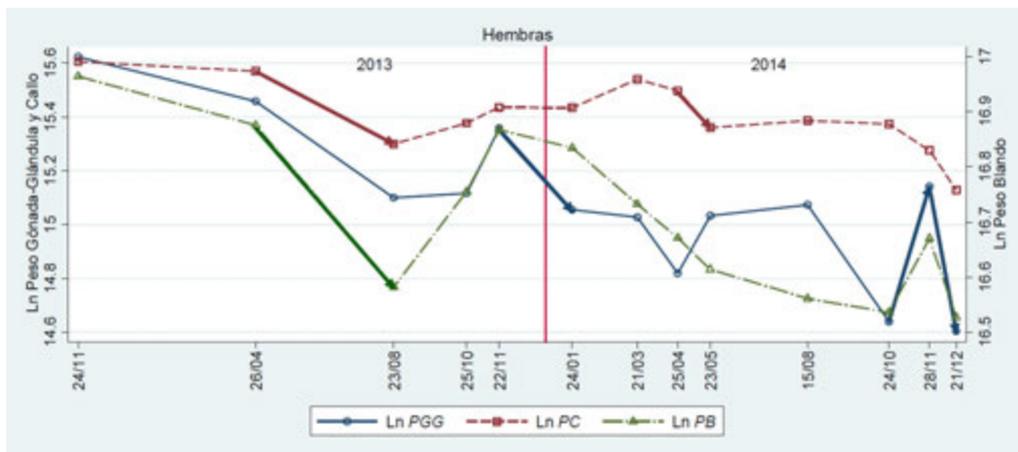


Figura 6.27. Medias ajustadas por covariada (Largo de la concha) de los pesos de gónada-glándula, callo y blando (en escala logarítmica) por fecha de colecta para las hembras de *H. princeps*. Las flechas indican sentido y significancia estadística del cambio de una fecha a otra.

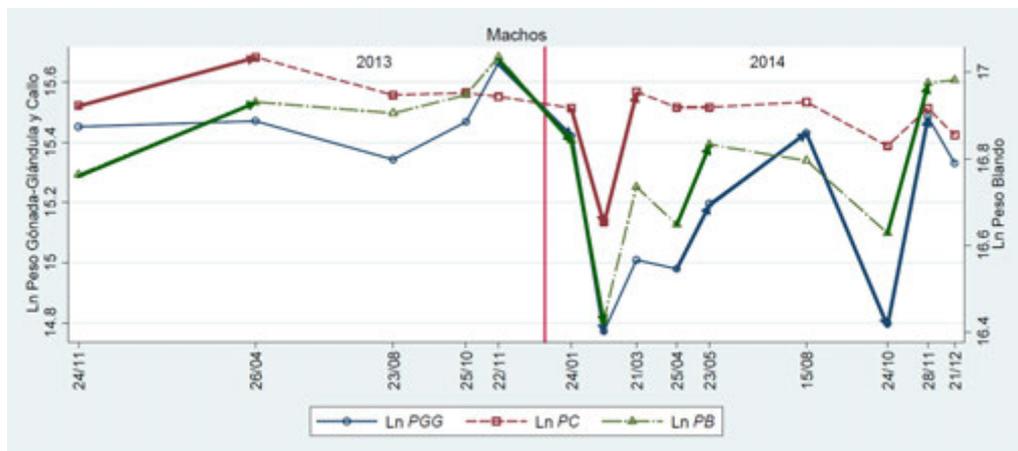


Figura 6.28. Medias ajustadas por covariada (Largo de la concha) de los pesos de gónada-glándula, callo y blando (en escala logarítmica) por fecha de colecta para los machos de *H. princeps*. Las flechas indican sentido y significancia estadística del cambio de una fecha a otra.



Cuadro 6.7. Análisis Multivariado de Covarianza (sin interacción) para datos de machos de *H. princeps*, N = 254.

Fuente	Criterio		gl	gfl	F	P	
Modelo	W	0.1200	14	(42,703.8)	17.49	0.0000	a
	P	1.1952		(42,717)	11.31	0.0000	a
	L	4.9223		(42,707)	27.62	0.0000	a
	R	4.4386		(14,239)	75.77	0.0000	u
Residual			239				
Fecha	W	0.5485	13	(39,702.6)	4.05	0.0000	a
	P	0.5269		(39,717)	3.92	0.0000	a
	L	0.6924		(39,707)	4.18	0.0000	a
	R	0.4471		(13,239)	8.22	0.0000	u
Ln Largo	W	0.2076	1	(3,237)	301.57	0.0000	e
	P	0.7924		(3,237)	301.57	0.0000	e
	L	3.8173		(3,237)	301.57	0.0000	e
	R	3.8173		(3,237)	301.57	0.0000	e
Residual			239				
Total			253				

e = exacto, a = aproximado, u = límite superior en F.

6.3.4. Valores suavizados

En las Figuras 6.29 y 6.30 se presenta una comparación entre el Índice de Condición (IC) el Índice Gónada-Glándula (IGG) para las hembras y los machos (respectivamente) de *H. princeps*.

Medias ajustadas suavizadas

En las siguientes figuras se comparan las medias ajustadas por covariada del ANMUCOVA (Figuras 6.31 y 6.32). En general puede apreciarse, más claramente en los machos, un comportamiento opuesto entre las medias ajustadas del peso de las gónadas con la condición con máximos gonádicos antes de los meses fríos (noviembre) y mínimos posteriormente (febrero, marzo) con una tendencia semejante pero adelantada en el peso de la glándula digestiva, con mínimos en enero-febrero.



Figura 6.29. Comparación de valores suavizados (4253EH,doble) del Índice de Condición y del Índice Gónada-Glándula por fecha de colecta para las hembras de *H. princeps*.



Figura 6.30. Comparación de valores suavizados (4253EH,doble) del Índice de Condición e Índice Gónada-Glándula por fecha de colecta para los machos de *H. princeps*.



Figura 6.31. Comparación de valores suavizados (4253EH,doble) de las medias ajustadas del (en escala logarítmica) peso del callo (PC) y del peso de Gónada-Glándula (PGG) por fecha de colecta para las hembras de *H. princeps*.



Figura 6.32. Comparación de valores suavizados (4253EH,doble) de las medias ajustadas del (en escala logarítmica) peso del callo (PC) y del peso Gónada-Glándula (PGG) por fecha de colecta para los machos de *H. princeps*.



6.3.5. Comparación de índices con Temperatura y Clorofila

A continuación se presentan los valores suavizados de los índices morfofisiológicos con la temperatura superficial del océano y la concentración de clorofila "a" en la zona. Las Figuras 6.33 y 6.34 muestran la variación del Índice de Condición a lo largo del periodo de estudio para hembras y machos (respectivamente).

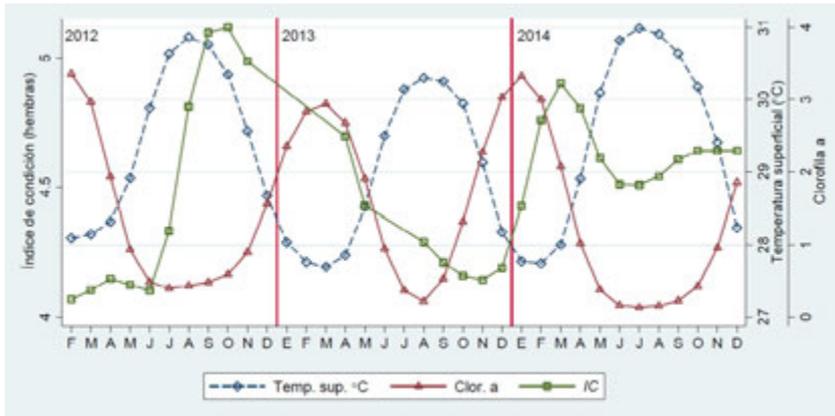


Figura 6.33. Índice de Condición, Temperatura superficial y Concentración de Clorofila a por fecha de colecta para hembras de *H. princeps*.

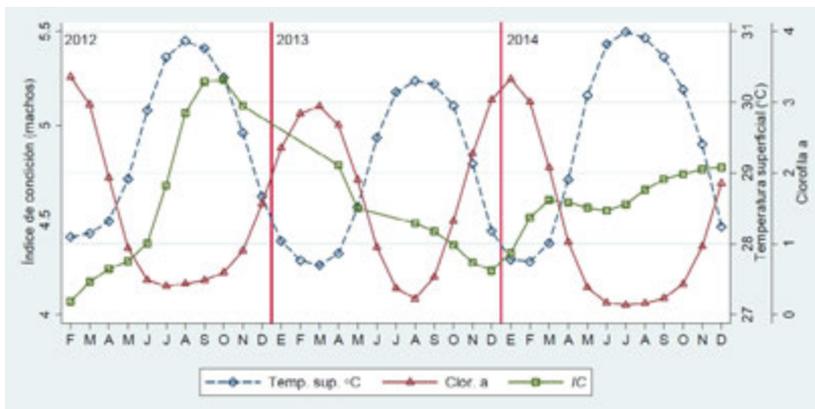


Figura 6.34. Índice de Condición, Temperatura superficial y Concentración de Clorofila a por fecha de colecta para machos de *H. princeps*.



El índice gónada-glándula suavizado tanto para hembras como para machos mostró valores altos durante los meses cálidos del año con máximos en noviembre y mínimos en enero y febrero, lo que sugiere una reproducción invernal (Figuras 6.35 y 6.36).

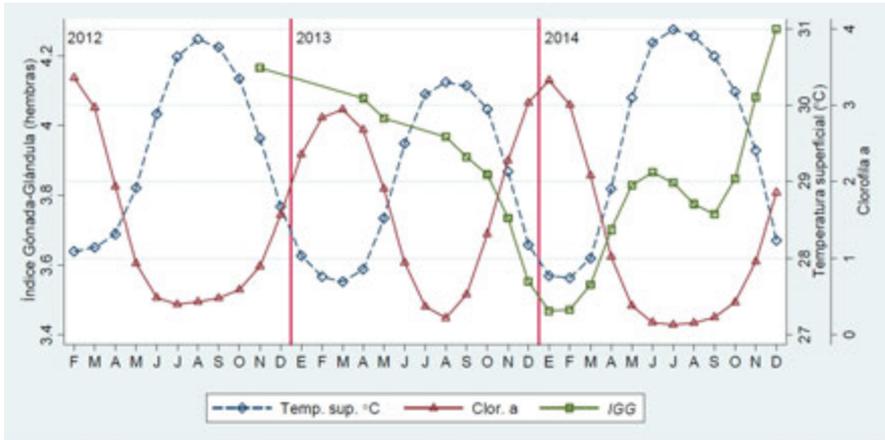


Figura 6.35. Índice Gónada-Glándula, Temperatura superficial y Concentración de Clorofila a por fecha de colecta para hembras de *H. princeps*.

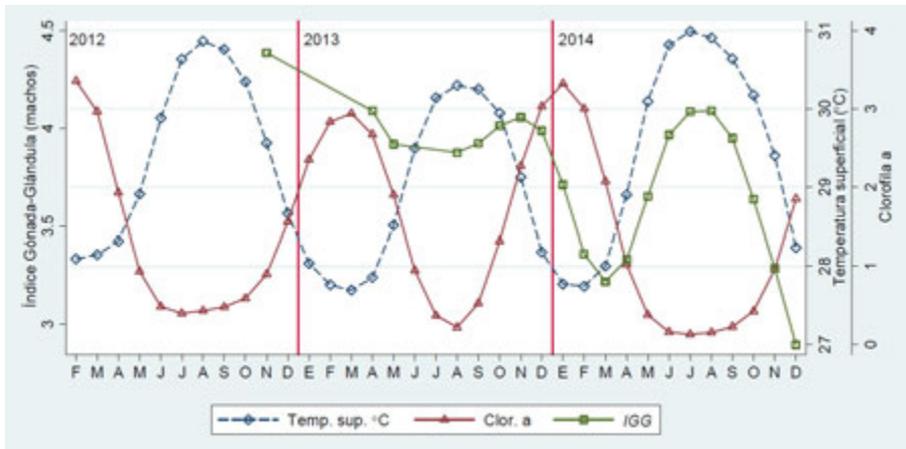


Figura 6.36. Índice Gónada-Glándula, Temperatura superficial y Concentración de Clorofila a por fecha de colecta para machos de *H. princeps*.



6.3.6. Consideraciones acerca de los resultados

Por lo general el *IGG* siempre estuvo por debajo del *IC*, sin embargo de manera contraria, en 2012 el valor de *IGG* en noviembre estuvo por encima del *IC*; en 2013 cuando el *IC* fue constante se observó que el *IGG* bajó y en el caso de octubre los valores de *IGG* y *IC* fueron altos con el *IGG* por encima; en 2014 se observó un comportamiento definido de cuando el *IC* es alto el *IGG* baja tanto en marzo, abril y octubre. En este mismo año en noviembre el mayor valor de *IGG* estuvo por arriba del *IC*, lo que sugirió que la reproducción se lleva a cabo en los meses más fríos y secos (invierno). En contraste con otros gasterópodos del mismo género (*Hexaplex trunculus*), Elhasni *et al.* (2010), en el sur de Túnez, por medio del índice gónada-glándula observaron que la reproducción se llevaba a cabo a finales de la primavera y principios de verano. Vasconcelos *et al.* (2008), reportaron que *H. trunculus* en Portugal, de acuerdo con el *IG* (complejo espiral visceral) la reproducción se presentaba en dos eventos: el primero a principios de primavera y el segundo (más intenso) en octubre. En Puerto Ángel, Oax. se observó la correspondencia de valores bajos de *IC* con valores altos de *IGG*, lo que concordó con lo reportado para *H. trunculus* tanto en Portugal (Mar Mediterráneo) como en Túnez (Océano Atlántico).

El comportamiento de los índices de condición para cada sexo mostró comportamiento semejante con valores altos de condición en el mes de octubre de cada año (2012 a 2014) y bajos a inicios y fin de año. El índice gónada-glándula fue más variable para las hembras pero con tendencias semejantes a las descritas previamente, con valores altos en octubre de 2013 y noviembre de 2014 y bajos en enero y octubre de 2014, en el caso de *H. trunculus* en Túnez, Elhasni *et al.* (2010) observaron que entre sexo existía asincronía de manera que cuando bajaba el *IC* era época en que no había reproducción, así mismo para Elhasni *et al.* (2010), donde ellos observaron que *H. trunculus* en Portugal tenía el mismo comportamiento del *IC*.

El *IGG* de las hembras presentó un valor alto en noviembre de 2012; para 2013 los valores se encontraron alrededor de la línea promedio con un valor alto en octubre; en 2014 los valores más bajos se observaron en marzo y octubre; en agosto los valores alcanzaron la línea media y el mayor valor se registró en noviembre. Para los machos, en 2012 noviembre tuvo valores altos. En 2013 abril manifestó los valores más bajos y los más altos en octubre; para 2014 los valores por debajo del promedio se observaron en febrero, marzo, abril y octubre mientras que los valores altos se observaron en enero, agosto y noviembre (el más alto). En otros estudios sobre el índice gonadosomático (complejo espiral visceral) de *H. trunculus* en Túnez, Elhasni *et al.* (2010) observaron sincronía entre sexos debido a que las hembras de esa especie guardan el esperma a semejanza de lo reportado por Vasconcelos *et al.* (2008) para esta especie en Portugal. En conclusión, se observó que ambos sexos presentan en general, una concordancia de valores altos de *IGG* con bajos de *IC* (y viceversa).



Se verificó la linealidad (en escala logarítmica) en las relaciones de las variables consideradas para el ANMUCOVA. A pesar de cierta dispersión de los datos, en general se aprecian claramente tendencias lineales de comportamiento en la relación de las variables consideradas, lo cual indica que el uso de este procedimiento es adecuado (Salgado-Ugarte *et al.*, 2005).

El diseño preliminar probado mostró que la interacción entre la covariada (longitud de la concha) con el factor (fecha de recolecta) fue significativo para las hembras y se dejó en el diseño final. Por otro lado, en los machos esta interacción no fue significativa por lo que se utilizó el análisis sin este término. Tanto en hembras como en machos, tanto el factor (fecha) como la covariada (largo de la concha) mostraron efectos estadísticos altamente significativos ($P < 0.001$; Cuadros 6.6 y 6.7). El peso medio de la gónada-glándula de las hembras disminuyó significativamente de noviembre 2013 a enero 2014, de agosto a octubre y noviembre a diciembre de 2014. La condición (peso del callo) disminuyó de abril a mayo y noviembre a diciembre de 2014. Los machos mostraron descensos significativos en el peso de la gónada-glándula de diciembre 2013 a febrero 2014. Valores altos se tuvieron en noviembre de ambos años. La condición, indicada por los valores medios del peso del callo, descendió significativamente de enero a febrero y subió de febrero a marzo de 2014.

La suavización aclaró de manera significativa el comportamiento de los valores de los índices morfofisiológicos. En las Figuras 6.31 6.32 puede apreciarse que a lo largo de los tres años de estudio ocurrieron variaciones en los ciclos, sin embargo pueden notarse comportamientos regulares. Tanto para las hembras como para los machos, valores bajos de condición ocurrieron en los meses fríos del año y se fueron recuperando hasta alcanzar valores altos antes de los mismos. En 2014 la condición fue alta también en los meses de primavera. El índice gónada-glándula mostró valores altos antes de los meses fríos y mínima después de los mismos, con una ligera tendencia a ser opuesta a la condición.

Las medias ajustadas suavizadas (Figuras 6.33 y 6.34) mostraron patrones más claros en el caso de los machos (Figura 6.34), con valores mínimos de peso de callo y gónada en los meses fríos e inicios de primavera y altos a fin de año. La condición fue alta durante los meses cálidos del año.

En dos de los tres años de estudio, valores bajos de condición se asociaron con altos de temperatura; valores altos de condición ocurrieron después o durante los altos de clorofila a. Los resultados parecen indicar una relación entre la clorofila a y el desarrollo gonadal y el desove, ya que los tres repuntes observados en los valores de clorofila a (finales de septiembre, mediados de enero y de mayo) coinciden con aumentos significativos en el



número de individuos en estos estados. En lo que respecta a las variaciones del índice de condición, valores máximos se relacionan con épocas de intenso desarrollo gametogénico o de crecimiento activo, mientras que los descensos suelen asociarse con liberación de gametos o con condiciones nutritivas desfavorables.

6.4. Comandos en Stata

Se decidió incluir una Sección aparte para la inclusión de los comandos de Stata utilizados para obtener los resultados arriba presentados. Debido a lo extenso de los comandos, se han incluido sólo algunos ejemplos representativos. Consideramos que el lector podrá usarlos como guía para llevar a cabo los propios.

```
. use suzukt15, clear
. gen pevi= totwei - (gonwei+livwei+stowe+intwei)
(149 missing values generated)
. gen lnle = log(bodlen*10^6)
. gen lnpe=log(pevi*10^6)
(149 missing values generated)
. gen lnpg=log(gonwei*10^6)
(149 missing values generated)
. gen lnph = log(livwei*10^6)
(128 missing values generated)
. graph matrix lnle lnpe lnpg lnph, half ms(Oh) msize(small) xlab(18.5 20)
diag(, size(vlarge))
```

Se obtiene la Figura 6.1.

```
. use manovrel
. tab edate2
```

edate2	Freq.	Percent	Cum.
320	1	6.25	6.25
338	1	6.25	12.50
355	1	6.25	18.75
367	1	6.25	25.00
397	1	6.25	31.25
468	1	6.25	37.50
493	1	6.25	43.75
522	1	6.25	50.00
551	1	6.25	56.25



593		1	6.25	62.50
647		1	6.25	68.75
732		1	6.25	75.00
760		1	6.25	81.25
832		1	6.25	87.50
864		1	6.25	93.75
923		1	6.25	100.00

Total		16	100.00	

```
. scatter meanlew meanlgw meanllw edate2 , c(1 1 1) ms(Oh Th Sh) xlab(320
"29/09" 338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "13/05" " 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12" 832
"18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1) rows(1)
lab(1 "Peso eviscerado") lab(2 "Peso gónada") lab(3 "Peso higado")) xtitle(" ")
ytitle("Ln del peso") xline(418 773) text(22 340 "1993") text(22 450 "1994")
text(22 810 "1995") xtitle(Fecha)
```

Se obtiene la Figura 6.2.

```
. scatter meanlgw edate2 , c(1) lcolor(maroon) ms(Th) mcolor(maroon) xlab(320
"29/09" 338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "1 3/05" 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12" 832
"18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1) rows(1)
lab(2 "Peso eviscerado") lab(1 "Peso gónada")) xtitle(" ") ytitle("Ln del peso
de la gónada") xline(418 773) text(18 380 "1993") text(18 450 "1994") text(18
810 "1995") xtitle(Fecha) || scatter meanlew edate2 , c(1) lcolor(navy) ms(Oh)
mcolor(navy) yaxis(2) ytitle("Ln del peso eviscerado", axis(2))
```

Se obtiene la Figura 6.3.

```
. use manovre2

. scatter meanlew meanlgw meanllw edate2 , c(1 1 1) ms(Oh Th Sh) xlab(320
"29/09" 338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "13/05" " 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12" 832
"18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1) rows(1)
lab(1 "Peso eviscerado") lab(2 "Peso gónada") lab(3 "Peso higado")) xtitle(" ")
ytitle("Ln del peso") xline(418 773) text(22 340 "1993") text(22 450 "1994")
text(22 810 "1995") xtitle(Fecha)
```

Se obtiene la Figura 6.4.

```
. scatter meanlgw edate2 , c(1) lcolor(maroon) ms(Th) mcolor(maroon) xlab(320
"29/09" 338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "13/05" 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12" 832
```



```
"18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1) rows(1)
lab(2 "Peso eviscerado") lab(1 "Peso gónada")) xtitle(" ") ytitle("Ln del peso
de la gónada") xline(418 773) text(20 380 "1993") text(20 450 "1994") text(20
810 "1995") xtitle(Fecha) yscale(range(13 20)) ylab(13(1)20) || scatter meanlew
edate2, c(1) lcolor(navy) ms(Oh) mcolor(navy) yaxis(2) ytitle("Ln del peso
eviscerado", axis(2)) yscale(range(20 21) axis(2)) ylab(20(.2)21, axis(2))
```

Se obtiene la Figura 6.5.

```
. use manovre3

. scatter meanlew meanlgw meanllw edate2 , c(1 1 1) ms(Oh Th Sh) xlab(320
"29/09" 338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "13/05" 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12" 832
"18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1) rows(1)
lab(1 "Peso eviscerado") lab(2 "Peso gónada") lab(3 "Peso hígado")) xtitle(" ")
ytitle("Ln del peso") xline(418 773) text(22 340 "1993") text(22 450 "1994")
text(22 810 "1995") xtitle(Fecha)
```

Se obtiene la Figura 6.6.

```
. scatter meanlgw edate2 , c(1) lcolor(maroon) ms(Th) mcolor(maroon) xlab(320
"29/09" 338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "13/05" 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12" 832
"18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1) rows(1)
lab(2 "Peso eviscerado") lab(1 "Peso gónada")) xtitle(" ") ytitle("Ln del peso
de la gónada") xline(418 773) text(20 380 "1993") text(20 450 "1994") text(20
810 "1995") xtitle(Fecha) yscale(range(14 20)) ylab(14(1)20) || scatter meanlew
edate2, c(1) lcolor(navy) ms(Oh) mcolor(navy) yaxis(2) ytitle("Ln del peso
eviscerado", axis(2)) yscale(range(20 21) axis(2)) ylab(20(.2)21, axis(2))
```

Se obtiene la Figura 6.7.

Para *Hexaplex princeps*

```
. use muricantotalc
. sum fc
```

Variable	Obs	Mean	Std. Dev.	Min	Max
fc	793	4.69532	1.008311	1.456996	10.20625

```
. scatter fc dias, ms(oh) xlab(0 "23/02/2012" 28 "22/03" 85 "18/05" 113 "15/06"
148 "20/07" 182 "23/08" 218 "28/09" 246 "26/10" 275 "24/11" 428 "26/04/2013"
547 "23/08" 610 "25/10" 638 "22/11" 701 "24/01/2014" 729 "21/02" 757 "21/03"
```



```
792 "25/04" 820 "23/05" 904 "15/08" 974 "24/10" 1009 "28/11" 1032 "21/12",
angle(90)) || mband fc dias, bands(45) ytitle("Índice de condición ({it:IC})")
xtitle(Fecha) legend(off) xline(313 678) text(10 150 "2012") text(10 490 "2013")
text(10 870 "2014") xsize(8) yline(4.69)
```

Se obtiene la Figura 6.16.

```
. sum igs
```

Variable	Obs	Mean	Std. Dev.	Min	Max
igs	495	3.883012	1.718986	.3424658	9.555556

```
. scatter igs dias, ms(oh) xlab(0 "23/02/2012" 28 "22/03" 85 "18/05" 113
"15/06" 148 "20/07" 182 "23/08" 218 "28/09" 246 "26/10" 275 "24/11" 428
"26/04/2013" 547 "23/08" 610 "25/10" 638 "22/11" 701 "24/01/2014" 729 "21/02"
757 "21/03" 792 "25/04" 820 "23/05" 904 "15/08" 974 "24/10" 1009 "28/11"
1032 "21/12", angle(90)) || mband igs dias, bands(45) ytitle("Índice Gónada-
Glándula ({it:IGG})") xtitle(Fecha) legend(off) xline(313 678) text(10 150
"2012") text(10 490 "2013") text(10 870 "2014") xsize(8) yline(3.88)
```

Se obtiene la Figura 6.17.

```
two mband fc dias, lw(medthick) bands(45) ytitle({it:IC} e {it:IGG}) xlab(0
"23/02/2012" 63 "26/04" 85 "18/05" 113 "15/06" 148 "20/07" 182 "23/08" 218
"28/09" 246 "26/10" 275 "24/11" 428 "26/04/2013" 547 "23/08" 610 "25/10" 638
"22/11" 701 "24/01/2014" 729 "21/02" 757 "21/03" 792 "25/04" 820 "23/05" 904
"15/08" 974 "24/10" 1009 "28/11" 1032 "21/12", angle(90)) || mband igs dias,
bands(45) lw(medthick) lpattern(dash) bands(45) lw(medthick) xtitle(Fecha)
xline(313 678) text(5.9 150 "2012") text(5.9 490 "2013") text(5.9 870 "2014")
legend(rows(1) label(1 "Índice de Condición ({it:IC})") label(2 "Índice Gónada-
Glándula {it:IGG}") ) xsize(8)
```

Se obtiene la Figura 6.18.

```
. sum fc if sex==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
fc	384	4.712266	.9938835	1.456996	10.20625

```
. scatter fc dias if sex==1, ms(oh) xlab(0 "23/02/2012" 28 "22/03" 85 "18/05"
113 "15/06" 148 "20/07" 182 "23/08" 218 "28/09" 246 "26/10" 275 "24/11"
> 428 "26/04/2013" 547 "23/08" 610 "25/10" 638 "22/11" 701 "24/01/2014" 729
"21/02" 757 "21/03" 792 "25/04" 820 "23/05" 904 "15/08" 974 "24/10" 1009
```

Métodos cuantitativos computarizados para biología pesquera



```
> "28/11" 1032 "21/12", angle(90)) || mband fc dias if sex==1, bands(45)
ytittle("Índice de Condición ({it:IC})") xtitle(Fecha) legend(off) xline(313
```

```
> 678) text(10 150 "2012") text(10 490 "2013") text(10 870 "2014") xsize(8)
yline(4.71)
```

```
. sum fc if sex==2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
fc	379	4.709252	.9879738	2.273929	8.044771

```
. scatter fc dias if sex==2, ms(oh) xlab(0 "23/02/2012" 28 "22/03" 85 "18/05"
113 "15/06" 148 "20/07" 182 "23/08" 218 "28/09" 246 "26/10" 275 "24/11"
> 428 "26/04/2013" 547 "23/08" 610 "25/10" 638 "22/11" 701 "24/01/2014" 729
"21/02" 757 "21/03" 792 "25/04" 820 "23/05" 904 "15/08" 974 "24/10" 1009
> "28/11" 1032 "21/12", angle(90)) || mband fc dias if sex==2, bands(45)
ytittle("Índice de Condición ({it:IC})") xtitle(Fecha) legend(off) xline(313
> 678) text(8 150 "2012") text(8 490 "2013") text(8 870 "2014") xsize(8)
yline(4.7)
```

Se obtienen las Figuras 6.19 y 6.20.

```
. two mband fc dias if sex==1, lw(medthick) bands(45) ytittle("{it:IC} e {it:IGG}
) (Hembras)") xlab(0 "23/02/2012" 63 "26/04" 85 "18/05" 113 "15/06" 148 "20/07"
182 "23/08" 218 "28/09" 246 "26/10" 275 "24/11" 428 "26/04/2013" 547 "23/08"
610 "25/10" 638 "22/11" 701 "24/01/2014" 729 "21/02" 757 "21/03" 792 "25/04"
820 "23/05" 904 "15/08" 974 "24/10" 1009 "28/11" 1032 "21/12", angle(90)) ||
mband igs dias, bands(45) lw(medthick) lpattern(dash) bands(45) lw(medthick)
xtitle(Fecha) xline(313 678) text(6.9 150 "2012") text(6.9 490 "2013") text(6.9
870 "2014") legend(rows(1) label(1 "Índice de Condición ({it:IC})") label(2
"Índice Gónada-Glándula {it:IGG}") ) xsize(8)
```

Se obtiene la Figura 6.23.

```
. two mband fc dias if sex==2, lw(medthick) bands(45) ytittle("{it:IC} e {it:IGG}
) (Machos)") xlab(0 "23/02/2012" 63 "26/04" 85 "18/05" 113 "15/06" 148 "20/07"
182 "23/08" 218 "28/09" 246 "26/10" 275 "24/11" 428 "26/04/2013" 547 "23/08"
610 "25/10" 638 "22/11" 701 "24/01/2014" 729 "21/02" 757 "21/03" 792 "25/04"
820 "23/05" 904 "15/08" 974 "24/10" 1009 "28/11" 1032 "21/12", angle(90)) ||
mband igs dias, bands(45) lw(medthick) lpattern(dash) bands(45) lw(medthick)
xtitle(Fecha) xline(313 678) text(5.9 150 "2012") text(5.9 490 "2013") text(5.9
870 "2014") legend(rows(1) label(1 "Índice de Condición ({it:IC})") label(2
"Índice Gónada-Glándula {it:IGG}") ) xsize(8)
```



Se obtiene la Figura 6.24.

```
. use muricantotalc
. manova lnpg lnpc lnps = dias c.lnlargo dias#c.lnlargo if sex==1
```

Diseño del análisis de covarianza con interacción para las hembras (código numérico = 1) de *H. princeps*

```
. margins dias
```

Se obtienen las medias ajustadas para las hembras.

```
. margins a.dias
```

Se obtienen los contrastes estadísticos entre medias consecutivas para las hembras

Para los machos (código numérico = 2)

```
. manova lnpg lnpc lnps = dias c.lnlargo if sex==2
. margins dias
```

Se obtienen las medias ajustadas para los machos.

```
. margins a.dias
```

Se obtienen los contrastes estadísticos entre medias consecutivas para los machos.

```
. use anmucovahpr
file anmucovahpr.dta saved

. scatter lnpg lnpc lnps dias, ytitle("Ln Peso Gónada-Glándula y Callo") ms(Oh Sh) c(1 1) lpattern(solid dash) lw(thick) xlab(275 "24/11" 428 "26/04" 547 "23/08" 610 "25/10" 638 "22/11" 701 "24/01" 757 "21/03" 792 "25/04" 820 "23/05" 904 "15/08" 974 "24/10" 1009 "28/11" 1032 "21/12", angle(90)) xline(678) text(15.6 530 "2013") text(15.6 870 "2014") legend(rows(1) label(1 "Ln {it:PGG}") label(2 "Ln {it:PC}") label(8 "Ln {it:PB}") order(1 2 8)) || scatter lnps lnpc dias, ytitle("Ln Peso Blando", axis(2)) ms(Th) yaxi(2) c(1) lpattern(longdash_dot) lw(thick) t1("Hembras") || pcarrowi 15.35 638 15.05 701, lc(navy) lw(thick) mcolor(navy) msize(2) barbsize(2) || pcarrowi 14.639 974 15.14 1009, lc(navy) lw(thick) mcolor(navy) msize(2) barbsize(2) || pcarrowi 15.14 1009 14.60 1032, lc(navy) lw(thick) mcolor(navy) msize(2) barbsize(2) || pcarrowi 15.57 428 15.3 547, lc(maroon) lw(thick) mcolor(maroon)
```



```
msize(2) barbsize(2) || pcarrowi 15.49 792 15.36 820, lc(maroon) lw(thick)
mcolor(maroon) msize(2) barbsize(2) || pcarrowi 16.87 428 16.58 547, yaxis(2)
lc(dkgreen) lw(thick) mcolor(dkgreen) msize(2) barbsize(2)
```

Se obtiene la Figura 6.27.

```
. scatter lnpgm lnpcm dias, ytitle("Ln Peso Gónada-Glándula y Callo") ms(Oh
Sh) c(1 1) lpattern(solid dash) lw(medthick medthick) xlab(275 "24/11" 428
"26/04" 547 "23/08" 610 "25/10" 638 "22/11" 701 "24/01" 757 "21/03" 792 "25/04"
820 "23/05" 904 "15/08" 974 "24/10" 1009 "28/11" 1032 "21/12", angle(90))
xline(678) text(15.67 530 "2013") text(15.67 870 "2014") legend(rows(1)
label(1 "Ln {it:PGG}") label(2 "Ln {it:PC}") label(12 "Ln {it:PB}") order(1 2
12)) || scatter lnpbm dias, ytitle("Ln Peso Blando", axis(2)) ms(Th) yaxis(2)
c(1) lpattern(longdash_dot) lw(medthick) t1("Machos") || pcarrowi 15.66 638
15.42 701, lc(navy) lw(thick) mcolor(navy) msize(2) barbsize(2) || pcarrowi
15.42 701 14.77 729, lc(navy) lw(thick) mcolor(navy) msize(2) barbsize(2)
|| pcarrowi 14.98 792 15.19 820, lc(navy) lw(thick) mcolor(navy) msize(2)
barbsize(2) || pcarrowi 15.19 820 15.43 904, lc(navy) lw(thick) mcolor(navy)
msize(2) barbsize(2) || pcarrowi 15.43 904 14.79 974, lc(navy) lw(thick)
mcolor(navy) msize(2) barbsize(2) || pcarrowi 14.79 974 15.48 1009, lc(navy)
lw(thick) mcolor(navy) msize(2) barbsize(2) || pcarrowi 15.52 275 15.68 428,
lc(maroon) lw(thick) mcolor(maroon) msize(2) barbsize(2) || pcarrowi 15.51 701
15.13 729, lc(maroon) lw(thick) mcolor(maroon) msize(2) barbsize(2) || pcarrowi
15.13 729 15.56 757, lc(maroon) lw(thick) mcolor(maroon) msize(2) barbsize(2)
|| pcarrowi 16.76 275 16.93 428, yaxis(2) lc(dkgreen) lw(thick) mcolor(dkgreen)
msize(2) barbsize(2) || pcarrowi 17.03 638 16.84 701, yaxis(2) lc(dkgreen)
lw(thick) mcolor(dkgreen) msize(2) barbsize(2) || pcarrowi 16.84 701 16.42 729,
yaxis(2) lc(dkgreen) lw(thick) mcolor(dkgreen) msize(2) barbsize(2) || pcarrowi
16.648 792 16.83 820, yaxis(2) lc(dkgreen) lw(thick) mcolor(dkgreen) msize(2)
barbsize(2) || pcarrowi 16.628 974 16.97 1009, yaxis(2) lc(dkgreen) lw(thick)
mcolor(dkgreen) msize(2) barbsize(2)
```

Se obtiene la Figura 6.28.

Datos suavizados

```
. use anmusmoothpc

. scatter smfch2 num, ms(Dh) c(1) yaxis(1) yscale(axis(1) ) ytitle("Índice de
Condición ({it:IC})", axis(1)) lpattern(dash) || scatter smigsh2 num, ms(Th)
c(1) yaxis(2) yscale(axis(2) ) ytitle("Índice Gónada-Glándula ({it:IGG})",
axis(2)) xlab(1 "F" 2 "M" 3 "A" 4 "M" 5 "J" 6 "J" 7 "A" 8 "S" 9 "O" 10 "N" 11 "D"
12 "E" 13 "F" 14 "M" 15 "A" 16 "M" 17 "J" 18 "J" 19 "A" 20 "S" 21 "O" 22 "N" 23
"D" 24 "E" 25 "F" 26 "M" 27 "A" 28 "M" 29 "J" 30 "J" 31 "A" 32 "S" 33 "O" 34 "N"
35 "D" ) legend(rows(1) lab(1 "{it:IC}") lab(2 "{it:IGG}") ) xline(11.5 23.5)
text(5.1 2.5 "2012") text(5.1 13.58 "2013") text(5.1 25.6 "2014") xttitle(" ")
title("Hembras") xszie(8)
```



Se obtiene la Figura 6.29.

```
. scatter smfcm2 num, ms(Dh) c(1) yaxis(1) yscale(axis(1) ) ytitle("Índice de
Condición ({it:IC})", axis(1)) lpattern(dash) || scatter smigsm2 num, ms(Th)
c(1) yaxis(2) yscale(axis(2) ) ytitle("Índice Gónada-Glándula ({it:IGG})",
axis(2)) xlab(1 "F" 2 "M" 3 "A" 4 "M" 5 "J" 6 "J" 7 "A" 8 "S" 9 "O" 10 "N" 11 "D"
12 "E" 13 "F" 14 "M" 15 "A" 16 "M" 17 "J" 18 "J" 19 "A" 20 "S" 21 "O" 22 "N" 23
"D" 24 "E" 25 "F" 26 "M" 27 "A" 28 "M" 29 "J" 30 "J" 31 "A" 32 "S" 33 "O" 34 "N"
35 "D" ) legend(rows(1) lab(1 "{it:IC}") lab(2 "{it:IGG}") ) xline(11.5 23.5)
text(5.1 2.5 "2012") text(5.1 13.58 "2013") text(5.1 25.6 "2014") xtitle(" ")
title("Machos") xsize(8)
```

Se obtiene la Figura 6.30.

```
. scatter smlnpch2 num, ms(Dh) c(1) yaxis(1) yscale(axis(1) ) ytitle("Ln Peso
Callo ({it:LnPC})", axis(1)) lpattern(dash) || scatter smlnpgh2 num, ms(Th)
c(1) yaxis(2) yscale(axis(2)) ytitle("Ln Peso Gónada-Glándula ({it:LnPGG})",
axis(2)) xlab(2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12
"D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23
"N" 24 "D" 25 "E" 26 "F" 27 "M" 28 "A" 29 "M" 30 "J" 31 "J" 32 "A" 33 "S" 34
"O" 35 "N" 36 "D" ) legend(rows(1) lab(1 "{it:LnPC}") lab(2 "{it:LnPG}") lab(3
"{it:LnGD}")) xline(12.5 24.5) text(15.6 2.5 "2012") text(15.6 13.58 "2013")
text(15.6 25.6 "2014") xtitle(" ") title("Hembras") xsize(8)
```

Se obtiene la Figura 6.31.

```
. scatter smlnpcm2 num, ms(Dh) c(1) yaxis(1) yscale(axis(1) ) ytitle("Ln Peso
Callo ({it:LnPC})", axis(1)) lpattern(dash) || scatter smlnpgm2 num, ms(Th)
c(1) yaxis(2) yscale(axis(2)) ytitle("Ln Peso Gónada-Glándula ({it:LnPGG})",
axis(2)) xlab(2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12
"D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23
"N" 24 "D" 25 "E" 26 "F" 27 "M" 28 "A" 29 "M" 30 "J" 31 "J" 32 "A" 33 "S" 34
"O" 35 "N" 36 "D" ) legend(rows(1) lab(1 "{it:LnPC}") lab(2 "{it:LnPG}") lab(3
"{it:LnGD}")) xline(12.5 24.5) text(15.7 2.5 "2012") text(15.7 13.58 "2013")
text(15.7 25.6 "2014") xtitle(" ") title("Machos") xsize(8)
```

Se obtiene la Figura 6.32.

```
. scatter smsst num, ms(Dh) c(1) yaxis(1) yscale(axis(1) alt) ytitle("Temperatura
superficial ({&degree}C)", axis(1)) lpattern(dash) || scatter smclor num, ms(Th)
c(1) yaxis(2) yscale(axis(2) alt) ytitle("Clorofila a", axis(2)) || scatter
smfch2 num, ms(Sh) c(1) yaxis(3) ytitle("Índice de condición (hembras)",
axis(3)) xlab(2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12
"D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23
"N" 24 "D" 25 "E" 26 "F" 27 "M" 28 "A" 29 "M" 30 "J" 31 "J" 32 "A" 33 "S" 34
"O" 35 "N" 36 "D" ) legend(rows(1) lab(1 "Temp. sup. {&degree}C") lab(2 "Clor.
```



```
a") lab(3 "{it:IC}")) xline(12.5 24.5) text(30.9 2.5 "2012") text(30.9 13.58  
"2013") text(30.9 25.6 "2014") xtitle(" ") xsize(8)
```

Se obtiene la Figura 6.33.

```
. scatter smsst num, ms(Dh) c(1) yaxis(1) yscale(axis(1) alt) ytitle("Temperatura  
superficial ({&degree}C)", axis(1)) lpattern(dash) || scatter smclor num,  
ms(Th) c(1) yaxis(2) yscale(axis(2) alt) ytitle("Clorofila a", axis(2)) ||  
scatter smfcm2 num, ms(Sh) c(1) yaxis(3) ytitle("Índice de condición (machos)",  
axis(3)) xlab(2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12  
"D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23  
"N" 24 "D" 25 "E" 26 "F" 27 "M" 28 "A" 29 "M" 30 "J" 31 "J" 32 "A" 33 "S" 34  
"O" 35 "N" 36 "D" ) legend(rows(1) lab(1 "Temp. sup. {&degree}C") lab(2 "Clor.  
a") lab(3 "{it:IC}")) xline(12.5 24.5) text(30.9 2.5 "2012") text(30.9 13.58  
"2013") text(30.9 25.6 "2014") xtitle(" ") xsize(8)
```

Se obtiene la Figura 6.34.

```
. scatter smsst num, ms(Dh) c(1) yaxis(1) yscale(axis(1) alt) ytitle("Temperatura  
superficial ({&degree}C)", axis(1)) lpattern(dash) || scatter smclor num, ms(Th)  
c(1) yaxis(2) yscale(axis(2) alt) ytitle("Clorofila a", axis(2)) || scatter  
smigsh2 num, ms(Sh) c(1) yaxis(3) ytitle("Índice Gónada-Glándula (hembras)",  
axis(3)) xlab(2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12  
"D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23  
"N" 24 "D" 25 "E" 26 "F" 27 "M" 28 "A" 29 "M" 30 "J" 31 "J" 32 "A" 33 "S" 34  
"O" 35 "N" 36 "D" ) legend(rows(1) lab(1 "Temp. sup. {&degree}C") lab(2 "Clor.  
a") lab(3 "{it:IGG}")) xline(12.5 24.5) text(30.9 2.5 "2012") text(3 0.9 13.58  
"2013") text(30.9 25.6 "2014") xtitle(" ") xsize(8)
```

Se obtiene la Figura 6.35.

```
. scatter smsst num, ms(Dh) c(1) yaxis(1) yscale(axis(1) alt) ytitle("Temperatura  
superficial ({&degree}C)", axis(1)) lpattern(dash) || scatter smclor num, ms(Th)  
c(1) yaxis(2) yscale(axis(2) alt) ytitle("Clorofila a", axis(2)) || scatter  
smigsm2 num, ms(Sh) c(1) yaxis(3) ytitle("Índice Gónada-Glándula (machos)",  
axis(3)) xlab(2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12  
"D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23  
"N" 24 "D" 25 "E" 26 "F" 27 "M" 28 "A" 29 "M" 30 "J" 31 "J" 32 "A" 33 "S" 34  
"O" 35 "N" 36 "D" ) legend(rows(1) lab(1 "Temp. sup. {&degree}C") lab(2 "Clor.  
a") lab(3 "{it:IGG}")) xline(12.5 24.5) text(30.9 2.5 "2012") text(30.9 13.58  
"2013") text(30.9 25.6 "2014") xtitle(" ") xsize(8)
```

Se obtiene la Figura 6.36.

Capítulo 7

Estadios de Madurez Gonádica



7.1. Introducción

El estudio de la reproducción es un aspecto fundamental en las ciencias pesqueras. Se han establecido varios métodos para llevar a cabo este objetivo que van desde procedimientos histológicos detallados hasta índices derivados de medidas corporales.

Los estudios y análisis de la madurez de los peces están dentro de los tópicos más importantes para la evaluación y manejo de los recursos pesqueros. La explotación irracional de las poblaciones naturales produce un decremento en el número relativo de especímenes adultos (en su momento sexualmente maduros) y cambios en la fisiología de los organismos (un decremento en el tamaño y edad a la que alcanzan la madurez sexual, inversión sexual, adaptaciones compensativas). Por lo tanto, el conocimiento del tamaño al que se alcanza la madurez, los periodos de reproducción y la proporción sexual (total y por clase de talla) es indispensable para conseguir un manejo sostenible de las pesquerías, en especial con referencia a la regulación de la selectividad del arte de pesca y a los periodos de veda (Ungaro, 2008).

7.2 Escalas macroscópicas de madurez

Los estadios de madurez gonadal macroscópicos correspondientes a una escala de madurez están basados en el aspecto general, forma, dimensiones relativas, posición y color de las gónadas.

Una escala de madurez generalizada tiene el riesgo de no incluir categorías de madurez de algunas especies. Es un hecho que el desarrollo de las gónadas varía de especie a especie. Por otra parte, una escala de madurez monoespecífica es útil para la especie estudiada pero no para otras. No se puede usar en proyectos para la evaluación de recursos pesqueros multiespecies.



Por lo general, las escalas de madurez tienen de tres a ocho categorías por sexo, pero parece que con cuatro a seis estadios es suficiente. A este respecto, si se definen pocas categorías se corre el riesgo de perder información. Las escalas con numerosas categorías permiten una mejor identificación del verdadero desarrollo gonadal pero su uso requiere de más tiempo y atención para la correcta especificación de las fases de madurez.

Es por esto, que los científicos pesqueros han hecho esfuerzos por definir escalas de madurez fáciles de usar pero tan detalladas como sea posible (Ungaro, 2008).

Cuadro 7.1. Escala de madurez de Nikolsky (1963) modificada.

Estadio I	Vírgenes	Individuos jóvenes que no se han reproducido
Estadio II	Inactivo J	Los gametos todavía no empiezan a desarrollarse; gónadas de tamaño muy pequeño
Estadio III	En maduración	Ovocitos visibles a simple vista; gónadas con incremento en peso rápido; testículos cambian de transparentes a rosa pálido
Estadio IV	Madurez	Gametos maduros; Las gónadas han alcanzado su peso máximo, pero los gametos no salen con presión ligera
Estadio V	Reproducción	Los gametos salen expulsados a la presión más ligera al cuerpo; el peso gonádico disminuye rápidamente por la expulsión de gametos
Estadio VI	Vaciado	Gónadas con aspecto de saco vacío por la expulsión de los gametos; cavidad celómica distendida; pocos óvulos o esperma residuales
*Estadio VII	Inactivo A	Los gametos ya se han descargado; la gónada se extiende completamente por la cavidad corporal; los óvulos no se aprecian a simple vista.

Una de las escalas de uso generalizado es la propuesta originalmente por Nikolsky (1963). Esta escala presenta seis categorías. En el Cuadro 7.1 se ha incluido una versión modificada para distinguir a los individuos no reproductivos como “inactivos juveniles” (Inactivo J) de los “inactivos adultos” (Inactivos A), los cuales fueron agrupados juntos en la propuesta original (Salgado-Ugarte, 1985).



7.3 Relación madurez talla

El análisis de la relación entre los estadios de madurez gonádico y el tamaño de los organismos proporciona información importante para definir el tamaño del organismo al alcanzar la madurez sexual así como la temporada y los lugares (si se dispone de información espacial) de la reproducción.

La representación gráfica de esta información puede hacerse de varias maneras. La más simple es el utilizar “diagramas univariados de dispersión” en combinación con la información de los componentes poblacionales (sexos) o de las condiciones ambientales (época y lugar). Ejemplos de estos esquemas pueden encontrarse en Lara-Domínguez *et al.* (1981) y Salgado-Ugarte (1985). En esta obra presentamos variaciones sobre estos gráficos y algunas alternativas.

Como ejemplo se hará uso del archivo “bagtotb.dta” que contiene datos biológicos del bagre *Cathorops melanopus* recopilados por Salgado-Ugarte (1985) en el sistema lagunar estuarino de Tampamachoco, Ver. Este archivo tiene registros de 2,439 individuos, con información sobre el estado de madurez gonádica.

Cuadro 7.2. Número de individuos por sexo y estadio de madurez gonádica (Nikolsky modificada). *Cathorops melanopus* (Salgado-Ugarte, 1985).

Estadio de madurez	Machos	Hembras	Indeterminados	Total
I	116	89	1,170	1,375
II	376	200	0	576
III	85	179	0	264
IV	4	32	0	36
V	1	12	0	13
VI	31	32	0	63
VII	100	12	0	112
Total	713	556	1,170	2,439



Para construir las gráficas de las siguientes figuras, es necesario generar una variable de agrupamiento que contenga los grupos (etapas de madurez) ordenados de menor a mayor. Ya que se usaron códigos numéricos para identificar al sexo (1, machos; 2, hembras; 3, indeterminados) y a la madurez (1, "I"; 2 "II"; 3 "III"; 4 "IV"; 5 "V"; 6 "VI", 7 "VII"), se requiere una variable que agrupe los siete estadios de hembras, uno de indeterminados y los siete de los machos. En Stata se hace:

```
.use bagtotb
. gen matuga = 15 if matugon==7 & sex==1
(2,339 missing values generated)
. replace matuga = 14 if matugon==6 & sex==1
(31 real changes made)
. replace matuga = 13 if matugon==5 & sex==1
(1 real change made)
. replace matuga = 12 if matugon==4 & sex==1
(4 real changes made)
. replace matuga = 11 if matugon==3 & sex==1
(84 real changes made)
. replace matuga = 10 if matugon==2 & sex==1
(376 real changes made)
. replace matuga = 9 if matugon==1 & sex==1
(117 real changes made)
. replace matuga = 8 if matugon==1 & sex==3
(1,170 real changes made)
. replace matuga = 7 if matugon==1 & sex==2
(89 real changes made)
. replace matuga = 6 if matugon==2 & sex==2
(200 real changes made)
. replace matuga = 5 if matugon==3 & sex==2
(179 real changes made)
. replace matuga = 4 if matugon==4 & sex==2
(32 real changes made)
. replace matuga = 3 if matugon==5 & sex==2
(12 real changes made)
. replace matuga = 2 if matugon==6 & sex==2
(32 real changes made)
. replace matuga = 1 if matugon==7 & sex==2
(12 real changes made)
```

Y una vez recodificados, se usa esta variable ("matuga") de "madurez ajustada" para hacer los gráficos. En primer lugar, la Figura 7.1 presenta una versión del "diagrama de puntos" (dotplot con la opción center) de Stata para estos datos. Se han cambiado a números romanos los valores numéricos de la madurez y se han copiado y pegado del MSWord el símbolo de Venus (hembras) y Marte (Machos). Las líneas rojas separan a los organismos de sexo indeterminado al centro.



```
. dotplot bodlen, over(matuga) center xline(7.5) xline(8.5) xline(1(1)7 9(1)15,
lcolor(blue) lpattern(shortdash)) ytitle("Longitud patrón (mm)") xtitle("Madurez
Gonádica (Nikolsky modificada)") xlab(1 "VII" 2 "VI" 3 "V" 4 "IV" 5 "III" 6 "II"
7 "I" 8 "I" 9 "I" 10 "II" 11 "III" 12 "IV" 13 "V" 14 "VI" 15 "VII") text(70 1.5
"♀", size(*2)) text(70 14.5 "♂", size(*2)) ms(o) text(265 8 "ind.")
```

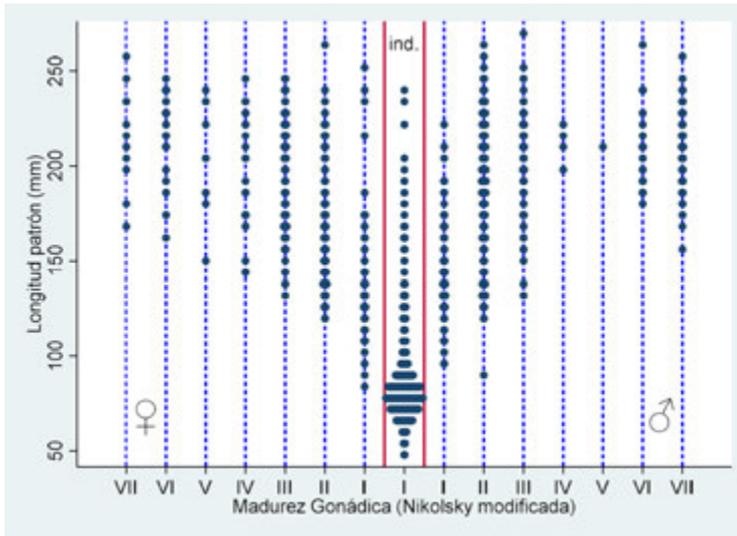


Figura 7.1. Diagrama de puntos (dotplot) para Longitud patrón por estadio de madurez gonádica, hembras, machos y organismos de sexo indeterminado de *Cathorops melanopus*.

Una versión alternativa es usar el comando “onewayplot” propuesto por Cox (2004), el cual se puede usar para proporcionar un gráfico muy semejante al anterior, con la diferencia de que este comando no lleva a cabo ningún agrupamiento sino que coloca los valores tal cual en la escala de medida (el eje vertical con la longitud patrón).

```
. onewayplot bodlen, by(matuga) xlab(1 "VII" 2 "VI" 3 "V" 4 "IV" 5 "III" 6 "II"
7 "I" 8 "I" 9 "I" 10 "II" 11 "III" 12 "IV" 13 "V" 14 "VI" 15 "VII") ms(o) stack
ytitle("Longitud estándar (mm)") text(60 14.5 "♂", size(*2)) text(60 1.5 "♀",
size(*2)) text(8.5 260 "ind.") xtitle("Madurez Gonádica (Nikolsky modificada)")
vertical center xline(7.5) xline(8.5) xline(1(1)7 9(1)15, lcolor(blue)
lpattern(shortdash)) text(265 8 "ind.")
```

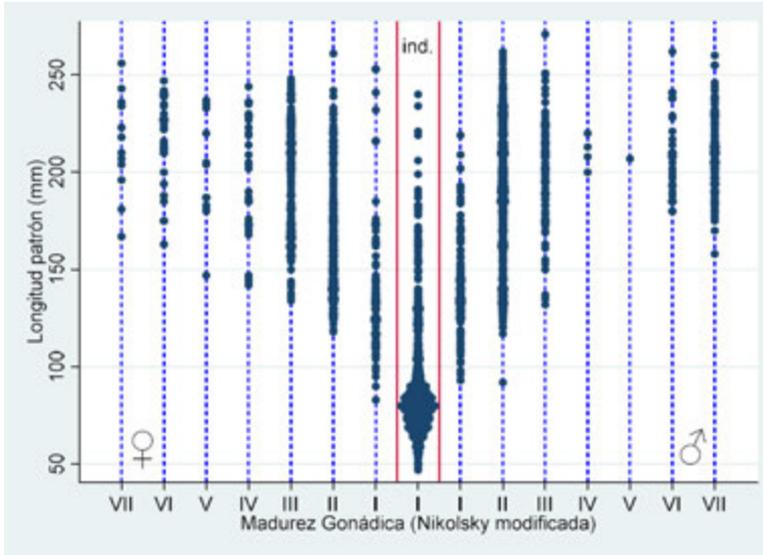


Figura 7.2. Gráfico univariado de dispersión (Oneway plot) para Longitud patrón por estadio de madurez gonádica, hembras, machos y organismos de sexo indeterminado de *Cathorops melanopus*.

Otra opción es utilizar diagramas de caja y bigotes, sólo que no será posible poner líneas verticales que separen a los sexos ni que conecten a las cajas-bigotes con el eje horizontal (Figura 7.3).

```
. graph box bodlen, over(matuga) ytitle("Longitud estándar (mm)") text(60 90 "♀", size(*2)) text(60 10 "♂", size(*2)) text(265 50 "ind.") subtitle("Madurez Gonádica (Nikolsky modificada)")
```

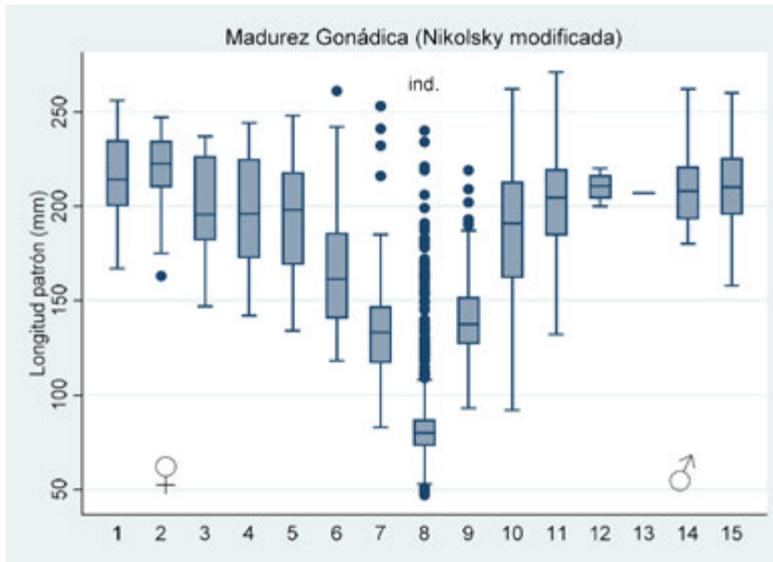


Figura 7.3. Diagrama de caja y bigotes (Stata) para Longitud patrón por estadio de madurez gonádica, hembras, machos y organismos de sexo indeterminado de *Cathorops melanopus*.

Con R, tanto líneas verticales que separen a los indeterminados, como líneas que conecten las cajas-bigotes con la categoría horizontal, pueden construirse. El primer paso es importar los datos en formato Stata (archivo “bagtoto11”, porque se guardó con el comando “saveold”, con la “opción Stata(11)”. Para ello se llama a la librería “foreign”:

```
setwd("c:/data")

library(foreign)

bagtot <- read.dta("bagtoto11.dta")

> madgon <- c("VII", "VI", "V", "IV", "III", "II", "I", "I", "I", "II", "III",
"IV", "V", "VI", "VII")

> boxplot(bodlen~matuga, data=bagtot, names=madgon, ylab="Longitud estándar
(mm)", xlab="Madurez Gonádica (Nikolsky modificada)")
```



```
> text(1.5,60,"\\VE",vfont=c("sans serif","bold"),cex=2)
> text(14.5,60,"\\MA",vfont=c("sans serif","bold"),cex=2)
> text(8,265,"ind.")
> abline(v = c(7.5,8.5), col="Red")
> abline(v = c(1,2,3,4,5,6,7,9,10,11,12,13,14,15), col="Blue", lty=3)
```

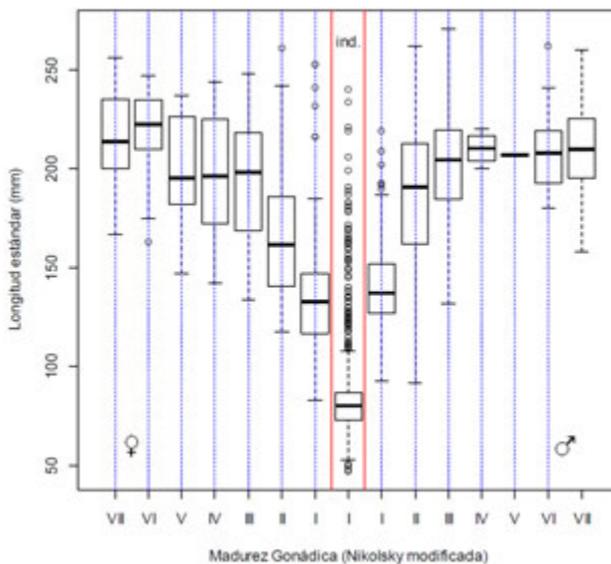


Figura 7.4. Diagrama de caja y bigotes (R) para Longitud patrón por estadio de madurez gonádica, hembras, machos y organismos de sexo indeterminado de *Cathorops melanopus*

Si se tienen numerosos datos y se quiere ver la distribución de los tamaños a cada estadio de madurez, un tipo de gráfico adicional lo constituyen los “Diagramas de violín”, los cuales son una combinación de los estimadores de densidad por kernel y los diagramas de caja y bigotes:

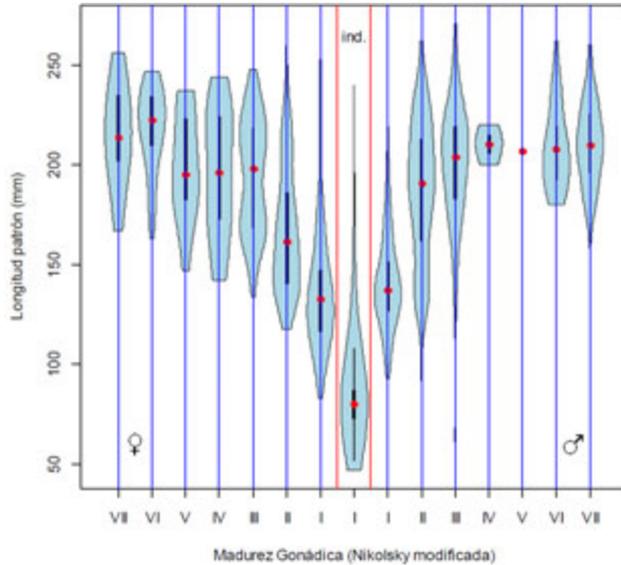


Figura 7.5. Diagramas de violín para Longitud patrón por estadio de madurez gonádica, hembras, machos y organismos de sexo indeterminado de *Cathorops melanopus*.

Si bien es posible hacerlos con Stata, el programa que los calcula da como resultado un gráfico con los grupos amontonados y sobrepuestos, lo que lo hace confuso e inapropiado. La versión de la Figura 7.5 se obtuvo con R:

```
> library(vioplot)

> violplot(bodlen~matuga, data=bagtot, names=madgon, col="Lightblue",
  ylab="Longitud patrón (mm)", xlab="Madurez Gonádica (Nikolsky modificada)",
  colMed="Red")

> text(1.5,60,"\\VE",vfont=c("sans serif","bold"),cex=2)

> text(14.5,60,"\\MA",vfont=c("sans serif","bold"),cex=2)

> text(8,265,"ind.",cex=1)
```



```
> abline(v = c(7.5, 8.5), col="Red")  
  
> abline(v = c(1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), col="Blue")  
  
> save(bagtot, file="bagtot.rda")
```

7.4 Escalas histológicas de madurez

Para la mayoría de las especies marinas explotadas, la determinación del sexo y la fase reproductiva tradicionalmente se basa en la inspección visual de características macroscópicas de las gónadas completas (West 1990). La ventaja de este método es que el estado reproductivo de los individuos puede determinarse inmediatamente y proporciona de manera rápida y económica una gran cantidad de datos que es requerida por los científicos pesqueros. Sin embargo, el sólo análisis macroscópico puede conducir a una clasificación inadecuada de las fases reproductivas ya que depende de apreciaciones subjetivas del tamaño, forma y color de las gónadas, características que pueden corresponder de manera parcial al desarrollo gonádico verdadero (Ferreri *et al.*, 2009; McPherson *et al.*, 2011; Midway y Scharf, 2012; Klibansky y Scharf, 2015).

El examen histológico proporciona información detallada acerca de los cambios que ocurren en el tejido gonádico y permite a los investigadores distinguir de manera precisa entre individuos inmaduros, en maduración o completamente maduros (Murua y Saborido-Rey 2003).

Este método puede usarse exitosamente como una herramienta poderosa para validar una escala macroscópica al enlazar categorías microscópicas de desarrollo con rasgos macroscópicos (Bianchini *et al.*, 2010; Prista *et al.*, 2014; Sieiro *et al.*, 2014) y posiblemente acopladas a un índice gonadosomático (Barreto *et al.*, 2008; McPherson *et al.*, 2011; Giacalone *et al.*, 2018).

La determinación de la madurez gonádica por medio sólo del IGS no es suficiente. La observación directa de la arquitectura histológica es el método más adecuado para conocer con exactitud el estado de maduración de las gónadas (Hosseinzadeh, *et al.* 2012).



7.5. Ciclo reproductivo del gasterópodo *Hexaplex princeps*

El gasterópodo *Hexaplex princeps* (Broderip 1833), conocido localmente como “caracol chino” es capturado de forma artesanal en Puerto Ángel Oaxaca. Este soporta el 80 por ciento de la captura de los gasterópodos en la localidad y tiene importancia pesquera ya que es consumido por la población local y la visitante (turistas).

En esta sección presentamos algunos resultados obtenidos por Saito-Quezada (2018) del capítulo con el objetivo de investigar el ciclo gonádico de *Hexaplex princeps* considerando la proporción sexual, maduración gonadal, períodos de desove, estadios de madurez y su variación con respecto a la temperatura superficial del agua de mar y la concentración de clorofila a lo largo de dos ciclos anuales en litorales rocosos aledaños a Puerto Ángel Oaxaca, México, información también incluida en Saito-Quezada *et al.* (2017) y Saito-Quezada *et al.* (2020).

Del total de ejemplares capturados (446) por buceo libre con ayuda de dos buzos locales, se separaron de 10 a 15 individuos para extraer sus partes blandas y fijarlas con formalina 10% con agua de mar (Ortíz-Ordóñez *et al.*, 2009). Después de 48 horas se lavaron en laboratorio y se conservaron en alcohol 70%. La presencia o ausencia de pene en las partes blandas permitió el sexado de los organismos y la proporción macho:hembra cuya significancia para cada muestreo y el periodo de estudio, se estableció por la prueba de bondad de ajuste a distribución uniforme, χ^2 con corrección de Yates (Cuadro 7.3).

Para el examen histológico, la espiral visceral (fusión de gónadas y glándula digestiva) se cortó transversalmente con un microtomo, se deshidrató con alcohol y xilol, se incluyó en Paraplast y parafina, se tiñó con Hematoxilina-Eosina y se fotografió de acuerdo a las recomendaciones de Uría-Galicia y Mora-Vázquez (1996).

El examen histológico de 244 individuos considerando el grado de desarrollo así como la presencia y abundancia de gametos permitió la definición de seis estadios gonádicos para las hembras (Cuadro 7.4, Figura 7.6) y cinco para los machos (Cuadro 7.5, Figura 7.7).



Cuadro 7.3. Número de individuos por sexo y fecha de colecta. Se incluye la proporción sexual y su significancia estadística (χ^2 con corrección de Yates).

Fecha de colecta	Total	Machos	Hembras	Proporción Macho:Hembra	χ^2 (Yates)	P
24/ENE/14	93	40	53	2.75 : 1.00	1.55	0.21
21/FEB/14	02	02	0	----- : -----	0.50	0.48
21/MAR/14	26	16	10	1.00 : 1.60	0.96	0.33
25/ABR/14	60	33	27	1.00 : 1.22	0.42	0.52
23/MAY/14	53	27	26	1.00 : 1.04	0.00	1.00
15/AGO/14	42	25	17	1.00 : 1.47	1.17	0.28
24/OCT/14	13	04	09	2.25 : 1.00	1.23	0.27
28/NOV/14	57	32	25	1.00 : 1.28	0.63	0.43
21/DIC/14	12	07	05	1.00 : 1.40	0.08	0.77
30/ENE/15	08	03	05	1.67 : 1.00	0.13	0.72
27/MAR/15	07	02	05	2.50 : 1.00	0.57	0.45
30/ABR/15	08	05	03	1.00 : 1.67	0.13	0.72
15/MAY/15	08	04	04	1.00 : 1.00	0.13	0.72
12/JUN/15	11	06	05	1.00 : 1.20	0.00	1.00
15/JUL/15	09	06	03	1.00 : 2.00	0.44	0.50
21/AGO/15	09	05	04	1.00 : 1.25	0.00	1.00
25/SEP/15	10	05	05	1.00 : 1.00	0.10	0.75
16/OCT/15	08	04	04	1.00 : 1.00	0.13	0.72
27/NOV/15	10	06	04	1.00 : 1.50	0.10	0.75
Total	446	232	214	1.00 : 1.08	0.65	0.42

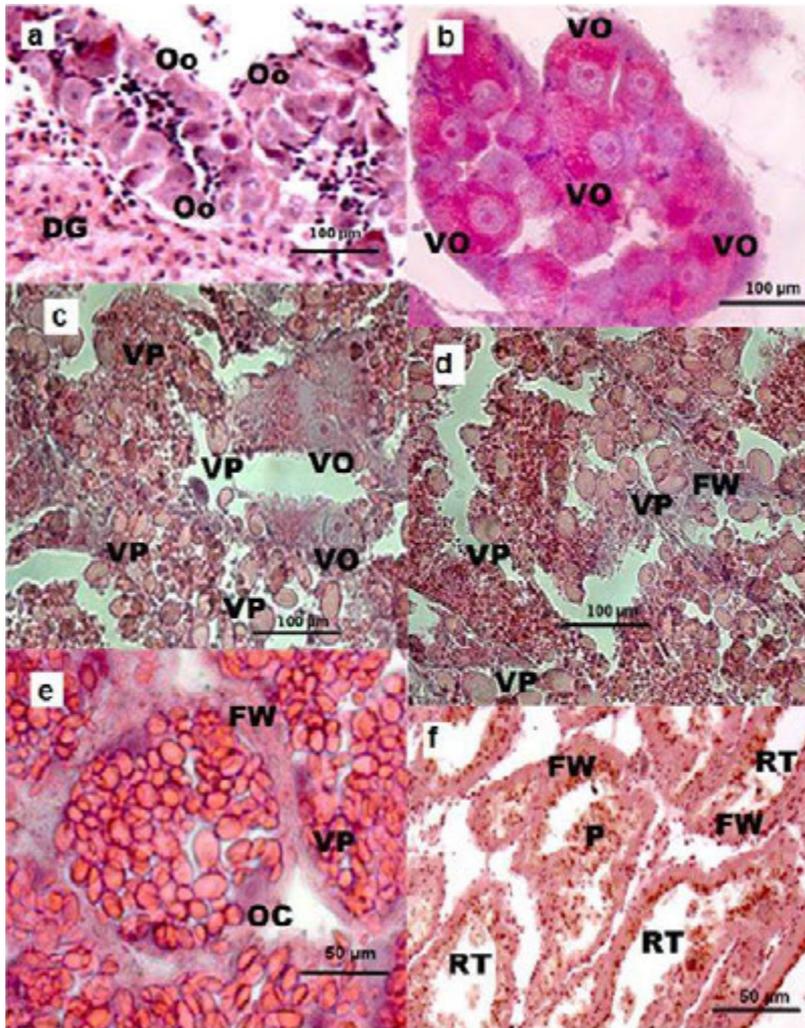


Figura 7.6. Secciones histológicas de las hembras de *Hexaplex princeps* que muestran los estadios ováricos: a: ovogénesis inicial (20X), b: madurez previtelogénica (10X), c: madurez vitelogénica (20X), d: madurez (10X), e: desove (10X), f: reposo (10X). Oo: oogonias Oc: ovocitos, DG: glándula digestiva, CT: tejido conectivo, PV: plaquetas vitelinas, PvO: ovocitos previtelogénicos, VO: ovocitos vitelogénicos, FW: pared folicular, RO: ovocitos residuales, P: fagocitos (Tomada de Saito-Quezada, 2018).



Cuadro 7.4. Caracterización de los estadios del desarrollo gonádico para hembras de *H. princeps*.

Hembras	Descripción
Estadio 1 (Ovogénesis)	Presencia ovogonias y ovocitos en desarrollo, la pared de los folículos gruesas.
Estadio 2 (Madurez vitelogénica)	Ovocitos llenos de gránulos de vitelo en algunos ovocitos se observa el núcleo y nucléolo, los folículos están completamente maduros llenos de ovocitos, con tejido conjuntivo evidente.
Estadio 3 (Madurez postvitelogénica)	Folículos con paredes delgadas y se observan ovocitos en desarrollo, se observan los gránulos de vitelo y la aparición plaquetas vitelinas.
Estadio 4 (Fin de la Madurez)	Los folículos están llenos de gránulos de vitelo y plaquetas de vitelo, las paredes de los folículos se observa delgadas.
Estadio 5 (Desove)	Se observa la luz en los folículos y parcialmente vacíos, las paredes de los folículos están delgadas, existen algunos ovocitos remanentes.
Estadio 6 (Reposo)	Se encuentran algunos folículos en reposo, así como también células o fagocitos y la pared folicular gruesas.

Cuadro 7.5. Caracterización de los estadios del desarrollo gonádico para machos de *H. princeps*.

Machos	Descripción
Estadio 1 (Espermatogénesis)	Folículos activamente en desarrollados con células inmaduras, Inicia con folículos pequeños alejados, numerosos espermatogonias y espermatocitos, las paredes de los folículos gruesas.
Estadio 2 (Madurez)	Los folículos están completamente llenos y se observa en ellos mayor cantidad de espermatozoides, Espermatogonias, espermatocitos, espermátidas.
Estadio 3 (Expulsión)	Espermatozoides maduros en expulsión, epitelio cilíndrico ciliado con pliegues.
Estadio 4 (Fin de expulsión, inicio de reposo)	Existen algunos folículos en expulsión, se observan folículos vacíos y folículos en reposo.
Estadio 5 (Reposo)	El lumen de los folículos se encuentra vacíos folículos y también se observan que los folículos están en reposo, debido a la expulsión de espermatozoides.

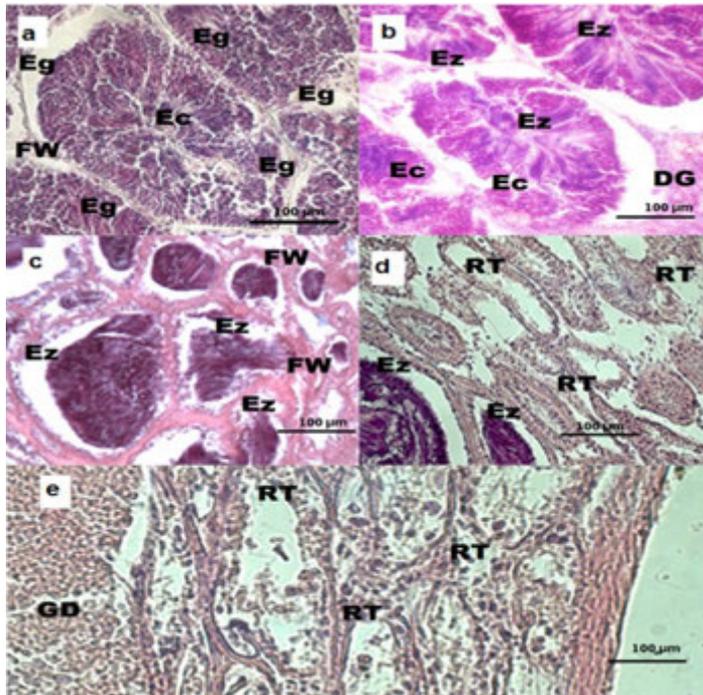


Figura 7.7. Secciones histológicas de los machos de *Hexaplex princeps* que muestran los estadios de los testículos. a: espermatogénesis inicial (50 µm), b: madurez (100 µm), c: expulsión (10x), d: fin de expulsión, inicio del reposo (100 µm), e: reposo (50 µm). Eg: espermatogonias, Ec: espermatocitos, Ez: espermatozoides, RS: espermatozoides residuales, DG: glándula digestiva, CT: tejido conectivo, P: fagocitos.

7.5.1. Gráficos de Área

La representación gráfica de esta información suele ser un “gráfico de área” (area plot). Para construirlo, es necesario seguir una serie de pasos. Dado que no es trivial, se incluyen las instrucciones para su construcción. En primer lugar se lleva la cuenta de cada individuo por fecha (año, clave numérica de mes) y estadio como en el Cuadro 7.6.



Cuadro 7.6. Estadio de desarrollo gonádico por año y por mes para hembras de *Hexaplex princeps*.

Num	Año	sexo	mes	estadio
1	2014	1	1	5
2	2014	1	1	5
3	2014	1	1	5
4	2014	1	1	4
(Se omiten filas)				
51	2014	1	12	2
52	2014	1	12	3
1	2015	1	1	4
2	2015	1	1	4
3	2015	1	1	5
(se omiten filas)				
39	2015	1	11	5
40	2015	1	11	5
41	2015	1	11	4
42	2015	1	11	5

Estos datos se introducen en Stata (por el “Editor de Datos”, “Data Editor”) y se utiliza el comando “tabulate” (que puede abreviarse “tab”) en combinación con “bysort:” como se muestra a continuación (notar que se substituyó la “ñ” por la combinación “nio”):

```
. bysort anio: tab mes estadio if sexo==1, row
```

```
-----
-> anio = 2014
```

```
+-----+
| Key      |
|-----|
|  frequency |
| row percentage |
+-----+
```



mes	estadio						Total
	1	2	3	4	5	6	
1	0	0	0	2	3	0	5
	0.00	0.00	0.00	40.00	60.00	0.00	100.00
3	0	0	0	2	4	0	6
	0.00	0.00	0.00	33.33	66.67	0.00	100.00
4	0	0	0	0	3	3	6
	0.00	0.00	0.00	0.00	50.00	50.00	100.00

(Se omiten filas)

. save hphistorepro

Cuadro 7.7. Porcentaje de individuos en cada estadio reproductivo; hembras de *Hexaplex princeps*.

Mes	f1	f2	f3	f4	f5	f6
1	0	0	0	40	60	0
2						
3	0	0	0	33.33	66.67	0
4	0	0	0	0	50	50
5	0	0	0	33.33	0	66.67
6	0	16.67	0	16.67	0	66.67
7	0	0	0	0	0	100
8	50	0	0	0	0	50
9						
10	0	20	20	20	40	0
11	0	0	16.67	16.67	66.67	0
12	0	16.67	33.33	50	0	0
13	0	0	0	40	60	0
14						
15	0	0	0	0	80	20
16	0	0	0	0	33.33	66.67



Cuadro 7.7. Porcentaje de individuos en cada estadio reproductivo; hembras de *Hexaplex princeps* (continuación).

Mes	f1	f2	f3	f4	f5	f6
17	0	25	25	0	0	50
18	0	0	0	20	20	60
19	0	0	0	0	33.33	66.67
20	0	0	0	25	25	50
21	40	40	0	0	0	20
22	0	0	50	25	25	0
23	0	0	0	25	75	0
24						

Los porcentajes de cada fila del desplegado anterior se copian y se organizan en una hoja de Excel (Cuadro 7.7) para calcular una suma acumulada para cada fila. Debido a que en algunos meses no fue posible la obtención de organismos es que aparecen filas vacías (2 = febrero 2014, 9 = septiembre 2014, 14 = febrero 2015 y 24 = diciembre 2015). Sin embargo es necesario indicarlas para mantener la continuidad de la serie de datos en el tiempo de estudio.

Cuadro 7.8. Porcentaje acumulado de individuos en cada estadio reproductivo; hembras de *Hexaplex princeps* y valores de temperatura superficial del mar (TSM, °C) y concentración de Clorofila a (CL, mg/m³) (GES DISC – NASA, 2016).

mes	f1	f2	f3	f4	f5	f6	TSM	CL
1	0	0	0	40	100	100	27.49	3.01
2							27.74	4.66
3	0	0	0	33.33	100	100	28.22	2.05
4	0	0	0	0	50	100	28.21	0.93
5	0	0	0	33.33	33.33	100	30.70	0.16
6	0	16.67	16.67	33.34	33.34	100	30.81	0.11



Cuadro 7.8. Porcentaje acumulado de individuos en cada estadio reproductivo; hembras de *Hexaplex princeps* y valores de temperatura superficial del mar (TSM, °C) y concentración de Clorofila a (CL, mg/m³) (GES DISC – NASA, 2016) (continuación).

mes	f1	f2	f3	f4	f5	f6	TSM	CL
7	0	0	0	0	0	100	31.15	0.17
8	50	50	50	50	50	100	30.83	0.44
9							29.88	0.23
10	0	20	40	60	100	100	30.44	0.38
11	0	0	16.67	33.33	100	100	29.73	0.59
12	0	16.67	50	100	100	100	27.97	2.27
13	0	0	0	40	100	100	28.05	1.62
14							27.44	1.23
15	0	0	0	0	80	100	28.18	1.09
16	0	0	0	0	33.33	100	28.12	11.89
17	0	25	50	50	50	100	30.33	0.00
18	0	0	0	20	40	100	30.43	0.20
19	0	0	0	0	33.33	100	30.59	0.19
20	0	0	0	25	50	100	31.19	0.21
21	40	80	80	80	80	100	31.25	0.35
22	0	0	50	75	100	100	30.39	0.19
23	0	0	0	25	100	100	30.58	0.19
24							29.51	1.14

La información del Cuadro 7.8 se transfiere a Stata y con el comando para gráficos de área (“tway area”) y una serie larga de opciones nos llevan a la Figura 7.8.

Para los machos se lleva a cabo la misma serie de pasos para construir la Figura 7.9.

En estos gráficos se han incluido los valores de temperatura superficial del mar (TSM) en grados Celsius y la concentración de Clorofila a (CL) en miligramos por metro cúbico.

. use hph1415b



```
. twoway area f6 f5 f4 f3 f2 f1 mes, xlab(1 "J" 2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J"
8 "A" 9 "S" 10 "O" 11 "N" 12 "D" 13 "J" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J"
20 "A" 21 "S" 22 "O" 23 "N" 24 "D") legend(order(6 5 4 3 2 1 7 8) rows(1) lab(8
"SST") lab(7 "Chlor")) col(gs0 gs3 gs6 gs9 gs12 gs15) yaxis(2) ylab(, axis(2))
ytitle("Reproductive Stage (%)" " ", axis(2)) || scatter cloro mes, c(1) yaxis(3)
lw(thick) ylab(, axis(3)) lcolor(black) mcolor(black) ms(Oh) xsize(9) ytitle("
" "Chlorophyll (mg/m{sup:3})" " ", axis(3)) xttitle(Mes) || scatter temp mes, c(1)
yaxis(1) lw(thick) ylab(27(1)31.5, axis(1)) lcolor(black) mcolor(black) ms(Sh)
lpattern(dash) ytitle(" " "Surface Sea Temperature ({&degree}C)" " ", axis(1))
xttitle(" ") yscale(axis(1) alt) yscale(axis(3) alt)
```

```
. use hpmal415
```

```
. twoway area f5 f4 f3 f2 f1 mes, xlab(1 "J" 2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7
"J" 8 "A" 9 "S" 10 "O" 11 "N" 12 "D" 13 "J" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J"
19 "J" 20 "A" 21 "S" 22 "O" 23 "N" 24 "D") legend(order(5 4 3 2 1 6 7) rows(1)
lab(7 "SST") lab(6 "Chlor")) col(gs0 gs3 gs6 gs9 gs12) yaxis(2) ylab(, axis(2))
ytitle("Reproductive Stage (%)" " ", axis(2)) || scatter clor mes, c(1) yaxis(3)
lw(thick) ylab(, axis(3)) lcolor(black) mcolor(black) ms(Oh) xsize(9) ytitle("
" "Chlorophyll (mg/m{sup:3})" " ", axis(3)) xttitle(" ") || scatter temp mes, c(1)
yaxis(1) lw(thick) ylab(27(1)31.5, axis(1)) lcolor(black) mcolor(black) ms(Sh)
lpattern(dash) ytitle(" " "Surface Sea Temperature ({&degree}C)" " ", axis(1))
xttitle(" ") yscale(axis(1) alt) yscale(axis(3) alt)
```

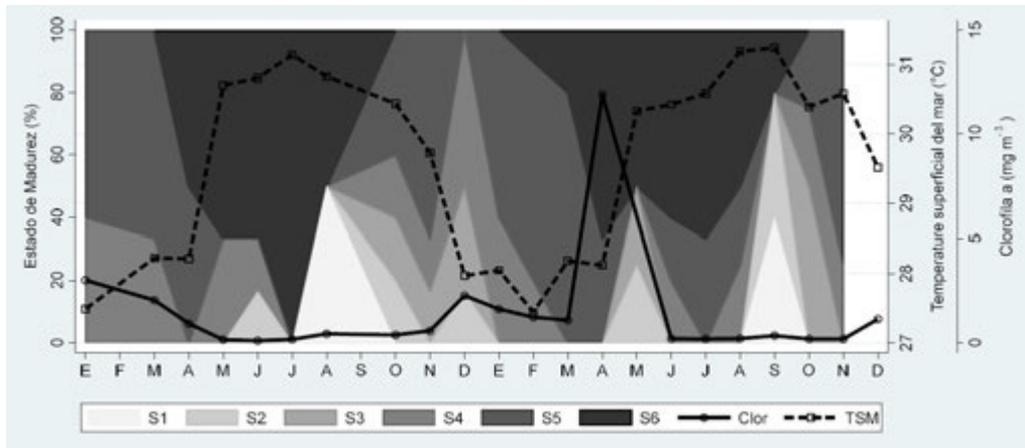


Figura 7.8. Frecuencia de estadios gonádicos 2014-2015 para las hembras de *H. princeps* por fecha de muestreo. Se muestran los valores de concentración de Clorofila a (mg m⁻³, CL) y Temperatura Superficial del Mar (°C).

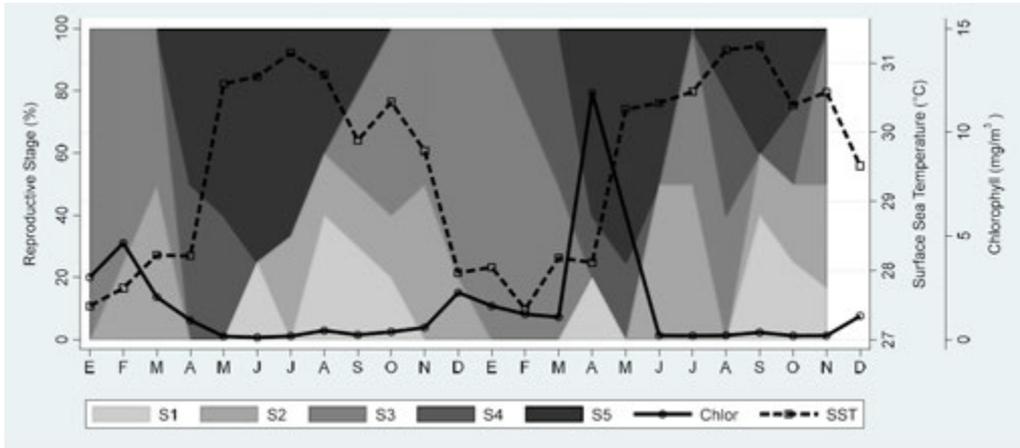


Figura 7.9. Frecuencia de estadios gonádicos 2014-2015 para los machos de *H. princeps* por fecha de muestreo. Se muestran los valores de concentración de Clorofila a (mg m^{-3} , CL) y Temperatura Superficial del Mar ($^{\circ}\text{C}$).

7.5.2. Suavización no lineal resistente

En el estudio de Saito-Quezada (2018), para evaluar la significancia estadística de las relaciones entre los estados de madurez y los valores medios de temperatura y clorofila se emplearon dos procedimientos: en primer lugar, para aclarar los patrones mostrados por los porcentajes de fases gonádicas a lo largo del tiempo, se aplicó un suavizador no lineal resistente de secuencias. El suavizador elegido fue el “4253eh, doble” que combina los resultados suaves de suavizadores de medianas corredizas de amplitud par (4,2) con la resistencia de las medianas corredizas de amplitud impar (5,3), el ajuste de puntos terminales (e), el “Hanning”, suavizador de medias móviles ponderadas (h) y el “re-roughing” (doble), que consiste en la aplicación de los suavizadores considerados (4253eh) a los residuos de los valores suavizados con los originales y su adición a los datos suavizados (Velleman, 1980; Velleman y Hoaglin, 1981; Goodall, 1990; Salgado-Ugarte, 1992; 2017; Salgado-Ugarte y Curts-García, 1992; 1993).

En segundo lugar, para comparar los datos de series de tiempo resultantes se utilizó el análisis de correlación cruzada (Davis, 2002; Beckett, 2013), entre los porcentajes de los estados de madurez contra los valores de temperatura y clorofila. También se aplicó la correlación-cruzada entre los estadios de madurez.



Los valores suavizados para los porcentajes acumulados de cada estadio de desarrollo gonadal se presentan en el cuadro 7.9, junto con los valores de clorofila también suavizados. Esto último debido a la ocurrencia en los datos recuperados de GES DISC – NASA, 2016 se presentó un caso extraordinario y con la suavización se recupera la tendencia general la cual no se ve afectada por este valor. Los valores de temperatura se dejaron sin modificación así como los porcentajes de la fase 1 de madurez ya que el suavizador los tomó por casos extraordinarios en una secuencia de ceros.

Con estos valores se construyen las gráficas para comparar la ocurrencia de los diferentes estadios de madurez a lo largo del ciclo de estudio y además para contrastarlos con la tendencia de temperatura y concentración de clorofila.

Se sustituyen los valores faltantes por ceros. Por la resistencia del método de suavización, el usar este valor no afecta a la tendencia general de los datos pero es necesario para que el método funcione.

En Stata los comandos son:

```
. replace f4 = 0 in 2  
. replace f4 = 0 in 9  
. replace f4 = 0 in 14  
. replace f4 = 0 in 24  
  
. smooth 4253eh,twice f4, gen(shf4)
```

Se repiten estos comandos para cada uno de los estadios de madurez y se generan los valores suavizados (shf2, shf3, shf4, shf5 y shf6) los cuales se listan en el Cuadro 7.9 en el cual se han agregado los valores suavizados de la concentración de Clorofila a. La fase de madurez 1 se copió del original ya que al ser sólo dos valores en la secuencia, el suavizador los elimina por considerarlos casos extraordinarios.



Cuadro 7.9. Valores suavizados de porcentajes acumulados de individuos en cada estadio reproductivo, hembras de *Hexaplex princeps*, y concentración de Clorofila a (CL, mg/m³) (GES DISC – NASA, 2016).

Mes	hf1	shf2	shf3	shf4	shf5	shf6	sCL
1	0	0	0	31.506	60	0	3.010
2	0	0	0	24.581	54.414	3.320	2.755
3	0	0	0	20.208	45.195	18.295	1.972
4	0	0	0	18.386	31.094	44.338	1.005
5	0	0	0	15.886	15.391	64.781	0.376
6	0	0	0	9.962	2.734	70.836	0.162
7	0	0	0	4.285	0	64.911	0.139
8	50*	0	1.120	2.371	1.914	44.662	0.166
9	0	1.758	6.485	4.585	13.086	18.398	0.242
10	0	5.274	13.856	14.038	26.094	2.799	0.428
11	0	7.033	16.981	26.069	35.625	0	0.823
12	0	5.274	13.856	30.938	40.527	0	1.242
13	0	1.758	6.485	26.602	39.967	0	1.407
14	0	0	1.120	14.375	35.630	6.562	1.391
15	0	0	0	2.930	31.164	22.344	1.231
16	0	0	0	0	27.648	42.370	0.862
17	0	0	0	1.289	25.135	56.329	0.454
18	0	0	0	6.562	24.088	60.418	0.226
19	0	0	0	12.051	23.782	57.657	0.177
20	0	0	0	15.176	23.796	45.378	0.182
21	40*	0	0	15.918	23.906	24.727	0.194
22	0	0	0	14.609	22.461	7.969	0.271
23	0	0	0	12.090	16.680	1.094	0.514
24	0	0	0	8.672	6.562	0	0.923



7.5.3. Correlación cruzada

Con esta información es posible comparar las secuencias. En la Figura 7.10 se comparan las hembras maduras (E4), en desove (E5) y en reposo (E6). Los correlogramas de las Figuras 7.11 y 7.12 muestran la sincronía de la maduración y el desove y el desfase de 4 a 5 meses entre el desove y el reposo.

```
. use hphm1415b

. scatter shf4 shf5 mes, ms(Oh Dh) mlw(medthick medthick) c(1 1) yscale(range(0
80)) ylab(0(20)80) legend(label(1 "Madurez (E4)") label(2 "Desove (E5)") label(3
"Reposo (E6)") rows(1)) ytitle("Estadios Reproductivos (%)") xlab(1 "E" 2 "F"
3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12 "D" 13 "J" 14 "F" 15
"M" 16 "A" 17 "M" 18 "E" 19 "J" 20 "A" 21 "S" 22 "O" 23 "N" 24 "D") xtitle("
") xline(12.5) text(78 10.8 "2014") text(78 14 "2015") text(78 20 "Hembras")
|| scatter shf6 mes, ms(Sh) c(1) mlw(medthick) yaxi(2) ytitle("Estadios
Reproductivos (%)", axis(2)) xsize(7)
```

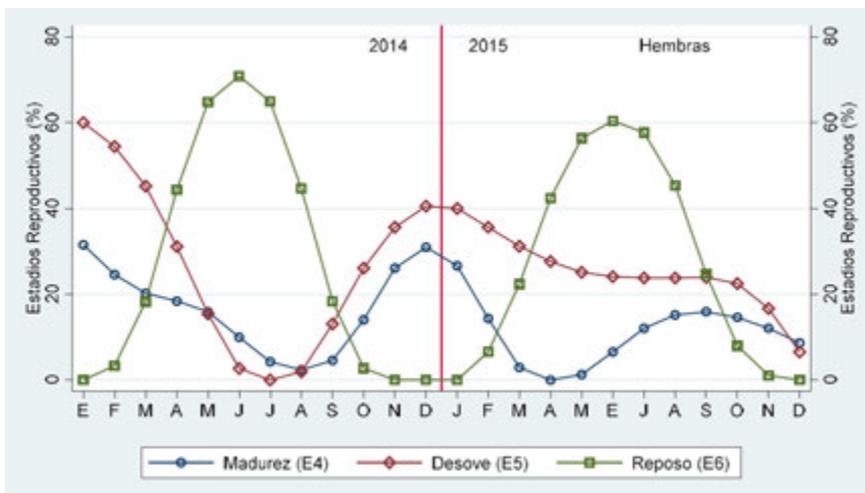


Figura 7.10. Frecuencia suavizada de hembras maduras (E4), en desove (E5) y en reposo (E6) durante el periodo de estudio.

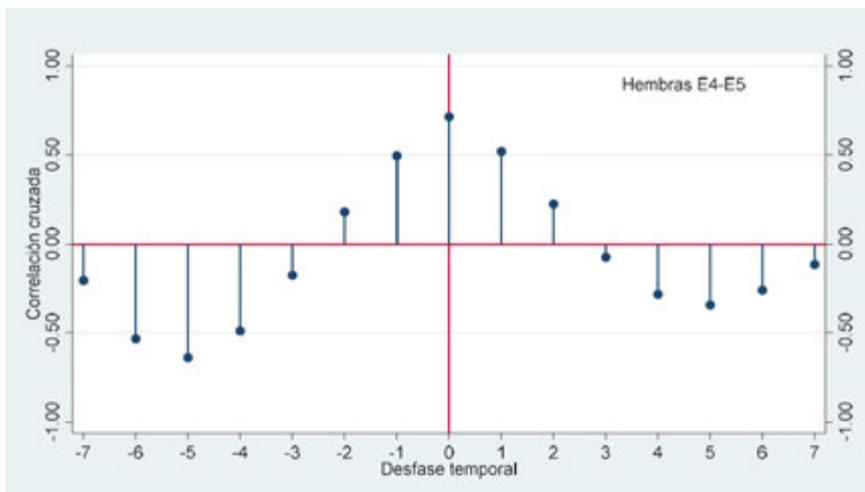


Figura 7.11. Correlograma de la secuencia de porcentajes suavizados de hembras maduras (E4) y en desove (E5).

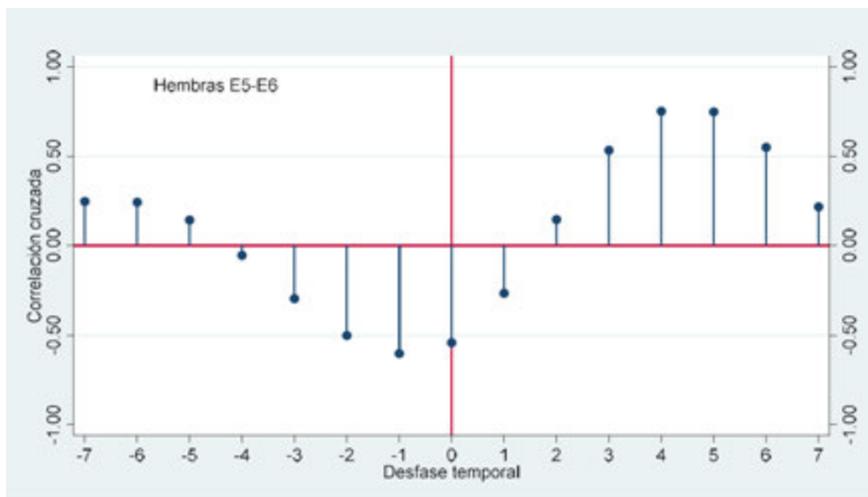


Figura 7.12. Correlograma de la secuencia de porcentajes suavizados de hembras en desove (E5) y en reposo (E6).



```
. xcorr shf5 shf6, lags(7) xlab(-7(1)7) xline(0) yline(0) text(.9 -4.5 "Hembras E5-E6") title(" ") ytitle("Correlación cruzada") xtitle("Desfase temporal") xsize(7)
```

```
. xcorr shf4 shf5, lags(7) xlab(-7(1)7) xline(0) yline(0) text(.9 4.5 "Hembras E4-E5") title(" ") ytitle("Correlación cruzada") xtitle("Desfase temporal") xsize(7)
```

Asimismo, se pueden correlacionar los porcentajes suavizados de estadios gonádicos con los valores de temperatura y concentración de clorofila. En la Figura 7.13 se comparan porcentajes suavizados de hembras en desove (E5) con los promedios de temperatura y concentración de clorofila. El comando para la gráfica, se incluye:

```
. scatter shf5 mes, c(1) ytitle("Estadio 5: Desove (%)" " ", axis(1)) ms(Sh) lw(.5) mlw(medthick) mlcolor(black) || scatter sst mes, ms(Dh) mlw(medthick) yaxi(2) c(1) mlcolor(black) xlab(1 "E" 2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12 "D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23 "N" 24 "D") legend(label(1 "% of Hembras en Desove") label(2 "Temperatura ({&degree}C)" " ", axis(2)) lw(.5) lpattern(dash) xtitle(" ") xsize(9) xline(12.5) text(58 10.8 "2014") text(58 14 "2015") || scatter sclor mes, ms(Th) mlw(medthick) c(1) yaxi(3) mlcolor(black) ytitle("Clorofila a (mg/m{superscript:3})", axis(3)) legend(label(1 "E5 Hembras") label(3 "Clorofila a") label(2 "TSM") rows(1) keygap(.1) symxsize(9)) lw(.5) yscale(axis(2) alt) yscale(axis(3) alt)
```



Figura 7.13. Frecuencia suavizada de hembras en desove (E5), temperatura superficial del mar (TSM) y concentración de clorofila a suavizada (sCL) durante el periodo de estudio.

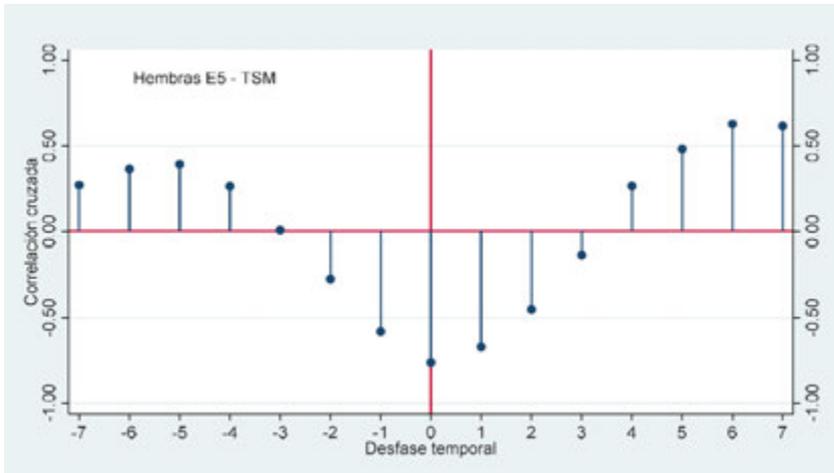


Figura 7.14. Correlograma de la secuencia de porcentajes suavizados de hembras en desove (E5) con la temperatura superficial del mar (TSM).

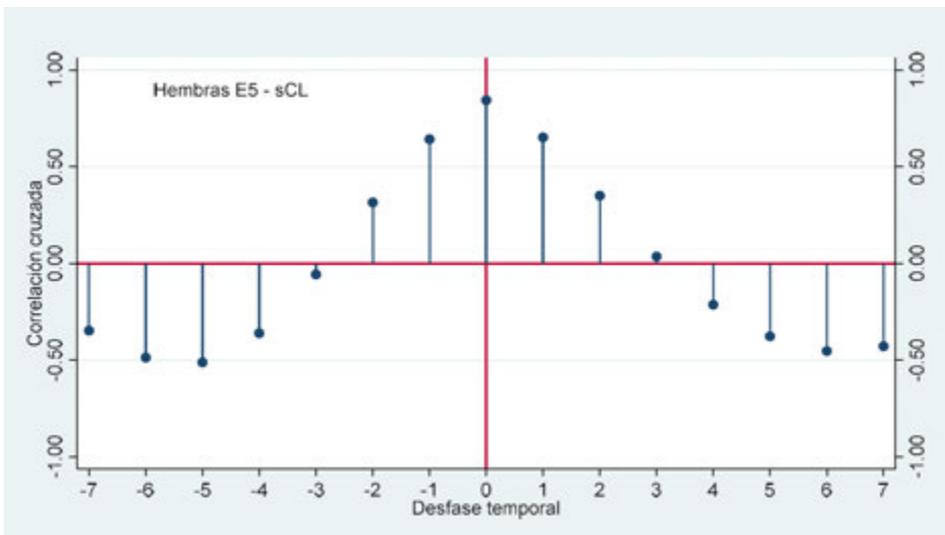


Figura 7.15. Correlograma de la secuencia de porcentajes suavizados de hembras en desove (E5) con la concentración de clorofila a suavizada (sCL).



Los comandos de Stata para las Figuras 7.14 y 7.15 se muestran a continuación:

```
. xcorr shf5 sst, lags(7) xlab(-7(1)7) xline(0) yline(0) text(.9 -4.5 "Hembras E5 - TSM") title(" ") ytitle("Correlación cruzada") xtitle("Desfase temporal") xsize(7)
```

```
. xcorr shf5 sclor, lags(7) xlab(-7(1)7) xline(0) yline(0) text(.9 -4.5 "Hembras E5 - sCL") title(" ") ytitle("Correlación cruzada") xtitle("Desfase temporal") xsize(7)
```

Es claro el desfase de seis meses entre el desove y la máxima temperatura así como la sincronía entre el desove y la máxima concentración de clorofila.

Para las hembras en reposo reproductivo se escriben los siguientes comandos:

```
. scatter shf6 mes, c(1) ytitle("Estadio 6: Reposo (%) " " ", axis(1)) ms(Sh) lw(.5) mlw(medthick) mlcolor(black) || scatter sst mes, ms(Dh) mlw(medthick) yaxis(2) c(1) mlcolor(black) xlab(1 "E" 2 "F" 3 "M" 4 "A" 5 "M" 6 "J" 7 "J" 8 "A" 9 "S" 10 "O" 11 "N" 12 "D" 13 "E" 14 "F" 15 "M" 16 "A" 17 "M" 18 "J" 19 "J" 20 "A" 21 "S" 22 "O" 23 "N" 24 "D") legend(label(1 "% of Hembras en Reposo") label(2 "Temperatura ({&degree}C)") ytitle(" " "Temperatura superficial del mar ({&degree}C)" " ", axis(2)) lw(.5) lpattern(dash) xtitle(" ") xsize(9) xline(12.5) text(68 10.8 "2014") text(68 14 "2015") || scatter sclor mes, ms(Th) mlw(medthick) c(1) yaxis(3) mlcolor(black) ytitle("Clorofila a (mg/m3)", axis(3)) legend(label(1 "E6 Hembras") label(3 "sCL") label(2 "TSM") rows(1) keygap(.1) symxsize(9)) lw(.5) yscale(axis(2) alt) yscale(axis(3) alt)
```

```
. xcorr shf6 sst, lags(7) xlab(-7(1)7) xline(0) yline(0) text(.9 -4.5 "Hembras E6 - TSM") title(" ") ytitle("Correlación cruzada") xtitle("Desfase temporal") xsize(7)
```

```
. xcorr shf6 sclor, lags(7) xlab(-7(1)7) xline(0) yline(0) text(.9 4.5 "Hembras E6 - sCL") title(" ") ytitle("Correlación cruzada") xtitle("Desfase temporal") xsize(7)
```

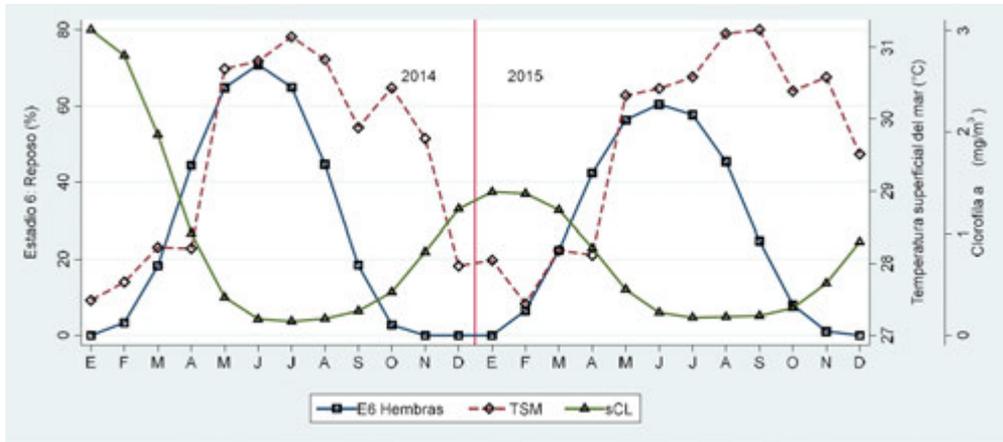


Figura 7.16. Frecuencia suavizada de hembras en reposo (E6), temperatura superficial del mar (TSM) y concentración de clorofila a suavizada (sCL) durante el periodo de estudio.

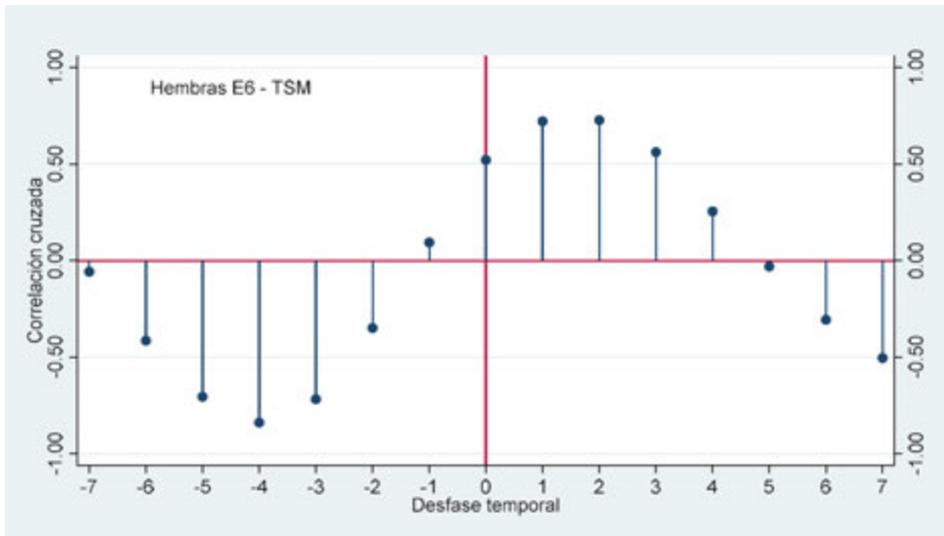


Figura 7.17. Correlograma de la secuencia de porcentajes suavizados de hembras en reposo (E6) con la temperatura superficial del mar (TSM).

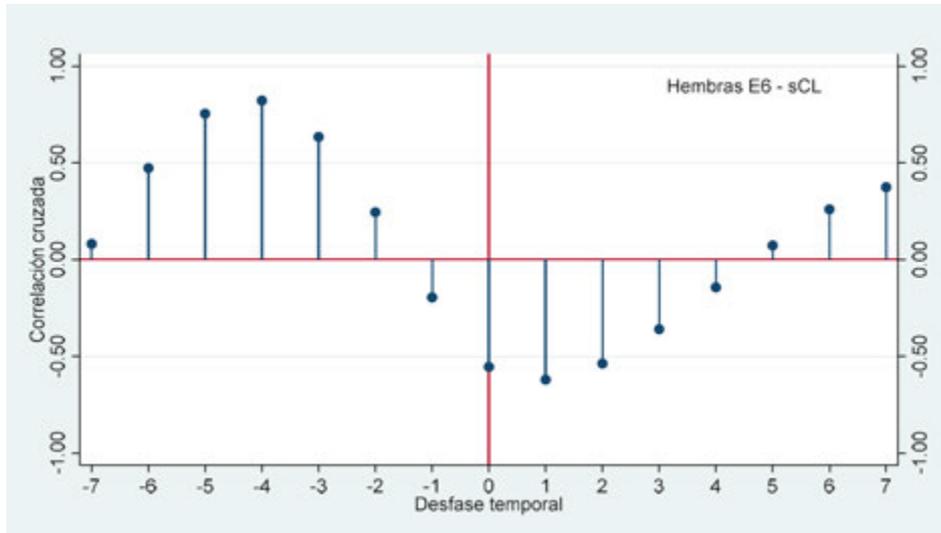


Figura 7.18. Correlograma de la secuencia de porcentajes suavizados de hembras en reposo (E6) con la concentración de clorofila a suavizada (sCL).

Es evidente, tanto de la figura con las series de datos de porcentaje suavizado de hembras en reposo con temperatura y clorofila como por los correlogramas, que las hembras se encuentran en reposo reproductivo durante los meses cálidos del año con baja concentración de clorofila. El reposo se manifiesta dos meses antes de la máxima temperatura y de cuatro a cinco meses después de las concentraciones altas de clorofila.

Gráficos y análisis equivalentes se pueden hacer para los machos. Una síntesis completa de los resultados para *Hexaplex princeps* se presentan en Saito-Quezada (2018) y Saito-Quezada *et al.* (2018).

7.6 Talla de primera madurez

Antes de poder reproducirse, los organismos deben alcanzar una cierta edad o tamaño. Esta etapa de vida depende de varios factores internos (genéticos propios de la especie) y externos (alimento, temperatura, entre otros) (Gómez-Márquez, *et al.* 2016). Para estimar el



tamaño al cual se alcanza la madurez por primera vez es usual el agrupar por tamaños a los organismos maduros sexualmente para construir una “ojiva de frecuencia acumulada”. Esta curva, suele tener una forma sigmoideal y permite la detección visual del valor de tamaño (eje horizontal) que intersecta con el porcentaje 50 del eje vertical, es decir el tamaño promedio de la mitad de los organismos maduros.

Para la determinación analítica de este valor es costumbre ajustar un modelo Logístico de tres parámetros. Existen varias formas de obtener los parámetros de esta curva, pero en la actualidad el método de preferencia es la Regresión no Lineal. El valor obtenido, al ser estimado por regresión, entonces representa un valor medio y se tiene por tanto “la talla promedio de primera madurez”.

Como ejemplo se usarán los datos del bagre estuarino *Cathorops melanopus* y el programa Stata. Primeramente se llaman los datos y se usa el comando “table” que permite obtener las frecuencias de los tamaños (longitud patrón) para los estadios considerados como maduros (III y IV). Es recomendable estimar la talla de primera madurez para cada sexo por separado, en este caso se mostrarán los resultados para los machos (código numérico de sexo = 1).

```
. use bagtoto
```

```
. table bodlen if sex==1 & (matugon ==3 | matugon==4), replace
```

El comando “table” permite reemplazar los datos originales por las frecuencias con la opción “replace”. Este método es más preciso que el derivado de primero construir un histograma (habría que elegir un origen y la amplitud de clase, problemas referenciados en el Capítulo 1 de esta obra) ya que toma los valores individuales de tamaño registrados. Al obtener las frecuencias directas (nombradas como “table1”) se hace necesario calcular sus porcentajes respecto del total (n) y su acumulado por una serie de pasos sencillos:

```
. gen sumfrec=sum(table1)
```

```
. gen porcien=table1/sumfrec[_N]*100
```

Para generar porcentajes acumulados

```
. gen pacum=sum(porcien)
```

Con estos resultados se construye la ojiva, y se ajusta un modelo logístico de tres parámetros (Ecuación 7.1):



$$Y = \frac{\beta_1}{[1 + e^{-\beta_2(X - \beta_3)}]} \tag{7.1}$$

Este modelo viene programado en Stata y se ajusta mediante el comando

```
. nl log3 pacum bodlen
(obs = 59)
```

```
Iteration 0: residual SS = 624.3003
Iteration 1: residual SS = 287.0863
Iteration 2: residual SS = 284.5307
Iteration 3: residual SS = 284.5067
Iteration 4: residual SS = 284.5064
Iteration 5: residual SS = 284.5064
```

Source	SS	df	MS	Number of obs =
Model	206860.38	3	68953.46	F(3, 56) =
Residual	284.506359	56	5.0804707	Prob > F =
Total	207144.886	59	3510.93028	R-squared =
				Adj R-squared =
				Root MSE =
				Res. dev. =

260.2546
3-parameter logistic function, pacum=b1/(1+exp(-b2*(bodlen-b3)))

pacum	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
b1	106.5702	1.390333	76.65	0.000	103.785 109.3554
b2	.0608055	.0015469	39.31	0.000	.0577067 .0639043
b3	205.1209	.5941286	345.25	0.000	203.9307 206.311

(SEs, P values, CIs, and correlations are asymptotic approximations)



Y con estos valores estimados se obtienen los calculados para la curva:

```
. predict pce
(option yhat assumed; fitted values)
```

Es posible hacer un ajuste visual dibujando diferentes líneas verticales que crucen con una línea a la altura del 50%, pero la forma analítica se obtiene con los parámetros estimados y despejando el valor de X (longitud) cuando el valor de Y (porcentaje acumulado de frecuencia) es igual a 50 (Ecuación 7.2). Con Stata, esto se logra:

$$X = -\frac{\ln\left(\frac{\beta_1}{Y} - 1\right)}{\beta_2} + \beta_3 \quad (7.2)$$

```
. di ((ln(106.5702/50 -1))/-0.0608055)+205.1209
203.0905
```

La estimación de la talla media de primera madurez para los machos de *Cathorops melanopus* es de 203.09 mm.

La Figura 7.19 presenta la ojiva de frecuencia acumulada (%) observada de longitud patrón, la curva logística ajustada y la intersección de la curva con el 50 %, punto que nos indica la talla (promedio) de la mitad de los individuos maduros, que a su vez se interpreta como de primera madurez.

```
. scatter pacum bodlen, ms(Oh) || mspline pce bodlen, bands(40) yline(50)
xline(203.09) legend(label(1 "Observados") label(2 "Calculados"))
xtitle("Longitud patrón (mm)") ytitle("Frecuencia acumulada (%)")
```

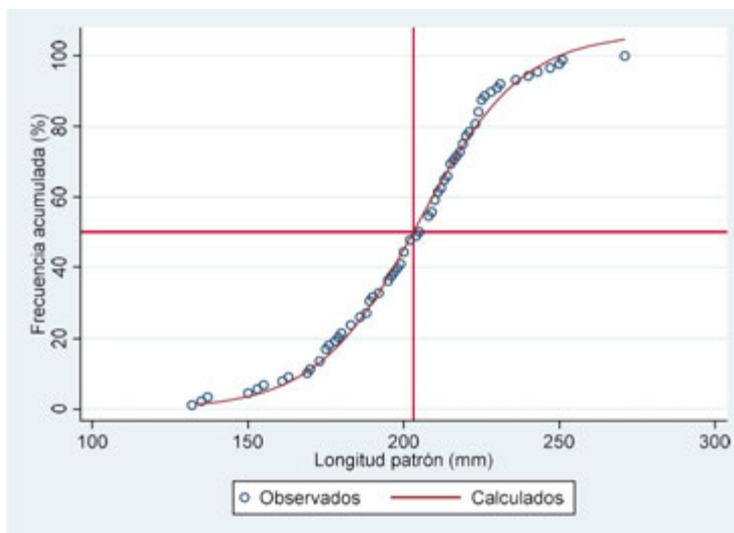


Figura 7.19. Ojiva de frecuencia acumulada en porcentaje de los tamaños, longitud patrón (mm) de individuos machos maduros (Estadios III y IV) del bagre estuarino *Cathorops melanopus*.

7.7 Comparación con curva logística de dos parámetros

El ajuste de la ojiva de frecuencia acumulada puede hacerse de varias maneras. El ajuste a una logística con tres parámetros para los datos presentados por King (2007) es:

```
. use pmking
. nl log3 pajus length
(obs = 10)
```

```
Iteration 0:   residual SS = .4570358
Iteration 1:   residual SS = .1952507
.
Iteration 10:  residual SS = .0394994
Iteration 11:  residual SS = .0394994
```



Source	SS	df	MS	Number of obs =	10
Model	4.79760051	3	1.59920017	F(3, 7) =	283.41
Residual	.039499408	7	.005642773	Prob > F =	0.0000
Total	4.83709991	10	.483709991	R-squared =	0.9918
				Adj R-squared =	0.9883
				Root MSE =	.0751184
				Res. dev. =	-26.96178

3-parameter logistic function, $pajus=b1/(1+exp(-b2*(length-b3)))$

pajus	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
b1	.8458394	.0344302	24.57	0.000	.7644249 .927254
b2	1.379613	.3422277	4.03	0.005	.5703736 2.188853
b3	14.5692	.2051721	71.01	0.000	14.08404 15.05435

(SEs, P values, CIs, and correlations are asymptotic approximations)

```
. di ((log(.8458394/.5 -1))/-1.379613)+14.5692
14.836401
```

```
. predict pae
(option yhat assumed; fitted values)
```

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	10	.	13.48089	3	-20.96178	-20.05402

Note: BIC uses N = number of observations. See [R] BIC note.

```
. scatter pajus length, ms(Oh) || mspline pae length, bands(40) yline(.5)
xline(14.83) legend(label(1 "Observados") label(2 "Calculados")) xtitle("Longitud
patrón (cm)") ytitle("Frecuencia acumulada (%)")
```

Ajuste que resulta en una talla media de primera madurez igual a 14.84 cm de longitud patrón.

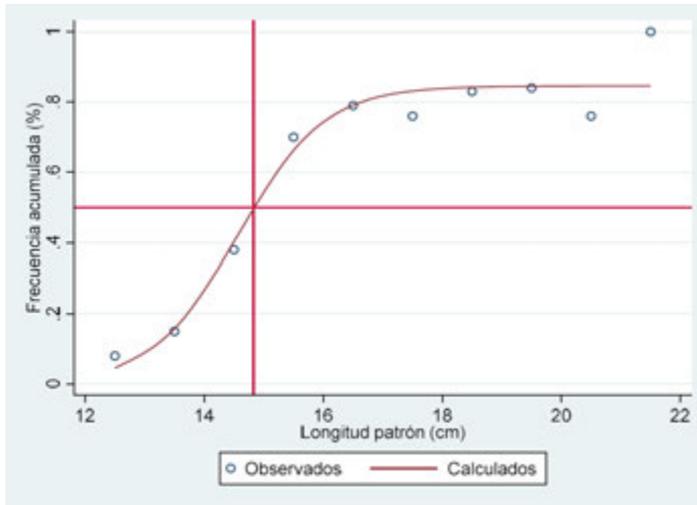


Figura 7.20. Modelo logístico de tres parámetros ajustado a los datos de madurez de hembras del camarón *Penaeus semisulcatus* (King, 2007). La talla de la mitad de los organismos maduros es de 14.84.

King (2007) propone el uso de una curva logística con dos parámetros:

$$P = \frac{1}{1 + e^{-r(L - L_m)}} \quad (7.3)$$

Donde r es el factor de curvatura y L_m es la talla media a la madurez (talla correspondiente a una proporción de 0.5). Esta ecuación no está pre-programada en Stata, pero se puede hacer el ajuste con la nueva sintaxis para la regresión no lineal que permite la especificación de la ecuación y de los valores iniciales de búsqueda de solución. Con los datos ejemplos de King (2007) se observa que valores iniciales razonables para r y L_m son 1 y 14 respectivamente. Estos valores se incluyen en la opción “initial” de “nl” y los resultados son:

```
. nl (pajus = 1/(1 + exp(-{r}*(length - {Lm}))), initial(r 1 Lm 14)
(obs = 10)
```

```
Iteration 0: residual SS = .1134799
Iteration 1: residual SS = .1083349
Iteration 2: residual SS = .1076053
```



```
Iteration 11: residual SS = .1073997
Iteration 12: residual SS = .1073997
Iteration 13: residual SS = .1073997
```

Source	SS	df	MS	
Model	4.7297002	2	2.36485011	Number of obs = 10
Residual	.10739969	8	.013424962	R-squared = 0.9778
Total	4.8370999	10	.483709991	Adj R-squared = 0.9722
				Root MSE = .1158661
				Res. dev. = -16.95906

pajus	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
/r	.6415312	.1445424	4.44	0.002	.3082158 .9748467
/Lm	15.19458	.3621655	41.95	0.000	14.35943 16.02974

```
. predict pae2
(option yhat assumed; fitted values)
```

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	10	.	8.47953	2	-12.95906	-12.35389

Note: BIC uses N = number of observations. See [R] BIC note.

La ecuación Logística con dos parámetros se automatizó en el programa “nllog2.ado” incluido.

```
. nl log2: pajus length
```

De donde la talla promedio de primera madurez con el modelo logístico de dos parámetros es de 15.19. La Figura 7.21 muestra la comparación de las curvas para ambos ajustes. Puede notarse que el ajuste a los puntos observados es mas cercano con el modelo de tres parámetros, el cual a su vez produce un valor de talla media de primera madurez mas conservador que el de dos parámetros. Como es sabido, el aumento de parámetros hace que la función se aproxime más a los datos (Verzani, 2014).



```
. scatter paju length, ms(Oh) || mspline pae2 length, bands(40) yline(.5)
xline(15.19, lpattern(dash)) lpattern(dash) legend(label(1 "Observados")
label(2 "Calculados")) xtittl
> e("Longitud patrón (cm)") ytitle("Proporción acumulada") || mspline pae
length, bands(40) xline(14.83) legend(label(1 "Observados") label(2 "FK")
label(3 "Log3") rows(1))
```

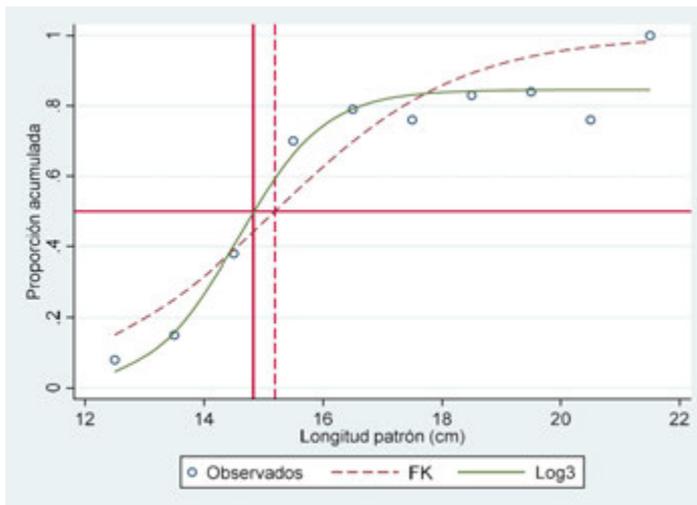


Figura 7.21. Comparación de modelos logísticos: FK modelo con dos parámetros (King, 2007); Log3 modelo con tres parámetros. Línea continua talla de primera madurez por Log3 (14.84), línea discontinua talla de primera madurez por FK (15.19).

Para elegir entre estos modelos se sugiere contrastar contra información biológica adicional (existencia de especímenes pequeños con gónadas maduras o evidencias de reproducción). Desde un punto de vista estadístico, la gráfica de las curvas ajustadas muestra que el modelo con tres parámetros sigue la tendencia de los datos de manera más cercana, lo que es de esperar al agregar un parámetro a la función logística.

Los criterios de información favorecen a la función logística con tres parámetros ya que las métricas para este modelo (AIC=-20.96; BIC= -20.05) son menores que para el Logístico con dos parámetros (AIC= -12.96; BIC=-12.35).

Bibliografía



- ABBOTT**, R.T. 1974. *American Seashells the Marine Mollusca of the Atlantic and Pacific Coasts of North America*. Second edition, Editorial Van Nostrand, NY USA, 663 p.
- ADMASU**, D. y J.M. Casselman, 2000. Otolith age determination for adult tilapia, *Oreochromis niloticus* L. from Lake Awassa (Ethiopian Rift Valley) by interpreting biannuli and differentiating biannual recruitment. *Hydrobiologia*, **418**(1): 15-24.
- AKAIKE**, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.
- AKAMINE**, T. 1985. Consideration of the BASIC programs to analyze polymodal frequency distribution into normal distributions. (in: Japanese with English abstract), *Bull. Jap. Sea Reg. Lab.* **35**: 129-160.
- AMEIJEIRAS-ALONSO**, J., R.M. Crujeiras y A. Rodríguez-Casal, 2018. *Multimode: An R Package for Mode Assessment*. arXiv preprint arXiv: 1803.00472.
- ANDREWS**, D.F. y A.M. Herzberg, 1985. *Data. A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag, Nueva York. 442 p.
- ANISLADO-TOLENTINO**, V., T. Ortiz-Pérez y G. González-Medina, 2016. *Breve Manual de Campo y Laboratorio para la Biología Pesquera de Peces*. Material Didáctico PROMEP 2010: "Dinámica pesquera de las poblaciones de peces demersales en la costa chica de Oaxaca, México". 24 p.
- AHRESTANI**, F.S. y H.T. Prins, 2011. Age and sex determination of gaur *Bos gaurus* (Bovidae). *Mammalia*, **75**: 151-155.
- ARAGÓN-NORIEGA**, E.A. 2013. Modelación del crecimiento individual del callo de hacha *Atrina maura* (Bivalvia: Pinnidae) a partir de la inferencia multi modelo. *Revista de Biología Tropical/International Journal of Tropical Biology*, **6**(3): 1167-1174.
- BAGENAL**, T.B. y F.W. Tesch, 1978. Age and growth. In: *Methods for assessment of fish production in freshwaters*. 3a. ed. Blackwell Scientific Publications Ltd. Oxford: 101-136.
- BARRETO**, A.D.V., J.R.F. Silva, M. Katsuragawa y M.C.F. Raposo, 2008. Development of the ovaries of the lobster *Panulirus echinatus* (Crustacea: Palinuridae) based on macroscopic and microscopic examination and gonadosomatic relation (GSR). *Rev Brasil Zool* **25**: 689-695
- BEAMISH**, R.J. y D.E. Chilton, 1981. Preliminary evaluation of a method to determine the age of sablefish (*Anoplopoma fimbria*). *Canadian Journal of Fisheries and Aquatic Sciences*. **39**: 277-287.
- BEAMISH**, R.J. y G.A. McFarlane, 1983. The forgotten requirement for age validation in fisheries biology. *Transactions of the American Fisheries Society*, **112**: 735-743.
-



- BECKMAN**, D. W., C.A. Wilson, y A.L. Stanley. 1988. Age and growth of red drum, *Sciaenops ocellatus* from offshore waters of the Gulf of Mexico. *Fish. Bull.* 87:17–28.
- BERNARD**, D.R. 1981. Multivariate analysis as a means of comparing growth in fish. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**: 233-236.
- BERNARD**, D.R. 1982. Reply to Misra, R.K. comment on “Multivariate Analysis as a Means of Comparing Growth in Fish”. Letters and Comments/Lettres et commentaires. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**(10): 1430-1431.
- BERTALANFFY**, L. von, 1938. A quantitative theory of organic growth. *Human Biology*, 10: 181-243.
- BEVERTON**, R.J.H. y S.J. Holt, 1957. On the dynamics of the exploited fish populations. *Fisheries Investigation of Ministry of Agriculture and Fisheries and Food Great Britain (2 Sea Fisheries)*. Vol. 19.
- BHATTACHARYA**, C. G. 1967. A simple method of resolution of a distribution into Gaussian components. *Biometrics*, **23**: 115-135.
- BIANCHINI** M.L., L. Di Stefano y S. Ragonese, 2010. Reproductive features of the deep-water rose shrimp, *Parapenaeus longirostris* (Crustacea: Penaeidae), in the Strait of Sicily. *Medit Mar Sci* 11: 81-92.
- BLACKER**, R.W. 1974. Recent advances in otolith studies. In: F.R. Harden Jones (ed.) *Sea Fisheries Research*. John Wiley, Nueva York: 67-90.
- BLACKWELL**, B.G., M.L. Brown y D.W. Willis, 2000. Relative weight (W_r) status and current use in fisheries assessment and management. *Reviews in Fisheries Science*, 8:1-44.
- BOLEN**, K.A. y R.W. Jackman, 1990. Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. En: Fox J. y J.S. Long (Eds.) *Modern methods of data analysis*. 257-291. Sage, Newbury Park, EUA.
- BOLGER**, T., y P.L. Conolly, 1989. The selection of suitable indices for the measurement and analysis of fish condition. *Journal of Fish Biology*, **34**: 171-182.
- BOWMAN**, A.W. 1984. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**: 353-360.
- BROTHERS**, E.B., 1987. Methodological approaches to the examination of otoliths in ageing studies. In: R.C. Summerfelt y G.E. Hall, (eds), *Age and growth of fish*. Iowa State University Press, Ames Iowa: 319-330.
- BURNHAM**, K.P. y D.R. Anderson, 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2a. ed. Springer, Nueva York. 448 p.
- BURNHAM**, K.P. y D.R. Anderson, 2004. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, **33**: 261-304.
- CADDY**, J.F., 1986. Size frequency analysis in stock assessment -some perspectives, approaches and problems. *Proceedings of the 37th Annual gulf and Caribbean Fisheries Institute*: 212-238.
- CADE**, B.S., J.W. Terrell y B.C. Neely, 2011. Estimating geographic variation on allometric growth and body condition of Blue Suckers with quantile regression. *Transactions of the American Fisheries Society*, 140: 1657-1669.



- CAPONI**, G. 2004. Los objetivos cognitivos de la Paleontología Cuvieriana. *Principia* 8(2): 233-258.
- CASSIE**, R.M. 1954. Some uses of probability paper for the graphical analysis of polymodal frequency distributions. *Aust. J. Mar. Freshw. Res.* 5: 513-522.
- CERRATO**, R.M. 1990. Interpretable statistical tests for growth comparisons using parameters in the von Bertalanffy equation. *Canadian Journal of Fisheries and Aquatic Sciences*, 47: 1416-1426.
- CAILLIET**, M.G., M.S. Love y A.W. Ebeling, 1986. Fishes. *A Field and Laboratory Manual of their Structure, Identification, and Natural History*. Wadsworth Publishing Company: 194 p.
- CHALE-MATSAU**, J.R., A. Govender y L.E. Beckley, 2001. Age growth and retrospective stock assessment of an economically extinct sparid fish, *Polysteganus undulosus*, from South Africa. *Fisheries Research*, 51(2001): 87-92.
- CHAMBERS**, J.M., W.S. Cleveland, B. Kleiner y P.A. Tukey, 1983. *Graphical Methods for Data Analysis*, Wadsworth & Brooks/Cole Pub. Company, Pacific Grove, p. 1-46; 26-29; 129-190.
- CHAPMAN**, D.G. 1961. Statistical problems in dynamics of exploited fisheries populations. Proceedings 4th Berkeley Symposium on Mathematics Statistics and Probability. *Cont. Biol. And Prob. Med.* 4: 153-168.
- CHECA** G.A. y P.A. Jiménez-Jiménez. 1998. Constructional morphology, origin and evolution of the gastropod operculum. *Paleobiology* 24(1): 109-132.
- CHEN**, Y., D.A. Jackson & H.H. Harvey, 1992. A comparison of von Bertalanffy and polynomial functions in modelling fish growth data. *Canadian Journal of Fisheries and Aquatic Sciences*, 49: 1228-1235.
- CHEN**, M.H. y K. Soong, 2002. Estimation of age in the sex-changing, coral-inhabiting snail *Coralliophila violacea* from the growth striae on opercula and a mark-recapture experiment. *Marine Biology* (2002) 140: 337-342.
- CHILTON**, D.E. y R.J. Beamish, 1982. Age determination methods for fishes studied by the Groundfish program at the Pacific Biological Station. *Canadian Special Publication of Fisheries and Aquatic Science*, 60: 102 p.
- COMPARINI**, A. y E. Gori, 1986. Estimating modes and antimodes of multimodal densities. *Metron (Italian Statistical Review)*, 44: 307-332.
- CONE**, R.S., 1989. The need to reconsider the use of condition indices in fishery science. *Transactions of the American Fisheries Society*, 118: 510-514.
- COX**, D.R. 1966. Notes on the analysis of mixed frequency distributions. *The British Journal of Mathematical and Statistical Psychology*, 19: 39-47.
- COX**, D.R. 1966. Notes on the analysis of mixed frequency distributions. *The British Journal of Mathematical and Statistical Psychology*, 19: 39-47.
- COX**, N.J., 2004. Speaking Stata: Graphing distributions. *The Stata Journal* 4(1): 66-68
- CRUZ-MARTÍNEZ**, A. 2000. Edad y crecimiento del tiburón toro *Carcharhinus leucas* (Valenciennes, 1839) en las aguas de Veracruz y Campeche, México. Tesis de Maestría, Instituto de Ciencias del Mar y Limnología, UNAM, México: 65 p.



- CUBILLOS**, L.A., D.F. Arcos, D.A. Bucarey y M.T. Canales, 2001. Seasonal growth of small pelagic fish off Talcahuano, Chile (37°S, 73°W): a consequence of their reproductive strategy to seasonal upwelling? *Aquatic Living Resources* 14: 115-124.
- CUPUL-MAGAÑA**, F.G. y G. Torres-Moye. 1996. Age and growth of *Astraea undosa* Wood (Mollusca: Gastropoda) in Baja California, Mexico. *Bulletin of Marine Science* 59(3): 490-497.
- DELMAS**, B. 2004. Pierre-Francois Verhulst et la loi logistique de la population. *Mathématiques et sciences humaines/Mathematics and Social Sciences*, 167: 51-81.
- DEVFLAMING**, V., G. Grossman, y F. Chapman, 1982. On the use of the gonosomatic index. *Comp. Biochem. Physiol.* 73A(1): 31-39.
- DOANE**, D.P. 1976. Aesthetic frequency classifications. *The American Statistician*, 30: 181-183.
- EBERT**, T.A. 1999. *Plant and animal populations. Methods in demography*. Academic Press. San Diego, EUA: 308 p.
- EFRON**, B. 1982. *The jackknife, the bootstrap, and other resampling plans*. Society for Industrial and Applied Mathematics, Filadelfia, EUA: 99 p.
- ELHASNI**, K., P. Vasconcelos, M. Ghorbel & O. Jarboui, 2013. Reproductive cycle of *Bolinus brandaris* (Gastropoda: Muricidae) in the Gulf of Gabès (southern Tunisia). *Mediterranean Marine Science*, 14: 24-35.
- EMERSON** J.D. y D.C. Hoaglin 1983. Stem-and-leaf displays. In: *Understanding robust and exploratory data analysis*, ed. Hoaglin, D.C., F. Mosteller and J.W. Tukey, 7-30. New York, John Wiley & sons.
- EMERSON**, J.D. y M.A. Stoto, 1983. Transforming Data. In: Hoaglin, D.C., F. Mosteller y J.W. Tukey (Eds.) *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, Nueva York, p. 97-128.
- EPANECHNIKOV**, V.A. 1969. Nonparametric estimation of a multidimensional probability density. *Theor. Probab. Appl.*, 14: 153-158.
- ERICKSON**, D.L., J.E. Hightower, y G.R. Grossman, 1985. The relative gonadal index: an alternative index for quantification of reproductive condition. *Comp. Biochem. Physiol.* 81A(1): 117-120.
- ERZINI**, K. 1990. Sample size and grouping of data for length-frequency analysis. *Fisheries Research* 9: 355-366.
- ESPINO-BARR**, E., A. González-Vega, H. Santana-Hernández y H. González-Vega, 2008. *Manual de Biología Pesquera*. Instituto Nacional de la Pesca y Universidad Autónoma de Nayarit: 131 p.
- EVERHART**, W.H. y W.D. Youngs, 1981. *Principles of Fishery Science*. 2a. ed. Cornell University Press, EUA: 350 p.
- EVERITT**, B.S. y T. Hothorn, 2006. *A Handbook of Statistical Analysis Using R*. Chapman & Hall/CRC, Boca Ratón, FL, EUA: 275 p.
- FABLET**, R., L. Pecquerie, H. de Pontual, H. Hoie, R. Millner, H. Mosegaard y S.A.L.M. Kooijman, 2011. Shedding Light on Fish Otolith Biomineralization Using a Bioenergetic Approach. *PLoS ONE*, 6(11): 1-7. e27055. doi:10.1371/journal.pone.0027055



- FAJARDO-LEÓN**, M.C., J.A. Vélez. (1996). Pesquería del Pepino de Mar. En: *Estudio del Potencial Pesquero y Acuícola de Baja California Sur. Vol (1)*
- FERRERI** R., G. Basilone, M. D'Elia, A. Traina, F. Saborido-Rey y S. Mazzola, 2009. Validation of macroscopic maturity stages according to microscopic histological examination for European anchovy. *Mar Ecol-Evol Persp* 30: 181-187.
- FISHER**, R.A., 1932. *Statistical methods for research workers*. Cuarta edición. Oliver and Boyd, Edinburgh.
- FISHER**, R.A., 1958. *Statistical methods for research workers*. XIII edición. Oliver and Boyd, Edinburgh.
- FORD**, E. 1933. An account of the herring investigations conducted at Plymouth during the years from 1924-1933. *Journal of Marine Biology Assessment*, 19: 305-384.
- FOURNIER**, D.A., J.R. Sibert, J. Majkowski y J. Hampton, 1990. MULTIFAN, a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for Southern bluefin tuna (*Thunnus maccoyii*). *Canadian Journal of Fisheries and Aquatic Sciences* 47: 301-317.
- FOX**, J. 1990. Describing univariate distributions. In: *Modern Methods of Data Analysis*, eds. J. Fox y J.S. Long: 58-125. Sage Publications, Newbury Park, CA.
- FRALEY**, C., A.E. Raftery y R. Wehrens, 2005. Mclust: Model-based Cluster Analysis, URL <http://www.stat.washington.edu/mclust>, R package version 2.1-11.
- FREEDMAN**, D. y P. Diaconis 1981a. On the histogram as a density estimator: L theory, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57: 453-476.
- FREEDMAN**, D. y P. Diaconis 1981b. On the maximum deviation between the histogram and the underlying density. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 58: 139-167.
- GARCÍA-BERTHOU**, E., y R. Moreno-Amich, 1993. Multivariate analysis of covariance in morphometric studies of the reproductive cycle. *Canadian Journal of Fisheries and Aquatic Science* 50: 1394-1399.
- FRIGGE**, M., D.C. Hoaglin, y B. Iglewicz, 1989. Some implementations of the boxplot. *The American Statistician*, 43: 50-54.
- FROESE**, R., A.C. Tsikliras y K.I. Stergiou, 2011. Editorial note on weight-length relations of fishes. *Acta Ichthyologica et Piscatoria*, 41(4): 261-263.
- GARCÍA-BERTHOU**, E. y R. Moreno-Amich, 1993. Multivariate analysis of covariance in morphometric studies of the reproductive cycle. *Canadian Journal of Fisheries and Aquatic Science* 50: 1394-1399.
- GARCÍA-BERTHOU**, G. Carmona-Catot, R. Merciai & D.H. Ogle, 2012. A technical note on seasonal growth models. *Rev. Fish Biol Fisheries*, marzo 2012: 6 p.
- GAYANILO**, F.C., Jr., M. Soriano, y D. Pauly. 1989. *A draft guide to the compleat ELEFAN*. ICLARM Software. International Center for Living Aquatic Resources Management, Manila Philippines, 2.



- GAYANILO**, F.C., Jr., P. Sparre, y D. Pauly, 1994. *The FAO-ICLARM Stock Assessment Tools (FiSAT) User's guide*. FAO Computerized Information Series (Fisheries), No. 7. Rome, FAO. 124 p.
- GHARSALLAH**, I.H., P. Vasconcelos, N. Zamouri-Langar y H. Missaoui, 2010. Reproductive cycle and biochemical composition of *Hexaplex trunculus* (Gastropoda: Muricidae) from Bizerte lagoon, northern Tunisia. *Aquatic Biology* 10: 155-166.
- GIACALONE**, V.M., R. Ferreri, V. Micale, C. Pipitone, G. Basilone, F. Sacco, S. Mazzola, A. Bonanno y G. D'Anna, 2018. Evaluation of macroscopic maturity analysis with histology in the digynic protandrous hermaphrodite white seabream *Diplodus sargus sargus* (Pisces, Sparidae). *Vie et Milieu – Life and Environment*, 68(2-3): 157-166.
- GLUYAS-MILLÁN**, M.G., C. Quiñonez-Velázquez, J.A. Massó-Rojas y F.N. Melo-Barrera, 1999. Diferencias en la relación talla-edad del caracol panocha *Astraea undosa* (Wood 1828) entre dos localidades de Bahía Tortugas, Baja California Sur, México. *Ciencias Marinas* 25(1): 91-106.
- GLUYAS-MILLÁN** M.G., C. Quiñonez Velázquez, & J. Talavera-Maya. 2000. Population parameters of the snail *Astraea undosa* (Wood 1828) from the West coast of the Baja California Peninsula. *Ciencias Marinas* 26(4): 643-658.
- GOEDEN**, G.B. 1978. A monograph of the coral trout, *Plectropomus leopardus* (Lacépède). *Res. Bull. Fish. Serv. Queensl.*, 1: 42 p.
- GÓMEZ-MÁRQUEZ**, J.L. 1994. *Métodos para determinar la edad en los organismos acuáticos*. F.E.S. Zaragoza U.N.A.M. México.
- GÓMEZ-MÁRQUEZ**, J.L., A. Cervantes-Sandoval, I.H. Salgado-Ugarte, B. Peña-Mendoza, Ma. Del C. Alejo-Plata, J.L. Guzmán-Santiago, E. Mendoza-Vallejo y C. Machuca-Rodríguez, 2016. *Temas Selectos de Biología Pesquera*. DGAPA (PAPIME PE205513) y FES Zaragoza, UNAM, México: 337 p.
- GOOD**, I.J. y R.A. Gaskins 1980. Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association*, 75: 42-73.
- GOODALL**, C. 1990. A survey of smoothing techniques. In: *Modern Methods of Data Analysis*, eds. J. Fox y J.S. Long, 58-125. Newbury Park, CA: Sage publications.
- GONZÁLEZ-CASTILLO**, A.G. y M.A. Rivera-Barrera, 2018. Aspectos biológicos de Equinoideos de Puerto Ángel, Oaxaca, México. Informe del Laboratorio de Biometría y Biología Pesquera, Carrera de Biólogo, FES Zaragoza, UNAM, México: 24 p.
- GOULD**, S.J. 1966. Allometry and size in ontogeny and phylogeny. *Biological Reviews*, 41: 587-640.
- GROSJEAN**, P. 2001. Growth model of the reared sea urchin *Paracentrotus lividus* (Lamarck, 1816). Tesis de doctorado, Facultad de Ciencias, Universidad Libre de Bruselas. 271 p.
- GULLAND**, J.A., 1969. *Manual of methods for fish stocks assessment. Part I. Fish population analysis*. FAO Manual Fisheries Science. Gulland, J.A., 1983. *Fish stock assessment; a manual of basic methods*. Wiley Interscience. Chichester.



- GULLAND**, J.A. y S.J. Holt, 1959. Estimation of growth parameters for data at unequal time intervals. *Journal Council CIEM*, 25(1): 47-49.
- HADDON**, M. 2001. *Modelling and Quantitative Methods in Fisheries*. Chapman & Hall/CRC, Boca Raton, 404 p.
- HALL**, P., S.J. Sheather, M.C. Jones, y J.S. Marron, 1991. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 27: 228-254.
- HAMILTON**, L.C., 1992. Quartiles, outliers, and normality: some Monte Carlo results. *Stata Technical Bulletin* 6. Reprinted in *Stata Technical Bulletin Reprints*, vol 3, pp. 92-95.
- HARDING**, J.F. 1949. The use of probability paper for the graphical analysis of polymodal frequency distributions. *J. Mar. biol. Ass. U.K.* 28: 141-153.
- HÄRDLE**, W.K., 1991. *Smoothing Techniques. With Implementations in S*. Springer-Verlag, Nueva York: 261 p.
- HÄRDLE**, W. y D.W. Scott. 1988. Smoothing in low and high dimensions by weighted averaging using rounded points. *Technical report* 88-16, Rice University.
- HÄRDLE**, W. y L. Simar, 2015. *Applied Multivariate Statistical Analysis*. 4ª ed. Springer-Verlag, Berlín. 580 p.
- HARRISON**, D.A. 2014. Stata tip 20: Generating Stata bin variables. In: Cox N.J. y H.J. Newton (eds.) 31-32. *One Hundred Nineteen Stata Tips*. 3a. ed. Stata Press, College Station, EUA: 326 p.
- HASSELBLAD**, V. 1966. Estimation of parameters for a mixture of normal distributions. *Technometrics*. 8: 431-444.
- HATANAKA**, M. y K. Sekino, 1962b. Ecological studies on the Japanese sea-bass, *Lateolabrax japonicus* - II. Growth. *Bulletin of the Japanese Society of Scientific Fisheries*, 28(9): 857-861. (en japonés con resumen en inglés).
- HAYASHI**, I., 1969. Some observations on the reproductive duct of the Japanese sea bass, *Lateolabrax japonicus*. *Japanese Journal of Ichthyology*, 16(2): 68-73. (En inglés con resumen en japonés).
- HAYASHI**, I. 1971. On the process of the testicular maturation of the Japanese sea bass, *Lateolabrax japonicus*. *Japanese Journal of Ichthyology*, 18(1): 39-50. (En japonés con resumen en inglés).
- HAYASHI**, I., 1972. On the ovarian maturation of the Japanese sea bass *Lateolabrax japonicus*. *Japanese Journal of Ichthyology*, 19(4): 243-254.
- HEADRICK**, M.R. y F.J. Margraf Jr. 1982. Comment on Multivariate Analysis as a Means of Comparing Growth in Fish. Letters and comments/Letres et commentaires. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(2): 529-530.
- HERRERO PEREZRUJ**, D. (2004). Análisis de la pesquería de pepino de mar en la Península de Baja California, México. Tesis de Doctorado. Centro Multidisciplinario de Ciencias Marinas, IPN, México: 143 p.
- HILBORN**, R. y C.J. Walters. 2001. *Quantitative Fisheries Stock Assessment: Choice, Dynamics, & Uncertainty*. Chapman & Hall, Nueva York. 570 p.



- HIRAI**, A., 1988. Fine structures of the micropyles of pelagic eggs of some marine fishes. *Japanese Journal of Ichthyology*, **35**(3): 351-357.
- HOAGLIN**, D.C. 1983. Letter values: A set of selected order statistics. In: *Understanding robust and exploratory data analysis*, ed. Hoaglin, D.C., F. Mosteller y J.W. Tukey, 33-57. Nueva York, John Wiley & sons.
- HOLDEN**, M.J. y D.F.S. Rait, 1975. Manual de Ciencia Pesquera Parte 2. Métodos para Investigar los Recursos y su Aplicación. Organización de las Naciones Unidas para la Agricultura y la Alimentación. Roma.
- HOSSEINZADEH**, M., M.R. Imanpoor & H. Nedoubin, 2012. Histology of ovarian development and maturity stages in the wild Persian sturgeon, *Acipenser persicus*. 1:483. doi:10.4172/scientificreports.483.
- HOWE**, J.C. 2002. Standard length: not quite so standard. *Fisheries Research* 56(2002): 1-7.
- IBÁÑEZ-AGUIRRE**, A.L., M. Gallardo-Cabello y X. Chiappa-Carrara, 1999. Growth analysis of striped mullet, *Mugil cephalus*, and white mullet, *M. curema* (Pisces: Mugilidae), in the Gulf of Mexico. *Fishery Bulletin* 97: 861-872.
- IGLEWICS**, B. 1983. Robust Scale Estimators and Confidence Intervals for Locations. In: Hoaglin, D.C., F. Mosteller y J.W. Tukey, (eds.) *Understanding robust and exploratory data analysis*, 404-431. Nueva York, John Wiley & Sons.
- IIZUKA**, K. y S. Katayama, 2008. Otolith morphology of teleost fishes of Japan. *Bulletin of Fisheries Research Agency*, 25: 1-222.
- IRIE**, T., 1955. The crystal texture of the otolith of a marine teleost *Pseudosciaena*. *Journal of the Faculty of Fisheries and Animal Husbandry, Hiroshima University*, **1**(1): 1-13.
- IRIE**, T., 1957. On the forming season of annual rings (opaque and translucent zones) in the otolith of several marine teleosts. *Journal of the Faculty of Fisheries and Animal Husbandry, Hiroshima University*, **1**(3): 311-319.
- IRIE**, T., 1960. The growth of the fish otolith. *Journal of the Faculty of Fisheries and Animal Husbandry, Hiroshima University*, **3**(1): 203-230.
- IZENMAN**, A.J., y C. Sommer, 1988. Philatelic mixtures and multimodal densities. *Journal of the American Statistical Association*, **83**(404): 941-953.
- JELLYMAN**, P.G., D.J. Booker, S.K. Crow y D.J. Jellyman, 2013. Does one size fit all? An evaluation of length-weight relationships for New Zealand's freshwater fish species. *New Zealand Journal of Marine and Freshwater Research*, 47:450-468.
- JONES**, M.C., 1990. Variable kernel density estimates and variable kernel density estimates. *Australian Journal of Statistics*, 32: 361-371.
- KATSANEVAKIS**, S. 2007. Growth and mortality rates of the fan mussel *Pinna nobilis* in Lake Vouliagmeni (Korinthiakos Gulf, Greece): a generalized additive modelling approach. *Marine Biology*, 152: 1319-1331.
- KEEN**, M. 1971. *Sea shells of tropical west America*. Stanford University Press, California. p. 353-355.



- KIMURA**, D.K. 1980. Likelihood methods for the von Bertalanffy growth curve. *Fishery Bulletin*, **77**(4): 765-776.
- King**, M. 1995. *Fisheries biology, assessment and management*. Fishing News Books. Blackwell Science, Ltd., 341 p.
- KING**, M., 2007. *Fisheries Biology, Assessment and Management*. Fishing News Books, Blackwell Science, Ltd., 341 p.
- KING**, M., 2007. *Fisheries Biology, Assessment and Management*. 2nd ed., Blackwell Publishing, Victoria, Australia: 172-238; 339-371.
- KLIBANSKY** N. y F.S. Scharf 2015. Success and failure assessing gonad maturity in sequentially hermaphroditic fishes: comparison between macroscopic and microscopic methods. *J Fish Biol* **87**: 930-957.
- KOSAKA**, M., 1969. Ecology of the common sea bass, *Lateolabrax japonicus* in Sendai Bay. *Collected Reprints of the Faculty of Marine Science and Technology, Tokai University*, **3**: 67-85. (En japonés con resumen en inglés).
- KUWATANI**, Y. 1962. Suzuki o taisho to suru gyosho no sogoteki kenkyu (The synthetic study on the fish bank for the Japanese sea bass, *Lateolabrax japonicus*). *Bull. Kyoto Pref. Fish. Res. Lab.*, **8**: 1-129. (En japonés)
- LARA-DOMÍNGUEZ**, A.L., A. Yáñez-Arancibia y F. Amezcua-Linares. 1981. Biología y ecología del bagre *Arius melanopus* Günther en la Laguna de Términos, sur del Golfo de México (Pisces: Ariidae). *Anales del Instituto de Ciencias del Mar y Limnología*. **8**(1): 267-304.
- LEISCH**, F. 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, **11**. URL <http://www.jstatsoft.org/v11/i08/>.
- LIU**, Q., T. Pitcher, y M. al-Hossaini, 1989. Ageing with fisheries length-frequency data, using information about growth. *Journal of Fish Biology*, **35**: 169-177.
- LÓPEZ-ROCHA**, J.A. (2012). Distribución y Abundancia del pepino de mar *Isostichopus badionotus* frente a la Costa de Sisal, Yucatán. Proceedings of the 64th Gulf and Caribbean Fisheries Institute, Octubre 31 - Noviembre 5, 2011, Puerto Morelos, Mexico:
- LOWERRE-BARBIERI**, S.K., M.E. Chittenden, Jr. y C.M. Jones, 1994. A comparison of a validated otolith method to age weakfish, *Cynoscion regalis*, with the traditional scale method. *Fishery Bulletin*, **92**: 555-568.
- LOZANO-CABO**, F. 1983. *Oceanografía, Biología Marina y Pesca. Tomo I. Tercera parte: Biometría y Bioestadística*. Paraninfo, Madrid. 445 p.
- MACDONALD** P.D.M. y J. Du, 2018. mixdist: Finite Mixture Distribution Models. R package version 0.5-5. <http://CRAN.R-project.org/package=mixdist>
- MACDONALD** P.D.M. y P.E.J. Green, 1988. *User's guide to program MIX: an interactive program for fitting mixtures of distributions*. Ichthus Data Systems, Hamilton, Ontario, Canada.
- MACDONALD** P.D.M. y T. Pitcher, 1979. Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. *Journal of the Fisheries Research Board of Canada* **36**: 987-1001.



- MALHADO**, C.H.M., A.A. Ramos, P.L.S. Carneiro, J.C. Souza, F.S. Wechsler, J.P. Eler, D.M.M.R. Azevêdo & J.R.B. Sereno, 2008. Modelos no lineales para describir el crecimiento de bufalinos de la raza Murrah. *Archivos de zootecnia*, 57(220): 497-503.
- MARRON**, J.S. 1986. Will the art of smoothing ever become a science? *Contemporary Mathematics*. 59: 169-178.
- MATSUMIYA**, Y., T. Mitani y M. Tanaka, 1982. Changes in distribution pattern and condition coefficient of the juvenile Japanese sea bass with the Chikugo River ascending. *Bulletin of the Japanese Society of Scientific Fisheries*, 48(2): 129-138.
- MATSUMIYA**, Y., M. Oka, K. Hiramatsu y K. Asano, 1991. Analysis of a mixed normal distribution by means of the first difference of the logarithmic frequency distribution. *Bulletin of the Faculty of Bioresources, Mie University*, 5: 69 - 72.
- McLACHLAN**, G.J. y K.E. Basford, 1988. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, Nueva York, EUA.
- McLACHLAN**, G. y D. Peel, 2000. *Finite Mixture Models*. John Wiley & Sons, Nueva York, EUA: 419 p.
- McMAHON**, T.A. y J.T. Bonner, 1986. *Tamaño y Vida*. Prensa Científica, Editorial Labor, Barcelona, España: 255 p.
- McPHERSON** L.R., K. Ganas y C.T. Marshall, 2011. Inaccuracies in routinely collected Atlantic herring (*Clupea harengus*) maturity data and correction using a gonadosomatic index model. *Journal of the Marine Biological Association of the United Kingdom* 91: 1477-1487.
- MELÉNDEZ-CONTRERAS**, M.E. 2015. Ciclo reproductivo de la madreperla *Pinctada mazatlanica* (Hanley, 1856) en zonas del litoral rocoso adyacentes a Puerto Ángel, Oaxaca, México. Tesis de licenciatura (Bióloga), Facultad de Estudios Superiores Zaragoza, Carrera de Biólogo, UNAM, México: 74 p.
- MIDWAY** S.R. y Scharf F.S. 2012. Histological analysis reveals larger size at maturity for southern flounder with implications for biological reference points. *Mar Coastal Fish* 4: 628-638.
- MINITAB** Inc., 2010. Minitab 17 Statistical Software [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com)
- MISRA**, R.K., 1980. Statistical comparisons of several growth curves of the von Bertalanffy type. *Canadian Journal of Fisheries and Aquatic Sciences*, 37: 920-926.
- MISRA**, R.K. 1982. Comment on "Multivariate Analysis as a Means of Comparing Growth in Fish". Letters and Comments/Letres et commentaires. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(10): 1430-1431.
- MITO**, S., 1957. On the egg development and larvae of a Japanese sea bass, *Lateolabrax japonicus* (Cuvier). *Science Bulletin of the Faculty of Agriculture, Kyushu University*, 16(1): 115-127. (En japonés con resumen en inglés)
- MOORE**, J.W. y I.A. Moore, 1974. Food and growth of arctic char *Salvelinus alpinus* in the Cumberland sound area of Baffin Island. *Journal of Fish Biology* 6: 79-92.



- MORALES-LINARES**, C.A. 2006. Estimaciones de la edad y modelos matemáticos del crecimiento de *Oreochromis niloticus* (Linnaeus, 1757) en la Presa “Emiliano Zapata”, Morelos, México. Tesis de Maestría, Posgrado en Ciencias Biológicas, FES Zaragoza, UNAM, México: 202 p.
- MORALES-NIN**, B. 1992. *Determination of growth in bony fishes from otolith microstructure*. FAO Fisheries Technical Paper No. 322, Roma: 51 p.
- MORALES-RAMÍREZ**, D.K. 2018. Edad y crecimiento en pepinos de mar de Puerto Ángel, Oaxaca, México. Reporte del Laboratorio de Investigación Formativa VIII. FES Zaragoza UNAM, México: 30 p.
- MOREAU**, J. 1987. Mathematical and biological expression of growth in fishes: Recent trends and further developments. En: R.C. Summerfelt y G. E. Hall (Eds.), *The Age and Growth of Fish*: 81-113. Iowa State University Press, Ames, EUA. 544 p.
- MORRISON**, D.F., 1978. *Multivariate Statistical Methods*. 2a. ed. McGraw-Hill International Book Co. Singapur: 128-169.
- MORRISON CASSIE**, R., 1962. Frequency distribution models in the ecology of plankton and other organism. *Journal of Animal Ecology*, **31**: 65-92.
- MORRIS**, P.A. 1969. *A field guide to Pacific coast shells. Including shells of Hawaii and the Gulf of California*. 2ª ed. Houghton Mifflin Co. Boston, EUA: 64-65; Lámina 32.
- MUGIYA**, Y. y T. Uchimura, 1989. Otolith resorption induced by anaerobic stress in the goldfish, *Carassius auratus*. *Journal of Fish Biology*, **35**: 813-818.
- MÜLLER**, D.W. y G. Sawitzki, 1991. Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, **86**(415): 738-746.
- MURRAY**, J.D. 2001. *Mathematical Biology*. 3a. ed. Springer-Verlag, Berlin. 551 p.
- MURUA** H. y F. Saborido-Rey, 2003. Female reproductive strategies of marine fish species of the North Atlantic. *J Northw Atl Fish Sci* **33**: 23-31.
- NASH**, R.D.M., A.Valencia-Hernández y A.J. Geffen. 2006. The Origin of Fulton’s Condition Factor – Setting the Record Straight. *Fisheries*, **31**(5): 236-238.
- NEUMANN**, R.M. y M.S. Allen, 2007. Size structure. En Guy, C.S. y M.L. Brown, eds. *Analysis and Interpretation of Freshwater Fisheries Data*, Capítulo 9: 375-421. American Fisheries Society, Bethesda, MD. E.U.A.
- NISPERUZA-PÉREZ**, C.A., J. Padilla-Cantero & J.A. Quirós-Rodríguez, 2018. Densidad poblacional y estructura de talla del pepino de mar *Holothuria (Halodeima) grisea* (Aspidochirotida: Holothuriidae) en aguas someras del sur del golfo de Morrosquillo, Caribe Colombiano. *Revista de Biología Tropical/International Journal of Tropical Biology*, **66**(2): 776-787.
- OGLE**, D.H., 2013. fishR – Von Bertalanffy Growth Model – Extra. <http://derekogle.com/fishR/examples/oldFishRVignettes/VonBertalanffyExtra.pdf>
- OGLE**, D.H., 2016. *Introductory Fisheries Analyses with R*. CRC Press, Boca Raton, EUA: 317 p.
- OGLE**, D.H. 2019. FSadata: Fisheries Stock Analysis, Datasets. R package version 0.3.8.



- OGLE**, D.H., P. Wheeler, & A. Dinno, 2020. *FSA: Fisheries Stock Analysis. R package, version 0.8.30*, <https://github.com/droglenc/FSA>.
- ORTÍZ ORDOÑEZ E.**, Mendoza-Santana E. L., Belmar-Pérez J., Padilla-Benavides T., 2009, Histological description of the Males and Female Gonads in *Tegula eiseni*, *T. funebraris*, *T. aureotincta*, *T. gallina* and *T. regina*, from Bahía Tortugas BCS., Mexico. *International Journal of Morphology*, 27(3):691-697.
- PANFILI J.**, H. De Pontual, H. Troadec & P.J. Wright (Eds), 2002. *Manual of Fish Sclerochronology*, IFREMER-IRD, Brest, France: 463 p.
- PATTERSON**, K.R., 1992. An improved method for studying the condition of fish, with an example using Pacific sardine *Sardinops sagax* (Jenyns). *Journal of Fish Biology*, 40: 821-831.
- PAULY**, D., 1984. *Fish Population Dynamics in Tropical Water: A Manual for Use With Programmable Calculator*. International Center for Living Aquatic Resources Management, ICLARM, Studies and Reviews 8, Manila Filipinas, 179 pags.
- PAULY**, D. y J.F. Caddy, 1985. A modification of Bhattacharya's method for the analysis of mixtures of normal distributions. *FAO Fish. Circ.* 781: 16.
- PAULY**, D. y G. Gaschütz, 1979. A simple method for fitting oscillating length growth data, with a program for pocket calculators. *International Council for the Exploration of the Sea. Council Meeting 1979, G:24. Demersal Fish Committee*: 26 p.
- PAULY**, D. y G.R. Morgan, 1987. *Length-based methods in fisheries research*. ICLARM Conference Proceedings 13, 468 p. International Center for Living Aquatic Resources Management, Manila, Philippines and Kuwait Institute for Scientific Research, Safat, Kuwait.
- PEARSON**, K. 1891. Lectures at Gresham College in London. Noviembre de 1891. Londres, R.U. (citado en Scott, 2015).
- PEARSON**, K. 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions A*, 185: 71-110.
- PETERSEN**, C.G.J., 1892. Fiskenes biologiske forhold i Holboek Fjord, 1890-91. *Beretning fra de Danske Biologiske Station for 1890-91*, 1: 121-183 (en Danés).
- PITCHER**, T.J., & P.D.M. MacDonald, 1973. Two models for seasonal growth in fishes. *Journal of Applied Ecology*, 10(2): 599-606
- PRIETO-RIOS E.**, F.A. Solís-Marín, G.H. Borrero-Pérez & P. Díaz-Jaimes, 2014. Filogeografía de *Holothuria (Halodeima) inornata* Semper, 1868 (Echinodermata: Holothuroidea). *Revista peruana de biología* 2(12): 155-162.
- PRISTAN.**, L. Gordo, J.L. Costa, M.J. Costa y C. Jones, 2014. Reproductive phase determination in male meagre (*Argyrosomus regius*, Sciaenidae): testis development and histologic corroboration of a gross anatomical scale. *Sci Mar* 78: 65-80.
- RAHMAN**, M.A. 2015. World Sea Cucumber Fisheries: Status, Culture, Application, Effectiveness of Management and Extinction Threats. *International Conference on Waste Management, Ecology and Biological Sciences (WMEBS'15) Mayo 13-14*. Kuala Lumpur, Malasia: 65-72.



- RAMÍREZ-GÓMEZ**, K.E. 2018. Determinación de edad y crecimiento de los Equinoideos *Equinometra vanbrunti* y *Toxopneustes roseus* en Puerto Ángel, Oaxaca. Reporte semestral, FES Zaragoza, UNAM, México: 27 p.
- R CORE TEAM**. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RIAÑO-RUFILANCHAS**, D. 2017. On the origin of Karl Pearson's term "histogram". *Estadística Española*, 59(192): 1-7.
- RICHARDS**, F.J. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany*, 10(29): 290-300.
- RICKER**, E.W., 1975. Computation and interpretation of biological statistics of fish populations. *Bulletin of the Fisheries Research Board of Canada*. (191): 145-157.
- ROEDER**, K., 1990. Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *Journal of the American Statistical Association*, 85(411):617-624.
- RODRÍGUEZ-GALICIA**, V.A. 2015. Determinación de edad y crecimiento del caracol *Leucozonia cerata* (Wood, 1828) del litoral rocoso cerca de Puerto Ángel Oaxaca. Reporte del Laboratorio de Investigación Formativa VIII. Carrera de Biología, FES Zaragoza, UNAM, México: 14 p.
- ROFF**, D.A. 2006. *Introduction to Computer-Intensive Methods of Data Analysis in Biology*. Cambridge University Press. Nueva York, EUA: 368 p.
- ROSENBLATT**, M. 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27: 832-837.
- ROUNSEFELL**, G.A. y W.H. Everhart, 1953. *Fishery Science. Its Methods and Applications*. John Wiley & Sons, Nueva York, EUA: 444 p.
- ROYSTON**, P. 1992. Sg1.2: Nonlinear regression command. *Stata Technical Bulletin* 7: 11-18.
- RUDEMO**, M., 1982. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9: 65-78.
- SAITO-QUEZADA**, V.M. 2004. Edad y crecimiento por lectura de otolitos de la tilapia (*Oreochromis niloticus*) en la presa Emiliano Zapata, Morelos. Tesis de licenciatura, Facultad de Estudios Superiores Zaragoza, UNAM, México: 87 p.
- SAITO-QUEZADA** V.M. 2010. Edad y Crecimiento del Caracol *Astrea undosa* (Wood, 1828), (Gastropoda: Turbinidae) de "La Bocana", Baja California Sur, México, Tesis de Maestría en Ciencias Biológicas, Facultad de Estudios Superiores Zaragoza UNAM, México: 122 p.
- SAITO-QUEZADA**, V.M. 2018. Aspectos biológico-pesqueros de los gasterópodos *Hexaplex princeps* (Broderip, 1833) y *Opeatostoma pseudodon* (Burrow, 1815) de la pesca artesanal en Puerto Ángel, Oaxaca, México. Tesis de Doctorado, Facultad de Estudios Superiores Zaragoza, UNAM, México: 230 p.
- SAITO-QUEZADA**, V.M., E. Uria-Galicia, J.L. Gómez-Márquez, A.B. Villaseñor-Martínez, Ma. De L. Jiménez-Badillo & I.H. Salgado-Ugarte, 2018. Reproductive cycle of *Hexaplex*



- princeps* (Broderip, 1833) from one artisanal fishery at the southern coast of Mexico. *Latin American Journal of Aquatic Research*, 46(1): 91-103.
- SAITO-QUEZADA**, V.M., E. Uria-Galicia, J.L. Gómez-Márquez, A.B. Villaseñor-Martínez, Ma. De L. Jiménez-Badillo & I.H. Salgado-Ugarte, 2020. Reproductive cycle of *Hexaplex princeps* (Broderip, 1833). En: Gardón-Poggy, J.C. y K. Satué-Ambrojo (Eds.): 91-103, *Reproductive Biology and Technology in Animals*. IntechOpen, Londres, R.U.
- SALGADO-UGARTE**, I.H., 1992. *El análisis exploratorio de datos biológicos. Fundamentos y aplicaciones*. Libro de texto, E.N.E.P. Zaragoza U.N.A.M. y Ediciones Marc, 250 p. (14 capítulos, 2 apéndices). I.S.B.N. 968-36-2439-1.
- SALGADO-UGARTE**, I.H., 1985. Algunos Aspectos Biológicos del Bagre *Arius melanopus* Günther (Osteichthyes: Ariidae) en el Sistema Lagunar de Tampamachoco, Veracruz. Tesis de licenciatura, E.N.E.P. Zaragoza, U.N.A.M. México, 108 p. 21 figuras, 20 tablas y un cuadro.
- SALGADO-UGARTE**, I.H., 1995. Nonparametric methods for fisheries data analysis and their application in conjunction with other statistical techniques to study biological data of the Japanese sea bass *Lateolabrax japonicus* in Tokyo Bay. Tesis de doctorado en Biociencia Acuática desarrollada en el Departamento de Pesquerías, Facultad de Agricultura de la Universidad de Tokio, Tokio, Japón, 389 p.
- SALGADO-UGARTE**, I.H., 2002. *Suavización No Paramétrica para Análisis de Datos*. F.E.S. Zaragoza U.N.A.M. 139 p.
- SALGADO-UGARTE**, I.H. 2013. *Métodos Estadísticos Exploratorios y Confirmatorios para Análisis de Datos. Un Enfoque Biométrico*. FES Zaragoza y DGAPA, UNAM, México: 307 p.
- SALGADO-UGARTE**, I.H. 2016. *Nonparametric Methods for Fisheries Data Analysis. The Case of the Japanese Sea Bass (Lateolabrax japonicus)*. Lambert Academic Publishing. Saarbrücken, Alemania: 393 p.
- SALGADO-UGARTE**, I.H., 2017. *Métodos Estadísticos Exploratorios y Confirmatorios para Análisis de Datos. Un Enfoque Biométrico*. D.G.A.P.A y F.E.S. Zaragoza U.N.A.M. [Open Access]. ISBN: 978-607-02-4231-1. Libro en formato ePub disponible en <http://www.librosoa.unam.mx/handle/123456789/296>
- SALGADO-UGARTE**, I.H. y J. Curts-García, 1992. Resistant smoothing using Stata. *Stata Technical Bulletin* 7: 8 - 11.
- SALGADO-UGARTE**, I.H. y J. Curts-García, 1993. Twice reroughing procedure for resistant nonlinear smoothing. *Stata Technical Bulletin* 11: 14 - 16.
- SALGADO-UGARTE**, I.H. y M.A. Pérez-Hernández, 2003. Exploring the use of variable bandwidth kernel density estimators. *The Stata Journal*, 3(2): 133-147.
- SALGADO-UGARTE**, I.H., J.L. Gómez-Márquez y B. Peña-Mendoza, 2005. *Métodos Actualizados Para Análisis de Datos Biológico-Pesqueros*. F.E.S. Zaragoza y D.G.A.P.A., UNAM, México: 240 p.
- SALGADO-UGARTE**, I.H., M. Shimizu, y T. Taniuchi. 1993. Exploring the shape of univariate data using kernel density estimators. *Stata Technical Bulletin* 16: 8-19.



- SALGADO-UGARTE**, I.H., M. Shimizu y T. Taniuchi. 1994. sg23: Semi-graphical determination of Gaussian components in mixed distributions. *Stata Technical Bulletin* 18: 15-27.
- SALGADO-UGARTE**, I.H., M. Shimizu, y T. Taniuchi, 1995a. ASH, WARPing, and kernel density estimation for univariate data. *Stata Technical Bulletin* 26: 2-10.
- SALGADO-UGARTE**, I.H., M. Shimizu, y T. Taniuchi, 1995b. Practical rules for bandwidth selection in univariate density estimation. *Stata Technical Bulletin*, 27: 5-19.
- SALGADO-UGARTE**, I.H., M. Shimizu y T. Taniuchi, 1997. snp13: Non parametric assessment of multimodality for univariate data. *Stata Technical Bulletin*, 38: 27-35.
- SALGADO-UGARTE**, I.H., J. Martínez-Ramírez, J.L. Gómez-Márquez y B. Peña-Mendoza, 2000. sg128: Some programs for growth estimation in Fisheries Biology. *Stata Technical Bulletin*, 53: 35 - 47.
- SAUCEDO-LASTRA** P.E. 1995. Crecimiento, Relaciones, Alométricas y Reproducción de las Ostras Perlas *Pinctada mazatlanica* y *Pteria stenia* (Bivalvia:Pteriidae) bajo condiciones de Repoblamiento en El mérito, Bahía de la Paz, Baja California Sur, México. Tesis de Maestría en Ciencias. Instituto Politécnico Nacional. Centro Interdisciplinario de Ciencias Marinas. México: 100 p.
- SCHNUTE**, J. y D. Fournier, 1980. A new approach to length-frequency analysis: growth structure. *Can. J. Fish. Aquat. Sci.* **37**: 1337-1351.
- SCHWARTZ**, G. 1987. Estimating the dimension of a model. *Annals of Statistics* 6: 461-464.
- SCOTT**, D.W. 1979. On optimal and data-based histograms. *Biometrika*, 66: 605-610.
- SCOTT**, D.W. 1985a. Frequency polygons. Theory and application. *Journal of the American Statistical Association*, **80**(390): 348-354.
- SCOTT**, D.W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, Nueva York.
- SCOTT**, D.W. 2015. *Multivariate Density Estimation: Theory, Practice, and Visualization*. 2a. ed. John Wiley & Sons, Nueva York.
- SCOTT**, D.W., A. Gebhardt y S. Kaluzny, 2015. Ash: David Scott's ASH Routines. R package version 1.0-15. <https://CRAN.R-project.org/package=ash>.
- SCRUCCA**, L., M. Fop, T.B. Murphy y A.E. Raftery, 2016 mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal* **8**(1): 205-233.
- SECOR**, D.H., J.M. Dean y E.H. Laban, 1991. *Manual for Otolith Removal and Preparation for Microstructural Examination. Technical Publication 1991-01*. Belle W. Baruch Institute for Marine Biology and Coastal Research, EUA: 85 p.
- SHAPOVALOV**, L. y H.C. Taft, 1954. The life histories of the Steelhead Rainbow Trout (*Salmo gairdneri gairdneri*) and Silver Salmon (*Oncorhynchus kisutch*) with special reference to Waddell Creek, California, and recommendations regarding their management. *Calif. Div. Fish Game, Fish Bull.*, **98**: 375 p.
- SHEATHER**, S.J., y M.C. Jones, 1991. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, B*, **53**(3): 683-690.
- SHRADER**, A.M., S.M. Ferreira, M.E. McElveen, P.C. Lee, C.J. Moss & R.J. van Aarde, 2006.



- Growth and age determination of African savanna elephants. *Journal of Zoology*, 270: 40-48.
- SIEIRO** P., J. Otero y A. Guerra, 2014. Contrasting macroscopic maturity staging with histological characteristics of the gonads in female *Octopus vulgaris*. *Hydrobiologia* 730: 113-125.
- SIEGEL**, S. y N.J. Castellan, 1995. *Estadística No Paramétrica. Aplicada a las Ciencias de la Conducta*. 4ª ed. Trillas, México: 437 p.
- SILLIMAN**, R.P., 1967. Analog computer models of fish populations. *Fishery Bulletin*, 66: 31-46.
- Silverman, B.W. 1978. Choosing the window width when estimating a density. *Biometrika*, 65: 1-11.
- SILVERMAN**, B.W. 1978. Choosing the window width when estimating a density. *Biometrika*, 65: 1-11.
- SILVERMAN**, B.W. 1981a. Density estimation for univariate and bivariate data. En: *Interpreting Multivariate Data*, ed. V. Barnett: 37-53, John Wiley and Sons, Chichester.
- SILVERMAN**, B.W. 1981b. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, B*, 43: 97-99.
- SILVERMAN**, B.W. 1983. Some properties of a test for multimodality based on kernel density estimates. In: *Probability, Statistics and Analysis*, ed. J.F.C. Kingman and G.E.H. Reuter: 248-259. Cambridge, Cambridge University Press.
- SILVERMAN**, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall. Londres: 176 p.
- SIMONOFF**, J.S. 1996. *Smoothing Methods in Statistics*. Springer-Verlag, Nueva York: 340 p.
- SOKAL**, R.R. y F.J. Rohlf, 1981. *Biometry, The Principles and Practice of Statistics in Biological Research*. 2a. ed. W.H. Freeman & Company, Nueva York: 859 p.
- SOMERS**, I.F. 1988. On Seasonally Oscillating Growth Function. *Fishbyte* 6: 8-11.
- SPARRE**, P., 1987b. *Computer programs for fish stock assessment*. FAO Fisheries Technical Paper. 101.2. Roma, FAO: 218 p.
- SPARRE**, P. y S.C. Venema, 1992. *Introduction to tropical fish stock assessment. Part 1. Manual*. FAO Fisheries Technical Paper. 306.1. Rev. 1. Roma, FAO: 376 p.
- SPARRE**, P., E. Ursin y S.C. Venema, 1989. *Introduction to tropical fish stock assessment. Part 1. Manual*. FAO Fisheries Technical Paper. 306.1. Roma, FAO: 57- 123.
- SPRUGEL**, D.G. 1983. Correcting for bias in log-transformed allometric equations. *Ecology*, 64:209-210.
- STATA CORP.** 1995. *Stata Statistical Software, Release 4.0*. College Station, TX; Stata Corporation.
- STATA CORP.** 1999. *Stata Statistical Software; Release 6.0*. College Station, TX; Stata Corporation.
- STATA CORP.** 2001. *Stata Statistical Software; Release 7.0*. College Station, TX; Stata Corporation.
- STATA CORP.** 2013. *Stata Statistical Software; Release 13.0*. College Station, TX; Stata Corporation.
- STATA CORP.** 2019. *Stata Statistical Software; Release 16.0*. College Station, TX; Stata Corporation.



- STEARNS**, S.C., 1983. A natural experiment in life-history evolution: field data on the introduction of mosquitofish (*Gambusia affinis*) to Hawaii. *Evolution* **37**(3): 601-617.
- STEVENSON**, R.D. y W.A. Woods, Jr., 2006. Condition indices for conservation: new uses for evolving tools. *Integrative and Comparative Biology*, **46**(6): 1169-1190.
- STURGES**, H.A. 1926. The choice of a class interval. *Journal of the American Statistical Association*, **21**: 65-66.
- SUKUMAR**, R., N.V. Joshi y V. Krishnamurthy, 1988. Growth in the Asian elephant. *Proceedings of the Indian Academy of Science. (Animal Science)*, **97**(6): 561-571.
- TANAKA**, S. 1962. A method of analyzing a polymodal frequency distribution and its application to the length distribution of the porgy, *Tauius tumifrons* (T. and S.). *Journal of the Fisheries Research Board of Canada*, **19**: 1143-1159.
- TARTER**, M.E. y R.A. Kronmal 1976. An introduction to the implementation and theory of nonparametric density estimation. *The American Statistician*, **30**: 105-112.
- TAYLOR**, C.C., 1989. Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, **76**: 705-712.
- TERRELL**, G.R., 1990. The maximal smoothing principle in density estimation. *Journal of the American Statistical Association*, **85**(410): 470-477.
- TERRELL**, G.R. and D.W. Scott, 1985. Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association*, **80**(389): 209-214.
- THOMPSON**, D'A.W. 1917; 1961, 2014. *On Growth and Form*. Cambridge University Press, Canto Classics edition, 5th printing (2018). Cambridge: 346 p.
- TORRES-ZEPEDA**, M.G. y M. Cruz-Pérez. 2002. Caracterización del ciclo reproductivo y épocas de desove de la Madreperla *Pinctada mazatlanica* (Hanley, 1856) en la Bahía de Acapulco (Guerrero, México). Congreso Iberoamericano Virtual de Acuicultura: 409-421.
- TUKEY**, J.W. 1977. *Exploratory Data Analysis*. Reading, MA. Addison-Wesley.
- TUCKER**, A.R., 1954. *American Seashells. The New Illustrated Naturalist*. D. Van Nostrand Company, Inc. Nueva York, EUA: 541 p.
- UNGARO**, N. 2008. Field manual on macroscopic identification of maturity stages for the Mediterranean fishery resources. GCP/RER/ITA/MSM-TD-21. MedSudMed Technical Documents No 21: 34 pp.
- URSIN**, E. 1968. A mathematical model of some aspects of fish growth, respiration and mortality. *Journal of the Fisheries Research Board of Canada*, **24**: 2355-2453.
- VALENCIA-HERNÁNDEZ**, A. 2003. *Assessment of the allometry and length-at age in the growth of fishes*. Tesis de doctorado, Universidad de Liverpool, RU.
- VASCONCELOS** P., B. Lopes, M. Castro & M.B. Gaspar. 2008. Gametogenic cycle of *Hexaplex (Trunculariopsis) trunculus* (Gastropoda: Muricidae) in the Ria Formosa lagoon (Algarve coast, southern Portugal). *J. Mar. Biol. Assoc. UK*, **82**: 321-329.
- VASCONCELOS**, P., P. Moura, C.M. Barroso & M.B. Gaspar. 2012. Reproductive cycle of *Bolinus brandaris* (Gastropoda: Muricidae) in the Ria Formosa lagoon (southern Portugal). *Aquatic Biology*, **16**: 69-83. doi: 10.3354/ab00434.



- VELLEMAN**, P.F., 1980. Definition and comparison of robust nonlinear data smoothing algorithms. *Journal of the American Statistical Association*, 75: 609-615.
- VELLEMAN** P.F. y D.C. Hoaglin, 1981. *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press, p. 41-63.
- VERHULST**, P.F. 1838. Notice sur la loi que suit la population dans son accroissement. *Correspondance Mathématique et Physique.*, 10: 113-121.
- VERZANI**, J. 2014. *Using R for Introductory Statistics*. CRC Press, Boca Raton, FL, EUA: 502 p.
- WALFORD**, L.A., 1946. A new graphic method of describing the growth of animals. *Biology Bulletin* 90(2): 141-147.
- WELDON**, W.F.R., 1893. On certain correlated variations in *Carcinus moenas*. *Proceedings of the Royal Society of London* 54: 318-329.
- WILSON**, C.A., D.L. Nieland y A.L. Stanley. 1993. *Variation of year-class strength and annual reproductive output of red drum Sciaenops ocellatus and black drum Pogonias cromis from the northern Gulf of Mexico. Final Report 1991–1992*, 31 p. U.S. Dept. Commerce Cooperative Agreement NA90AAH-MF724. Marine Fisheries Initiative (MARFIN) Program. Coastal Fisheries Institute. Louisiana State University, Baton Rouge, La 70803-7503
- WINSOR**, C.P. 1932. The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences*, 18(1): 1-8.
- WOLFF**, M., 1989. A proposed method for standardization of the selection of class intervals for length frequency analysis. *Fishbyte*, 7: 5.
- WONG**, M.A., 1985. A bootstrap testing procedure for investigating the number of subpopulations. *Journal of Statistical Computation and Simulation*, 22: 99-112.
- ZULLINGER**, E.M., R.E. Ricklefs, K.H. Redford & G.M. Mace, 1984. Fitting sigmoidal equations to mammalian growth curves. *Journal of Mammalogy*, 56(4): 607-636.
- ZWEIFEL**, J.R. y L. Lasker, 1976. Prehatch and posthatch growth of fishes. A general model. *Fishery Bulletin*, 74: 609-621.

Apéndices



Apéndice 1

Uso del programa AED2020



A continuación se presenta un breve tutorial para el programa AED2020 tal como se presenta en el archivo readme.txt que lo acompaña.



El programa AED2020 es un ejecutable escrito en Visual Basic Ver. 6.0 (Microsoft Corporation, 1998) que integra algunos de los procedimientos más conocidos para el cálculo de métodos exploratorios, tales como:

- Suavización no lineal resistente (Velleman y Hoaglin, 1981; Goodall, 1990)
- Estimadores de densidad por kernel
- “Trazas de densidad” presentadas en Chambers, *et al.* (1983)



- Estimador de densidad por kernel gaussiano de amplitud variable (Fox, 1990; Salgado-Ugarte, *et al.* 1993; Salgado-Ugarte y Pérez-Hernández, 2003).

Esta versión (derivada de sus predecesores los programas EDK2000 y AED2010) sigue siendo muy simple pero con la ventaja de funcionar en el ambiente Windows e integra además métodos no incluidos en el anterior:

- Tres suavizadores no lineales resistentes (4253EH, 3RRSH y 53EH) con la opción de re-suavización (procedimiento “doble”) para cada uno
- Dos métodos de validación cruzada (por mínimos cuadrados y sesgada) para determinación de amplitud de banda de estimadores de densidad por kernel y para regresión no paramétrica (por kernel)
- Rutina auxiliar para encontrar las bandas críticas, el método de bootstrap suavizado para la evaluación no paramétrica de la multimodalidad y regresión no paramétrica por kernel.

Por lo anterior, consideramos que el programa AED2020 es un auxiliar valioso y muy poderoso para el análisis exploratorio (y en el caso de la prueba de multimodalidad, confirmatorio) de datos. A continuación la explicación puntual de las rutinas.



**Programa AED2020 Versión 1.1.0****Mayo, 2020**

Conjunto de programas exploratorios y de cómputo intensivo para el análisis de datos uni y bivariados

- Estimación no paramétrica de la densidad por medio de diferentes funciones ponderales “kerneles”
- Determinación de ancho de banda por validación cruzada
- Evaluación de multimodalidad por método bootstrap
- Regresión no paramétrica por kernel y
- Suavización no lineal resistente.

Derechos Reservados (Copyright): Isaías Hazarmabeth Salgado Ugarte, Verónica Mitsui Saito Quezada y José R. Rodríguez Rojas, D.G.A.P.A. y F.E.S. Zaragoza, UNAM Proyectos P.A.P.I.I.T. IN217596, P.A.P.I.M.E. 192031, EN221403, PE205407 y PE207417, México, 2010-2020. Patente en trámite.

Requerimientos:

Sistema 486 o superior (Pentium recomendado)

S.O. Windows 95, 98, Millenium, XP o 10.

Guía breve de uso:

A la fecha se proporciona la siguiente información. Versiones posteriores contarán con ayuda dentro del programa.

El programa de instalación debe copiarse a una carpeta propia en el disco duro, ejecutarse y seguir las instrucciones de instalación.

Para su ejecución puede utilizarse el Explorador de Windows abriendo la carpeta donde se encuentra el programa y oprimiendo dos veces (doble click) el botón del ratón en el icono del archivo ejecutable “aed2020.exe” o bien en el ícono creado al instalarse.

Se abre una ventana de presentación que se cierra al oprimir el botón del ratón en el centro de dicha ventana. Enseguida aparece la pantalla principal del programa con las siguientes opciones:



Archivo

- Abrir (Una Variable)...
- Abrir (Valores x,y)...
- Guardar resultados...
- Guardar gráfica...
- Cerrar

Validación Cruzada

- Mínimos cuadrados
- Sesgada
- Para Regresión

Estimar

- Traza de densidad
- EDK
- EDKVariable(Gauss)
- WARP (PPPR)
- Amplitud Crítica
- Prueba de Silverman
- Regresión por kernel (PPPR)
- Suavización no Lineal Resistente

Ayuda

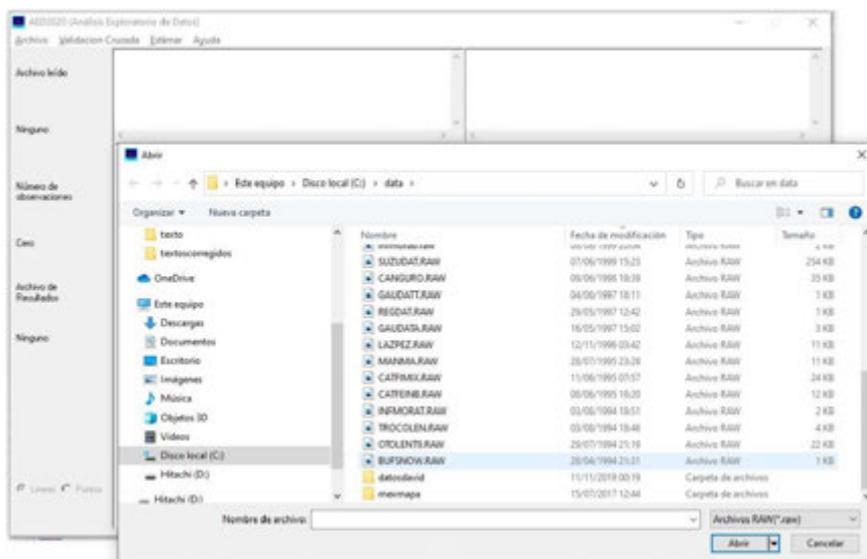
- Acerca de





El menú “Archivo-Abrir (Una Variable)” permite la lectura de un archivo de texto con los datos numéricos de una sola variable mediante la apertura de un cuadro de diálogo en el que se pueden seleccionar archivos con tres extensiones: “.raw” (extensión para archivos de texto creados con Stata), “.txt” (pre-establecido) y cualquier otra siempre y cuando el formato sea de texto en una sola columna y con el fin de archivo debajo del último número.

Al abrir el archivo, en la parte izquierda de la pantalla se actualizan los datos sobre “Archivo leído” y “Número de observaciones”, y aparece una nueva ventana con un resumen estadístico básico de los datos leídos. La ventana principal se divide en tres partes: en la superior izquierda se presenta la lista de los datos y en la superior derecha aparece la lista de resultados cuando se aplica un método.



Con los datos en la memoria es posible abrir el Menú “Estimar”. La opción “Traza de densidad” permite calcular los estimadores presentados por Chambers, *et al.* 1983.

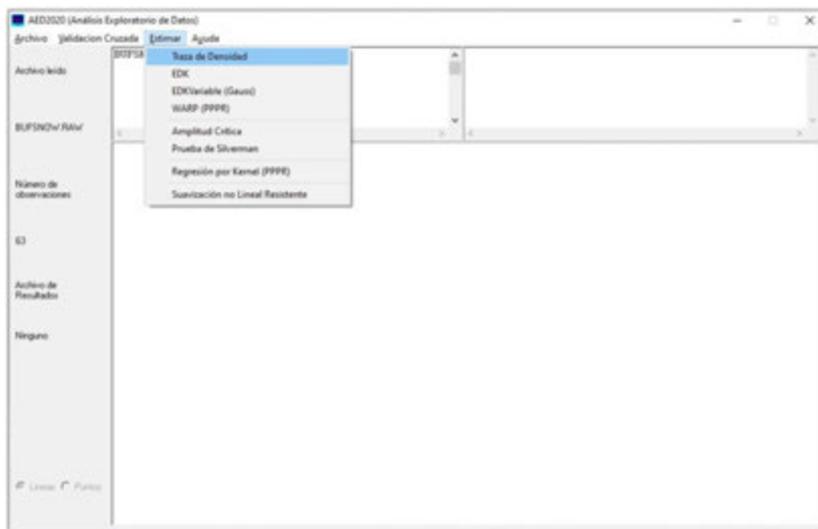


Al seleccionarla se abre un cuadro de diálogo que permite elegir función ponderal de dos opciones “Cuadrada” y “Coseno”; además se debe escoger la forma de cálculo entre “Directa” para estimar la densidad de cada dato ó “Discreta” que utiliza una trama de puntos uniformemente espaciados desde un valor un poco menor al mínimo y hasta un poco más allá del máximo. De manera pre-establecida, el número de puntos es de 50, pero ese valor puede cambiarse por el usuario en el campo que aparece al escoger esta opción de cálculo. Asimismo, se debe especificar la amplitud de ventana (ancho de banda) para lo cual el cuadro presenta un botón para el cálculo de valores de referencia: dos óptimos (Silverman y Härdle) y un sobresuavizado (Scott). Al presionar el botón “Calcular ventanas” aparecen los tres valores y en el campo de entrada se dispone del valor intermedio (óptima de Härdle).

La estimación se lleva a cabo oprimiendo el botón “Calcular Densidad” lo que produce el listado de resultados en el campo superior derecho y un gráfico con la densidad (eje vertical) para la variable (eje horizontal). En la parte inferior izquierda de la ventana principal existen dos selectores de opción para la gráfica: la opción “línea” es la pre-establecida y muestra los valores de densidad unidos por líneas acompañados de un gráfico univariado de dispersión en el eje de las abscisas (horizontal). La opción “puntos” presenta los valores de densidad en su respectiva posición de acuerdo a su valor como puntos aislados. Estas opciones son válidas para todas las gráficas de los estimadores de densidad subsiguientes.



Los resultados pueden almacenarse en un archivo de texto mediante la opción “Archivar Guardar Resultados...” opción que hace aparecer un cuadro de diálogo estándar de Windows en el que se puede elegir el lugar de almacenaje, crear una carpeta nueva y especificar el nombre del archivo el cual es de texto (extensión “txt”). Este archivo tiene dos columnas: la primera son los valores observados “Valor de X” (Directa) ó los “Valores medios” (Discreta) y la segunda es la estimación de la densidad. Si no se especifica valor para el ancho de ventana, aparece un mensaje de error y se regresa al cuadro de diálogo.



En la presente versión no se ha implementado la opción “Guardar Gráfica...”, sin embargo, es posible utilizar la función de copia de pantalla (Impr Pant) para obtener una imagen de la pantalla actual (con la gráfica) y pegar ésta en un programa de edición gráfica, seleccionar la parte de la gráfica, copiar la selección y volverla a pegar en como imagen distinta en el mismo programa o en otro (como procesador de textos). Esta forma de copia de gráfica es posible para todas las gráficas producidas por el programa (como el caso de la regresión por kernel y las gráficas de validación cruzada).



Traza de Densidad

Función Ponderal (Kernel)

Cuadrada

Coseno

Densidad

Directa

Discreta

Amplitud de ventana

Optima de Silveiman

Optima de Haerde

Sobresuavizada de Scott

Calcular ventanas

Calcular Densidad

Traza de Densidad

Función Ponderal (Kernel)

Cuadrada

Coseno

Densidad

Directa

Discreta

Amplitud de ventana

38.2057

Optima de Silveiman

Optima de Haerde

Sobresuavizada de Scott

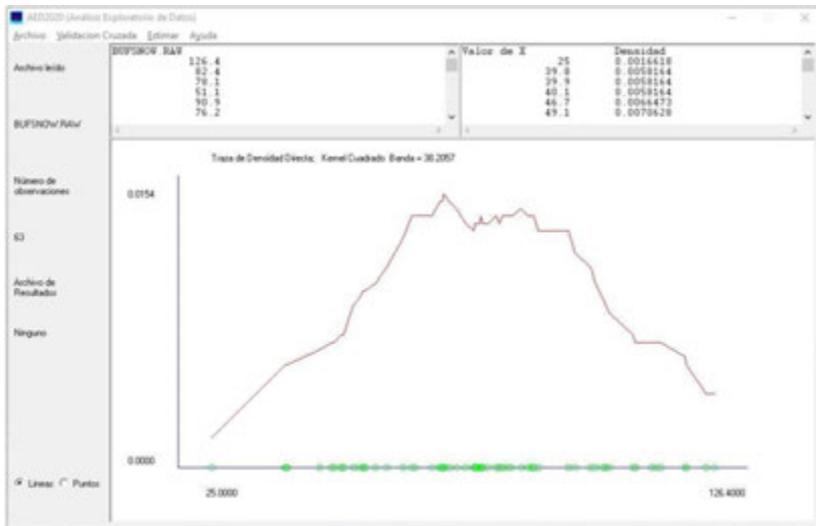
32.4388

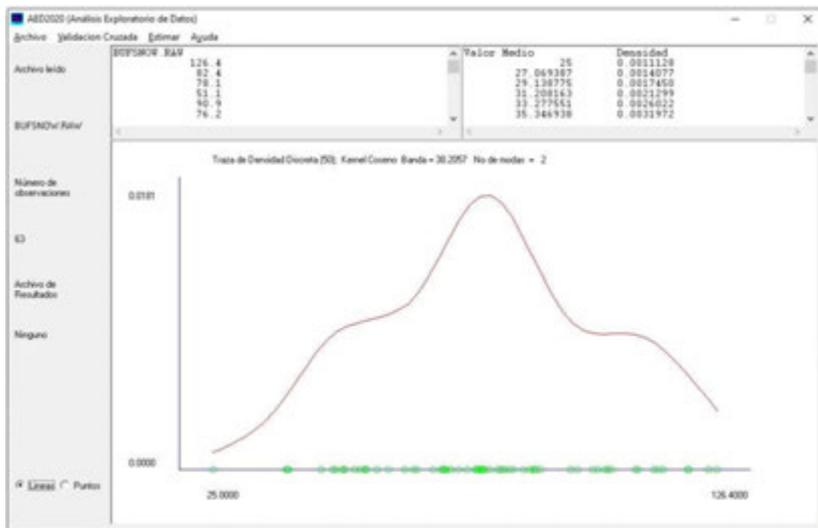
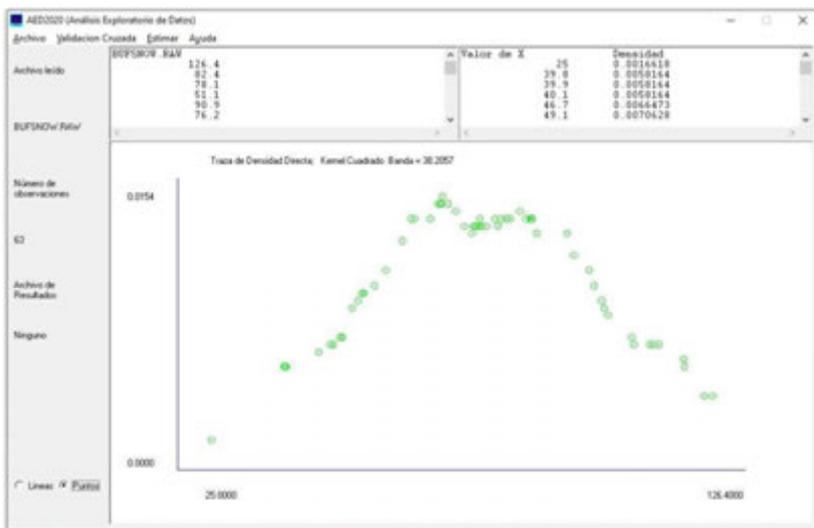
38.2057

41.2333

Calcular ventanas

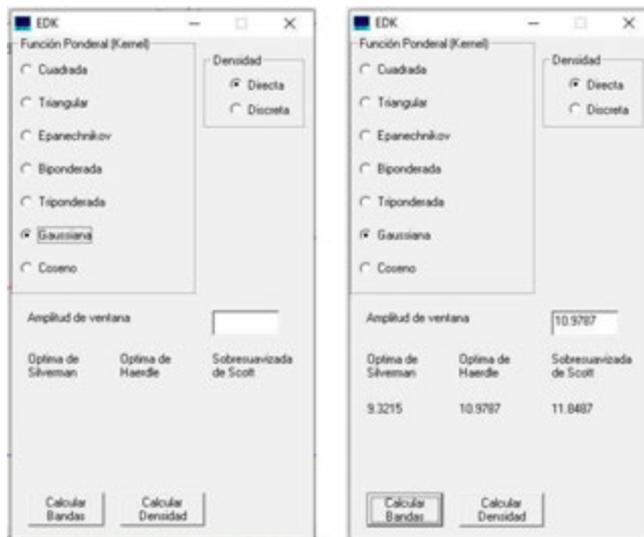
Calcular Densidad





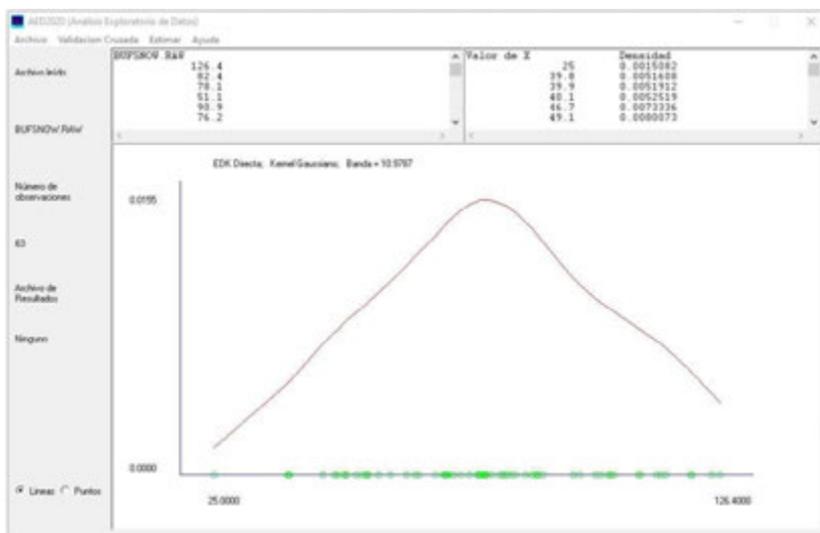


La opción “EDK” permite calcular los Estimadores de Densidad por Kernel presentados por Härdle (1991) y Scott (1992). Para detalles ver los artículos de Salgado-Ugarte (1993; 1995a; 1995b) y el texto de Salgado-Ugarte (2002). Al elegir esta opción aparece un cuadro de diálogo en donde es posible establecer la función ponderal (kernel) de un total de siete opciones: Cuadrada, Triangular, Epanechnikov, Biponderada, Triponderada, Gaussiana (pre-establecida) y Coseno. Se puede elegir también la forma de cálculo: “Directa” (pre-establecida) o “Discreta”. En este último caso el número de puntos pre-establecido es de 50. El usuario puede utilizar un número mayor si lo requiere.



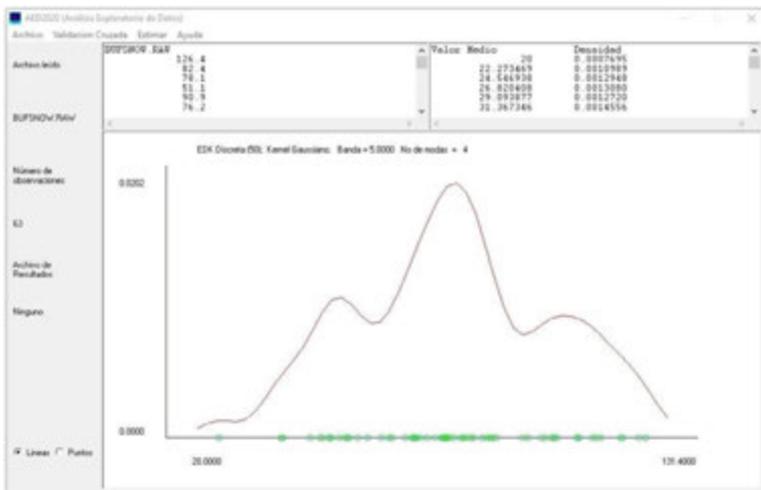
De nuevo se tienen los botones para calcular las bandas y la densidad. Los resultados (lista y gráfica) aparecen como en la opción “Traza de Densidad”. Es posible crear el archivo de resultados pero no se ha implementado aún “Guardar Gráfica”. Las bandas calculadas (dos óptimas y una sobresuavizada) automáticamente se ajustan al tipo de función ponderal (kernel) especificado de acuerdo a la idea de los “kerneles equivalentes” de Scott (1992) para producir estimaciones con el mismo grado de suavización.

Si no se especifica valor para el ancho de banda y se oprime el botón de “Calcular Densidad” aparece un mensaje de error y se regresa al cuadro de diálogo.



La opción “EDKVariable(Gauss)” permite calcular el Estimador de Densidad por Kernel Gaussiano con banda de amplitud variable (ver Fox, 1990, Salgado-Ugarte, *et al.* 1993 y Salgado-Ugarte y Pérez Hernández, 2003).





Este estimador resulta muy conveniente ya que proporciona detalles de la distribución donde los datos abundan (empleando bandas angostas) y disminuye el ruido donde los datos son escasos (con bandas más anchas). Para lograr esto el algoritmo realiza una estimación preliminar de la densidad, la cual es usada como referencia para en un segundo paso, calcular la densidad tomando a los valores iniciales como factores de ajuste a la banda.

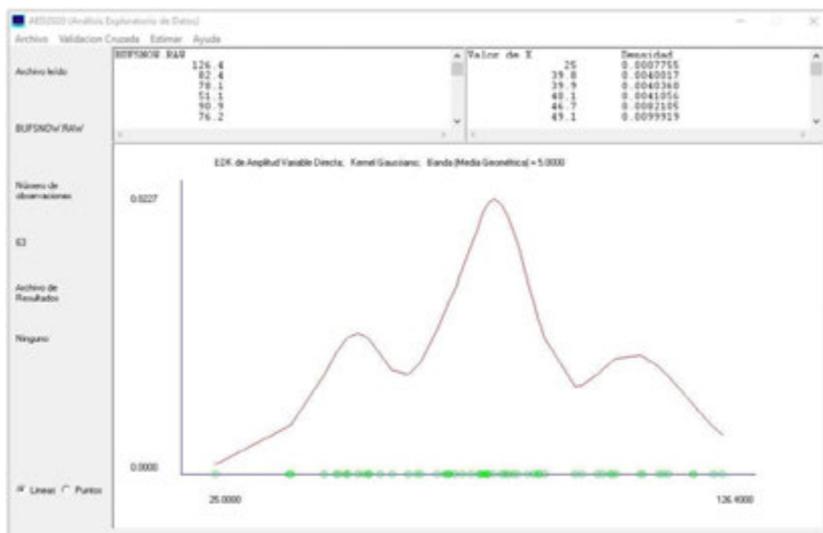
This dialog box is titled 'EDK Gaussiano de amplitud variable'. It has a 'Densidad' section with radio buttons for 'Directa' (selected) and 'Discreta'. Below this are three options: 'Óptima de Silverman', 'Óptima de Haerde', and 'Sobresuavizada de Scott'. At the bottom, there are buttons for 'Calcular Bandas', 'Calcular Densidad', and 'Terminado'. The 'Amplitud de banda (Meda geométrica)' field contains the value '5'.

This dialog box is titled 'EDK Gaussiano de amplitud variable'. It has a 'Densidad' section with radio buttons for 'Directa' and 'Discreta' (selected). Below this are three options: 'Óptima de Silverman', 'Óptima de Haerde', and 'Sobresuavizada de Scott'. At the bottom, there are buttons for 'Calcular Bandas', 'Calcular Densidad', and 'Terminado'. The 'Amplitud de banda (Meda geométrica)' field contains the value '5' and the 'Número de puntos' field contains the value '100'.



A semejanza de los procedimientos anteriores al elegirla se abre un cuadro de diálogo que permite escoger la forma de cálculo: “Directa” o “Discreta”. En la “Directa” se utilizan todos los datos para el cálculo y en la “Discreta” se usa un número de puntos uniformemente espaciados en la segunda parte del algoritmo de cálculo (lo que resulta más eficiente). El valor pre-establecido es de 50, pero recomendamos usar 100 puntos. Este método sólo se ha implementado para el kernel gaussiano y permite establecer el ancho de banda (media geométrica de todas las bandas utilizadas) y de nuevo es posible aplicar el botón “Calcular Bandas” para que aparezcan los valores en el cuadro para utilizarlos como referencia. De nuevo, en el campo de entrada aparece el valor intermedio (óptimo de Härdle).

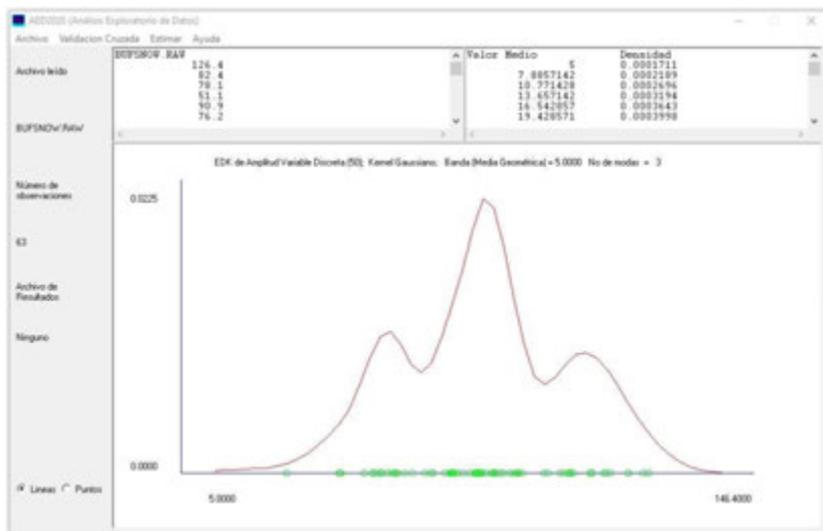
Debido a la cantidad de cálculos requeridos, esta opción puede llevar un tiempo considerable si el número de los datos es elevado (miles de observaciones) por lo que se pide paciencia al usuario. Este método está considerado como cerca de ser el mejor posible para estimar la distribución de un conjunto de datos (Simonoff, 1996). Como con las opciones anteriores es posible guardar en un archivo de texto los resultados, pero no la gráfica.



La opción “WARP” permite el cálculo de “histogramas desplazados promediados”, una forma eficiente para obtener a los Estimadores de densidad por kernel. Este procedimiento reduce considerablemente el número de operaciones. Al elegirla aparece un cuadro de



diálogo para especificar una de las siguientes funciones ponderales (kerneles): Cuadrada, Triangular, Epanechnikov, Biponderada, Triponderada y Gaussiana. Se cuenta con los botones para calcular las bandas y la densidad.



WARPing

Función Ponderal (Kernel)

- Cuadrada
- Triangular
- Epanechnikov
- Biponderada
- Triponderada
- Gaussiana

Amplitud de banda: 5

Óptimo de Silverman: 9.3215 Óptimo de Haerde: 10.9787 Sobresupercada de Scott: 11.8487

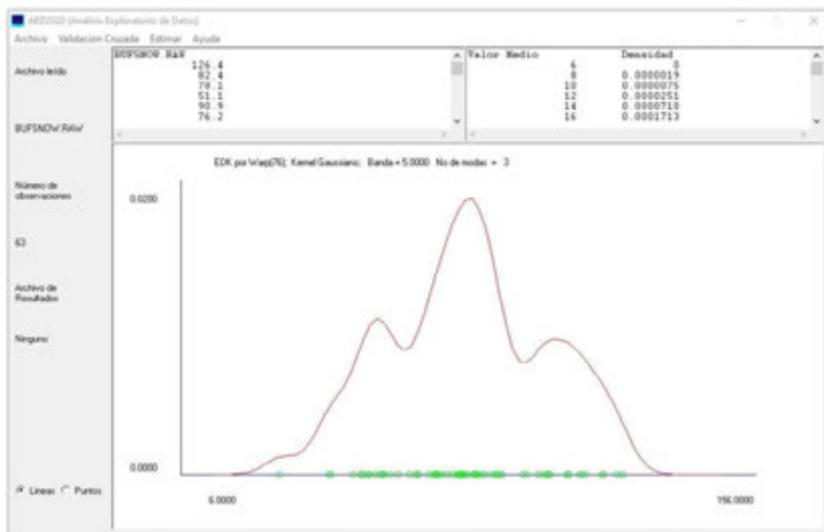
Número de histogramas: 10

Calcular Bandas Calcular Densidad Terminado



Este método requiere especificar el número de histogramas a promediar el cual de manera pre-establecida es de 5, valor que conduce a un histograma desplazado promediado. Utilizando un valor mayor (10 o más) el resultado es equivalente a un Estimador de Densidad por Kernel. Se recomienda utilizar un valor de 10 para no aumentar el número de operaciones y disminuir la eficiencia del método. Nuevamente al oprimir el botón “Calcular Densidad” aparece la lista de resultados y la gráfica (densidad y univariado de dispersión).

Es posible guardar los resultados en un archivo de texto. No está implementado aún almacenar la gráfica. En todos los casos, las rutinas gráficas incluidas son muy simples (presentan valores numéricos máximos y mínimos para cada eje y un encabezado con el estimador utilizado, el número de puntos usado para la estimación en el caso discreto, el kernel y la amplitud de banda y el número de modas) pero son de gran utilidad ya que permiten observar de una manera muy clara la forma de la distribución de los datos. Si se desean gráficos de mejor presentación, los resultados almacenados como texto pueden importarse a otro programa capaz de realizar gráficos para presentaciones.

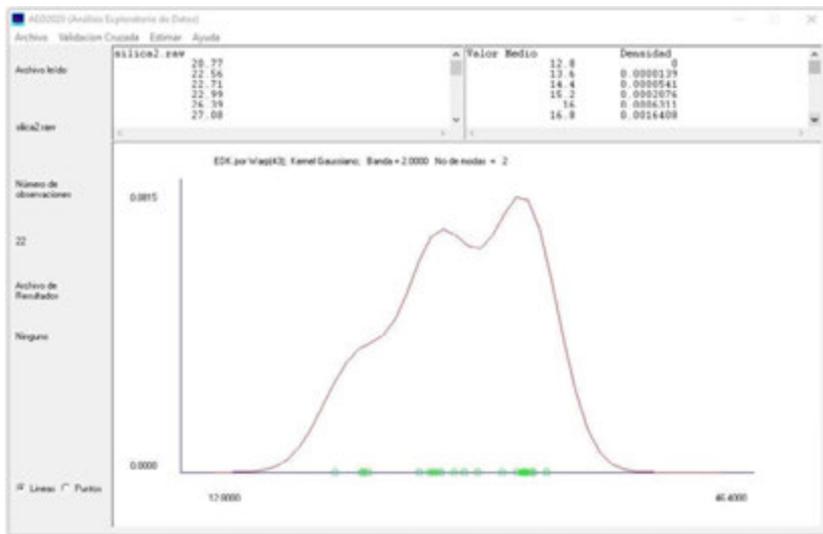


El siguiente grupo de opciones es útil para llevar a cabo la prueba de multimodalidad de Silverman (bootstrap suavizado). La opción “Amplitud Crítica” facilita la identificación de



las amplitudes de banda críticas, es decir el último valor compatible con un número dado de modas en la estimación. Al escogerla se abre un cuadro de diálogo con varios campos para especificar: el límite superior e inferior del intervalo de amplitudes en el cual buscar el valor crítico; el número de histogramas promediados para la estimación (se recomienda el valor pre-establecido de 40) y el valor para la distancia entre los puntos (“step”) dentro de los intervalos. Este valor depende de la precisión original de los datos. Esta rutina calcula un EDK con la amplitud de banda determinada por el intervalo y el valor de cada paso y cuenta el número de modas basándose en primeras diferencias.

Se recomienda usar primero valores amplios y posteriormente valores pequeños una vez que se han identificado las regiones donde se localizan los valores críticos. Aquí es pertinente anotar que se deben utilizar un poco más de posiciones decimales que los presentes en las observaciones generales (una o dos a lo más).



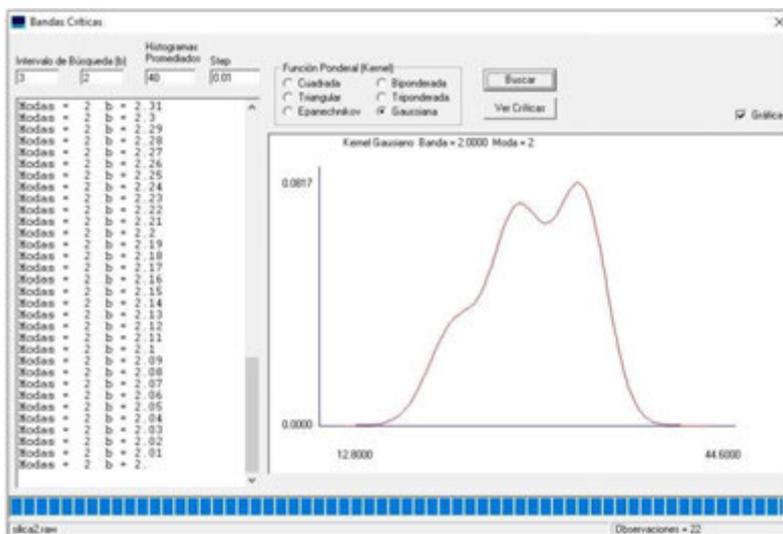
En el mismo cuadro es posible escoger la función ponderal de un total de 6: Cuadrada, Triangular, Epanechnikov, Biponderada, Triponderada y Gaussiana (pre-establecida). Es recomendable utilizar la Gaussiana ya que la prueba de multimodalidad se implementó y valoró con este kernel. Es posible también activar la opción “Gráfica”, la cual permite visualizar la gráfica de densidad para cada uno de los pasos en el intervalo especificado.



El resultado es una “animación” que permite observar como al disminuir el ancho de banda el número de modas en el resultado se va incrementando. En el campo izquierdo aparece un listado con resultados que consisten en el número de modas y la amplitud asociada. El último valor compatible con un número de modas especificado es la “amplitud crítica”.

En la práctica es posible observar cierta fluctuación en la vecindad del valor crítico, por lo que se recomienda hacer un recorrido amplio alrededor del posible valor crítico. Eventualmente se llega a un aumento en el número de modas que ya no disminuye, y el valor crítico es el valor anterior con el valor modal menor especificado.

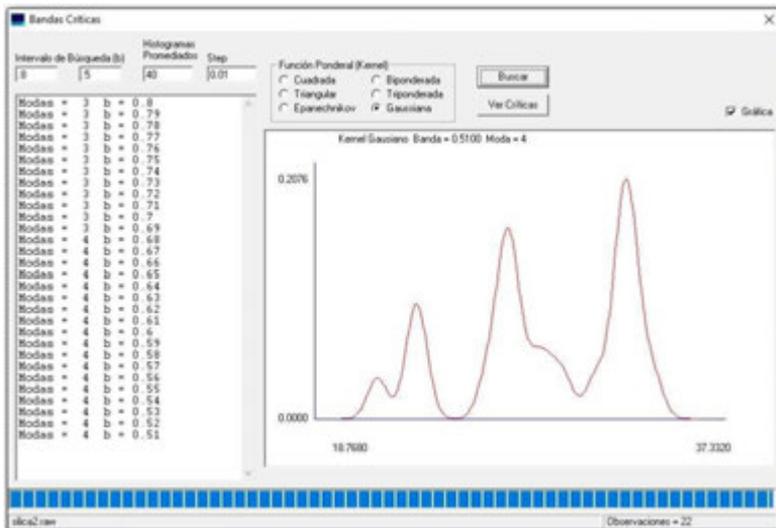
Una vez identificadas las amplitudes críticas es posible proseguir con la prueba de multimodalidad de Silverman. Esta prueba se basa en un esquema de muestreo con repetición “bootstrap”. Al activar la opción “Prueba de Silverman” se abre un cuadro de diálogo con campos para especificar la “banda crítica” (el valor encontrado con la opción anterior); el “No. Crítico de Modas” (el número de modas a probar); el número de histogramas promediados (se recomienda el valor pre-establecido de 40); el “No. de Replicaciones” el cual se ha pre-establecido como 100 (valor utilizado en el artículo original del método) pero actualmente es posible utilizar valores mayores.

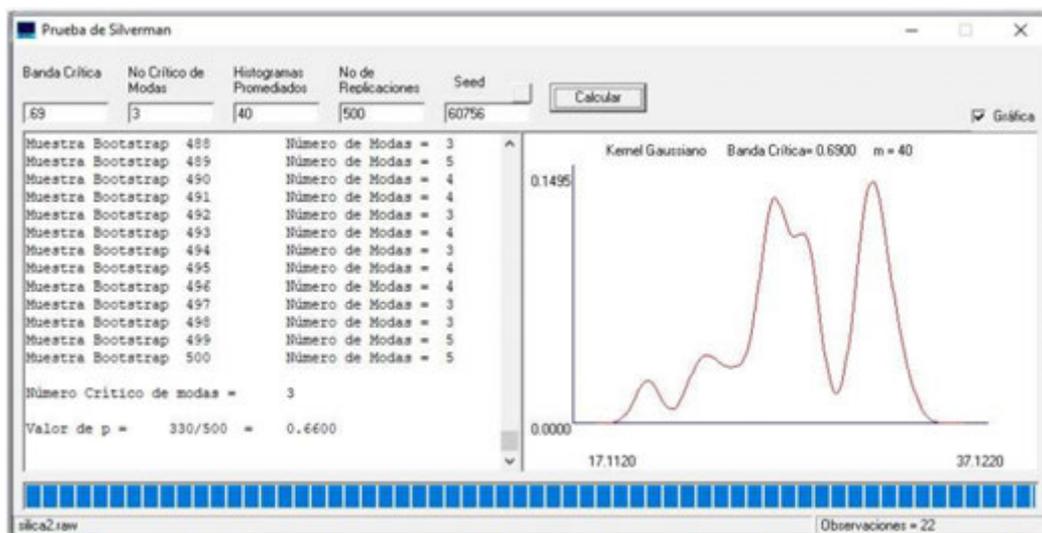




Dependiendo del número de datos este puede ser de 500, 1000 ó más aunque se recomienda prudencia en el número de replicaciones. De todas formas, es posible acumular los resultados y si se hacen 10 corridas de 1000 cada una, se finalizará con un total de 10000 replicaciones.

Cabe señalar que las repeticiones requeridas para la estimación eficiente de valores de nivel y dispersión así como de coeficientes de regresión son de unos pocos miles. No es necesario un número mayor si bien, con este programa es posible investigar la cuestión. Es importante también establecer el valor inicial para los números pseudo-aleatorios generados. Estas “semillas” deben ser anotadas en todo estudio serio de simulación ya que al utilizar el mismo número permite la reproducción exacta del muestreo realizado (cosa posible con las computadoras pero imposibles por otros métodos aleatorios). El cuadro contiene un botón que permite la generación “aleatoria” de valores para ser utilizados como semilla. Se puede activar la opción “Gráfica” para observar el EDK para cada una de las muestras bootstrap las cuales aparecerán al oprimir el botón “Calcular”.



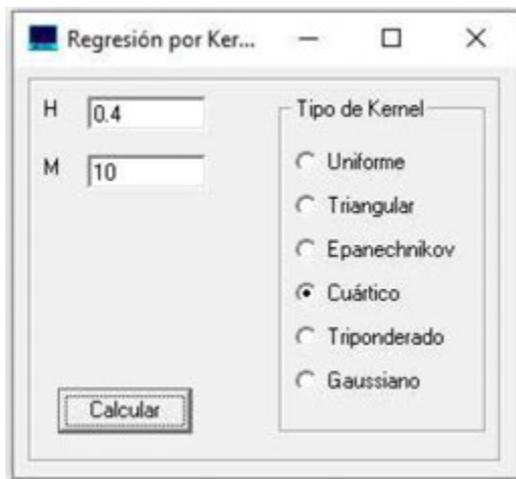


El resultado de la prueba se presenta en el campo izquierdo como una lista de valores con el número de muestra y el número de modas que cada una posee. Al final se obtiene un valor de P obtenido de la división del número de estimaciones con un número mayor al probado entre el total de replicaciones. El valor de referencia utilizado proviene de la recomendación de Izenman y Sommer (1988) de 0.4 para señalar el número de modas en la muestra. Al mostrar los valores para el cociente se hace posible acumular resultados de varias corridas. Cabe señalar que estas rutinas son hasta donde hemos podido investigar, la única versión de la prueba disponible aparte de la presentada en Salgado-Ugarte (1997 y 2002) con rutinas para el programa estadístico Stata.

La opción Regresión por Kernel (PPPR) permite calcular regresión no paramétrica por kernel mediante el método de promedio de histogramas desplazados (WARP por sus siglas en inglés ó PPPR por referirse al Promedio Ponderado de Puntos Redondeados, su equivalente en español, Salgado-Ugarte, 2002).

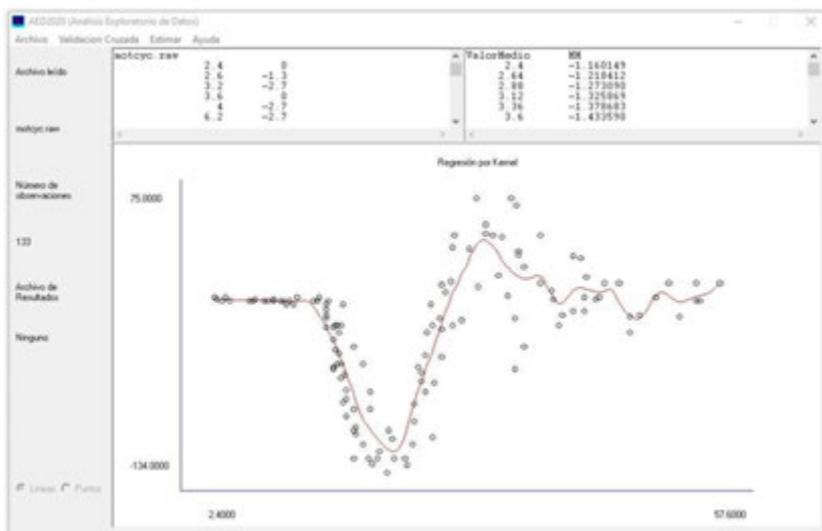


Para ello se requiere tener datos bivariados (x,y). Con los datos en la memoria, al seleccionar esta opción se abre un cuadro de diálogo que solicita especificar el ancho de banda (el valor pre-establecido de 0.4 tan sólo es un ejemplo y no es utilizable en general) y el número de histogramas a promediar (se recomienda utilizar el valor pre-establecido de 10). En el mismo se requiere especificar la función ponderal. En la literatura (Härdle, 1990) se recomienda el uso del kernel cuártico por lo que se ha puesto como opción pre-establecida.





Cómo recomendación para encontrar un valor de banda adecuado se recomienda utilizar los valores óptimos y el sobreesuavizado para los datos del eje horizontal, los cuales pueden estimarse separando a esta variable en un archivo de texto y leyéndolo para su uso con las rutinas de EDKs. Una estrategia más directa es utilizar la opción de “Validación Cruzada-Para Regresión” en el AED2020.



Al especificar los anteriores valores y función kernel se oprime el botón “Calcular” y aparecen los resultados: en el campo superior derecho se tiene la lista de los puntos de cálculo (Valor Medio) y los valores estimados (MM) y en el campo principal la gráfica x,y de los valores observados y los estimados por la regresión por kernel. Para esta gráfica, los selectores de “línea” y “puntos” no aplican. Es posible guardar estas dos columnas de datos en un archivo de texto mediante la opción “Guardar Resultados...”, pero no la gráfica.



Suavización no Lineal Resistente

Suavizadores:

- 4253EH
- Doble
- 3RSSH
- Doble
- 53EH
- Doble

39	60	55.5703125
40	54	55.1054687
41	55	54.6601562
42	50	54.203125
43	55	53.140625
44	54	51.7226562
45	46	50.3242187
46	48	48.8164062
47	54	47.6992187
48	42	47.1601562
49	44	46.8242187
50	62	46.625
51	49	46.0429687
52	41	44.8125
53	45	43.8476562
54	39	43.96875
55	49	45.7578125
56	46	49.21875
57	54	53.0234375
58	54	55.53125
59	60	56.28125
60	58	55.7929687
61	52	53.8710937
62	47	51.1796875
63	52	49.484375
64	38	48.5195312
65	55	47.2773437
66	45	46.25
67	47	45.609375
68	41	44.953125
69	48	44.625
70	42	44.9453125
71	45	45.8359375
72	48	47.1953125
73	52	48.8867187
74	49	50.7734375
75	53	53

Realizar

TEMPCOM.TXT

Obs = 75

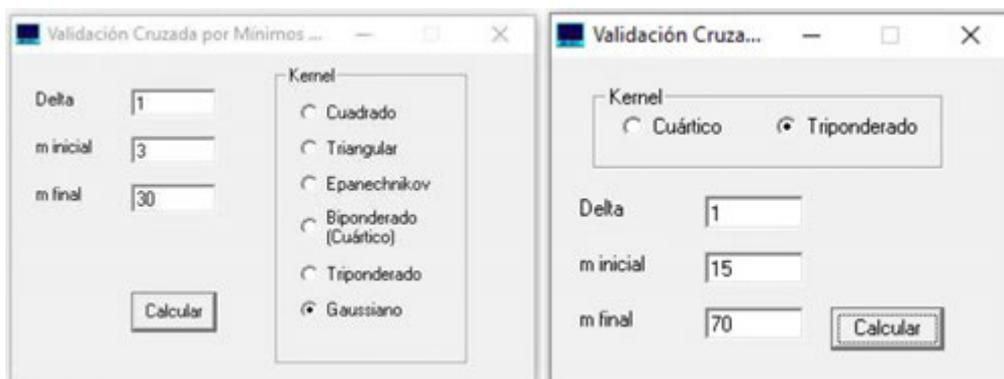
La rutina de “Suavización no Lineal Resistente” abre un cuadro de diálogo que permite elegir el método de suavización compuesto: 4253EH, 3RSSH y 53EH. Cada uno de ellos puede hacerse “doble” para recuperar información de los valores rugosos (residuales) y reincorporarla a la suavización.

Los resultados aparecen listados después de oprimir el botón “Realizar”. Esta lista con el número de dato, valor original y valores suavizados puede seleccionarse con el “ratón”, copiarse y pegarse en una hoja de cálculo o un procesador de textos para luego copiarse a un paquete de gráficas. La opción “Guardar Resultados” no está implementada aún para este procedimiento.



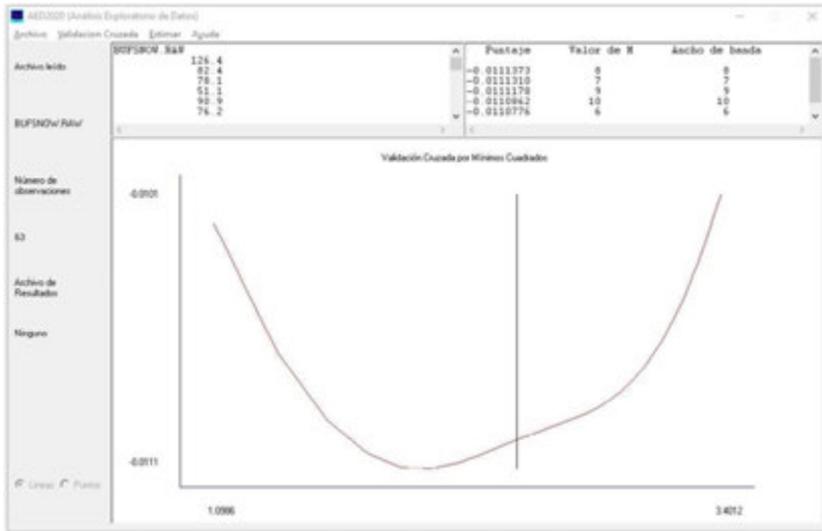
La opción “Validación Cruzada” puede aplicarse en la búsqueda del ancho de banda apropiado para datos univariados (EDKs) y bivariados (regresión por kernel).

Para el primer caso se tiene la opción de VC por “Mínimos Cuadrados” la cual al ser elegida abre un cuadro de diálogo que permite establecer el valor de delta (equivalente al paso de búsqueda en un intervalo) cuyo valor pre-establecido es recomendado para utilizarlo en primera instancia; asimismo se tiene que especificar un intervalo dentro del cual buscar el cual puede entenderse como intervalo de anchos de banda dentro de los cuales buscar un óptimo, equivalencia uno a uno cuando delta es 1; si delta es 0.5 entonces la banda se obtiene multiplicando al doble los valores del intervalo; si delta es 0.1 entonces los valores del intervalo se multiplican por diez, etc.



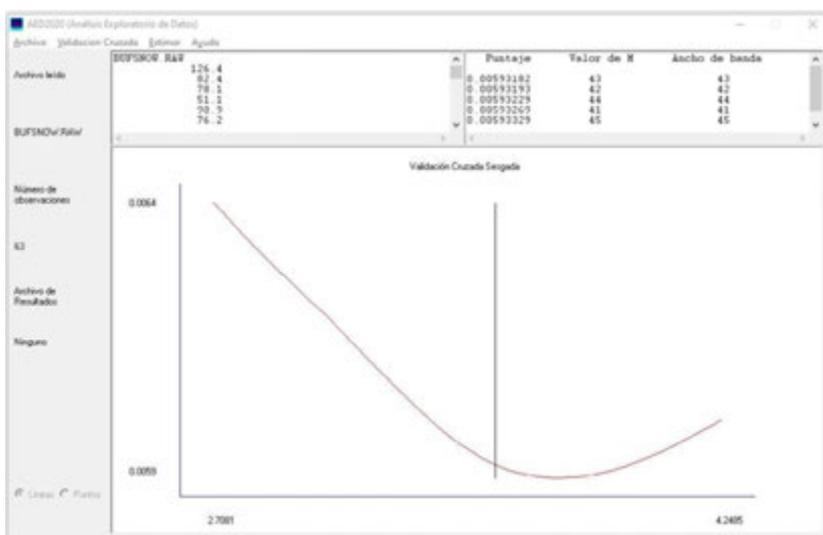
El resultado de la rutina es una lista en el campo superior derecho que contiene los puntajes, el valor de M (escalado por delta) y el ancho de banda (M multiplicado por delta de acuerdo a lo arriba señalado). Estos valores están ordenados por el puntaje de menor a mayor, por lo que el ancho de banda recomendado por el método es el que aparece en el primer renglón. Este valor se puede utilizar como ancho de banda en las rutinas para estimación de EDKs (EDK, WARP o inclusive EDKVariable).

En el campo principal aparece la gráfica con los valores del puntaje en el eje vertical y los de M en el horizontal (en escala logarítmica neperiana). Se marca además como una línea vertical la banda sobresuavizada. Se pretende encontrar un mínimo en los puntajes y tan sólo será de interés si este mínimo es menor al valor señalado por la banda sobresuavizada. La lista completa de puntajes, valores de M y ancho de banda puede almacenarse en un archivo de texto con la opción “Guardar Resultados” no así la gráfica.



La VC “Sesgada” muestra un cuadro de diálogo que permite escoger una de las dos opciones implementadas para la función ponderal (kernel): cuártico y triponderado, así como el valor de delta (se recomienda en primera instancia el valor pre-establecido de uno) y los límites inferior (inicial) y superior (final) para la búsqueda del ancho óptimo en términos de M (número de histogramas promediados). Los resultados se presentan de manera semejante a la VC por mínimos cuadrados, una lista en el campo superior derecho y la gráfica en el campo principal de la ventana.

Los resultados están ordenados de acuerdo al puntaje y por tanto la anchura de banda recomendada por el método aparece en el primer renglón. La lista completa de resultados puede salvarse en un archivo de texto pero no la gráfica.



The dialog box "V.C. P. Regresión" contains the following settings:

- Delta:** 0.2
- M. Inicio:** 5
- M. Final:** 20
- Boundary:** 0.1
- Kernel:**
 - Uniforme
 - Triangular
 - Epanechnikov
 - Cuártico
 - Triponderado
 - Gaussiano
- Selector:**
 - Shibata
 - V.C.G.
 - Akaike
 - E.F.P.
 - T. de Rice

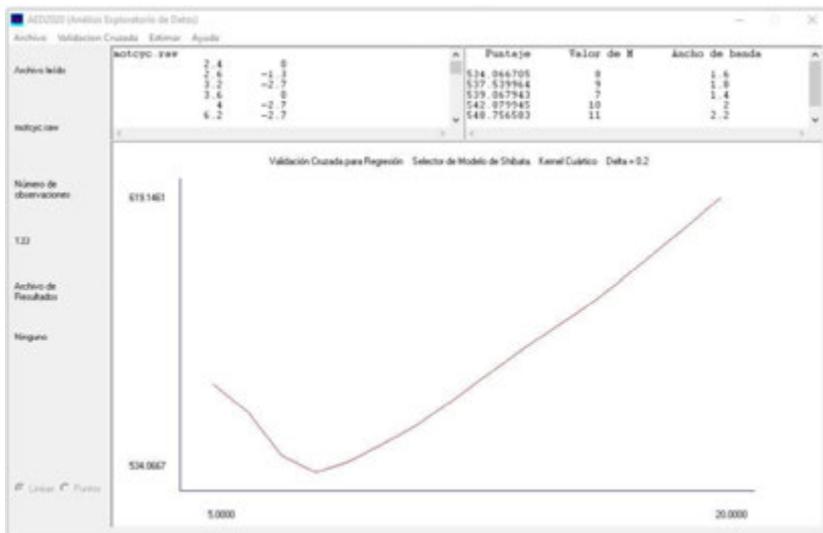
A "Calcular" button is located at the bottom left of the dialog.

La opción Validación Cruzada “Para Regresión” requiere de cargar en la memoria datos bivariados (“Abrir (Valores x,y)...”). Al seleccionarla se muestra un cuadro de diálogo con campos para especificar el parámetro delta (1), m de inicio (5), m final (20) y valor de frontera (0.1) (boundary). Se recomienda utilizar en primera instancia los valores pre-establecidos

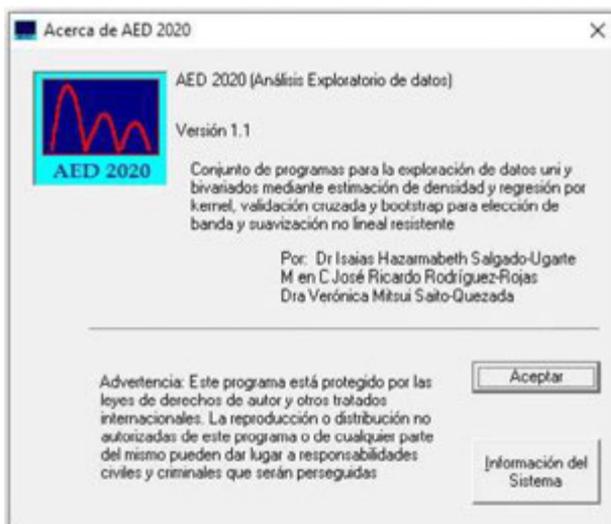


(entre paréntesis) e ir afinando las estimaciones en pasos posteriores. Se debe especificar el kernel (cuártico como pre-establecido), y la función penalizante de cinco opciones: Shibata, Validación cruzada generalizada, criterio de información de Akaike, error finito de predicción y T de Rice. La primera reduce el sesgo mientras que la última reduce la varianza, pero sin importar las diferencias, cualquiera de estos selectores conduce substancialmente a la misma amplitud de banda “óptima” (Härdle, 1991; Salgado-Ugarte, 2002).

Los resultados aparecen en el mismo patrón que para la VC univariada: lista de valores mínimos con la banda óptima recomendada en el primer renglón y la gráfica en el campo principal. La lista completa de puntajes, valores de M y amplitudes de banda puede grabarse en un archivo de texto.



Finalmente el menú “Ayuda” contiene una opción única (“Acerca de”) que abre un cuadro de diálogo con información del programa AED2020, las advertencias sobre su protección como obra intelectual, y autoría. Este cuadro tiene dos botones: “Aceptar”, que cierra el cuadro de informativo e “Información del Sistema” que proporciona características del sistema de cómputo que se está utilizando.



Para información adicional así como reporte de fallas, contactar a

Dr. Isaías Hazarmabeth Salgado Ugarte

isalgado@unam.mx

ihsalgadougarte@gmail.com

Apéndice 2

Listado de Programas para Stata



A continuación se incluye una lista de los programas (archivos con extensión “ado”) que los autores escribieron para utilizarse en esta obra. Se presentan organizados por Capítulo.

Capítulo 1:

1) bandw1.ado	bandw1.sthlp
2) warpdnm1.ado	warpdenm1.hlp
3) varwiker1.ado	varwiker1.hlp
4) varwike21.ado	varwike21.hlp
5) kerneld1.ado	kerneld1.hlp
6) l2cvwarp.ado*	l2cvwarp.hlp
7) bcvwarp.ado*	bcvwarp.hlp
8) l2cvwarp2.ado	
9) bcvwarp2.ado	
10) l2cvwarpy.ado	
11) bcvwarpy.ado	
12) critiband1.ado	critiband1.hlp
13) bootsamb.ado	
14) bhataplt1.ado	bhataplt1.hlp
15) bhatgauc1.ado	bhatgauc1.hlp
16) gaussgen1.ado	gaussgen1.hlp
17) bhatado.hlp	

* Programas sustituidos por rutinas equivalentes de programa “AED2020”. Funcionales en versiones anteriores de Windows (XP). Los programas nuevos corren con archivos .py y .exe de programas Python (sufijo “2” y “py”, respectivamente).

**Capítulo 2:**

1) nlpotfun.ado nlpotfun.hlp

Capítulo 4:

1) fordwaln.ado fordwaln.hlp
2) gullptn.ado gullptn.hlp
3) gulholtn.ado gulholtn.hlp
4) bevholtn.ado bevholtn.hlp
5) nlsvbgff.ado
6) nlvbgfn.ado nlvbgfn.hlp
7) nlsvbf2.ado
8) nlsvbfm.ado
9) nlsvbfs.ado
10) mlvbgf.ado
11) mlvbgf2.ado

Capítulo 5:

1) hotegfun.ado hotegfun.hlp
2) nlvbgfno.ado nlvbgfno.hlp
3) nlvbgfel.ado
4) nlvbgfek.ado
5) nlvbgfet.ado
6) nllog3no.ado
7) nllog3el.ado
8) nllog3ek.ado
9) nllog3et.ado
10) nlgom3no.ado
11) nlgom3el.ado
12) nlgom3ek.ado
13) nlgom3et.ado
14) vblrtest.ado vblrtest.sthlp



- 15) vbmltest.ado vbmltest.sthlp
- 16) l3lrtest.ado
- 17) l3mltest.ado
- 18) g3lrtest.ado
- 19) g3mltest.ado

Capítulo 6:

- 1) sm4253eh.ado
- 2) smtwice.ado
- 3) s4253eh.ado
- 4) s4253eht.ado

Capítulo 7:

- 1) nllog2.ado

Apéndice 3

Listado de Archivos de Datos



Capítulo 1:

- 1) bagretallas.dta
- 2) trocolen.dta
- 3) kidai.dta
- 4) ksbagretallas.dta
- 5) baghsmi.raw
- 6) baghsmi.dta

Capítulo 2:

- 1) formadatos.dta
- 2) bagtoto.dta
- 3) erizosgma.dta
- 4) erizotogama.dta
- 5) leuco2.dta
- 6) pinmazadatos.dta
- 7) evantros.dta
- 8) hexaprincepsdatos.dta
- 9) garraleon.dta
- 10) bagtoto2.dta
- 11) trofipeces.dta

Capítulo 3:

- 1) scale111.dta
- 2) bodoto2.dta
- 3) otoage113.dta



- 4) sinpulir.dta
- 5) astraeanillos.dta
- 6) miotome7.dta
- 7) modainc.dta

Capítulo 4:

- 1) catfilen.dta
- 2) vbcatage.dta
- 3) vbsuzuki.dta
- 4) anillostotales.dta
- 5) validahp.dta
- 6) anillostotales2.dta
- 7) pepinodatos.dta
- 8) seasonvbdata.dta
- 9) seasonking.dta
- 10) kimudat.dta

Capítulo 5:

- 1) kimudat.dta
- 2) vbrchpcomb.dta

Capítulo 6:

- 1) suzukt15.dta
- 2) manovre2.dta
- 3) manovre3.dta
- 4) muricantotalc.dta
- 5) anmucovahpr.dta
- 6) anmusmoothpc.dta
- 7) manov3.dta
- 8) anmucovsk.dta



Capítulo 7:

- 1) bagtotb.dta
- 2) bagtoto11.dta
- 3) hphistorepro.dta
- 4) hph1415b.dta
- 5) hpm1415.dta
- 6) hphm1415b.dta
- 7) bagtoto.dta
- 8) pmking.dta

Apéndice 4

Listado de Objetos para R



Capítulo 1:

Paquetes

- 1) foreign (paquete para importar datos en formato Stata)
- 2) ash (paquete para cálculo de histogramas desplazados promedio)
- 3) mclust (paquete para ajuste de componentes gaussianos en distribuciones mezcladas)
- 4) flexmix (paquete para ajuste de componentes gaussianos en distribuciones mezcladas)

Funciones escritas para el capítulo

- 1) bandw.R (función para cálculo de número y amplitud de intervalos/banda para histogramas, polígonos de frecuencia y estimadores de densidad por kernel)
- 2) l2cvwarpy.R (función para cálculo de validación cruzada por mínimos cuadrados para elección de amplitud de banda de EDKs)
- 3) bcvarpy.R (función para cálculo de validación cruzada sesgada para elección de amplitud de banda de EDKs)

Datos

- 1) bagtot.rda (ejemplo de grabado de datos en formato de R)



Capítulo 2:

Paquetes

- 1) foreign (paquete para importar datos en formato Stata)

Funciones

- 1) lm (regresión lineal de R)
- 2) potfun (función potencial especificada en texto)
- 3) nls (regresión no lineal: función especificada en texto)

Capítulo 4:

Paquetes

- 1) UsingR
- 2) Hmisc
- 3) FSA
- 4) FSAdata

Ejemplos (secuencia de comandos)

Ejemplo 1: crecimiento de erizo de mar del libro "Using R" datos de Grosjean (2001)

```
> library(foreign)
> library(Hmisc)
> library(UsingR)

> names(urchin.growth)

[1] "age" "size"

> plot(size~age, data=urchin.growth)
> plot(jitter(size)~jitter(age,3), data=urchin.growth)

> vbgf <- function(t,Li,k,t0) Li*(1-exp(-k*(t-t0)))
```



```
> vb.res <- nls(size~vbgf(age, Li,k,t0), start=c(Li=60,k=1,t0=0),
data=urchin.growth)
> vb.res
Nonlinear regression model
  model: size ~ vbgf(age, Li, k, t0)
  data: urchin.growth
      Li      k      t0
66.2598 0.3653 0.3721
  residual sum-of-squares: 7445

Number of iterations to convergence: 6
Achieved convergence tolerance: 2.67e-06

> curve(vbgf(x,Li=66.26, k=.37, t0=.37), add=T)

> AIC(vb.res)
[1] 1565.929

> logf <- function(t, Li, k, t0) Li*(1+exp(-k*(t-t0)))^(-1)
> log.res <- nls(size~logf(age, Li,k,t0), start=c(Li=60,k=1,t0=0),
data=urchin.growth)

> log.res
Nonlinear regression model
  model: size ~ logf(age, Li, k, t0)
  data: urchin.growth
      Li      k      t0
53.903 1.393 1.958
  residual sum-of-squares: 7299

Number of iterations to convergence: 10
Achieved convergence tolerance: 1.953e-06

> curve(logf(x,Li=53.9, k=1.39, t0=1.96), add=T, lty=2)

> AIC(log.res)
[1] 1560.963

> rgf <- function(x,Li,k,t0,m) Li*(1-exp(-k*(t-t0)))^m
> rgf.res <- nls(size~rgf(age, Li,k,t0,m), start=c(Li=53,k=.5,t0=0,m=1),
data=urchin.growth)
> rgf.res
Nonlinear regression model
  model: size ~ rgf(age, Li, k, t0, m)
  data: urchin.growth
      Li      k      t0      m
57.2649 0.7843 -0.8587 6.0636
```



```
residual sum-of-squares: 6922
```

```
Number of iterations to convergence: 9
```

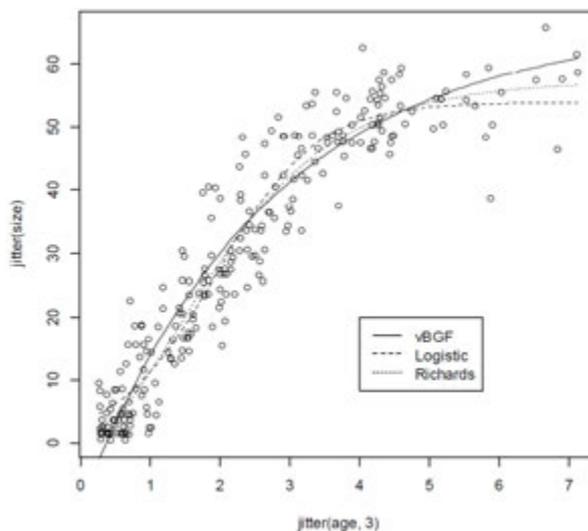
```
Achieved convergence tolerance: 1.833e-06
```

```
> curve(rgf(x,Li=57.3, k=.78, t0=-0.86,m=6.1), add=T, lty=3)
```

```
> AIC(rgf.res)
```

```
[1] 1549.703
```

```
> legend(4.20, legend=c("vBGF", "Logistic", "Richards"), lty=1:3)
```



El modelo recomendado (AIC menor) es el de Richards

Ejemplo 2: Problema 12.9 crecimiento de la corvina roja Beckman, et al. (1988); Wilson, et al. (1993) en Verzani (2014).

```
> names(reddrum)
```

```
[1] "age" "length"
```

```
> plot(length~age, data=reddrum)
```



```
> vbgf.res <- nls(length~vbgf(age, Li,k,t0), start=c(Li=45,k=.5,t0=0),
data=reddrum)
> vbgf.res
Nonlinear regression model
  model: length ~ vbgf(age, Li, k, t0)

  data: reddrum
      Li      k      t0
44.14021  0.06481 -12.50935
residual sum-of-squares: 237.3

Number of iterations to convergence: 14
Achieved convergence tolerance: 4.343e-06

> curve(vbgf(x,Li=44.14, k=.06, t0=-12.51), add=T)

> lvbgf <- function(t,Li,b1,k,t0,m) (Li+b1*t)*(1-exp(-k*(t-t0)))
> lvbgf.res <- nls(length~lvbgf(age, Li,b1,k,t0), start=c(Li=32,b1=.25,k=.5
,t0=0), data=reddrum)

> lvbgf.res
Nonlinear regression model
  model: length ~ lvbgf(age, Li, b1, k, t0)
  data: reddrum
      Li      b1      k      t0
32.1750  0.2983  0.5978  0.3002
residual sum-of-squares: 116.1

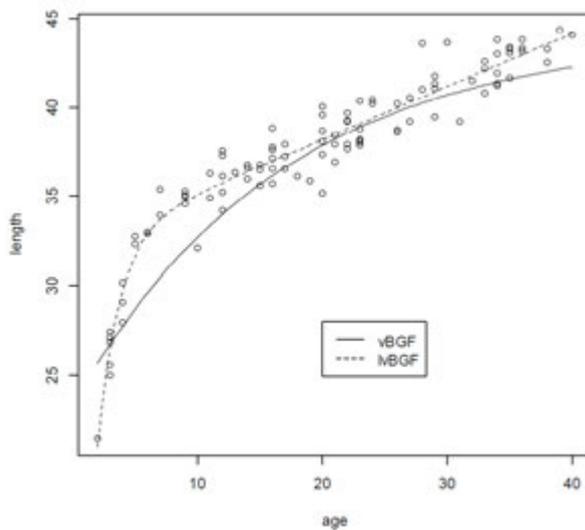
Number of iterations to convergence: 5
Achieved convergence tolerance: 9.951e-06

> curve(lvbgf(x,Li=32.18, b1 = .3, k=.6, t0=.3), add=T, lty=2)

> AIC(vbgf.res)
[1] 378.1829

> AIC(lvbgf.res)
[1] 308.6806

> legend(20,28, legend=c("vBGF", "lvBGF"), lty=1:2)
```



El modelo recomendado (AIC menor) es la curva "lineal" de von Bertalanffy

Capítulo 6:

Funciones

- 1) `s4253eht.R` (Suavizador no lineal resistente compuesto)

Apéndice 5

Comandos de Stata para

Análisis Multivariado de Covarianza



Capítulo 6:

Análisis multivariado de covarianza con Stata (nueva versión). Se presenta el análisis para todos los individuos del “Suzuki” (*Lateolabrax japonicus*) reportados en Salgado-Ugarte (1995; 2016). Con estos datos, para continuar el análisis con cada sexo es necesario especificar el selector “if” y la condición “sex==1” para utilizar los datos de los machos, “sex==2” para las hembras y “sex==3” para los indeterminados.

```
. use manov3
```

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
num	257	14.27237	18.21947	1	117
sex	257	1.719844	.6954899	1	3
sample	257	8.377432	3.955015	1	16
logebl	257	19.5975	.3219472	18.69071	20.44883
logetw	257	20.04746	.9118758	17.49181	22.2793
logegw	257	14.85294	1.962049	9.21034	18.52774
logelw	257	15.63326	.9278898	11.77529	17.6448
logescw	222	15.259	1.532721	10.81978	19.53746
logeew	257	20.1127	.9070545	17.53298	22.32358

```
. tab sample
```

sample	Freq.	Percent	Cum.
1	11	4.28	4.28
2	9	3.50	7.78
3	9	3.50	11.28
4	14	5.45	16.73
5	19	7.39	24.12
6	25	9.73	33.85



7		25	9.73	43.58
8		32	12.45	56.03
9		16	6.23	62.26
10		16	6.23	68.48
11		20	7.78	76.26
12		23	8.95	85.21
13		2	0.78	85.99
14		3	1.17	87.16
15		31	12.06	99.22
16		2	0.78	100.00

Total		257	100.00	

. tab sample sex

sample	sex			Total		
	1	2	3			
1		2	9	0		11
2		3	6	0		9
3		4	5	0		9
4		6	8	0		14
5		7	8	4		19
6		2	4	19		25
7		11	13	1		25
8		27	4	1		32
9		6	10	0		16
10		7	7	2		16
11		6	10	4		20
12		14	8	1		23
13		0	2	0		2
14		0	2	1		3
15		13	15	3		31
16		0	2	0		2

Total		108	113	36		257

. graph matrix logebl logeew logegw logelw, half ms(Oh) msize(small)
 xlab(18.5 20) diaopt(size(vlarge))

. manova logeew logegw logelw = c.logebl sample

Number of obs = 257

W = Wilks' lambda L = Lawley-Hotelling trace
 P = Pillai's trace R = Roy's largest root

Source | Statistic df F(df1, df2) = F Prob>F



1 11		20.11901	.0241439	833.30	0.000	20.07145	20.16658
1 12		20.03779	.0228746	875.99	0.000	19.99273	20.08285
1 13		19.90127	.0769981	258.46	0.000	19.74959	20.05294
1 14		20.15765	.0628333	320.81	0.000	20.03388	20.28143
1 15		20.14965	.0193631	1040.62	0.000	20.11151	20.1878
1 16		20.2552	.0773506	261.86	0.000	20.10283	20.40757
2 1		14.35217	.3042252	47.18	0.000	13.75288	14.95146
2 2		14.20309	.3031334	46.85	0.000	13.60595	14.80023
2 3		15.59499	.334313	46.65	0.000	14.93642	16.25355
2 4		15.71096	.2456767	63.95	0.000	15.22701	16.19492
2 5		15.79626	.2126322	74.29	0.000	15.3774	16.21513
2 6		14.76775	.3022864	48.85	0.000	14.17228	15.36323
2 7		14.62936	.1812185	80.73	0.000	14.27238	14.98635
2 8		16.134	.1615891	99.85	0.000	15.81568	16.45231
2 9		13.82028	.2312423	59.77	0.000	13.36476	14.2758
2 10		13.90943	.2326091	59.80	0.000	13.45122	14.36765
2 11		13.5899	.2030668	66.92	0.000	13.18988	13.98992
2 12		15.44849	.192391	80.30	0.000	15.0695	15.82748
2 13		16.78668	.647607	25.92	0.000	15.51096	18.0624
2 14		14.22045	.5284715	26.91	0.000	13.17941	15.26148
2 15		14.32354	.1628574	87.95	0.000	14.00273	14.64436
2 16		14.39837	.6505717	22.13	0.000	13.11681	15.67993
3 1		15.70405	.1169885	134.24	0.000	15.47359	15.9345
3 2		15.45075	.1165687	132.55	0.000	15.22112	15.68038
3 3		15.77218	.1285586	122.68	0.000	15.51893	16.02543
3 4		15.47895	.0944739	163.84	0.000	15.29284	15.66505
3 5		15.47533	.0817668	189.26	0.000	15.31426	15.6364
3 6		16.13098	.116243	138.77	0.000	15.90199	16.35996
3 7		15.65928	.0696868	224.71	0.000	15.52201	15.79656
3 8		15.28481	.0621384	245.98	0.000	15.16241	15.40722
3 9		15.76671	.0889233	177.31	0.000	15.59154	15.94188
3 10		15.70685	.0894488	175.60	0.000	15.53064	15.88305
3 11		15.29416	.0780885	195.86	0.000	15.14033	15.44798
3 12		15.32501	.0739831	207.14	0.000	15.17927	15.47075
3 13		16.02345	.2490345	64.34	0.000	15.53288	16.51402
3 14		15.9428	.2032215	78.45	0.000	15.54247	16.34312
3 15		15.97453	.0626261	255.08	0.000	15.85116	16.0979
3 16		16.18485	.2501746	64.69	0.000	15.69203	16.67767

. contrast a.sample

Contrasts of marginal linear predictions

Margins : asbalanced

 | df F P>F



```

-----
logeew
  sample |
(1 vs 2) |          1          9.99      0.0018
(2 vs 3) |          1          6.80      0.0097
(3 vs 4) |          1         13.11      0.0004
(4 vs 5) |          1          0.14      0.7049
(5 vs 6) |          1          0.59      0.4437
(6 vs 7) |          1          4.03      0.0458
(7 vs 8) |          1          2.97      0.0859
(8 vs 9) |          1         26.14      0.0000
(9 vs 10)|          1          1.41      0.2365
(10 vs 11)|         1          3.30      0.0705
(11 vs 12)|         1          6.04      0.0147
(12 vs 13)|         1          2.93      0.0881
(13 vs 14)|         1          6.52      0.0113
(14 vs 15)|         1          0.01      0.9031
(15 vs 16)|         1          1.75      0.1876
  Joint  |         15          6.10      0.0000
-----
Denominator |         240
-----

```

```

-----
          | Contrast   Std. Err.   [95% Conf. Interval]
-----
logeew
  sample |
(1 vs 2) | .1582988   .0500844   .0596376   .25696
(2 vs 3) | -.1373463  .0526596  -.2410804  -.0336122
(3 vs 4) | .1723325   .0475907   .0785836   .2660814
(4 vs 5) | .014901    .0393062  -.0625281   .0923302
(5 vs 6) | .0309199   .0403022  -.0484713   .1103111
(6 vs 7) | -.0837425  .0417131  -.1659131  -.001572
(7 vs 8) | .0498303   .0288924  -.0070847   .1067454
(8 vs 9) | -.1693362  .0331183  -.2345758  -.1040966
(9 vs 10)| .0452039   .038093   -.0298355   .1202433
(10 vs 11)| .0661929   .0364283  -.0055671   .1379529
(11 vs 12)| .0812277   .0330437   .0161349   .1463205
(12 vs 13)| .1365208   .0797154  -.0205104  .293552
(13 vs 14)| -.2563887  .1003896  -.454146   -.0586314
(14 vs 15)| .0080014   .0656431  -.1213088   .1373116
(15 vs 16)| -.1055476  .0798687  -.2628809   .0517856
-----

```

Para obtener los contrastes entre valores sucesivos de los pesos de gónadas e hígado se repiten los comandos pero con la variable deseada en la primera posición del comando. Por ejemplo, ahora para los pesos gonadales:



```
. manova logegw logeew logelw = c.logebl sample
```

Salida omitida

```
. contrast a.sample
```

Contrasts of marginal linear predictions

Margins : asbalanced

	df	F	P>F
logegw			
sample			
(1 vs 2)	1	0.13	0.7237
(2 vs 3)	1	9.88	0.0019
(3 vs 4)	1	0.08	0.7723
(4 vs 5)	1	0.07	0.7966
(5 vs 6)	1	9.21	0.0027
(6 vs 7)	1	0.16	0.6936
(7 vs 8)	1	38.34	0.0000
(8 vs 9)	1	69.00	0.0000
(9 vs 10)	1	0.08	0.7810
(10 vs 11)	1	1.09	0.2980
(11 vs 12)	1	44.72	0.0000
(12 vs 13)	1	3.98	0.0471
(13 vs 14)	1	9.24	0.0026
(14 vs 15)	1	0.03	0.8520
(15 vs 16)	1	0.01	0.9114
Joint	15	14.76	0.0000
Denominator	240		

	Contrast	Std. Err.	[95% Conf. Interval]	
logegw				
sample				
(1 vs 2)	.1490776	.4212446	-.6807312	.9788864
(2 vs 3)	-1.391892	.4429039	-2.264368	-.5194171
(3 vs 4)	-.1159783	.4002708	-.9044709	.6725142
(4 vs 5)	-.0852983	.3305922	-.7365311	.5659344
(5 vs 6)	1.028509	.3389693	.3607741	1.696244
(6 vs 7)	.1383886	.350836	-.5527223	.8294996
(7 vs 8)	-1.504634	.2430048	-1.983329	-1.02594
(8 vs 9)	2.31372	.2785475	1.76501	2.86243

Métodos cuantitativos computarizados para biología pesquera



(9 vs 10)		-.0891533	.3203888	-.7202864	.5419797
(10 vs 11)		.3195324	.306387	-.2840187	.9230834
(11 vs 12)		-1.858588	.2779206	-2.406063	-1.311113
(12 vs 13)		-1.338192	.6704617	-2.658933	-.0174509
(13 vs 14)		2.56623	.8443461	.9029547	4.229505
(14 vs 15)		-.1030934	.5521037	-1.190681	.9844944
(15 vs 16)		-.0748303	.6717511	-1.398111	1.248451

Debido a que el análisis de varianza y la regresión son procedimientos estadísticos relacionados, es posible llevar a cabo el anterior análisis con los comandos “mvreg”, “margins” y “contrast” (Salidas omitidas):

```
. mvreg logeew logegw logelw = logebl i.sample  
  
. margins sample  
  
. contrast a.sample
```

En versiones anteriores de Stata, el contraste entre medias ajustadas puede llevarse a cabo con el comando “regress”

```
. regress logeew logebl i.sample  
  
. margins sample  
  
. contrast a.sample
```

Para las gráficas, una forma de hacerlas es pasar las medias ajustadas a una hoja de Excel, agregar la variable de los días transcurridos entre muestreos y pasarlas de nuevo a Stata con la memoria sin datos. Una vez realizado lo anterior se guardan y se hacen las gráficas.

```
. clear  
  
. *(8 variables, 48 observations pasted into data editor)  
  
. *(1 variable, 48 observations pasted into data editor)  
  
. sum
```

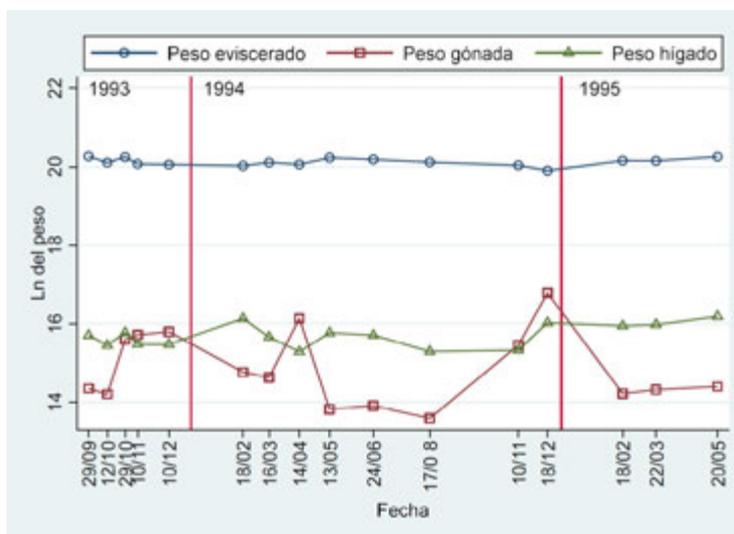
Variable		Obs	Mean	Std. Dev.	Min	Max
var		48	2	.825137	1	3
muestra		48	8.5	4.658554	1	16



ma	48	16.89263	2.400552	13.5899	20.26627
ee	48	.153285	.1504725	.0192123	.6505717
t	48	296.0037	307.3268	22.13	1044.18

p	48	0	0	0	0
li	48	16.59068	2.582724	13.11681	20.19501
ls	48	17.19459	2.242891	13.98992	20.40757
edias	48	572.625	195.8721	320	923

```
. scatter ma edias if var==1, c(l) ms(Oh) || scatter ma edias if var==2,
ms(Sh) c(l) || scatter ma edias if var==3, ms(Th) c(l) xlabel(320 "29/09"
338 "12/10" 355 "29/10" 367 "10/11" 397 "10/12" 468 "18/02" 493 "16/03"
522 "14/04" 551 "13/05" 593 "24/06" 647 "17/0 8" 732 "10/11" 760 "18/12"
832 "18/02" 864 "22/03" 923 "20/05", angle(vertical) ) legend(position(1)
rows(1) lab(1 "Peso eviscerado") lab(2 "Peso gónada") lab(3 "Peso hígado"))
ytitle("Ln del peso") xline(418 773) text(22 340 "1993") text(22 450 "1994")
text(22 810 "1995") xtitle("Fecha") yscale(range(14 22)) ylab(14(2)22)
```



```
. save anucovszk
file anucovszk.dta saved
. saveold anucovszk, version(11) replace
(saving in Stata 12 format, which Stata 11 can read)
file anucovszk.dta saved
```

Apéndice 6

Hojas de Datos en Excel



Para uso de algunos datos se introdujeron-tomaron directamente de archivos Excel.

Capítulo 2:

- 1) forma.xlsx

Capítulo 6:

- 1) anmucovszki.xlsx

Capítulo 7:

- 1) grafiarea.xlsx

Apéndice 8:

- 1) multinormal.xlsx

Apéndice 7

Programa PAC (Programas para el Análisis del Crecimiento)



Programación por José Ricardo Rodríguez Rojas

El programa PAC (Programas para el análisis del Crecimiento) consta de una serie de rutinas para la estimación y comparación de funciones de crecimiento a través de diferentes métodos. Para su ejecución se requiere de la instalación del programa al disco duro, y se realiza haciendo doble clic con el ratón en el icono del archivo ejecutable "Instalar.exe". Una vez instalado se ejecuta haciendo clic con el ratón en el menú inicio→programas→PAC→PAC v1.0, y se abre una ventana de presentación que se cierra al oprimir el botón Aceptar. Enseguida aparece la pantalla principal del programa con los siguientes menús:

Archivo

Nuevo

- Estimación de Parámetros
- T^2 de Hotelling
- Razón de Verosimilitud

Abrir

- Estimación de Parámetros
- T^2 de Hotelling
- Razón de Verosimilitud

Guardar Matriz de Varianza-Covarianza

Salir

Estimación de Parámetros

Regresión Lineal

- Regresión No Lineal

Comparación de Curvas

T^2 de Hotelling

Razón de Verosimilitud



En el submenú Nuevo del menú Archivo se pueden crear diferentes archivos correspondientes a los datos disponibles.

Estimación de Parámetros: se requieren datos de edad y talla, y opcionalmente número de individuos. En caso de no tener éste último dato se coloca el número 1 en la celda correspondiente.

T² de Hotelling: se requiere la matriz de varianza-covarianza, el valor estimado de los parámetros y el número de individuos.

Razón de Verosimilitud: se requieren datos de edad, talla y una variable categórica que indique a que grupo pertenecen dichos datos, ésta variable debe ser un número entero. De manera opcional se puede ingresar el valor de número de individuos, y en caso de no tener éste último dato se coloca el número 1 en la celda correspondiente. En los tres casos se guardan los archivos en formato .csv o .txt.

El submenú Abrir del menú Archivo se encuentra disponible solo si está activa la ventana del método correspondiente. Para crear los archivos .csv, en una hoja de cálculo se ingresan los valores colocando en columnas primero la edad, talla y al final el número de individuos, si es que se cuenta con este último dato, para el caso de estimación de parámetros.

Para el caso de razón de verosimilitud, se colocan en columnas primero la edad, talla, número de individuos (opcional), y al final la variable categórica que indique a que grupo pertenecen los datos. Al guardar el documento, se selecciona como tipo de documento el formato .csv.

Es importante verificar que no exista información aparte de la mencionada, como títulos, columnas de datos extras, etc. En caso de trabajar con archivos de texto se recomienda usar el programa Notepad. Para crearlos, en una línea por muestra se colocan los valores en el orden indicado anteriormente, separándolos por una coma o una tabulación y dejar una línea en blanco al final del archivo de texto.

El método de regresión lineal consta de cuatro procedimientos que permiten el cálculo de los parámetros de la función de crecimiento de von Bertalanffy. Cada procedimiento despliega información referente a los cálculos realizados, así como los valores calculados de los parámetros, y una ventana en la cual se puede visualizar el gráfico correspondiente.

Además, en la ventana del gráfico de Beverton-Holt se encuentra un botón que permite visualizar el gráfico de von Bertalanffy. En la ventana del gráfico hay un botón que permite



guardar la imagen del gráfico en formato .wmf, o copiarla al portapapeles. Para recuperar la imagen del portapapeles, por ejemplo, en un procesador de texto, se utiliza “pegado especial” del menú edición, y se selecciona imagen.

En la regresión no lineal, la ventana cuenta con un cuadro de texto con información referente a los cálculos realizados, y tres tablas con los valores de los parámetros calculados, la matriz de correlaciones y la matriz de varianza-covarianza, ésta última puede guardarse seleccionando “Guardar Matriz de Varianza-Covarianza” en el menú Archivo. También se cuenta con la opción de realizar la regresión no lineal con o sin ponderación, dependiendo de si se cuenta con el valor de número de individuos; además, se puede visualizar el gráfico con la curva de ajuste.

En el menú Comparación de Curvas existen dos métodos, la T^2 de Hotelling y la Razón de Verosimilitud. En la T^2 de Hotelling se requieren las matrices de varianza-covarianza, el valor de los parámetros y el número de individuos de los grupos a comparar, y se puede seleccionar el nivel de confianza para realizar la comparación. En el primer cuadro de texto se observan los datos ingresados, y en el segundo los resultados, como la matriz de varianza-covarianza conjunta, la diferencia de los vectores de los parámetros, los valores de la T^2 calculada y la de “tablas” y el valor de F, y en caso de existir diferencia, los intervalos de confianza y el valor de F crítica.

En el método de Razón de Verosimilitud se requieren los valores de edad, talla, número de individuos (opcional) y el indicador de pertenencia a un grupo. En la ventana se puede seleccionar la función de crecimiento, la ponderación con el número de individuos, los grupos a comparar y los valores iniciales para la regresión. Como resultado se obtiene una tabla de la prueba, con los valores de Razón de verosimilitud, los grados de libertad y el valor de P.

Si se pasa la información obtenida del programa a un procesador de textos, para conservar el formato de los resultados se debe usar una fuente de ancho fijo como “Courier”.

Apéndice 8

Ajuste de distribuciones Gaussianas por Máxima Verosimilitud de la distribución Multinomial



Recientemente, Haddon (2001; 2011) ha presentado un procedimiento para ajustar distribuciones gaussianas por medio de la maximización de la verosimilitud de la distribución multinomial utilizando la hoja computarizada de cálculo “Excel” y el procedimiento de optimización iterativa “Solver” disponible para este programa.

En este apéndice se aplicará el método a los datos de longitud patrón de hembras e indeterminados del bagre estuarino *Cathorops melanopus* que se han presentado en secciones anteriores (Capítulo 1).

El método parte de un cuadro de frecuencia con límites (inferiores y superiores) de clase y frecuencia absoluta. Se agregan frecuencias gaussianas probabilidades gaussianas (normales) tantas como componentes se sugieren en la distribución multimodal calculadas con parámetros gaussianos (media, desviación estándar y frecuencia) estimados previamente. Las frecuencias de cada componente se suman para dar una frecuencia total esperada cuya suma a su vez se utiliza para ajustar la frecuencia total calculada al número de observaciones (N , frecuencia absoluta total) y calcular la verosimilitud logarítmica de cada clase. La suma (con signo negativo) de la columna de verosimilitud (ajustada por N) es el objetivo de la minimización (maximización negativa) de la función “Solver”. El Cuadro A.8.1 resume estos pasos de acuerdo con las indicaciones de Haddon (2011) utilizando los datos de bagre analizados por medio del estimador de densidad por kernel gaussiano con la amplitud de banda (9) recomendada por la prueba bootstrap de multimodalidad (Silverman, 1981) y con parámetros de componentes Gaussianos estimados por Bhattacharya presentados en la Sección 1.13 del Capítulo 1 de esta obra.

La Figura A.8.1 muestra el ajuste inicial con los componentes estimados en el Capítulo 1 por Bhattacharya. El cuadro A.8.1 presenta un extracto de la hoja de cálculo para el procedimiento de acuerdo con las indicaciones (modificadas en la segunda edición) de Haddon (2011). La hoja completa se incluye en los archivos adjuntos (multinormal.xlsx). Como se ha mencionado, la distribución obtenida con la suma de los componentes Gaussianos es muy cercana a la distribución multimodal observada.



Los resultados del método de máxima verosimilitud de la distribución multinomial para el ajuste de componentes gaussianos se presenta en la Figura A.8.2. Las diferencias gráficas son casi imperceptibles, si bien una observación cuidadosa las revela. El Cuadro A.8.1 presenta la comparación de los valores iniciales introducidos al método y los resultantes de su aplicación.

Cuadro A.8.1. Comparación de parámetros Gaussianos obtenidos por Bhattacharya y Máxima Verosimilitud de Multinomial.

	Bhattacharya				MV Multinomial			
	CG1	CG2	CG3	CG4	CG1	CG2	CG3	CG4
Media	75.42	138.87	178.54	219.46	75.82	139.11	176.72	219.41
DE	12.98	11.26	12.93	12.82	13.03	11.01	12.89	12.96
Frec.	97	26	19	38	100	26	18	38

CG: Componente Gaussiano; DE: Desviación Estándar; Frec: Frecuencia.

El Cuadro A.8.1 incluye al final una columna con los residuales cuadrados y en la celda I3 se incluye la suma para utilizarla como criterio de minimización para el Solver. Los resultados al ser tan semejantes no se incluyeron, pero el lector puede verificarlos.

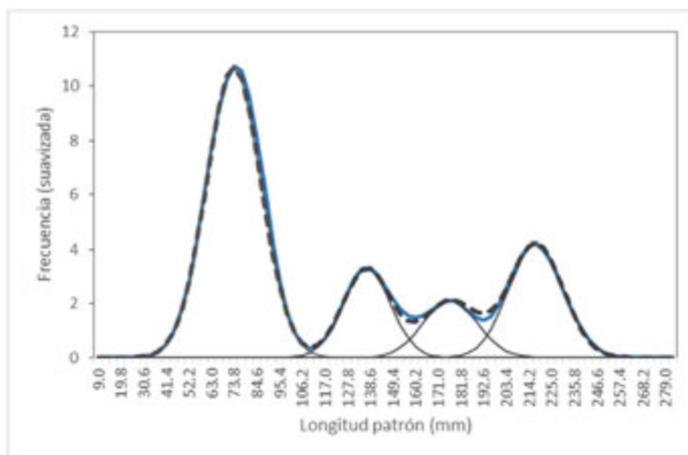


Figura A.8.1. Frecuencia suavizada observada y frecuencia estimada por suma de Componentes Gaussianos caracterizados por Método de Bhattacharya.



Cuadro A.3.2. Extracto de la hoja de cálculo para optimización de componentes gaussianos por máxima verosimilitud de distribución Multinomial (Haddon, 2011). Datos de longitud patrón de hembras e indeterminados de *Cathorops melanopus* N = 183.

A	B	C	D	E	F	G	H	I	J	K
1	Media	75.42	138.87	178.57	219.46					
2	DE	12.98	11.26	12.93	12.82			FMV	699.30618	
3	N	97	26	19	38	180		SRC	11.12025	
4										
5						180.00	183.00			
6	LIC	LSC	CG1	CG2	CG3	CG4	Esp	PL	LL	RC
7	9	12.6	4.8E-05	4.4E-28	9.5E-37	2.7E-57	4.8E-05	0.0000317	2.7E-07	-0.00048
8	12.6	16.2	0.00018	1.5E-26	3.3E-35	2.4E-55	0.00018	0.0001546	1.0E-06	-0.00213
9	16.2	19.8	0.00064	4.9E-25	1.1E-33	2.0E-53	0.00064	0.0006478	3.5E-06	-0.00818
10	19.8	23.4	0.00209	1.4E-23	3.2E-32	1.5E-51	0.00209	0.0023418	1.2E-05	-0.02661
11	23.4	27	0.00630	3.6E-22	8.9E-31	1.1E-49	0.00630	0.0073649	3.5E-05	-0.07556
12	27	30.6	0.01761	8.6E-21	2.3E-29	7.5E-48	0.01761	0.020402	9.8E-05	-0.18835
13	30.6	34.2	0.04562	1.8E-19	5.5E-28	4.6E-46	0.04562	0.0506642	0.00025	-0.419512
14	34.2	37.8	0.10946	3.4E-18	1.2E-26	2.6E-44	0.10946	0.1152454	0.00061	-0.85341
15	37.8	41.4	0.24330	5.9E-17	2.4E-25	1.3E-42	0.24330	0.24514	0.00136	-1.61951
16	41.4	45	0.50100	9.3E-16	4.63E-24	6.64E-41	0.50100	0.493708	0.002783	-2.90503
17	45	48.6	0.95574	1.3E-14	8.08E-23	2.97E-39	0.95574	0.941291	0.00531	-4.93069
18	48.6	52.2	1.68904	1.7E-13	1.31E-21	1.23E-37	1.68904	1.682736	0.009384	-7.85635
19	52.2	55.8	2.76531	1.9E-12	1.95E-20	4.7E-36	2.76532	2.788146	0.015363	-11.6427
20	55.8	59.4	4.19424	2.0E-11	2.71E-19	1.66E-34	4.19424	4.248782	0.023301	-15.9722
21	59.4	63	5.89341	1.9E-10	3.47E-18	5.43E-33	5.89341	5.944246	0.032741	-20.3241
22	63	66.6	7.67161	1.6E-09	4.12E-17	1.64E-31	7.67161	7.662989	0.04262	-24.18
23	66.6	70.2	9.25145	1.2E-08	4.52E-16	4.58E-30	9.25145	9.166581	0.051397	-27.208
24	70.2	73.8	10.3357	8.4E-08	4.6E-15	1.18E-28	10.3357	10.24855	0.057421	-29.2837
25	73.8	77.4	10.6973	5.2E-07	4.33E-14	2.82E-27	10.6974	10.75291	0.05943	-30.355

LIC: Límite inferior de clase; LSC: Límite superior de clase; CGn: Componente Gaussiano n; Esp: Frecuencia esperada; Frec: Frecuencia observada; PL: G7/G\$5; LL: Logaritmo de Verosimilitud; H7*LN(I7); RC: Residuales Cuadrados; FMV: Función de Máxima Verosimilitud; SRC: Suma de Residuales al Cuadrado.

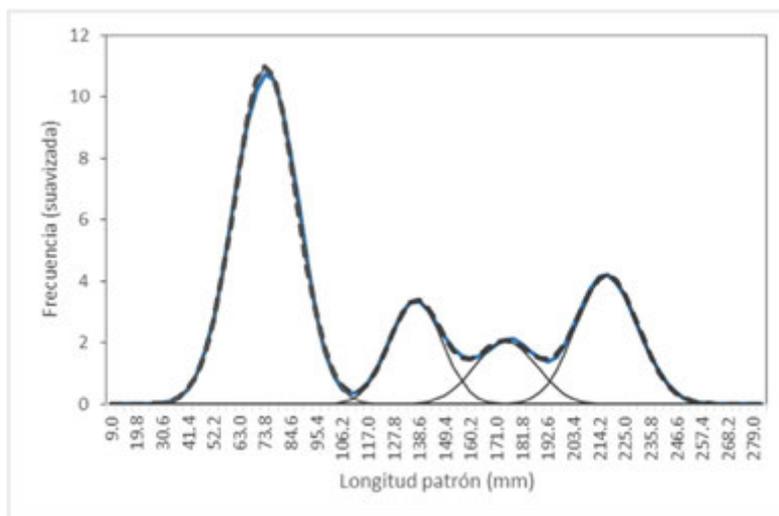


Figura A.8.2. Frecuencia suavizada observada, frecuencia estimada por Máxima Verosimilitud de Distribución Multinomial y componentes Gaussianos estimados

El uso de la hoja de cálculo y su resolución resultan ser un auxiliar didáctico muy interesante para la comprensión y aplicación del análisis de frecuencia de tallas en la Biología Pesquera.

El diseño de la misma y la disponibilidad de las herramientas de cálculo del programa hacen posible la experimentación y efecto de las magnitudes de los parámetros de los componentes Gaussianos.

Cabe destacar que la culminación satisfactoria del procedimiento (iterativo por aproximaciones sucesivas) depende de la introducción de valores iniciales adecuados. El histograma o el estimador de densidad utilizado como fuente de los valores para el ajuste, debe de mostrar la distribución de los datos resultante de valores de intervalo o banda apropiados.

Es claro del ejemplo mostrado, que con el uso de los EDKs para mostrar la distribución de las tallas, el acercamiento a valores apropiados se consigue de manera más directa y sobre todo con un respaldo cuantitativo estadístico significativo.



A

Alometría

coeficiente alométrico, 150, 157, 159, 163, 164, 168, 171, 173, 174, 180, 182, 184, 186, 187, 188

crecimiento hiperalométrico, 161, 163, 164, 187

crecimiento hipalométrico, 171, 182, 186

crecimiento isométrico, 166, 168, 314

amplitud

crítica, 78, 85, 88, 90, 92, 92, 94, 95, 424, 435, 437

de banda (intervalo), 21, 22, 30, 33, 34, 38, 39, 40, 41, 43, 44, 46, 47, 49, 50, 51, 52, 53, 54, 56, 58, 59, 60, 61, 63, 64, 67, 68, 69, 70, 72, 73, 74, 76, 78, 79, 81, 82, 84, 87, 89, 92, 123, 126, 219, 225, 231, 235, 422, 435, 436, 446, 454, 472

de intervalo, 21, 22, 30, 31, 33, 41, 46, 47, 49, 50, 51, 52, 53, 58, 59, 64, 69, 81, 126, 454

de ventana (banda), 38, 78, 85, 89, 426

fija, 38

variable, 38, 39, 422, 431

óptima, 49, 54, 59, 60, 61, 64, 65, 69, 105, 107

óptima Gaussiana de Silverman, 60, 64, 105

óptima mejorada de Härdle, 60, 64

óptima asintótica, 49

óptima Gaussiana de Scott, 60, 64

óptima Gaussiana en PF, 60, 64

óptima robusta de Freedman-Diaconis, 60, 64, 96, 106

regla de Friedman-Diaconis, 50, 51

regla de referencia gaussiana, 50

regla de Scott, 50, 51

regla de Sturges, 50

regla mejorada del pulgar, 55

regla óptima adaptativa, 55

reglas prácticas, 14, 48, 60, 63, 64, 225

reglas prácticas para selección, 49

sobresuavizada, 51

sobresuavizada de Terrel y Scott, 51, 55

sobresuavizada en PF, 51, 55

sobresuavizada homoscedástica, 51

sobresuavizada para kernel Gaussiano de Scott, 51, 55

sobresuavizada robusta, 51, 55

Análisis

análisis de covarianza (ANCOVA), 285, 315, 359

Análisis de frecuencia de tallas, 10, 21, 81, 120, 224, 231, 251, 475

cohorte, 14, 45, 100, 101, 189, 225, 239, 330

estimador de densidad por kernel, 23, 36, 37, 40, 69, 78, 105, 107, 121, 147

Cathorops melanopus, 22, 24, 26, 27, 28, 81, 83, 84, 92, 102, 120, 122, 143, 147



reclutamiento, 21, 46, 223

Análisis Multivariado de Covarianza

(ANMUCOVA), 315, 316, 319, 320, 321, 329, 333, 334, 342, 347, 353

covariada, 315, 319, 320, 321, 323, 324, 325, 326, 327, 328, 329, 333, 334, 343, 346, 347, 353

Archivos de datos

Excel

forma para Alometría, 451

para gráfica de área (estadios reproductivos), 451

R

biológicos del bagre estuarino *Cathorops melanopus*, 172, 185, 372, 454

Stata

biológicos del bagre estuarino *Cathorops melanopus*, 365, 369, 451, 452

crecimiento de la merluza del Pacífico *Merluccius productus* (Kimura), 452

crecimiento estacional de almeja blanca (King), 452

crecimiento estacional de Picther y MacDonald en Haddon, 452

crecimiento por retrocálculo del caracol marino *Hexaplex princeps*, 452

Dentex tumifrons, 451

distribución de tallas de bagre estuarino *Cathorops melanopus*, 143, 147, 451

erizos de mar, 451

forma para Alometría, 451

madreperla *Pinctada mazatlanica*, 451

opérculos del caracol marino *Astraea undosa*, 207, 452

otolitos de la Tilapia *Oreochromis niloticus*, 218, 221, 452

pepino de mar *Holothuria inornata*, 452

tallas del bagre estuarino *Cathorops melanopus* en formato texto, 72, 451

bagre estuarino *Cathorops melanopus*, 452

caracol marino *Hexaplex princeps*, (ANMUCOVA), 180, 451, 452

gasterópodo marino *Leucozonia cerata*, 167, 451

opérculos del caracol marino *Astraea undosa*, 203

otolitos de la Tilapia *Oreochromis niloticus*, 214, 216

trucha coralina, 451

Hexaplex princeps

primera madurez King, 453

reproductivos del caracol marino *Hexaplex princeps*, 453

suavizados del caracol marino *Hexaplex princeps* (ANMUCOVA), 453

talla-edad del bagre estuarino *Cathorops melanopus*, 452

tróficos de peces de la captura artesanal de Puerto Ángel, Oax. México, 451

B

Bahía de Sendai, 317

Bahía de Tokio, 191, 202, 316, 318, 333



Bahía de Wakasa, 317, 332

Beverton-Holt (Método de), 230, 241, 242, 248, 249, 470

Bhattacharya, 101, 102, 103, 108, 109, 110, 112, 114, 116, 121, 122, 126, 130, 140, 225, 234, 237, 238

biomasa, 172, 223

C

Capítulo 1

Análisis de frecuencia de tallas, 21, 81, 120, 224, 231, 251, 475

Capítulo 2

Ejemplos

Echinodermata, Echinoidea, erizos de mar, 159, 179, 252, 266, 455

Gasteropoda, *Leucozonia cerata*, 164, 165, 166, 167, 168,

Pelecípoda, *Pinctada mazatlanica*, 169, 170, 171

relaciones entre longitudes y pesos, 149

Capítulo 3

Ejemplos

Tilapia, *Oreochromis niloticus*, 214

Capítulo 4

crecimiento de bóvidos, 251

crecimiento de elefantes, 251

crecimiento de mamíferos en general, 251

crecimiento estacional, 267

minnow *Phoxinus phoxinus*, 268, 270, 273

stock pesquero, 101, 224, 225, 251, 402, 403, 405, 406, 407, 411, 412, 416

Capítulo 6

Índices morfofisiológicos, 313, 334, 337, 350, 353

Capítulo 7

Ciclo reproductivo, 313, 373

Correlación cruzada, 383, 386, 388, 390

curva logística de dos parámetros, 396

Escalas histológicas de madurez, 372

Escalas macroscópicas de madurez, 363

Gráficos de área, 377

Talla de primera madurez, 393

Cathorops melanopus, 22, 24, 26, 27, 81, 83, 84, 92, 102, 120, 122, 135, 143, 147

Comparación de dos funciones de crecimiento

prueba de máxima verosimilitud, 285, 289

prueba de razón de verosimilitud, 285, 289, 290

componentes gaussianos, 100, 101, 102, 107, 109, 110, 112, 113, 114, 118, 119, 120, 123, 124, 125, 126, 127

condritas, 77, 80, 92, 94

carbonosas, 77

crecimiento, 21, 45, 46, 81, 100, 101, 149, 150, 151, 153, 155, 161, 163, 164, 166, 168, 171, 172, 177, 180, 182, 184, 186, 187, 189, 194, 196, 197, 198, 202, 203, 204, 206, 214, 220, 223, 224, 225, 226, 227, 228, 229, 230, 237, 239, 241, 243, 244, 245, 246, 249, 250, 251, 252, 253, 254, 257, 260, 262, 264, 266, 267, 268, 273, 274, 275, 277, 278, 281, 285, 86, 288, 289, 290, 291, 292, 293, 294, 295, 300, 301, 302, 303, 305, 306, 307, 308, 309, 310, 314, 332, 333, 354, 469, 470, 471



curvas, 289, 290, 291, 292, 300, 302, 303

modelos, 223, 224, 226, 243

Criterio de Información Bayesiano, 136, 259, 262, 266, 270, 283

Criterio de información de Akaike, 246, 259, 260, 261, 262, 265, 266, 269, 270, 272, 276, 283, 297, 299

curtosis, 50, 53, 54

D

Datos de frecuencia de tallas, 21, 224

Datos multimodales

grosor de timbres postales antiguos de México, 92, 94

velocidad de galaxias, 92, 93

densidad

función de, 29, 49, 78

crítica, 79

Dentex tumifrons, 110, 111, 112, 113

desviación estándar, 41, 47, 49, 50, 53, 58, 87, 88, 102, 105, 110, 113, 123, 125, 134, 140, 194, 195, 207, 210, 213, 225, 233, 234, 255, 472, 473

diagrama de caja, 369, 370

diferencias logarítmicas, 101, 103, 108, 109, 112, 114, 115, 121, 125, 234

discontinuidad, 28, 29, 33, 34

dispersión de los cuartos, 55

distribución, 21, 22, 23, 27, 38, 40, 41, 45, 46, 47, 49, 50, 54, 55, 60, 61, 70, 72, 74, 76, 77, 78, 79, 80, 81, 82, 87, 91, 100, 101, 102, 108, 109, 118, 120, 124, 126, 127, 128, 129, 130, 133, 140, 141, 142,

143, 144, 145, 211, 218, 220, 224, 225, 227, 234, 235, 262, 263, 264, 278, 286, 288, 289, 290, 292, 302, 314, 315, 316, 318, 370, 374

acumulada, 78

asimétrica, 49

bimodal, 23, 50, 54

binomial, 48

curtosis, 50, 53, 54

de Cauchy, 58

de los datos, 70, 81, 82

gaussiana, 41, 47, 49, 50, 60, 88, 100, 225

modas, 22, 27, 45, 46, 50, 61, 66, 72, 76, 77, 78, 79, 80, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 100, 101, 112, 211, 218, 219, 220, 225, 227, 231, 232, 233, 263, 435, 436, 437, 439

multimodal, 27, 41, 45, 49, 74, 126, 127, 141, 224, 225

multinomial, 140, 472, 473, 474, 475

multivariada, 289

no gaussiana, 49, 50, 51, 54, 314

normal (gaussiana), 41

sesgada, 41, 50, 54,

t de Student, 50

tetramodal, 82, 127, 129, 130

trimodal, 23, 27

unimodal, 211

distribuciones

comparación

Método de Kolmogorov-Smirnov, 141

de frecuencia de talla, 143



mezcladas (mezcla de componentes individuales), 47, 100, 101, 102, 126, 128, 140, 454

multimodales (mezcla de componentes individuales), 100, 126, 127, 224

distribuciones Gaussianas, 47, 50, 100, 129, 140, 225, 472

E

edad, 45, 46, 77, 101, 189, 190, 194, 195, 197, 298, 203, 204, 206, 207, 211, 213, 220, 223, 224, 225, 226, 228, 229, 237, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 252, 253, 257, 258, 259, 260, 261, 262, 265, 266, 267, 268, 270, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 287, 293, 295, 296, 301, 302, 303, 305, 307, 310, 311, 319, 321, 325, 326, 327, 328, 329, 330, 331, 332, 333, 363, 392

error estadístico

cuadrado integrado, 56

cuadrado integrado medio, 41, 49, 51, 58

cuadrado integrado medio asintótico, 51, 58

escala probabilística, 101

Estadística exploratoria, 80

estimación de constantes de función de crecimiento de von Bertalanffy

gráfico de Ford-Walford, 227, 228, 237

Método (gráfico) de Gulland, 228, 238, 239, 240

Método (gráfico) de Gulland-Holt, 229, 239, 240, 246, 247, 248

de los parámetros de la función de crecimiento de von Bertalanffy, 226

de crecimiento, 46, 100, 101, 149, 153, 155, 172, 189, 196, 197, 198, 204, 206, 214, 220, 223,

224, 225, 226, 227, 228, 229, 230, 237, 239, 241, 243, 244, 245, 246, 249, 250, 251, 252, 253, 254, 257, 259, 260, 262, 264, 266, 267, 268, 273, 274, 275, 277, 278, 281, 285, 286, 288, 289, 290, 291, 292, 293, 294, 300, 301, 302, 303, 305, 306, 307, 308, 309, 310, 354, 469, 470, 471

de edad, 46, 77, 101, 189, 197, 220, 223, 239, 245, 246, 251, 252, 273, 287, 318, 319, 330, 332, 333

de edad (interpretación de estructuras duras), 224, 225, 245

Estimación de densidad para cada dato, 426, 433

Estimación de densidad por trama de puntos, 426, 430, 433

estimación de edad 197

análisis de frecuencia de tallas, 21, 81, 120

interpretación de partes duras, 225

opérculo, 190, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 253, 254, 255, 256, 311

por escamas, 190, 191, 192, 196, 197, 198, 224, 225, 245

por otolitos, 190, 197, 198, 199, 201, 202, 203, 214, 216, 217, 218, 224, 225

estimadores de densidad por kernel, 33, 34, 38, 46, 47, 48, 54, 55, 58, 69, 77, 81, 211, 218, 224, 225, 370, 421, 422, 430, 433, 454

F

F-pseudosigma, 41

factores de condición

de Fulton, 313, 314

hepatosomático, 314

relativa, 314



frecuencia, 21, 22, 23, 25, 27, 29, 30, 31, 33, 38, 39, 40, 45, 46, 47, 48, 52, 60, 61, 62, 63, 64, 65, 66, 77, 81, 101, 102, 103, 104, 105, 107, 108, 109, 110, 112, 113, 114, 115, 116, 118, 119, 120, 121, 122, 123, 124, 125, 126, 130, 133, 135, 136, 138, 139, 140, 141, 142, 143, 144, 145, 189, 194, 195, 211, 217, 224, 225, 231, 233, 234, 235, 236, 244, 251, 274, 383, 386, 388, 391, 393, 393, 395, 396, 397, 454, 472, 473, 474, 475

función

densidad de probabilidad, 78

kernel, 34, 43, 47, 79, 441,

ponderal

biponderada (cuártica), 68, 84, 120, 430, 434, 436

coseno, 34, 56, 72, 426, 430

Epanechnikov, 34, 47, 56, 68, 430, 434, 436

gaussiana, 43, 45, 68, 107, 430, 434, 436

triangular, 34, 44, 56, 68, 430, 434, 436

uniforme, 33, 44

Función de crecimiento de von Bertalanffy (FCvB), 224, 226, 237, 239, 241, 243, 244, 246, 249, 251, 260, 267, 278, 281, 285, 293, 470

Función de crecimiento de Gompertz, 224, 251, 264, 306, 309, 310

Función de crecimiento Logística, 303, 306, 305, 307

funciones penalizantes

criterio de información de Akaike, 446

error finito de predicción, 446

T de Rice, 446

G

gallineta del Pacífico, 293

géiser, 93

Grupos de edad, 77, 101, 252

Gráfico de Gulland, 228, 238, 239, 240

Gráfico de Gulland-Holt, 229, 239, 240, 246, 247, 248

H

histograma, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 36, 37, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 58, 59, 60, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 85, 88, 102, 127, 225, 233, 234, 393

centros de clase, 29, 116, 234

discontinuidad, 28, 29, 33, 34

intervalo, 21, 22, 27, 28, 29, 30, 31, 33, 34, 36, 37, 38, 40, 41, 42, 46, 47, 48, 49, 50, 51, 52, 53, 59, 60, 64, 66, 67, 69, 76, 81, 86, 87, 88, 91, 102, 104, 105, 110, 112, 113, 114, 123, 126, 150, 157, 159, 163, 164, 166, 168, 171, 173, 177, 187, 225, 227, 228, 235, 243, 246, 275, 288, 293, 436, 443, 454, 471, 475

número de intervalos, 27, 48, 49, 51, 52

regla de Sturges, 50

número sobreesuavizado de intervalos, 52, 53, 59,

origen, 22, 46

problemas del, 47

histogramas desplazados promedio (HDP), 48, 57, 69



I

índices morfométricos

factor de condición de **Fulton**, 313, 314

factor de condición relativa, 314

gonadosomático, 313, 314, 318, 332, 352, 372

índice de repleción, 291

índice gonadal relativo, 315

índice hepatosomático, 314

índices morfométricos, 314, 333

intervalo, 21, 22, 27, 28, 29, 30, 31, 33, 34, 36, 37, 38, 40, 41, 42, 46, 47, 48, 49, 50, 51, 52, 53, 59, 60, 64, 66, 67, 69, 76, 81, 86, 87, 88, 91, 102, 104, 105, 110, 112, 113, 114, 123, 126, 150, 157, 159, 163, 164, 166, 168, 171, 173, 177, 187, 225, 227, 228, 235, 243, 246, 275, 288, 293, 436, 443, 454, 471, 475

amplitud, 21, 22, 30, 31, 33, 41, 46, 47, 49, 50, 51, 52, 53, 58, 59, 64, 69, 81, 126, 454

Interpretación de marcas en partes duras, 224

K

kernel

biponderado, 56, 69, 70, 71, 73, 74, 75, 84, 120

cuártico (biponderado), 58, 440

Epanechnikov, 34, 47, 56, 68, 430, 434, 436

gaussiano, 72, 79, 120

rugosidad, 55, 56

triangular, 34, 44, 56, 68, 69, 72,

triponderado, 34, 56, 70, 444

kerneles

equivalentes, 67, 430

Kyushu, 316

L

Laguna de Tampamachoco, Ver. México, 81, 231, 234, 365

Lateolabrax japonicus, suzuki, 191, 193, 194, 195, 197, 200, 201, 202, 245, 246, 247, 248, 249, 250, 316, 318, 333

lenguajes de programación

Visual Basic, 421

localización, 28, 53, 76, 77, 141

M

matriz conjunta de varianza-covarianza, 287

máxima verosimilitud, 47, 56, 101, 127, 140, 277, 278, 281, 283, 289, 291, 292, 307, 308, 310, 472, 473, 474

Merluza norteña, 294, 295

mezclas filatélicas, 77

mínimos cuadrados, 56, 57, 67, 70, 73, 74, 84, 110, 120, 126, 159, 225, 230, 243, 290, 292, 315, 422, 443, 444, 454

modelos lineales generalizados, 315, 320

Mollusca

Gasteropoda

caracol marino *Astraea undosa*, 203, 204, 210, 212, 213

caracol marino *Hexaplex princeps*, 180, 181, 182, 252, 253, 254, 255, 256, 259, 303, 304, 306, 307, 308, 309, 310, 311, 333, 334, 338, 339, 340, 341, 342, 356, 373, 374, 375, 377, 378, 379, 380, 381, 385, 392

Pelecípoda

almeja, 273, 274

mortalidad infantil súbita



cot death, 77

muestreo repetitivo con remplazo

bootstrap, 48, 76, 77, 79, 80, 84, 87, 88, 90, 93, 93, 100, 120, 225, 232, 233, 422, 423, 435, 437, 438, 472

multimodalidad, 61, 66, 76, 77, 78, 79, 81, 82, 84, 85, 87, 91, 92, 93, 120, 141, 225, 231

N

nlog2.ado, 352, 390

nlpotfun.ado, 158, 449

nlpfun2.ado, 158

nls, 167, 176, 181, 185, 261, 262

nlsvbgff.ado, 449

nlsvbffm.ado, 449

nlsvbffs.ado, 449

nlvbgf.ado, 243, 294

nlvbgfn.ado, 270, 449

nlvbgfek.ado, 295, 449

nlvbgfel.ado, 294, 449

nlvbgfet.ado, 295, 449

nlvbgfno.ado, 295, 449

O

Objetos para R

ajuste de componentes gaussianos en distribuciones mezcladas, 47, 100, 101, 102, 128, 454

FSA (Fish Stock Assessment), 277, 455

FSAdata datos para, 275, 277, 455

función para cálculo de función potencial por regresión no lineal, 158, 162, 166, 167, 169, 176, 181, 183, 185, 449, 455

función para regresión lineal, 153, 154, 173, 455

función para regresión no lineal, 167, 176, 181, 185, 261, 455, 456, 458

función para suavización no lineal resistente, suavizador compuesto 4253eh doble, 256

histogramas desplazados promedio (ASH de Scott), 42, 48, 454

importación de datos en otros formatos (Stata), 22, 36, 39, 59, 67

selección de amplitud de intervalo/banda, 48, 60, 81

P

Paquete estadístico R (R Core Team), 24, 26, 36, 37, 60, 66, 93, 94, 95, 96, 127, 128, 129, 134, 137, 138, 139, 145, 153, 155, 166, 168, 172, 175, 180, 261, 262, 266, 275, 277, 369, 370, 371, 404, 411, 412, 454, 459

parábola, 101

Plectropomus leopardus, 59

polígonos de frecuencia, 21, 22, 29, 30, 48, 52, 60, 62, 63, 64, 66, 102, 211, 454

probabilidad, 78, 126, 277, 278, 286, 288, 289, 302, 472

Programas para estimación de constantes de funciones de crecimiento

FCvB por Máxima Verosimilitud (mlvbgf.ado), 277, 449

FCvB por Máxima Verosimilitud (versión 2) (mlvbgf2.ado), 449



- Regresión no lineal, 287
- FCvB estacional un ciclo (semanal) por función evaluadora (nlsvbgff.ado), 449
- FCvB estacional un ciclo (mensual) por función evaluadora (nlsvbffm.ado), 275, 449
- Función Logística de dos parámetros, 396, 398, 399
- para estimación de índice de alometría por regresión no lineal, 158, 449
- suavización no lineal resistente
- procedimiento doble (twice) para suavizador compuesto 4253eh, 450
- Programas actualizados
- para estimación de constantes de la FCvB gráfico de Ford-Walford, 227, 237
 - Método de Beverton y Holt, 226, 230, 240
 - Método de Gulland, 228, 238, 239, 240
 - Método de Gulland-Holt, 229, 239, 240, 246, 247, 248
 - Regresión no lineal, 287
- Programas actualizados para estimación de densidad
- warpdenm1.ado, 61, 85, 121, 263, 448
 - kerneld1.ado, 448
 - para estimación de densidad directa por kernel de banda variable
 - varwiker1.ado, 39, 448 - para estimación de densidad discreta por kernel de banda variable
 - varwike21.ado, 39, 40, 448
- para descomposición gaussiana
- bhatgauc1.ado, bhataplt1.ado, 103, 105, 110, 112, 113, 116, 121, 448
- prueba de multimodalidad de Silverman
- bootsamb.ado, critiband1.ado, 87, 89, 90, 448
 - silvtest1.ado, 88
- selección de amplitud de intervalo/banda
- bandw1.ado, 59, 66, 219, 448
- Programa para Análisis Exploratorio de Datos, 48, 67, 68, 69, 70, 71, 72, 73, 75, 85, 93, 94, 95, 100, 421, 422, 441, 446, 448
- Programa para localización de bandas críticas para estimadores de densidad por kernel, 448
- Programas para comparación de dos funciones de crecimiento
- Método de razón de verosimilitud (iguales K_s), 289, 290, 295
 - (iguales L_{∞} s), 289, 290, 294
 - (iguales t_s), 289, 290, 295
 - (sin restricciones), 294
- prueba T^2 de Hotelling, 255, 260, 414
- Programas para comparación de dos funciones de crecimiento de Gompertz de tres parámetros
- prueba de Máxima Verosimilitud, 307, 450
 - prueba de razón de Verosimilitud, 307, 450
- Programas para comparación de dos funciones de crecimiento de von Bertalanffy
- prueba de Máxima Verosimilitud, 301, 450
 - prueba de razón de Verosimilitud, 297, 298, 450



Programas para comparación de dos funciones de crecimiento Logísticas de tres parámetros

prueba de Máxima Verosimilitud, 307, 450

prueba de razón de Verosimilitud, 307, 450

Programas para pesquerías

ELEFAN I, 101

LFSA, 101

MIX, 101, 127

MULTIFAN, 101

programas para selección de amplitud de intervalo/banda

bcvwarp.ado, 67, 448

l2cvwarp.ado, 67, 448

promedio ponderado de puntos redondeados,

42, 48, 57, 439

prueba de Silverman para multimodalidad, 89,

90, 91, 96, 233, 424, 437

R

razón de verosimilitud, 285, 289, 290, 291, 293, 294, 296, 297, 300, 301, 303, 307, 308, 309, 469, 470, 471

recorrido intercuartílico, 41, 50, 54

regresión no lineal, 158, 159, 160, 161, 163, 164, 165, 166, 168, 170, 171, 175, 176, 177, 178, 180, 224, 230, 243, 244, 245, 246, 247, 250, 257, 266, 268, 271, 287, 293, 298, 455, 469, 471

regresión no paramétrica, 422, 423, 439

por kernel (Nadaraya-Watson), 421, 424, 439

reproducción

estrategias y tácticas, 45, 223, 225, 313, 332, 351, 352, 363, 364, 400

S

Sebastes alutus, 293

sesgo, 22, 49, 50, 53, 58, 81, 82, 175, 211, 256, 289, 302, 446

sobresuavización, 50, 53, 55, 56, 58

Stage, editor de gráficos, 382

Stata Statistical Package (Paquete estadístico Stata) Stata Corporation, 22, 24, 36, 39, 40, 50, 55, 59, 61, 66, 67, 81, 85, 100, 101, 102, 103, 108, 112, 120, 121, 125, 130, 134, 139, 140, 141, 144, 158, 160, 165, 170, 178, 179, 180, 183, 192, 209, 210, 211, 217, 232, 233, 243, 259, 262, 268, 271, 278, 292, 293, 354, 366, 369, 371, 378, 381, 384, 390, 393, 394, 395, 398, 425, 439, 448, 454, 455, 460, 467, 468

suavización

elección del parámetro de, 47, 48, 76

parámetro de, 34, 46, 47, 48, 50, 52, 56, 57, 69, 70, 76

principio de suavización máxima, 52

Suavización no lineal resistente, 421, 423, 424, 442

suavizador compuesto 4253EH, doble, 256, 257, 336, 348, 349, 383, 422, 442

suavizador compuesto 3RRSHe, 422

suzuki, 191, 193, 194, 195, 197, 200, 201, 202, 245, 246, 247, 248, 249, 250, 316, 318, 333

T

trazas de densidad, 421

transformación rápida de Fourier, 57

trucha coralina, 59, 60, 62, 63, 67, 68, 72

Plectropomus leopardus, 59



U

univariada, 286, 287, 289, 303, 446

V

Validación

análisis de bordes de estructuras duras, 216

Análisis del incremento marginal, 218, 219, 220, 221

validación de periodicidad de marcas de crecimiento en estructuras duras, 225

validación cruzada, 56, 57, 58, 59, 67, 68, 69, 70, 71, 73, 74, 75, 76, 84, 91, 225

por máxima verosimilitud, 56

por mínimos cuadrados, 56, 57, 67, 73, 74, 84, 225, 422, 424, 454

por muestreo repetitivo (bootstrap), 76

sesgada, 58, 70, 71, 74, 75, 84, 225, 422, 424, 454

varianza, 52, 55, 56, 79, 129, 130, 173, 175, 285, 286, 287, 288, 289, 298, 302, 446, 466, 469, 470, 471

von Bertalanffy (función de crecimiento), 224, 226, 237, 239, 241, 243, 244, 246, 249, 251, 260, 267, 278, 281, 285, 293, 470

W

WARP (Weighted Averaging of Rounded Points) Promedio ponderado de puntos redondeados (PPPR), 23, 42, 43, 44, 48, 61, 66, 67, 68, 69, 82, 88, 107, 146, 263

Métodos cuantitativos computarizados para biología pesquera

Isaías Hazarmabeth Salgado Ugarte
Verónica Mitsui Saito Quezada



Facultad de Estudios Superiores Zaragoza,
Campus I. Av. Guelatao No. 66 Col. Ejército de Oriente,
Campus II. Batalla 5 de Mayo s/n Esq. Fuerte de Loreto.
Col. Ejército de Oriente.
Iztapalapa, C.P. 09230 Ciudad de México.
Campus III. Ex fábrica de San Manuel s/n,
Col. San Manuel entre Corregidora y Camino a Zautla,
San Miguel Contla, Santa Cruz Tlaxcala.

<http://www.zaragoza.unam.mx>

