

UTILIZACION DEL PROCEDIMIENTO REGRESIONAL PONDERADO APLICADO EN UN MODELO LOGISTICO DE SELECTIVIDAD

Patricio Pavez Carrera
Escuela de Ciencias del Mar
Universidad Católica de Valparaíso
Casilla 1020, Valparaíso, Chile

RESUMEN: Se aplica un procedimiento regresional ponderado a una curva de selectividad, suponiendo una función de respuesta logística en la relación de retención por longitud.

El procedimiento puede ser desarrollado por cualquier sistema computacional y considera una secuencia simple, coherente y de fácil manejo, que resulta ser superior a los procedimientos de regresión de probitas, suponiendo una función de respuesta a la ojiva de selección del tipo sigmoide; se presenta una aproximación del intervalo de confianza para la longitud de retención al 50% resultante.

ABSTRACT: A weighted least squares procedure is applied to a selectivity curve assuming a logistic response function on the relationship of retention by length.

The procedure can be developed by any computational standard package; it considers a simple, coherent and easy to handle sequence which shows better results than the probit regression procedure, assuming a sigmoid response function for the selective ogive; an approximated confidence interval for the resulting retained length at 50% point is presented.

Recibido el 10 de septiembre de 1986

Aceptado el 10 de diciembre de 1986

INTRODUCCION

En las experiencias realizadas para evaluar la selectividad de artes de pesca, especial atención se presta al diseño mismo de la secuencia experimental y al análisis de la información resultante. Se entiende como proceso selectivo al efecto del cambio probabilístico de la retención por talla de los individuos, atribuido a las características inherentes al arte.

En el análisis de los resultados experimentales, se ha aplicado una variada gama de procedimientos estadísticos de diversa complejidad, los que son empleados en concordancia a la calidad del dato mismo (POPE, 1975).

La aplicación de ajustes regresionales resultan apropiados para determinar la talla de retención al 50% y han conducido a extensas consideraciones que emanan de la especificación misma de la curva de selección, la que puede ser de tipo sigmoide o de tipo logístico. En el primer caso se ha propuesto la regresión de las desviaciones normalizadas, las cuales se transforman en las denominadas probitas mediante la adición de una constante y el segundo caso, que considera la regresión de las logitas en función de la talla.

Se ha supuesto que la consideración logística es apropiada para la especificación de la curva de selección y en base a ello, el método de máxima probabilidad (aunque laborioso) sería el más eficiente. Es evidente que la calidad de los resultados no podrá superar la calidad inicial de la información ingresada, cualquiera sea el procedimiento analítico utilizado; la mayor eficiencia atribuida al procedimiento basado en consideraciones logísticas, se basa en evaluaciones a un mismo nivel de calidad de información (POPE, 1975).

Se presenta la aplicación de una alternativa analítica para ajustar por métodos regresionales la curva de selección, en el contexto de una consideración logística. El procedimiento requiere de una regresión del tipo ponderada, debido a la capacidad que tiene esta alternativa, de corregir el error producido por la variación del muestreo en la estimación de la proporción de retención, en las distintas tallas. Se aplica un procedimiento de estimación de la varianza de la talla de retención al 50%, que resulta ser especialmente valioso en análisis no seriados.

ANTECEDENTES

En los experimentos de selectividad, el ajuste de la curva de selección y la determinación de la talla o longitud al 50% de retención cobra especial interés.

POPE (1975) concede en su síntesis metodológica una gran importancia a la calidad de la información generada para el ajuste de una curva de selectividad; este autor reiteradamente se refiere a la magnitud de error experimental e indican que es éste el que fundamenta el empleo de técnicas que permitan evaluar con exactitud la talla al 50% de retención. Cuando las experiencias son proximales, los ajustes a "mano alzada" son aceptables para obtener una idea de la talla de retención requerida.

Los métodos de "mano alzada" consideran la posibilidad de ajustar la función curvilínea, ya sea uniendo puntos o suavizando la trayectoria más equilibrada; el ajuste a mano alzada también es posible cuando se emplea el papel de probabilidades (HARDING 1949; CASSIE 1954, 1955). Resulta altamente improbable, sin embargo, que se pretenda montar una experiencia de evaluación de la selectividad con intenciones de obtener resultados proximales únicamente. El costo y la complejidad de tales experimentos demanda gran cuidado en su ejecución y la alternativa más probable es que los resultados requerirán de un adecuado trato.

Los ajustes analíticos van desde una suavización de la función curvilínea mediante promedios móviles, hasta técnicas de ajuste por máxima probabilidad y regresionales por mínimos cuadrados ponderados. POPE (1975) indica que el problema no se refiere únicamente al ajuste de la función sino también a la elección de la función curvilínea que explica a la curva de selectividad.

El procedimiento de promedios móviles se aplica por igual a las tallas como a los porcentajes de retención y la talla al 50% de retención se obtiene por interpolación.

El procedimiento más común considera la regresión de los porcentajes de retención en la forma de sus desviaciones normalizadas y revalorizadas (probitas), en el supuesto de que la función curvilínea corresponde a una especificación de la forma sigmoide. Esta función se obtiene por la acumulación de las frecuencias de

una distribución normal y se encuentra tabulada en función de porcentajes (FISHER y YATES 1963); el procedimiento tiene la desventaja de no poder corregir las distorsiones de los extremos de la curva sigmoide.

Algunos autores trabajan dicha función, ajustando la regresión de probitas entre el primer y tercer cuartil donde se encuentra la menor distorsión; esta alternativa disminuye la información a procesar, lo que debe agregarse a la imposibilidad de corregir el efecto de variabilidad en la representación de la fracción vulnerable en las tallas.

Otra alternativa está representada por la consideración de una explicación logística, que constituye una función simétrica y que permite un ajuste regresional, a pesar de los efectos de distorsión de los extremos y que además permite la corrección del efecto de variabilidad de representación de la fracción vulnerable en las tallas.

POPE (1975) argumenta que el ajuste de la explicación logística es más eficiente, sobre todo si se emplea el método de máxima probabilidad; este autor indica que el método de estimación de la talla a un 50% de retención es igualmente eficiente, aunque no indica claramente si esta eficiencia se mantiene al contrastar este procedimiento con el método de regresión ponderada.

DERIVACION TEORICA

La situación derivada del ajuste de una función logística por procedimientos regresionales es relativamente simple cuando existe una observación asociada a cada nivel de la variable independiente X_1 . Esto es válido en el caso de la explicación logística a la curva de selectividad, ya que los datos experimentales son analizados por separado debido a la imposibilidad de homologar las condiciones del muestreo.

La función de respuesta logística a la curva de selección adopta de esta manera la siguiente forma:

$$E(Y) = \frac{\text{EXP}[b_0 + b_1 X]}{1 + \text{EXP}[b_0 + b_1 X]} \quad 1)$$

función que puede ser igualmente expresada como sigue:

$$\hat{p}_j = \frac{\text{EXP}[l_j]}{1 + \text{EXP}[l_j]} \quad 2)$$

en este caso, la variable regresada (Y_j) corresponde a:

$$Y_j = \ln \left[\frac{\hat{p}_j}{1 - \hat{p}_j} \right] \quad 3)$$

\hat{p}_j vale R_j / n_j

donde

EXP	=	base de logaritmos naturales (2.718282)
b_0	=	intercepto de la función de regresión
b_1	=	pendiente de la función de regresión
\hat{p}_j	=	proporción de retención en la j-esima talla
R_j	=	retención neta en la j-esima talla
n_j	=	total ingresado a la red en la j-esima talla

la varianza respectiva de la estimación de la proporción de retención en el j-esimo nivel de la talla (p_j) es igual a

$$\frac{\hat{p}_j [1 - \hat{p}_j]}{n_j}$$

La transformación logística, al ser linearizada, no corrige el efecto de varianzas desiguales en cada nivel de j-esima talla y por lo tanto, se presenta un efecto de heterocedasticidad que debe ser corregido. Debe recordarse que a cada nivel de talla los n_j son diferentes y por ello NETER *et al.* (1983) recomiendan el empleo de un ajusté regresional ponderado. Para valores moderadamente mayores de n_j , la proporción p_j adopta aproximadamente una distribución normal con una varianza estimada por:

$$s^2 (\hat{p}_j) = \frac{1}{n_j \hat{p}_j [1 - \hat{p}_j]} \quad 4)$$

NETER *et al.* (1983) de igual forma indica que la ecuación (4) es aproximadamente igual a σ^2/w_j . Es así como se determina que los pesos w_j son iguales a $n_j \hat{p}_j [1 - \hat{p}_j]$ y la varianza es igual a 1. Esto resulta de especial interés para el manejo de la matriz de varianza-covarianza, ya que ésta, se obtiene multiplicando la varianza por la matriz $[X'X]^{-1}$; si la varianza es igual a la unidad, la matriz de constantes actúa como matriz de varianza-covarianza.

Por lo tanto:

$$C = \begin{bmatrix} \text{Var}(b_0) & \text{Cov}(b_0, b_1) \\ \text{Cov}(b_0, b_1) & \text{Var}(b_1) \end{bmatrix} \quad 5)$$

El ajuste en sí no resulta ser dificultoso para rutinas estadísticas de uso corriente; interesante resulta sin embargo el manejo del estimador de varianza de la talla al 50% de retención.

En una curva ajustada de esta manera, la talla al 50% ($L_{50\%}$), se estima a una logita $p = 0.50$; el valor queda determinado por:

$$L_{50\%} = -b_0 / b_1 \quad 6)$$

lo que demanda una estimación de la varianza, aproximada por la expansión de Taylor de la siguiente forma:

$$\text{Var}[g(b_0, b_1)] \approx \left[-\frac{1}{b}\right]^2 \text{var}(b_0) + \left[\frac{b_0}{b_1}\right]^2 \text{var}(b_1) + 2\left[-\frac{1}{b_1}\right]\left[\frac{b_0}{b_1^2}\right] \text{Cov}(b_0, b_1) \quad 7)$$

de esta manera se puede evaluar una excelente aproximación a los límites de confianza del punto $L_{50\%}$ de la manera usual.

EJEMPLO DE APLICACION

Se presenta un ejemplo de aplicación para mostrar como opera el ajuste regresional ponderado. La información procede de las experiencias de selectividad realizadas a bordo el B/E TIBERIADES (UCV) durante el 18 - 26 abril de 1978 en las aguas de la zona de Valparaíso, con un arte de dos paneles operando en la captura de la merluza común (*Merluccius gayi* Guichenot 1848). Los datos fueron procesados mediante un IBM DOS Communication Program y se hizo uso del Statistics Interactive Programming System (SIPS), versión CYBER-NOS (DEPARTMENT OF STATISTICS 1978).

SIPS VERSION 52

VAR, 20

NAME (variables)

READ, MERLUZA 60 mm; MACHOS

SET, P=P/N

SET, W=N*P* (1.-P)

SET, Y=LN (P/ (1. -P))
 REGRESS, W, Y, X\$WT

** ENTERING REGRESS SUBSYSTEM **

Y = 1.4551 (equivalente al promedio ponderado de Y)

ADD, X

Y =

-4.238340 (CONSTANTE)

.200558 X

INVERSION (equivalente a la matriz de constantes C)

(CONSTANTE , CONSTANTE) = .246939

(X , CONSTANTE) = -.848622E-02

(X , X) = .298939E-03

AVTABLE

TABLA DE ANALISIS DE VARIANZA

FUENTE	GL	SUMA DE CUADRADOS	CUADRADOS MEDIOS
TOTAL	21	158.645	7.55455
REGRESION	1	134.554	134.55400
RESIDUAL	20	24.0911	1.20456

R CUADRADO = .8481

TVALUES

VARIABLE	ERROR ST.	COEFF. REGRES.	T	P
CONSTANTE		.54539	-7.771	.0000
X		.18976E-01	10.569	.0000

END

** LEAVING REGRESS **

EXIT

De la presentación sintetizada de una salida de SIPS se puede deducir que el modelo ponderado de regresión del ajuste logístico es:

$$Y = -4.23834 + .200558 X$$

el modelo resulta ser apropiado, ya que su F calculado es altamente significativo y este modelo explica el 84.81 % de la variabilidad observada

F =	111.703859 >	F* (0.05; 1, 21 gl)	= 4.32
F =	111.703859 >	F* (0.01; 1, 21 gl)	= 8.02

La estimación de la talla al 50 % de retención se obtiene haciendo $p = 0.5$, lo que determina que:

$$Y = -4.23834 + .200558 X = 0$$

de esta manera:

$$X = L_{50\%} = -b_0 / b_1$$

$$L_{50\%} = -(-4.23834) / .200558 = 21.132740 \text{ cm}$$

El intervalo de confianza se obtiene determinando la aproximación de varianza mediante la expansión de Taylor, lo que corresponde a:

$$\text{VAR} [-b_0 / b_1] = \{A + B - C\}$$

donde:

$$A = (-1/b_1)^2 * \text{Var} (b_0)$$

$$B = [(b_0 / (b_1)^2)]^2 * \text{Var} (b_1)$$

$$C = (2 * b_0 / (b_1)^3) * \text{Cov} (b_0, b_1)$$

de la matriz de constantes, que en este caso se comporta como la matriz de varianza-covarianza, se tiene que:

$$C_{00} = .246939$$

$$C_{11} = .000298939$$

$$C_{01} = -.008486220$$

así, el valor de la aproximación de la varianza es 18.374849 y un intervalo de confianza aproximado al 95 % resulta ser para el parámetro calculado ($L_{50\%}$):

$$L_{50\%} \pm 1.96 * \text{VAR} [-b_0 / b_1] / n_1$$

donde n_1 = número de datos pareados empleados
y,

$$21.132740 \pm 1.8334046$$

de esta manera el límite superior se localiza en 22.97 cm y el límite inferior en 19.30 cm. (La amplitud del rango está en función de la calidad de la información).

DISCUSION

Se ha presentado una secuencia analítica modelada, para el ajuste y manejo de información de retención descrita en la forma de una curva de selectividad, explicada por una función logística. Los procedimientos no revisten una complejidad

extrema toda vez que se refieren a desarrollos propios de una linearización ponderada.

La secuencia modelada obedece a una concepción simplificada que se circunscribe solamente a los procedimientos de interés; éstos conducen a la evaluación de cantidades primarias (tallas de retención al 50 % con sus respectivas bandas de confianza), las que a su vez conducen a través de un conjunto seriado de tamaños de mallas, al establecimiento de relaciones secundarias que describen la retención en función del tamaño de malla.

Las relaciones de la secuencia modelada, pueden ser obtenidas con la ayuda de cualquier sistema analítico computacional y tienen la ventaja de no trabajar con secciones de la relación curvilineal de selección y ser de fácil manejo.

Los resultados están sujetos a un esquema de causalidad de difícil interpretación e insertos en una red de contingencias que motiva una reflexión. Como una expresión de capacidad sistémica, un comportamiento (en el presente caso, selectivo) está asociado a un estado ambiental o de relaciones ambientales (considerando como tales las relaciones subsistémicas de la especie, el medio, el aparejo y la intencionalidad del hombre). La capacidad real del sistema se refiere al comportamiento de los comportamientos en todos los estados ambientales, por ello es difícil suponer una validez y correspondencia universal a los resultados que se logren. Ellos sólo pueden ser considerados como una expresión de la capacidad (explicativa) del sistema.

El uso de procedimientos regresionales ponderados permite corregir los efectos de variabilidad en la representatividad de la fracción vulnerable a cada nivel de talla; esto constituye un aspecto del comportamiento que está necesariamente inserto en una trama de estados ambientales que no necesariamente se repite en el futuro. Más aún, es del todo imposible pretender controlar la fracción vulnerable, homogenizarla o pretender establecer la validez de una función para una relación de estado ambiental estable.

No fue posible contrastar este procedimiento con otros como para fundamentar recomendaciones basadas en una mayor eficiencia cuantitativa; no obstante, se considera que el procedimiento es simple, armónico, constituye un todo coherente y fácil de operar. Particularmente valioso resulta la posibilidad de estimar una aproximación de la varianza de la talla de retención al 50 %, lo que individualiza a este procedimiento como superior a la regresión de probitas de retención u otros de menor elaboración o de mayor complejidad.

Reconocimiento: El procedimiento regresional teórico aplicado en este trabajo a la consideración logística de la curva de selectividad, formó parte del material de discusión y lecturas impartidas por el Dr. Dave Thomas, del Departamento de Estadística de la Oregon State University (Corvallis, USA.).

REFERENCIAS

- CASSIE, R.M. 1954. Some uses of probability paper in the analysis of size frequency distributions. *Aus. J. Mar. Freshw. Res.*, 5(3):513-522.
- CASSIE, R.M. 1955. The escapement of small fish from trawl nets and its application to the management of the New Zealand snapper fisheries. *Fisheries Bulletin of New Zealand*, 11:7-99.
- DEPARTMENT OF STATISTICS, 1978. Statistical Interactive Programming System (SIPS). Command Reference Manual. Oregon State University, Corvallis (ORE), 109 pp.
- FISHER, R.A. y F. YATES, 1963. Statistical tables for biological, agricultural and medical research. Oliver and Boyd, Edinburgh, 156 pp.
- JONES, R. 1984. Mesh size regulation and its role in fisheries management. In FAO. Papers presented at the Expert Consultation on the regulation of fishing effort (fishing mortality). Rome, 17 - 26 January 1983. A preparatory meeting for the FAO World Conference on fisheries management and development. FAO Fisheries Report, (289) Supp. 2: 214 pp.
- HARDING, J.P. 1949. The use of probability paper for the graphical analysis of polymodal frequency distributions. *J. Mar. Biol. Assoc. U.K.*, 28:141-153.
- NETER, J., W. WASSERMAN y M.H. KUTNER, 1983. Applied Linear Regression Models., Richard D. Irving, Inc., Illinois (USA), 547 pp.
- POPE, J.A. 1975. Manual of methods for fish stock assessment. Part III. Selectivity of Fishing Gear. FAO Fisheries Technical Paper 41 (Rev 1): 41 pp.