

# UNA ESTRATEGIA EN DOS DIRECCIONES PARA LA SELECCIÓN DE MODELOS DE REGRESIÓN PARA EL CASO DE UNA VARIABLE EXPLICATIVA CUALITATIVA.

E. Castells\* y O. Juárez\*\*

\*Facultad de Matemática y Computación, Universidad de La Habana, Cuba

\*\*Facultad de Matemáticas, Universidad Autónoma de Guerrero, México

## RESUMEN:

La modelación estadística de fenómenos reales es una problemática que continúa teniendo gran vigencia. Los modelos lineales constituyen una poderosa herramienta, no sólo por lo sencilla que resulta su aplicación sino porque aun en casos donde la verdadera función desconocida no es lineal resultan una buena aproximación. En el presente trabajo se propone una estrategia de selección de modelos aplicable cuando se tiene sólo un regresor y éste es cualitativo. La estrategia propuesta es una modificación de la propuesta por Castells (1999) (ver además Bunke y Castells, 1999). La bondad de la estrategia se estudia utilizando datos simulados.

## ABSTRACT:

Statisticians continue trying to model real phenomenon. Linear models are a powerful tool even when the true unknown model is non linear, because they bring us good approximations. We propose a model selection strategy sensible to apply in situation where we have a unique regressor which is a qualitative one. The strategy seems to be a modification of one proposed by Castells (1999) (see also Bunke y Castells, 1999). A validation in terms of simulated data is presented.

Palabras claves: selección de modelos, variables cualitativas, estimación mínimo cuadrática, cuadrado medio del error y cuadrado medio del error de predicción.

**KEY WORDS:** Models Choice, qualitative variables, least square estimation, mean squared error and mean squared error of prediction.

MSC: 62J05

## 1. INTRODUCCIÓN

La modelación estadística de fenómenos reales, es una problemática que continúa teniendo gran vigencia. Los modelos lineales constituyen una poderosa herramienta, no sólo por lo sencilla que resulta su aplicación sino porque aun en casos donde la verdadera función desconocida no es lineal, resultan una buena aproximación. En el presente trabajo se propone un método de selección de modelos aplicable en la situación que se describe a continuación:

Considérese una variable cualitativa  $X$ , con  $m$  categorías, las cuales van a ser denotadas por un conjunto de índices, por comodidad se denotarán por los primeros  $m$  naturales,  $\Omega = \{1, 2, \dots, m\}$  y en cada una de estas categorías, también llamadas puntos de diseño, se realizan  $n_i$  observaciones de una variable de respuesta  $Y$ , estas observaciones se denotarán por  $Y_{ij}$ , con  $j = 1, 2, \dots, n_i$ ;  $i = 1, 2, \dots, m$ ; donde  $n = \sum_{i=1}^m n_i$  y sus valores esperados se denotan por  $\mu_1, \mu_2, \dots, \mu_m$ . Como las variables explicativas son cualitativas, se puede considerar un modelo como el que se da en (1.1).

$$E\{Y_{ij}\} = \mu_i = \mu + \alpha_i \quad (1.1)$$

$$\text{Var}\{Y_{ij}\} = \sigma^2 \quad (1.2)$$

Se supone que las  $Y_{ij}$  no están correlacionadas, se añade la condición

$$\sum_{i=1}^m \alpha_i = 0 \quad (1.3)$$

Usualmente el modelo (1.1) es tratado como un modelo de análisis de varianza (Mood et al., 1974), y su objetivo es establecer si existe diferencia entre las medias  $\mu_i$ , equivalentemente entre los efectos  $\alpha_i$  o entre contrastes que dependen de ellos. En el análisis de varianza se parte de un modelo con todos los parámetros y luego se hacen dóxicas para determinar cuáles de ellos son iguales.

De la teoría de Modelos Lineales se sabe que, el mejor estimador lineal para los parámetros  $\alpha_i$ , está dado por:

$$\hat{\alpha}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} \quad \text{para } i = 1, 2, \dots, m \quad (1.4)$$

donde

$$\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{y} \quad \bar{Y}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} Y_{ij}$$

(ver Scheffé, 1982; Draper y Smith, 1998; Rao, 1973 y Searle, 1971 ).

La idea que se desarrolla en el presente trabajo, es proponer un método para seleccionar modelos de regresión, donde el proceso de comparación de parámetros es incorporado al proceso de selección del modelo, lo que se propone es una modificación de lo propuesto por Castells (1999). La estrategia se diseña evitando tener que hacer una búsqueda exhaustiva sobre el conjunto de todos los modelos posibles.

## 2. UNA ESTRATEGIA EN DOS DIRECCIONES PARA LA SELECCIÓN DE MODELOS DE REGRESIÓN PARA EL CASO DE UNA VARIABLE EXPLICATIVA CUALITATIVA.

La estrategia propuesta por Castells (1999), va en sentido ascendente en referencia al número de parámetros del modelo, sin embargo, un análisis de los métodos clásicos de selección de variables sugiere una modificación del mismo para convertirlo en un método que explore en las dos direcciones, primero ascendente y luego descendente. Las características primordiales de la propuesta son seleccionar el modelo interrelacionando el proceso de estimación y selección y explorar el espacio de modelos posibles visitando sólo algunos puntos de éste. Cada uno de los modelos visitados es evaluado con un criterio que apunta hacia el uso entendido del mismo.

Como se muestra más adelante, la cantidad de modelos posibles depende de manera directa del número de categorías. Otro elemento que también influye en el número de modelos es la existencia o no de cierta monotonía de la respuesta respecto a un ordenamiento de las categorías de la variable independiente.

Cuando  $m = 3$ , el conjunto de categorías queda bien representado como  $\Omega = \{1, 2, 3\}$ . Para este caso, un modelo posible con 2 parámetros es:

$$g(x) = \begin{cases} \mu + \alpha & \text{si } x \in \{1\} \\ \mu - \alpha & \text{si } x \in \{2, 3\} \end{cases}$$

A través de este sencillo ejemplo se observa que la función  $g(\circ)$  genera una partición  $\pi : \{\{1\}, \{2,3\}\}$  sobre el conjunto de categorías de la variable explicativa, y cualquier otra función  $g(\circ)$  generará una partición diferente sobre  $\Omega$ . Las clases correspondientes a una partición  $\pi^g$  determinada por una ecuación  $g(x)$  se denotarán por  $C_j^g$ :

$$\pi_g : \Omega = \bigcup_j C_j^g$$

En forma recíproca, para todo  $i \in C_j^g, i' \in C_k^g$ , se tiene:

$$\begin{aligned} \text{Si } j=k &\Rightarrow g(i)=g(i') \\ \text{Si } j \neq k &\Rightarrow g(i) \neq g(i') \end{aligned}$$

En consecuencia existe una correspondencia uno a uno entre las funciones  $g$  y las posibles particiones del espacio  $\Omega$ .

Se supone que las observaciones  $Y_{ij}$  son realizaciones de una variable aleatoria que satisfacen la ecuación de regresión:

$$Y_{ij} = f(i) + \varepsilon_{ij} \quad (2.1)$$

donde los  $\varepsilon_{ij}$  son errores aleatorios (no observables) con media cero y varianza común  $\sigma^2$  y no están correlacionados. La función  $f$  pertenece a un conjunto  $M$  tal que:

$$M = \{g(x, \beta) : \beta \in B\} \quad (2.2)$$

las funciones  $g(x, \beta)$  son de la forma:

$$g(x, \beta) = \sum_{j=1}^p 1_{C_j^g}(x) \mu_j^g \quad p=1,2,\dots,m \quad (2.3)$$

siendo  $\beta^t = (\mu, \alpha_1, \alpha_2, \dots, \alpha_p)$  un vector de parámetros desconocidos con  $p \leq m$ ,  $\mu_j^g$  tiene la forma  $\mu_j^g = \mu + \alpha_j$  y  $1_{C_j^g}(x)$  es la función indicadora de la clase  $C_j^g$ .

Dado que  $f$  es desconocida, en base a las observaciones se debe calcular un estimador  $\hat{\beta}$  para el vector de parámetros  $\beta$  y así determinar un elemento  $g(x, \hat{\beta})$  del conjunto  $M$ , que aproxime a la función  $f$ . La elección de la función  $g(x, \hat{\beta})$  es el resultado de un proceso de minimización de un estimador del cuadrado medio del error (CME) o del cuadrado medio del error de predicción (CMEP). El estimador mínimo cuadrático ordinario del vector de parámetros  $\beta$  se denotará por  $\hat{\beta}$ .

Sea  $p_o$  un número entero positivo, en el primer paso de la estrategia se analizarán los modelos que tengan un número de parámetros  $p \leq p_o$ , donde  $p_o$  lo define el usuario considerando sus recursos de tiempo y cómputo disponibles.

**Proposición 2.1.**

Sea  $X$  una variable cualitativa con  $m$  categorías diferentes y sean las observaciones  $Y_{ij}$ , con  $i = 1, 2, \dots, m$ , y  $j = 1, 2, \dots, n_i$  correspondientes a los  $m$  puntos de diseño diferentes de  $X$ . Entonces el número de modelos diferentes en  $\mathbf{M}$  con  $p \leq m$  parámetros está dado por

$$\sum_{j=1}^p \frac{(-1)^{p-j}}{j!(p-j)!} j^m \quad (p \leq m) \tag{2.4}$$

**Demostración** : (ver Castells, 1999 siguiendo demostración que aparece en Kovacs, 1980; incluida en el Anexo B)).

En el caso anterior, el orden de las categorías es irrelevante, sin embargo, existen situaciones donde el orden de las categorías cobra importancia, es el caso en el que se tiene información sobre un comportamiento monótono de la variable de respuesta  $Y$ , ante un ordenamiento de las categorías de  $X$ . Castells (1999) hace comentarios acerca de cómo puede darse en la práctica esta monotonía.

Incorporando esta idea, las funciones  $g(x)$  tendrían que cumplir ciertos criterios de monotonía con respecto a una permutación  $\rho$  del conjunto de las categorías de  $X$ , por lo que también es importante conocer el número de modelos existentes bajo estas consideraciones.

**Proposición 2.2.**

Se supone que  $Y$  es una variable aleatoria que tiene un comportamiento monótono con respecto a un ordenamiento de las categorías de  $X$  dado por una permutación  $\rho$ . Entonces el número de modelos diferentes con  $p$  parámetros está dado por

$$\binom{m-1}{p-1} \tag{2.5}$$

**Demostración:** (Castells 1999; incluida en el Anexo B)

Uno de los elementos centrales en la estrategia propuesta para la búsqueda del mejor modelo, son los llamados modelos vecinos que a continuación se definen:

**Definición 2.1.** Sean las ecuaciones  $g_o(x)$  y  $g_v(x)$  que generan las particiones  $\pi_o$  y  $\pi_v$  del espacio  $\Omega$  y que definen los modelos  $M_o$  y  $M_v$  respectivamente. Se denotan por  $C_{oj}$  y  $C_{vj}$ ; las clases correspondientes a  $\pi_o$  y  $\pi_v$ . Se dice que  $M_v$  es un modelo vecino de  $M_o$  mayor que él, si existe una clase y sólo una clase  $C_{ok}$  tal que:

$$C_{ok} = C_{vk'} \cup C_{vk''}$$

además, para todo  $r \neq k$  debe existir  $q$  distinto de  $k'$  y  $k''$  de tal forma que  $C_{or} = C_{vq}$ . Entonces se dirá que  $M_o$  es un vecino menor de  $M_v$ .

La diferencia entre dos modelos vecinos se establece en la clase  $C_{ok}$ , la cual en el vecino mayor aparece subdividida en las clases  $C_{vk'}$  y  $C_{vk''}$ . Esto se ve reflejado en el modelo  $M_v$  con un parámetro más que el modelo vecino menor  $M_o$ . Como  $M_o$  es un modelo con ecuación  $g_o(x)$  que tiene  $p$  parámetros, esto es,

$$g_o(x) = \begin{cases} \mu + \alpha_k & \text{si } x \in C_{ok}, \quad k = 1, 2, \dots, p-1 \\ \mu - \sum_{i=1}^{p-1} \alpha_k & \text{si } x \in C_{op} \end{cases} \quad (2.6)$$

la partición  $\pi_o$  asociada tiene  $p$  clases diferentes. Como existe una y sólo una clase  $C_{ok}$  para la cual existen  $k'$  y  $k''$  de tal forma que

$$C_{ok} = C_{vk'} \cup C_{vk''}$$

y para cualquier otro  $r \neq k$  existe  $q$  distinto de  $k'$  y  $k''$  tal que  $C_{or} = C_{vq}$ , la ecuación  $g_v(x)$  tiene la forma

$$g_v(x) = \begin{cases} \mu + \alpha_k & \text{si } x \in C_{vk}, \quad k = 1, 2, \dots, p \\ \mu - \sum_{i=1}^p \alpha_k & \text{si } x \in C_{v,p+1} \end{cases} \quad (2.7)$$

esto es, define un modelo con  $p+1$  parámetros.

La relación entre el conjunto de las particiones posibles y el conjunto de las funciones de regresión puede ser explotada convenientemente para definir la estrategia de selección. La partición que corresponde a la ecuación de regresión verdadera se denotará por  $\pi_o$ . Obsérvese que esta partición no tiene que ser la mejor en el sentido del *CMEP* o del *CME* del estimador  $\hat{\mu}^{\pi_o}$ .

**Definición 2.2.** Se dirá que  $\pi_*$  es una partición óptima si cumple que:

$$CMP(M(\pi_*)) = \text{Min}_{\pi \in \Pi} CMEP(m(\pi)) \quad (2.8)$$

La partición  $\pi_*$  puede conducir a un modelo con un *CMEP* menor que el del modelo  $m(\pi_o)$  y a un estimador con un *CME* menor que el de  $\hat{\mu}^{\pi_o}$ . Por supuesto, este *CMEP* puede ser menor que el correspondiente a la partición más fina, que es la que determina al estimador mínimo cuadrático (ver Bunke y Castells, 1999 y Castells, 1999. Ver además Hocking, 2003; Bunke y Strüby, 1975 y Bunke y Bunke, 1989).

La determinación de la partición óptima se debe hacer de manera dependiente de los datos puesto que hay una fuerte relación entre el estimador óptimo  $\hat{\mu}^{\pi_*}$  y los valores  $\mu$  y  $\sigma^2$ . Bunke y Castells (1999) proponen un estimador adaptivo basado en su estrategia que en los ejemplos que reportó mejoró al estimador mínimo cuadrático.

## 2.1 Estrategia para la selección del modelo.

Se denota por  $M(p)$  el modelo con  $p$  parámetros distintos. Una vez que se ha fijado el valor de  $p_0$ , el conjunto de modelos con un número de parámetros menor o igual que  $p_0$  se denota por:

$$M_o = M_o(p_o) = \{M(p) : p \leq p_o\}$$

Sea  $\hat{r}(M) = \hat{r}(M(p))$  un estimador del CMEP para el modelo  $M(p)$ . Estudios interesantes sobre el CMEP y sus estimadores pueden encontrarse en Allen (1971); Mallows (1973 y 1995) y Bunke y Droge (1984)

### Etapas de la estrategia:

- I. Cálculo del número de modelos posibles.
- II. Determinación del valor de  $p_0$ .

Considerando el número de modelos posibles, el tiempo disponible y los recursos de cómputo disponibles por el usuario, se determina el valor de  $p_0$ . Al conjunto  $M_o(p_o)$  se le llama conjunto de modelos básicos

III. Seguir los siguientes pasos:

1) Determinar  $M_1(p_1)$  como:

$$M_1(p_1) = \text{ArgMin}_{M \in M_o} \hat{r}(M)$$

$$p_1 = p_0 + 1$$

2) Determinar el conjunto de modelos vecinos mayores de  $M_1(p_1)$ , los cuales se denotan por  $M_{1v} = M_{1v}(p_1 + 1)$ .

Determinar  $M_2(p_2)$  como:

$$M_2(p_2) = \text{ArgMin}_{M \in M_{1v}} \hat{r}(M)$$

$$p_2 = p_1 + 1$$

3) Determinar el conjunto de modelos vecinos mayores de  $M_2(p_2)$ , los cuales se denotan por  $M_{2v} = M_{2v}(p_2 + 1)$ .

Determinar  $M_3(p_3)$  como:

$$M_3(p_3) = \text{ArgMin}_{M \in M_{2v}} \hat{r}(M)$$

$$p_3 = p_1 + 2$$

: : :

i) Determinar el conjunto de modelos vecinos mayores de  $M_{(i-1)}(p_{(i-1)})$ , los cuales se denotan por  $M_{(i-1)v} = M_{(i-1)v}(p_{(i-1)} + 1)$ .

Determinar  $M_i(p_i)$  como:

$$Mi(p_i) = \text{ArgMin}_{M \in M_{(i-1)o}} \hat{r}(M)$$

$$p_i = p_1 + i - 1 \\ \vdots \quad \vdots$$

$m + 1 - p_1$ ). Determinar el conjunto de modelos vecinos mayores de  $M_{(m-p_1)}(p_{(m-p_1)})$  los cuales se denotan por  $M_{(m-p_1)v} = M_{(m-p_1)v}(m)$

Determinar  $M_{(m+1-p_1)}(p_{(m+1-p_1)}) = M_{(m+1-p_1)}(m)$  como:

$$M_{(m+1-p_1)}(p_{(m+1-p_1)}) = \text{ArgMin}_{M \in M_{(m-p_1)v}} \hat{r}(M)$$

$$p_{m+1-p_1} = m - 1 \quad \vdots \quad \vdots$$

En este caso el conjunto  $M_{(m-p_1)v}$  tiene como único elemento el modelo saturado o modelo con  $m$  parámetros. El paso número  $m + 1 - p_1$  es el último de la etapa donde la estrategia va incrementando el número de parámetros, en el siguiente se inician los pasos donde el número de parámetros es decreciente. En los pasos que siguen en lugar de determinar el conjunto de modelos vecinos mayores, se determinará el conjunto de vecinos menores.

$m + 2 - p_1$ ). Determinar el conjunto de modelos vecinos menores de  $M_{(m+1-p_1)}(m)$  los cuales se denotan  $M'_{(m+1-p_1)v} = M'_{(m+1-p_1)v}(m-1)$

Determinar  $M_{m+2-p_1}(m-1)$  como:

$$M_{m+2-p_1}(m-1) = \text{Arg Min}_{M \in M'_{(m+1-p_1)v}} \hat{r}(M)$$

$$p_{m+2-p_1} = m - 1$$

$m + 3 - p_1$ ). Determinar el conjunto de modelos vecinos menores de  $M_{(m+2-p_1)}(m-1)$  los cuales se denotan  $M'_{(m+2-p_1)v} = M'_{(m+2-p_1)v}(m-2)$

Determinar  $M_{(m+3-p_1)v}(m-2)$  como:

$$M_{m+3-p_1}(m-2) = \text{Arg Min}_{M \in M'_{(m+2-p_1)v}} \hat{r}(M)$$

$$p_{m+3-p_1} = m - 2$$

$\vdots \quad \vdots$

$2(m - p_1) + 1$ ). Determinar el conjunto de modelos vecinos menores de  $M_{2(m-p_1)}(p_1 + 1)$  los cuales se denotan por  $M'_{[2(m-p_1)]v} = M'_{[2(m-p_1)]v}(p_1 + 1)$

Determinar  $M_{2(m-p_1)+1}(p_1)$  como:

$$M_{2(m-p_1)+1}(p_1) = \text{Arg Min}_{M \in M_{(m+2-p_1)v}} \hat{r}(M)$$

El resultado de aplicar esta estrategia es un conjunto de modelos con distinto número de parámetros. El criterio de selección es el CMEP y la estrategia está enfocada hacia el uso predictivo del modelo.

### 3. VALIDACIÓN DE LA ESTRATEGIA PROPUESTA.

Para validar la estrategia propuesta se programó la misma y diferentes rutinas necesarias para el análisis, se utilizó el paquete estadístico S-Plus (Venables et al., 1994).

Los datos utilizados fueron simulados. Se consideraron variables independientes con 5, 7, 9 y 10 categorías. Los valores correspondientes a los elementos que intervienen en la especificación del modelo general (1.1) se reportan en la Tabla No. 1 del Anexo; en el caso de la media general  $\mu$  se mantuvo constante e igual a la unidad y se fijaron de igual tamaño las submuestras  $n_1 = n_2 = \dots = n_9$ . Para cada una de las variantes del modelo se generaron 20 juegos de datos.

Como es de esperarse, el valor promedio de las estimaciones del cuadrado medio del error de predicción crece cuando crece la varianza y se comporta de manera aproximadamente igual cuando se mantiene la varianza fija y crece  $p$  (Tabla 2 y Tabla 2A del Anexo). La estrategia seleccionó el 70.58% de las veces al modelo optimal. Castells (1999), obtuvo el 80.5%, lo que parecía indicar que ahora debería obtenerse un porcentaje superior. En trabajos futuros se aumentará el número de simulaciones con el objetivo de revalorar estos dos resultados.

La diferencia entre los porcentajes de selección del modelo optimal, cuando hay monotonía o no en la variable dependiente, es grande: de un 87% cambia a un 54% (según se desprende de la Tabla 3 del Anexo). Cuando va creciendo la varianza no hay diferencia notable entre los porcentajes de selección del *modelo optimal*, aunque se observa una ligera disminución: cuando hay monotonía cambia de 89.1% a 84.4%, en caso contrario cambia de 59.5% a 44.4%.

Fijando la varianza, los valores del cuadrado medio del error de predicción estimado no dependen mucho del número de categorías.

También se analizó el promedio de la variación entre las estimaciones del cuadrado medio del error de predicción de los modelos seleccionados y los optimales, obteniéndose que: en el caso en que no hay monotonía cambia de 0.0000, con 5 categorías, a 0.3413, con 10 categorías, y en caso contrario cambia de 0.0000, con 5 categorías, a 0.0711, con 10 categorías (ver Tabla No. 4 del Anexo).

RECEIVED JANUARY 2008  
REVISED JANUARY 2009

### REFERENCIAS

- [1] ALLEN, D.M. (1971): Mean Square Error of Prediction as a Criterion for Selecting Variables. *Technometrics*, 13, .469-474.
- [2] BUNKE, O y BUNKE, H (EDITORS) (1989): **Statistical Methods of Models Building**. John Wiley & Sons, Inc, Chichester.
- [3] BUNKE, O. y CASTELLS, E. (1999): Regression and Contrast Estimates based on Adaptive Regressograms depending on Qualitative Explanatory Variables. *Statistics*, 33, .37-56.
- [4] BUNKE, O. Y STRÜBY, R. (1975): Estimation procedures in inadequate models: Comparisons and empirical two step procedure. *Mathematische Operationsforschung und Statistik* 6, .167-177.

- [5] BUNKE, O. Y DROGE, B. (1984): Estimators of the mean squared error of prediction in linear regression. **Technometrics**, .26, .145-154.
- [6] CASTELLS GIL, E. (1999): Selección de Modelos de Regresión con Regresores Cualitativos. **Tesis doctoral** Universidad de La Habana, Cuba.
- [7] DRAPER, N. Y SMITH, H. (1998): **Applied Regression Analysis**. John Wiley & Sons, Inc. New York.
- [8] HOCKING, R. R. (2003): **Methods and Applications of Linear Models: Regression and the Analysis of Variance. 2<sup>nd</sup> Edition**. John Wiley & Sons, Inc. New York.
- [9] KOVACS, L.B. (1980): **Combinatorial Methods of Discrete Programing. Mathematical Methods of Operations Research**, Vol.2. Akadémiai Kiadó. Budapest.
- [10] MALLOWS, C.L. (1973): Some comments on  $C_p$ . **Technometrics**, .15, .661-675
- [11] MALLOWS, C. L. (1995): More comments on  $C_p$ . **Technometrics**, 37,.362-372
- [12] MATHSOFT (1999): **S-PLUS 2000**. Mathsoft, Inc., Seattle, Washington.
- [13] MOOD, A. M., GRAYBILL, F. Y BOES, D. (1974): **Introduction to the Theory of Statistics**. McGraw-Hill International Book Company. New York.
- [14] RAO, C.R. (1973): **Linear Statistical Inferences and its Applications**. John Wiley & Sons, Inc. New York.
- [15] SCHEFFÉ, H. (1982): **The Analysis of Variance**. Segunda edición. John Wiley & Sons, Inc. New York.
- [16] SEARLE, S. R. (1971): **Linear Models**. John Wiley & Sons, Inc. New York.
- [17] VENABLES, W. N. Y RIPLEY, B. D. (1994): **Statistics and Computing. Modern Applied Statistics with S-Plus**. Springer-Verlag. Berlín.

## ANEXO A

**Tabla No. 1.-** Valores de los parámetros de los modelos con que se generaron los datos simulados.

No. de ejercicio	Parámetros diferentes	Desv. estándar	Valores de los parámetros del modelo									
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_9$	$\alpha_{10}$
Modelos con 5 categorías												
1	2	1	-1.7	-1.7	-1.7	1.7	1.7					
2	3	1	-1.5	-0.2	-0.2	1.7	1.7					
3	4	1	-1.0	-0.9	0.2	1.7	1.7					
Modelos de 7 categorías												
4	4	1	-2.2	0.4	0.4	0.7	0.7	0.7	1.1			
5	6	1	-2.8	-2.1	-2.1	-1.4	1.4	2.1	2.8			
6	4	5	-7.0	-5.0	3.0	3.0	3.0	3.0	9.0			
7	6	5	-8.9	-6.0	-6.0	-5.0	6.2	6.7	7.0			
8	4	10	-16.0	-15.0	-15.0	13.0	13.0	18.0	18.0			
9	6	10	-15.0	-15.0	-14.0	-13.0	13.0	14.0	15.0			
Modelos de 9 categorías												
10	4	1	-1.0	-0.2	-0.2	-0.2	0.2	0.2	0.2	0.2	1.0	
11	6	1	-0.8	-0.5	-0.2	0.2	0.2	0.2	0.2	0.3	1.0	
12	7	1	-3.0	-2.0	-1.8	1.5	1.5	1.6	1.6	1.7	2.0	
13	4	5	-10.0	-10.0	-7.0	-7.0	8.0	8.0	9.0	9.0	9.0	
14	6	5	-8.0	-7.9	-6.0	-6.0	-6.0	6.0	6.0	6.9	9.0	
15	7	5	-9.8	-9.7	-8.6	5.2	5.2	5.2	6.1	7.0	9.8	
16	4	10	-15.0	-14.0	-14.0	-14.0	-14.0	-14.0	14.0	14.0	15.0	
17	6	10	-15.0	-15.0	-14.0	-14.0	-13.0	13.0	14.0	14.0	15.0	
18	7	10	-21.0	-19.0	-15.0	-15.0	10.0	10.0	13.0	15.0	17.0	
Modelos de 10 categorías												
19	5	1	-3.0	-2.0	-2.0	-0.9	2.9	2.9	2.9	3.0	3.0	3.0
20	6	1	-2.5	-2.0	-1.6	1.1	1.1	1.1	2.0	2.0	3.0	3.0
21	8	1	-2.7	-2.0	-2.0	-1.5	-1.4	1.4	1.5	2.0	2.0	2.7
22	5	5	-10.0	-10.0	-10.0	-9.3	5.0	5.0	5.3	9.0	9.0	9.0
23	6	5	-9.0	-7.0	-5.0	6.0	6.0	7.0	7.0	7.0	8.0	8.0
24	8	5	-8.0	-7.0	-6.0	-6.0	-6.0	-5.0	5.0	6.0	7.0	8.0
25	5	10	-12.0	-11.0	-10.0	-10.0	-10.0	-10.0	15.0	15.0	15.0	18.0
26	6	10	-15.0	-13.0	-10.0	-10.0	-10.0	10.0	13.0	13.0	15.0	15.0
27	8	10	-20.0	-15.0	-12.0	-11.0	10.0	10.0	13.0	17.0	18.0	18.0

**Tabla No. 2.-** Valor mínimo, máximo, rango y promedio del estimador del Cuadrado medio del error de predicción para cada combinación de  $m$ ,  $p$  y  $\sigma$ , (Y sin monotonía)

No. de ejercicio	Parámetros diferentes	Desv. estándar	Estimador del CMEP			
			Promedio	Mínimo	Máximo	Rango
Modelos con 5 categorías.						
1	2	1	1.967	1.328	2.354	1.026
2	3	1	1.386	0.845	1.696	0.851
3	4	1	1.194	0.909	1.586	0.677
Modelos con 7 categorías						
4	4	1	0.994	0.681	1.355	0.674
5	6	1	1.033	0.781	1.370	0.590
6	4	5	27.816	20.080	35.914	15.833
7	6	5	26.999	16.314	40.972	24.658
8	4	10	97.611	70.428	126.839	56.411
9	6	10	112.598	72.775	151.709	78.934
Modelos con 9 categorías						
10	4	1	1.012	0.740	1.441	0.701
11	6	1	1.065	0.838	1.347	0.508
12	7	1	1.097	0.764	1.506	0.742
13	4	5	26.310	17.030	33.482	16.452
14	6	5	25.085	13.420	38.095	24.675
15	7	5	27.167	19.545	41.988	22.442
16	4	10	94.525	57.153	135.105	77.952
17	6	10	109.092	80.788	142.395	61.607
18	7	10	108.869	72.567	157.067	84.499
Modelos con 10 categorías						
19	5	1	1.080	0.736	1.583	0.847
20	6	1	1.016	0.732	1.352	0.620
21	8	1	1.054	0.458	1.668	1.210
22	5	5	27.050	16.065	35.834	19.769
23	6	5	28.387	17.108	44.050	26.942
24	8	5	27.986	18.398	36.044	17.646
25	5	10	99.144	63.708	120.819	57.110
26	6	10	101.541	79.137	145.086	65.949
27	8	10	107.064	64.408	146.560	82.152

**Tabla No. 2A.-** Valor mínimo, máximo, rango y promedio del estimador del Cuadrado medio del error de predicción para cada combinación de  $m$ ,  $p$  y  $\sigma$ . (Y con monotonía)

No. de ejercicio	Parámetros diferentes	Desv. estándar	Estimador del CMEP			
			Promedio	Mínimo	Máximo	Rango
Modelos con 5 categorías.						
1	2	1	1.943	1.328	2.353	1.025
2	3	1	1.386	0.845	1.696	0.851
3	4	1	1.197	0.942	1.586	0.644
Modelos con 7 categorías						
4	4	1	0.992	0.688	1.353	0.665
5	6	1	1.034	0.784	1.370	0.586
6	4	5	27.862	20.161	35.296	15.136
7	6	5	27.058	16.314	40.972	24.658
8	4	10	97.939	72.601	126.839	54.238
9	6	10	112.637	72.775	151.709	78.934
Modelos con 9 categorías						
10	4	1	1.043	0.751	1.571	0.820
11	6	1	1.071	0.840	1.387	0.547
12	7	1	1.122	0.770	1.527	0.757
13	4	5	26.243	17.030	31.943	14.913
14	6	5	25.297	13.469	38.136	24.667
15	7	5	27.659	19.625	42.185	22.560
16	4	10	95.607	59.057	136.172	77.115
17	6	10	110.664	80.900	147.723	66.824
18	7	10	113.774	77.319	171.739	94.421
Modelos con 10 categorías						
19	5	1	1.080	0.753	1.582	0.829
20	6	1	1.020	0.732	1.352	0.620
21	8	1	1.085	0.467	1.764	1.297
22	5	5	27.103	17.210	35.834	18.624
23	6	5	28.520	17.215	43.667	26.451
24	8	5	29.919	18.985	41.893	22.907
25	5	10	100.739	64.501	119.856	55.355
26	6	10	102.449	79.137	145.866	66.728
27	8	10	109.862	66.395	172.092	105.697

**Tabla No. 3.-** Número de veces que la estrategia seleccionó al modelo optimal de acuerdo a los valores de  $m$  y  $\sigma$ .

m	$\sigma$	Número de veces que resultó seleccionado el modelo optimal	
		Y con monotonía	Y sin monotonía
5	1	60	60
7	1	39	26
7	5	40	36
7	10	40	29
9	1	57	29
9	5	59	41
9	10	58	31
10	1	58	28
10	5	57	25
10	10	54	20

**Tabla No. 4.-** Diferencia promedio entre el valor del estimador del cuadrado medio del error de predicción del modelo seleccionado y el del modelo óptimo de acuerdo a los valores de  $m$ .

m	diferencia promedio	
	Y con monotonía (%)	Y sin monotonía (%)
5	0.0000	0.0000
7	0.0000	0.2078
9	0.0197	0.1937
10	0.0711	0.3413

Anexo B

**Demstración de la Proposición 2.1 con  $p$  parámetros**

Sean:  $N(m, p)$  - número de modelos con  $p$  parámetros.

$\Omega$  - conjunto formado por las categorías de la variable  $X$ .

$D$  - un conjunto de cardinal  $d$  ( $d \geq p$ )

Kovacs (1980) da una demostración de valor general para (2.04) que se puede retomar en este caso particular, donde el interés está en determinar el número de modelos con  $p$  parámetros: Los elementos de  $\Omega$  se pueden hacer corresponder con los elementos de  $D$  en  $d^m$  formas diferentes. Pero estas formas de correspondencia podrían contarse así:

Sea  $f : \Omega \rightarrow D$  tal que el rango de  $f$  tenga cardinal  $p$ , entonces las correspondencias de este tipo pueden determinar  $N(m, p)$  particiones diferentes en los elementos de  $\Omega$ . Sea  $\pi$  una de estas particiones, pero fija. Entonces a cada clase de  $\pi$  le corresponde una imagen en  $D$  y para clases diferentes las imágenes son diferentes. Por tanto, a una clase le pueden corresponder  $d$  imágenes, a la otra  $d - 1$  y así sucesivamente. De manera que el número de formas diferentes en que se puede hacer la correspondencia con rango  $p$  de  $\Omega$  en  $D$ , es el dado por la expresión (B.01):

$$N(m, p).d.(d - 1) \cdots (d - p + 1) \tag{B1}$$

Entonces las funciones polinomiales  $d^m$  y la dada en (B.1) coinciden para infinitos valores enteros de  $d$  y extendiendo el dominio de éstas al conjunto de los números reales,  $R$ , los polinomios que resultan tienen que ser iguales.

Por otro lado,

$$\begin{aligned} N(m, p) &= N(m, p) + \sum_{r=0}^{p-1} \frac{N(m, r)}{(p-r)!} (1-1)^{p-r} = \sum_{r=0}^p \frac{N(m, r)}{(p-r)!} \sum_{j=r}^p (1-1)^{p-r} \\ &= \sum_{r=0}^p \frac{N(m, r)}{(p-r)!} \sum_{j=r}^p (-1)^{p-j} \binom{p-r}{p-j} = \sum_{j=0}^p \sum_{r=0}^j \frac{(-1)^{p-j} N(m, r)}{(p-j)!(j-r)!} \\ &= \sum_{j=0}^p \frac{(-1)^{p-j}}{j!(p-j)!} \sum_{r=0}^j N(m, r) j(j-1) \cdots (j-r+1) = \sum_{j=0}^p \frac{(-1)^{p-j}}{j!(p-j)!} j^m \end{aligned}$$

que es lo que se quería probar. El número total de modelos que se pueden obtener es el dado en (B.2):

$$N(m) = \sum_{p=1}^m \sum_{j=0}^p \frac{(-1)^{p-j}}{j!(p-j)!} j^m \quad (\text{B2})$$

### **Demostración de la Proposición 2.2**

Se sabe ya que contar el número de modelos con  $p$  parámetros, equivale a contar el número de particiones con  $p$  clases que se pueden hacer sobre  $\Omega$ , pero ahora las clases de la partición hay que tomarlas en un orden fijo, es decir, no se pueden permutar para que las funciones  $g(x)$  así encontradas reflejen el comportamiento monótono indicado. De esta forma, si se consideran las categorías ordenadas,  $\rho(1), \rho(2), \dots, \rho(m)$  como  $m$  puntos sobre un segmento, la pregunta que debe formularse es: ¿De cuántas formas diferentes se puede dividir ese segmento en  $p$  subsegmentos? Está claro que como hay que respetar el orden, eso se puede hacer de combinaciones de  $m-1$  en  $p-1$  maneras diferentes. Para contar el total de modelos posibles sólo hay que hacer variar  $p$  desde 1 hasta  $m$  y por tanto se tendrá:

$$N(m) = \sum_{p=1}^m \binom{m-1}{p-1}$$