

## Alternativas metodológicas al análisis de regresión lineal

**Vidal Díaz de Rada Igúzquiza**

*Universidad Pública de Navarra*

### RESUMEN

Este trabajo comienza con una exposición de las hipótesis que deben tenerse en cuenta con objeto de aplicar correctamente la técnica estadística de la Regresión Lineal. Una vez expuestas estas condiciones, se realiza una explicación del método de Regresión Logística surgido a mediados de la década de los 60 como una alternativa metodológica con objeto de superar las limitaciones del modelo lineal.

### SUMMARY

This work begins by showing the hypothesis which must be taken into account in order to apply the Lineal Regression statistics technique correctly. Once these conditions are shown, the method of Logistic Regression is explained. It was first used in the mid-60's as a methodologic alternative which aimed to overcome the limitations of the lineal model.

El análisis de Regresión es una técnica estadística que estudia la naturaleza de las relaciones entre un conjunto de variables con el objetivo de hacer predicciones; tratando de conocer el influjo que una serie de variables independientes  $X_i$  ejercen sobre una variable dependiente  $Y$ : Esta relación puede ser total –cuando se desea conocer el influjo conjunto que todas las variables independientes tienen en la dependiente– o parcial; cuando el objeto de estudio es analizar la influencia particular que cada variable independiente  $X$  ejerce sobre el término dependiente  $Y$ . No obstante, debe tenerse presente en todo momento que esta relación entre las variables no implica que unas sean la causa de otras, puesto que los diversos coeficientes aportan información sobre la probabilidad con que puede ocurrir un fenómeno determinado y sobre la relación existente entre varios fenómenos, no siendo posible extraer de ellos *leyes de causalidad* (Tabachnick y Fidell, 1989: 127).

El objetivo de este trabajo no es explicar la técnica de la Regresión Lineal<sup>1</sup>, sino analizar las condiciones que deben tenerse en cuenta cuando ésta es aplicada a una matriz de datos, así como exponer las distintas alternativas existentes cuando no es posible la aplicación correcta de la misma.

El trabajo comienza con una exposición de los criterios que deben considerarse cuando se desean aplicar distintas técnicas estadísticas, para analizar a continuación los supuestos que deben tenerse en cuenta cuando se desea aplicar correctamente la técnica estadística de la Regresión Lineal. Posteriormente, y de las distintas alternativas disponibles, nos centraremos en el modelo de Regresión Logística que requiere unos supuestos menos restrictivos, y que su utilización ha aumentado notablemente en los últimos años (Amemiya 1981: 1480).

## 1. INTRODUCCIÓN

Aunque en un primer momento se deseaba realizar un planteamiento general de las técnicas multivariantes a fin de contextualizar dentro de éstas la técnica de la Regresión Lineal, eludiremos el análisis pormenorizado de esta temática debido a las limitaciones de espacio y –fundamentalmente– a la gran cantidad de textos dedicados específicamente a esta tarea. Realizada esta salvedad, el punto de partida de este trabajo –una vez definida la técnica de regresión– es una exposición de los distintos

1. Si se desea una exposición pormenorizada de las bases teóricas de la técnica de Regresión aconsejamos la lectura de: W.D. Berry y S. Feldman (1984), Bowman y Robinson (1990), Edwards (1984), Guillén (1992), Shroeder (1986).

criterios a tener en cuenta para aplicar correctamente una determinada técnica estadística, utilizando la ya clásica clasificación realizada por Sheth (1971) en la que clasifica los métodos multivariantes según las respuestas a tres cuestiones<sup>2</sup>:

- ¿Son algunas variables dependientes de otras?
- En caso de ser así, ¿Cuántas variables son dependientes?
- ¿Cuáles son las propiedades –en cuanto a la métrica– de los datos disponibles?

En una publicación realizada ese mismo año, T.C. Kinnear y J.R. Taylor introdujeron una *cuarta pregunta* que solicita información sobre la métrica en la que están medidas las variables independientes.

El resultado de estas preguntas es la ya conocida «estructura arbórea» propuesta por estos autores en las que se presentan nueve técnicas multivariantes en el esquema de Sheth y dieciseis en el propuesto por Kinnear y Taylor. No obstante, en los últimos años han aparecido modelos de clasificación más complejos, como el propuesto por Andrews y otros (1981), en los que las preguntas hacen referencia no sólo al número de variables que aparecen en el problema, el tipo de medida de cada una de ellas y si hay distinción entre variables dependientes e independientes, sino también al objetivo de la investigación, la naturaleza de la relación entre las variables, etc. Con esas *cuestiones*, los autores no sólo elaboran una clasificación de los métodos multivariantes y sus condiciones de aplicación, sino que van más allá al realizar un análisis pormenorizado de la utilización de una gran cantidad de test estadísticos<sup>3</sup>.

Elegiendo el último de estos esquemas, nos interesaremos en primer lugar por las técnicas que analizan la relación de dependencia entre las variables, para preguntar después por el tipo de métrica que emplean: cuando todas las variables –dependientes e independientes– son cuantitativas y el investigador busca una relación de dependencia respecto a una variable, habrá que utilizar la técnica estadística de la Regresión Simple (con una variable independiente) o múltiple (cuando las variables independientes son más de una). La robustez del modelo permite incluso que algunas variables independientes sean cualitativas y que reciben el nombre de «Dummy variables» (García Ferrando, 1985: 410-411; Lewis-Beck, 1990: 66-71). No obstante, y una vez

2. Recogido de Sanz de la Tajada, 1990: 410-411. Véase también García Ferrando 1985: 385-393.

3. No obstante, en la página 60 los autores indican que determinadas técnicas estadísticas quedan fuera de esta clasificación debido fundamentalmente a su novedad y a la escasa literatura sobre las mismas, que genera una escasa utilización de éstas.

definidas las variables a analizar, el investigador debe verificar si el modelo de regresión cumple las hipótesis y supuestos de inferenciabilidad que requiere esta técnica estadística.

## 2. SUPUESTOS QUE DEBE SATISFACER EL MODELO DE REGRESIÓN LINEAL

En la consideración de las condiciones necesarias para poder aplicar correctamente la técnica de la Regresión Lineal se va seguir el modelo planteado por W.D. Berry (1993: 3-13) y Lewis-Beck (1990: 26). Desde el punto de vista de estos autores, y una vez está correctamente especificado el modelo de regresión, se deben cumplir los siguientes supuestos:

- Todas la información contenida en las variables utilizadas, tanto las independientes como la dependiente, ha sido recogida sin error.
- La relación entre las variables independientes y la dependiente es lineal.
- La media de la distribución de todos los términos de error es igual a cero.
- Cumplir el principio de Homocedasticidad o igualdad de varianzas.
- Los términos de error están incorrelacionados.
- Los residuos se distribuyen normalmente.
- Ausencia de multicolinealidad.

Una explicación detallada de los significados de cada hipótesis y de los procesos de confirmación se exponen en las siguientes páginas; aunque antes de explicarlos es necesario dejar claro que la primera condición a tener en cuenta al realizar un análisis de regresión es realizar una *correcta especificación del modelo de regresión*. Aunque esta no es propiamente una hipótesis de partida, consideramos que es necesario prestarle una atención especial ya que si el diseño de la investigación es un aspecto de vital importancia cuando hay que realizar cualquier investigación, con el método de la Regresión Múltiple la elaboración del diseño adquiere una importancia fundamental. Tal y como apunta Bisquerra (1989: 200-201), la selección del modelo es el primer paso del Análisis de Regresión, y «es un elemento que supera los aspectos puramente estadísticos puesto que requiere una profundización en los aspectos teóricos del problema a tratar».

En este sentido, la correcta especificación del modelo de Regresión significa que no se han excluido variables relevantes en el modelo y que no han sido incluidas variables irrelevantes en el mismo. Uno de los procedimientos utilizados se fundamenta en un análisis de aquellas variables que mejor correlacionan con el término

dependiente e ir añadiéndolas al modelo, aunque el más utilizado es realizar la Regresión Múltiple «por etapas»; seleccionando paulatinamente –de todas las variables independientes– aquellas que mejor explican la variable dependiente: El proceso comienza con la selección de la variable independiente que tenga la correlación parcial mayor con el término dependiente. Si la ecuación resultante al seleccionar esta variable no resulta significativa al 0.05 (valor por defecto) termina el proceso de selección<sup>4</sup>. En caso contrario se vuelve a introducir en el análisis aquella que muestre el coeficiente de correlación parcial más elevado con objeto de maximizar el porcentaje de varianza explicado por el modelo.

Una vez escogida la segunda variable se vuelven a calcular los coeficientes de la ecuación y la significación de los mismos, y en caso que no cumplan los criterios establecidos esta variable quedará excluida del análisis. Este proceso se repite hasta que todas las variables independientes que satisfacen los criterios expuestos anteriormente y tengan un nivel de significación inferior al 0.05 estén introducidas en el modelo, mientras que las restantes quedarán excluidas de la ecuación. Aunque a grandes rasgos ésta es la lógica del procedimiento, dentro de los modelos de selección «por etapas» podemos diferenciar:

- Forward: Es el método explicado anteriormente, las variables van introduciéndose en la ecuación de modo ascendente.
- Backward: Es el proceso inverso, en un primer momento se introducen todas las variables en el modelo y posteriormente se van eliminando aquellas que no cumplen los requisitos.
- Stepwise: En los métodos anteriores puede ocurrir que una variable seleccionada en un primer momento reduzca su poder explicativo al introducir otras variables. Frente a esta deficiencia, el método «Stepwise» analiza en cada etapa la capacidad explicativa de TODAS las variables introducidas en el modelo, de modo que si en un momento determinado los coeficientes de una variable (ya introducida en el análisis) disminuyeran a un nivel de significación menor que el permitido, ésta quedaría excluida del análisis<sup>5</sup>.

4. Normalmente, para que una variable forme parte de la ecuación debe tener una alta tolerancia con objeto de evitar la multicolinealidad (concepto que será explicado en el epígrafe 2.6), un valor « $F$ » (calculado dividiendo la suma de cuadrados explicada por la ecuación entre la suma de cuadrados residual) mayor a 3.84 y una «probabilidad de  $F$ -to enter» inferior a 0.05. (Norusis, 1990: 278, Schroeder, 1986: 53).

5. No obstante, y pese a que este es uno de los procedimientos más utilizados, numerosos autores (Schroeder, Sjoquist y Stephan, 1986: 69-70) aconsejan prudencia con estos «métodos automáticos»

## 2.1. La relación entre las variables independientes y la dependiente es lineal

Un requisito imprescindible en la aplicación del modelo de Regresión Lineal es que las variables tengan una relación lineal; es decir, que el efecto de la variable independiente sobre la dependiente sea el mismo en cualquier valor de la primera (Bowman y Robinson, 1990: 23). De los distintos métodos que pueden utilizarse a fin de conocer si la relación entre una variable independiente y la variable dependiente es lineal, numerosos autores aconsejan utilizar una representación gráfica construyendo un diagrama de dispersión (scatterplot) entre las variables (Dunn y Clark, 1987). Si las variables se ajustan a la distribución normal y la relación entre ellas es lineal, el gráfico tendrá una forma ovalada. Si la forma de éste es distinta a la ovalada, la relación entre las variables no será lineal.

Cuando tenemos más de una variable independiente existe otra forma de comprobar la naturaleza de la relación entre las variables a través de un gráfico de dispersión entre los residuales y los valores predichos por la ecuación. Si la relación entre éstas es lineal, los residuales estarán distribuidos aleatoriamente (formando una especie de rectángulo) alrededor del valor horizontal 0. (Bowman, 1990: 215-228).

No obstante, si después de estos procedimientos descubrimos que las variables seleccionadas tienen una relación no lineal (parabólica, exponencial, etc.), puede optarse por realizar una transformación de éstas (por medio de logaritmos, raíces cuadradas, etc.) a fin de linealizar el modelo (Alvira Martín, 1985: 325-358; Freixa i Blanxart y otros 1992: 75-119). No obstante, J. Etxeberría aconseja prudencia en la transformación de los datos puesto que «todo modelo tiene una parte aleatoria o de error, y con las transformaciones, los errores también puede resultar profundamente alterados» (Etxeberría, 1990: pp. 79-92)<sup>6</sup>.

---

puesto que una variable considerada como fundamental para el planteamiento hipotético del investigador (pongamos que sea  $X_1$ ) puede quedar fuera del modelo debido a que tenga correlación con otra variable independiente ( $X_2$ ): Como será explicado más adelante –apartado 2.6– si la variable  $X_2$  tiene más relación con el término  $Y$  que la  $X_1$ , esta última quedará eliminada del modelo y en su lugar quedará seleccionada una variable muy adecuada en términos estadísticos, pero que aportará información poco relevante para nuestro planteamiento teóricos.

6. Los efectos derivados de una incorrecta asunción de la linealidad, así como una exhaustiva exposición de las distintas alternativas para corregir este problema pueden verse en Aldrich-Nelson (1980: 27-30 y 31-35).

## 2.2. La media de la distribución de todos los términos de error es igual a cero

Para cada valor de  $X$ , la media de los errores que se comete al predecir  $Y$  debe ser igual a cero; ó dicho de otro modo, que la media de todos los residuos (diferencia entre el valor observado y el valor predicho por el modelo) debe ser igual a cero. Según Michael Lewis-Beck este supuesto tiene una importancia secundaria en investigación social ya que cuando se viola únicamente «se produce un sesgo en la constante»; pero como en la investigación social generalmente estamos interesados en el análisis de los coeficientes de regresión a fin de conocer la influencia de cada variable en el término dependiente, el no cumplimiento de este supuesto no tendrá consecuencias prácticas (Lewis-Beck, 1980: 28).

## 2.3. Principio de homocedasticidad

El principio de homocedasticidad o igualdad de varianza viene a decir que la varianza de los errores debe ser constante; es decir, que el error de la variable dependiente « $Y$ » es igual en cada uno de los distintos valores de las variables dependientes  $X$ . Tal y como es expuesto por Jobson, «si la dispersión de los residuos aumenta o disminuye con los valores de las variables independientes o con los valores predichos, hay que cuestionar la hipótesis de igualdad de varianzas de  $Y$  para todos los valores de  $X$ » (1991: 168).

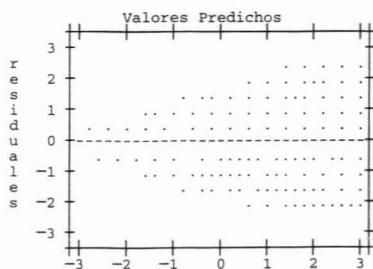


Figura A.

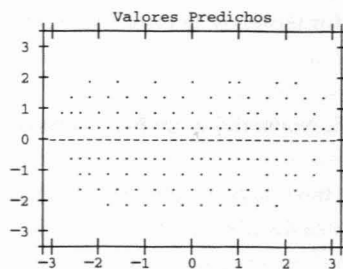


Figura B.

Fuente: Jobson (1991: 169)

En la figura A la distribución de los residuos se incrementa con la magnitud de los valores predichos, mientras que en la B esta distribución se mantiene constante con independencia de los valores que adopten la variable independiente  $Y$ .

#### 2.4. Los términos de error deben estar incorrelacionados: Ausencia de autocorrelación

Esta hipótesis viene a significar que los términos de error no están correlacionados, o lo que es lo mismo, que la respuesta de cada individuo es independiente de la respuesta de otros individuos y por lo tanto no existen problemas sistemáticos de medición. Con objeto de detectar la existencia de dependencia entre las observaciones suele utilizarse el test de *Durbin-Watson*, y cuya fórmula es:

$$D = \frac{\sum_{n=1}^n (E_t - E_{t-1})^2}{\sum_{n=1}^n E_t^2}$$

Donde:

$E_t$  es el residual de ecuación de regresión.

Las diferencias entre los residuales tienden a ser pequeñas cuando los terminos de error están correlacionados y son mayores cuando los terminos de error están incorrelacionados. De este modo, valores de  $D$  cercanos a cero indican una alta correlación, del mismo modo que altos valores de  $D$  expresan una ausencia de correlación entre los términos de error (Neter, Wasserman, Whitmore, 1988: 880-883).

#### 2.5. Normalidad de los residuos

Se trata de comprobar si la distribución de los residuos de la variable dependiente sigue una distribución normal, siendo éste uno de los supuestos más importantes puesto que si los residuos no tienen una distribución normal no se pueden emplear los test  $F$  de Snedecor y  $T$  de Student (Seber, 1977: 149-154). De los diversos métodos para comprobar si los residuos siguen la distribución normal, la más rigurosa consiste en la utilización de las pruebas estadísticas que permiten comprobar la normalidad. De todas las pruebas existentes (Pearson, Kolmogorov-Smirnov, Lilliefors, etc.) el más adecuado es el test de Lilliefors puesto que no requiere el conocimiento de los parámetros poblacionales, al tiempo que requiere que las variables tengan una métrica paramétrica (San Martín y Pardo, 1989: 91) como son las utilizadas en la Regresión Lineal Múlti-



ple. El test de Lilliefors, basado en la prueba de Kolmogorov-Smirnov no requiere -a diferencia de este último- el conocimiento de las medias y varianzas poblacionales (San Martín y Pardo, 1989: 91). No obstante, y al tener en cuenta que el Teorema Central del Límite enuncia que la distribución de una variable se aproxima a la normalidad a medida que el tamaño de la muestra aumenta, este supuesto no es muy importante en investigación social ya que generalmente se trabaja con grandes tamaños de muestra que hacen cumplir este Teorema (Calvo, 1990: 285-311).

## 2.6. Ausencia de multicolinealidad

Cuando se realiza un Análisis de Regresión debe considerarse muy detenidamente la existencia de relaciones entre las variables independientes. La existencia de relación entre ellas se define como Multicolinealidad. Que exista multicolinealidad significa que varias variables independientes están aportando una información similar, y que no es posible conocer el influjo separado de cada una de ellas en la ecuación de regresión. Uno de los criterios utilizados para detectar este fenómeno es el concepto de Tolerancia estadística, definida como uno menos el Cuadrado del Coeficiente de Determinación Múltiple;  $1 - R_1^2$ . Si una variable tiene una tolerancia pequeña significa que es una combinación lineal de las otras variables independientes (Norusis, 1990a: 257). No obstante, cuando hay que introducir una nueva variable independiente en el análisis de Regresión, la mayoría de los programas estadísticos analizan por defecto si puede existir relación entre ésta y las ya existentes en el modelo.

Antes de dar por concluido este apartado, señalar que la gran dificultad en el cumplimiento de estos requisitos, unido a la escasa utilización de variables de intervalo en investigación social reducen considerablemente las posibilidades de utilización de esta técnica<sup>7</sup>. Debido a esta limitación, a mediados de la década de los 60 comenzaron a surgir modelos de regresión con variables dependientes cualitativas y que son conocidos como modelos de RESPUESTA CUALITATIVA. Amemiya (1981: 1485) destaca los factores que han influido decisivamente en el desarrollo de estos modelos:

- El gran número de científicos (Economistas, Biólogos, Sociólogos, etc.) que trabajan con este tipo de variables.

7. Aldrich-Nelson señalan que esta es una de las técnicas en la que más se abusa dentro de las Ciencias Sociales (1980: 9).

El incremento de las bases de datos en soportes magnéticos que pueden ser analizados con los potentes programas informáticos disponibles para ordenadores personales.

### 3. ALTERNATIVAS METODOLÓGICAS AL ANÁLISIS DE REGRESIÓN LINEAL

De todos los modelos de Respuesta Cualitativa este trabajo se centrará en la REGRESIÓN LOGÍSTICA, ya que según los expertos es el más utilizado en investigación sociológica (Zhang y Hoffman, 1993: 195)<sup>8</sup>. El propósito de la Regresión Logística es conocer la ocurrencia o no de un determinado fenómeno definiendo una variable dependiente que adoptará el valor 1 si el suceso ocurre, y el 0 si no ocurre. Con una única variable dependiente el modelo de regresión logística es definido con la expresión:

$$\text{Prob (ocurra)} = \frac{e^{B_0+B_1X}}{1 + e^{B_0+B_1X}}$$

o lo que es lo mismo,

$$\text{Prob (ocurra)} = \frac{1}{1 + e^{-(B_0+B_1X)}}$$

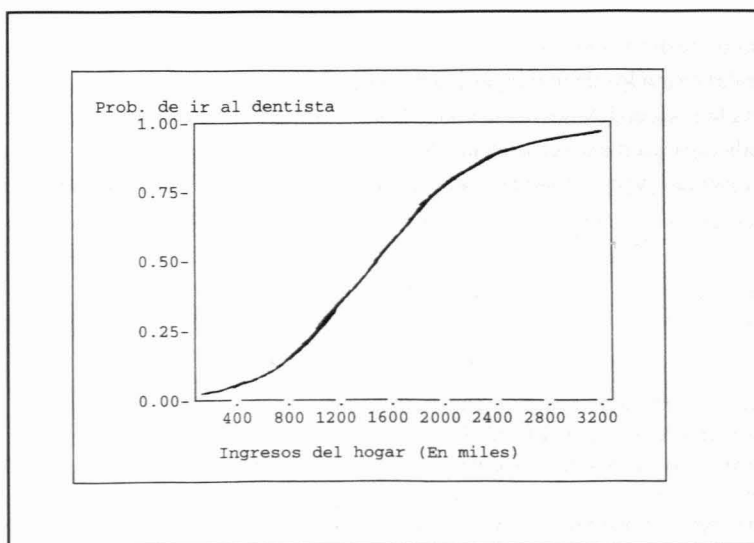
donde  $B_0$  y  $B_1$  son los coeficientes estimados,  $X$  la variable independiente y  $E$  es la base del logaritmo neperiano.

Como señala Mauro F. Guillén cuando analiza la forma simoidal que se obtiene al representar la función logística (gráfico 1), «las propiedades de ésta la configuran como un modelo consistente con un amplio grupo de supuestos sociológicos sobre la

8. Por otro lado, numerosos autores consideran que algunos de estos modelos proporcionan resultados similares, en especial el modelo Logit y el Probit (Anemiya, 1981: 178-179). M.F. Guillén señala que una de las mayores ventajas del modelo Logístico «es que puede utilizarse para variables dependientes con múltiples categorías, mientras que el modelo probit presenta problemas con variables que tengan más de cuatro categorías» (1992: 66).

probabilidad de que ciertos sucesos ocurran» (1992: 70-71); ya que refleja con gran exactitud ciertos fenómenos sociales que se caracterizan por una relación no lineal entre dos (o más) variables en la que el aumento de una supone en un primer momento un gran aumento en la segunda, para después ralentizar su crecimiento. Este autor pone el ejemplo de la relación entre la probabilidad de ir al dentista y el nivel de ingresos: con bajos niveles de renta cabe esperar que un aumento de 10.000 pts apenas aumentará la probabilidad de visitar al dentista, mientras que en personas con ingresos medios la probabilidad de acudir al dentista al obtener este mismo «sobreingreso será» mayor. Por otro lado, en las familias de rentas muy elevadas la probabilidad de acudir al dentista al aumentar el salario en 10.000 pts será prácticamente nula (Guillén, 1992: 70-71). Figura que es válida también para explicar la relación entre determinados fenómenos comerciales como las ventas de un nuevo producto a lo largo de un período de tiempo: ventas que al principio crecen con gran rapidez hasta que alcanzan el nivel de saturación del mercado en el cual las ventas tienden a estabilizarse (Mahajan, 1985: 8-10, Grande, 1992: 81-82).

### Gráfico I. Representación de la Función Logística



Fuente: Mauro F. Guillén, 1992: 71

Por otro lado, y haciendo referencia a la métrica de las variables, señalar que esta técnica aún siendo mucho mucho más compleja que la Regresión Lineal es aplicable en un número mayor de situaciones puesto que únicamente requiere –como condición indispensable para ser utilizado– que la variable dependiente sea cualitativa, y carece de restricciones sobre la forma de medición del resto de variables. Del mismo modo, los supuestos en los que se sustenta son menos exigentes que el modelo de regresión lineal, como puede verse en Aldrich-Nelson, 1980: 48-49.

No obstante, señalar que este trabajo no trata de exponer pormenorizadamente los sofisticados procesos estadístico-matemáticos en los que se fundamenta el modelo de regresión logística, sino que nuestro interés se centra en la explicación de los distintos procedimientos utilizados en la construcción del modelo de regresión, comprobar en que medida éste se ajusta a los datos, y realizar una interpretación del mismo. En definitiva, este artículo pretende «acercar» los criterios de utilización de la regresión logística a aquellas personas interesadas en la investigación social. Para los lectores interesados en una mayor profundización teórica de los aspectos técnicos en los que se fundamenta esta técnica aconsejamos la consulta de los textos de Brown (1980) y Jobson (1992).

### **3.1. Definición operacional del problema a investigar**

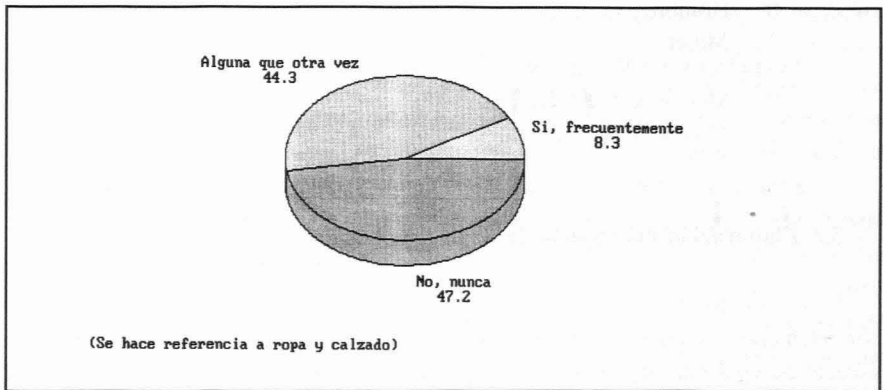
A fin de realizar una exposición más fluida realizaremos la explicación utilizando datos reales extraídos de una investigación sociológica sobre hábitos de consumo realizada en la sociedad Navarra en Junio de 1993, cuyas características y ficha técnica del estudio aparecen explicadas a pié de página<sup>9</sup>. De toda la información suministrada por este estudio nos centraremos únicamente en el análisis de cuales son los rasgos

9. El tamaño muestral estadísticamente requerido para estudiar de forma representativa la población navarra entre 16 y 65 años de edad se ha establecido en 600 entrevistas, que en el supuesto 50/50 tiene un error muestral máximo para datos globales de  $\pm 4\%$  a un nivel de confianza del 95%. El tipo de muestreo utilizado es Aleatorio Estratificado por afijación proporcional en virtud de la variable Hábitat o tamaño del municipio y Zona Geográfica y corregido por cuotas de sexo y edad. El trabajo de campo fue realizado entre Mayo y Junio de 1993 mediante encuesta personal en el hogar de cada entrevistado.

sociodemográficos que inciden significativamente en comprar productos que no son utilizados, utilizando para ello como término dependiente (Y) la pregunta expuesta en el gráfico II<sup>10</sup>.

## Gráfico II

### ¿Suele comprar cosas que después no utiliza?



Puesto que la opción «Sí frecuentemente» tiene muy pocos individuos (50 personas, un 8,3% de la muestra) y que el objeto de estudio es analizar las variables que influyen en los individuos que compran cosas que no utilizan –independientemente de la frecuencia que lo hagan– se procederá a agruparla con la categoría «Alguna que otra vez», de modo que la variable dependiente va a tener dos opciones: los que alguna vez compran cosas que no utilizan (categorizada con el dígito 1), y los que nunca lo hacen (valor 0).

Por otro lado, las preguntas mostradas en el gráfico III serán utilizadas como términos independientes a fin de conocer qué rasgos sociodemográficos indiquen de modo notable en comprar productos que no son utilizados. Como puede comprobarse, éstas han sido categorizadas en dos valores: 1 y 0.

10. Debe quedar claro que el propósito de utilización de este ejemplo es puramente pedagógico a fin de hacer más fluida esta exposición, no pretendiéndose con ello realizar grandes interpretaciones sociológicas.

### Gráfico III

#### Exposición de las variables independientes

Nombre	Definición	Frecuencias
<b>EDAD</b>		
0	Entre 16 y 35 años	316
1	Entre 36 y 65 años	284
<b>SEXO</b>		
0	Hombre	306
1	Mujer	294
<b>TAMAÑO MUNICIPIO DE RESIDENCIA</b>		
0	Menos de 5.000 habitantes	245
1	Más de 5.001 habitantes	355

### 3.2. Elaboración del modelo de Regresión Logística

Una vez definidas las variables a utilizar, el paso siguiente consiste en especificar correctamente el modelo de regresión, puesto que si el diseño de la investigación es un aspecto de vital importancia cuando hay que realizar cualquier investigación, con el método de la Regresión la elaboración de éste adquiere una importancia fundamental, tal y como ha sido expuesto anteriormente en el apartado 2.

Recordemos que el modelo de Regresión Lineal trata de ajustar, utilizando el procedimiento de los mínimos cuadrados, una nube de puntos a una recta. En la regresión logística, al utilizar variables dependientes cualitativas, los estimadores por mínimos cuadrados no cumplen las condiciones de insesgadez y consistencia (Gracia Díez, 1988: 11) ya que el método pierde sus propiedades cuando se aplica a este tipo de variables. De las distintas alternativas metodológicas para solucionar este problema, P.J. Dhymes (1986: 1575) considera que el procedimiento de estimación más adecuado para el modelo Logit es el de Máxima Verosimilitud, que proporciona estimadores consistentes e insesgados<sup>11</sup>. La probabilidad que los parámetros estimados se ajusten a los resultados observados es conocida como Verosimilitud, y debido a que suele adoptar un valor menor que la unidad normalmente se utiliza su logaritmo multiplicado por -2 (-2LL); que sigue una distribución Ji-Cuadrado con grados de libertad igual al número de individuos menos el número de parámetros incluidos en el modelo (N-p).

11. Una explicación pormenorizada de éste método y de sus condiciones de aplicación puede encontrarse en el trabajo de Xie-Manski (1989: 283-302) y Aldrich-Nelson (1984: 14-35).

Así, la correcta especificación del modelo significa que no se han excluido variables relevantes en el modelo y que no han sido incluidas variables irrelevantes en el mismo. De este modo, el planteamiento teórico utilizado indicaba que -a grandes rasgos- los colectivos caracterizados por comprar cosas que después no serán utilizadas son fundamentalmente mujeres (1984) y personas jóvenes (A. del Pino 1990: 32-38, R. Lopez Pintor 1975: 86). Esta caracterización nos sugirió analizar la influencia de otro término; el tamaño de municipio en el que vive el entrevistado: La idea subyacente es comprobar hasta que punto el comprar cosas que no son utilizadas tiene relación con vivir en un hábitat rural y urbano, puesto que cada uno de éstos tiene diferentes equipamientos comerciales (Banesto 1992, Dirección General de Comercio Interior 1987). Una vez formulados los objetivos, y a fin de comprobar la importancia de esta última variable en la configuración del modelo, utilizamos una Regresión por etapas («Stepwise») puesto que nuestro interés es seleccionar paulatinamente -de todas las variables independientes- aquellas que mejor explican la variable dependiente.

No obstante, mientras que la Regresión Lineal selecciona las variables utilizando el test de la F -ya que parte del hecho que los errores están distribuidos normalmente (Bisquerra, 1989: 215)- la Regresión Logística utiliza el test proporcional de verosimilitud (Likelihood-Ratio test), que sigue una distribución Ji-Cuadrado con grados de libertad igual a la diferencia entre los grados de libertad del modelo completo y los del modelo formado únicamente por la constante (Demaris, 1992: 46).

Una vez construido el modelo, y antes de comenzar con la interpretación de los coeficientes del mismo, debe comprobarse si está correctamente especificado, es decir, si los parámetros estimados se adaptan correctamente a los datos. Al igual que en el modelo lineal, en la regresión logística existen un conjunto de tests que analizan las diferencias entre los valores observados y los estimados por el modelo.

Un análisis de éstos pueden realizarse observando el cuadro 1, que muestra el ajuste del modelo completo, es decir, con todas las variables independientes. La primera fila (Modelo Ji-Cuadrado) muestra el valor del test proporcional de verosimilitud, que se calcula restando el -2 LL del modelo formado únicamente por la constante al valor de -2 LL con todas las variables independientes: 790.2969 - 741.944 (Aldrich-Nelson, 1984: 55). Los bajos niveles de significación inducen a rechazar la hipótesis nula de que los coeficientes de estas dos variables (Sexo y Edad) son iguales a 0; o dicho de otro modo, muestran que el modelo compuesto por dos variables independientes y la constante es mejor que el modelo que contiene únicamente la constante. Dicho de otro modo, es similar al valor de la F en la regresión lineal ya que viene a ser la resta de la Suma de Cuadrados Total menos la Suma de Cuadrados Residual (Demaris, 1992: 53).

## Cuadro 1. Ajuste del modelo

	Ji-Cuadrado	gl	Significación
Modelo Ji-Cuadrado	48.353	2	.0000
Mejora	19.497	1	.0000

En segunda línea se expone el cambio del -2LL cuando la última variable (SEXO) ha sido incluida en el modelo (761.441 - 741.944). Su bajo nivel de significación indica que la selección de esta última variable mejora notablemente el ajuste del modelo de regresión. Las ventajas y condiciones de la utilización de otros test como la Pseudo R<sup>2</sup> de Mcfadden, R<sup>2</sup> de Cragg-Uhler, etc. pueden encontrarse en Millán-Ruiz (1987: 118) y Aldrich-Nelson (1984: 25-29).

Además de estos tests, en la regresión Logística es frecuente utilizar métodos gráficos con objeto de comprobar la bondad del ajuste del modelo. De todos los métodos existentes analizaremos uno de los más utilizados y que se fundamenta en una comparación –mediante una tabla de doble entrada– entre los valores observados y los predichos por la ecuación.

El proceso comienza calculando la probabilidad de que cada individuo compre cosas que no utiliza, analizando la puntuación de este individuo en los coeficientes de la ecuación. Si el valor obtenido es mayor que 0.5 este individuo queda clasificado en la «celdilla» 1; si por el contrario este resultado es menor que 0.5 se incluye en la casilla de la izquierda, bajo el valor 0. Posteriormente se comparan estos valores predichos con los observados, dando lugar a una tabla de doble entrada en la que se muestran las diferencias entre las predicciones realizadas por el modelo y los valores observados (Jobson, 1992: 294-297).

## Cuadro 2. Tabla de aciertos/errores

Valores Observados		Val. Predichos		Porcentaje
0	1	0	1	
.00	0	157	115	57.74%
1.00	1	45	245	84.48%
Total				71.53%



Como se observa en el cuadro 2, 157 personas que nunca compran cosas que no usan (utilizan todo lo que compran) están correctamente «predichos por el modelo». Por otro lado, 245 entrevistados que «frecuentemente y alguna que otra vez» compran cosas que después no utilizan están correctamente predichas por esta ecuación de regresión. El resto de «celdillas» muestran los valores no correctamente especificados: hay 115 personas que utilizan todo lo que compran y el modelo considera que no es así, y 45 que adquieren objetos que no llegan a utilizar. La tasa de clasificación correcta, dentro del subgrupo de los entrevistados que utilizan todo lo que compran es del 57.72% ( $157/272 * 100$ ). Este porcentaje alcanza el 84.48% cuando son analizados aquellos que compran elementos que no llegan a utilizar. En conjunto, un 71.53% de la población total está correctamente clasificado por el modelo.

No obstante, y pese a las ventajas mostradas por este método, no es adecuado cuando se estudian fenómenos con alta probabilidad de ocurrencia, ya que los resultados no son muy fiables (Millán-Ruiz, 1987: 118). En esta línea, Millán y Ruiz aconsejan utilizar conjuntamente todos los diversos métodos disponibles a fin de conocer con exactitud el ajuste del modelo construido.

### 3.3 Interpretación de los coeficientes de la ecuación

En el cuadro 3 se exponen los resultados de los parámetros estimados para el modelo de regresión logística: La segunda columna muestra los coeficientes de la ecuación, la tercera el error estándar y la cuarta el test de Wald. Los coeficientes Beta que aparecen en la segunda columna cuantifican la incidencia que cada variable independiente tiene en la ocurrencia del fenómeno. El signo de estos coeficientes (positivo ó negativo) señalan el incremento (o disminución) en la relación logarítmica de probabilidad de que ocurra el fenómeno analizado ante un aumento (o disminución) unitario en la variable independiente (Hosmer y Lemeshow, 1989: 38). Así, un valor positivo de B indica que esta variable contribuye positivamente a la ocurrencia del suceso, mientras que valores negativos de B inciden negativamente en la percepción subjetiva sobre la satisfacción con la situación económica de su hogar. No obstante, Guillén aconseja precaución en la interpretación de estos coeficientes; ya que «cuanto mayor es el valor absoluto de Beta más importante es el efecto, pero evaluar su importancia proporcional no es tan sencilla como en la regresión lineal» (1992: 75)<sup>12</sup> puesto que como ha quedado reflejado en el Gráfico 1, el efecto es menos importante para valores extremos –curva con poca pendiente– que para valores medios.

12. Véase también Aldrich-Nelson, 1990, p. 43.

**Cuadro 3.**  
**Coefficientes de las variables independientes**  
**tomando como variable dependiente**  
**¿Suele comprar cosas que después no llega a utilizar?**

Variable	B	S.E.	Wald	gl	Sig	R
SEXO	.7674	.1756	19.0924	1	.0000	.1471
EDAD	.9366	.1756	28.4443	1	.0000	.1829
Constante	-.7262	.1507	23.2349	1	.0000	

La cuarta columna muestra los resultados del test de Wald para cada término, que en variables con dos categorías se calcula elevando al cuadrado el cociente entre el valor de los coeficientes y el error estándar. Bajo el título GL se encuentran los grados de libertad (número de categorías menos 1) y a su derecha la significación del test de Wald. La última columna muestra la correlación parcial de cada variable con el término dependiente: pequeños valores de R indican que la contribución de esa variable es relativamente pequeña. Por otro lado, valores positivos indican que a medida que la variable aumenta de valor hay más probabilidad de comprar cosas que no sean utilizadas.

Una vez explicado el origen de cada uno de los coeficientes se procederá a la interpretación de los resultados considerando que es la edad, mucho más que el sexo, la variable que incide en adquirir productos que después no serán utilizados:

La interpretación de este fenómeno debe hacerse analizando la evolución de la sociedad de consumo en España (Mínguez: 1969), teniendo en cuenta que autores como F. Andrés Orizo consideran que es a partir de la década de los setenta cuando la sociedad española adquiere niveles de desarrollo económico y pautas de consumo similares a las de los países más avanzados; de modo que es a partir de esta fecha cuando puede considerarse a España como una sociedad de consumo masivo:

*«Ante la pregunta final de si en España se ha alcanzado o no la cota del consumo de masas, nuestros datos nos indican que estamos en camino para unos cuantos bienes y servicios básicos, aunque interrumpido por la crisis económica de los tres últimos años» (Andrés Orizo, 1977: 148).*

La relativa cercanía temporal de este hecho (hace tan sólo dos décadas), sobre todo comparado con otros ámbitos geográficos como Estados Unidos y Europa, nos lleva a

plantear el hecho que **las generaciones jóvenes son las primeras que se han educado y socializado en una sociedad de consumo masivo**, frente a las generaciones anteriores que han sido socializadas en el espíritu del ahorro y del sacrificio de la España franquista.

Por otro lado, hay que recordar que hasta hace unos años correspondía exclusivamente a la familia la función de socializar a sus hijos, y nada «escapaba» del control de la unidad familiar. Frente a esto, las nuevas generaciones son educadas dentro de un sistema institucional complejo no sólo por la temprana edad a la que el niño acude a guarderías y a escuelas (frente al niño tradicional que permanecía «encerrado» en casa hasta tener la edad de ir a la escuela) sino por la gran **influencia** que ejercen sobre ellos los **patrones de comportamiento difundidos por la televisión**. La televisión, a través de la publicidad y de los modos de vida que refleja, ha influido tremendamente en las actitudes de los jóvenes (y de los adultos) hacia el consumo. Así, estas nuevas generaciones han vivido en hogares llenos de electrodomésticos y no se plantean (quizás porque no lo conocen) como sería su vida sin el televisor, sin el teléfono, etc. Los regalos de los padres han sido frecuentes y han disfrutado desde muy jóvenes de un cierto «dinerillo» para gastar en «sus cosas». En este sentido, Ronald Inglehart considera que el factor que subyace a esta influencia de la edad es el impacto de su socialización primaria:

*«...si diferentes generaciones de la misma sociedad manifiestan reacciones significativamente diferentes ante los mismos estímulos, ello implicaría la presencia de diferencias en el aprendizaje cultural»* (Inglehart, 1991: 9).

El análisis del segundo término del modelo muestra que el hecho de ser mujer tiene una gran relación con la compra de productos que no son utilizados, aspecto que obliga a recordar investigaciones realizadas en los últimos años en las que se manifiesta que las diferencias entre los hombres y las mujeres españolas se han reducido notablemente. No obstante, y aunque en términos generales estas diferencias están desapareciendo<sup>13</sup>, en el ámbito específico del consumo –y a tenor de los resultados hasta ahora expuestos– esta tendencia general se está produciendo muy despacio, como manifiesta José Castillo (1991: 58-59) cuando señala que aunque el hombre se está incorporando a tareas caseras, esta incorporación se lleva a cabo muy lentamente, y sobretodo se realiza no en tareas rutinarias sino cuando se cumplen determinadas

13. Son muchos los estudios que han verificado esta tendencia, como son las investigaciones de Andrés Orizo en 1992 (p. 249) y 1979, pp. 239-240.

circunstancias que dan a esas tareas un «rango distinto a la propia cotidianidad». Según afirma este autor: «la rutinaria e innoble tarea de la mujer es enaltecida al rango de ocupación seria y digna por medio del concurso del varón» (1991: 59).

Antes de finalizar señalar la escasa importancia del tamaño del municipio en el que se reside, explicación que habrá que buscarla en la mejora de las comunicaciones, la alta tasa de posesión del vehículo privado y la acción homogeneizadora de los medios de comunicación de masas, que si bien no han eliminado las diferencias existentes entre las zonas rurales y las urbanas, sí han homogeneizado la población al modificar los hábitos de las personas que residen en las zonas rurales.

#### 4. CONCLUSIÓN

La técnica estadística de la regresión es utilizada en investigación social en áreas tan diversas como el análisis de las variables que inciden en la abstención electoral (Font 1992) y en el tipo de voto (Linz y otros, 1976, 35-73), los factores que intervienen en el abandono de las enseñanzas superiores (Davies, 1994), señalar los elementos que inciden significativamente en las respuestas de las encuestas por correo (Heberlein-Baumgartner, 1978: 454), analizar la predicción del comportamiento de compra de los consumidores en base a ciertas variables y así anticipar las ventas de un determinado producto (Sanz de la Tajada, 1990: 424), determinar la asociación existente entre la calidad global de cada producto y el precio (Cruz Roche y Múgica Grijalba, 1993: 25-35), enumerar qué factores influyen de modo significativo en los efectos de las promociones de precios (Fernández Nogales, 1993: 36-44), analizar los elementos que inciden en la compra de un microondas, (Oropesa, 1993), etc.

No obstante, el objetivo perseguido por este trabajo ha sido realizar una exposición sobre las condiciones que deben tenerse en cuenta cuando la técnica de la regresión lineal es aplicada a una matriz de datos; así como exponer una de las alternativas que pueden utilizarse cuando no es posible la aplicación correcta de la misma. El punto de arranque es la argumentación, ya señalada por numerosos autores, que si bien el uso de los ordenadores ha permitido un enorme desarrollo de determinadas técnicas estadísticas, debe tenerse en cuenta que puede generar también determinados problemas debido a que el ordenador proporciona siempre un resultado; incluso si se ha realizado una mala elección de la codificación de los datos o una incorrecta elección del método; condición que puede dar lugar a empleos abusivos. Por otro lado, la facilidad de empleo puede hacer olvidar las reglas de interpretación (Clapier, 1983: 63).

Así, la simplificación en la realización de los cálculos producida por la enorme proliferación de programas estadísticos ha traído consigo —en numerosas ocasiones un

uso y abuso de ciertas técnicas estadísticas, eludiendo realizar análisis adecuados de las condiciones que deben cumplir los datos para poder aplicar en cada momento la técnica estadística correspondiente.

Una vez comprobados los supuestos que debe cumplir esta técnica para ser aplicada correctamente, y aunque la alternativa metodológica expuesta en este trabajo ha sido la utilización de la Regresión Logística –siguiendo un modelo similar al propuesto por Aldrich-Nelson (1980: 31-35)– deseamos exponer otras alternativas propuestas por diversos autores: Así, y según Manuel García Ferrando, en *Investigación Social* es muy difícil que los datos cumplan con exactitud todos los supuestos anteriormente descritos, y expone una serie de recursos metodológicos a emplear en estas circunstancias. En referencia a «la ausencia de relación entre las variables», «la interpretación que suele hacer el investigador social de este requisito es la existencia de una «intercorrelación baja», requisito más fácil de cumplir realmente por parte de las variables sociológicas» (García Ferrando, 1985: 397).

Por otro lado, numerosos autores indican que raramente se cumplen todos los supuestos, aunque esto nunca debe ser una justificación para ignorarlos. Mediante un cuidadoso análisis de los residuos y utilizando transformaciones u otros métodos de análisis –en caso que sea necesario– es posible corregir ciertas tendencias de los datos y verificar de este modo estos supuestos (Bowerman, 1990: 631-728). En esta línea, Etxeberria considera que si el incumplimiento de los supuestos no es muy significativo las conclusiones pueden seguir siendo acertadas debido a la robustez del método» (Etxeberria, 1990: 82).

## REFERENCIAS BIBLIOGRÁFICAS

- AMEMIYA, T. (1981): «Qualitative Response Models: A Survey», en *Journal of Economic Literature*, Vol. XIX, pp. 1483-1536.
- ANDRÉS ORIZO, F.: *Los nuevos valores de los españoles*, S.M., Madrid 1992.
- ANDRÉS ORIZO, F.: *Cambio sociocultural y comportamiento económico*, Centro de Investigaciones Sociológicas, Madrid 1979.
- ANDRÉS ORIZO, F.: (1977). *Las bases sociales del consumo y del ahorro en España*. Madrid: Confederación Española de Cajas de Ahorro.
- ANDREWS, F.M. y otros (1981): *A guide for selecting Statistical Techniques for analyzing Social Science Data*, Survey Research Center, University of Michigan.
- ALDRICH, J.J.; NELSON, F.D. (1984): *Linear Probability, Logit and Probit Models*. Londres: Sage University Paper.
- ALVIRA MARÍN, F.: «Introducción al análisis de datos», en M. García Ferrando y otros (1985). *El análisis de la realidad social*, Madrid: Alianza Universidad, pp. 325-358.
- BABUSIAUX, C.; LANTZ, F.; RODRÍGUEZ-CALAZA, J.J. (1987): «Regresión en presencia de colinealidad», en *Revista Española de Economía*, Vol. 4, nº1, pp. 11-42.
- BANESTO, (1992): *Anuario de Mercado Español 1992*, Apéndices estadísticos sobre los municipios españoles mayores de 1.000 habitantes, Madrid.
- BARNETT, V.; LEWIS, T.: *Outliners in statistical data*, Wiley and Sons, New York 1984.
- BERRY, W.D.; FELDMAN, S. (1984): *Multiple Regression in Practique*. Londres: Sage University Paper.
- BISQUERRA, R. (1989): *Análisis Multivariable*, Barcelona: P.P.U.
- BOWMAN, A.W.; ROBINSON, D.R. (1990): *Introduction to Regression and Analysis of Variance*, Cambridge: IOP.
- BOWERMAN O'CONNEL, A: *Linear Statistical Models*, Pws-Kent Publishing Company, Boston 1990.
- BROWN, B.W. (1980): «Predictor Analyses for Binary Data», en Miller, R.G. Efron, B. Brown, y L.E. Moses *Biostatistics Casebooks*. New York: John Wiley and Sons.
- CALVO, F. (1990): *Estadística Aplicada*, Bilbao: Deusto.
- CALVO GÓMEZ, F. (1993): *Técnicas Estadísticas Multivariantes*, Bilbao: Univ. de Deusto.
- CASTILLO CASTILLO, J. (1991): «Consumo y Bienestar», en J. Vidal-Beneyto, *España a debate. Tomo II: La Sociedad*. Madrid: Tecnos.

- CRUZ ROCHE, I.; MÚGICA GRIJALBA, J.M. (1993): «La relación precio-calidad objetiva en los mercados de productos de consumo», en *Información Comercial Española*, N° 716, pp. 25-35.
- DAVIES, S. (1994): «In search of resistance and rebellion among high school dropouts», en *Canadian Journal of Sociology*, Vol. 19, N°3, pp. 331-350.
- DEMARIS, A. (1992): *Logit Modeling*, Londres: Sage University Paper.
- DIRECCIÓN GENERAL DE COMERCIO INTERIOR, (1987): «*El comprador español, hábitos de compra*», Madrid: Ministerio de Economía y Hacienda.
- DRHYMES, P.J. (1986): «Handbook of Econometrics», Vol. III, cap. 27, pp. 1567-1631, en Z. Griliches y M.D. (eds), *Intriligator*. Elsevier Science Publishers.
- DUNN, O.J.; CLARK, V.A. (1987): *Applied Statistics: Analysis of Variance and Regression*, New York: John Wiley & Sons.
- EDWARDS, A.L. (1984): *An introduction to Linear Regression and Correlation*, Nueva York: Freeman.
- ETXEBERRÍA, J. (1990): «Análisis de las condiciones de aplicación de la Regresión haciendo uso de los paquetes de programas», en *Revista de Investigación Educativa*, Vol. 8. N° 15 pp. 79-92.
- ETXEBARRIA, J.; JOARISTI, L.; LIZASOAIN, L. (1990): *Programación y análisis estadísticos básicos con SPSS/PC(+)*, Madrid: Paraninfo.
- FERNÁNDEZ NOGALES, A: «La utilización del precio como variable comercial por parte del minorista», en *Infor. Comercial Española*, N° 716, Abril 1993, pp. 36-44.
- FIENBERG, S. (1980): *The Analysis of Cross-Classified Categorical Data*, Cambridge, Massachusetts: The MIT Press.
- FREIXA I BLANXART, M. y otros, (1992): *Análisis Exploratorio de datos: Nuevas técnicas estadísticas*, Barcelona: PPU.
- FONT, J. (1992): «La abstención en las Grandes Ciudades: Madrid y Barcelona», en *Revista Española de Investigaciones Sociológicas*, N° 58, pp. 123-139
- GARCÍA FERRANDO, M. (1985): *Socioestadística*, Madrid: Alianza Universidad Textos.
- GLANTZ, S.A.: «Todo está en las cifras», en *Journal of the American college of Cargiology* (Edición española), Vol. 2, Núm. 5, septiembre-octubre 1993.
- GRACIA DIEZ, M. (1988): «Modelos con variable dependiente cualitativa y de variación limitada», en *Cuadernos económicos del I.C.E.*, Vol 39, n° 2, pp. 6-50.
- GUILLÉN, M.F. (1992): *Análisis de Regresión Múltiple*, Madrid: CIS, Cuadernos metodológicos N° 4.
- GUIKEY, D.K.; RINDFUSS, D.R. (1987): «Logistic Regression: Multivariate Life Tables», en *Sociological Methods Research*, Vol. 16, N° 2, Noviembre 1987, pp. 276-300.

- HANUSHEK, E.A.; JACKSON, J.E. (1977). *Statistical Method for Social Scientists*, Orlando, Florida: Academic Press.
- HAUCK, W.W.; DONNER, A.(1977). «Wald's test as applied to hypotheses in logit analysis» en *Journal of the American Statistics Associations*, Nº 72, pp. 851-853.
- HEBERLEIN, T.A.; BAUMGARTNER, R. «Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature», en *Amerian Sociological Review*, Nº 43, 1978, p. 447-462.
- HOSMER, D.W.; LEMESHOW, S. (1989), *Applied logistic regression*, New York: John Wiley and Sons.
- INGLEHART, R. (1991). *El Cambio Sociocultural en las Sociedades Industriales Avanzadas*. Madrid: CIS-Siglo XXI.
- JOBSON, J.D. *Applied Multivariate Data Analysis*, Volume II, Springer-Verlag, Nueva York 1991.
- KINNEAR, T.C.; TAYLOR, J.R. «Multivariate Methods in Marketing Research: A Futher Attempt at Classification», en *Journal of Marketing*, Vol. 34, Nº 3, octubre de 1971.
- LEWIS-BECK, M.S. (1980): *Applied Regression: an introduction*. Londres: Sage University Paper.
- LINZ, J.J.; GÓMEZ REINO, M.; ANDRÉS ORIZO, F. *Informe Sociológico sobre el Camcio Político en España*, Fundación FOESSA, Madrid 1976.
- LOPEZ PINTOR, R. (1975): *Los españoles de los años 70*, Madrid: Tecnos.
- MAHAJAN, V.; PETERSON, R.A. (1985). *Models for Innovarion Diffusion*. Londres: Sage University Paper.
- MILLÁN, J.A.; RUIZ, P. (1987) «Modelos Logit de adopción de innovaciones en invernaderos de Almería» en *Investigación Agraria*, Nº 2, Vol II, Diciembre, pp. 115-125.
- MÍNGUEZ, A. (1969). *España: ¿Una Sociedad de Consumo?*. Madrid: Guadiana de Publicaciones.
- NETER, J.; WASSERMAN, W.; WHITMORE, G.A. (1988). *Applied Statistics*, London: Allyn and Bacon.
- NORUSIS, M.S. (1990 a): *SPSS Base System. User's Guide*, Chicago: SPSS inc.
- NORUSIS, M.S. (1990 b): *SPSS Advanced Statistics. User's Guide*, Chicago: SPSS inc.
- DEL PINO MERINO, A. (1990): *Los nuevos consumidores españoles*, Bilbao: Deusto.
- OROPESA, R.S. «Female Labor Force Participation and Timesaving Household Technology: A Case Study of the Microwave from 1978 to 1989», en *Journal of Consumer Research*, Vol. 19, nº 4, marzo 1993.



- SAN MARTÍN, R.; PARDO, A. (1989): *Psicoestadística: Contrastes Paramétricos y No Paramétricos*, Madrid: Pirámide.
- SCHROEDER L.D.; SJOQUIST, D.L.; STEPHAN, P.L. (1990): *Understanding Regression Analysis: An introductory Guide*, Londres: Sage University Paper.
- SEBER, G.A.F.: *Linear Regression Analysis*, Willey and Sons, New York, 1977.
- SHETH, J.N.: «Multivariate Analysis for Marketing Research», en *Journal of Marketing*, Vol. 34, N° 1, enero de 1971.
- TABACHNICK, B.G.; FIDELL, L.S. (1989): *Using Multivariate Statistics*, New York: Harper & Publishers.
- XIE, X.; MANSKI, C.H. (1989): «The Logit Model and Response-Based Samples», en *Sociological Methods & Research*, Vol 17, N° 3, February, pp. 283-302.
- ZHANG, J.; HOFFMAN, S.D. (1993): «Discrete-Choice Logit Models», en *Sociological Methods & Research*, Vol 22, N° 2, November, pp. 193-213.