Variables dependientes cualitativas

Manuel Artís Ortuño
Montserrat Guillén Estany

PID_00160621

Universitat Oberta de Catalunya
www.uoc.edu

Índice

Introducción

Hasta el momento hemos estudiado el modelo de regresión, es decir, hemos supuesto la existencia de una relación unidireccional de causalidad entre una variable llamada *variable dependiente* y un conjunto de variables que explican su comportamiento. Pues bien, todo lo que hemos hecho hasta ahora ha sido trabajar con un supuesto implícito: siempre hemos considerado la variable dependiente siempre como una variable cuantitativa.

Ejemplos de modelos con variables dependientes cuantitativas

A lo largo de la asignatura hemos ido proponiendo modelos para estudiar variables como, por ejemplo, las ventas de una empresa, el consumo de las familias, los costes de las bajas por enfermedad del personal, etc. Fijémonos en que todas estas variables tienen en común el hecho de que toman valores numéricos en un rango de valores suficientemente amplio como para considerar que tienen un comportamiento similar al de una variable aleatoria con distribución normal.

Limitar los modelos al entorno mencionado es insuficiente, puesto que no cabe duda de que muy a menudo podemos estar interesados en analizar el comportamiento de una variable de naturaleza cualitativa. Un ejemplo típico de este tipo de situación es el caso de una empresa que esté interesada en conocer si un cliente, cuando se interesa por un producto, lo comprará o no. En este caso, la variable dependiente es una variable cualitativa que registra dos situaciones: si el cliente compra el producto o si no lo compra. No podemos decir, en cualquier caso, que la variable dependiente toma dos valores, ya que, de hecho, las variables cualitativas, y ésta en particular, se refieren a **atributos** o **categorías**, y se caracterizan precisamente porque no indican valores.

Ved las variables cualitativas en los subapartados 1.1 y 1.2 del módulo "Variables exógenas cualitativas".

Ejemplos de variables cualitativas

Hay muchos ejemplos de variables cualitativas, que pueden ser de dos tipos:

- Saber si una empresa invertirá o no en el extranjero es un ejemplo de **variable dicotómica** (variable cualitativa que indica dos categorías).
- Si tuviésemos la posibilidad de invertir en Francia, Italia o Portugal, podríamos decir que la variable que determina en qué país invertimos es una variable cualitativa de tres categorías. Hemos llamado a estos casos **variables politómicas**.

Como es habitual, codificaremos la opción de compra con un 1 y la de ausencia de compra, con un 0. En este ejemplo, las variables explicativas pueden ser tanto las características personales del comprador (sexo, edad, nivel de renta, estudios, etc.), como las características del producto (precio, atractivo, diseño, etc.). Del mismo modo, pueden influir otras variables relativas al entorno, como, por ejemplo, la existencia de productos parecidos ofrecidos por la competencia y el precio y las características de éstos.

Como vemos, las variables dependientes cualitativas aparecen en la práctica tan a menudo como las cuantitativas y, por lo tanto, vamos a dedicarles este módulo completo.

El uso de modelos para variables dependientes cualitativas abre un nuevo marco que supone una ampliación con respecto a lo que se ha practicado en los módulos anteriores, ampliando el estudio de modelos de regresión en este caso concreto. Teniendo en cuenta la dificultad de este nuevo grupo de modelos, analizaremos a fondo los modelos para variable dependiente dicotómica, ya que son los que se utilizan más y, a pesar de su dificultad, son el punto de referencia para modelos más complicados con la variable dependiente politómica, que tendríamos que estudiar más tarde.

Con el fin de conseguir los principales objetivos, empezaremos el módulo realizando una introducción a los **modelos de variable dependiente cualitativa**; a continuación, presentaremos el llamado **modelo de probabilidad lineal** y estudiaremos sus rasgos característicos más importantes; acto seguido, ampliaremos el análisis a los **modelos logit** y los **modelos probit**. Explicaremos cómo tenemos que contrastar diferentes hipótesis (en particular nos interesaremos por el contraste de significación individual de un parámetro y de significación global del modelo) y estudiaremos diferentes medidas para valorar la bondad del ajuste. Para acabar, veremos un caso práctico en el que desarrollaremos todo el contenido de este módulo y, por otra parte, indicaremos cómo se tratan situaciones en las que la variable dependiente cualitativa es politómica.

Objetivos

Tras haber seguido este módulo didáctico, el estudiante tiene que haber alcanzado los siguientes objetivos:

1. Detectar en qué situación tenemos que utilizar un modelo de variable dependiente cualitativa adecuado y por qué.

2. Utilizar el método de estimación adecuada para estimar un modelo de este tipo: tanto el modelo de probabilidad lineal como los modelos logit y probit.

3. Interpretar los resultados del modelo en el contexto de las variables cualitativas, identificando que se modeliza la probabilidad de la ocurrencia de un fenómeno y no el valor esperado de una variable condicionado a los valores observados en las variables explicativas.

4. Evaluar la bondad del ajuste en los modelos con variable dependiente cualitativa y su capacidad predictiva.

5. Saber entender los resultados de las estimaciones y utilizar los modelos en la práctica.

1. Modelos con variable dependiente cualitativa

> Los modelos de variable endógena cualitativa, conocidos genéricamente como *modelos de elección discreta*, permiten explicar las decisiones de un individuo a partir de un conjunto de variables explicativas que identifican sus características.

Daniel McFadden

Daniel McFadden nació en Raileigh (Carolina del Norte) en el año 1937. Se licenció en física por la Universidad de Minesota a los 19 años. Su interés por hacer un uso riguroso de las herramientas matemáticas y estadísticas en las ciencias sociales lo llevó a especializarse en economía. Sus contribuciones a la teoría econométrica incluyen los modelos de variable dependiente cualitativa.

Según cuál sea el número de decisiones alternativas entre las que puede elegir el individuo, podemos diferenciar dos tipos de **modelos de elección discreta**, que presentamos acto seguido: los modelos de elección binaria y los modelos de elección múltiple.

1) Modelos de elección binaria

> Los modelos de elección binaria son aquellos en los que la variable endógena cualitativa es dicotómica: podemos decir que la variable endógena únicamente toma dos categorías posibles.

En consecuencia, en los modelos de elección binaria se supone que los individuos sólo pueden elegir entre dos opciones alternativas, es decir, entre dos éxitos mutuamente excluyentes. Como veremos, en los modelos de elección múltiple la variable dependiente es politómica y, por lo tanto, presenta más de dos opciones.

Para ver un ejemplo de variable dependiente dicotómica, supongamos que nos interesa modelizar la decisión de una empresa acerca de la adquisición de una nave industrial en una nueva zona de un polígono muy bien situado, que le es atractiva desde un punto de vista logístico, ya que es muy interesante desde un punto de vista estratégico para distribuir sus productos. Esta empresa se encuentra ante dos alternativas excluyentes entre sí: comprar la nave o no comprarla. Así pues, la variable dependiente sólo registra dos alternativas: la empresa decide comprar la nave o decide no comprarla.

Como es habitual cuando se utilizan variables cualitativas, a menudo tenemos que codificar con valores las diferentes posibilidades. Tal y como solemos hacer en general para cualquier variable cualitativa, codificaremos, aquí, con un valor igual a 1 la decisión de comprar y con un valor igual a 0, la de no comprar. La única diferencia en estos momentos es que la variable que acabamos de mencionar será objeto de una modelización y, por lo tanto, es la variable dependiente de un modelo. Definiremos la variable dependiente de la manera siguiente:

$$Y_i = \begin{cases} 1 \text{ si se compra la nave,} \\ 0 \text{ si no se compra la nave,} \end{cases}$$

donde Y_i hace referencia a la decisión que toma la empresa *i*-ésima.

2) Modelos de elección múltiple

Los modelos de elección múltiple son aquellos en los que la variable dependiente cualitativa puede hacer referencia a más de dos categorías.

En este caso, el individuo puede elegir entre más de dos sucesos, es decir, tiene más de dos opciones que, de la misma manera que en el caso de los modelos de elección binaria, también son mutuamente excluyentes.

Para ver ahora un ejemplo de variable dependiente de elección múltiple y siguiendo con el caso anterior, supongamos que en el mercado inmobiliario hay tres naves industriales en venta, todas de características similares a las que requiere la empresa. En tal caso, llamaremos A, B y C a las diferentes opciones de nave al alcance y consideraremos, también, que la empresa pueda no comprar ninguna de ellas. En este caso, la empresa se enfrenta a cuatro alternativas: comprar la nave A, la B, la C o no comprar ninguna. Como consecuencia, la variable endógena puede codificarse de manera que tome cuatro valores indicativos de cada una de las decisiones posibles:

$$Y_i^c = \begin{cases} 0 \text{ si se compra la nave A,} \\ 1 \text{ si se compra la nave B,} \\ 2 \text{ si se compra la nave C,} \\ 3 \text{ si no se compra ninguna nave.} \end{cases}$$

Denotamos la variable endógena con el nombre Y_i^c, para distinguirla del caso anterior.

A continuación, estudiaremos los modelos para variables como la del primer tipo. Es decir, supondremos que tenemos información sobre *N* empresas que han decidido acerca de la compra de la nave. Para cada empresa, además de conocer su decisión final, conocemos algunas características individuales que utilizaremos como explicaciones de la decisión tomada. De esta manera, obtendremos un modelo que nos servirá para estudiar la relación entre las variables explicativas y la alternativa elegida por la empresa.

1.1. El modelo de probabilidad lineal

El modelo de probabilidad lineal (MPL) es el modelo más sencillo que podemos utilizar para modelizar una variable dependiente dicotómica, aunque, debido a su gran simplicidad, también poseerá numerosos inconvenientes.

De entrada, el modelo supone que los individuos se enfrentan a una elección con dos alternativas mutuamente excluyentes, de manera que el hecho de elegir una invalida la elección de la otra y que la elección entre una opción o la otra se puede explicar por una serie de variables explicativas, como variables que registran características personales de los individuos y/o variables que registran las condiciones de mercado. En definitiva, todas aquellas variables que determinan el comportamiento (la elección) del individuo.

Características personales de los individuos

Algunos ejemplos de variables explicativas que registran características personales de los individuos son el nivel de renta, el estado civil, la profesión, etc.

Ejemplo de modelo de probabilidad lineal

En el ejemplo de la empresa que se plantea comprar una nave industrial propuesto al inicio de este apartado, recordamos que las posibilidades de elección son las siguientes:

$$Y_i = \begin{cases} 1 \text{ si se compra la nave,} \\ 0 \text{ si no se compra la nave,} \end{cases}$$

donde Y_i se refiere a la decisión que toma la empresa *i*-ésima.

Para cada empresa, podemos considerar que puede influir en la decisión su situación financiera en el momento de la compra, sus perspectivas de expansión, el rendimiento que espera obtener del establecimiento de la nueva localización, el número de metros cuadrados disponibles, el tipo de producto que fabrica, etc.

El **modelo de probabilidad lineal** supone que la relación que hay entre las variables es de tipo lineal. Por lo tanto, podemos escribir el MPL de la siguiente forma:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, N, \tag{1.1}$$

donde $\mathbf{X}_i$ es el vector que registra los valores que toman las variables explicativas para el individuo *i*-ésimo (suponiendo un total de $k - 1$ variables explicativas y un término independiente) y $\mathbf{B}$ es el vector de parámetros, de dimensión k. Para cada individuo podemos escribir el modelo anterior de la siguiente manera:

$$Y_i = \mathbf{X}_i'\mathbf{B} + u_i, \quad i = 1, \ldots, N, \text{ donde } \mathbf{X}_i = \begin{bmatrix} 1 \\ X_{2i} \\ \vdots \\ X_{ki} \end{bmatrix} \text{ y } \mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

En este momento, se supone que Y_i puede tomar los valores 0 y 1, respectivamente, para cada individuo, y que estos valores se pueden explicar por una combinación lineal de regresores más un término de perturbación aleatorio.

Debido a que la variable dependiente sólo puede tomar dos valores posibles, si mantenemos la hipótesis sobre el término perturbación, que dice que se comporta según una distribución de probabilidad normal, tendremos una serie de inconvenientes de este modelo porque ambos comportamientos no son compatibles. Así, a pesar de su aparente simplicidad, en la práctica, el modelo no será demasiado utilizado a causa de sus características y todavía menos si utilizamos el método de estimación de los mínimos cuadrados ordinarios.

1.1.1. Especificación del modelo de probabilidad lineal

Si mantenemos el supuesto de que la esperanza de los términos de perturbación es cero $E[u_i] = 0$, y que las variables explicativas* son deterministas, la esperanza matemática de la variable dependiente es la parte sistemática del modelo. Para verlo, bastará con que tomemos esperanzas en la expresión 1.1, y obtenemos:

*** Las variables explicativas en este modelo son las características de los individuos.**

$$E[Y_i] = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki}, \quad i = 1, ..., N. \tag{1.2}$$

Por otra parte, sin embargo, teniendo en cuenta que la variable endógena es una variable cualitativa y, por lo tanto, discreta, su valor esperado lo da la suma de los valores que toma la variable multiplicados por la probabilidad asociada a cada uno de éstos. Es decir:

$$E[Y_i] = 0 \cdot P(Y_i = 0) + 1 \cdot P(Y_i = 1), \quad i = 1, ..., N,$$

o, simplificando,

$$E[Y_i] = P(Y_i = 1), \quad i = 1, ..., N. \tag{1.3}$$

Utilizaremos mucho esta última expresión a lo largo de este módulo, motivo por el que haremos uso de la notación P_i para indicar la probabilidad de que el individuo i-ésimo elija la opción 1, es decir, $P(Y_i = 1)$. Evidentemente, este último resultado es consecuencia de haber definido Y_i de manera que tome valores 1 y 0; así, la probabilidad de que valga cero, al estar multiplicada por cero, no aparece en la expresión. Si, en lugar de la codificación anterior, utilizásemos otra, el modelo también podría definirse, pero su interpretación sería bastante más complicada.

A partir de ahora denotaremos con P_i la expresión $P(Y_i = 1)$.

A partir de los resultados establecidos en 1.2 y 1.3, podemos escribir lo siguiente:

$$P_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki}, \quad i = 1, ..., N. \tag{1.4}$$

Esta última es la **especificación del modelo de probabilidad lineal**.

En esta expresión es donde se entiende por qué el modelo recibe este nombre. En primer lugar, podemos ver que el modelo describe la probabilidad de una elección de la empresa i-ésima*. La probabilidad de elección se pone en función de una combinación lineal de determinadas características de esta empresa. Podremos estimar los parámetros desconocidos mediante un procedimiento estadístico y así podremos utilizar este sencillo modelo econométrico.

*** Por ejemplo, la probabilidad de elección puede ser comprar la nave industrial.**

Un de los hechos más importantes es la falta del término aleatorio en la expresión 1.4. El hecho de entender que, aunque no esté, el modelo contiene un grado de incertidumbre, es un de los aspectos fundamentales de este módulo; y, para entenderlo, utilizamos el ejemplo de decisión de compra de una nave industrial. El modelo establece que la empresa comprará la nave industrial con una probabilidad que está determinada por una serie de factores (combinación lineal de características y parámetros). Notad que la decisión de comprar puede tener un cierto grado de aleatoriedad, ya que, por ejemplo, aunque el modelo establezca que la probabilidad de comprar la nave sea muy baja, finalmente puede comprarla.

Cuando hay aleatoriedad...

... en todo fenómeno como el que hemos descrito en este subapartado, aunque la probabilidad de ocurrencia sea baja, el acontecimiento puede tener lugar. Pensemos, por ejemplo, en un dado trucado en el que sale el seis con el 90% de probabilidad; a pesar de todo, puede ser que tiremos y no salga el seis.

Según la explicación del parágrafo anterior, la decisión de la empresa contiene un cierto grado de incertidumbre y el modelo sólo establece una dependencia entre la probabilidad de que efectúe la compra y una serie de variables explicativas (factores que la determinan).

En general, tendremos que interpretar que se modeliza la probabilidad de que el individuo elija una alternativa (una de las señaladas por la variable dependiente), dado su vector de características personales.

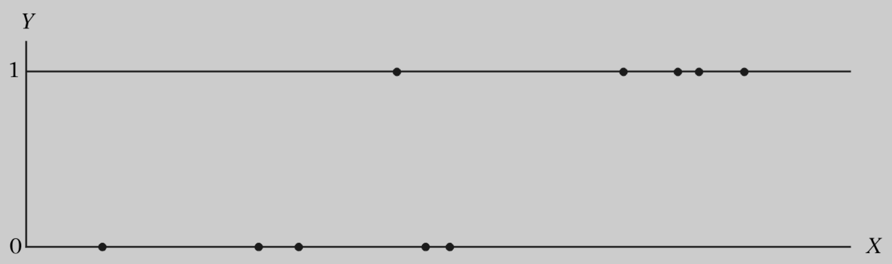
Así, en la expresión 1.4, el vector de parámetros (β_1, β_2, ..., β_k) indicará la variación esperada que tendrá la probabilidad de que el individuo elija la alternativa de interés ante variaciones unitarias en las características incluidas en las variables explicativas. Por lo tanto, el modelo de probabilidad lineal relaciona la probabilidad de ocurrencia de un determinado éxito (o de elección de una alternativa) con las características propias de cada individuo mediante una combinación lineal.

1.1.2. Propiedades del modelo de probabilidad lineal

El modelo de probabilidad lineal se especifica con mucha facilidad, pero tiene una serie de problemas que exponemos a continuación, que, en la práctica, nos conducirán a utilizar otros modelos:

1) Como ya se ha mencionado, la variable Y_i sólo puede tomar dos valores: 0 o 1. Por tanto, si se representan gráficamente los pares de observaciones de la variable mencionada con una de las variables explicativas, obtendremos todos los pares sobre las rectas $Y = 1$ y $Y = 0$.

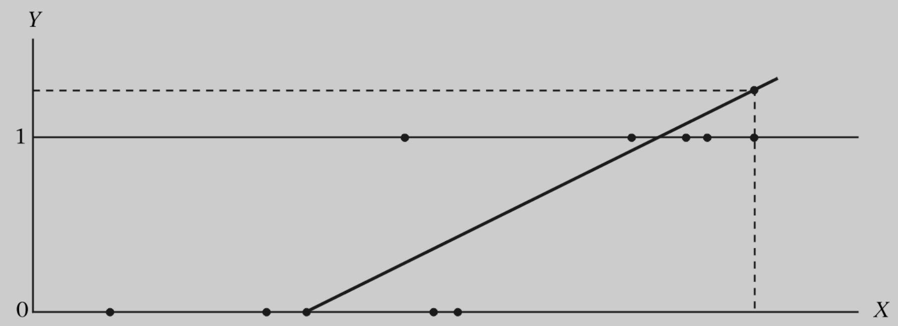
Pares de observaciones asociados a un modelo de probabilidad lineal



Cuando planteamos el modelo de probabilidad lineal, lo que estamos haciendo es ajustar una recta a la nube de puntos presentada en el gráfico anterior, de manera que, como podemos intuir a partir del gráfico de la página siguiente, los puntos se encontrarán lejos de la recta y, como ésta cortará las rectas $Y = 1$ y $Y = 0$, entonces algunos valores ajustados serán naturalmente diferentes de 0 y 1, y, como es lógico, en muchos casos se situarán fuera del intervalo [0,1]. En consecuencia, no siempre podremos interpretar los resultados de los valores ajustados como ajustes de la probabilidad.

Nota: recordemos que la probabilidad de un acontecimiento siempre se encuentra acotada por 0 y 1.

Ajuste a una recta del modelo de probabilidad lineal



El ajuste de una recta a la situación anterior producirá predicciones fuera del rango esperado para la probabilidad. Si, en lugar de considerar sólo una variable explicativa, consideramos más de una, el comportamiento sigue siendo el mismo, aunque no podemos dibujarlo en una gráfica de dos dimensiones.

Podéis ver las hipótesis sobre el término de perturbación del MRLM en el subapartado 2.2.2 del módulo "Modelo de regresión lineal múltiple...".

2) El término de perturbación u_i del modelo de probabilidad lineal no es una variable aleatoria continua, sino una variable aleatoria discreta, puesto que sólo puede tomar dos valores determinados. Por lo tanto, no podremos suponer que sigue una distribución normal tal y como sucedía en el modelo de regresión lineal múltiple (MRLM).

En concreto, u_i sólo puede tomar dos valores, respectivamente, en función de los valores de Y_i, y, por consiguiente, los tomará con la misma probabilidad que Y_i.

- Cuando Y_i vale 1, entonces, utilizando 1.1, obtendremos que:

$$u_i = 1 - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} = 1 - P_i, \quad i = 1, \ldots, N.$$

- Cuando Y_i vale 0, entonces, por analogía, obtendremos que:

$$u_i = -\beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} = -P_i, \quad i = 1, \ldots, N.$$

Como Y_i vale 1 con probabilidad P_i, y vale 0 con probabilidad $1 - P_i$, y además establecemos que $P_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$, entonces $u_i = 1 - P_i$ con probabilidad $\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} = P_i$ y $u_i = -P_i$ con probabilidad $1 - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} = 1 - P_i$, como se ve en la tabla, en la que resumimos toda la información:

Valores posibles en el modelo de probabilidad lineal		
y_i	$P(Y_i = y_i)$	u_i
1	P_i	$1 - P_i$
0	$1 - P_i$	$-P_i$

Interpretación de la tabla

En la tabla, Y_i representa la variable que nos sirve para construir un modelo a partir de los datos observados, representados por la variable y_i. Gracias a los resultados de la tabla sabemos que el término de perturbación se distribuye según una distribución binomial, al tomar dos valores posibles. Además, la distribución es diferente para cada individuo, ya que esta probabilidad depende de los valores observados en las variables explicativas, los cuales no tienen por qué ser los mismos para todos los individuos considerados en el modelo.

En consecuencia, la hipótesis de normalidad del término de perturbación no se mantiene, sino que la distribución de probabilidad que rige el comportamiento del término de perturbación es una binomial*. De todos modos, como el método de estimación de mínimos cuadrados ordinarios no supone ningún tipo de comportamiento con respecto al término de perturbación, sino que sólo se basa en el hecho de minimizar la suma de los cuadrados de los errores (la *SCR*), tanto la estimación de mínimos cuadrados de los parámetros del modelo (β_1, β_2, ..., β_k), como la varianza del término de perturbación, σ_u^2, y la predicción puntual de la variable endógena siguen siendo válidas. Sin embargo, por otra parte, dejan de serlo los contrastes de significación individual de los parámetros, así como el contraste de significación conjunta de los parámetros y, en general, los contrastes de restricciones lineales; es decir, queda invalidada cualquier inferencia y también la estimación por intervalo, teniendo en cuenta que los errores estándar estimados tendrán sesgo.

*** La distribución binomial, cuando hace referencia al caso dicotómico, se denomina *distribución de Bernouilli*.**

En resumidas cuentas, podemos estimar el modelo como si de un modelo de regresión lineal múltiple se tratase, poniendo en la variable dependiente los valores codificados con ceros y unos. De todos modos, los contrastes estadísticos que obtendríamos de manera automática mediante los procedimientos habituales no son correctos.

Homoscedasticidad

Recordemos que el término de perturbación es homoscedástico si tiene la misma varianza para todos los individuos considerados.

3) El término de perturbación del modelo de probabilidad lineal no cumple la hipótesis básica de homoscedasticidad. Y, para verlo, nos disponemos a calcular la esperanza matemática y la varianza del término de perturbación:

a) Empecemos, pues, por calcular la esperanza matemática del término de perturbación:

$$\begin{aligned} E[u_i] &= (1 - \beta_1 - \beta_2 X_{2i} - \dots - \beta_k X_{ki})P_i + \\ &+ (-\beta_1 - \beta_2 X_{2i} - \dots - \beta_k X_{ki})(1 - P_i), \quad i = 1, \dots, N. \end{aligned}$$

También llegamos a $E[u_i] = 0$...

... si lo vemos de esta otra manera:

$$E[u_i] = (1 - P_i)P_i - P_i(1 - P_i) = 0.$$

Efectuando las operaciones correspondientes, deducimos que:

$$E[u_i] = P_i - \beta_1 - \beta_2 X_{2i} - \dots - \beta_k X_{ki} = P_i - P_i = 0, \quad i = 1, \dots, N.$$

Varianza del término de perturbación

Recordemos que $\text{Var}[u_i] = E[(u_i - E[u_i])^2]$; en consecuencia, si $E[u_i] = 0$, entonces $\text{Var}[u_i] = = E[u_i^2]$.

Por lo tanto, sí que es correcta la hipótesis que establece que el término de perturbación tiene esperanza nula; no obstante, y como veremos acto seguido, su varianza no será constante (no será la misma para todos los individuos).

b) Calculamos la varianza de los términos de perturbación y omitimos cada vez la expresión i = 1, ..., N. A partir de la definición:

$$\text{Var}[u_i] = \text{E}[u_i^2] = (1 - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})^2 P_i + \\ + (-\beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})^2 (1 - P_i),$$

la última expresión es igual en $\text{Var}[u_i] = (1 - P_i)^2 P_i + (-P_i)^2 (1 - P_i)$. Para finalizar, efectuando una serie de operaciones, podemos demostrar que:

$$\text{Var}[u_i] = P_i(1 - P_i).$$

Debido a que el término de perturbación del modelo de probabilidad lineal es no homoscedástico, como podemos ver mediante el resultado que acabamos de obtener, la estimación de los modelos por mínimos cuadrados ordinarios (MCO) no es eficiente y, por lo tanto, si nos interesa garantizar esta propiedad, estaremos obligados a utilizar el método de los mínimos cuadrados generalizados (MCG).

Con el fin de obtener estimadores eficientes que además no tengan sesgo y sean consistentes, hay que utilizar el estimador de mínimos cuadrados generalizados:

$$\hat{\mathbf{B}}_{\text{MCG}} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{Y},$$

donde $\hat{\Omega}$ es una matriz diagonal con los elementos de la diagonal iguales a $P_i(1 - P_i)$. Para estimar estos elementos podemos aproximarlos por los ajustes obtenidos mediante la estimación por mínimos cuadrados ordinarios, es decir, si $\hat{\mathbf{B}}$ es la estimación MCO, entonces se puede estimar P_i con $\mathbf{X}_i'\hat{\mathbf{B}}$.

Otro método de estimación posible

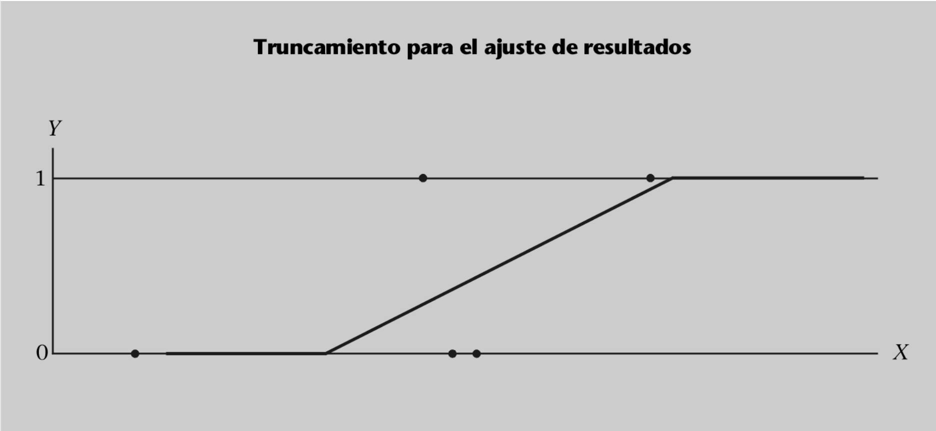
Recordemos que es equivalente hacer la estimación por mínimos cuadrados generalizados o bien transformar todas las observaciones de todas de las variables, incluyendo la variable dependiente y el término independiente, dividiéndolas por la raíz cuadrada de $P_i(1 - P_i)$.

A pesar de esto, es posible que el valor de P_i $(1 - P_i)$, al aproximarse por un ajuste, no sea positivo y, entonces, dejará de tener sentido su uso, ya que no correspondería a una varianza.

1.1.3. Limitaciones del modelo de probabilidad lineal y modelos alternativos

Con el método de estimación de los mínimos cuadrados generalizados anterior se soluciona el problema del heteroscedasticidad del modelo de probabilidad lineal, pero no podemos garantizar que los valores ajustados de la variable dependiente, después de efectuar la estimación, se encuentren dentro del intervalo [0,1].

Cuando los valores ajustados tras haber estimado el modelo no sean directamente interpretables como una probabilidad (es decir, cuando sean inferiores a 0 o superiores a 1), se acostumbra a efectuar el truncamiento de los valores, igualando a los valores extremos las predicciones que se encuentran fuera del intervalo, esto es, a 0 o a 1, respectivamente. De manera gráfica, la solución propuesta consiste en lo que mostramos a continuación:



Esta solución, de hecho, es una manipulación de los ajustes obtenidos, sin que puedan establecerse las propiedades y, por lo tanto, es más una solución cómoda que un método bien fundamentado.

Truncamiento para el ajuste de resultados

El modelo de probabilidad lineal proporciona el ajuste de una recta. El valor de la ordenada que corresponde a un valor concreto de la variable X representa la probabilidad de que se dé el acontecimiento representado por Y en la circunstancia descrita por X. Pero, dado que la probabilidad sólo toma valores entre 0 y 1, los valores de X que, por ajuste, se correspondan con valores de Y situados fuera del intervalo [0,1] se truncarán, de manera que pasarán a tener, bien el valor 0, bien el valor 1, y su interpretación tendrá sentido.

Uno de los principales problemas de fondo del modelo de probabilidad lineal se encuentra en su propia definición. Si suponemos que la probabilidad tiene un comportamiento lineal, entonces estamos suponiendo que la variación de la probabilidad como consecuencia de cambios en sus factores explicativos no depende de la situación inicial del individuo.

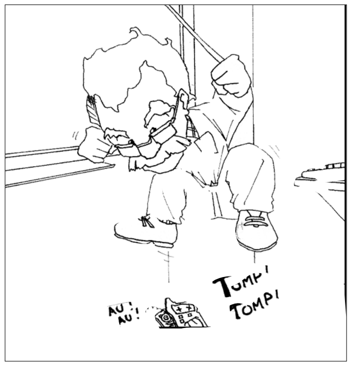
Supongamos que la empresa que quiere comprar una nave industrial efectúa la compra a partir del conocimiento de sus beneficios en el ejercicio anterior. Supongamos, también, un modelo muy sencillo que nos dice que la probabilidad de comprar la nave viene determinada por el siguiente modelo:

$$P_i = 0{,}3 + 0{,}01 benef_i,$$

donde $benef_i$ indica los beneficios del ejercicio anterior de la empresa i-ésima. El modelo establece que la probabilidad de que compre la nave se obtiene sumando a 0,3 el producto de los beneficios del ejercicio anterior por el coeficiente 0,01. Si los beneficios de una empresa son 10, entonces estimamos que la probabilidad de comprar la nave es del 40% (es decir, 0,3 + 0,01 · 10). Para una empresa con un beneficio de 50 unidades monetarias, la probabilidad de comprar la nave es del 80% (que se obtiene a partir del cálculo de 0,3 + 0,01 · 50). El significado del parámetro estimado que acompaña a la variable explicativa es que, para cada aumento de una unidad de los beneficios, la probabilidad de comprar la nave aumenta de 0,01 (es decir, del 1%), sea cual sea la situación de la empresa. Sin embargo, en la práctica se sabe que el efecto puede no ser el mismo para cualquier nivel de beneficios.

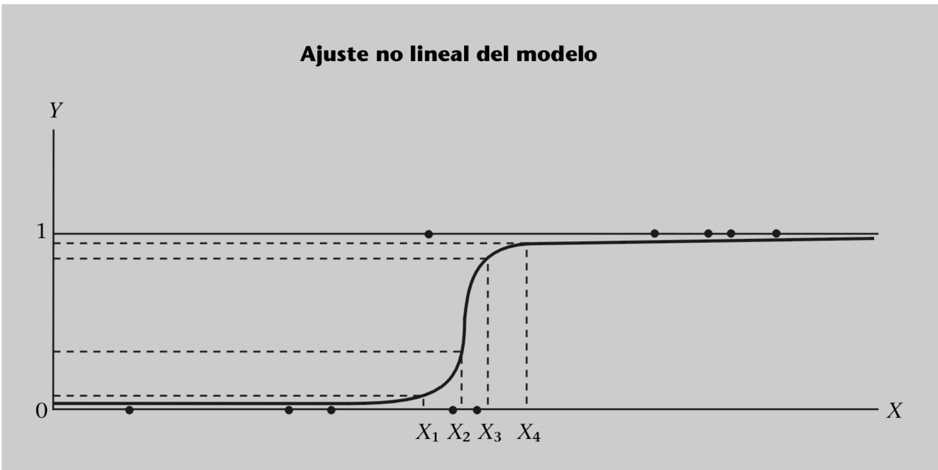
Ejemplo de independencia de la probabilidad con respecto a las condiciones iniciales en el MPL

Para ilustrar con un ejemplo que en un modelo de probabilidad lineal la probabilidad que se obtiene es independiente de las condiciones iniciales, suponemos que nos interesa modelizar la probabilidad que hay de que ciertos individuos tengan una vivienda en propiedad (Y_i = 1) o no (Y_i = 0). Para simplificar, suponemos que la renta, medida en millares de euros, es la única característica (variable explicativa) que determina si la vivienda es de propiedad o no lo es. Imaginemos que a partir de una muestra se ha estimado un modelo en el que el valor del coeficiente que acompaña a la renta se estima en 0,2.



Como es sabido, el estimador 0,2 quiere decir que una variación unitaria de la renta aumenta en 0,2 (es decir, en el 20%) la probabilidad de tener vivienda en propiedad. Por lo tanto, el incremento en la probabilidad de tener vivienda en propiedad por un individuo que en un principio tiene una renta de un millar de euros y pasa a tener una de dos millares es la misma que la de un individuo que inicialmente tiene una renta de cien millares y pasa a tener una de ciento un millares. Así que la variación en la probabilidad es la misma, con independencia de la situación inicial del individuo. No obstante, en la práctica este supuesto de linealidad no es cierto. En términos del modelo, un aumento unitario de renta no provocará el mismo aumento en la probabilidad de tener vivienda en propiedad si el individuo tiene en un principio una renta baja que si tiene una renta alta.

Con el fin de garantizar que los valores se encontrarán dentro del intervalo [0,1], podemos ajustar una función que no sea una recta, es decir, podemos especificar un modelo que no sea lineal y que en el caso de una única variable explicativa tenga la forma del gráfico que presentamos a continuación:



Ajuste no lineal del modelo

Como podemos ver en el gráfico, esta curva soluciona ambos problemas: tanto el hecho de mantener las predicciones de la probabilidad dentro del intervalo acotado por 0 y 1, como el hecho de que los incrementos en la probabilidad dependan de la situación inicial del individuo. Podemos ver esta propiedad en el hecho de que un incremento de la variable X comporta más incremento en la probabilidad de que Y valga 1 si el valor inicial pasa de X_1 a X_2 que si pasa de X_3 a X_4.

De entre las funciones que presentan una forma similar a la anterior, hay dos que son las que se utilizan con mayor frecuencia: la función logística y la función de distribución de una normal estándar, que vemos a continuación:

1) Función logística

La función logística es una función de este tipo:

$$F(z) = \frac{\exp(z)}{1 + \exp(z)};$$

por lo tanto, en nuestro caso utilizaremos la siguiente expresión:

$$F(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki}) = \frac{\exp(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki})}{1 + \exp(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki})}.$$

También podríamos escribirla de esta otra manera:

$$F(\mathbf{X}_i'\mathbf{B}) = \frac{\exp(\mathbf{X}_i'\mathbf{B})}{1 + \exp(\mathbf{X}_i'\mathbf{B})},$$

donde $\mathbf{X}_i'\mathbf{B}$ indica la combinación lineal anterior.

Otra expresión de la función logística

En algunas ocasiones la función logística aparece expresada de esta otra manera:

$$F(z) = \frac{1}{1 + \exp(-z)},$$

que, en nuestro caso, resulta ser:

$$F(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}) = \frac{1}{1 + \exp(-\beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})}.$$

Podemos ver que ambas funciones son equivalentes multiplicando el numerador y el denominador por $\exp(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})$ en la expresión anterior.

2) Función de distribución de una ley normal estándar

La función de distribución de una ley normal estándar es una función de este tipo:

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\mathrm{d}t.$$

En el caso que nosotros consideramos, esta función toma la siguiente forma:

$$F(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}) = \int_{-\infty}^{\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\mathrm{d}t.$$

Igual que en el caso anterior, también podríamos escribirla de esta otra manera:

$$F(\mathbf{X}_i'\mathbf{B}) = \int_{-\infty}^{\mathbf{X}_i'\mathbf{B}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\mathrm{d}t.$$

Cuando utilizamos la función logística para especificar el modelo de variable dependiente dicotómica, y se ajusta esta función en lugar de una recta, estamos especificando un modelo que se conoce con el nombre de *modelo logit*, mientras que, cuando utilizamos la función de distribución de una normal estándar, el modelo recibe el nombre de *modelo probit*.

Ved el modelo logit y el modelo probit en los subapartados 1.2 y 1.3, respectivamente, de este módulo didáctico.

A pesar de que ambas funciones tienen una forma muy similar (son simétricas alrededor de cero), por regla general es más sencillo trabajar con una función logística que con la función de distribución de la ley normal, puesto que, por una parte, la interpretación de los parámetros es más directa y, por la otra, el modelo es más fácil de linealizar.

El modelo de probabilidad lineal no resulta atractivo para muchos valores de las variables explicativas, porque predice con certeza la ocurrencia de la alternativa elegida, mientras que no efectúa predicciones sobre la probabilidad de que suceda.

Cuando no nos interesa utilizar un método de estimación diferente de la minimización de cuadrados, tenemos que utilizar el modelo de probabilidad lineal, aunque debemos corregir la heteroscedasticidad. Sin embargo, cuando disponemos de los medios técnicos necesarios, es recomendable utilizar los modelos logits y probits, estimándolos por máxima verosimilitud.

1.2. El modelo logit

Como acabamos de ver, una solución a los problemas que presenta el modelo de probabilidad lineal consiste en ajustar una función logística. El modelo no es lineal, aunque contiene una combinación lineal de parámetros y observaciones de las variables explicativas.

Encontraréis el ajuste en una función logística como solución a los problemas del modelo de probabilidad lineal en el subapartado 1.1.3 de este módulo didáctico.

La función logística se encuentra acotada entre 0 y 1, y, por lo tanto, siempre proporcionará valores dentro de este intervalo. Su forma sinusoidal hace que el aumento de probabilidad no sea lineal ante incrementos de las variables explicativas, sino que dependa de la situación inicial del individuo.

Cotas de la función logística

La función logística es siempre positiva e inferior a la unidad. Para ver esto, bastará con que nos fijemos en su expresión:

$$F(\beta_1 + \beta_{21}X_{2i} + ... + \beta_k X_{ki}) = \frac{\exp(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki})}{1 + \exp(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki})}.$$

El numerador y el denominador son siempre positivos porque involucran funciones exponenciales, que siempre son positivas.

Por otra parte, siempre se cumple que:

$$\exp(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki}) < [1 + \exp(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki})],$$

y, en consecuencia, el cociente nunca superará la unidad.

1.2.1. Desarrollo del modelo

Suponemos, como desde el principio, que una determinada empresa se plantea tomar una decisión en cuanto a si compra una nave industrial en una zona de un polígono o no. Como ya hemos venido diciendo, la elección entre una opción o la otra está acondicionada por una serie de variables.

Ejemplos de variables explicativas

Algunas variables explicativas del modelo pueden ser, por ejemplo, la ganancia en términos de tiempo que se producirá en la fabricación y distribución del *output* de la empresa como consecuencia de la incorporación de la nueva planta, el precio de la nave, la ganancia en términos de servicio (reducción de gastos de distribución, más espacio para guardar los excedentes, etc.), etc.

Si, por otra parte, sabemos que la utilidad puede entenderse como la propensión a comprar, y se supone que la utilidad media derivada de la elección de una opción o la otra depende de las variables que determinan la elección, se define la utilidad derivada de cada opción como la utilidad media más un término de perturbación. Esta variable no es directamente observable y, por lo tanto, lo único que podremos observar es si una empresa compra o no compra una nave industrial.

La utilización del modelo pasa por definir una variable dicotómica *Y*, que es la variable observada, y, por consiguiente, toma el valor 1 o 0 en función de la elección elegida. De hecho, sin embargo, el modelo supone que hay una variable no observable que definimos como la propensión a comprar, o la utilidad de la compra, que no podremos medir, pero que es la que está determinada por una serie de factores explicativos.

Ved la función de utilidad en la asignatura *Microeconomía*.

La especificación de un modelo logit dice que, dada la decisión de la empresa *i*-ésima representada por la variable Y_i:

$$Y_i = \begin{cases} 1 \text{ si se compra la nave,} \\ 0 \text{ si no se compra la nave,} \end{cases}$$

su esperanza condicionada a los valores de las variables explicativas o, de manera equivalente, la probabilidad de que la empresa efectúe la compra dadas las características explicativas es la siguiente:

$$\mathrm{E}[Y_i] = P_i = F(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}) =$$
$$= \frac{\exp(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})}{1 + \exp(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})},$$

donde *F* es la función logística. En algunas ocasiones, este modelo se escribe de manera más compacta como, por ejemplo, la siguiente:

$$logit(P_i) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}.$$

Más adelante veremos cuál es la interpretación de $logit(P_i)$.

Ved la interpretación de $logit(P_i)$ en el subapartado 1.2.3 de este módulo didáctico.

Formalmente, podemos suponer la existencia de una variable Y_i^* que mide la utilidad de compra. De hecho, si esta utilidad es positiva, entenderemos que se efectuará la compra; pero, si por otra parte, la utilidad es negativa, entonces no se habrá llevado a cabo la acción de comprar. El modelo logit supone que la utilidad de comprar, Y_i^*, no es observable, pero se comporta según un modelo de regresión lineal múltiple en el que incorporamos un término de perturbación, u_i^*, de manera que el modelo queda de la siguiente manera:

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i^*, \quad i = 1, \ldots, N.$$

Nunca podremos estimar este modelo, ya que no podremos medir Y_i^*. Supondremos que la variable dicotómica, que sí que se observa, está definida por la expresión que vemos a continuación:

$$Y_i = \begin{cases} 1 \text{ si } Y_i^* > 0, \\ 0 \text{ si } Y_i^* \leq 0, \end{cases}$$

donde el hecho de poner el umbral en 0 carece de implicación alguna, ya que podría tratarse de cualquier otro nivel prefijado, porque no podremos medir la variable latente. Según la definición anterior, podemos decir que:

$$P(Y_i = 1) = P(Y_i^* > 0) = \mathrm{P}(\mathbf{X}_i'\mathbf{B} > -u_i^*).$$

Si suponemos que la función de distribución aleatoria del término de perturbación está determinada por la función logística, y dada su simetría en torno a cero, podemos decir que:

$$P(Y_i = 1) = \mathrm{P}(\mathbf{X}_i'\mathbf{B} \leq u_i^*) = F(\mathbf{X}_i'\mathbf{B}),$$

y, de esta manera, obtener el modelo logit que hemos especificado inicialmente:

$$P_i = \frac{\exp(\mathbf{X}_i'\,\mathbf{B})}{1 + \exp(\mathbf{X}_i'\,\mathbf{B})}.$$

Según el supuesto que se realice sobre la distribución de probabilidad del término de perturbación, tendremos un modelo u otro. Es decir, si se toma F como la función logística, obtendremos el modelo logit y, si se toma como función de distribución una normal estándar, obtenemos el modelo probit.

1.2.2. Estimación del modelo logit

Tendremos que llevar a cabo la estimación del modelo logit mediante el método de la máxima verosimilitud. Sea el modelo logit siguiente:

$$P_i = \frac{\exp(\mathbf{X}_i'\,\mathbf{B})}{1 + \exp(\mathbf{X}_i'\,\mathbf{B})}.$$

Cualquier sistema de tratamiento estadístico avanzado proporciona las estimaciones de un modelo logit, tanto de los coeficientes (el vector de parámetros **B**) como de sus errores estándar, lo cual nos permitirá efectuar los contrastes de significación individual de los parámetros del modelo.

A grandes rasgos, dadas N observaciones independientes, podemos decir que el método de estimación de la máxima verosimilitud sigue los pasos que aparecen a continuación:

1) Especificar la función de verosimilitud del modelo, que, como se sabe, es la probabilidad conjunta de las N observaciones independientes y denotaremos* por L:

* La notación L proviene de la palabra inglesa *likelihood*.

$$\mathrm{L}(\mathbf{B}) = \prod_{i=1}^{N} P_i^{y_i}(1 - P_i)^{(1-y_i)} = \prod_{i=1}^{N} F(\mathbf{X}_i'\mathbf{B})^{y_i}[1 - F(\mathbf{X}_i'\mathbf{B})]^{(1-y_i)},$$

donde F representa, en cada caso, la función de distribución logística (o bien la de una ley de distribución normal estándar en el modelo probit). Indicamos con

y_i el valor observado de la variable dependiente para cada individuo. De esta manera, cuando el individuo elija la opción 1, entonces la verosimilitud incluirá el término $P(Y_i = 1)$, es decir, P_i; en cambio, cuando elija la opción 0, entonces incluirá el término $(1 – P_i)$.

2) Calcular el logaritmo neperiano de la función de verosimilitud, ya que su máximo se encontrará en el mismo lugar que el máximo de la función sin el logaritmo:

$$\ln L(\mathbf{B}) = \sum_{i=1}^{N} \{y_i \ln[F(\mathbf{X}_i'\mathbf{B})] + (1 - y_i) \ln[1 - F(\mathbf{X}_i'\mathbf{B})]\}.$$

3) Derivar el logaritmo neperiano de la función de verosimilitud con respecto a los parámetros que queremos estimar e igualar estas derivadas a cero. Notad que si derivamos respecto de cada β_j y si definimos f como la derivada de F, entonces llegamos al siguiente resultado:

$$\frac{\partial \ln L(\mathbf{B})}{\partial \beta_j} = \sum_{i=1}^{N} \{y_i \frac{f(\mathbf{X}_i'\mathbf{B})}{F(\mathbf{X}_i'\mathbf{B})} + (1 - y_i) \frac{-f(\mathbf{X}_i'\mathbf{B})}{[1 - F(\mathbf{X}_i'\mathbf{B})]}\} X_{ji} = 0.$$

Así pues, obtenemos k ecuaciones.

4) Solucionar el sistema de las k ecuaciones que resultan de considerar las β_j como incógnitas. Las soluciones de este sistema son los estimadores de máxima verosimilitud. Para solucionar este sistema, solemos utilizar un algoritmo iterativo, que no describiremos. Para finalizar, podemos demostrar que a partir de la matriz de las derivadas segundas podremos obtener estimadores de la varianza de las estimaciones de los parámetros.

1.2.3. Interpretación de los parámetros

Tanto en el modelo logit como en el modelo probit, como veremos después, no podemos interpretar las estimaciones de los parámetros directamente. No es cierto que el valor estimado de $\hat{\beta}_j$ indique la estimación del cambio esperado en la probabilidad $P(Y_i = 1)$ cuando la variable X_j cambia una unidad. Podemos calcular este incremento esperado en la probabilidad mediante la derivada de $P(Y_i = 1)$, o P_i, respeto de X_j, que no es igual al parámetro. Para el caso del modelo logit, esta derivada es igual a $P_i(1 - P_i)\beta_j$.

Actividad

1.1. Comprobad que el resultado de derivar P_i respecto de la variable X_j coincide con la expresión que acabamos de ver.

De todos modos, aunque las derivadas anteriores se pueden evaluar, tenemos que fijar para qué valores se estimará P_i. De manera habitual, se suele estimar fijando los valores de las variables explicativas en el vector de medias de la muestra. En algunas ocasiones, las variables explicativas no son continuas y, por lo tanto, no tiene sentido hablar de variaciones infinitesimales de éstas. En este caso, es más práctico calcular la estimación de P_i antes y después del incremento unitario de la variable explicativa, para, de esta manera, ver su efecto.

Lo único que podemos hacer en la práctica son interpretaciones del signo de los parámetros. Es decir, si $\hat{\beta}_j$ es positivo, entonces entenderemos que la variable a la que acompaña tiene un efecto que hace que sus aumentos impliquen aumentos en la probabilidad de elegir la opción 1. Mientras que, si el parámetro estimado es negativo, esto querrá decir que aumentos de la variable explicativa a la que acompaña implican disminuciones de la probabilidad de que la variable dependiente tenga un valor de 1. No podemos hablar de la magnitud de los parámetros, ya que el impacto o el efecto de incrementos de las variables explicativas implican incrementos en la probabilidad estimada de elegir la alternativa 1, pero estos aumentos tendrán una mayor o menor intensidad dependiendo del valor inicial de la variable explicativa.

Tened en cuenta que la interpretación de los signos de los parámetros estimados es similar en el modelo logit y en el probit.

Ved la interpretación de los parámetros estimados del modelo probit en el subapartado 1.3 de este módulo didáctico.

En el caso del modelo logit hay un par de medidas que se utilizan muy a menudo en la interpretación de los parámetros, y que presentamos a continuación:

1) Por una parte, tenemos el siguiente cociente de probabilidades:

$$\frac{P(Y_i = 1)}{P(Y_i = 0)} = \frac{P_i}{1 - P_i}. \qquad (1.5)$$

Se conoce como **riesgo** y también ***odds****. Para el modelo logit, el cociente de probabilidades tiene la expresión siguiente:

*** La palabra *odds* es el término inglés que indica precisamente el cociente de probabilidades opuestas.**

$$\frac{P_i}{1 - P_i} = \exp(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}).$$

Así pues, si tomamos logaritmo neperiano en ambos lados de la expresión anterior, obtenemos el siguiente resultado:

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki},$$

que a veces aparece escrito así:

$$logit(P_i) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}.$$

Esta expresión se utiliza muy a menudo, ya que, en este caso, podemos ver la similitud con el modelo de regresión lineal simple, debido que se establece una clara identificación de la parte sistemática del modelo (es decir, de $\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$) con el llamado *logit*, el logaritmo del cociente de probabilidades.

Podéis ver el modelo de regresión lineal simple en las asignaturas de estadística.

2) Por otra parte, denominaremos ***odds-ratio*** los cocientes de expresiones como la que hemos visto en 1.5. Cuando hacemos referencia a un incremento de una unidad en una de las variables explicativas, aparece de forma muy clara el *odds-ratio*. Si decimos que, al incrementar una unidad la variable X_j, estimamos que la probabilidad P_i pasa a ser P_i', entonces sabemos que:

$$logit(P_i') = \beta_j + logit(P_i),$$

y, efectuando operaciones, podemos obtener un resultado fundamental:

$$\frac{P_i'/(1 - P_i')}{P_i/(1 - P_i)} = \exp(\beta_j).$$

Por lo tanto, hemos encontrado una interpretación para la cantidad $\exp(\beta_j)$, que se recibe el nombre de *odds-ratio* y indica el cambio relativo que experimenta el cociente de probabilidades (la expresión 1.5) cuando la variable X_j aumenta una unidad.

El modelo logit se utiliza a menudo como una técnica estadística de clasificación entre dos grupos. Suponiendo que la variable dicotómica representa la pertenencia o no al primer grupo, el modelo establece una especificación de las variables que afectan a la probabilidad de pertenecer a este primer grupo. De hecho, supera muchas desventajas del análisis discriminando, ya que no requiere que los factores que afectan a la probabilidad de pertenecer al grupo deban tener una distribución normal multivariante para extraer inferencias. Además, si disponemos de una observación nueva, el modelo puede predecir la probabilidad de que la observación pertenezca al grupo, y, a partir de esta estimación, podemos llevar a cabo la clasificación.

1.3. El modelo probit

El modelo probit se especifica de manera análoga al modelo logit, con la única diferencia de que utiliza la función de distribución de una ley de distribución normal estándar; por lo tanto:

$$E[Y_i] = P_i = F(\beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki}) = \int_{-\infty}^{X_i'B} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\mathrm{d}t.$$

La estimación del modelo se hace de la misma manera que hemos presentado en el subapartado anterior, y, en este momento, sólo es preciso que hagamos hincapié en la gran dificultad que presenta la interpretación directa de los parámetros del modelo probit. Lo único que podemos hacer fácilmente es interpretar los signos de los parámetros y su significación.

1.4. Medidas de bondad del ajuste en los modelos de elección dicotómica

Los modelos logit y probit no se estiman por el método de los mínimos cuadrados ordinarios, por lo que tendremos que valorar los resultados de la estimación con medidas estadísticas adecuadas.

Una forma análoga al coeficiente de determinación, que tiene un objetivo similar a R^2, es la definición de una nueva medida que se conoce con el nombre de *pseudo*-R^2. Esta medida se define tal y como vemos a continuación:

$$\rho^2 = \left[1 - \frac{\ln L(\hat{\mathbf{B}})}{\ln L_0(\hat{\mathbf{B}}^0)}\right].$$

En esta definición utilizamos el logaritmo de la verosimilitud evaluada en los valores de los coeficientes obtenidos en el proceso de estimación para dos modelos. El del numerador es el modelo que queremos evaluar, y en el denominador colocamos un modelo que no contiene ninguna variable explicativa y sólo contiene el término independiente. Por lo tanto, en este caso, sólo habrá estimado un parámetro. Si, por otra parte, utilizamos un modelo logit, el modelo del numerador sería $logit(P_i) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$, y el del denominador, $logit(P_i) = \beta_1$.

Podemos ver con facilidad...

... que, para el modelo que sólo tiene término constante, el valor de la verosimilitud se obtiene multiplicando la proporción de ceros elevada a una potencia igual al número de ceros, por la proporción de unos elevada a la potencia igual al número de unos.

La interpretación del coeficiente es parecida a la del coeficiente de determinación (R^2) en un modelo de regresión lineal múltiple. Si se encuentra próximo a la unidad, diremos que el modelo ajusta bien y, si se encuentra cerca de cero, concluiremos que el ajuste es inadecuado.

Para llevar a cabo un **contraste de significación global de los parámetros**, es decir, para contrastar si todos los parámetros que acompañan a las variables explicativas (sin considerar el del término independiente) son iguales a cero o si, por lo menos uno, es significativamente diferente de cero, se realiza el siguiente contraste:

$$2[\ln L(\hat{\mathbf{B}}) - \ln L_0(\hat{\mathbf{B}}^0)],$$

en el que los logaritmos de las verosimilitudes corresponden a los dos modelos que acabamos de mencionar. Este estadístico se compara con unas tablas de la distribución χ-cuadrado con $(k - 1)$ grados de libertad –la diferencia entre el número de parámetros del modelo inicial y el modelo sin variables explicativas–. Si el estadístico supera el valor de las tablas, se concluye que como mínimo uno de los parámetros del modelo es significativamente diferente de cero. Así pues, podemos deducir que por lo menos una de las variables explicativas tiene un efecto significativo en la explicación de la probabilidad de elegir la alternativa 1.

De manera análoga al MRLM, podemos realizar los **contrastes sobre la significación individual de los parámetros**. En algunas ocasiones, para hacer el contraste, en lugar de utilizar el contraste de la t de Student, se utiliza el **estadístico de Wald**, cuyo comportamiento con la hipótesis nula sigue una distribución χ-cuadrado con un grado de libertad.

Estadístico de Wald

En los paquetes econométricos es habitual utilizar, para este tipo de modelos, el estadístico de Wald en lugar del estadístico de la t de Student. Podemos entender este estadístico –aunque en realidad no lo es– como el estadístico de la t de Student al cuadrado.

Y ya para finalizar, con la intención de evaluar correctamente los resultados de los modelos de elección binaria, podemos elaborar **tablas de clasificación**. Es decir, se comparan las observaciones de la variable dependiente (que toman los valores 0 y 1) con las probabilidades que obtenemos por ajuste del modelo.

Si, por ejemplo, un individuo tiene un valor igual a 1 en la variable dependiente y su probabilidad obtenida por ajuste del modelo supera el valor 0,5, este resultado se considera un acierto del modelo. Del mismo modo, si el valor observado es 0 y la probabilidad ajustada (probabilidad de elegir 1) es baja, también se considera un acierto. De manera que, con las observaciones consideradas para la estimación del modelo, podemos considerar una tabla como la que vemos continuación:

Tabla de clasificación

		Predicciones del modelo	
		$P(Y_i = 1) \geq 0{,}5$	$P(Y_i = 1) < 0{,}5$
Observaciones	$y_i = 1$		
	$y_i = 0$		

En las casillas vacías de la tabla que acabamos de mostrar se cuenta la frecuencia de observaciones que cumplen las condiciones indicadas por la fila y la columna correspondientes. Así, las observaciones que cumplen las condiciones de las casillas de la diagonal son aciertos. Por otra parte, las observaciones contadas en las casillas de fuera de la diagonal son equivocaciones del modelo.

En algunos casos podemos variar el valor de 0,5 para así mejorar la clasificación final del modelo. Este hecho se justifica diciendo que, con vistas a utilizar el modelo con finalidades predictivas, conviene modificar el punto a partir del cual se considera suficientemente grande la probabilidad de la primera elección para asignar el individuo a esta categoría. Esto es interesante cuando la muestra contiene más elecciones de un tipo que de otro.

1.5. Ejemplo de utilización del modelo logit

En este subapartado analizaremos los resultados de la estimación de un modelo de elección dicotómica.

Un gremio ha efectuado un estudio entre sus asociados para saber si cumplen estrictamente la normativa de prevención de riesgos laborales. En total hay una representación de 150 pequeñas empresas. Para cada empresa, la variable dependiente indica con el valor 1 si la empresa cumple la normativa, y con el valor 0, si la incumple. En tanto que variables explicativas de este comportamiento se utilizan algunas variables relativas a las dimensiones y características de la empresa: número de trabajadores *(trab)*, volumen de subcontratación respecto de la facturación total en porcentaje *(subcontr)*, gasto en seguro de responsabilidad civil con respecto al total de facturación en tanto por diez mil *(respciv)* y una variable dicotómica que indica si han recibido alguna inspección en materia de seguridad en los últimos seis meses *(inspec)*. A continuación, reproducimos los resultados de la estimación:

Resultados de la estimación del modelo de elección

Variable dependiente: Y

Número de observaciones: 150

VARIABLE	COEFICIENTE (**B**)	ERROR ESTAD.	WALD	DF	SIGNIFICACIÓN	exp (β_j)
treb	0,9419	0,2101	20,0821	1	0.0000	2,5648
subcontr	-0,0951	0,0350	7,3869	1	0,0066	0,9093
respciv	0,0039	0,0198	0,0387	1	0,8440	1,0039
inspec	0,9711	0,8715	1,2417	1	0,2651	2,6409
Constant	-14,3163	3,2033	19,9734	1	0,0000	

-2 log-verosimilitud con sólo la constante: 196,02546
-2 log-verosimilitud con todas las explicativas: 45,400
Grados de libertad = 4
Significación = 0,0000
Contraste χ-cuadrado: 150,625

Tabla de clasificación

		Predicciones del modelo	
		$P(Y_i = 1) \geq 0{,}5$	$P(Y_i = 1) < 0{,}5$
Observaciones	$y_i = 1$	91	5
	$y_i = 0$	7	47

Tabla de clasificación

Los resultados que nos muestra la tabla ofrecen una idea de la validez del modelo en cuanto a predicción. Los elementos de la diagonal de la tabla son aciertos del modelo, mientras que los que no aparecen en la diagonal son valores que no se ajustan a las predicciones del modelo. Así, podemos observar que el modelo predice correctamente las observaciones en el 94,79%, en el caso de las empresas que cumplen la normativa ($y_i = 1$), y en el 87,04%, en el de las empresas que no la cumplen ($y_i = 0$). De manera global, el porcentaje de aciertos del modelo es del 92,00%.

Podemos deducir los puntos que anotamos a continuación a partir de la tabla anterior:

- Globalmente, el modelo hace un buen ajuste, ya que el contraste de significación global (χ-cuadrado) nos lleva a aceptar la hipótesis alternativa que dice que por lo menos alguno de los parámetros que acompañan a las variables explicativas es significativamente diferente de cero.

- Por otra parte, el pseudo-R^2, si se calcula con los datos de los cuadros anteriores, conduce al valor del 76,84%, que podemos considerar bueno, aunque no excesivamente elevado.

- El porcentaje de clasificaciones correctas del modelo es muy bueno, ya que clasifica bien el 92% de las observaciones, aunque consigue un mejor porcentaje en el grupo de empresas que cumplen la normativa, es decir, cuando $Y_i = 1$.

- Los signos de los coeficientes son los esperados, pero el contraste de significación individual de los parámetros provoca la imposibilidad de rechazar que el parámetro que acompaña a la variable *respciv* (que indica el tanto por diez mil dedicado a seguro de responsabilidad civil) no sea significativamente diferente de cero. Este hecho nos lleva a concluir que esta variable no indica nada sobre la propensión que hay a cumplir la normativa.

- Con respecto al hecho de haber recibido recientemente alguna inspección (algo que aparece registrado en la variable *inspec*) tampoco podemos afirmar que su coeficiente sea significativamente diferente de cero y, por lo tanto, concluiremos que este hecho no tiene influencia en la probabilidad existente de que una empresa cumpla la normativa de seguridad.

- Como los primeros coeficientes son significativos, podemos aceptar su significación individual:

– Por una parte, como el coeficiente que acompaña a la variable del número de trabajadores es positivo y significativamente diferente de cero, diremos que, cuantos más trabajadores haya (empresa mayor), más alta es la probabilidad de que se cumpla la normativa de seguridad.

– Por otra parte, cuanto mayor es el porcentaje de subcontratación respecto del total de la facturación, menor es la propensión de la empresa a cumplir la normativa.

En el ejemplo anterior también podemos interpretar los *odds-ratios*. Tomamos, por ejemplo, 2,5648, que es *el odds-ratio* de la variable *trab* (número de trabajadores). Este valor nos indica que el cociente de probabilidades entre la probabilidad de cumplir la normativa y la de no cumplirla en una empresa dada se multiplicaría por 2,5648, cuando la empresa pasase a tener un trabajador más, y no cambiara ninguna otra característica de las consideradas.

Imaginaos, ahora, que comparamos dos empresas, una de quince trabajadores y una de dieciséis (con el resto de las características iguales), y que la primera tiene un cociente de probabilidades de 2:1 (es decir, que tiene el doble de probabilidad de cumplir la normativa de seguridad que de no cumplirla –de hecho, decimos que tendría una probabilidad de cumplir la normativa del 66,66% y, en consecuencia, una probabilidad de no cumplirla del 33,34%–). Entonces, si tomásemos la empresa de dieciséis trabajadores, este cociente pasaría a ser 5,1296:1 (que proviene de multiplicar 2:1 por 2,5648). Finalmente, lo que acabamos de ver quiere decir que la empresa de dieciséis trabajadores tiene una probabilidad de cumplir la normativa del 83,69%, y una de no cumplirla del 16,31%. Apreciad, en cualquier caso, que 83,69 se obtiene de calcular $100 \cdot 5{,}1296 / (5{,}1296 + 1)$.

El modelo que hemos estudiado posee la siguiente utilidad práctica: de entrada, sabemos que las empresas de más trabajadores muestran una mayor propensión a cumplir la normativa de seguridad y que aquellas que subcontratan más tienen menos probabilidades de cumplirla. Por otra parte, parece que a partir de los datos se verifica que, si una empresa dedica más dinero a cubrirse de los riesgos derivados de la responsabilidad civil (respecto del total facturado), este hecho no aporta ningún indicio en torno a su actividad preventiva en el cumplimiento de la normativa de seguridad. Para acabar, las empresas que con anterioridad han recibido inspecciones de seguridad no parece que tengan un comportamiento diferenciado del resto, en relación con el hecho de que varíe su comportamiento esperado (es decir, con respecto a que varíe la probabilidad de que cumplan la normativa).

Si nuestro objetivo fuese hacer un escenario y se diera una nueva empresa con unas características determinadas –15 trabajadores, el 10% de subcontratación respecto de su facturación total, 5 unidades de cada 10.000 dedicadas al seguro de responsabilidad civil y que no hubiese recibido ninguna inspección–, podríamos calcular la probabilidad de que ésta cumpliese la normativa vigente. Para hacer el cálculo, tenemos que escribir la ecuación del modelo:

$$P_i = \frac{\exp(-14{,}3163 + 0{,}9419 \cdot trab - 0{,}0951 \cdot subcontr + 0{,}0039 \cdot respciv + 0{,}9711 \cdot inspec)}{1 + \exp(-14{,}3163 + 0{,}9419 \cdot trab - 0{,}0951 \cdot subcontr + 0{,}0039 \cdot respciv + 0{,}9711 \cdot inspec)}$$

y el resultado final es que esta empresa tiene una probabilidad estimada del 24,61% de cumplir la normativa de seguridad.

Presentación alternativa de los resultados del modelo logit

En algunas ocasiones se presentan los resultados del modelo en la forma alternativa siguiente:

$$logit(P_i) = -14{,}3163 + 0{,}9419 trab - 0{,}0951 subcontr + 0{,}0039 respciv + 0{,}9711 inspec.$$

En este caso, para la empresa que utilizamos como ejemplo, y sustituyendo sus características, obtenemos un *logit* igual a –1,1193. Por lo tanto, haciendo el exponencial, obtenemos que exp(–1,1193) = 0,3265 y, en consecuencia, su cociente de probabilidades es 0,3265, lo cual quiere decir que la probabilidad de cumplir la normativa dividida por la probabilidad de no cumplirla es igual a 32,65. Así pues, el cociente de probabilidades es 0,3265:1, o bien 32,65:100; pero, utilizando la siguiente expresión:

$$\frac{P_i}{1 - P_i} = \frac{32{,}65}{100},$$

y aislando P_i, esto quiere decir que la empresa tiene una probabilidad de cumplir la normativa del 24,61% (es decir, 32,65 / (32,65 + 100)).

Actividad

1.2. Utilizad los datos de este subapartado para estimar un modelo de probabilidad lineal y comparad los resultados con los que ha facilitado el modelo logit. Tened en cuenta los inconvenientes que supone el uso del modelo de probabilidad lineal y observad si algún ajuste para alguna observación se encuentra fuera del intervalo [0,1].

1.6. Modelos de elección múltiple

En los modelos de elección múltiple se considera, además, que es posible tener más de dos alternativas de elección; es decir, la variable dependiente que queremos explicar tiene más de dos categorías.

En este tipo de modelo hay existen tres formas posibles de plantear la parte estructural del modelo*:

*** La parte estructural del modelo es la parte que corresponde a las variables explicativas.**

1) Las variables explicativas son características del individuo que efectúa la elección. Por ejemplo, en la decisión de ir a un espectáculo o no ir, pueden influir sólo las características personales del espectador.

2) Las variables explicativas son características de la elección que hay que efectuar y no del individuo. En la elección, por ejemplo, de una marca de leche puede influir el precio del litro, si es leche fresca o no lo es, y la fecha de caducidad que tiene impresa.

3) Las variables explicativas contienen características tanto individuales como de la elección que hay llevar a cabo. Cuando decidimos utilizar un tipo de transporte, influye en ello la renta del individuo que se desplaza, la distancia a la quiere desplazarse, el precio y la velocidad del medio de transporte, y, por lo tanto, hay características de la persona así como del elemento acerca del que está decidiendo.

La variable dependiente (que es politómica) a menudo se expresa mediante el uso de una codificación que va desde 0 hasta el número total de alternativas menos una.

Ved las variables politómicas en el subapartado 1.2 del módulo "Variables exógenas cualitativas".

El problema de los modelos de elección múltiple consiste en saber reconocer la independencia de las alternativas irrelevantes. Es decir, hay que tener modelos que no sean sensibles a la presencia de alternativas superfluas.

Ejemplo de presencia de alternativas superfluas en un modelo de elección múltiple

Supongamos que una persona tiene tres posibilidades de contratar un seguro de automóvil: a terceros, a todo riesgo sin franquicia o a todo riesgo con franquicia. La elección se hace con probabilidad 3/6, 2/6 y 1/6, respectivamente. Pero, supongamos que se le ofrece una alternativa de elección que es completamente irrelevante. Supongamos, entonces, que la compañía ofrece la posibilidad de comprar seguro a todo riesgo con franquicia que consta escrita en euros o en pesetas. Asimismo, supongamos que al asegurado le es completamente indiferente con qué tipo de unidad monetaria se le tramite la contratación del seguro. Entonces, esta alternativa sería irrelevante para el asegurado y, por lo tanto, podríamos pensar que la probabilidad de elegir entre las cuatro alternativas ahora sería: 3/6, 2/6, 1/12 y 1/12 (terceros, todo riesgo sin franquicia, todo riesgo con franquicia expresada en euros y todo riesgo con franquicia expresada en pesetas). En cambio, el modelo multinomial asignaría las siguientes probabilidades: 3/7, 2/7, 1/7 y 1/7. La razón de esto es que, al hacerlo así, podemos mantener el mismo cociente relativo que se asignaba al principio, cuando había tres elecciones. Fijémonos en que con esta última asignación el cociente relativo entre elegir a terceros y a todo riesgo sin franquicia es 3/2. El cociente entre terceros y el primer tipo de todo riesgo con franquicia en euros es 3/1 (igual que al inicio). En cambio, comprobamos que, si tomamos la primera asignación de probabilidad, este último cociente sería (3/6) / (1/12) y daría 6/1.

El modelo también se formula mediante la modelización de la probabilidad de elección de cada alternativa. La estimación del vector de parámetros se lleva a cabo con el procedimiento de máxima verosimilitud, y la interpretación de los parámetros es análoga al caso de los modelos logit.

Podemos estudiar todos los modelos uniecuacionales (tanto de variable dependiente cualitativa como el MRLM) desde una perspectiva global que ahora pasamos a describir de forma breve. Los **modelos lineales generalizados** constituyen un tipo de aproximación más amplia que el modelo de regresión lineal múltiple. La formulación que consideraremos a continuación tiene como caso particular modelos conocidos, como, por ejemplo, el modelo de regresión, los modelos de análisis de la varianza, el modelo de regresión logística, el modelo probit, y también otros modelos que quedan fuera del alcance de este material: los modelos para datos de enumeración, los modelos de diseño de experimentos y los modelos de análisis de la supervivencia.

En los últimos tiempos, los modelos lineales generalizados se han introducido con fuerza, ya que constituyen un marco de referencia que unifica la teoría de la modelización lineal clásica que hemos estudiado a lo largo de las asignaturas de econometría.

Los **componentes de un modelo lineal generalizado** son los que enumeramos a continuación:

1) El **componiendo aleatorio**: Y_i, con una determinada distribución de la familia exponencial de distribuciones (normal, de Poisson, binomial, gama, etc.) y esperanza matemática $E[Y_i] = \mu_i$.

2) El **componiendo sistemático**: $\eta_i = \beta_1 + \beta_2 X_{2i} + \beta_k X_{ki}$, donde las variables explicativas $X_2, ..., X_k$ producen el predictor lineal.

3) El **vínculo* entre el componente sistemático y el componente aleatorio**: de manera que hay una función monótona diferenciable g tal que $\eta_i = g(\mu_i)$.

* En inglés, *link*.

En todos los casos, es posible encontrar la estimación de máxima verosimilitud del vector de parámetros **B** mediante un procedimiento iterativo. Una vez obtenidas las estimaciones, obtenemos la estimación de la matriz de varianzas y covarianzas, para, de esta manera, poder valorar la significación de los parámetros.

Si tomamos de forma adecuada la distribución del componente aleatorio y la forma particular de la función de vínculo, podemos obtener los modelos lineales clásicos como caso particular. En concreto, para obtener el modelo de regresión lineal múltiple, es suficiente con que supongamos que la distribución es la normal y que la función de vínculo es la identidad.

Glosario

modelo de probabilidad lineal (MPL)
Modelo de regresión lineal múltiple que se aplica cuando la variable dependiente es dicotómica.

modelo logit
Modelo utilizado para casos en los que la variable dependiente es dicotómica, y que está basado en la función logística. También se conoce con el nombre de *modelo logístico*.

modelo probit
Modelo usado para casos en los que la variable dependiente es dicotómica, que se basa en la función de distribución normal estándar.

modelos de elección discreta
Modelos utilizados en aquellos casos en los que la variable dependiente es cualitativa.

odds-ratio
En el modelo logit, cociente en dos escenarios diferentes de los cocientes entre la probabilidad de elegir la primera alternativa y la segunda.

pseudo-R^2
Medida de bondad del ajuste en modelos de elección discreta.

Bibliografía

Novales, A. (1993). *Econometría* (cap. 16, parte I; 2.ª ed.). Madrid: McGraw–Hill.