Regresión con variable dependiente discreta

Aplicaciones Econométricas para la empresa

Javier Nievas

Índice

1. Características de los modelos de elección discreta.
2. Modelos binarios.
3. El Modelo de Probabilidad Lineal.
4. Regresión Probit y Logit.
5. Estimación e Inferencia en Probit y Logit.
6. Ejemplo sobre si la discriminación racial condiciona la concesión de préstamos hipotecarios.
7. Modelos de respuesta múltiple.

Modelos de elección discreta

- Hasta ahora hemos considerado que la variable dependiente (Y) era de tipo continuo:
  - gasto en consumo familiar
  - tasa de mortalidad en accidentes de tráfico
- En los modelos de elección discreta la variable endógena es una variable aleatoria que toma un conjunto finito de valores.
  - Modelos de respuesta binaria:
    - Busca trabajo o no busca trabajo, fuma o no fuma, etc.
  - Modelos de respuesta múltiple no ordenada:
    - Medios de transporte: coche, tren, autobús.
  - Modelos de respuesta múltiple ordenada:
    - Notas: suspenso, aprobado, notable, sobresaliente.

Variables dependientes binarias, ¿qué las hace diferentes?

Ejemplos de modelos binarios:

- Y = tiene título universitario, o no
  X = especialidades en bachillerato, notas en bachillerato, variables demográficas.
- Y = fumador, o no
  X = impuestos al tabaco, renta, variables demográficas.
- Y = hipoteca concedida, o no
  X = raza, ingresos, características de la casa, estado civil.

Ejemplo: Motivos de denegación de concesión de hipoteca en Boston

- Datos de las solicitudes individuales de hipotecas unifamiliares realizadas en 1990 en el área metropolitana de Boston.
- 2380 observaciones, recogidas de la base de datos sobre hipotecas en Boston (HMDA).

Variables:

- Variable dependiente:
  - ¿La hipoteca se aceptó o se denegó?
- Variables explicativas:
  - Renta, riqueza, situación laboral.
  - Raza del solicitante.

Variables dependientes binarias y el Modelo de Probabilidad Lineal

Un punto de partida natural es comenzar con un modelo lineal con un solo regresor:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Cuestiones que nos plantearemos:

- ¿Cómo se interpreta β_1 cuando Y es binaria?

  ¿Es $\beta_1 = \dfrac{\Delta Y}{\Delta X}$?
- ¿Qué significa la línea $\beta_0 + \beta_1 X$ cuando Y es binaria?
- ¿Cómo se interpreta la predicción $\hat{Y}$ cuando Y es binaria? Por ejemplo, ¿qué significa que $\hat{Y} = 0.26$?

El Modelo de Probabilidad Lineal

Cuando Y *es binaria, el modelo de regresión lineal es:*

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Tomando esperanza matemática al modelo:
  $$E(Y|X) = 1\times\Pr(Y=1|X) + 0\times\Pr(Y=0|X) = \Pr(Y=1|X)$$
- Bajo la hipótesis $E(u_i|X_i) = 0$, se cumple que:
  $$E(Y_i|X_i) = E(\beta_0 + \beta_1 X_i + u_i|X_i) = \beta_0 + \beta_1 X_i,$$

  Por lo tanto, $\Pr(Y=1|X) = \beta_0 + \beta_1 X_i$

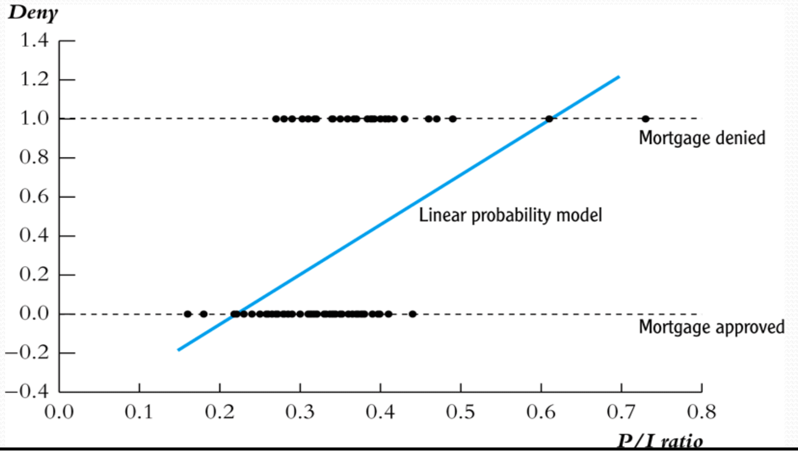
Por eso se llama **Modelo de Probabilidad Lineal (MPL)**

El Modelo de Probabilidad Lineal

- El valor predicho de Y es una **probabilidad**:
  - $E(Y|X=x) = \Pr(Y=1|X=x)$ = prob. de que $Y = 1$ dado x
  - $\hat{Y}$ = la **probabilidad** de que $Y_i = 1$, dado X
- En el modelo de probabilidad lineal, el valor predicho de Y se interpreta como la probabilidad de que Y=1.
- β_1 es el cambio en la probabilidad de que Y=1 para un cambio unitario en x:

$$\beta_1 = \frac{\Pr(Y=1 \mid X = x + \Delta x) - \Pr(Y=1 \mid X = x)}{\Delta x}$$

Ejemplo: modelo de probabilidad lineal, datos HMDA. Denegación de hipoteca frente al ratio pagos de deudas respecto al ingreso (P/I ratio) en un subgrupo de los datos HMDA (n = 127).



Modelo de probabilidad lineal con los datos completos HMDA

$$\widehat{den} = \text{-0.080} + 0.604\, P/I\ ratio \qquad (n = 2380)$$
$$(.032) \quad (.098)$$

- ¿Cuál es el valor predicho para un P/I *ratio* = 0.3?

$$\widehat{\Pr(den.=1 \mid P/Iratio=0.3)} = \text{-0.080} + 0.604 \times 0.3 = 0.151$$

- El "efecto" de aumentar el ratio P/I de 0.3 a 0.4:

$$\widehat{\Pr(den=1 \mid P/Iratio=0.4)} = \text{-0.080} + 0.604 \times 0.4 = 0.212$$

El efecto sobre la probabilidad de denegar la hipoteca de un incremento en el ratio P/I *de* 0.3 a 0.4 es un incremento de la probabilidad de un 0.061, es decir, un 6.1% (¿por qué?).

Modelo de probabilidad lineal con los datos completos HMDA

Si ahora incluimos un nuevo regresor, "raza negra":

$$\widehat{den} = \underset{(.032)}{-0.091} + \underset{(.098)}{0.559}\ P/I\ ratio + \underset{(.025)}{0.177}\ negro$$

La probabilidad estimada de denegación es:

- Para una persona negra con un ratio P/I = 0.3:

$$\widehat{\Pr(den=1)} = -0.091 + 0.559\times0.3 + 0.177\times1 = 0.254$$

- Para un blanco, con un ratio P/I = 0.3:

$$\widehat{\Pr(den=1)} = -0.091 + 0.559\times0.3 + 0.177\times0 = 0.077$$

- Diferencia = 0.177 = 17.7%
- El coeficiente estimado de "negro" es significativo al 5%
- *Pero aún puede haber mucho sesgo de variables omitidas…*

El modelo de probabilidad lineal: resumen

- El Modelo de probabilidad lineal (MPL) considera la $\Pr(Y=1|X)$ como una función lineal de X.
- Ventajas:
  - Es simple de estimar e interpretar.
  - La inferencia es igual que en la regresión múltiple (se necesita que las desviaciones típicas sean robustas en homoscedasticidad).
- Desventajas:
  - El impacto de un aumento unitario de X en la probabilidad es constante, eso no tiene sentido (pensar en el ej. HMDA).
  - El MPL predice probabilidades que pueden ser <0 ó >1!
  - Las perturbaciones son heteroscedásticas.
- Estas desventajas se resuelven usando un Modelo de probabilidad **no lineal**: regresión Probit y Logit.

Regresión Probit y Logit

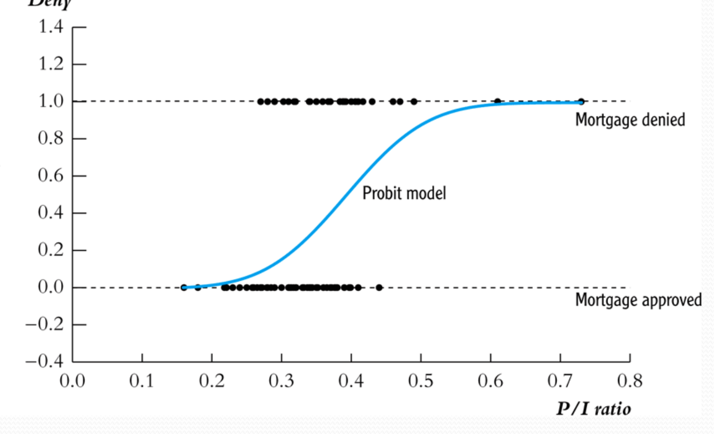
El problema del modelo de probabilidad lineal es que asume que la probabilidad de que Y=1 es lineal:

$$\Pr(Y = 1|X) = \beta_0 + \beta_1 X$$

En cambio, buscamos:

1. Que $\Pr(Y = 1|X)$ se incremente con X si β_1>0.
2. Que $0 \leq \Pr(Y = 1|X) \leq 1$ para todo X.

Estas condiciones requieren usar una forma funcional no lineal para la probabilidad. Como una curva en forma de S.



- El modelo Probit cumple estas condiciones:
  - I. $\Pr(Y = 1|X)$ se incrementa en X cuando β_1>0, y
  - II. $0 \leq \Pr(Y = 1|X) \leq 1$ para todo X

Regresión Probit

- ***La regresión Probit*** modeliza la probabilidad de que Y=1 usando la función de distribución acumulada estándar, $\Phi(z)$, evaluada como $z = \beta_0 + \beta_1 X$.

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

donde Φ es la función de distribución Normal acumulada y $z = \beta_0 + \beta_1 X$ es el "valor-z" del modelo Probit.

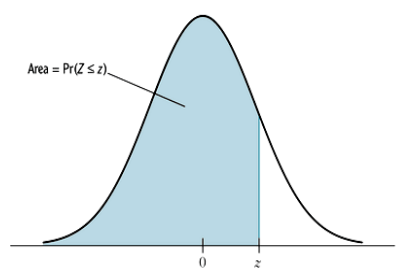
$$\Phi(\beta_0 + \beta_1 X) = \int_{-\infty}^{X'\beta} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \partial t$$

Ejemplo: Suponer β_0 = -2, β_1= 3, X = 0.4, entonces

$\Pr(Y = 1|X=0.4) = \Phi(-2 + 3\times 0.4) = \Phi(-0.8)$

$\Pr(Y = 1|X=0.4)$ = área a la izquierda de la función de densidad de una Normal estándar para z = -0.8, que es…

TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = \Pr(Z \leq z)$



z	Second Decimal Value of z: 0	1	2	3	4	5	6	7	8	9
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121

$\Pr(z \leq -0.8) = 0.2119$

Regresión Probit

¿Por qué usamos la distribución de probabilidad acumulada Normal estándar?

- La forma en S nos da lo que queremos:
  - La $\Pr(Y = 1|X)$ aumenta con X si β_1>0
  - $0 \leq \Pr(Y = 1|X) \leq 1$ para todo X
- Fácil de usar (las probabilidades están tabuladas en las tablas de la Normal), también es fácil de calcular con el software econométrico habitual.

Regresión Probit. Más ventajas.

- Tiene una interpretación relativamente sencilla:
  - $\beta_0 + \beta_1 X$ = valor-z
  - $\hat{\beta}_0 + \hat{\beta}_1 X$ es la estimación del valor-z, dado X
  - β_1 es el cambio en el valor-z para un cambio unitario en X.
- En estos modelos los coeficientes no indican el cambio en la probabilidad de Y dado un cambio unitario en las X's, salvo el signo, que si es + indica que la variable favorece la ocurrencia del evento y si es – indica que la variable es una barrera para la ocurrencia del evento.

Ejemplo de estimación: datos HMDA

```
. probit deny p_irat, r;
Iteration 0:   log likelihood = -872.0853              We'll discuss this
   later
Iteration 1:   log likelihood = -835.6633
Iteration 2:   log likelihood = -831.80534
Iteration 3:   log likelihood = -831.79234
Probit estimates                                  Number of obs   =       2380
                                                  Wald chi2(1)    =      40.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -831.79234                       Pseudo R2       =     0.0462
------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      p_irat |   2.967908   .4653114     6.38   0.000     2.055914    3.879901
       _cons |  -2.194159   .1649721   -13.30   0.000    -2.517499    -1.87082
------------------------------------------------------------------------------
```

$$\widehat{\Pr(deny=1|P/I\,ratio)} = \Phi(-2.19 + 2.97 \times P/I\ ratio)$$
$$(.16) \quad (.47)$$

Ejemplo de estimación: datos HMDA

$$\widehat{\Pr(deny=1|P/I\ ratio)} = \Phi(-2.19 + 2.97 \times P/I\ ratio)$$
$$(.16) \quad (.47)$$

- Coeficiente positivo: ¿*Tiene sentido*?
- Las desviaciones típicas se interpretan como siempre.
- Probabilidades estimadas:

$$\Pr(deny=1|P/I\ ratio=0.3) = \Phi\ (-2.19+2.97\times 0.3)$$
$$= \Phi\ (-1.30) = 0.097$$

- Efecto de un cambio en el ratio *P/I de* 0.3 a 0.4:

$$\Pr(deny=1|P/I\ ratio=0.4) = \Phi\ (-2.19+2.97\times 0.4)$$
$$= \Phi\ (-1.00) = 0.159$$

- La probabilidad estimada de denegación aumenta de 0.097 a 0.159

Regresión Probit con múltiples regresores

$$\Pr(Y = 1|X_1, X_2) = \Phi\ (\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

- Φ es la función de distribución acumulada normal.
- $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ es el "z-valor" del modelo Probit.
- β_1 es el efecto sobre el z-valor de un cambio unitario en X_1, manteniendo constante X_2

Ejemplo de estimación probit: datos HMDA

```
. probit deny p_irat black, r;
Iteration 0:   log likelihood =  -872.0853
Iteration 1:   log likelihood = -800.88504
Iteration 2:   log likelihood =  -797.1478
Iteration 3:   log likelihood = -797.13604
Probit estimates                                  Number of obs   =       2380
                                                  Wald chi2(2)    =     118.18
                                                  Prob > chi2     =     0.0000
Log likelihood = -797.13604                       Pseudo R2       =     0.0859
------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      p_irat |   2.741637   .4441633     6.17   0.000     1.871092    3.612181
       black |   .7081579   .0831877     8.51   0.000      .545113    .8712028
       _cons |  -2.258738   .1588168   -14.22   0.000    -2.570013   -1.947463
------------------------------------------------------------------------------
```

Analizaremos los detalles de la estimación a continuación…

Ejemplo con datos HMDA. Probabilidades estimadas con probit

$$\Pr(deny=1|P/I, black) = \Phi(-2.26 + 2.74 \times P/I\ ratio + .71 \times black)$$

$$(.16) \quad (.44) \quad (.08)$$

- El coeficiente de *black* es significativo?
- Efecto estimado de la raza para un ratio *P/I* = 0.3:

$$\Pr(deny=1|0.3,1) = \Phi(-2.26+2.74\times0.3+0.71\times1) = 0.233$$

$$\Pr(deny=1|0.3,0) = \Phi(-2.26+2.74\times0.3+0.71\times0) = 0.075$$

- Diferencia en las probabilidades de rechazo = 0.158 (15.8%).
- *Aunque todavía puede haber sesgo de variables omitidas.*

Regresión Logit

- ***La regresión Logit*** calcula la probabilidad de Y=1, dado X, usando la función de distribución acumulada *logística* estándar, calculada en $z = \beta_0 + \beta_1 X$:

$$\Pr(Y = 1|X) = F(\beta_0 + \beta_1 X)$$

- F es la función de distribución acumulada *logística*:

$$F(\beta_0 + \beta_1 X) = \frac{1}{1+e^{-(b_0+b_1X)}}$$

- Como los modelos Logit y Probit usan diferentes funciones de probabilidad, los coeficientes (β 's) son distintos.

Regresión Logit

$\Pr(Y = 1|X) = F(\beta_0 + \beta_1 X)$; donde $F(\beta_0 + \beta_1 X) = \dfrac{1}{1+e^{-(\beta_0+\beta_1 X)}}$

Ejemplo: β_0 = -3, β_1= 2, X = 0.4,

entonces $\beta_0 + \beta_1 X$ = -3 + 2x0.4 = -2.2, luego

$\Pr(Y = 1|X{=}0.4) = 1/(1+e^{-(-2.2)}) = 0.0998$

¿Por qué usamos el Logit si ya tenemos el Probit?

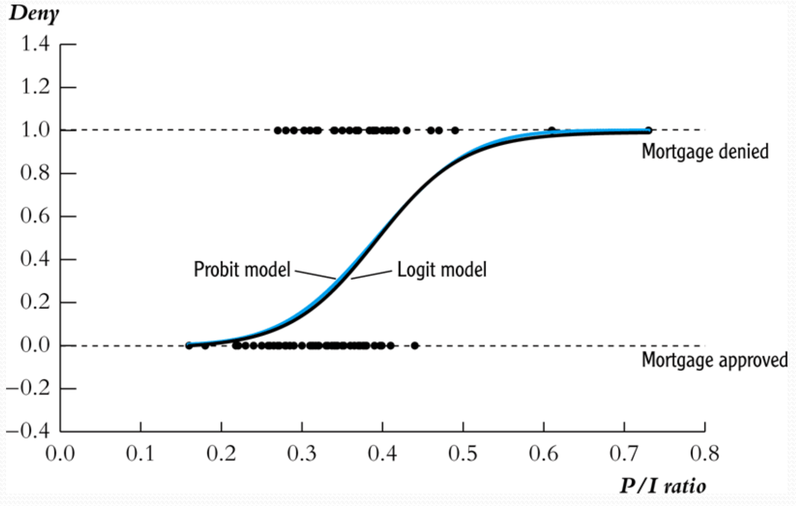
- La principal razón es histórica: la distribución logística se podía calcular más rápido que la distribución normal acumulada, pero desde que tenemos ordenadores, esto ya no es importante.
- En la práctica, Logit y Probit son muy similares, los resultados no dependen del modelo usado, por lo que ambos se pueden usar indistintamente.

Ejemplo de estimación logit: datos HMDA.

```
. logit deny p_irat black, r;
Iteration 0:   log likelihood =  -872.0853                Later…
Iteration 1:   log likelihood =  -806.3571
Iteration 2:   log likelihood = -795.74477
Iteration 3:   log likelihood = -795.69521
Iteration 4:   log likelihood = -795.69521
Logit estimates                                   Number of obs   =       2380
                                                  Wald chi2(2)    =     117.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -795.69521                       Pseudo R2       =     0.0876
------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      p_irat |   5.370362   .9633435     5.57   0.000     3.482244    7.258481
       black |   1.272782   .1460986     8.71   0.000     .9864339     1.55913
       _cons |  -4.125558    .345825   -11.93   0.000    -4.803362   -3.447753
------------------------------------------------------------------------------
.  dis "Pred prob, p_irat=.3, white: "                                        >
   1/(1+exp(-(_b[_cons]+_b[p_irat]*.3+_b[black]*0)));

Pred prob, p_irat=.3, white: .07485143
          NOTE:  the probit predicted probability is .07546603
```

Las probabilidades estimadas a partir de modelos logit y probit son muy parecidas en las regresiones con datos HMDA:



Estimación e Inferencia en los modelos Logit y Probit

Nos centraremos en el modelo probit:

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

- Estimación e inferencia, tres aspectos:
  - ¿Cómo podemos estimar β_0 y β_1?
  - ¿Cuál es la distribución muestral de los estimadores?
  - ¿Por qué podemos seguir usando los métodos habituales de inferencia?
- Primero justificaremos la estimación no lineal.
- Después hablaremos de la estimación máximo verosímil (que es lo que se hace en la práctica).

Estimación Probit por mínimos cuadrados no lineales

Recordemos, en la estimación MCO: $\min_{\beta_0,\beta_1} \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 X_i)]^2$

y obtenemos los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$

- *Los MCO no lineales se usan cuando tenemos parámetros que aparecen como no lineales:*

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n}[Y_i - \Phi(\beta_0 + \beta_1 X_i)]^2$$

¿Cómo resolver este problema de minimización?

- En la práctica, los MCO no lineales no se usan.
- Se usan algoritmos de minimización especiales, con la ayuda de ordenadores.
- Tenemos un estimador más eficiente (con menor varianza) que es…

El estimador Máximo Verosímil

- La **función de verosimilitud** es la distribución de probabilidad conjunta de los datos $Y_1,\ldots,Y_n$ dados $X_1,\ldots,X_n$, considerada como una función de los coeficientes desconocidos β_o y β_1.
- El estimador máximo verosímil (MV) es el valor de (β_o, β_1) que maximiza la función de verosimilitud.
- El estimador MV es el valor de (β_o, β_1) que mejor describe la distribución conjunta de los datos.
- En muestras grandes, el estimador MV es:
  - consistente
  - distribuido normalmente
  - eficiente (el estimador de menor varianza)

La función de verosimilitud Probit

$f(\beta_0, \beta_1; Y_1,\ldots,Y_n|X_1,\ldots,X_n) =$

$$= \{\ \Phi(\beta_0 + \beta_1 X_1)^{Y_1}[1-\Phi(\beta_0 + \beta_1 X_1)]^{1-Y_1}\}\times$$

$$\ldots\times\{\ \Phi(\beta_0 + \beta_1 X_n)^{Y_n}[1-\Phi(\beta_0 + \beta_1 X_n)]^{1-Y_n}\ \}$$

- $\hat{\beta}_0^{MLE}$, $\hat{\beta}_1^{MLE}$ maximizan esta función de verosimilitud.
- Pero lo anterior se hace con métodos complejos.
- Resultados en muestras grandes para $\hat{\beta}_0^{MLE}$ y $\hat{\beta}_1^{MLE}$
  - Son consistentes.
  - Se distribuyen normalmente.
  - Son asintóticamente eficientes (asumiendo que el modelo Probit es el modelo correcto).

Estimación del modelo Probit

- Las desviaciones típicas de $\hat{\beta}_0^{MLE}$ y $\hat{\beta}_1^{MLE}$ se calculan automáticamente.
- Los contrastes e intervalos de confianza se siguen realizando igual que siempre.
- Todo lo anterior se puede extender al caso con varias *X*'s.
- De forma análoga se puede hacer todo para el **caso Logit**, la única diferencia es la forma de la función de probabilidad, ya que ahora Φ se sustituye por la distribución acumulada logística. Las propiedades de los estimadores son las mismas.

Medidas de ajuste para Logit y Probit

El R^2 y R^2 corregido no tienen sentido porque al ser la dependiente de tipo binario, no podemos valorar que los datos se ajusten a la línea de regresión.

Dos medidas usadas para evaluar el ajuste son:

1. La **proporción correctamente estimada** = proporción de *Y*'s cuya probabilidad es >50% cuando Y_i=1, o es <50% cuando Y_i=0.
2. El ***pseudo-R^2*** mide el ajuste del modelo mediante la verosimilitud, comparándola respecto al caso en que no hubiera *X*'s.

Modelo de respuesta múltiple ordenada.

- Se usa cuando las decisiones de los individuos pueden ordenarse.
  - Por ejemplo, considerando la decisión de compra de una vivienda según el tamaño de la misma.

$$Y_i = \begin{cases} 0 & \text{si el tamaño es menor de 70 m}^2 \\ 1 & \text{si el tamaño es entre 70 y 90 m}^2 \\ 2 & \text{si el tamaño es mayor de 90 m}^2 \end{cases}$$

  - El individuo elige una alternativa si la utilidad que le produce es mayor que la del resto de alternativas.
  - Se estiman los coeficientes del modelo junto con unos umbrales que ordenan las opciones.

Modelos Probit y Logit multinomiales.

- Se usan cuando la variable Y toma más de dos alternativas.
- Se modelizan tantas ecuaciones como alternativas tiene Y.
- Para cada variable se estiman tantos parámetros como alternativas de Y menos una.
- Es necesario identificar una categoría de referencia.

35

Logit multinomial

- Por ejemplo, para el caso de 3 alternativas de Y (la primera es la que se toma de referencia) y 2 variables explicativas:

$$\Pr ob(Y_i = 1) = \frac{1}{1 + e^{a_2 + b_{12}X_{1i} + b_{22}X_{2i}} + e^{a_3 + b_{13}X_{1i} + b_{23}X_{2i}}}$$

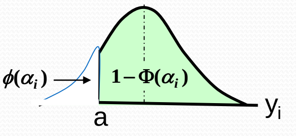
$$\Pr ob(Y_i = 2) = \frac{e^{a_2 + b_{12}X_{1i} + b_{22}X_{2i}}}{1 + e^{a_2 + b_{12}X_{1i} + b_{22}X_{2i}} + e^{a_3 + b_{13}X_{1i} + b_{23}X_{2i}}}$$

$$\Pr ob(Y_i = 3) = \frac{e^{a_3 + b_{13}X_{1i} + b_{23}X_{2i}}}{1 + e^{a_2 + b_{12}X_{1i} + b_{22}X_{2i}} + e^{a_3 + b_{13}X_{1i} + b_{23}X_{2i}}}$$

36

Otros modelos de variable limitada

- Modelo de variable censurada: los valores de la endógena superiores o inferiores a un determinado valor no se conocen, aunque sí se tiene información de los regresores.



- Modelo de variable truncada: no se dispone de información a partir de un valor tanto de la endógena como de los regresores.
- Modelo Tobit: la variable endógena toma valores positivos o cero según si Y > Y* o si Y <= Y*.

37

Resumen

- Si Y_i es <u>binaria</u>, entonces $E(Y| X) = \Pr(Y=1|X)$.
- La recta de regresión muestra la probabilidad de que Y=1, dados los valores de las X's.
- Tres modelos:
  - **Modelo de probabilidad lineal** (regresión lineal múltiple).
  - **Probit** (distribución acumulada normal estándar).
  - **Logit** (distribución acumulada logística).
- El efecto de un ΔX es el cambio en la probabilidad condicional de que Y=1. Para el caso Logit y Probit, esto depende del valor inicial de X.

Resumen

- Probit y Logit se estiman por máxima verosimilitud.
  - Los coeficientes se distribuyen asintóticamente como normales.
  - La inferencia es la habitual con tamaños muestrales grandes.
- Cuando la variable endógena no es binaria, sino que toma varios valores cualitativos, por ej. el medio de transporte elegido (a pie, coche, autobús, metro, tranvía), el análisis de estos modelos se basa en los fundamentos de la maximización de la utilidad.
  - Las probabilidades de elección se expresan en forma de Probit o Logit y a estos modelos se les llama *multinomiales*.