Microeconometría I

Maestría en Econometría

Lecture 3

Variables Categóricas

Agenda

Variables Categóricas No Ordenadas

- Hasta ahora hemos considerado modelos cuya variable dependiente es binaria. Sin embargo, en la práctica, son mucho más comúnes aquellos modelos que tratan de explicar una variable categórica con más de dos opciones.
- Por ejemplo, es usual tratar de explicar el tipo de trabajo que tiene un individuo sobre la base de características personales y del trabajo en si.
- Para extender el análisis que hemos hecho consideremos el caso de una variable dependiente categórica no ordenada.
- Es decir, ahora asumimos que un individuo i puede elegir entre J categorías, con $J > 2$.
- La elección del individuo i se denota por y_i, que ahora puede adoptar valores discretos: $1, 2, \ldots, J$.

Variables Categóricas No Ordenadas

- Como en los modelos de variable dependiente binaria, vamos a relacionar la elección de las categorías con diferentes variables explicativas.
- En general, tenemos tres tipos de variables explicativas:
  1. variables que son diferentes entre individuos pero iguales entre categorías (edad, ingreso, género, etc.). Vamos a llamar a estas variables x_i.
  2. variables diferentes entre individuos y además diferentes entre categorías (por ejemplo, en la elección de un tipo de trabajo una de estas variables podría ser el salario de cada tipo de trabajo j que enfrenta el individuo i). Llamaremos a estas variables $w_{i,j}$.
  3. variables iguales entre individuos pero diferentes entre categorías (por ejemplo, diferentes características de cada trabajo j). Llamaremos a estas variables z_j.

Modelo Multinomial

- Cuando la variable dependiente se explica utilizando solo el primer tipo de variables explicativas, x_i (edad, género, etc.) el modelo recibe el nombre de MULTINOMIAL.
- Un posible modelo en términos de utilidades estocásticas es,

$$U_i^j = x_i'\beta_j + \epsilon_{i,j}$$

Donde, U_i^j representa la utilidad que tiene para el individuo i seleccionar la categoría j. Las x_i representan diferentes características del individuo i; y los β_j ponderan las diferentes caraterísticas del individuo para dar la utilidad total de la categoría j. Los $\epsilon_{i,j}$ son las preferencias específicas del individuo i no modeladas (i.e. los errores del modelo).

Modelo Multinomial

- Se asume que el individuo i selecciona aquella categoría que le brinda mayor utilidad:

$$
\begin{aligned}
p_{i,j} &= Pr[y_i = j|\cdot] = Pr[U_i^j > U_i^k|\cdot] \ \forall k \neq j \\
&= Pr[x_i'\beta_j + \epsilon_{i,j} > x_i'\beta_k + \epsilon_{i,k}|\cdot] \ \forall k \neq j \\
&= Pr[\epsilon_{i,k} - \epsilon_{i,j} \leq x_i'(\beta_j - \beta_k)|\cdot] \ \forall k \neq j \\
&= Pr[\epsilon_i \leq x_i'(\beta_j - \beta_k)|\cdot] \ \forall k \neq j
\end{aligned}
$$

- Dependiendo de la función de distribución asumida (logística o normal) el modelo recibe el nombre de Logit Multinomial o Probit Multinomial.

Agenda

1. Variables Categóricas No Ordenadas
   - Clasificación de las Variables Explicativas
   - Modelo Logit Multinomial
   - Modelo Logit Condicional
   - Modelo Logit General
   - Modelo Probit Multinomial, Condicional y General
2. Variables Categóricas Ordenadas
   - Modelo Logit Ordenado
   - Modelo Probit Ordenado
3. Variables Categóricas Secuenciales
   - Modelo Logit Secuencial
   - Modelo Probit Secuencial

Modelo Logit Multinomial

- Asumiendo que $\epsilon_{i,j}$ tiene distribución doble exponencial, se puede demostrar (Amemiya, 1995 pg. 297):

$$p_{i,j} = Pr[y_i = j|\cdot] = \frac{e^{x_i'(\beta_j - \beta_k)}}{\sum_{l=1}^{J} e^{x_i'(\beta_l - \beta_k)}} \quad j = 1, 2, \ldots, J$$

  y $\sum_{j=1}^{J} p_{i,j} = 1$.
- La última condición asegura que cada individuo selecciona alguna de las categorías de la variable dependiente. Esto implica que una de las categorías actúa como base.

Modelo Logit Multinomial

- Es decir, en el denominador de la expresión anterior tenemos,

$$
\begin{aligned}
\sum_{l=1}^{J} e^{x_i'(\beta_l-\beta_k)} &= e^{x_i'(\beta_1-\beta_k)} + e^{x_i'(\beta_2-\beta_k)} + \cdots + e^{x_i'(\beta_k-\beta_k)} + \cdots \\
&+ e^{x_i'(\beta_J-\beta_k)} = 1 + \sum_{l=1}^{J} e^{x_i'(\beta_l-\beta_k)} \ \forall l \neq k.
\end{aligned}
$$

- Si pérdida de generalidad, asumamos que la categoría base es la categoría J.

Modelo Logit Multinomial

- Entonces, el modelo Logit multinomial queda,

$$p_{i,j} = Pr[y_i = j|\cdot] = \frac{e^{x_i' \alpha_j}}{1 + \sum_{l=1}^{J-1} e^{x_i' \alpha_l}} \quad j = 1, 2, \ldots, J-1$$

y

$$p_{i,J} = Pr[y_i = J|\cdot] = \frac{1}{1 + \sum_{l=1}^{J-1} e^{x_i' \alpha_l}}$$

donde $\alpha_j = \beta_j - \beta_J$ tiene la interpretación de ser el efecto del coeficiente de la categoría j sobre la categoría base.

- No es obvio que se pueda hacer una interpretación directa de estos parámetros. El efecto de x_i sobre la probabilidad de elegir la categoría j es claramente una función no lineal de los α_j.

Modelo Logit Multinomial: Interpretación

- El efecto de un cambio en x_i sobre la probabilidad de que se cumpla el evento analizado surge de la derivada parcial de $Pr[y_i = j|\cdot]$ con respecto a x_i:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial x_i} = Pr[y_i = j|\cdot]\{\alpha_j - \sum_{l=1}^{J-1} \alpha_l Pr[y_i = l|\cdot]\}$$

- El signo de la derivada depende ahora no solo del signo del coeficiente que acompaña a x_i, sino también del signo del término entre llaves.
- Este último punto contrasta con el modelo Logit para variables binarias donde las probabilidades eran monótonas crecientes o decrecientes.

Modelo Logit Multinomial: Interpretación

- Note que para $J = 2$ la derivada parcial anterior se reduce a:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial x_i} = Pr[y_i = j|\cdot]\{1 - Pr[y_i = j|\cdot]\}\alpha_j$$

- Además de la derivada parcial, la semi-elasticidad de x_i, también se utiliza para interpretar el modelo:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial x_i} x_i = Pr[y_i = j|\cdot]\{\alpha_j - \sum_{l=1}^{J-1} \alpha_l Pr[y_i = l|\cdot]\} x_i$$

- La semi-elasticidad mide el cambio en puntos porcentuales en la probabilidad de que la categoría j sea elegida ante un cambio porcentual en x_i.

Agenda

Modelo Logit Condicional

- En el modelo Logit multinomial, la elección de los individuos está relacionada con variables explicativas específicas de esos individuos.
- En otros casos, uno puede tener variables explicativas que adoptan diferentes valores para cada categoría. La versión del modelo Logit que se ajusta a este tipo de variables recibe el nombre de modelo Logit Condicional (McFadden 1973).
- En el modelo Logit condicional, la probabilidad de que el individuo i seleccione la categoría j es:

$$p_{i,j} = Pr[y_i = j|\cdot] = \frac{e^{\gamma_0 + w_{i,j}'\beta}}{\sum_{l=1}^{J} e^{\gamma_0 + w_{i,l}'\beta}} \quad j = 1, 2, \ldots, J$$

y $\sum_{j=1}^{J} p_{i,j} = 1$.

Modelo Logit Condicional: Interpretación

- Igual que antes, fijamos a J como categoría base. Al hacerlo, la constante del modelo se hace igual a cero, $\gamma_0 = 0$; y las probabilidades del modelo condicional quedan:

$$p_{i,j} = Pr[y_i = j|\cdot] = \frac{e^{(w_{i,j}-w_{i,J})'\beta}}{1+\sum_{l=1}^{J-1} e^{(w_{i,l}-w_{i,J})'\beta}} \quad j = 1,\, 2,\, \ldots,\, J-1$$

y

$$p_{i,J} = Pr[y_i = J|\cdot] = \frac{1}{1+\sum_{l=1}^{J-1} e^{(w_{i,l}-w_{i,J})'\beta}}$$

- Como sucedía en el modelo multinomial, las probabilidades son funciones no lineales de los parámetros del modelo y por lo tanto la interpretación de los mismos no es directa.

Modelo Logit Condicional: Interpretación

- El efecto de un cambio en $w_{i,j}$ sobre la probabilidad de selección surge de la derivada parcial de $Pr[y_i = j|\cdot]$ con respecto a $w_{i,j}$:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial w_{i,j}} = Pr[y_i = j|\cdot]\{1 - Pr[y_i = j|\cdot]\}\beta$$

- La derivada parcial de la probabilidad de que el individuo i elija la categoría j con respecto a $w_{i,l}$ para $l \neq j$. Esto es,

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial w_{i,l}} = -Pr[y_i = j|\cdot]Pr[y_i = l|\cdot]\beta$$

- El signo de estas derivadas está determinado por el signo de β y entonces la probabilidad varía monotónicamente.

Modelo Logit Condicional: Interpretación

- En el ejemplo de la elección de diferentes tipos de trabajo donde $w_{i,j}$ es el salario del trabajo j que enfrenta la persona i, las dos derivadas anteriores muestran que para $\beta > 0$, un incremento en el salario del trabajo j incrementa la probabilidad de elección de j y disminuye la probabilidad de elección de otros trabajos ($l \neq j$).
- De las dos derivadas anteriores surgen las semi-elasticidades directas y cruzadas. El cambio porcentual en la probabilidad de que la categoría j sea elegida ante un cambio porcentual en $w_{i,j}$ es:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial w_{i,j}} w_{i,j} = Pr[y_i = j|\cdot]\{1 - Pr[y_i = j|\cdot]\} w_{i,j}\beta$$

Modelo Logit Condicional: Interpretación

- El cambio porcentual en la probabilidad de elegir la categoría j debido a un cambio porcentual en $w_{i,l}$ es:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial w_{i,l}} w_{i,l} = -Pr[y_i = j|\cdot] Pr[y_i = l|\cdot] w_{i,l} \beta$$

Agenda

Modelo Logit General

- Una especificación más general se da cuando se combinan los dos modelos anteriores (el multinomial y el condicional) y se agrega el último tipo de variable explicativa. Esto es, variables z_j que adoptan valores diferentes entre categorías pero iguales entre individuos.
- Por lo tanto, el modelo especificado queda:

$$p_{i,j} = Pr[y_i = j|\cdot] = \frac{e^{x_i'\alpha_j + \gamma_0 + w_{i,j}'\beta + z_j'\delta}}{\sum_{l=1}^{J} e^{x_i'\alpha_l + \gamma_0 + w_{i,l}'\beta + z_l'\delta}} \quad j = 1, 2, \ldots, J$$

y $\sum_{j=1}^{J} p_{i,j} = 1$.

Modelo Logit General

- Igual que en los dos modelos anteriores, para identificar los parámetros del modelo hay que tomar una categoría como base. Fijando a J como esta categoría las probabilidades del modelo general quedan:

$$p_{i,j} = Pr[y_i = j|\cdot] = \frac{e^{x_i'\alpha_j + (w_{i,j} - w_{i,J})'\beta + (z_j - z_J)'\delta}}{1 + \sum_{l=1}^{J-1} e^{x_i'\alpha_l + (w_{i,l} - w_{i,J})'\beta + (z_l - z_J)'\delta}}$$

para $j = 1, 2, \ldots, J - 1$, y

$$p_{i,J} = Pr[y_i = J|\cdot] = \frac{1}{1 + \sum_{l=1}^{J-1} e^{x_i'\alpha_l + (w_{i,l} - w_{i,J})'\beta + (z_l - z_J)'\delta}}$$

Modelo Logit General: Interpretación

- Como sucedía en los modelos anteriores, las probabilidades son funciones no lineales de los parámetros del modelo y por lo tanto la interpretación de los mismos no es directa.
- El efecto de un cambio en z_j sobre la probabilidad de selección surge de la derivada parcial de $Pr[y_i = j|\cdot]$ con respecto a z_j:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial z_j} = Pr[y_i = j|\cdot]\{1 - Pr[y_i = j|\cdot]\}\delta$$

- El signo de esta derivada está totalmente determinado por el signo de δ y por lo tanto la probabilidad varía monotónicamente.

Modelo Logit General: Interpretación

- El cambio porcentual en la probabilidad de que la categoría j sea elegida ante un cambio porcentual en z_j determina la cuasi-elasticidad:

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial z_j} z_j = Pr[y_i = j|\cdot]\{1 - Pr[y_i = j|\cdot]\} z_j \delta$$

Modelo Logit General: Interpretación

- Para cualquiera de los tres modelos se puede definir la tasa de probabilidad entre dos alternativas j y k.
- En el modelo multinomial, el logaritmo natural de la tasa de probabilidad entre j y k viene dado por,

$$log\left(\frac{Pr[y_i = j|\cdot]}{Pr[y_i = k|\cdot]}\right) = x_i'(\alpha_j - \alpha_k)$$

- En el modelo condicional, el logaritmo natural de la tasa de probabilidad entre j y k viene dado por,

$$log\left(\frac{Pr[y_i = j|\cdot]}{Pr[y_i = k|\cdot]}\right) = (w_{i,j} - w_{i,k})'\beta$$

Modelo Logit General: Interpretación

- En el modelo general, el logaritmo natural de la tasa de probabilidad entre j y k viene dado por,

$$log\left(\frac{Pr[y_i = j|\cdot]}{Pr[y_i = k|\cdot]}\right) = x_i'(\alpha_j - \alpha_k) + (w_{i,j} - w_{i,k})'\beta + (z_j - z_k)'\delta$$

- Note que la tasa de probabilidad entre la elección de las alternativas j y k no está afectada por el resto de las alternativas.
- Esta propiedad de los tres modelos se conoce como la independencia de alternativas irrelevantes.
- Esto es, comparando las alternativas j y k, las otras opciones son irrelevantes.

Modelo Logit General: Estimación

- La estimación de los parámetros de cualquiera de los modelos descriptos se realiza mediante el método de maximización de la función de verosimilitud.
- La función de verosimilitud para cualquiera de los tres modelos anteriores es la misma, excepto porque la forma funcional de la probabilidad es diferente.
- La función de verosimilitud viene dada por:

$$L(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{J} Pr[y_i = j|\cdot]^{I[y_i=j]}$$

donde $I[\cdot]$ es la función indicador que asume el valor 1 cuando el argumento de la función es verdadero y asume el valor 0 cuando el argumento de la función es falso.

Modelo Logit General: Estimación

- El logaritmo de la función de verosimilitud es:

$$l(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} I[y_i = j] log(Pr[y_i = j|\cdot])$$

- El estimador de máxima verosimilitud del modelo es el valor de $\hat{\theta}$ que maximiza la función anterior.
- Las condiciones de primer orden para un máximo vienen dadas por,

$$\begin{aligned} \frac{\partial l(\hat{\theta})}{\partial \hat{\theta}} &= \sum_{i=1}^{n} \sum_{j=1}^{J} I[y_i = j] \frac{\partial log(Pr[y_i = j|\cdot])}{\partial \hat{\theta}} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{J} \frac{I[y_i = j]}{Pr[y_i = j|\cdot]} \frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\theta}} = 0 \qquad (1) \end{aligned}$$

Modelo Logit General: Estimación

- Para el modelo general $\hat{\theta}' = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots \hat{\alpha}_{J-1}, \hat{\beta}', \hat{\delta}')$.
- Derivando con respecto a $\hat{\alpha}_j$ tenemos,

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\alpha}_j} = Pr[y_i = j|\cdot](1 - Pr[y_i = j|\cdot])x_i \quad j = 1\,2, \ldots, J-1$$

y

$$\frac{\partial Pr[y_i = l|\cdot]}{\partial \hat{\alpha}_j} = -Pr[y_i = j|\cdot]Pr[y_i = l|\cdot]x_i \quad j = 1\,2, \ldots, J-1 \neq l$$

Modelo Logit General: Estimación

- Reemplazando en (1),

$$
\begin{aligned}
\frac{\partial l(\hat{\theta})}{\partial \hat{\alpha}_j} &= \sum_{i=1}^{n} \sum_{j=1}^{J} \frac{I[y_i = j]}{Pr[y_i = j|\cdot]} \frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\alpha}_j} \\
&= \sum_{i=1}^{n} \left\{ \frac{I[y_i = j]}{Pr[y_i = j|\cdot]} \frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\alpha}_j} \right. \\
&+ \left. \sum_{l \neq j} \frac{I[y_i = l]}{Pr[y_i = l|\cdot]} \frac{\partial Pr[y_i = l|\cdot]}{\partial \hat{\alpha}_j} \right\} \\
&= \sum_{i=1}^{n} \left(I[y_i = j] - Pr[y_i = j|\cdot] \right) x_i = 0 \qquad (2)
\end{aligned}
$$

Modelo Logit General: Estimación

- Derivando con respecto a $\hat{\beta}$ tenemos,

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\beta}} = Pr[y_i = j|\cdot](w_{i,j} - \sum_{l=1}^{J-1} Pr[y_i = l|\cdot]w_{i,l})$$

- Reemplazando en (1),

$$\begin{aligned} \frac{\partial l(\hat{\theta})}{\partial \hat{\beta}} &= \sum_{i=1}^{n}\sum_{j=1}^{J} \frac{I[y_i = j]}{Pr[y_i = j|\cdot]} \frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\beta}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} I[y_i = j]\left(w_{i,j} - \sum_{l=1}^{J-1} Pr[y_i = l|\cdot]w_{i,l}\right) = 0 \end{aligned} \tag{3}$$

Modelo Logit General: Estimación

- Finalmente, derivando con respecto a $\hat{\delta}$ tenemos,

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\delta}} = Pr[y_i = j|\cdot](z_j - \sum_{l=1}^{J-1} Pr[y_i = l|\cdot] z_l)$$

- Reemplazando en (1),

$$\begin{aligned}\frac{\partial l(\hat{\theta})}{\partial \hat{\delta}} &= \sum_{i=1}^{n}\sum_{j=1}^{J} \frac{I[y_i = j]}{Pr[y_i = j|\cdot]} \frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\delta}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} I[y_i = j] \left(z_j - \sum_{l=1}^{J-1} Pr[y_i = l|\cdot] z_l \right) = 0\end{aligned} \tag{4}$$

Modelo Logit General: Estimación

- Las ecuaciones (2), (3) y (4) representan las condiciones de primer orden para la maximización de la función de verosimilitud.

$$\frac{\partial l(\hat{\theta})}{\partial \hat{\alpha}_j} = \sum_{i=1}^{n} \left(I[y_i = j] - Pr[y_i = j|\cdot] \right) x_i = 0 \quad j = 1, 2, \ldots, J-1$$

$$\frac{\partial l(\hat{\theta})}{\partial \hat{\beta}} = \sum_{i=1}^{n} \sum_{j=1}^{J} I[y_i = j] \left(w_{i,j} - \sum_{l=1}^{J-1} Pr[y_i = l|\cdot] w_{i,l} \right) = 0$$

$$\frac{\partial l(\hat{\theta})}{\partial \hat{\delta}} = \sum_{i=1}^{n} \sum_{j=1}^{J} I[y_i = j] \left(z_j - \sum_{l=1}^{J-1} Pr[y_i = l|\cdot] z_l \right) = 0$$

Modelo Logit General: Estimación

- Es inmediatamente obvio que no se pueden resolver las condiciones de primer orden despejando las incógnitas por tratarse de ecuaciones no lineales en las mismas.
- Para obtener los estimadores de máxima verosimilitud del modelo hay que recurrir a algoritmos no lineales.
- El procedimiento de maximización requiere que se cumplan las condiciones de segundo orden representadas por la siguiente matriz Hesiana:

Modelo Logit General: Estimación

- $$H(\hat{\theta}) = \begin{bmatrix} \frac{\partial^2 l(\cdot)}{\partial \alpha_1^2} & \cdots & \frac{\partial^2 l(\cdot)}{\partial \alpha_1 \partial \alpha_{J-1}} & \frac{\partial^2 l(\cdot)}{\partial \alpha_1 \partial \beta} & \frac{\partial^2 l(\cdot)}{\partial \alpha_1 \partial \delta} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial^2 l(\cdot)}{\partial \alpha_{J-1} \partial \alpha_1} & \cdots & \frac{\partial^2 l(\cdot)}{\partial \alpha_{J-1}^2} & \frac{\partial^2 l(\cdot)}{\partial \alpha_{J-1} \partial \beta} & \frac{\partial^2 l(\cdot)}{\partial \alpha_{J-1} \partial \delta} \\ \frac{\partial^2 l(\cdot)}{\partial \beta \partial \alpha_1} & \cdots & \frac{\partial^2 l(\cdot)}{\partial \beta \partial \alpha_{J-1}} & \frac{\partial^2 l(\cdot)}{\partial \beta^2} & \frac{\partial^2 l(\cdot)}{\partial \beta \partial \delta} \\ \frac{\partial^2 l(\cdot)}{\partial \delta \partial \alpha_1} & \cdots & \frac{\partial^2 l(\cdot)}{\partial \delta \partial \alpha_{J-1}} & \frac{\partial^2 l(\cdot)}{\partial \delta \partial \beta} & \frac{\partial^2 l(\cdot)}{\partial \delta^2} \end{bmatrix}$$
- Puede demostrarse que la función de verosimilitud es globalmente cóncava y por lo tanto las condiciones de segundo orden se satisfacen.
- El estimador de máxima verosimilitud del vector θ es insesgado, consistente y eficiente.

Modelo Logit General: Evaluación

- El estimador de máxima verosimilitud tiene distribución normal,

$$\hat{\theta} \xrightarrow{d} N(\theta, [E\{-H(\hat{\theta})\}]^{-1})$$

- Como medidas de bondad del ajuste se pueden usar los $Pseudo - R^2$, como el R^2_{MF} de McFadden:

$$R^2_{MF} = 1 - \frac{l(\hat{\theta})}{l(\hat{\theta}_0)}$$

- Donde, $l(\hat{\theta})$ es el valor del logaritmo de la función de verosimilitud evaluada en los estimadores de MV y $l(\hat{\theta}_0)$ es el valor del logaritmo de la función de verosimilitud de un modelo que tiene solo una constante.

Modelo Logit General: Evaluación

- Otra forma de medir la bondad del ajuste es observar como clasifica a las observaciones el modelo en comparación con los datos realmente observados.
- Para generar esta clasificación debemos estimar las probabilidades de seleccionar cada categoría:

$$\hat{p}_{i,j} = Pr\widehat{[y_i = j}|\cdot] = \frac{e^{x_i'\hat{\alpha}_j + (w_{i,j}-w_{i,J})'\hat{\beta} + (z_j - z_J)'\hat{\delta}}}{1 + \sum_{l=1}^{J-1} e^{x_i'\hat{\alpha}_j + (w_{i,j}-w_{i,J})'\hat{\beta} + (z_j - z_J)'\hat{\delta}}}$$

para $j = 1, 2, \ldots, J - 1$, y

$$\hat{p}_{i,J} = Pr\widehat{[y_i = J}|\cdot] = \frac{1}{1 + \sum_{l=1}^{J-1} e^{x_i'\hat{\alpha}_j + (w_{i,j}-w_{i,J})'\hat{\beta} + (z_j - z_J)'\hat{\delta}}}$$

- El siguiente paso consiste en trasladar estas estimaciones en elecciones discretas.

Modelo Logit General: Evaluación

- En la práctica la regla es asignarle a $\hat{y}_i$ el valor j correspondiente a la categoría con la mayor probabilidad. Esto es,

$$\hat{y}_i = j \quad si \;\; \hat{p}_{i,j} = max\{\hat{p}_{i,1},\; \hat{p}_{i,2},\; \ldots \hat{p}_{i,J}\}$$

- Después se puede construir la tabla de predicción-realización.

	Predicción				
Realización	$\hat{y}_i = 1$	$\cdots$	$\hat{y}_i = j$	$\cdots$	$\hat{y}_i = J$
$y_i = 1$	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1J}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$y_i = j$	p_{j1}	$\cdots$	p_{jj}	$\cdots$	p_{jJ}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$y_i = J$	p_{J1}	$\cdots$	p_{Jj}	$\cdots$	p_{JJ}

Modelo Logit General: Evaluación

- La proporción $h = p_{11} + \cdots + p_{jj} + \cdots + p_{JJ}$ se intrepreta como la tasa de aciertos.
- La tasa de aciertos puede ser comparada con una predicción totalmente aleatoria, donde para cada individuo i, la alternativa j se predice con probabilidad igual a la proporción de observaciones en la categoría j en la muestra: n_j/n. La tasa esperada de aciertos con estas predicciones aleatorias es $\hat{q} = \sum_{j=1}^{J}(n_j/n)^2$.
- El modelo general tiene mejores predicciones que las hechas aleatoriamente siempre que

$$z = \frac{h - \hat{q}}{\sqrt{\hat{q}(1-\hat{q})/n}} = \frac{nh - n\hat{q}}{\sqrt{n\hat{q}(1-\hat{q})}}$$

es lo suficientemente grande (mayor a 1.645 al 5% de significación).

Agenda

Modelo Probit Multinomial, Condicional y General

- Finalmente, si en los tres modelos desarrollados, el Multinomial, el Condicional y el General en lugar de $F(\cdot)$ y $f(\cdot)$ se utilizan $\Phi(\cdot)$ y $\phi(\cdot)$ tenemos los modelos Probit Multinomial, Condicional y General, respectivamente.

Agenda

Modelo Logit Ordenado

- Ahora nuestro interés se basa en una variable dependiente discreta ordenada.
- Este tipo de variables surge frecuentemente en marketing cuando una empresa, antes de lanzar un producto al mercado, realiza un estudio para que distintas personas evalúen las características del mismo.
- A las personas se les pide, por ejemplo, que indiquen si el producto les gusta mucho, poco o nada. En este caso tenemos tres categorías ordenadas.
- Otro caso surge cuando los individuos mismos son asignados a categorías ordenadas de acuerdo a su actitud acerca de cierto fenómeno y el objetivo del investigador es ver que variables explican la clasificación de los individuos en esas categorías.

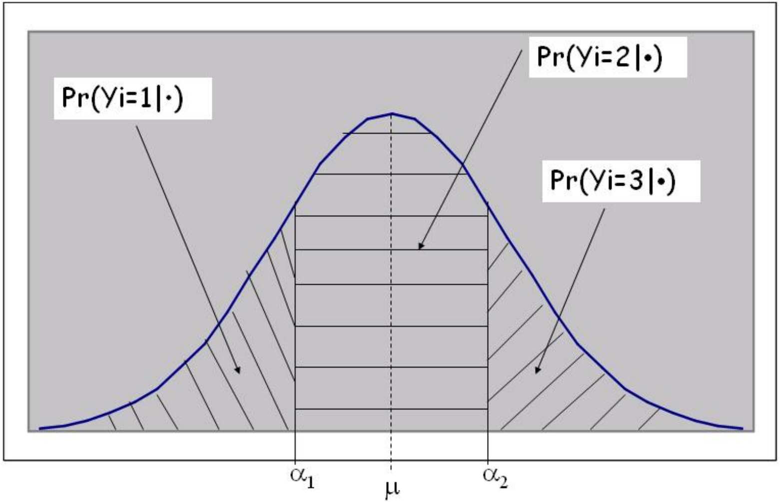
Modelo Logit Ordenado

- Por ejemplo, personas que son clientes del mismo banco o compañía financiera y que son asignadas a tres categorías de acuerdo a su perfil de riesgo.
- Aquellos que solo tienen plazos fijos corresponden a una categoría de bajo riesgo; aquellas personas que compran bonos del gobierno pertenecen a una categoría de mayor riesgo; y aquellos que invierten en la bolsa son las de más alto riesgo.
- Tal como hicimos con los otros modelos asumamos que la variable y_i representa las diferentes categorías a las que, por ejemplo, pueden asignarse los individuos.
- En nuestro caso de los clientes del banco y_i adoptará tres valores: 1 si es un cliente que asume poco riesgo, 2 si es un cliente que asume algo de riesgo y 3 si es un cliente que asume mucho riesgo en sus inversiones.

Modelo Logit Ordenado

- El objetivo de nuestro análisis es relacionar esta variable categórica ordenada con algunas variables explicativas.
- Tal como hicimos en los casos anteriores el ajuste se realizará asumiendo un modelo de probabilidad para la elección de la categoría j.
- A diferencia de lo que ocurría con los modelos multinomiales y condicionales ahora debemos preservar el orden de las categorías de la variable dependiente al calcular las probabilidades.
- Esto se puede lograr, por ejemplo, dividiendo a la distribución de probabilidad que se decida usar en partes ordenadas como sigue:

Modelo Logit Ordenado



Modelo Logit Ordenado

- La figura anterior asigna probabilidades a tres categorías. La generalización al caso de J categorías se realiza de la siguiente manera:
  - Definamos los valores $\alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{J-1} < \alpha_J$.
  - Usualmente los límites inferior y superior se determinan como $\alpha_0 = -\infty$ y $\alpha_J = +\infty$.
  - Las categorías se asignan de la siguiente manera: $\alpha_{j-1} < y_i < \alpha_j$, implica $y_i = j$, $j = 1, 2, \ldots, J$.
- Esto es,

$$
\begin{aligned}
Pr[y_i = j|\cdot] &= Pr[\alpha_{j-1} < y_i < \alpha_j] = Pr[\alpha_{j-1} < x_i'\beta + u_i < \alpha_j] \\
&= Pr[\alpha_{j-1} - x_i'\beta < u_i < \alpha_j - x_i'\beta] \\
&= F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta)
\end{aligned}
$$

para $j = 1,\ 2,\ 3,\ \ldots,\ J$.

Modelo Logit Ordenado

- Note que para que los parámetros de este modelo estén identificados las x_i no deben tener un término constante.
- Cuando la función de distribución utilizada es la logística, el modelo recibe el nombre de Logit Ordenado.
- En este caso,

$$F(\alpha_j - x_i'\beta) = \frac{e^{\alpha_j - x_i'\beta}}{\sum_{j=1}^{J} e^{\alpha_j - x_i'\beta}}$$

- No es obvio que se pueda hacer una interpretación directa de los parámetros del modelo ordenado. El efecto de x_i sobre la probabilidad de elegir la categoría j es claramente una función no lineal de los α_j y los β.

Modelo Logit Ordenado

- La derivada parcial de la probabilidad de selección con respecto a alguna variable explicativa es:

$$\begin{aligned}\frac{\partial Pr[y_i = j|\cdot]}{\partial x_i} &= \frac{\partial F(\alpha_j - x_i'\beta)}{\partial x_i} - \frac{\partial F(\alpha_{j-1} - x_i'\beta)}{\partial x_i} \\ &= [f(\alpha_{j-1} - x_i'\beta) - f(\alpha_j - x_i'\beta)]\beta\end{aligned}$$

  donde $f(\cdot)$ denota la función de densidad de la distribución logística.
- Puede verse que esta última expresión no sólo depende del signo de β, sino también del valor de $f(\alpha_{j-1} - x_i'\beta) - f(\alpha_j - x_i'\beta)$.

Modelo Logit Ordenado

- Como en los modelos anteriores la estimación se realiza mediante la maximización de la función de verosimilitud.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n}\prod_{j=1}^{J} Pr[y_i = j|\cdot]^{I[y_i=j]} \\
&= \prod_{i=1}^{n}\prod_{j=1}^{J} [F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta)]^{I[y_i=j]}
\end{aligned}
$$

- Aquí, θ representa a $(\alpha_1, \alpha_2, \ldots, \alpha_{J-1})$ y β. La función indicador $I(y_i = j)$ adopta el valor 1 cuando $y_i = j$ y 0 para los demás casos.

Modelo Logit Ordenado

- El logaritmo de la FV es,

$$l(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} I[y_i = j] log\{F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta)\}$$

- El estimador de máxima verosimilitud del modelo es el valor de $\hat{\theta}$ que maximiza la función anterior.

Modelo Logit Ordenado

- Las condiciones de primer orden para un máximo vienen dadas por,

$$\begin{aligned} \frac{\partial l(\hat{\theta})}{\partial \hat{\theta}} &= \sum_{i=1}^{n}\sum_{j=1}^{J} I[y_i = j]\frac{\partial log(Pr[y_i = j|\cdot])}{\partial \hat{\theta}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} \frac{I[y_i = j]}{Pr[y_i = j|\cdot]}\frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\theta}} = 0 \end{aligned}$$

- Donde,

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\beta}} = [f(\alpha_{j-1} - x_i'\beta) - f(\alpha_j - x_i'\beta)]x_i$$

Modelo Logit Ordenado

- Y,

$$\frac{\partial Pr[y_i = j|\cdot]}{\partial \hat{\alpha}_s} = \begin{cases} f(\alpha_s - x_i'\beta) & \text{si } s = j \\ -f(\alpha_s - x_i'\beta) & \text{si } s = j-1 \\ 0 & \text{en otros casos.} \end{cases}$$

- Reemplazando las expresiones de las derivadas de la probabilidad en la derivada del logaritmo de la función de verosimilitud se obtienen las condiciones de primer orden para el Logit ordenado.
- La función de verosimilitud es globalmente cóncava por lo tanto las condiciones de segundo orden se cumplen.
- Como pasaba en los modelos anteriores, aquí se pueden utilizar como medidas de bondad del ajuste los denominados $Pseudo - R^2$.

Modelo Logit Ordenado

- Uno de esos estadísticos es el R^2_{MF} de McFadden:

$$R^2_{MF} = 1 - \frac{l(\hat{\theta})}{l(\hat{\theta}_0)}$$

- Donde, $l(\hat{\theta})$ es el valor del logaritmo de la función de verosimilitud evaluada en los estimadores de MV y $l(\hat{\theta}_0)$ es el valor del logaritmo de la función de verosimilitud de un modelo que tiene solo una constante.
- Otra forma de medir la bondad del ajuste es observar como clasifica a las observaciones el modelo en comparación con los datos realmente observados.

Modelo Logit Ordenado

- Para generar esta clasificación debemos estimar las probabilidades de seleccionar cada categoría:

$$\hat{p}_{i,j} = Pr\widehat{[y_i = j}|\cdot] = F(\hat{\alpha}_j - x_i'\hat{\beta}) - F(\hat{\alpha}_{j-1} - x_i'\hat{\beta})$$

- El siguiente paso consiste en trasladar estas estimaciones en elecciones discretas.
- En la práctica la regla es asignarle a $\hat{y}_i$ el valor j correspondiente a la categoría con la mayor probabilidad. Esto es,

$$\hat{y}_i = j \quad si \;\; \hat{p}_{i,j} = max\{\hat{p}_{i,1},\, \hat{p}_{i,2},\, \ldots \hat{p}_{i,J}\}$$

Modelo Logit Ordenado

- Una vez estimados estos valores se puede construir la tabla de predicción-realización como hacíamos en el caso de una variable binaria.

	Predicción				
Realización	$\hat{y}_i = 1$	$\cdots$	$\hat{y}_i = j$	$\cdots$	$\hat{y}_i = J$
$y_i = 1$	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1J}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$y_i = j$	p_{j1}	$\cdots$	p_{jj}	$\cdots$	p_{jJ}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$y_i = J$	p_{J1}	$\cdots$	p_{Jj}	$\cdots$	p_{JJ}

- La proporción $h = p_{11} + \cdots + p_{jj} + \cdots + p_{JJ}$ se intrepreta como la tasa de aciertos.

Modelo Logit Ordenado

- La tasa de aciertos puede ser comparada con una predicción totalmente aleatoria, donde para cada individuo i, la alternativa j se predice con probabilidad igual a la proporción de observaciones en la categoría j en la muestra: n_j/n. La tasa esperada de aciertos con estas predicciones aleatorias es $\hat{q} = \sum_{j=1}^{J}(n_j/n)^2$.
- El modelo general tiene mejores predicciones que las hechas aleatoriamente siempre que

$$z = \frac{h - \hat{q}}{\sqrt{\hat{q}(1-\hat{q})/n}} = \frac{nh - n\hat{q}}{\sqrt{n\hat{q}(1-\hat{q})}}$$

es lo suficientemente grande (mayor a 1.645 al 5% de significación).

Agenda

Modelo Probit Ordenado

- Finalmente, si en lugar de $F(\cdot)$ y $f(\cdot)$ se utilizan $\Phi(\cdot)$ y $\phi(\cdot)$ tenemos el modelo Probit Ordenado.

Agenda

Modelo Logit Secuencial

- Este es un caso fácil de analizar porque la maximización de la función de verosimilitud en este caso particular puede realizarse maximizando la función de verosimilitud de variables binarias que ya estudiamos.
- A modo de ilustración, considere que los individuos pueden ser clasificados en cuatro categorías educativas como sigue:
  - $y_i = 1$ si la persona i tiene hasta educación primaria completa.
  - $y_i = 2$ si la persona i tiene más que educación primaria completa y hasta secundaria completa.
  - $y_i = 3$ si la persona i tiene más que educación secundaria completa y hasta universitaria completa.
  - $y_i = 4$ si la persona i tiene educación post-universitaria.
- Entonces las probabilidades correspondientes a las distintas categorías son:

Modelo Logit Secuencial

- $$\begin{aligned} Pr[y_i = 1|\cdot] &= F(x_i'\beta_1) \\ Pr[y_i = 2|\cdot] &= [1 - F(x_i'\beta_1)] \times F(x_i'\beta_2) \\ Pr[y_i = 3|\cdot] &= [1 - F(x_i'\beta_1)] \times [1 - F(x_i'\beta_2)] \times F(x_i'\beta_3) \\ Pr[y_i = 4|\cdot] &= [1 - F(x_i'\beta_1)] \times [1 - F(x_i'\beta_1)] \times [1 - F(x_i'\beta_3)] \end{aligned}$$
- Entonces, β_1 puede estimarse maximizando un modelo de variable binaria definida adoptando el valor 1 para aquellos que tienen hasta primaria completa y 0 para el resto.
- β_2 puede estimarse maximizando un modelo de variable binaria para una sub-muestra integrada por todas las personas que terminaron la escuela primaria.

Modelo Logit Secuencial

- Para estas personas se define una variable binaria que adopta el valor 1 para aquellos que tienen educación hasta secundaria completa y 0 para el resto.
- β_3 puede estimarse maximizando un modelo de variable binaria para una sub-muestra integrada por todas las personas que terminaron el colegio secundario.
- Para estas personas se define una variable binaria que adopta el valor 1 para aquellos que tienen educación hasta universitaria completa y 0 para el resto.

Agenda

Modelo Probit Secuencial

- Si las probabilidades se definen con $\Phi(\cdot)$ en lugar de $F(\cdot)$ tenemos el modelo Probit Secuencial.