SP1639-Análisis de Datos Categóricos.

Modelos de Escogencia Discreta.

Como les he comentado en clases anteriores, los modelos de escogencia discreta son modelos con variable dependiente categórica. Pero a diferencia de los modelos logísticos comunes y corrientes que ustedes conocen, las variables independientes no se refieren a características de las unidades estadísticas, sino:

A CARACTERÍSTICAS DE LAS CATEGORÍAS DE LA VARIABLE RESPUESTA.

Ejemplo:

Imagínese que a usted lo contrata la Cervecería Costa Rica, y tiene que averiguar características que hacen que la gente tome Imperial, Pilsen o Bavaria Negra. Imagínese que usted plantea empezar la investigación con los clientes de algún bar cercano a la Universidad, un viernes por la noche.

Una característica que puede hacer que la gente escoja entre Imperial, Pilsen o Bavaria Negra es si el cliente es estudiante de Bachillerato, estudiante de Posgrado, o profesor. Suponga que tiene esta tabla:

	F abs Marca			F relat Marca			
Tipo de cliente	Imperial	Pilsen	Bavaria Negra	Imperial	Pilsen	Bavaria Negra	Total
Est Bach.	50	50	20	41.7	41.7	16.7	100.0
Est. Posgrado	40	20	20	50.0	25.0	25.0	100.0
Profesor	15	5	30	30.0	10.0	60.0	100.0
Total	105	75	70	42.0	30.0	28.0	100.0

Para analizar estos datos, se tiene que utilizar una regresión logística común y corriente como se les ha enseñado en otros cursos.

Pero, como ustedes saben, en el consumo de cervezas, no sólo importan las características del consumidor (segmentación de mercado), sino también las características del producto. Se puede también tener el mismo cuadro, pero con los precios de la cerveza.

Precio de cerveza	F abs Marca Imperial	Pilsen	Bavaria Negra	F relat Marca Imperial	Pilsen	Bavaria Negra	Total
800.00 col	105	75	0	58.3	41.7	0.0	100.0
1000.00 col	0	0	70	0.0	0.0	100.0	100.0
Total	105	75	70	42.0	30.0	28.0	100.0

El precio es una característica del producto o marca, no del consumidor, así que el efecto del precio sobre el consumo de ciertas marcas de cerveza se analiza con modelos de escogencia discreta.

De hecho, si se desea analizar la interacción entre el precio y las características de los consumidores, esto se hace también con modelos de escogencia discreta. La tabla sería igual a la primera, pero tomando en cuenta el precio.

Tipo de cliente	F abs Marca Imperial 800 col	Pilsen 800 col	Bavaria Negra 1000 col	F relat Marca Imperial 800 col	Pilsen 800 col	Bavaria Negra 1000 col	Total
Est Bach.	50	50	20	41.7	41.7	16.7	100.0
Est. Posgrado	40	20	20	50.0	25.0	25.0	100.0
Profesor	15	5	30	30.0	10.0	60.0	100.0
Total	105	75	70	42.0	30.0	28.0	100.0

La teoría de los modelos de escogencia discreta.

El modelo considera respuestas observadas en un conjunto de alternativas:

- discretas,
- mutuamente excluyentes
- y exhaustivas
- j=1,2,...,J

El modelo supone que los sujetos que escogen las alternativas en forma determinística según las utilidades que los individuos perciben de cada alternativa:

U_{nj}, $n=1,2,..., N$, $j=1,2,..., J$

Además de las respuestas observadas, podemos tener también características medidas de cada una de las alternativas, y algunas características medidas de los sujetos.

Especificamos una función de estos dos tipos de características, el cual representa la UTILIDAD REPRESENTATIVA, o la PARTE SISTEMÁTICA DE LA UTILIDAD:

$$V_{nj} = V(x_{nj}, s_n) \quad \forall\, j$$

La utilidad se especifica como la suma de las partes sistemáticas, más una parte aleatoria, que en general representa a variables no observadas que hacen diferente nuestra predicción con lo que en realidad se observa (o sea, el error).

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad \forall\, j$$

La clave en el modelo estadístico es definir una distribución para este término de error, o sea una distribución de probabilidad conjunta para las variables aleatorias ε_{nj}.

Vamos a decir que el vector aleatorio $\varepsilon_{nj} = \{\varepsilon_{1j} ,..., \varepsilon_{nJ}\}$ tiene una función de densidad a la que vamos a llamar: $f(\varepsilon_{nj})$

> (En la mayoría de los modelos estudiados por ustedes, los errores se consideran independientes, por lo que su distribución conjunta es un multiplicatoria de sus distribuciones marginales: Recuerdan la función de verosimilitud?).

El modelo supone que los sujetos siguen una regla para decidir entre las alternativas. Esta regla es maximizar su utilidad basados en la información disponible. Entonces, la probabilidad de que la persona escoja la alternativa i en lugar de la alternativa J se define como:

$$P_{ni} = Pr(U_{ni} > U_{nj} \quad \forall\, j \neq i)$$

$$P_{ni} = Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \quad \forall\, j \neq i)$$

$$P_{ni} = Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall\, j \neq i)$$

Utilizando la función de densidad, se puede plantear también esta probabilidad como:

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall\, j \neq i) * f(\varepsilon_n)\, d\, \varepsilon_n$$

donde I() es una función indicadora, igual a 1 si la expresión a lo interno del paréntesis es verdadera, y 0 de lo contrario.

Ahora bien, si se tienen J categorías en la muestra y el evento de la probabilidad se planteó como las diferencias entre eventos (utilidades), o bien, como las diferencias entre errores, entonces en lugar de tener una muestra de J categorías, tenemos una muestra de J-1 diferencias, con un categoría tomado como base.

$P_{ni} = Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall\ j\neq i)$
$P_{ni} = Pr(\tilde{\varepsilon}_{nji} < V_{ni} - V_{nj} \ \forall\ j\neq i)$

$$P_{ni} = \int_{\varepsilon} I(\tilde{\varepsilon}_{nji} < V_{ni} - V_{nj} \ \forall\ j\neq i) * f(\tilde{\varepsilon}_{nji})\, d\, \tilde{\varepsilon}_{nji}$$

<u>Características de los modelos de escogencia discreta:</u>

- Solo las diferencias entre utilidades importan. En el ejemplo de las cervezas, como vamos a ver, si solo tuviéramos Imperial y Pilsen, y las dos valen igual, no vamos a poder estimar ningún modelo, porque las dos tienen la misma utilidad sistemática.
- Para incorporar variables de individuos, es necesario incluir interacciones, entre variables independientes de individuo y variables independientes de las alternativas.

El Modelo logístico de Escogencia discreta:

Supone que las variables de error (los J errores de los N individuos) se distribuyen conjuntamente como la productoria de J distribuciones “valor extremo” distribuidas independientemente e idénticamente. A la distribución valor extremo también se le conoce como distribución “Gumbel”.

La distribución “valor extremo” tiene función de densidad

$$f\left(\varepsilon_{nj}\right) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}$$

Su función de probabilidad acumulada es:

$$F\left(\varepsilon_{nj}\right) = e^{-e^{-\varepsilon_{nj}}}$$

La media y la variancia de variables con distribución "valor extremo" son: 0.57722 y $\pi^2/6$

Ahora bien, recuerden que lo que nos interesaba es la diferencia entre errores (o bien, que es lo mismo, la diferencia entre individuos). Se podría demostrar que la distribución de probabilidad acumulada de la diferencia entre dos errores (variables) con distribución "valor extremo", tienen distribución logística estándar:

$$F\left(\tilde{\varepsilon}_{nji}\right) = \frac{\exp\left(\tilde{\varepsilon}_{nji}\right)}{1 + \exp\left(\tilde{\varepsilon}_{nji}\right)}$$

La media de la distribución de densidad logística es 0, pero la variancia es: $\pi^2/3$

Si las diferencias de errores tienen distribución logística, entonces las utilidades también tienen distribución logística:

$$P_{ni} = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \ldots \beta_I x_I\right)}{\sum_j \exp\left(\beta_0 + \beta_1 x_1 + \ldots \beta_I x_I\right)}$$

El modelo se puede plantear más complejo, si se toman también las características de los individuos.

$$P_{ni} = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \ldots \beta_I x_I + \lambda_i y_n\right)}{\sum_j \exp\left(\beta_0 + \beta_1 x_1 + \ldots \beta_I x_I + \lambda_i y_n\right)}$$

donde los β son los coeficientes para las características de las alternativas, y los λ son los coeficientes para las características de los individuos.

Características del modelo logístico sobre otros modelos de escogencia discreta:

- El modelo logístico puede representar variación en la parte sistemática de los gustos, pero no variación en la parte aleatoria. Esto quiere decir que si yo no incluyo una variable importante para definir el gusto por las alternativas (por ejemplo, gusto por lo amargo de la cerveza), puede que encuentre sesgos en mis estimaciones de los parámetros.

- Si los factores no incluidos en el modelo y por lo tanto no observados, NO varían en el tiempo, entonces el modelo logístico se puede utilizar cuando hay distintas observaciones de cada individuo en el tiempo. Esto quiere decir que si hago el estudio en el mismo bar 5 viernes seguidos, y un mismo individuo va al bar esas 5 veces, no importa que lo tome como una persona diferente en mi modelo, siempre y cuando las variables que no estoy midiendo no varíen en el tiempo, por ejemplo, que de un momento a otro, a este individuo le llegue a gustar cervezas más amargas.
- SUSTITUCIÓN PROPORCIONAL: El modelo logístico supone sustitución proporcional entre opciones. Esto quiere decir que, si en el ejemplo de las cervezas, en el bar no hay Pilsen, la probabilidad de Pilsen se prorratea entre la probabilidad de Imperial y la probabilidad de Bavaria Negra. Nótese que esto no necesariamente es correcto. Si en el bar no hay Pilsen, el típico tomador de cerveza costarricense va a pedir Imperial en lugar de Bavaria Negra.

LA SUSTITUCIÓN PROPORCIONAL ESTÁ RELACIONADA CON EL PRINCIPIO DE IIA.

<u>El principio de Independencia entre Alternativas Irrelevantes (IIA por Independence of Irrelevant Alternatives).</u>

Este principio supone que, para un individuo específico, la razón de probabilidades de dos alternativas cualesquiera no está afectada por la parte sistemática de las utilidades de las otras alternativas.

$$\frac{P_{ni}}{P_{nk}} = \frac{\exp(\beta x_{ni}) \Big/ \sum_j \exp(\beta x_{nj})}{\exp(\beta x_{nk}) \Big/ \sum_j \exp(\beta x_{nj})}$$

$$\frac{P_{ni}}{P_{nk}} = \frac{\exp(\beta x_{ni})}{\exp(\beta x_{nk})} = \exp\left[\beta(x_{ni} - x_{nk})\right]$$

El principio de IIA se puede probar con el test de Hausman.

Estudiaremos el test de Hausman cuando veamos los otros modelos.

<u>Representación de elasticidades:</u>

En Economía, la elasticidad se define como el cambio relativo en cierta variable dependiente, debido al cambio relativo en cierta variable independiente.

La fórmula de elasticidad en un modelo logístico sería:

$$E_{iz_{ni}} = \frac{\partial P_{ni}}{\partial z_{ni}} * \frac{z_{ni}}{P_{ni}} = \beta_z z_{ni} \left(1 - P_{ni}\right)$$

si la utilidad es lineal en zni.

Estimación:

La fórmula de estimación por máxima verosimilitud es equivalente a la fórmula de estimación de máxima verosimilitud para logísticas condicionales, ya que estamos CONDICIONANDO a grupos discordantes: en otras palabras, a diferencias en utilidades, porque sólo IMPORTAN las diferencias en utilidades.

Poner el ejemplo de la página 76.

Ejemplo de p.76.

Table 3.1. Logit model of work trip mode choice

Explanatory Variable	Coeff	z-Statistic
Cost-divided by post-tax wage, cents/minute (1-4)	-0.0284	4.31
Auto on vehicle time, minutes (1,3,4)	-0.0644	5.65
Transit on-vehicle time, minutes (2,3)	-0.0259	2.94

El Modelo logístico ANIDADO (UN CASO DEL GEV):

El GEV son las siglas de los modelos de “valor extremo” anidado.

Supone que las porciones no observadas de utilidad para todas las alternativas están distribuidas conjuntamente con una distribución de "valor extremo generalizado".

Más específicamente, permite plantear correlaciones entre alternativas, algo que el modelo logístico no permitía. Si las correlaciones entre alternativas o distribuciones es cero, entonces, este modelo es el modelo logístico de escogencia discreta.

El modelo más usado de GEV es el modelo logístico anidado porque es el más simple de implementar. Es el que está implementado en STATA, por ejemplo.

Las características del modelo logístico anidado son:

- Apropiado cuando el conjunto total de alternativas se pueden agrupar en subconjuntos. En nuestro ejemplo, Pilsen e Imperial se agrupan, mientras que la Bavaria negra es totalmente diferente, porque sería otro subconjunto.
- En este caso, para cada par de alternativas dentro del mismo "nido", la razón de probabilidades es independiente de los atributos o existencia de las otras alternativas. En otras palabras, el IIA se mantiene dentro de cada "nido".
- Para dos alternativas diferentes en "nidos" diferentes, la razón de probabilidades puede depender en los atributos de otras alternativas de los dos "nidos". IIA no se mantiene entre alternativas de diferente "nido".

Distribuciones del modelo logístico anidado.

Suponga que las j alternativas se pueden reagrupar en K subconjuntos mutuamente excluyentes denominados B1, B2, ..., B_k, llamados "nidos".

En este caso, el vector de errores (o utilidad no observada) $\varepsilon_n = \{\varepsilon_1, ..., \varepsilon_{nJ}\}$ tiene distribución de probabilidad acumulada igual a:

$$F(\varepsilon_{nj}) = \exp\left[-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right]$$

Las ε_{nJ} están correlacionadas a lo interno de cada "nido". Para cualequiera dos alternativas j y m dentro del nido $_{Bk}$, ε_{nj} está correlacionado con $\varepsilon_{nm.}$ Para cualesquiera dos alternativas en "nidos" diferentes, la parte no observada de la utilidad se supone que están NO correlacionada.

El parámetro λ_k es una medida del grado de independencia en la utilidad no observada (el error) entre alternativas a lo interno de cada "nido". Un valor de λ_k= 1 significa que las alternativas dentro del mismo "nido" son totalmente independientes, o sea, no correlacionadas.

Con base en esta fórmula, la probabilidad de selección de cada alternativa sería:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k}\left[-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}\right]}{\sum_{l=1}^{K}\left[-\sum_{k=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_k}\right)^{\lambda_k}\right]}$$

Nuevamente, sustituyendo donde corresponda y multiplicando las probabilidades, se obtiene la función de verosimilitud, con el fin de maximizarla.

La expresión de Pni se puede reexpresar como:

$$P_{ni} = P_{ni|B_k} * P_{nB_k}$$

En otras palabras, la probabilidad de cada escogencia se puede descomponer en la probabilidad marginal de cada "nido" por la probabilidad de la escogencia ni condicional a estar en el nido Bk.

Las características de los modelos logísticos anidados son:

- Al igual que el modelo logístico simple, sólo se puede representar variación en la parte sistemática de los gustos, pero no variación en la parte aleatoria. Nuevamente, esto quiere decir que si yo no incluyo una variable importante para definir el gusto por las alternativas (por ejemplo, gusto por lo amargo de la cerveza), puede que encuentre sesgos en mis estimaciones de los parámetros.

- NO SUPONE SUSTITUCIÓN PROPORCIONAL: Precisamente, este modelo flexibiliza este supuesto, ya que solo va a suponer sustitución proporcional entre "nidos". En otras palabras, si en el bar no hay Pilsen, la gente preferiría Imperial que Bavaria Negra, y esto lo podemos modelar dejando a Imperial y Pilsen en el mismo nido. En otras palabras, en este modelo no necesitamos suponer IIA: Independencia entre Alternativas Irrelevantes.

Estudiaremos el test de Hausman con este modelo.

La hipótesis nula del test de Hausman es que el conjunto de coeficientes de un modelo es igual al conjunto de coeficientes de otro modelo.

La hipótesis alternativa es que son diferentes.

La fórmula del test de Hausman es:

$$\boldsymbol{H} = \left(\beta_{\boldsymbol{c}} - \beta_{\boldsymbol{e}}\right)' \left(\boldsymbol{V}_{\boldsymbol{c}} - \boldsymbol{V}_{\boldsymbol{e}}\right)^{-1} \left(\beta_{\boldsymbol{c}} - \beta_{\boldsymbol{e}}\right) \qquad \boldsymbol{H} \sim \chi^2_{\boldsymbol{n}}$$

donde:
βc es el vector de estimaciones de la ecuación más grande
βe es el vector de estimaciones de la ecuación más pequeña
Vc es la matriz de var-covar de la ecuación más grande
Ve es la matriz de var-covar de la ecuación más pequeña
n son los grados de libertad de la $\chi^2_{\boldsymbol{n}}$ (número de variables incluida la constante)

Las elasticidades se calculan igual.

El Modelo probit:

Las ventajas del modelo probit es que permiten flexibilizar los supuestos de:

- Puede representar o incluir variación aleatoria en los gustos.
- No tiene restricciones en la violación del supuesto de IIA
- Puede usarse en estudios de panel cuando factores no observados están correlacionados en el tiempo para cada tomador de decisiones.

El único supuesto restrictivo que tiene el modelo probit es que el componente no observado de la utilidad se distribuye normalmente.

Supone que las variables de error (los J errores de los J categorìas) se distribuyen normalmente con media 0 y matriz var-covar Ω. La función de densidad de ε_n es:

$$\boldsymbol{f}\left(\varepsilon_{\boldsymbol{n}}\right)=\Phi(\varepsilon_{\boldsymbol{n}})=\frac{1}{\left(2\pi\right)^{\boldsymbol{J}/2}\left|\Omega\right|^{1/2}}\boldsymbol{e}^{-\frac{1}{2}\varepsilon'_{\boldsymbol{n}}\Omega^{-1}\varepsilon_{\boldsymbol{n}}}\boldsymbol{e}^{-\boldsymbol{e}^{-\varepsilon_{\boldsymbol{nj}}}}$$

Su función de probabilidad acumulada es entonces la distribución normal estándar acumulada.

La matriz de var-covar tiene dimensión JxJ, aunque si el modelo se plantea como diferencias entre alternativas, se puede estimar una matriz (J-1)x(J-1).

Nótese que, al estar los errores distribuidos normalmente, los componentes de la matriz de variancia covariancia también pueden considerarse como parámetros por estimar. Sin embargo, la ventaja del modelo probit es que también se puede imponer una estructura de variancias-covariancias justificadas teóricamente. Por ejemplo, se puede plantear una matriz de var-covariancia que suponga un modelo autorregresivo. En otras palabras, que el error de la alternativa i está correlacionado con el error en la alternativa i-1. La estructura de la matriz de var-covar también se podría derivar con otros métodos, como análisis factorial, o con matriz de proximidades.

Variación aleatoria en el gusto.

Se puede demostrar que, si uno supone que los coeficientes se distribuyen normalmente, uno puede suponer que los coeficientes estimados están compuestos por una parte sistemática y también una parte aleatoria que representaría la variación aleatoria en el gusto. Esto permite suponer variación aleatoria en el gusto.

Esto se puede ejemplificar con ir a una heladería. Hay personas que siempre que van a una heladería piden el helado de un sabor particular: chocolate, ron con pasas, "brownie dinamita". Sin embargo, hay otros que a veces quieren un sabor y a veces otro. Por ejemplo, yo a veces pido una nieve de limón, a veces una nieve de guanábana, y a veces una nieve de naranja. Hay un componente sistemático en mis preferencias que podrían pronosticar el que yo prefiera nieves ácidas, en lugar de un helado de chocolate.

Con un modelo logito o logito anidado, esta variación aleatoria en mis gustos queda mal definida. Sin embargo, el modelo probit supone que esta variación aleatoria va contenida en el error.

Consideración del supuesto IIA.

Como dijimos, el modelo logístico supone que la propiedad IIA, o sea, que si falta alguna alternativa, las probabilidades se prorratean entre las alternativas restantes.

En el modelo logístico anidado, se supone que el supuesto IIA sólo está contenido dentro de cada "nido", pero el IIA se flexibiliza entre nidos.

En cambio en el modelo probit, los patrones de sustitución se pueden determinar en la matriz de var-covar Ω. El investigador puede determinar los patrones de sustitución imponiendo una matriz var-covar teórica.

Consideración de la correlación a través del tiempo

La correlación a través del tiempo también se incorpora al modelo imponiendo una estructura de la matriz var-covar que tome en cuenta la correlación de la alternativa i e tiempo t con la alternativa i en tiempo t-1.