Modelos de Clasificación / Variable Dependiente Cualitativa

Métodos Cuantitativos en Estudios Urbanos II - MEU UTDT

Ricardo Pasquini

4 de julio de 2025

Modelos de Clasificación

- Hasta aquí hemos trabajado con variables continuas como variables a explicar/predecir, pero en una gran cantidad de aplicaciones las predicciones son discretas.
- Estos modelos también son conocidos como modelos de clasificación

Modelos de Clasificación Binaria

- Ejemplo: Elección de transporte
- Variable objetivo: 1 si toma transporte público, 0 de otro modo.

$$P(\text{Transporte Publico}) = \beta_0 + \beta_1 \text{Ingreso}_i + \beta_2 \text{Genero}_i + \epsilon_i$$

Modelos de Clasificación Binaria

- En general la variable $y_i \in \{0, 1\}$, representando un resultado "Sí/ "No".
- El objetivo es describir $P(y_i = 1|x_i)$ (Notar que esto determinaría la completa distribución condicional)
- Si proponemos un modelo lineal:

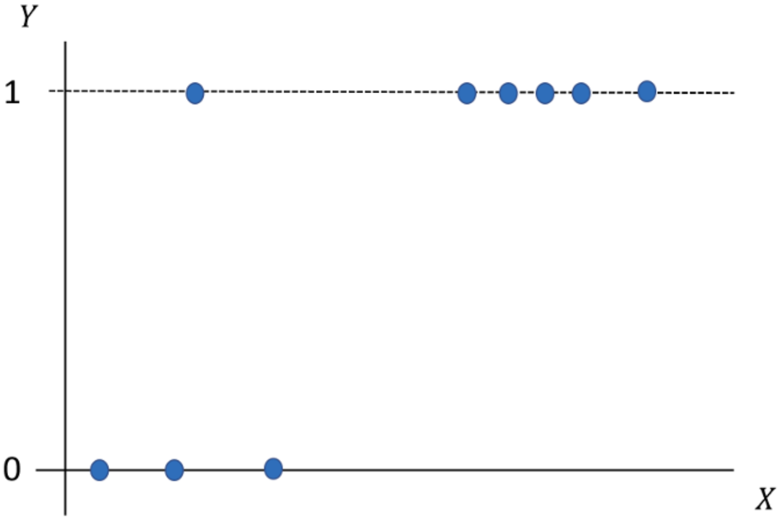
$$P(y_i = 1|x_i) = x_i'\beta$$

- Notemos también que $P(y_i = 1|x_i) = E(Y_i|x_i)$ por lo tanto podríamos estimar

$$y_i = x'\beta + e_i$$

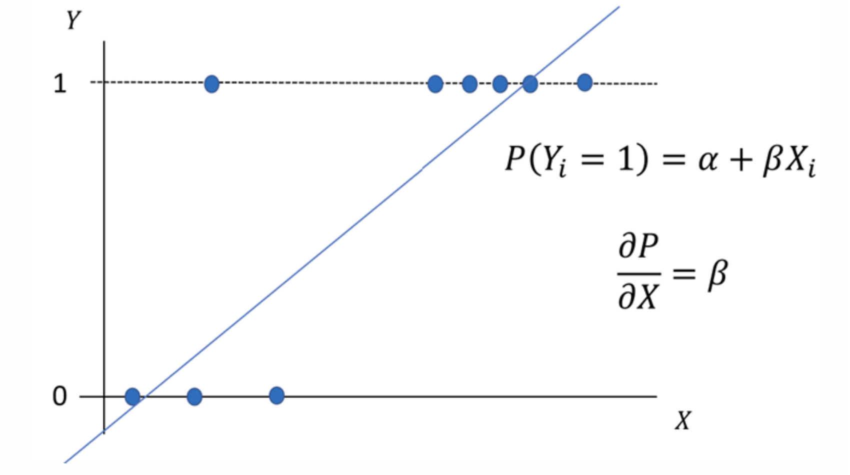
Modelos de Clasificación Binaria

- Sin embargo, notemos que en un caso de observaciones binarias tendríamos algo como:



Modelo Lineal para Clasificación

- El resultado devolvería:



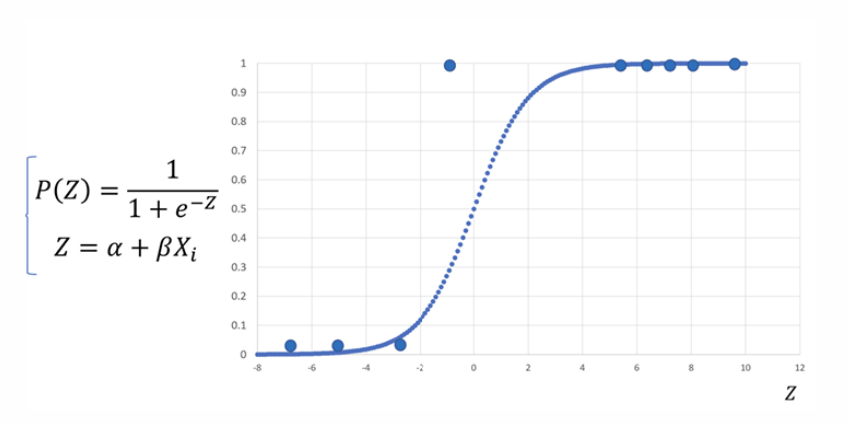
Modelos de Clasificación

- Es deseable imponer una restricción adicional, que $0 \leq P(y_i = 1|x_i) \leq 1$
- Por eso la alternativa estándar es proponer un modelo como:

$$y_i = F(x'\beta) + e_i$$

  donde $F(\cdot)$ es una función que cumple la restricción, como por ejemplo una Función de Densidad Acumulativa.
- Dos funciones comúnmente utilizadas son:
  - Logística: $F(u) = \frac{1}{(1+e^{-u})} = \frac{e^u}{1+e^u}$ (resultando en el modelo llamado **Logit**)
  - Normal: $F(u) = \Phi(u)$ (resultando en el modelo llamado **Probit**)

Ejemplo: Logit



Logit

- Logit también es el nombre que toma el logaritmo del ratio de chances entre dos resultados.
- En este caso, tal como fue definido, el modelo logit propone que el logaritmo del ratio de chances es la función lineal $\alpha + \beta X$. Esto se puede ver facilmente:

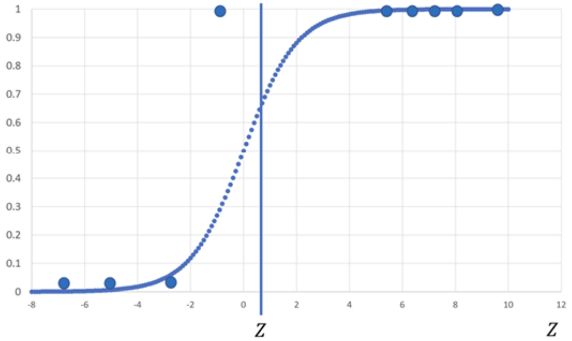
$$\ln \frac{P(y_i = 1|x_i)}{P(y_i = 0|x_i)} = \beta X$$

Efectos marginales

- Comúnmente nos interesará conocer $\frac{\partial p}{\partial x}$.
- Notar que debido a la transformación funcional de $F(\cdot)$ el efecto marginal de un incremento en x_i no es constante. Por lo tanto, para conocer el efecto de un incremento marginal en x adicionalmente tendremos que especificar el nivel (de Z) de interés.

$$\begin{cases} P(Z) = \dfrac{1}{1+e^{-Z}} \\ Z = \alpha + \beta X_i \end{cases}$$

$$\frac{\partial p}{\partial X} = \frac{e^{-Z}}{(1+e^{-Z})^2}\beta$$



Efectos marginales

- Típicamente los paquetes estadísticos devuelven un efecto marginal que:
  - Podría estar calculado por default en el *valor promedio de las variables* (esto podria ser de interés o no).
  - Nos permiten determinar los valores de las variables explicativas donde calcular el efecto.
  - Permiten calcular un *Efecto Marginal Promedio* entendido como el promedio de los efectos marginales en los valores definidos por las observaciones en nuestra muestra.

Estimación

- La estimación se realiza mediante Máxima Verosimilitud.
- Recordemos que la estimación mediante Máxima Verosimilitud propone buscar el valor del parámetro que maximiza la probabilidad conjunta de ocurrencia de los datos que sugieron en la muestra.

$$\max_{\beta} \prod_{i=1}^{n} P(Y_i, \beta)$$

  donde $p(Y_i, \beta)$ es la probabilidad de la observación i-esima en cuestión.
- Notemos en este caso las probabilidades son

$$\begin{cases} P(Y=0|X) & \text{si } Y_i = 0 \\ P(Y=1|X) & \text{si } Y_i = 1 \end{cases}$$

Estimación

- Es decir los términos de la productoria serán

$$\begin{cases} 1 - F(x_i'\beta) & \text{si } Y_i = 0 \\ F(x_i'\beta) & \text{si } Y_i = 1 \end{cases}$$

- Siguiendo el método de máxima verosimilitud, será conveniente maximizar la función en logaritmos

$$\max_{\beta} L = \ln(\prod_{i=1}^{n} P(Y_i, \beta)) = \max_{\beta} \sum \ln(P(Y_i, \beta))$$

- Típicamente se utilizarán algoritmos de maximización para encontrar el máximo.

Ejemplo: Crecimiento urbano

Urban Growth Pattern Modeling Using Logistic Regression

NONG Yu[1], **DU Qingyun**[1, 2]

1. School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
2. Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

Figura: Yu and Qingyun (2011)

Ejemplo: Crecimiento urbano

Table 1 List of variables included in logistic regression model

Variable	Broad definition	Unit	Nature of variable
Dependent			
Y	0. no urban growth; 1. urban growth		Dichotomous
Independent			
X_1	Population density	Population (km^2)	Continuous
X_2	Gross industrial output value	Billion Yuan (town)	Continuous
X_3	Gross agricultural output value	Billion Yuan (town)	Continuous
X_4	Index of distance to economic center	-	Continuous
X_5	Index of distance to the major road	-	Continuous
X_6	Slope (5 categories) 1. $0° \leqslant$ slope $< 2°$ 2. $2° \leqslant$ slope $< 6°$ 3. $6° \leqslant$ slope $< 15°$ 4. $15° \leqslant$ slope $< 25°$ 5. slope $\geqslant 25°$	Degree	Design

Table 4 Variables in the equation in SPSS 11.5

	Coefficient	Standard error	Wald	Df	Sig.(Pr > chi-square.)
X_1	0.023	0.002	116.928	1	0.000
X_2	0.002	0.000	97.834	1	0.000
X_3	0.001	0.000	1.718	1	0.019
X_4	0.040	0.003	173.902	1	0.000
X_5	0.022	0.001	1074.271	1	0.000
X_6	−0.051	0.018	7.978	1	0.005
Constant	−1.873	0.051	1329.625	1	0.000

Figura: Yu and Qingyun (2011)

Multinomial Logit para elecciones múltiples

- Un caso de interés es la elección entre alternativas múltiples. Por ejemplo:
  - Elección medio de transporte (auto, colectivo, tren, etc.).
  - Elección residencial (barrio 1, barrio 2,...,barrio n).
- Analizaremos aquí un modelo con alternativas que no tienen un orden implícito. Para alternativas ordenadas, existen otros modelo disponibles.

Multinomial Logit para elecciones múltiples

- El modelo multinomial se puede fundamentar suponiendo que cada individuo tiene una utilidad por la j-esima opcion dada por

$$U_{ij} = \beta' z_{ij} + \epsilon_{i,j}$$

- Si el consumidor elige la j-esima opción sobre las $J - 1$ restantes, es porque esa opción le dió mayor utilidad. La idea de la modelización probabilística es

$$P(Y_i = j) = P(U_{ij} > U_{ik}) \; \forall k \neq j$$

- Y se puede obtener que la probabilidad de elegir la categoría j-esima es:

$$P(Y_i = j) = \frac{e^{\beta' z_{ij}}}{\sum_{j=1}^{J} e^{\beta' z_{ij}}}$$

Multinomial Logit para elecciones múltiples

- Un supuesto implícito del modelo es la independencia de alternativas irrelevantes.
- Este supuesto pide que la preferencia por una categoría sobre otra no cambie si se considera una tercera como parte de la comparación.
- El uso de este supuesto permite que este modelo pueda implementarse como la comparación de K-1 categorías contra una K-esima. Conociendo estas preferencias puedo derivar todas las demás.

Derivación como un conjunto de elecciones binarias

$$\ln \frac{Pr(Y_i = 1)}{Pr(Y_i = K)} = \beta_1 \cdot X_i$$
$$\ln \frac{Pr(Y_i = 2)}{Pr(Y_i = K)} = \beta_2 \cdot X_i$$
$$\vdots$$
$$\ln \frac{Pr(Y_i = K-1)}{Pr(Y_i = K)} = \beta_{K-1} \cdot X_i$$

Multinomial Logit para elecciones múltiples

$$Pr(Y_i = 1) = Pr(Y_i = K)e^{\beta_1 \cdot X_i}$$

$$Pr(Y_i = 2) = Pr(Y_i = K)e^{\beta_2 \cdot X_i}$$

$$\vdots$$

$$Pr(Y_i = K - 1) = Pr(Y_i = K)e^{\beta_{K-1} \cdot X_i}$$

$$Pr(Y_i = K) = 1 - Pr(Y_i = K) \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}$$

$$\rightarrow Pr(Y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

Multinomial Logit para elecciones múltiples

Reemplazando para obtener expresiones para $Pr(Y_i = 1)$, $Pr(Y_i = 2)$ ...

$$Pr(Y_i = 1) = \frac{e^{\beta_1 \cdot X_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$
$$Pr(Y_i = 2) = \frac{e^{\beta_2 \cdot X_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$
$$\vdots$$
$$Pr(Y_i = k) = \frac{e^{\beta_k \cdot X_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot X_i}}$$

Estimación

- Como en el caso del Logit, se aplicará un método similar a Máxima Verosimilitud para identificar las condiciones del óptimo, y se utilizarán algoritmos numéricos iterativos para encontrar la solución.

Ejemplo: Elección Modalidad Transporte

A multinomial logistic regression model for public transportation use in a medium-sized Brazilian city

Marianna Lucinda de Oliveira[a*], Josiane Palma Lima[a**]
[a]Universidade Federal de Itajubá, Itajubá, MG, Brasil
*mariannaoliveira@unifei.edu.br, **jplima@unifei.edu.br

4.3. Defining the dependent and exploratory variables

The independent variables were developed according to relevance in literature, and data availability. Altogether, 34 variables that could influence PT frequency were considered. Frequency-of-use was set as a dependent variable, and classified into three categories: frequent users (from 3 to 7 days a week); occasional users (1 or 2 days a week); and rare users (up to 3 days per month).

Ejemplo: Elección Modalidad Transporte

$$Z = \mathrm{Ln}\left[\frac{Probability(Ocassional\ PT)}{Probability(Frequent\ PT)}\right] = \alpha_{1,0} + \sum_{i=1}^{k} B_{1,i}.X_i \qquad (1)$$

$$Z = \mathrm{Ln}\left[\frac{Probablity(\mathrm{Rare}\ PT)}{Probability(Frequent\ PT)}\right] = \alpha_{2,0} + \sum_{i=1}^{k} B_{2,i}.X_i \qquad (2)$$

Where:

a) Xi are the independent variables that predict the frequency of PT use;

b) Bji are the estimated regression coefficients for category j of the dependent variable; and

c) α is the intercept.

Ejemplo: Elección Modalidad Transporte

Table 3. Multinomial Logistic Regression - Significant results in bold.

	Occasional				Rare			
	B1	EP B1	Z	p-value	B2	EP B2	Z	p-value
Constant	-4.21	3.23	1.70	0.19	-10.02	3.78	7.04	0.01
Female	-0.51	0.75	0.45	0.50	-2.03	0.92	4.88	**0.03****
Age Group (15-19)	3.67	1.71	4.62	**0.03****	1.91	1.91	1.01	0.32
Age Group (20-35)	3.74	1.29	8.44	**0.00*****	3.43	1.38	6.14	**0.01*****
Level of Education (High School)	-1.59	0.94	2.88	**0.09***	-0.18	1.09	0.03	0.87
Household Income (>$553.00)	-3.24	1.17	7.68	**0.01*****	-2.07	1.33	2.41	0.12
Occupation (Retired)	-0.39	1.95	0.04	0.84	-4.55	2.24	4.11	**0.04****
Occupation (Unemployed)	-1.30	1.22	1.13	0.29	-2.43	1.41	2.98	**0.08***
Occupation (Student)	1.48	1.20	1.51	0.22	2.71	1.45	3.51	**0.06***
Payment (Normal)	7.07	1.48	22.78	**0.00*****	8.17	1.74	21.99	**0.00*****
Payment (Free)	7.68	2.36	10.61	**0.00*****	10.97	2.73	16.19	**0.00*****
Automobile/Motorcycle	2.28	0.83	7.59	**0.01*****	3.57	0.93	14.72	**0.00*****
PT to Work	-5.26	1.16	20.65	**0.00*****	-7.37	1.45	25.68	**0.00*****
PT to Study	-6.19	1.45	18.29	**0.00*****	-10.30	2.56	16.24	**0.00*****
Punctuality (Neutral)	2.29	1.10	4.34	**0.04****	3.93	1.29	9.27	**0.00*****
Fare (Unsatisfied)	5.94	2.06	8.31	**0.00*****	6.43	2.15	8.93	**0.00*****
Fare (Neutral)	7.39	2.29	10.42	**0.00*****	9.01	2.39	14.22	**0.00*****
Accessibility (Good)	-6.07	2.55	5.68	**0.02****	-3.78	2.75	1.89	0.17
Accessibility (Regular)	-8.29	3.01	7.59	**0.01*****	-6.69	3.27	4.19	**0.04****
Distance to CBD	-0.17	0.22	0.63	0.43	-0.52	0.27	3.62	**0.06***
POI to a Supermarket	2.26	1.36	2.73	0.10	6.02	1.78	11.41	**0.00*****

Note: *Significance Level 0.1; ** Significance Level 0.05; *** Significance Level 0.01.

Figura: Oliveira and Lima (2023)

Resumen

- Los modelos de clasificación binaria incorporan una transformación no-lineal que garantizar que las predicciones vivan en el rango apropiado.
- Esto implica que los efectos marginales de un factor explicativo serán especificos a un nivel deseado.

Referencias

- Hansen, B. 2018, Econometrics , ch. 21
- Greene Análisis Econométrico ch.19