

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{\beta_0}^2 & \hat{\sigma}_{\beta_0 \beta_1} \\ \hat{\sigma}_{\beta_1 \beta_0} & \hat{\sigma}_{\beta_1}^2 \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

Carlos de Anta Puig

$$E[u_t | X] = 0$$

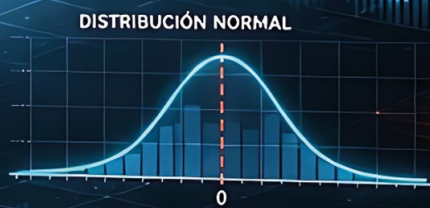
(Exogeneidad)

$$\text{Var}(\tilde{u}) = \sigma^2 I$$

(Homoscedasticidad)

$$\text{Cov}(u_t, u_s) = 0$$

(No autocorrelación)



#### PRUEBAS DE HIPÓTESIS

|                    | ESTADÍSTICO t | p-VALOR |
|--------------------|---------------|---------|
| $H_0: \beta_1 = 0$ | 3.85          | 0.0005  |
| $H_0: \beta_2 = 0$ | -2.17         | 0.030   |
| $H_0: \beta_3 = 0$ | 4.12          | 0.0001  |

**RECHAZAR  $H_0$**

#### DIAGNÓSTICOS

|                                    |   |
|------------------------------------|---|
| $R^2 = 0.842$                      | ✓ |
| Breusch-Pagan (Heterocedasticidad) | ✓ |
| Durbin-Watson (Autocorrelación)    | ✓ |
| Jarque-Bera (Normalidad)           | ✓ |

# Manual de Microeconometría

# MANUAL DE MICROECONOMETRÍA

**CARLOS DE ANTA PUIG**

# **MANUAL DE MICROECONOMETRÍA**



*A mis alumnos,  
por su curiosidad, su esfuerzo  
y las preguntas que me obligan  
a pensar mejor las respuestas.*

*A quienes creen que la econometría  
es difícil: lo es.  
Pero merece la pena.*

Manual de Microeconometría  
© Carlos de Anta Puig, 2026  
© Digital Reasons, 2026  
Serrano, 51 - 28006 Madrid (España)  
bibliotecaonline.es  
pedidos@bibliotecaonline.es  
Diseño de cubierta: BibliotecaOnline.

ISBN (papel): 978-84-10093-39-3  
D.L.: M-9998-2026

Ficha bibliográfica: de Anta Puig, Carlos (2026): *Manual de Microeconometría*,  
Digital Reasons.

Imprime: Podiprint.  
Impreso en España - *Printed in Spain*

Las afirmaciones incluidas en el libro son responsabilidad exclusiva de los autores. Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita reproducir algún fragmento de esta obra ([www.conlicencia.com](http://www.conlicencia.com); 91 702 19 70 / 93 272 04 47).

|                                                                                          |           |
|------------------------------------------------------------------------------------------|-----------|
| <b>Prólogo</b>                                                                           | <b>15</b> |
| <b>1 Datos con Variable Dependiente Limitada</b>                                         | <b>17</b> |
| 1.1 Cuando la variable dependiente no es libre . . . . .                                 | 17        |
| 1.1.1 Por qué MCO no es válido con variables dependientes limitadas . . . . .            | 18        |
| 1.2 La taxonomía de las variables dependientes limitadas . . . . .                       | 19        |
| 1.3 Variables binarias: la decisión sí o no . . . . .                                    | 20        |
| 1.3.1 La distribución de Bernoulli . . . . .                                             | 20        |
| 1.3.2 La distribución Binomial . . . . .                                                 | 21        |
| 1.3.3 La variable latente: el mecanismo que no vemos . . . . .                           | 22        |
| 1.4 Variables multinomiales: más de dos opciones . . . . .                               | 23        |
| 1.5 Variables continuas limitadas: censura y truncamiento . . . . .                      | 25        |
| 1.6 El análisis exploratorio de datos como primer paso . . . . .                         | 25        |
| 1.6.1 Carga e inspección inicial de los datos . . . . .                                  | 26        |
| 1.6.2 Análisis visual de distribuciones univariantes . . . . .                           | 28        |
| 1.6.3 Análisis de la variable dependiente: patrones por tipo de VDL . . . . .            | 28        |
| 1.6.4 Análisis bivalente: relaciones entre variables . . . . .                           | 30        |
| 1.7 Casos prácticos . . . . .                                                            | 30        |
| 1.7.1 Caso práctico 1: Participación laboral — variable binaria . . . . .                | 32        |
| 1.7.2 Caso práctico 2: Gasto en seguros del hogar — variable censurada . . . . .         | 32        |
| 1.7.3 Caso práctico 3: Visitas al médico — variable de recuento . . . . .                | 34        |
| 1.8 Guía de comandos R para este capítulo . . . . .                                      | 35        |
| 1.8.1 load() — Carga de datos en formato RData . . . . .                                 | 35        |
| 1.8.2 str() — Estructura interna del objeto . . . . .                                    | 35        |
| 1.8.3 table() y prop.table() — Frecuencias y proporciones . . . . .                      | 35        |
| 1.8.4 hist() — Histograma de distribución . . . . .                                      | 36        |
| 1.8.5 boxplot() — Diagrama de caja por grupos . . . . .                                  | 36        |
| 1.8.6 mean(y == 0) — Proporción de ceros (diagnóstico de censura) . . . . .              | 36        |
| 1.8.7 var(y) / mean(y) — Ratio de sobredispersión (diagnóstico para recuentos) . . . . . | 37        |
| <b>2 Modelos de Elección Discreta</b>                                                    | <b>38</b> |
| 2.1 El problema económico: cuando el agente dice sí o no . . . . .                       | 38        |
| 2.2 El Modelo Lineal de Probabilidad . . . . .                                           | 38        |
| 2.3 El modelo Probit: derivación y estimación . . . . .                                  | 39        |
| 2.3.1 El enfoque de variable latente . . . . .                                           | 39        |
| 2.3.2 La función de verosimilitud del Probit . . . . .                                   | 41        |
| 2.4 El modelo Logit: derivación y odds ratios . . . . .                                  | 42        |

|          |                                                                                                    |           |
|----------|----------------------------------------------------------------------------------------------------|-----------|
| 2.4.1    | La distribución logística . . . . .                                                                | 42        |
| 2.4.2    | La probabilidad Logit y la función log-odds . . . . .                                              | 43        |
| 2.4.3    | La función de verosimilitud del Logit . . . . .                                                    | 43        |
| 2.5      | Efectos marginales: AME y MEM . . . . .                                                            | 44        |
| 2.5.1    | Derivación matemática . . . . .                                                                    | 44        |
| 2.5.2    | AME y MEM: dos formas de resumir el efecto marginal . . .                                          | 46        |
| 2.6      | Modelo ilustrativo: participación laboral . . . . .                                                | 46        |
| 2.6.1    | Estimación comparada: MPL, Probit y Logit . . . . .                                                | 46        |
| 2.6.2    | Efectos marginales AME y MEM . . . . .                                                             | 48        |
| 2.6.3    | Odds ratios del modelo Logit . . . . .                                                             | 48        |
| 2.6.4    | Significación individual y conjunta . . . . .                                                      | 48        |
| 2.7      | Comparación final: MPL, Probit y Logit . . . . .                                                   | 50        |
| 2.8      | Casos prácticos . . . . .                                                                          | 51        |
| 2.8.1    | Caso práctico 1: Concesión de crédito bancario . . . . .                                           | 51        |
| 2.8.2    | Caso práctico 2: Decisión de compra online . . . . .                                               | 52        |
| 2.8.3    | Caso práctico 3: Matrícula universitaria . . . . .                                                 | 52        |
| 2.9      | Guía de comandos R para este capítulo . . . . .                                                    | 54        |
| 2.9.1    | glm() con family = binomial — Estimación Probit y Logit . .                                        | 54        |
| 2.9.2    | lm() + coefest(vcovHC()) — Modelo Lineal de Probabili-<br>dad (MPL) con errores robustos . . . . . | 54        |
| 2.9.3    | exp(coef()) — Odds Ratios del modelo Logit . . . . .                                               | 55        |
| 2.9.4    | Efectos marginales AME y MEM — Cálculo directo con de-<br>rivadas . . . . .                        | 55        |
| 2.9.5    | logLik() y pchisq() — Test LR y Pseudo-R <sup>2</sup> de McFadden . .                              | 56        |
| <b>3</b> | <b>Modelos para Datos Censurados: el Modelo Tobit</b>                                              | <b>57</b> |
| 3.1      | El problema: variables continuas con restricciones en el rango . . . .                             | 57        |
| 3.2      | El modelo Tobit tipo I . . . . .                                                                   | 58        |
| 3.2.1    | La estructura del modelo . . . . .                                                                 | 58        |
| 3.2.2    | Proporción de censura . . . . .                                                                    | 58        |
| 3.2.3    | La reparametrización de Olsen (1978) . . . . .                                                     | 59        |
| 3.2.4    | Por qué MCO es inconsistente con datos censurados . . . . .                                        | 61        |
| 3.2.5    | La función de verosimilitud del Tobit . . . . .                                                    | 61        |
| 3.3      | Efectos marginales en el modelo Tobit . . . . .                                                    | 62        |
| 3.4      | Modelo ilustrativo: gasto en seguros médicos . . . . .                                             | 62        |
| 3.4.1    | Análisis exploratorio . . . . .                                                                    | 62        |
| 3.4.2    | Estimación del modelo Tobit y comparación con MCO . . . .                                          | 65        |
| 3.4.3    | Efectos marginales AME (McDonald-Moffitt) . . . . .                                                | 65        |
| 3.5      | Interpretación de los coeficientes: las tres preguntas . . . . .                                   | 66        |
| 3.5.1    | Efecto sobre la variable latente: $E(y^*   \mathbf{x})$ . . . . .                                  | 66        |

|          |                                                                                    |           |
|----------|------------------------------------------------------------------------------------|-----------|
| 3.5.2    | Efecto sobre los no censurados: $E(y \mid \mathbf{x}, y > 0)$ . . . . .            | 66        |
| 3.5.3    | Efecto sobre toda la muestra: $E(y \mid \mathbf{x})$ . . . . .                     | 66        |
| 3.6      | Contrastes de especificación . . . . .                                             | 67        |
| 3.7      | Casos prácticos . . . . .                                                          | 68        |
| 3.7.1    | Caso práctico 1: Gasto en seguros médicos . . . . .                                | 68        |
| 3.7.2    | Caso práctico 2: Donaciones benéficas . . . . .                                    | 68        |
| 3.7.3    | Caso práctico 3: Gasto mensual en formación profesional . . .                      | 70        |
| 3.8      | Guía de comandos R para este capítulo . . . . .                                    | 70        |
| 3.8.1    | tobit() — Estimación del modelo Tobit (paquete AER) . . .                          | 70        |
| 3.8.2    | predict(tob, type = "lp") — Predictor lineal del Tobit . . . .                     | 71        |
| 3.8.3    | Efectos marginales McDonald-Moffitt — Cálculo manual . . .                         | 71        |
| 3.8.4    | jb_test y bp_test — Contrastes de especificación sobre los re-<br>siduos . . . . . | 72        |
| <b>4</b> | <b>Modelos para Datos de Recuento</b> . . . . .                                    | <b>74</b> |
| 4.1      | Cuando el resultado es un número entero no negativo . . . . .                      | 74        |
| 4.2      | El modelo de regresión de Poisson . . . . .                                        | 74        |
| 4.2.1    | La distribución de Poisson . . . . .                                               | 74        |
| 4.2.2    | El modelo de regresión de Poisson . . . . .                                        | 75        |
| 4.2.3    | Ejemplos de cálculo . . . . .                                                      | 76        |
| 4.2.4    | Función de log-verosimilitud . . . . .                                             | 76        |
| 4.2.5    | Tasas de incidencia (IRR) y efectos marginales . . . . .                           | 76        |
| 4.3      | La sobredispersión y el modelo Binomial Negativa . . . . .                         | 77        |
| 4.3.1    | El problema de la sobredispersión . . . . .                                        | 77        |
| 4.3.2    | El test formal de sobredispersión . . . . .                                        | 78        |
| 4.3.3    | El modelo Binomial Negativa (NB2) . . . . .                                        | 78        |
| 4.4      | Modelo ilustrativo: visitas al médico . . . . .                                    | 78        |
| 4.4.1    | Análisis exploratorio y diagnóstico previo . . . . .                               | 79        |
| 4.4.2    | Estimación: Poisson, Quasi-Poisson y Binomial Negativa . . .                       | 79        |
| 4.4.3    | Test de sobredispersión y comparación de modelos . . . . .                         | 79        |
| 4.4.4    | IRR y efectos marginales del modelo NB . . . . .                                   | 79        |
| 4.5      | Extensiones: modelos inflados en cero . . . . .                                    | 81        |
| 4.6      | Casos prácticos . . . . .                                                          | 81        |
| 4.6.1    | Caso práctico 1: Delitos por municipio . . . . .                                   | 81        |
| 4.6.2    | Caso práctico 2: Accidentes de tráfico por tramo . . . . .                         | 82        |
| 4.6.3    | Caso práctico 3: Publicaciones académicas . . . . .                                | 82        |
| 4.7      | Guía de comandos R para este capítulo . . . . .                                    | 84        |
| 4.7.1    | glm(..., family = poisson) — Regresión de Poisson . . . . .                        | 84        |
| 4.7.2    | AER::dispersiontest() — Test de sobredispersión de<br>Cameron-Trivedi . . . . .    | 84        |

|          |                                                            |           |
|----------|------------------------------------------------------------|-----------|
| 4.7.3    | MASS::glm.nb() — Regresión Binomial Negativa . . . . .     | 84        |
| 4.7.4    | exp(coef()) — Tasas de Incidencia (IRR) . . . . .          | 85        |
| 4.7.5    | mean(fitted(nb)) * coef(nb) — Efectos marginales AME . . . | 85        |
| 4.7.6    | AIC() — Comparación de modelos por criterio de información | 86        |
| <b>5</b> | <b>Datos de Panel</b>                                      | <b>87</b> |
| 5.1      | De la fotografía a la película . . . . .                   | 87        |
| 5.2      | Estructura de los datos de panel . . . . .                 | 88        |
| 5.2.1    | Notación y dimensiones . . . . .                           | 88        |
| 5.2.2    | Ejemplo numérico . . . . .                                 | 88        |
| 5.3      | La descomposición within/between . . . . .                 | 89        |
| 5.4      | El modelo general y las tres estrategias . . . . .         | 90        |
| 5.4.1    | Especificación del modelo . . . . .                        | 90        |
| 5.4.2    | Pooled OLS: ignorar la estructura de panel . . . . .       | 92        |
| 5.5      | Efectos Fijos: el estimador within . . . . .               | 93        |
| 5.5.1    | La transformación within . . . . .                         | 93        |
| 5.5.2    | Ejemplo numérico paso a paso . . . . .                     | 93        |
| 5.5.3    | Propiedades del estimador de efectos fijos . . . . .       | 94        |
| 5.5.4    | Grados de libertad . . . . .                               | 95        |
| 5.6      | Primeras diferencias . . . . .                             | 95        |
| 5.7      | Efectos Aleatorios: el estimador GLS . . . . .             | 95        |
| 5.7.1    | El supuesto clave . . . . .                                | 95        |
| 5.7.2    | La transformación cuasi-within . . . . .                   | 96        |
| 5.8      | El test de Hausman . . . . .                               | 96        |
| 5.8.1    | Formulación . . . . .                                      | 96        |
| 5.8.2    | Regla de decisión . . . . .                                | 97        |
| 5.9      | Batería de diagnósticos . . . . .                          | 98        |
| 5.9.1    | Test F de efectos individuales . . . . .                   | 98        |
| 5.9.2    | Test LM de Breusch-Pagan . . . . .                         | 98        |
| 5.9.3    | Test de Wooldridge para autocorrelación serial . . . . .   | 98        |
| 5.9.4    | Errores robustos (clustered) . . . . .                     | 98        |
| 5.10     | Ejemplo práctico: datos de Grunfeld . . . . .              | 98        |
| 5.10.1   | Estimación comparativa . . . . .                           | 99        |
| 5.10.2   | Tests de especificación . . . . .                          | 100       |
| 5.10.3   | Efectos fijos individuales . . . . .                       | 100       |
| 5.10.4   | Comparación visual: Pooled OLS vs Efectos Fijos . . . . .  | 100       |
| 5.11     | Extensiones . . . . .                                      | 102       |
| 5.11.1   | Efectos fijos bidireccionales . . . . .                    | 102       |
| 5.11.2   | Diferencias en Diferencias (DiD) . . . . .                 | 102       |
| 5.11.3   | Paneles dinámicos . . . . .                                | 102       |

|          |                                                                |            |
|----------|----------------------------------------------------------------|------------|
| 5.12     | Esquema de decisión . . . . .                                  | 102        |
| 5.13     | Casos prácticos . . . . .                                      | 103        |
| 5.13.1   | Caso Práctico 1: Inversión empresarial . . . . .               | 103        |
| 5.13.2   | Caso Práctico 2: Desempleo regional . . . . .                  | 104        |
| 5.13.3   | Caso Práctico 3: Salarios y formación . . . . .                | 104        |
| 5.14     | Lo esencial del capítulo . . . . .                             | 105        |
| 5.15     | Funciones esenciales de R para panel . . . . .                 | 105        |
| 5.15.1   | plm() — Estimación de modelos de panel . . . . .               | 105        |
| 5.15.2   | fixef() y ranef() — Efectos individuales estimados . . . . .   | 106        |
| 5.15.3   | phtest() — Test de Hausman . . . . .                           | 106        |
| 5.15.4   | pFtest() y plmtest() — Tests de efectos individuales . . . . . | 106        |
| 5.15.5   | pwartest() — Test de autocorrelación de Wooldridge . . . . .   | 107        |
| 5.15.6   | coeftest() con vcovHC() — Errores robustos clustered . . . . . | 107        |
| <b>6</b> | <b>Estimación con Variables Instrumentales</b>                 | <b>108</b> |
| 6.1      | El problema de la endogeneidad . . . . .                       | 108        |
| 6.2      | El sesgo de endogeneidad: demostración formal . . . . .        | 109        |
| 6.2.1    | Inconsistencia de MCO . . . . .                                | 109        |
| 6.2.2    | Dirección del sesgo . . . . .                                  | 109        |
| 6.2.3    | Ejemplo numérico . . . . .                                     | 110        |
| 6.3      | La idea del instrumento . . . . .                              | 111        |
| 6.3.1    | Las dos condiciones . . . . .                                  | 111        |
| 6.3.2    | Ejemplos clásicos de instrumentos . . . . .                    | 111        |
| 6.4      | El estimador IV simple . . . . .                               | 112        |
| 6.4.1    | Derivación paso a paso . . . . .                               | 112        |
| 6.4.2    | Ejemplo numérico completo . . . . .                            | 112        |
| 6.4.3    | Consistencia del estimador IV . . . . .                        | 113        |
| 6.4.4    | El coste de IV: mayor varianza . . . . .                       | 113        |
| 6.5      | Mínimos Cuadrados en Dos Etapas (MC2E) . . . . .               | 114        |
| 6.5.1    | Generalización a múltiples regresores . . . . .                | 114        |
| 6.5.2    | Formulación matricial . . . . .                                | 114        |
| 6.5.3    | Identificación y sobreidentificación . . . . .                 | 115        |
| 6.6      | Diagnósticos fundamentales . . . . .                           | 115        |
| 6.6.1    | Test de instrumentos débiles (F de la primera etapa) . . . . . | 115        |
| 6.6.2    | Test de endogeneidad (Durbin-Wu-Hausman) . . . . .             | 116        |
| 6.6.3    | Test de Sargan (sobreidentificación) . . . . .                 | 116        |
| 6.7      | Ejemplo práctico: retorno de la educación . . . . .            | 117        |
| 6.7.1    | Comparación MCO vs IV . . . . .                                | 117        |
| 6.7.2    | Diagnósticos . . . . .                                         | 117        |
| 6.8      | Casos prácticos . . . . .                                      | 117        |

|          |                                                                |            |
|----------|----------------------------------------------------------------|------------|
| 6.8.1    | Caso Práctico 1: Retorno de la educación . . . . .             | 117        |
| 6.8.2    | Caso Práctico 2: Gasto público y crecimiento . . . . .         | 119        |
| 6.8.3    | Caso Práctico 3: Estimación de la curva de demanda . . . . .   | 119        |
| 6.9      | Lo esencial del capítulo . . . . .                             | 120        |
| 6.10     | Funciones esenciales de R para IV . . . . .                    | 120        |
| 6.10.1   | ivreg() — Estimación por Variables Instrumentales . . . . .    | 120        |
| 6.10.2   | summary() con diagnostics — Diagnósticos automáticos . . . . . | 121        |
| 6.10.3   | coeftest() con vcovHC() — Errores robustos para IV . . . . .   | 121        |
| <b>7</b> | <b>Modelos de Duración</b>                                     | <b>122</b> |
| 7.1      | Cuando el tiempo importa . . . . .                             | 122        |
| 7.2      | Conceptos fundamentales . . . . .                              | 123        |
| 7.2.1    | La función de supervivencia . . . . .                          | 123        |
| 7.2.2    | La función de riesgo (hazard) . . . . .                        | 123        |
| 7.2.3    | El riesgo acumulado . . . . .                                  | 124        |
| 7.2.4    | Ejemplo numérico: cálculo de $S(t)$ y $h(t)$ . . . . .         | 124        |
| 7.3      | Tipos de censura . . . . .                                     | 125        |
| 7.4      | Estimación no paramétrica: Kaplan-Meier . . . . .              | 126        |
| 7.4.1    | El estimador . . . . .                                         | 126        |
| 7.4.2    | Ejemplo numérico paso a paso . . . . .                         | 126        |
| 7.4.3    | Test de log-rank . . . . .                                     | 127        |
| 7.5      | Modelos paramétricos . . . . .                                 | 127        |
| 7.5.1    | El modelo exponencial . . . . .                                | 127        |
| 7.5.2    | El modelo Weibull . . . . .                                    | 127        |
| 7.5.3    | Otros modelos paramétricos . . . . .                           | 127        |
| 7.5.4    | Verosimilitud con datos censurados . . . . .                   | 128        |
| 7.6      | El modelo de Cox . . . . .                                     | 129        |
| 7.6.1    | Riesgos proporcionales . . . . .                               | 129        |
| 7.6.2    | Interpretación: los Hazard Ratios . . . . .                    | 129        |
| 7.6.3    | La verosimilitud parcial de Cox . . . . .                      | 130        |
| 7.6.4    | Diagnósticos: test de proporcionalidad . . . . .               | 130        |
| 7.7      | Ejemplo práctico: duración del desempleo . . . . .             | 130        |
| 7.7.1    | Estimación del modelo de Cox . . . . .                         | 131        |
| 7.7.2    | Verificación de proporcionalidad . . . . .                     | 132        |
| 7.7.3    | Comparación con modelo Weibull . . . . .                       | 132        |
| 7.8      | Casos prácticos . . . . .                                      | 132        |
| 7.8.1    | Caso Práctico 1: Duración del desempleo . . . . .              | 132        |
| 7.8.2    | Caso Práctico 2: Supervivencia de empresas . . . . .           | 132        |
| 7.8.3    | Caso Práctico 3: Tiempo hasta impago crediticio . . . . .      | 133        |
| 7.9      | Lo esencial del capítulo . . . . .                             | 134        |

|          |                                                                |            |
|----------|----------------------------------------------------------------|------------|
| 7.10     | Funciones esenciales de R para duración . . . . .              | 134        |
| 7.10.1   | Surv() — Crear el objeto de supervivencia . . . . .            | 134        |
| 7.10.2   | survfit() — Estimación de Kaplan-Meier . . . . .               | 134        |
| 7.10.3   | survdif() — Test de log-rank . . . . .                         | 135        |
| 7.10.4   | coxph() — Modelo de riesgos proporcionales de Cox . . . . .    | 135        |
| 7.10.5   | cox.zph() — Test de proporcionalidad de Schoenfeld . . . . .   | 135        |
| 7.10.6   | survreg() — Modelos paramétricos . . . . .                     | 136        |
| <b>8</b> | <b>Datos Cualitativos</b>                                      | <b>137</b> |
| 8.1      | Variables cualitativas en econometría . . . . .                | 137        |
| 8.2      | PARTE I: Variables cualitativas como regresores . . . . .      | 137        |
| 8.2.1    | Variables ficticias (dummies) . . . . .                        | 137        |
| 8.2.2    | Interacciones y cambio estructural . . . . .                   | 139        |
| 8.2.3    | Modelos ANOVA . . . . .                                        | 139        |
| 8.2.4    | Modelos ANCOVA . . . . .                                       | 140        |
| 8.3      | PARTE II: Variable dependiente cualitativa . . . . .           | 140        |
| 8.3.1    | Más allá de la elección binaria . . . . .                      | 140        |
| 8.3.2    | El Logit Multinomial . . . . .                                 | 141        |
| 8.3.3    | Efectos marginales en el Logit Multinomial . . . . .           | 141        |
| 8.3.4    | La propiedad IIA y sus implicaciones . . . . .                 | 141        |
| 8.3.5    | Modelos de elección ordenada . . . . .                         | 142        |
| 8.3.6    | Bondad del ajuste en modelos cualitativos . . . . .            | 143        |
| 8.4      | Ejemplo práctico: elección de transporte . . . . .             | 144        |
| 8.5      | Casos prácticos . . . . .                                      | 144        |
| 8.5.1    | Caso Práctico 1: Elección de transporte . . . . .              | 144        |
| 8.5.2    | Caso Práctico 2: Satisfacción laboral . . . . .                | 146        |
| 8.5.3    | Caso Práctico 3: Situación laboral . . . . .                   | 146        |
| 8.6      | Lo esencial del capítulo . . . . .                             | 147        |
| 8.7      | Funciones esenciales de R . . . . .                            | 147        |
| 8.7.1    | factor() y lm() — Variables ficticias y ANOVA/ANCOVA . . . . . | 147        |
| 8.7.2    | multinom() — Logit Multinomial . . . . .                       | 148        |
| 8.7.3    | avg_slopes() — Efectos marginales medios . . . . .             | 148        |
| 8.7.4    | polr() — Probit y Logit Ordenado . . . . .                     | 148        |

|                     |            |
|---------------------|------------|
| <b>Bibliografía</b> | <b>150</b> |
|---------------------|------------|

# Prólogo

Este manual nació de una frustración y de una convicción.

La frustración es la que siente cualquier profesor que, tras años de enseñanza, constata que los materiales disponibles no terminan de funcionar. Los manuales internacionales de referencia — Cameron y Trivedi, Wooldridge, Greene — son obras maestras, pero están escritos para investigadores, no para alumnos que se enfrentan por primera vez a un modelo Logit o a un panel de datos. Los apuntes de las universidades, por su parte, suelen pecar de lo contrario: son telegráficos, dan por supuestos conceptos que el alumno no domina, y saltan de la fórmula al código sin detenerse a explicar el porqué de cada paso.

La convicción es que la microeconometría puede explicarse de otra manera. Con rigor, sí — las derivaciones formales están aquí, completas y sin atajos —, pero también con claridad pedagógica, con ejemplos numéricos resueltos paso a paso, con gráficos que iluminan la intuición detrás de las fórmulas, y con un hilo narrativo que conecta cada modelo con los problemas económicos reales que lo motivaron.

Este manual está escrito pensando en mis alumnos. En los que llegan al curso con buena base de estadística pero nunca han visto una variable latente. En los que saben ejecutar `lm()` en R pero no entienden qué pasa cuando la variable dependiente es binaria. En los que necesitan no solo saber *qué* hace un modelo, sino *por qué* lo hace, *cuándo* usarlo y *cómo* interpretar sus resultados de forma que un economista — no un estadístico — pueda tomar decisiones informadas.

Cada capítulo sigue una estructura deliberada: primero la motivación económica del problema, después la formulación del modelo con sus supuestos explícitos, a continuación la estimación y sus propiedades, luego la interpretación detallada de los resultados (que es donde la mayoría de los manuales se quedan cortos), los diagnósticos necesarios para validar el modelo, y finalmente un ejemplo práctico completo en R que el alumno puede reproducir íntegramente. Los scripts de R que acompañan cada capítulo están disponibles en el repositorio del manual:

<https://github.com/carlanta/MicroEconometrics>

Este manual cubre los modelos fundamentales de la microeconometría: elección discreta (Probit y Logit), datos censurados (Tobit), datos de recuento (Poisson y Binomial Negativa), datos de panel (efectos fijos y aleatorios), variables instru-

mentales (MC2E), modelos de duración (Cox, Weibull) y datos cualitativos (Logit Multinomial y Ordenado).

Escribir este manual ha sido un trabajo enorme — mucho mayor de lo que imaginé cuando comencé. Pero cada vez que un alumno me dice que por fin entiende qué son los efectos marginales, o que ahora sabe interpretar un odds ratio, o que ha conseguido replicar un paper con los modelos del curso, siento que el esfuerzo ha merecido la pena. Si este libro consigue que algún alumno más disfrute aprendiendo microeconometría, habrá cumplido su propósito.

*Carlos de Anta Puig  
Madrid, marzo de 2026*

# Capítulo 1

## Datos con Variable Dependiente Limitada

### 1.1 Cuando la variable dependiente no es libre

En la asignatura de econometría se aprende a estimar modelos de regresión lineal por MCO, donde la variable dependiente  $y$  podía tomar cualquier valor real: el salario, las ventas, el PIB. Sin embargo, en una proporción enorme de las aplicaciones microeconómicas, la variable que queremos explicar **no es libre de tomar cualquier valor**. Cuando su rango está restringido de alguna forma, hablamos de una **variable dependiente limitada** (*limited dependent variable*, LDV), y esa limitación tiene consecuencias profundas para la estimación.

Pensemos en tres situaciones concretas que todo economista aplicado encuentra de forma habitual. Cuando analizamos la participación laboral, la variable de interés es **binaria**: la persona trabaja o no trabaja, codificado como 0 o 1. Cuando estudiamos la elección de medio de transporte, la variable es **multinomial**: el viajero selecciona entre coche, autobús, tren o avión, y ninguna de esas opciones tiene un valor numérico intrínseco. Cuando modelamos el gasto en seguros médicos, la variable es **continua pero censurada**: muchas familias no gastan nada (el valor observado es exactamente cero), mientras que las demás presentan un gasto estrictamente positivo. En los tres casos, el modelo de regresión lineal clásico genera problemas graves que hacen inválida la inferencia estadística.

Este capítulo establece el **mapa del territorio** que recorreremos en el resto del manual. Presentamos los distintos tipos de variables dependientes limitadas, explicamos por qué MCO falla en cada caso y desarrollamos las distribuciones estadísticas que subyacen a cada familia de modelos. Al final del capítulo introducimos la caja de herramientas del **análisis exploratorio de datos** (EDA), que el alumno deberá aplicar sistemáticamente antes de estimar cualquier modelo.

### 1.1.1 Por qué MCO no es válido con variables dependientes limitadas

El modelo lineal clásico  $y_i = \mathbf{x}_i' \beta + u_i$  produce predicciones  $\hat{y}_i = \mathbf{x}_i' \hat{\beta}$  que pueden tomar cualquier valor en  $\mathbb{R}$ . Cuando  $y_i$  solo puede ser 0 o 1, esto genera tres problemas simultáneos e irremediables.

En primer lugar, las **predicciones son incoherentes**: nada impide a MCO predecir probabilidades negativas o mayores que uno. Una probabilidad de  $-0.12$  o de  $1.34$  carece de interpretación. En segundo lugar, los **errores son heterocedásticos por construcción**: si  $y_i \in \{0, 1\}$  y  $p_i = P(y_i = 1 \mid \mathbf{x}_i)$ , entonces  $\text{Var}(u_i) = p_i(1 - p_i)$ , que depende de  $\mathbf{x}_i$ , violando de forma sistemática los supuestos del teorema de Gauss-Markov. En tercer lugar, los **contrastes de hipótesis son inválidos**: al no satisfacerse los supuestos distribucionales, los estadísticos  $t$  y  $F$  no siguen sus distribuciones teóricas en muestras finitas, lo que hace que los  $p$ -valores sean incorrectos.

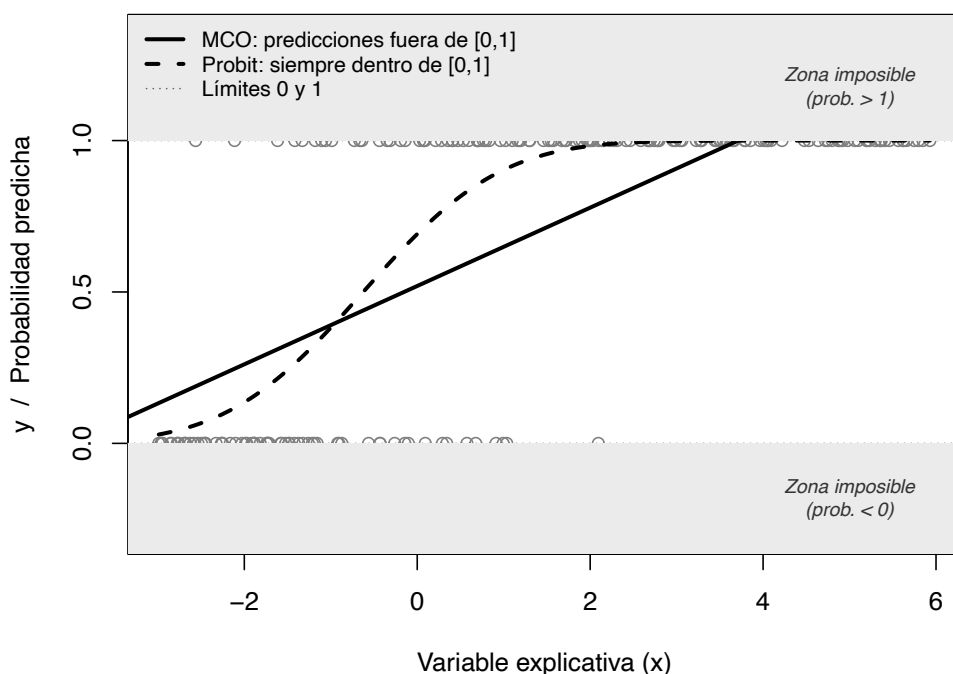


Figura 1.1: El modelo lineal de probabilidad (MCO) aplicado a una variable binaria produce predicciones fuera del intervalo  $[0, 1]$  — zonas sombreadas — mientras que el modelo Probit correcto mantiene las predicciones dentro del rango válido.

La Figure 1.1 lo ilustra sin ambigüedad: la recta de MCO traspasa las barreras naturales de la probabilidad. El modelo Probit, en cambio, utiliza la función de distri-

bución acumulada de la Normal para garantizar que todas las predicciones estén en  $[0, 1]$ . Cuando la variable dependiente está limitada, necesitamos modelos que respeten esas restricciones por diseño.

## 1.2 La taxonomía de las variables dependientes limitadas

Las variables dependientes limitadas se organizan en tres grandes familias según la naturaleza de la restricción que sufre  $y_i$ . La clasificación no es meramente académica: cada familia conduce a un tipo de modelo diferente, y entender bien esta arquitectura es el primer paso para elegir la herramienta correcta ante un problema econométrico real.

La **primera familia** comprende las variables **binarias** o dicotómicas, donde  $y_i \in \{0, 1\}$ . La restricción es total: solo existen dos resultados posibles. El decisor trabaja o no trabaja, contrata o no contrata, compra o no compra. Los modelos apropiados son el **Probit** y el **Logit**, que estudiaremos en el Chapter 2.

La **segunda familia** son las variables **multinomiales**, donde  $y_i \in \{0, 1, \dots, J\}$  con  $J \geq 2$ . Dentro de esta familia se distingue entre variables ordenadas — el nivel de satisfacción va de 1 a 5, y existe un orden natural entre las categorías — y variables no ordenadas, como la elección de modo de transporte, donde las categorías no tienen jerarquía. Esta familia exige modelos de elección múltiple que van más allá del alcance de este manual.

La **tercera familia** son las variables **continuas limitadas**, donde  $y_i$  es continua pero solo puede tomar valores en un subconjunto de  $\mathbb{R}$ : típicamente  $y_i \geq 0$  (censura por la izquierda en cero). El gasto en seguros, las horas trabajadas y la cantidad demandada de un bien son ejemplos clásicos. El modelo **Tobit** y sus extensiones son la herramienta adecuada, desarrollada en el el Capítulo 3.

Una cuarta familia, que comparte rasgos con la anterior, son las **variables de recuento**:  $y_i \in \{0, 1, 2, \dots\}$ , donde  $y_i$  solo toma valores enteros no negativos. El número de visitas al médico, el número de accidentes de tráfico o el número de patentes son ejemplos habituales. Los **modelos de Poisson y Binomial Negativa** son los instrumentos adecuados, objeto del el Capítulo 4.

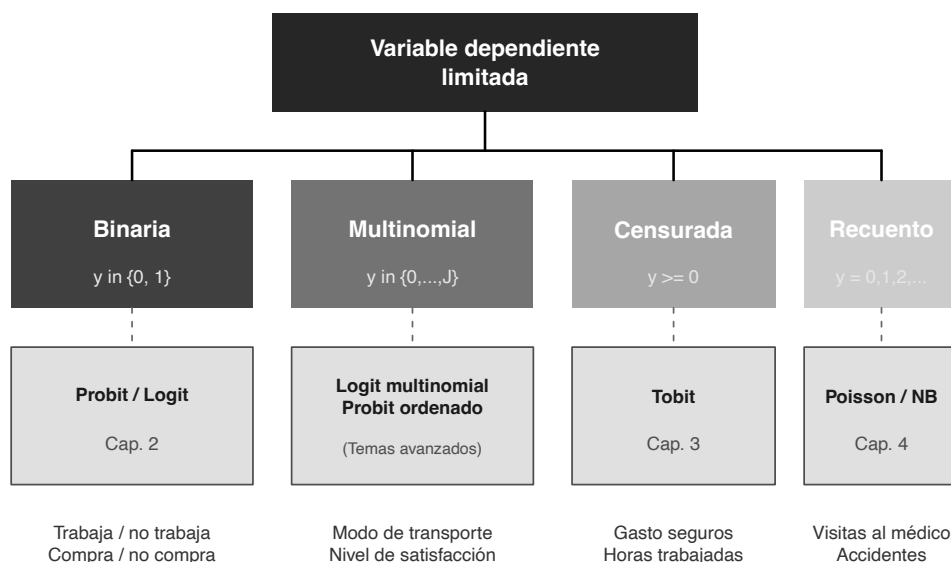


Figura 1.2: Clasificación de las variables dependientes limitadas. Cada familia conduce a una familia de modelos diferente. Las flechas indican el capítulo del manual donde se desarrolla cada modelo.

## 1.3 Variables binarias: la decisión sí o no

### 1.3.1 La distribución de Bernoulli

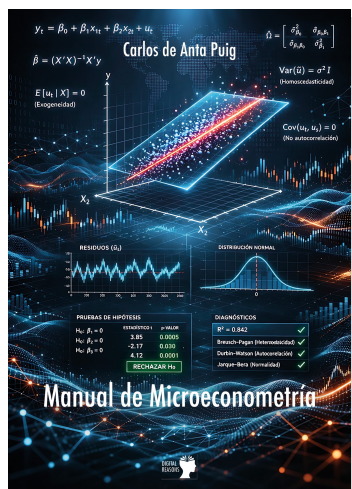
Una variable dependiente **binaria** solo puede tomar dos valores:  $y_i = 1$  si ocurre el evento de interés (trabajar, comprar, aprobar),  $y_i = 0$  en caso contrario. Estadísticamente, sigue una distribución **Bernoulli**( $p$ ):

$$P(y = 1) = p \quad P(y = 0) = 1 - p$$

En forma compacta, la función de masa de probabilidad es  $f(y; p) = p^y (1-p)^{1-y}$  con  $y \in \{0, 1\}$ . Sus momentos son  $E(y) = p$  y  $\text{Var}(y) = p(1-p)$ . El hecho de que la varianza dependa de  $p$  — que a su vez depende de  $\mathbf{x}_i$  en cualquier modelo econométrico — confirma directamente que la homocedasticidad se viola de forma inherente con variables binarias. No es un fallo del modelo: es una característica matemática de la distribución Bernoulli.

La clave de los modelos que construiremos en el siguiente capítulo es hacer que  $p$  **no sea la misma para todos los individuos**, sino que varíe en función de sus características observables:  $p_i = F(\mathbf{x}'_i \beta)$ , donde  $F$  es una función de distribución acumulada que garantiza  $p_i \in [0, 1]$  para cualquier valor del índice lineal  $\mathbf{x}'_i \beta$ .

# ¿Quiere seguir leyendo?



Acaba de leer el comienzo del *Manual de Microeconometría*, de Carlos de Anta Puig. El libro completo recorre, en 8 capítulos, los modelos fundamentales de la microeconometría, con derivaciones completas, interpretación detallada de los resultados y ejemplos reproducibles en R.

**Comprar el libro completo ›**

[bibliotecaonline.es/manual-de-microeconometria](http://bibliotecaonline.es/manual-de-microeconometria)

## QUÉ ENCONTRARÁ EN EL LIBRO

- **Variable dependiente limitada.** Modelo lineal de probabilidad, Probit y Logit con efectos marginales AME y MEM, y el modelo Tobit para datos censurados.
- **Datos de recuento.** Regresión de Poisson, sobredispersión, Binomial Negativa y modelos inflados en cero.
- **Panel e instrumentos.** Efectos fijos y aleatorios, test de Hausman, diferencias en diferencias y MC2E con sus diagnósticos.
- **Duración y datos cualitativos.** Kaplan-Meier, Weibull y modelo de Cox; Logit multinomial y modelos de elección ordenada.

*Con ejemplos numéricos resueltos paso a paso, tres casos prácticos por capítulo, una guía de comandos R y los scripts disponibles en el repositorio del manual.*

## Digital Reasons

[bibliotecaonline.es](http://bibliotecaonline.es)

[pedidos@bibliotecaonline.es](mailto:pedidos@bibliotecaonline.es)

ISBN (papel): 978-84-10093-39-3



Escanee para comprar