Elementos de Probabilidad y Estadística

Diplomatura en Ciencias de Datos

2024

Entrega 5

Dr. Matías Hisgen – Lic. Celine Cabás – Lic. Fernando Álvarez

FACENA - UNNE

Modelos con respuesta Binaria

- La idea consiste en modelar la *elección* entre dos opciones (*Modelo de Elección Binaria*) o la ocurrencia de un *evento* de interés (Modelo de probabilidad).

- El objetivo es conocer cómo cambia la probabilidad de ocurrencia de una elección o evento en función de características *observables* de los agentes intervinientes.

- Otro objetivo es poder clasificar observaciones nuevas en uno de los dos grupos según la probabilidad predicha (Problema de Clasificación).

Ejemplos

- **Ejemplos de elección binaria:**

  - Trabajadores: Participar (o no) del Sindicato

  - Empresas: Invertir (o no) en I&D

  - Estudiantes: Trabajar (o no) mientras estudian

- **Ejemplos de otros eventos:**

  - Empresas: Accedan (o no) a Crédito bancario

  - Estudiantes: Sean admitidos (o no) al Doctorado

  - Precio de acciones: Superen (o no) cierto límite

Problema econométrico

- Modelar la esperanza de una variable dependiente cualitativa (con dos categorías), condicional a un vector ***x*** de variables independientes.
- La variable explicada es una variable ***binaria*** *y* que puede tomar valores de 0 o 1 (“éxito” o “fracaso”), cuyo valor esperado o media es igual a la probabilidad de que *y* tome el valor 1 (“éxito”).
- Entonces, el objetivo concreto es definir

  $E(y|\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$

  a través de un modelo de regresión.

Modelo Lineal de Probabilidad (MLP)

- El MLP surge de modelar la probabilidad anterior como una función lineal del vector ***x***:

$$E(y|\boldsymbol{x}) = P(y = 1|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

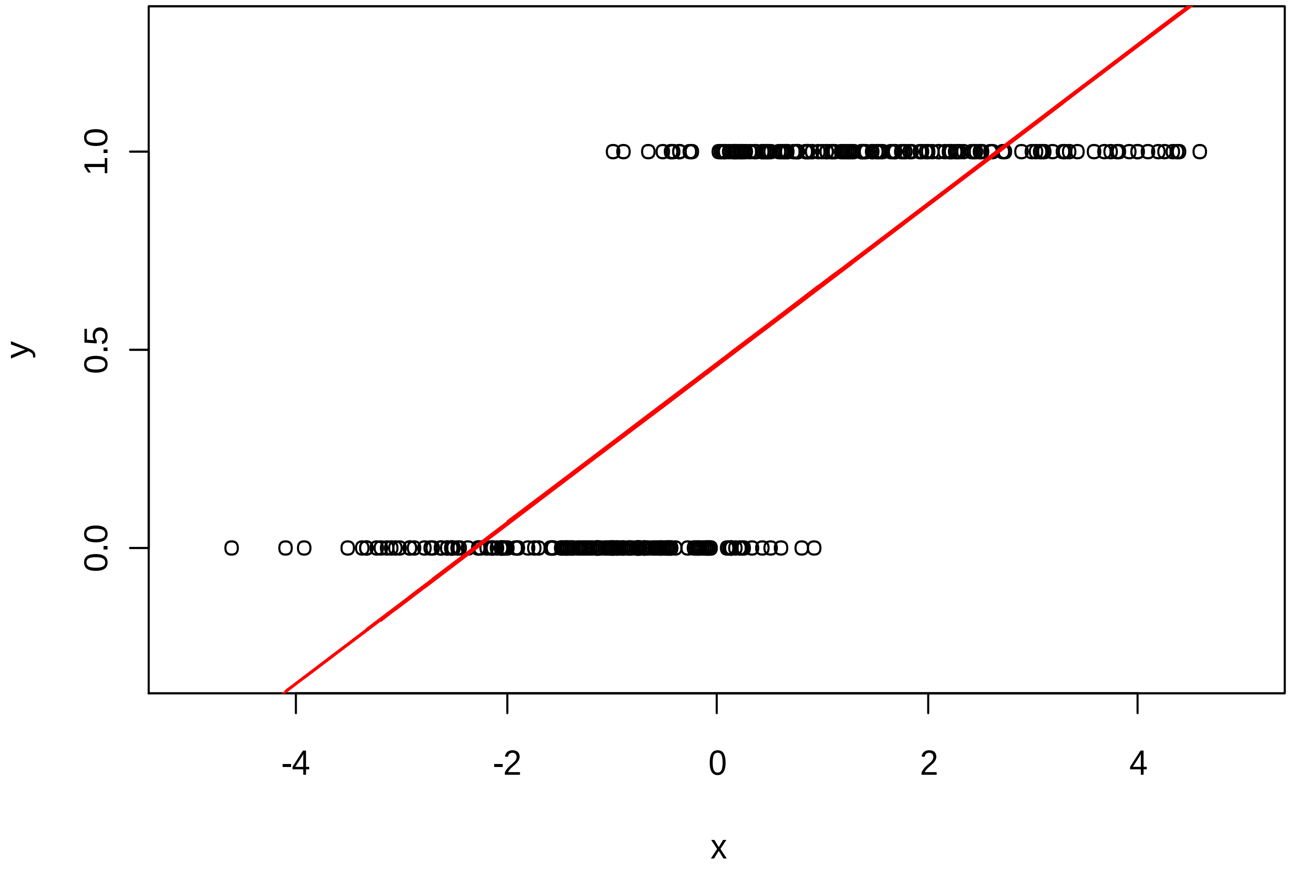
- Cada valor de y vendrá generado por el modelo

$$y = P(y = 1|\boldsymbol{x}) + u = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- El MLP puede ser estimado fácilmente por MCO mediante el ajuste del modelo muestral:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i$$

Ejemplo: ajuste del MLP



Interpretación e Inferencia

- Entonces, β_j representa el cambio en la probabilidad de "éxito" cuando x_j cambia en 1 unidad.
- El MLP se ajusta aplicando Mínimos Cuadrados Ordinarios (MCO), por lo que se aplica todo lo visto previamente en análisis de regresión simple y múltiple.
- En particular, los procedimientos de pruebas de hipótesis y construcción de intervalos de confianza son los mismos vistos previamente.

Series de Tiempo

- $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$
- Tendencia
- Ciclo
- Estacionalidad

Series de Tiempo vs. Corte Transv.

- Las series de tiempo tienen una *ordenación* temporal, a diferencia del corte transversal.
- Ejemplos: las mediciones de la tasa de desempleo mensual, el índice Merval diario, las ventas semanales, etc.
- Puede ser de interés: 1) estudiar la dinámica de una única serie a lo largo del tiempo (Ciclo, Tendencia y Estacionalidad) y 2) estudiar cómo se relacionen 2 o más series de tiempo.

Ciclo: Modelo Autoregresivo (AR)

- El modelo más básico para modelar una serie de tiempo es el ***Autoregresivo*** de orden 1 o AR(1):

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

- En él, el valor de *y* el período *t* solo depende del valor del período pasado *t-1*.
- Ésta es la forma más básica de modelar un ciclo dentro de la serie de tiempo.

Modelo AR(p)

- Extendiendo la cantidad de rezagos a "*p*" se tiene

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + u_t$$

- En éste caso, el componente cíclico tiene una duración mayor que en el caso AR(1).

Series con tendencia temporal

- Las Series muchas veces presentan tendencia.
- A veces, las series tendrán tendencia a causa de factores inobservables.
- Aunque tales factores sean no observables, es posible “dejarlos fijos” incluyendo una variable temporal como regresor.

Tipos de Tendencia

- Una tendencia *lineal* sería modelada como

$$y_t = \alpha_0 + \alpha_1 t + e_t, \ t = 1, 2, \ldots$$

- Otra posibilidad es la tendencia *exponencial*, la cual puede ser modelada como

$$\log(y_t) = \alpha_0 + \alpha_1 t + e_t, t = 1, 2, \ldots$$

- Una tendencia *cuadrática* sería modelada como $y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t, t = 1, 2, \ldots$

Estacionalidad

- Algunas veces las series presentan periodicidades, llamadas *estacionalidad*.
- Ejemplo: gastos en turismo, venta de calefactores, venta de helados.
- La estacionalidad puede ser incluída utilizando un conjunto de variables *binarias* (d)
- La binaria debe indicar si cada la observación de la serie pertenece (d=1) o no pertenece (d=0) al período estacional.

Ejemplo: Modelo AR(1) con tendencia y estacionalidad

- Un modelo autorregresivo de orden 1 con tendencia lineal y una estacionalidad, sería:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \alpha_1 t + \delta_1 d_1 + u_t$$

- En éste caso, se tiene el componente de *tendencia*, el componente *cíclico* y el de *estacionalidad*.

Combinando Series Temporales

- Un segundo objetivo de las regresiones con series temporales es relacionar dos o más series.
- Como antes. Una serie será el regresor (y_t) y las demás los regresores (x_{1t}, x_{2t} … x_{kt}):
- Un modelo *estático* sería:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_3 x_{3t} + u_t$$

el cual incorpora variables contemporáneas, es decir sin rezagos.

Combinando Series Temporales II

- Un modelo dinámico (con variables rezagadas) sería por ejemplo:

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

- En general, un modelo dinámico puede incluir, además de variables rezagadas, los componentes de tendencia, ciclo y estacionalidad:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_1 x_{1t} + \beta_2 x_{1t-1} + \alpha_1 t + \delta_1 d_1 + u_t$$