Regresión simple

Gabriel V. Montes-Rojas

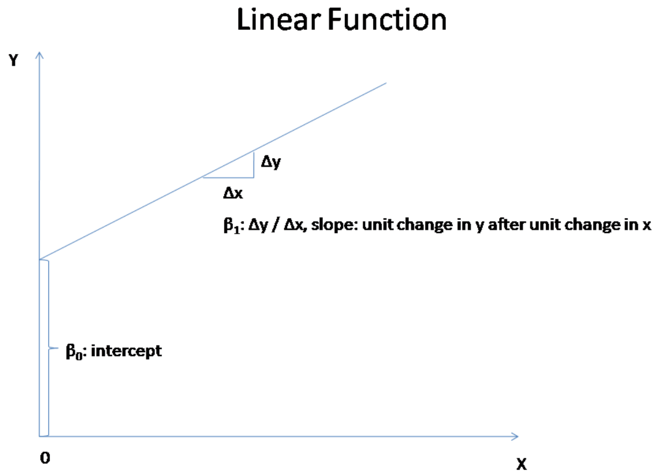
Regresión simple

Regresión simple significa de una sola variable de control (x es un escalar):

$$y_i = \beta_0 + \beta_1 x_i + u_i, \ \ i = 1, 2, ..., n$$

Elementos básicos:

1. Muestra/datos $\{y_i, x_i\}_{i=1}^{n}$.
2. Modelo lineal $y = \beta_0 + \beta_1 x$
   - y variable dependiente, lo que queremos explicar.
   - x variable independiente/de control/explicativa, cómo la vamos a explicar.
   - β_0 intercepto, valor de y cuando $x = 0$
   - $\beta_1 = \frac{\Delta y}{\Delta x}$ pendiente, cuánto se incrementa y al incrementarse x por **1 unidad**.
3. u error o residuo, aquello que no podemos observar pero que afecta y.



(¿dónde están los errores?)

Regresión simple

Democracia y crecimiento.

- Datos de n países.
- y variable dependiente, PBI per capita.
- x variable independiente, índice de democracia.
- β_0 intercepto, valor de y cuando $x = 0$.
- $\beta_1 = \frac{\Delta y}{\Delta x}$ pendiente, cuánto se incrementa y al incrementarse x por **1 unidad**.
- u error o residuo, aquello que no podemos observar pero que afecta y.

$$PBIpercap_i = \beta_0 + \beta_1 Democracia_i + u_i, \;\; i = 1, 2, ..., n$$

Regresión simple

Educación y salarios.

- Datos de n individuos.
- y variable dependiente, salario.
- x variable independiente, años de educación.
- β_0 intercepto, valor de y cuando $x = 0$.
- $\beta_1 = \frac{\Delta y}{\Delta x}$ pendiente, cuánto se incrementa y al incrementarse x por **1 unidad**.
- u error o residuo, aquello que no podemos observar pero que afecta y.

$$Salario_i = \beta_0 + \beta_1 Educ_i + u_i, \ i = 1, 2, ..., n$$

Regresión simple

- Una forma de ver los modelos de regresión es la siguiente. Notemos que

$$\beta_1 = \frac{cov(y,x)}{var(x)},$$

bajo el supuesto de que $cov(x,u) = 0$, o sea que la variable explicativa no tiene relación con los errores.

- La prueba es sencilla:

$$\frac{cov(y,x)}{var(x)} = \frac{cov(\beta_0 + \beta_1 x + u, x)}{var(x)} = \frac{\beta_1 cov(x,x) + cov(x,u)}{var(x)}$$

(dado que $cov(.,.)$ se puede distribuir linealmente y $cov(cte, variable\ aleatoria) = 0$)

$$= \beta_1 + \frac{cov(x,u)}{var(x)} = \beta_1$$

(porque $cov(x,x) = var(x)$ y $cov(u,x) = 0$) esto significa que β_1 mide cuanto y se relaciona (covaría) con x, estandarizado por la varianza de x.

Mínimos cuadrados ordinarios

¿Cómo estimamos β_0 and β_1?

Tomemos los residuos (recordar que son no observables...), $u_i \equiv y_i - \beta_0 - \beta_1 x_i$, $i = 1, 2, ..., n$.

Ahora....
Cuadrados..: $\sum_i^n u_i^2 = \sum_i^n (y_i - \beta_0 - \beta_1 x_i)^2$
$+$Mínimos...: β_0 y β_1 que **minimiza** $\sum_i^n (y_i - \beta_0 - \beta_1 x_i)^2$
$+$Ordinarios... puede ser más complicado...
$=$ **Mínimos cuadrados ordinarios (MCO)**
OLS en inglés: Ordinary Least Squares

Método de los momentos: Mínimos cuadrados ordinarios

Momentos en la población	Momentos en la muestra
$E[u] = E[y - \beta_0 - \beta_1 x] = 0$	$n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$
$E[xu] = E[x(y - \beta_0 - \beta_1 x)] = 0$	$n^{-1} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$

- Sistema de 2 ecuaciones y 2 incógnitas... se puede resolver.
- β es un parámetro, $\hat{\beta}$ un estimador. β es un valor fijo (no lo sabemos...), $\hat{\beta}$ una variable aleatoria (depende de cada muestra...).
- Conceptos a repasar: esperanza o valor esperado $E[\cdot]$. Esperanza incondicional vs. esperanza condicional.
- Notación: $\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + ... + x_n$ (sumatoria); $n^{-1} \sum_{i=1}^{n} x_i$ (promedio).

Consideremos las dos condiciones de primer orden:

$$n^{-1}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)=0 \tag{1}$$

$$n^{-1}\sum_{i=1}^{n}x_i(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)=0 \tag{2}$$

De la primera ecuación

$$\bar{y}=\hat{\beta}_0+\hat{\beta}_1\bar{x}\ (demostrar)$$

- Notación: $\bar{x}=n^{-1}\sum_{i=1}^{n}x_i=n^{-1}(x_1+x_2+x_3+...+x_n)$ (promedio)

Entonces,

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}.$$

De la segunda ecuación

$$\sum_{i=1}^{n} x_i[y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i] = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} x_i(x_i - \bar{x})$$

Finalmente,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

El siguiente resultado lo vamos a usar muchas veces:

$$\sum_{i=1}^{n} a_i(b_i - \bar{b}) = \sum_{i=1}^{n} b_i(a_i - \bar{a}) = \sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})$$

(¡demostrar!)

Resumiendo:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Modelo sin intercepto

Supongamos un modelo que satisface $y_i = \beta x_i + u_i$, para $i = 1, 2, ..., n$ con $E(u_i|x_i) = 0$.

- Graficar una muestra $\{y_i, x_i\}_{i=1}^{n}$ con estas condiciones. ?'Qué restricciones tiene este modelo con respecto al modelo general? Este modelo también se llama de ordenada al origen.
- Plantear el estimador de MCO sin intercepto como una minimización y mostrar que
$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$
- ¿Qué debería encontrar si el modelo generador de datos es este pero se estima el modelo con intercepto?

Teorema de Gauss-Markov

- Supuesto 1: Lineal en los parámetros y se relaciona con x a traves de una función lineal.
- Supuesto 2: Muestra aleatoria $\{(y_i, x_i)\}_{i=1}^{n}$ es una muestra aleatoria del modelo del Supuesto 1.
- Supuesto 3: Variación muestral en x: $\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0$
- Supuesto 4: Media condicional cero $E(u|x) = 0$.

MCO es insesgado Si los Supuestos 1-4 se cumplen, entonces $E(\hat{\beta}_0|x) = \beta_0$ *and* $E(\hat{\beta}_1|x) = \beta_1$

Teorema de Gauss-Markov

- Supuesto 5: Homoscedasticidad $Var(u|x) = \sigma^2$

Teorema de Gauss-Markov: Si los Supuestos 1-5 se cumplen, el estimador MCO $\hat{\beta}_0, \hat{\beta}_1$ es el mejor estimador insesgado de β_0, β_1. Nota: MEJOR= menor varianza (repasar concepto de varianza $V[\cdot]$). Se llama EFICIENTE a un estimador que cumple esta propiedad.

Insesgadez

Los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ son insesgados.
Esto es, $E[\hat{\beta}_0|x] = \beta_0$ y $E[\hat{\beta}_1|x] = \beta_1$.
La prueba se puede hacer en pocos pasos.... a continuación.

Insesgadez

Para simplificar la notación escribimos $E(.)$ en vez de $E(.|x)$, o sea que las esperanzas incondicionales son en realidad esperanzas condicionales.

$$E[\hat{\beta}_1] = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

por la propiedad $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$.

$$... = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

by Supuesto 1: Lineal en los parámetros y se relaciona con x a través de una función lineal. O sea, $y = \beta_0 + \beta_1 x + u$.

$$... = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + E[u_i])}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

por propiedades de la esperanza. (Notemos que $E[u_i]$ es en realidad $E[u_i|x]$.)

- $E[\sum_{i=1}^{n}(.)] = \sum_{i=1}^{n} E[(.)]$
- $E[\beta_0 + \beta_1 x_i + u_i] = \beta_0 + \beta_1 x_i + E[u_i]$

Insesgadez

Por el Supuesto 4: Media Condicional Cero $E(u|x) = 0$.

$$... = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + 0)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Luego de algo de álgebra...

$$... = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\beta_0}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\beta_1 x_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0 + \beta_1 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1$$

Entonces probamos que $E[\hat{\beta}_1] = \beta_1$

Sesgo

Probar que $E[\hat{\beta}_0|x] = \beta_0$ es más fácil.
De la primera condición de momento de MCO

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Usando esperanzas en los dos lados,

$$E[\hat{\beta}_0] = E[\bar{y}] - E[\hat{\beta}_1\bar{x}]$$

Sabemos que $E[\bar{y}] = E[\beta_0 + \beta - 1\bar{x} + \bar{u}] = \beta_0 + \beta_1\bar{x} + E[\bar{u}] = \beta_0 + \beta_1\bar{x}$ y que $E[\hat{\beta}_1\bar{x}] = E[\hat{\beta}_1]\bar{x} = \beta_1\bar{x}$. Así obtenemos,

$$E[\hat{\beta}_0] = \beta_0.$$

Predicción

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ es el **valor de predicción** de y dado x_i, esto es, un estimador de $E(y|x_i)$
- $\hat{u}_i = y_i - \hat{y}_i$ es el **residuo de la regresión** o **error de predicción** para la observación i, o sea un estimador de $y_i - \beta_0 - \beta_1 x_i$.
- Usar gráficos para distinguir claramente $y_i, \hat{y}_i, u_i, \hat{u}_i$.
- Demostrar que $\sum_{i=1}^{n} \hat{u}_i = 0$ y $\sum_{i=1}^{n} x_i \hat{u}_i = 0$. ¿Qué implica?

Varianza de los estimadores MCO

¡¡Todo estimador se merece su varianza!!

$$Var(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba...

$$Var(\hat{\beta}_0|x) = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba...

Pregunta: $Var(\beta_1|x)$=??

Varianza de los estimadores MCO

$$Var(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba: (para simplificar la notación $var(.)$ corresponde a $var(.|x)$)

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_0}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= \frac{Var\left[\sum_{i=1}^n (x_i - \bar{x}) u_i\right]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var[u_i]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Varianza de los estimadores MCO

Usamos

- Supuesto 1: Modelo lineal en los parámetros y se relaciona con x por una función lineal.

O sea, $y = \beta_0 + \beta_1 x + u$.

Varianza de los estimadores MCO

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba:

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_0}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= \frac{Var\left[\sum_{i=1}^n (x_i - \bar{x})u_i\right]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var[u_i]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Varianza de los estimadores MCO

Usamos

- Propiedad de la varianza: $Var[aX + bY] = a^2 \times Var[X] + b^2 \times Var[Y] + 2ab \times Cov[X, Y]$, donde $Cov[X, Y] = E[XY] - E[X]E[Y]$
- Propiedad de la covarianza: $Cov[a, Y] = 0$, donde a es una constante y Y una variable aleatoria (también $Cov[a, b] = 0$, donde tanto a como b son constantes...)

Varianza de los estimadores MCO

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba:

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) \beta_0}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) \beta_1 x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= 0 + 0 + \frac{Var\left[\sum_{i=1}^n (x_i - \bar{x}) u_i\right]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var[u_i]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Varianza de los estimadores MCO

Usamos

- Propiedad de la varianza: $Var[a] = 0$ donde a es una constante.

Las X's son consideradas como constantes.

Varianza de los estimadores MCO

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba:

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_0}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= \frac{Var\left[\sum_{i=1}^n (x_i - \bar{x}) u_i\right]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var[u_i]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Varianza de los estimadores MCO

Usamos

- Supuesto 2: Muestreo aleatorio $\{(y_i, x_i)\}_{i=1}^{n}$ es una muestra aleatoria del modelo dado en el Supuesto 1.

Hacemos $Var[\sum_{i=1}^{n} u_i] = \sum_{i=1}^{n} Var[u_i] + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} Cov[u_i, u_j]$.

Pero, por la propiedad de muestreo aleatorio $Cov[u_i, u_j] = 0, \; i \neq j$

Entonces, $Var[\sum_{i=1}^{n} u_i] = \sum_{i=1}^{n} Var[u_i]$.

Varianza de los estimadores MCO

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba:

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_0}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] + Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$$

$$= \frac{Var\left[\sum_{i=1}^n (x_i - \bar{x}) u_i\right]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var[u_i]}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Varianza de los estimadores MCO

Usamos

- Supuesto 5: Homoscedasticidad $Var(u|x) = \sigma^2$

donde $Var[u_i] = Var[u_i|x] = \sigma^2$ for all $i = 1, 2, ..., n$

Varianza de los estimadores MCO

$$Var(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Prueba:

$$Var(\hat{\beta}_0) = Var\left[\bar{y} - \bar{x}\hat{\beta}_1\right] = Var\left[\bar{y}\right] + Var\left[\bar{x}\hat{\beta}_1\right] - 2Cov\left[\bar{y}, \bar{x}\hat{\beta}_1\right]$$

$$= \frac{\sigma^2}{n} + \bar{x}^2 Var\left[\hat{\beta}_1\right] - 2\frac{\bar{x}}{n\sum_{i=1}^n (x_i - \bar{x})^2} Cov\left[\sum_{i=1}^n y_i, \sum_{i=1}^n (x_i - \bar{x}) y_i\right]$$

$$= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - 2\frac{\bar{x}}{n\sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2 \sum_{i=1}^n (x_i - \bar{x})$$

$$= \frac{\sigma^2}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \bar{x}^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Inferencia

¡$\hat{\beta}_0$ y $\hat{\beta}_1$ son variables aleatorias!

- Supuesto 6: Normalidad u es independiente de x y $u \sim N(0, \sigma^2)$.

Distribución normal: Bajo los supuestos 1-6,

$$\hat{\beta}_0 \sim N(\beta_0, Var[\hat{\beta}_0])$$

$$\hat{\beta}_1 \sim N(\beta_1, Var[\hat{\beta}_1])$$

Entonces,

$$(\hat{\beta}_0 - \beta_0)/se(\hat{\beta}_0) \sim N(0, 1)$$
$$(\hat{\beta}_1 - \beta_1)/se(\hat{\beta}_1) \sim N(0, 1)$$

Contrastes de hipótesis (tests)

Los estimadores de MCO son variables aleatorias. Dependiendo de la muestra lo que estimamos podría estar cerca o lejos de los parámetros de la población. Lo importante es cuán cerca o lejos?
Consideremos la **hipótesis nula**

$$H_0 : \beta_1 = \beta_{10},$$

y contrastemos con la **hipótesis alternativa**

$$H_A : \beta_1 \neq \beta_{10}$$

Un ejemplo MUY usado es $H_0 : \beta_1 = 0$. ¿Hay relación entre y con x?
En la práctica tenemos que hacer inferencia acerca de si H_0 es verdad o no usando $\hat{\beta}_1$.

Contrastes de hipótesis (tests)

Si H_0 es verdad, entonces $\hat{\beta}_1$ debería estar cerca de β_{10}. Pero ¿por cuánto? ¿Cuán cerca es cerca?

Bajo $H_0 : \beta_1 = \beta_{10}$ y asumiendo que u tiene distribución normal estándar, tenemos el siguiente resultado importante

$$(\hat{\beta}_1 - \beta_{10}) / \widehat{se(\hat{\beta}_1)} \sim t_{n-2}$$

donde $se(.)$ son los errores estándar (standard errors) y t_{n-2} es la distribución "t de estudiante" (t-Student) con $n - 2$ grados de libertad.

Nota: Para obtener $Var(\hat{\beta})$ necesitamos estimar σ^2, la varianza del error. Usamos $\hat{\sigma}^2$.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}^2}{n-2}$$

El número 2 de los grados de libertad dice cuántos parámetros estamos estimando. Por otro lado $\widehat{se(.)}$ es el estimador del error estándar.

Contrastes de hipótesis (tests)

- Paso 1: ¿Qué hipótesis?

  En general queremos ver la significatividad estadística (**statistical significance**) de un coeficiente de regresión. O sea, $H_0 : \beta_1 = 0$ en el modelo $y = \beta_0 + \beta_1 x + u$.

  También pueder haber hipótesis nulas compuestas $H_0 : \beta_2 = 0, \beta_3 = 0$ en el modelo (ver más adelante)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \, x_2 + \beta_3 \, x_3 + u$$

Contrastes de hipótesis (tests)

- Paso 2: Nivel de significancia, α. En general:
  $\alpha = .1$ OR $\alpha = .05$ OR $\alpha = .01$
  Cuanto mas pequeño es α mas confianza se tiene en los resultados.

En Estadística se llama Error de Tipo I, el error de rechazar H_0 cuando es verdadera. Dado que estamos trabajando con variables aleatorias siempre podemos cometer errores.

Bajo H_0, $S = \hat{\beta}_1 - \beta_{10}$ debería estar cercano a 0. Entonces, la evidencia de que esto no es cierto debería estar asociado a un alto valor de S. Llamemos a S_α al valor tal que $P[|S| > |S_\alpha|] = \alpha$.
Nota:
- Modelo en **una dirección**: $H_A : \beta_1 > \beta_{10}$ (o $H_A : \beta_1 < \beta_{10}$). Entonces tenemos $P[S > S_\alpha] = \alpha$ (o $P[S < S_\alpha] = \alpha$).
- Modelo en **dos direcciones**: $H_A : \beta_1 \neq \beta_{10}$. Entonces tenemos dos valores críticos tal que $P[S > S^1_{\alpha/2} > 0] = \alpha/2$ y $P[S < S^2_{\alpha/2} < 0] = \alpha/2$. Si la distribución de S es simétrica, $S^1_{\alpha/2} = -S^2_{\alpha/2}$.

Contrastes de hipótesis (tests)

- Paso 3: Mirar el $p-valor$.
- El p-valor (para hipótesis en dos direcciones) es $P[|\hat{\beta}_1| > |\hat{\beta}_1^{obs}|]$ bajo la hipótesis nula, donde $\hat{\beta}_1^{obs}$ es el valor observado, es decir, en la muestra.
- Intuitivamente nos dice que probabilidad hay de encontrar un valor que nos más evidencia que el realmente observado. Si esta probabilidad es pequeña, entonces tenemos un valor muy distinto al que se asume en H_0.
  REGLA:
- Si $p-valor < \alpha$ entonces **rechazar** la hipótesis nula.
- Si $p-valor \geq \alpha$ entonces **aceptar** la hipótesis nula.

Contrastes de hipótesis (tests)

- Ejemplo: Si la hipótesis nula es $H_0 : \beta_1 = 0$ en el modelo $y = \beta_0 + \beta_1 x + u$, entonces
- Si $p - valor < \alpha$, **rechazar** $\Rightarrow \beta_1 \neq 0$, x tiene un efecto lineal sobre y.
- Si $p - valor \geq \alpha$, **aceptar** $\Rightarrow \beta_1 = 0$, x no tiene efecto lineal sobre y.

Contrastes de hipótesis (tests)

- Si la hipótesis es $H_0 : \beta_2 = 0, \beta_3 = 0$ para modelos de regresión múliple $y = \beta_0 + \beta_1 x_1 + \beta_2\, x_2 + \beta_3\, x_3 + u$, entonces
- Si $p - valor < \alpha$, **rechazar** $\Rightarrow \beta_2 \neq 0$ o $\beta_3 \neq 0$, x_2 y x_3 tienen **conjuntamente** un efecto lineal sobre y.
- Si $p - valor \geq \alpha$, **aceptar** $\Rightarrow \beta_2 = 0$ y $\beta_3 = 0$, x_2 y x_3 no tienen un efecto lineal sobre y.

Contrastes de hipótesis (tests)

- Paso 3 (alternativo): Mirar el estimador dividido el error estándar. Muchos trabjos empíricos reportan los coeficientes estimados y los errores estándar de esos estimadores. La idea es que $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ tiene aproximadamente una distribución normal, y para un $\alpha = 0.05$ el valor crítico es 2 (en una variable aleatoria $Z\ N(0,1)$, $P[Z > 1.96] = 0.025$.
  REGLA:
- Si $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} > 2$ entonces **rechazar** la hipótesis nula.
- Si $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \leq 2$ entonces **aceptar** la hipótesis nula.

Ejemplo: Retornos a la educación

$$wage = \beta_0 + \beta_1 educ + u$$

1976 Current Population Survey (CPS) de los Estados Unidos

use http://fmwww.bc.edu/ec-p/data/wooldridge/wage1, clear
(para abrir la base de datos)

reg wage educ (para correr la regresión)

Ejemplo: Retornos a la educación

$$
\begin{array}{rcc}
wage = & -.905+ & .541^{***} educ \\
 & (.685) & (.053) \\
 & < 0.187 > & < 0.000 > \\
 & [-1.321] & [10.2]
\end{array}
$$

(errores estándar); $< p - valor >$; $[t - valor]$; * significancia 10%; ** significancia 5%; *** significancia 1%;

- ¿Qué significa $\hat{\beta}_1 = .541$? cada año de educación incrementa el salario horario en promedio 54 centavos de dólar.
- ¿Es estadísticamente significativo? Ver el p-valor.
  Esto tiene implícita la hipótesis $H_0 : \beta_1 = 0$. $se(\hat{\beta}_1) = .053$,
  $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{.541}{.053} = 10.2$. Rechazar con el p-valor de 0.000.
- ¿Qué significa $\hat{\beta}_0 = -.905$? ¿Es significativo?

¿Cómo aparecen los resultados en STATA?

http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge2.html

Otros comandos en STATA

- Para obtener estadísticos de la base de datos tipear:

  `summ`
  (reporta para todas las variables: nro. de observaciones, promedio, desviaciones estándar, mínimo, máximo)
  `summ wage educ`
  (sólo para las variables especificadas)

- Más información para una variable (mediana, cuantiles, asimetría, curtosis)

  `summ VARIABLE, detail`
  (en VARIABLE va la variable de interés)

Otros comandos en STATA

- Valor predicho, $\hat{y}$, de una regresión,

  `predict NEWNAME`
  (en NEWNAME va el nombre que se le quiere dar a la nueva variable, por ejemplo *ypred*)
  (nota: antes hay que correr la regresión)

- Residuos de la regresión, $\hat{u} = y - \hat{y}$

  `predict NEWNAME, resid`
  (en NEWNAME va el nombre que se le quiere dar a la nueva variable, por ejemplo *upred*)

Gráficos en STATA

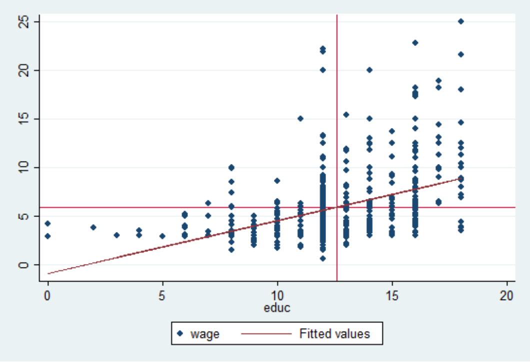
- Nube de puntos
  `scatter YVAR XVAR`
  (YVAR es la variable del eje vertical, XVAR es la variable del eje horizontal)
- Línea (conecta los puntos)
  `sort XVAR`
  `line YVAR XVAR`
  (YVAR es la variable del eje vertical, XVAR es la variable del eje horizontal)

Ejemplo:

predict wage_hat (para predecir los salarios, $\widehat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ$)

scatter wage educ || line wage_hat educ, xline(12.57) yline(5.90)

(hace un gráfico con la nube de puntos y la línea de regresión)




Julius
and Isabella
Dufferburger,
statistically average
family with
their 2.5
children