Notas de Econometría

Pavel Solís
2025

1 Modelo de Regresión Lineal Simple (RLS)

El modelo explica una variable en términos de otra

- El álgebra y la interpretación del modelo son relativamente sencillas
- Permite cubrir temas importantes de forma aislada
  - Cambios en unidades de medición, efectos no lineales

1.1 Definición del modelo

y y x son 2 variables que representan alguna población

- Queremos explicar y en términos de x

Preguntas que debemos responder:

- ¿Cómo permitimos que otros factores afecten a y si la relación no es exacta?
- ¿Cuál es la forma funcional de la relación entre x y y?
- ¿Cómo sabemos si capturamos una relación ceteris paribus entre x y y?

Empezamos con una ecuación simple para capturar la relación:

$$y = \beta_0 + \beta_1 x + u$$

- Suponemos que la ecuación se cumple en la población

La ecuación se conoce como:

- Modelo de **regresión lineal simple** (RLS)
- Modelo de regresión lineal de 2 variables
- Modelo de regresión lineal bivariado

La variable y se conoce como variable:

- **dependiente**
- explicada
- de respuesta
- predicha
- regresada

La variable x se conoce como variable:

- **independiente**
- explicativa
- de control
- predictor o predictiva
- **regresor**
- covariable

Los **parámetros** son β_0 y β_1

- β_0 es el parámetro del intercepto o la **constante** (rara vez es de interés)
- β_1 es el parámetro de la **pendiente** (interés principal)

La variable u:

- Se conoce como el **término de error** o de pertubación
- Captura todos los factores no observados distintos a x que afectan a y
- Ej. $y = salario$, $x = educ$, $u = experiencia,\ habilidad,\ antigüedad, ética\,laboral, \ldots$

La relación expresada en cambios es:

$$\Delta y = \beta_1 \Delta x + \Delta u$$

Si los factores en u se mantienen fijos,

$$\Delta u = 0 \implies \Delta y = \beta_1 \Delta x$$

- x tiene un efecto lineal en y
  - Mismo efecto en y sin importar el valor inicial de x

¿Cómo sabemos el efecto ceteris paribus de x sobre y si ignoramos todos los otros factores?

- Necesitamos hacer un supuesto para restringir la relación entre x y u
  - Usamos conceptos de probabilidad porque x y u son variables aleatorias
- Un supuesto poco restrictivo es
$$\mathbb{E}\left(u\right) = 0$$
  - Siempre podemos redefinir el intercepto β_0 para esto
- Posibilidades para que $\mathbb{E}\left(u\right) = 0$:
  - Suponer que u y x no estan correlacionadas
    - No es útil porque solo mide dependencia lineal
  - Definir una distribución condicional de u dado x
- Supuesto clave:
$$\mathbb{E}\left(u|x\right) = \mathbb{E}\left(u\right)$$
  - Dice que u es independiente en media de x
  - El valor esperado de u no depende de x y es igual al promedio de u para toda la población
- Combinando ambos supuestos, obtenemos el supuesto de **media condicional cero**
$$\mathbb{E}\left(u|x\right) = 0$$
  - Ej. $\mathbb{E}\left(habilidad|8\right) = \mathbb{E}\left(habilidad|6\right)$
  - El supuesto de media condicional cero da una nueva interpretación para β_1

La **función de regresión de la población** (FRP) es:

$$\mathbb{E}\left(y|x\right) = \mathbb{E}\left(\beta_0 + \beta_1 x + u|x\right) = \beta_0 + \beta_1 x$$

- *Promedio* de y (no la y que varía con u) es una función lineal de x
  - Si $\Delta x = 1$, $\mathbb{E}\left(y\right)$ sube en β_1

[Gráfica]

1.2 Derivación de los estimadores de MCO

Observamos x y y, pero no observamos β_0, β_1 ni u
Podemos estimar los parámetros β_0 y β_1 a partir de una muestra aleatoria de la población de tamaño n, $\{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$
En la población,

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad \forall\, i \implies u_i = y_i - \beta_0 - \beta_1 x_i$$

- u_i es el término de error de la observación i
  - Factores que afectan a y_i distintos de x_i

Con los supuestos $\mathbb{E}\left(u\right) = 0$ y $\mathbb{E}\left(u|x\right) = \mathbb{E}\left(u\right)$ podemos concluir que $\mathrm{Cov}\left(x, u\right) = 0$

$$\mathrm{Cov}\left(x, u\right) = \mathbb{E}\left(xu\right) - \mathbb{E}\left(x\right)\mathbb{E}\left(u\right) = \mathbb{E}\left(xu\right) = 0$$

- Entonces, obtenemos 2 restricciones para la distribución de probabilidad conjunta de x y y:
  - $\mathbb{E}\left(u\right) = \mathbb{E}\left(y - \beta_0 - \beta_1 x\right) = 0$
  - $\mathbb{E}\left(xu\right) = \mathbb{E}\left[x\left(y - \beta_0 - \beta_1 x\right)\right] = 0$

Escogemos $\widehat{\beta}_0$ y $\widehat{\beta}_1$ de forma que resuelvan las versiones muestrales de esas 2 restricciones (sistema 2×2):

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0 \implies \bar{y} - \widehat{\beta}_0 - \widehat{\beta}_1 \bar{x} = 0 \implies \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0 \implies \sum_{i=1}^{n} x_i\left[y_i - \left(\bar{y} - \widehat{\beta}_1 \bar{x}\right) - \widehat{\beta}_1 x_i\right] = 0$$

$$\implies \sum_{i=1}^{n} x_i\left(y_i - \bar{y}\right) - \widehat{\beta}_1 \sum_{i=1}^{n} x_i\left(x_i - \bar{x}\right) = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) - \widehat{\beta}_1 \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 0$$

$$\implies \widehat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2} = \widehat{\rho}_{xy}\frac{\widehat{\sigma}_y}{\widehat{\sigma}_x}$$

Estos estimadores se conocen como los **estimadores de mínimos cuadrados ordinarios** (MCO)

- El nombre viene de una forma alternativa de obtener las versiones muestrales de las 2 restricciones anteriores

- Escogemos $\widehat{\beta}_0$ y $\widehat{\beta}_1$ de forma que minimicen la suma de los residuales al cuadrado $\min \sum_{i=1}^{n} \widehat{u}_i^2 = \min \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2$
  - Se obtienen las mismas condiciones de primer orden que antes

- ¿Cuál es la única condición para calcular $\widehat{\beta}_0$ y $\widehat{\beta}_1$?
  - Variación en x_i

- Regresión simple es un análisis de correlación entre 2 variables
  - No implica causalidad
  - A veces es suficiente

Al calcular los estimados de MCO, podemos obtener 2 cosas:

- **Valores ajustados** o estimados para y cuando $x = x_i$:

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

- **Residuales** (no son los errores u_i):

$$\widehat{u}_i = y_i - \widehat{y}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$$

[Gráfica]

La línea de regresión de MCO es

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x \implies \Delta\widehat{y} = \widehat{\beta}_1 \Delta x$$

- Es la **función de regresión de la muestra** (FRM), varía con cada muestra
- Versión estimada de la FRP: $\mathbb{E}\left(y|x\right) = \beta_0 + \beta_1 x$, fija y desconocida en la población

[Gráfica]

Ejemplo. wage1.dta

$$\widehat{salario} = -0.90 + 0.54 educ$$

- Si educ = 8,

$$\widehat{salario} = -0.90 + 0.54(8) = 3.42$$

Terminología: Correr una regresión de y (variable dependiente) sobre x (variable independiente) para obtener $\widehat{\beta}_0$ y $\widehat{\beta}_1$

- Ej. Correr una regresión del *salario* sobre *educ*

1.3 Propiedades algebraicas de MCO

Estas propiedades se cumplen para cualquier muestra (por construcción)
Ayudan a entender qué le pasa a los estimados de MCO cuando manipulamos los datos

- Ej. Si cambiamos las unidades de medición

1.3.1 Valores ajustados y residuales

Cada valor ajustado de $\widehat{y}_i$ está sobre la línea de regresión de MCO
El residual de MCO para cada observación i es $\widehat{u}_i = y_i - \widehat{y}_i$

- Generalmente, $\widehat{u}_i \neq 0$ por lo que los puntos y_i no caen sobre la línea de MCO
- Si $\widehat{u}_i > 0$, la línea de regresión subestima y_i
- Si $\widehat{u}_i < 0$, la línea de regresión sobreestima y_i

1.3.2 Propiedades algebraicas

1. $\sum_{i=1}^{n} \widehat{u}_i = 0$
   - Viene de la primera condición
   - $\widehat{\beta}_0$ y $\widehat{\beta}_1$ son escogidos tal que la suma (y promedio) de los residuales sea cero
2. $\sum_{i=1}^{n} x_i \widehat{u}_i = 0$
   - Viene de la segunda condición
   - Equivale a $\operatorname{Cov}(x_i, \widehat{u}_i) = 0$ porque el promedio muestral de $\widehat{u}_i$ es cero
3. $\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}$
   - Viene de la primera condición
   - El punto $(\bar{x}, \bar{y})$ siempre está sobre la línea de regresión de MCO

Para una interpretación alternativa, observa que MCO descompone y_i en 2 partes:

$$y_i = \widehat{y}_i + \widehat{u}_i$$

- $\bar{\widehat{y}} = \bar{y}$ porque $\bar{\widehat{u}} = 0$
- $\operatorname{Cov}(\widehat{y}_i, \widehat{u}_i) = 0$
  - Los valores ajustados y los residuales no están correlacionados

Podemos definir la variación de cada una de las partes:

- Suma de cuadrados total: $\text{SCT} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
  - Mide la variación total de (o qué tan dispersa está) y_i en la muestra
- Suma de cuadrados explicada: $\text{SCE} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$
  - Mide la variación de $\widehat{y}_i$ en la muestra
- Suma de cuadrados de los residuales: $\text{SCR} = \sum_{i=1}^{n} \widehat{u}_i^2$
  - Mide la variación de $\widehat{u}_i$ en la muestra

La variación total en y siempre se puede expresar como la suma de la variación explicada y la no explicada (porque $\operatorname{Cov}(\widehat{y}_i, \widehat{u}_i) = 0$):

$$\text{SCT} = \text{SCE} + \text{SCR}$$

1.3.3 Bondad de ajuste

Número que resume qué tan bien se ajusta la línea de regresión de MCO a los datos
Si SCT $\neq 0$ $(y_i \neq c \ \forall i)$,

$$1 = \frac{\text{SCE}}{\text{SCT}} + \frac{\text{SCR}}{\text{SCT}} \implies \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} = \text{R}^2$$

- R^2 se conoce como el **coeficiente de determinación**
- $0 \leq \text{R}^2 \leq 1$ porque SCE $\not> $ SCT
  - R^2 es la fracción de la variación de y en la muestra explicada por x
- Interpretación en porcentaje $(100 \times \text{R}^2)$
  - Porcentaje de variación de y en la muestra explicada por x
- $\text{R}^2 \approx 0$ significa que la línea de regresión de MCO ajusta mal
  - Aún así puede llegar a ser útil

Pregunta. ¿Qué significaría una $\text{R}^2 = 1$?

1.4 Unidades de medición y forma funcional

¿Qué le pasa a los estimados de MCO cuando manipulamos los datos?

1.4.1 Efectos del cambio en unidades de medición

¿Qué pasa si cambiamos las unidades de medición de las variables en el modelo RLS?
Podemos multiplicar por una constante c a la variable dependiente o a la independiente

- Si $\widehat{y}_1 = \widehat{\beta}_0 + \widehat{\beta}_1 x$ y $y_2 = cy_1$, entonces
$$\widehat{y}_2 = c\widehat{\beta}_0 + c\widehat{\beta}_1 x$$
  - Afecta a ambos estimados
- Si $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1$ y $x_2 = \frac{x_1}{c}$, entonces
$$\widehat{y} = \widehat{\beta}_0 + c\widehat{\beta}_1 x_2$$
  - No afecta al intercepto

R^2 es invariante a cambios en unidades de x o y

- R^2 no depende de las unidades de medición

Preguntas. Una regresión del salario expresado en miles de dólares (*salariomil*) sobre el rendimiento del capital expresado en porcentaje (*roepct*) arroja lo siguiente:

$$\widehat{salariomil} = 963.191 + 18.501 roepct$$

- ¿Cómo cambian los estimados si expresamos el salario en dólares (*salariodol*) (y el rendimiento del capital se queda en porcentaje, *roepct*)?
- ¿Cómo cambian los estimados si expresamos el rendimiento del capital en proporción (*roedec*) (y el salario se queda en miles de dólares, *salariomil*)?

1.4.2 Efectos no lineales en RLS

Hasta ahora solo hemos visto la relación lineal entre x y y

- Es fácil capturar efectos no lineales si definimos apropiadamente x y y

Podemos usar el logaritmo natural para capturar:

- Crecimiento porcentual constante (rendimientos crecientes)
- Elasticidad constante

Modelo nivel-nivel

$$y = \beta_0 + \beta_1 x + u \implies \Delta y = \beta_1 \Delta x$$

- Mismo efecto en y sin importar el valor inicial de x

Modelo log-nivel

$$\log(y) = \beta_0 + \beta_1 x + u \implies \Delta \log(y) = \beta_1 \Delta x \implies \% \Delta y \approx (100 \times \beta_1)\, \Delta x$$

- Equivalente a $y = \exp(\beta_0 + \beta_1 x + u)$

[Gráfica]

- Captura un crecimiento porcentual constante
- Cambio en la variable dependiente crece por cada cambio *unitario* en x
- $100 \times \beta_1$ se conoce como la **semi-elasticidad** de y respecto a x

Pregunta. Una regresión del logaritmo del salario en dólares por hora sobre los años de educación arroja lo siguiente:

$$\widehat{\log(salario)} = 0.584 + 0.083 educ$$

- ¿Cómo cambia el *salario* por cada año adicional de educación?

Modelo log-log o de elasticidad constante

$$\log(y) = \beta_0 + \beta_1 \log(x) + u \implies \Delta \log(y) = \beta_1 \Delta \log(x) \implies \% \Delta y \approx \beta_1 \% \Delta x$$

- β_1 se conoce como la **elasticidad** de y con respecto a x

Pregunta. Una regresión del logaritmo del salario sobre el logaritmo de los ventas arroja lo siguiente:

$$\widehat{\log(salario)} = 4.822 + 0.257 \log(ventas)$$

- ¿Cómo cambia el *salario* con las *ventas*?

Modelo nivel-log (poco usado)

$$y = \beta_0 + \beta_1 \log(x) + u \implies \Delta y = \beta_1 \Delta \log(x) = \frac{\beta_1}{100} 100 \Delta \log(x) \implies \Delta y \approx \frac{\beta_1}{100} \% \Delta x$$

- $\frac{\beta_1}{100}$ es el cambio unitario en y cuando x aumenta 1%

Preguntas. Una regresión de las horas trabajadas a la semana sobre el logaritmo del salario por hora arroja lo siguiente:

$$\widehat{horas} = 33 + 45.1 \log(salario)$$

- ¿Cómo cambian las horas trabajadas con un aumento de 1% en el salario?
- ¿Y si el salario aumenta en 10%?

1.4.3 Significado de regresión lineal

El modelo $y = \beta_0 + \beta_1 x + u$ permite capturar relaciones no lineales entre x y y

- El modelo es lineal en los parámetros β_0 y β_1
- No hay restricciones en cómo se definen x y y
  - Importante: Definiciones de x y y sí afectan la interpretación
- Ej. Modelo de regresión no lineal

$$y = \frac{1}{\beta_0 + \beta_1 x} + u$$

1.5 Valores esperados y varianzas de los estimadores de MCO

Interesados en propiedades de los estimadores $\widehat{\beta}_0$ y $\widehat{\beta}_1$ para los parámetros β_0 y β_1

- Propiedades de las distribuciones muestrales (valor esperado, varianza)
- Propiedades estadísticas (falta sesgo, consistencia)

Para estudiar esas propiedades, hacemos supuestos sobre la población (modelo RLS)

- Supuestos Gauss-Markov (G-M)
  - Supuestos ideales para obtener propiedades deseables
  - Después desviaciones de esos supuestos
  - Ej. Línea de producción de una fábrica

1.5.1 Falta de sesgo de MCO

¿Cuál es el centro de la distribución de $\widehat{\beta}_1$?

Supuesto RLS.1. El modelo es lineal en parámetros

Este supuesto define el modelo poblacional

$$y = \beta_0 + \beta_1 x + u$$

- β_0 y β_1 son los parámetros poblacionales (desconocidos)
- x, y, u son variables aleatorias
- Permite capturar relaciones no lineales

Supuesto RLS.2. La muestra es aleatoria

Obtenemos una muestra aleatoria de tamaño n, $\{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$, del modelo poblacional en RLS.1

- Usamos los datos para estimar los parámetros

El modelo poblacional en términos de la muestra aleatoria es:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad \forall i$$

- u_i son los factores no observados que afectan a y_i
- $u_i \neq \widehat{u}_i$

[Gráfica]

Supuesto RLS.3. Hay variación en la variable independiente en la muestra

Los valores de x en la muestra, $\{x_i \mid i = 1, 2, \ldots, n\}$, no son todos iguales

- Se puede verificar revisando que desvest$(x_i) \neq 0$
- Recordatorio: Los estimados de MCO para el intercepto y la pendiente están definidos solo si hay variación en la variable independiente

Supuesto RLS.4. Media condicional cero

$$\mathbb{E}\left(u|x\right) = 0$$

Todos los otros factores que afectan a y no se correlacionan con x

- Para una muestra aleatoria, RLS.4 implica que $\mathbb{E}\left(u_i|x_i\right) = 0 \quad \forall i$
- Permite obtener estimadores de β_0 y β_1 insesgados

Bajo supuestos RLS.1-RLS.4, $\widehat{\beta}_0$ y $\widehat{\beta}_1$ son insesgados

- Primero definimos: $\text{SCT}_x = \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Reescribimos $\widehat{\beta}_1$ (variable aleatoria)

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\text{SCT}_x} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\text{SCT}_x}$$

$$\widehat{\beta}_1 = \frac{\beta_1 \text{SCT}_x + \sum_{i=1}^{n} (x_i - \bar{x}) u_i}{\text{SCT}_x} \implies \mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1 + \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \mathbb{E}(u_i)}{\text{SCT}_x} = \beta_1$$

- Para $\widehat{\beta}_0$, usamos $\bar{y} = \beta_0 + \beta_1 \bar{x}$

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} + \bar{u} - \widehat{\beta}_1 \bar{x} = \beta_0 + \left(\beta_1 - \widehat{\beta}_1\right) \bar{x} + \bar{u}$$

$$\mathbb{E}\left(\widehat{\beta}_0\right) = \beta_0 + \mathbb{E}\left(\beta_1 - \widehat{\beta}_1\right) \bar{x} + \mathbb{E}(\bar{u}) = \beta_0$$

Comentarios sobre la falta de sesgo:

- Falta de sesgo depende de RLS.1-RLS.4
  - Ej. Si RLS.4 no se cumple, $\widehat{\beta}_0$ y $\widehat{\beta}_1$ son sesgados
- En RLS siempre preocupa que $\text{Corr}(x, u) \neq 0$
  - Puede resultar en correlación espuria entre x y y
  - Ej. Variable omitida

1.5.2 Varianza de los estimadores de MCO

La distribución de $\widehat{\beta}_1$ está centrada en β_1, ¿qué tan dispersa es?

> Supuesto RLS.5. Homocedasticidad
>
> $$\text{Var}(u \mid x) = \sigma^2$$

La varianza del error no observado condicional en x es constante (**homocedasticidad**)

- Es diferente del valor esperado de u dado x: $\mathbb{E}(u|x) = 0$
- No se utiliza para demostrar falta de sesgo
- De la definición de la varianza,

$$\text{Var}(u \mid x) = \mathbb{E}\left(u^2 \mid x\right) - \left[\mathbb{E}(u \mid x)\right]^2 = \mathbb{E}\left(u^2 \mid x\right) = \sigma^2$$

- No depende de x, entonces $\sigma^2 = \mathbb{E}(u^2) = \text{Var}(u)$
- σ^2 es la varianza incondicional de u y se llama **varianza del error**
- σ es la desviación estándar del error
- Si $\text{Var}(u \mid x) = g(x)$, el término de error exhibe **heterocedasticidad**
  - Varianza no constante

Podemos expresar RLS.4 y RLS.5 en términos de la media y varianza condicionales de y

- RLS.4: $\mathbb{E}(y \mid x) = \beta_0 + \beta_1 x$, es lineal en x
- RLS.5: $\text{Var}(y \mid x) = \sigma^2$, es constante

[Gráfica]

Las **varianzas muestrales** de los estimadores de MCO son

$$\text{Var}\left(\widehat{\beta}_1\right) = \frac{\sigma^2}{\text{SCT}_x} \quad \text{y} \quad \text{Var}\left(\widehat{\beta}_0\right) = \frac{\sigma^2}{n}\frac{\sum_{i=1}^n x_i^2}{\text{SCT}_x}$$

- Válidas solo bajo homocedasticidad
- Más variación en lo no observado (u), hace más difícil estimar β_1
- Se prefiere más varianza en x_i porque es más fácil estimar β_1, así como identificar la relación entre $\mathbb{E}(y \mid x)$ y x
- Una muestra más grande debería aumentar SCT_x
  - Menor varianza para $\widehat{\beta}_1$

¿Por qué no podemos calcular las varianzas muestrales?

1.5.3 Estimación de la varianza del error

Las fórmulas de las varianzas dependen de σ^2, generalmente se desconoce
Podemos estimar σ^2 para, a su vez, estimar $\text{Var}\left(\widehat{\beta}_0\right)$ y $\text{Var}\left(\widehat{\beta}_1\right)$
Importante distinguir entre:

- Modelo poblacional: $y_i = \beta_0 + \beta_1 x_i + u_i$
  - Usa parámetros (β_0 y β_1) y los errores u_i nunca se observan
- Descomposición entre valores ajustados y residuales: $y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \widehat{u}_i$
  - Usa estimados ($\widehat{\beta}_0$ y $\widehat{\beta}_0$) y los residuales $\widehat{u}_i$ se obtienen de los datos

Podemos reescribir los residuales como función de los errores:

$$\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i = (\beta_0 + \beta_1 x_i + u_i) - \widehat{\beta}_0 - \widehat{\beta}_1 x_i = u_i - \left(\widehat{\beta}_0 - \beta_0\right) - \left(\widehat{\beta}_1 - \beta_1\right) x_i$$

- Entonces $\widehat{u}_i \neq u_i$, aunque $\mathbb{E}(u_i - \widehat{u}_i) = 0$

Sabemos que $\sigma^2 = \mathbb{E}(u^2)$, entonces un 'estimador' insesgado sería $\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n u_i^2$

- ¿Cuál es el problema con ese estimador?

Tenemos estimados de u_i ($\widehat{u}_i$), entonces podríamos usar $\check{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 = \frac{\text{SCR}}{n}$

- ¿Problema? Es sesgado
- No satisfice las restricciones de los residuales de MCO: $\sum_{i=1}^n \widehat{u}_i = 0$ y $\sum_{i=1}^n x_i \widehat{u}_i = 0$
- ¿Cuántos grados de libertad hay?

Un estimador insesgado de σ^2 es

$$\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{\text{SCR}}{n-2}$$

- $\mathbb{E}\left(\widehat{\sigma}^2\right) = \sigma^2$
- Si sustituimos $\widehat{\sigma}^2$ arriba, tenemos estimadores insesgados para $\text{Var}\left(\widehat{\beta}_0\right)$ y $\text{Var}\left(\widehat{\beta}_1\right)$

$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$ se llama el **error estándar de la regresión**

- $\mathbb{E}\left(\widehat{\sigma}\right) \neq \sigma$ pero es consistente
- Es un estimado de la desviación estándar de los factores no observados que afectan a y
- Propósito principal: Usar $\widehat{\sigma}$ para estimar las desviaciones estándar de $\widehat{\beta}_0$ y $\widehat{\beta}_1$

Dado que $\text{desvest}\left(\widehat{\beta}_1\right) = \frac{\sigma}{\sqrt{\text{SCT}_x}}$, su estimador es

$$\text{errest}\left(\widehat{\beta}_1\right) = \frac{\widehat{\sigma}}{\sqrt{\text{SCT}_x}}$$

- Se conoce como el **error estándar de $\widehat{\beta}_1$**
- Representa tanto a una variable aleatoria y como al estimado
- Nos da una idea de la precisión del estimador
- Sirven para construir estadísticos de prueba e intervalos de confianza

1.6 Regresión por el origen y sobre una constante

Si imponemos la restricción $\beta_0 = 0$, escogemos $\widetilde{\beta}_1$ tal que $\widetilde{y} = \widetilde{\beta}_1 x$

$$\widetilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

- Igual que $\widehat{\beta}_1$ pero con $\bar{x} = 0$
- Al menos un $x_i \neq 0$, ¿por qué?
- La línea pasa por el origen ($\widetilde{y} = 0$ cuando $x = 0$)
- Si $\beta_0 \neq 0$, $\widetilde{\beta}_1$ estará sesgado $[\mathbb{E}\left(\widetilde{\beta}_1\right) \neq \beta_1]$
- $\text{R}^2 < 0$ es posible ($\bar{y}$ ajusta mejor a y que $\widetilde{y}$)

Si imponemos la restricción $\beta_1 = 0$, $\widetilde{\beta}_0 = \bar{y}$

- Produce la menor suma de desviaciones cuadradas
- Si $\beta_1 \neq 0$, $\widetilde{\beta}_0$ estará sesgado $[\mathbb{E}\left(\widetilde{\beta}_0\right) \neq \beta_0]$

1.7 Resumen del modelo RLS

- Dada una muestra aleatoria, el método de MCO se usa para estimar los parámetros del intercepto y la pendiente del modelo poblacional
- Algebra de MCO: Podemos calcular valores ajustados, residuales y cambios predichos en variable dependiente para un cambio dado en la variable independiente
- Cuestiones prácticas:
  - Comportamiento de MCO cuando cambian las unidades de medición
  - Uso del logaritmo natural para modelar elasticidad constante y semi-elasticidad
- Bajo supuestos RLS.1 a RLS.4, los estimadores de MCO son insesgados
  - RLS.4 no se cumple si valores omitidos en u se correlacionan con x
- Si suponemos homocedasticidad, podemos estimar las varianzas muestrales de MCO
- Regresión por el origen
- ¿Qué falta?
  - Eficiencia de MCO, hacer pruebas de hipótesis sobre los parámetros poblacionales, intervalos de confianza
  - RLS es un caso particular de RLM