

Actividad 1. Regresión Simple

Econometría LECO

Dr. Francisco Cabrera

Entregable: Un archivo de Word o PDF (e.g. LaTeX, R-Markdown) contestando correctamente a las preguntas abajo presentadas y gravado con su nombre, así como el archivo “Script” de R.

1. Derive OLS utilizando las condiciones de primer orden que surgen de $\sum(y_i - \bar{y})^2 = 0$

2. Genere sus propios datos para probar la teoría:

****Instrucciones en R-Studio:** Vaya a la pestaña “File” en R. Click en “New Project”. Click en “New Directory”. Click nuevamente en “New Project”. Nombre el proyecto como “Actividad 1 LECO”. Click “Browse” e indique el directorio preferido. Click “Create Project” (esto crea un nuevo folder dónde trabajar también).

Asuma que la ecuación poblacional está dada por $y_i = 15 + 7X_i + \mu_i$, con el término estocástico (μ) distribuido normalmente con media de cero. X toma valores aleatorios. *Recuerde que en la práctica no observamos esta ecuación, pero podemos estimarla con una muestra de datos para X e Y .*

- a. Genere una muestra aleatoria de 100 observaciones y asuma una desviación estándar de 10, conforme al script siguiente:

```
help(rnorm)
help(runif)

#creamos variables aleatorias usando las funciones "rnorm" y "runif"
set.seed(123) #esto sirve para que siempre cree los mismos números aleatorios
x <- rnorm(100)
error <- rnorm(100, mean=0, sd=10)
y <- 15 + (7*x) + error
```

- b. Muestre un scatterplot de x e y , y describa brevemente la relación entre estas dos variables.
c. Estime la regresión conforme al script abajo y escriba la ecuación estimada.

```
reg <- lm(y ~ x)
summary(reg)
anova(reg)

#find sse
sse <- sum((fitted(reg) - y)^2)
sse
```

- d. interprete el coeficiente de regresión

- e. ¿cuál es el valor predicho de Y cuando $X = \bar{x}$?
- f. calcule manualmente R^2
- g. ejecute:

```
summary(fitted(reg))
summary(y)
```

3. Creando sus propios datos para indagar R^2 .

- a. Ejecute el siguiente código, muestre los resultados e interprete R^2 .

```
set.seed(123) #esto sirve para que siempre cree los mismos números aleatorios
x <- runif(100)
error <- rnorm(100, mean=0, sd=5)
y <- 2 + (5*x) + error
hist(y)

#regresión:
regression <- lm(y ~ x)
summary(regression)

#scatterplot (gráfico) de nuestra regresión.
plot(x, y, main = "Main title",
     xlab = "X axis title", ylab = "Y axis title",
     pch = 19, frame = FALSE) +
  abline(lm(y ~ x, data = mtcars), col = "blue")
```

- b. En el código anterior, modifique: $SD = 15$ e interprete R^2 .
- c. Muestre que R^2 es menor con $SD = 15$ en el error (hint: use SST, SSR) ¿Esto quiere decir que es “peor” la segunda regresión para calcular $\hat{\beta}$ en términos de sesgo y precisión? ¿Por qué sí o por qué no?

4. Regresión lineal con datos de Wooldridge.

- a. Ejecute los siguientes comandos en R-Studio y muestre los resultados:

```
library(wooldridge)
data("wage1")
?wage1 #esto explica la base de datos.

wagereg <- lm(lwage ~ educ, data = wage1)
summary(wagereg)
```

- a. Interprete los coeficientes β_0 y β_1 ¿Es el signo de β_1 el esperado? ¿por qué?
- b. Obtenga R^2 mediante el cálculo de $SST/SSE/SSR$.
- c. ¿Bajo qué supuestos, de los estudiados en clase hasta ahora, podemos interpretar β_1 como el efecto de la educación en ingresos? Describa cada uno formalmente.
- d. ¿Qué supuesto estándar estamos violando y por qué?
- e. Demuestre formalmente el sesgo del estimador β_1 bajo violación de ese supuesto.

5. Utilizando datos de 2018 sobre ventas de casas en México, la siguiente ecuación es estimada, referente al precio de la casa y en relación con la distancia, en kilómetros, a un basurero municipal recientemente instalado:

$$\log(\text{Price}) = 9.40 + 0.312(\text{dist})$$

$$n = 135$$

$$R^2 = 0.162$$

- a. Interprete el coeficiente de dist .
- b. ¿Cree que esta estimación nos ofrece un efecto insesgado la elasticidad “*ceteris paribus*” del precio respecto a la distancia? ¿Intuitivamente explique por qué? (piense en la decisión del municipio de colocar el basurero)
- c. ¿Qué otros factores sobre la casa pueden afectar su precio dado $\text{Cov}(u, \text{dist}) \neq 0$?