

Regresión múltiple y otras técnicas multivariadas

Tarea 02

Rivera Torres Francisco de Jesús

Rodríguez Maya Jorge Daniel

Samayoa Donado Víctor Augusto

Trujillo Barrios Georgina

Febrero 20, 2019

Ejercicio 1

Mostrar que los EMCO de β_0 y β_1 son lineales e insesgados.

Ejercicio 2

Mostrar que $\hat{\mu}_0 = \hat{\beta}_1 + \hat{\beta}_1 x_0$ es un estimador insesgado de $\mu_0 = \beta_1 + \beta_1 x_0$ y que

$$V(\hat{\mu}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right) \quad (1)$$

Demostración. Procedemos a calcular el valor esperado de $\hat{\mu}_0$.

$$E(\hat{\mu}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 x_0)$$

aplicando la linealidad de la esperanza y considerando que

X es una variable NO aleatoria, se tiene

$$= E(\hat{\beta}_0) + x_0 E(\hat{\beta}_1)$$

pero por el ejercicio 1, sabemos que $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados

$$= \beta_0 + \beta_1 x_0 = \mu_0$$

por lo tanto se tiene que $E(\hat{\mu}_0) = \mu_0$, lo cual significa que $\hat{\mu}_0$ es un estimador insesgado de μ_0 . \square

Ejercicio 3

Suponer que se ajusta un modelo RLS por MCO a las observaciones $(x_i, y_i), i = 1, \dots, n$. Verificar que se cumplen las siguientes igualdades:

Para los siguientes incisos, será necesario tener las siguientes identidades:

$$\begin{array}{lll}
e_i \stackrel{(i)}{=} y_i - \hat{y}_i & \hat{y}_i \stackrel{(ii)}{=} \hat{\beta}_0 + \hat{\beta}_1 x_i & e_i \stackrel{(iii)}{=} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\
\hat{\beta}_0 \stackrel{(iv)}{=} \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x} & \hat{\beta}_1 \stackrel{(v)}{=} \frac{S_{xy}}{S_{xx}} & \hat{\beta}_0 \stackrel{(vi)}{=} \bar{y} - \hat{\beta}_1 \bar{x}
\end{array}$$

Inciso (a)

$$\sum_{i=1}^n e_i = 0 \quad (2)$$

Demostración.

$$\begin{aligned}
\sum_{i=1}^n e_i &\stackrel{(iii)}{=} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i = n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1 \bar{x} \\
&= n(\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}) = 0 \quad \text{usando la identidad (vi)}
\end{aligned}$$

□

Inciso (b)

$$\sum_{i=1}^n x_i e_i = 0 \quad (3)$$

Demostración.

$$\sum_{i=1}^n x_i e_i \stackrel{(iii)}{=} \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i^2 = \sum_{i=1}^n x_i y_i - n\hat{\beta}_0 \bar{x} - \hat{\beta}_1 \sum_{i=1}^n x_i^2$$

usando la identidad (vi) se tiene

$$\begin{aligned}
&= \sum_{i=1}^n x_i y_i - n(\bar{y} - \hat{\beta}_1 \bar{x}) \bar{x} - \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} + n\hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \\
&= \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) - \hat{\beta}_1 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)
\end{aligned}$$

usando los incisos 6.(a), 6.(b) de la tarea 1 y la identidad (v), se tiene

$$= S_{xy} - \frac{S_{xy}}{S_{xx}} S_{xx} = 0$$

□

Inciso (c)

$$\sum_{i=1}^n \hat{y}_i e_i = 0 \quad (4)$$

Demostración.

$$\begin{aligned} \sum_{i=1}^n \hat{y}_i e_i &\stackrel{(ii),(iii)}{=} \sum_{i=1}^n \left[(\hat{\beta}_0 + \hat{\beta}_1 x_i) (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \right] \\ &= \sum_{i=1}^n \hat{\beta}_0 (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) + \sum_{i=1}^n \left[\hat{\beta}_1 x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \right] \\ &= \hat{\beta}_0 \left(\sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i \right) + \hat{\beta}_1 \left(\sum_{i=1}^n x_i y_i - \sum_{i=1}^n \hat{\beta}_0 x_i - \sum_{i=1}^n \hat{\beta}_1 x_i^2 \right) \\ &= \hat{\beta}_0 (n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \left(\sum_{i=1}^n x_i y_i - n\hat{\beta}_0 \bar{x} - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \right) \\ &= n\hat{\beta}_0 (\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \left(\sum_{i=1}^n x_i y_i - n\hat{\beta}_0 \bar{x} - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \right) \end{aligned}$$

por la identidad (vi) se cancela el primer término

y sustituimos el valor de $\hat{\beta}_0$ en el segundo término

$$\begin{aligned} &= n\hat{\beta}_0 \left(\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} \right) + \hat{\beta}_1 \left(\sum_{i=1}^n x_i y_i - n(\bar{y} - \hat{\beta}_1 \bar{x}) \bar{x} - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \right) \\ &= \hat{\beta}_1 \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} + n\hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \right) \\ &= \hat{\beta}_1 \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) - \hat{\beta}_1^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \end{aligned}$$

usando los incisos 6.(a), 6.(b) de la tarea 1 y la identidad (v), se tiene

$$= \frac{S_{xy}}{S_{xx}} (S_{xy}) - \frac{S_{xy}^2}{S_{xx}^2} (S_{xx}) = 0$$

□

Inciso (d)

$$\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i \quad (5)$$

Demostración.

$$0 \stackrel{(a)}{=} \sum_{i=1}^n e_i \stackrel{(i)}{=} \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{y}_i, \text{ lo anterior implica que } \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i$$

□

Inciso (e)

$$\sum_{i=1}^n y_i e_i = \sum_{i=1}^n e_i^2 \quad (6)$$

Demostración.

$$\sum_{i=1}^n y_i e_i \stackrel{(i)}{=} \sum_{i=1}^n (e_i + \hat{y}_i) e_i = \sum_{i=1}^n e_i^2 + \sum_{i=1}^n \hat{y}_i e_i \stackrel{(c)}{=} \sum_{i=1}^n e_i^2 + \sum_{i \neq 1}^n \hat{y}_i e_i \overset{0}{=} \sum_{i=1}^n e_i^2$$

□

Inciso (f)

$$\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 = \hat{\beta}_1^2 S_{xx} \quad (7)$$

Demostración.

□

Ejercicio 4

Calcular el error cuadrático medio de $\hat{\sigma}_{\text{MCO}}^2$ y de $\hat{\sigma}_{\text{MV}}^2$. A partir de los resultados decidir que estimador de σ^2 es mejor.

Ejercicio 5

El conjunto de datos `mtcars` del paquete `datasets` de R contiene información sobre el rendimiento y otras características de 32 vehículos. El rendimiento se encuentra en la variable `mpg` y está medido en millas por galón y el peso del vehículo está en la variable `wt` que está medida en miles de libras.

Inciso (a)

Ajustar un modelo RLS para explicar `mpg` en términos de `wt`. Reportar las estimaciones de β_0 y β_1 .

```
# reutilizando la función definida en la tarea 1
rls <- function(x, y) {
  x.bar <- mean(x, na.rm = TRUE)
  y.bar <- mean(y, na.rm = TRUE)
  S.xx <- sum((x - x.bar)^2, na.rm = TRUE)
  S.xy <- sum((x - x.bar) * (y - y.bar), na.rm = TRUE)

  b0.hat <- y.bar - x.bar * S.xy / S.xx
```

```

b1.hat <- S.xy / S.xx

y.adj <- b0.hat + b1.hat * x
residuos <- y - y.adj
sigma2.hat <- sum(residuos^2) / (length(residuos) - 2)

return(list(b0.hat = b0.hat, b1.hat = b1.hat, sigma2.hat = sigma2.hat))
}

rls(mtcars$wt, mtcars$mpg)

## $b0.hat
## [1] 37.28513
##
## $b1.hat
## [1] -5.344472
##
## $sigma2.hat
## [1] 9.277398

```

Los valores estimados de β_0 y β_1 son los siguientes:

```

paste(rls(mtcars$wt, mtcars$mpg)$b0.hat, " y ", rls(mtcars$wt, mtcars$mpg)$b1.hat)

## [1] "37.285126167342 y -5.34447157272268"

```

Inciso (b)

Transformar la variable `wt` a toneladas y repetir el inciso anterior. ¿Cómo se relacionan estas estimaciones de β_0 y β_1 con las anteriores?

Según la información de la base, la unidad de `wt` es 1000 libras. Por lo que para transformarla a toneladas basta multiplicarla por 0.453592

```
mtcars$wt_ton <- 0.453592 * mtcars$wt
```

Ajustando una regresión lineal, los valores estimados de β_0 y β_1 son los siguientes:

```

paste(rls(mtcars$wt_ton, mtcars$mpg)$b0.hat, " y ", rls(mtcars$wt_ton, mtcars$mpg)$b1.hat)

## [1] "37.285126167342 y -11.7825525422024"

```

En comparación con el modelo 1, notamos que este modelo tiene el mismo intercepto y su valor estimado para β_1 es 0.453592 veces el valor de la β_1 del primer modelo.

Inciso (c)

Repetir el inciso anterior pero ahora con `mpg` transformada a kilómetros por litro y `wt` en las unidades originales (miles de libras).

Según la información de la base, *mpg* representa las millas por galón. Para transformar el rendimiento a kilómetros por litro, usaremos la siguiente información: una milla es 1.609344 kilómetros y un galón es 4.54609 litros. Por lo que basta multiplicarlo por $0.35400618 = 1.609344/4.54609$.

```
mtcars$kp1 <- 0.35400618 * mtcars$mpg
```

Ajustando una regresión lineal para el rendimiento medido por kilómetros por litros y peso en miles de libras, los valores estimados de β_0 y β_1 son los siguientes:

```
paste(rls(mtcars$wt, mtcars$kp1)$b0.hat, " y ", rls(mtcars$wt, mtcars$kp1)$b1.hat)

## [1] "13.1991650853188 y -1.89197596557815"
```

En comparación con el modelo 1, notamos que este modelo tiene valores distintos tanto para β_0 como para β_1 . En ambos casos estas nuevas estimaciones corresponden a la 0.3500618 parte de las primeras estimaciones.

Inciso (d)

Concluir sobre el efecto de los cambios de escala en las estimaciones de los parámetros del modelo RLS.

Si en un modelo RLS la variable independiente cambia de escala, la nueva estimación del intercepto no cambiará mientras que la de β_1 sí se modifica y lo hace en la misma razón en la que cambia la escala. En el caso del cambio de escala de la variable dependiente ambas estimaciones se modifican (su nuevo valor es exactamente al viejo valor multiplicado por el cambio de escala). En el primer caso (cambio de escala de variable independiente), cuando el valor de la variable independiente es cero (y tiene sentido interpretarla), la interpretación para la variable dependiente es la misma que sin cambio de escala; no obstante, la interpretación del cambio ante una variación en la variable independiente se modifica en razón del cambio de escala (pero en realidad podría decirse que es la misma). En el caso en que se modifica la escala de la variable dependiente, ambas estimaciones se modifican pero nuevamente este cambio sólo refleja el cambio de escala en los valores estimados.

Ejercicio 6

Con el conjunto de datos *mtcars* ajustar un modelo RLS para explicar *mpg* en términos de *wt*

```
rls(mtcars$wt, mtcars$mpg)
```

```
## $b0.hat
## [1] 37.28513
##
## $b1.hat
## [1] -5.344472
##
## $sigma2.hat
## [1] 9.277398
```

Inciso (a)

Reportar las estimaciones de $V(\hat{\beta}_0)$ y $V(\hat{\beta}_1)$.

Beta 1

Primero veamos que ocurre $V(\hat{\beta}_1)$.

$$V(\hat{\beta}_1) = V\left(\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) = \frac{1}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \sum_{i=1}^n (x_i - \bar{x})^2 V(y_i) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

La primera igualdad está dada por la definición de β_1 como la razón entre S_{xy} y S_{xx} . La segunda por las propiedades de la varianza y porque $V(\sum_{i=1}^n (y_i))$ puede verse de la siguiente manera:

$$V\left(\sum_{i=1}^n y_i\right) = V\left(\sum_{i=1}^n \beta_0 + \beta_1 X_i + \epsilon_i\right) = V\left(\sum_{i=1}^n \epsilon_i\right) = \sum_{i=1}^n \sum_{j=1}^n Cov(\epsilon_i, \epsilon_j) = \sum_{i=1}^n Cov(\epsilon_i, \epsilon_i) \sum_{i=1}^n V(\epsilon_i) \quad (9)$$

$$= \sum_{i=1}^n V(\beta_0 + \beta_1 X_i + \epsilon_i) = \sum_{i=1}^n V(y_i) \quad (10)$$

$$(11)$$

En este caso usamos la forma funcional que define a nuestro modelo, las propiedades de la varianza sobre constantes y el supuesto de que los errores son independientes (es decir, que $Cov(\epsilon_i, \epsilon_j) = 0$)

Beta 0

En el caso de β_0 , podemos escribirlo en términos de β_1 según la ecuación 8. De tal manera, ocurre lo siguiente:

$$V(\hat{\beta}_0) = V(\bar{y} - \hat{\beta}_1 \bar{x}) = V(\hat{\beta}_0) = V(\bar{y}) + (\bar{x}^2)V(\bar{\beta}_1) - 2(\bar{x})Cov(\bar{Y}, \hat{\beta}_0)$$

El primer término de la ecuación 12 a su vez puede desarrollarse como sigue:

$$V(\bar{y}) = V\left(\frac{1}{n} \sum_{i=1}^n y_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(y_i) = \frac{\sigma^2}{n}$$

El segundo término de 12 ya se desarrolló anteriormente en 8. Finalmente, la covarianza entre \bar{y} y $\hat{\beta}_1$ puede verse como sigue:

$$Cov(\bar{Y}, \hat{\beta}_1) = Cov\left\{\frac{1}{n} \sum_{i=1}^n Y_i, \frac{\sum_{j=1}^n (x_j - \bar{x}) Y_j}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\} = \frac{1}{n} \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} Cov\left\{\sum_{i=1}^n Y_i, \sum_{j=1}^n (x_j - \bar{x}) Y_j\right\} \quad (12)$$

$$= \frac{1}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x}) \sum_{j=1}^n Cov(Y_i, Y_j) \quad (13)$$

$$= \frac{1}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x}) \sigma^2 = 0 \quad (14)$$

Esto último ocurre pues $\sum_{i=1}^n (x_i - \bar{x}) = 0$

Así

$$V(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (15)$$

$$= \frac{\sigma^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 + n \bar{x}^2 \right\} \quad (16)$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \quad (17)$$

Inciso (b)

Reportar la estimación de la medida del rendimiento de un vehículo con peso de 4,300 libras.

Para un coche con 4.3 miles de libras, la medida de rendimiento estimada está dada por $37.285126167342 - 5.34447157272268 * 4.3 = 14.3039 \frac{\text{miles}}{\text{gallon}}$.

Inciso (c)

Reportar la estimación de la varianza de la estimación del inciso anterior.