Modelos Estadísticos I **(Fecha: 03/03/2021)**

Tarea #2

Estudiante: Roberto Vásquez Martínez *NUA:* 424662

Problema 1

Sea

$$\begin{aligned} Y_1 =& \theta + \epsilon_1 \\ Y_2 =& 2\theta - \phi + \varepsilon_2 \\ Y_3 =& \theta + 2\phi + \varepsilon_3, \end{aligned}$$

donde $\mathbb{E}[\epsilon_i] = 0$. Encuentre las estimacions mínimos cuadráticas de θ y ϕ.

(Solución)

Sea

$$X = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix},$$

$Y = [Y_1, Y_2, Y_3]^T$, $\epsilon = [\epsilon_1, \epsilon_2, \epsilon_3]^T$, consideremos $\beta = [\theta, \pi]$, en notación matricial el problema se reduce a

$$Y = X\beta + \epsilon,$$

por lo que la estimación de mínimos cuadrados $\hat{\beta}$ de β es

$$\hat{\beta} = (X^T X)^{-1} X^T Y,$$

y haciendo las cuentas con el código de Python tenemos que

$$\hat{\beta} = \begin{bmatrix} \hat{\theta} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} \frac{1}{6}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \\ -\frac{1}{5}Y_2 + \frac{2}{5}Y_3, \end{bmatrix}$$

de lo que concluimos que las estimaciones de mínimos cuadrados de θ y ϕ son

$$\begin{aligned} \hat{\theta} =& \frac{1}{6}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \\ \hat{\phi} =& -\frac{1}{5}Y_2 + \frac{2}{5}Y_3, \end{aligned}$$

que es lo que estábamos buscando.

$\square$

Problema 2

Considere el modelo de regresión

$$\mathbb{E}[Y_i] = \beta_0 + \beta_1 x_i + \beta_2(3x_i^2 - 2), \quad (i = 1, 2, 3),$$

donde $x_1 = -1, x_2 = 0, x_3 = 1$. Encuentre las estimaciones de mínimos cuadrados de β_0, β_1 y β_2. Pruebe que las estimaciones de mínimos cuadrados de β_0 y β_1 no cambian si $\beta_2 = 0$.

(Solución)

Sea

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix},$$

y $w_i = 3x_i^2 - 2$ para cada $i = 1, 2, 3$. Consideremos

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

La matriz de diseño $\mathbf{Z}$ tomando en cuenta la variable β_2 viene dada por la matriz bloque

$$\mathbf{Z} = [\mathbf{X}, \mathbf{w}].$$

Observamos que

$$\mathbf{Z}^T\mathbf{Z} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{w} \\ \mathbf{w}^T\mathbf{X} & \mathbf{w}^T\mathbf{w}. \end{bmatrix}$$

Realizando unas cuentas matriciales se llega a que

$$\mathbf{X}^T\mathbf{w} = \begin{bmatrix} w_1 + w_2 + w_3 \\ x_1w_1 + x_2w_2 + x_3w_3. \end{bmatrix}$$

Como $x_1 = -1, x_2 = 0, x_3 = 1$ entonces

$$w_1 = 1, \quad w_2 = -2 \ \text{ y } w_3 = 1.$$

Por lo tanto

$$\mathbf{X}^T\mathbf{w} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

además

$$\mathbf{w}^T\mathbf{X} = [0, 0],$$

y

$$\mathbf{w}^T\mathbf{w} = 6$$

por lo que

$$\mathbf{Z}^T\mathbf{Z} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{0} \\ \mathbf{0} & 6. \end{bmatrix},$$

en donde $\mathbf{0}$ tiene los tamaños apropiados según el bloque en que se encuentre.
Si $\boldsymbol{\beta}_{01} = [\beta_0, \beta_1]^T$ entonces el modelo de regresión considerando las variables $\boldsymbol{\beta} = [\boldsymbol{\beta}_{01}, \beta_2]$ es

$$\mathbb{E}[\mathbf{Y}] = \mathbf{Z}\boldsymbol{\beta},$$

luego la estimación por mínimos cuadrados $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ viene dada por

$$\begin{aligned}
\hat{\boldsymbol{\beta}} =& (\mathbf{Z}^T\mathbf{Z})\mathbf{Z}^T\mathbf{Y} \\
=& \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{0} \\ \mathbf{0} & 6 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{w}^T \end{bmatrix} \mathbf{Y} \\
=& \begin{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \\ \frac{1}{6}\mathbf{w}^T\mathbf{Y} \end{bmatrix}.
\end{aligned}$$

Por lo tanto las estimaciones de mínimos cuadrados de β_0, β_1 en este caso son

$$\hat{\boldsymbol{\beta}}_{01} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

que coinciden con la del modelo de regresión considerando $\beta_2 = 0$ y la estimación de mínimos cuadrados de β_2 es

$$\hat{\beta}_2 = \frac{1}{6}\mathbf{w}^T\mathbf{Y}.$$

Realizando las cuentas con el código en Python llegamos a que

$$\hat{\beta}_0 = \frac{1}{3}(Y_1 + Y_2 + Y_3), \quad \hat{\beta}_1 = \frac{1}{2}(Y_3 - Y_1), \quad \text{y } \hat{\beta}_2 = \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3.$$

□

Problema 3

Pruebe que

$$(Y - X\beta)^T(Y - X\beta) = (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (\hat{\beta} - \beta)^T X^T X(\hat{\beta} - \beta),$$

y enseguida deduzca que el lado izquierdo se minimiza cuando $\beta = \hat{\beta}$.

(Solución)

En primer lugar notamos que

$$(Y - X\beta)^T(Y - X\beta) = Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta.$$

como todos los términos son matrices de 1×1 podemos simplificar a

$$(Y - X\beta)^T(Y - X\beta) = Y^TY - 2Y^TX\beta + \beta^TX^TX\beta.$$

De manera análoga haciendo las cuentas pertinentes se obtiene

$$(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = Y^TY - 2\hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta}, \tag{1.1}$$

y

$$(\hat{\beta} - \beta)^TX^TX(\hat{\beta} - \beta) = \hat{\beta}^TX^TX\hat{\beta} - 2\hat{\beta}^TX^TX\beta + \beta^TX^TX\beta. \tag{1.2}$$

Recordamos que $\hat{\beta} = (X^TX)^{-1}X^TY$, por lo que

$$\hat{\beta}^TX^TX\hat{\beta} = Y^TX(X^TX)^{-1}X^TY,$$

y

$$\hat{\beta}^TX^TY = Y^TX(X^TX)^{-1}X^TY,$$

por lo que

$$\hat{\beta}^TX^TX\hat{\beta} = \hat{\beta}^TX^TY$$

Además se tiene que

$$\begin{aligned}\hat{\beta}^TX^TX\beta =& Y^TX(X^TX)^{-1}X^TX\beta \\ =& Y^TX\beta,\end{aligned}$$

De estas dos últimas identidades, sumando (1.1) y (1.2) se sigue que

$$\begin{aligned}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (\hat{\beta} - \beta)^TX^TX(\hat{\beta} - \beta) =& Y^TY + 2(\hat{\beta}^TX^TX\hat{\beta} - \hat{\beta}^TX^TY) - 2Y^TX\beta + \beta^TX^TX\beta \\ =& Y^TY - 2Y^TX\beta + \beta^TX^TX\beta \\ =& (Y - X\beta)^T(Y - X\beta),\end{aligned}$$

que era lo que queríamos ver.

Notamos que

$$(\hat{\beta} - \beta)^TX^TX(\hat{\beta} - \beta) = \langle X(\hat{\beta} - \beta), X(\hat{\beta} - \beta)\rangle \geq 0,$$

por lo que

$$(Y - X\beta)^T(Y - X\beta) \geq (Y - X\hat{\beta})^T(Y - X\hat{\beta}),$$

para todo vector β de las dimensiones apropiadas, y la igualdad se da si y sólo si

$$\langle X(\hat{\beta} - \beta), X(\hat{\beta} - \beta)\rangle = 0,$$

que pasa si y sólo si

$$X(\hat{\beta} - \beta) = 0,$$

y como suponemos X de rango completo esto sólo sucede si

$$\hat{\beta} - \beta = 0,$$

lo que implica que

$$\beta = \hat{\beta}.$$

Concluimos que

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}(Y - X\beta)^T(Y - X\beta),$$

que es a lo que queríamos llegar.

$\square$

Problema 4

Pruebe que $\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0$.

(Solución)

Sea $Y = [Y_1, \ldots, Y_n]^T$. Estamos considerando el modelo de regresión lineal múltiple

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i.$$

Si $\mathbf{X}$ es la matriz de diseño entonces

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \ldots & X_{1p} \\ 1 & X_{21} & \ldots & X_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & X_{n1} & \ldots & X_{np} \end{bmatrix}.$$

Si $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]^T$ es el vector de errores que se distribuyen normales con media cero y varianza constante el modelo de regresión lineal en notación matricial es

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

donde $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_p]^T$ es el vector de parámetros.

Sea $\mathbf{P} = X(X^T X)^{-1} X^T$, sabemos que esa es una matriz de proyección y las columnas de $\mathbf{P}$ son combinación lineales de las columnas de $\mathbf{X}$, por lo que

$$\operatorname{Col}\mathbf{P} = \operatorname{Col}\mathbf{X},$$

donde $\operatorname{Col}\mathbf{X}$ representa el espacio de columnas de la matriz $\mathbf{X}$.

Sea $\mathbf{1} = [\underbrace{1, 1, \ldots, 1}_{\text{n veces}}]^T$, de la definición de $\mathbf{X}$ es claro que $\mathbf{1} \in \operatorname{Col}\mathbf{X}$ pues es una columna de la matriz $\mathbf{X}$, por lo que

$$\mathbf{1} \in \operatorname{Col}\mathbf{P}. \tag{1.3}$$

Por otra parte como $\hat{\mathbf{Y}} = \mathbf{P}\mathbf{Y}$ entonces

$$\mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P})\mathbf{Y},$$

y también $\mathbf{I} - \mathbf{P}$ es una matriz de proyección pero al espacio $(\operatorname{Col}\mathbf{P})^{\perp}$, de esto se tiene

$$\mathbf{Y} - \hat{\mathbf{Y}} \in (\operatorname{Col}\mathbf{P})^{\perp}. \tag{1.4}$$

Si $\langle \mathbf{x}, \mathbf{y} \rangle$ denota el producto interior de los vectores $\mathbf{x}$ y $\mathbf{y}$, entonces por (1.3) y (1.4) se sigue que

$$\langle \mathbf{Y} - \hat{\mathbf{Y}}, \mathbf{1} \rangle = \sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0,$$

que es lo que queríamos probar.

$\square$

Problema 5

Sean $Y_1, \dots, Y_n$ variables aleatorias independientes, con distribución común $N(\mu, \sigma^2)$. Sean

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \text{y} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

(a) ¿Cuál es la distribución de $\overline{Y}$?

(Solución)

Como $Y_i \sim N(\mu, \sigma^2)$ y son independientes entonces

$$Y := (Y_1, \dots, Y_n) \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{I}),$$

donde $\mathbf{1} = [\underbrace{1, 1, \dots, 1}_{\text{n veces}}]^T$ y $\mathbf{I}$ es la matriz identidad de $n \times n$.

Sea $\mathbf{u} = \frac{1}{n}\mathbf{1}$. Observamos que

$$\overline{Y} = \mathbf{u}^T Y,$$

por lo que

$$\overline{Y} \sim N(\mathbf{u}^T\mu\mathbf{1}, \mathbf{u}^T\sigma^2\mathbf{I}\mathbf{u}).$$

Como

$$\mathbf{u}^T\mu\mathbf{1} = \mu,$$

y

$$\mathbf{u}^T\sigma^2\mathbf{I}\mathbf{u} = \frac{\sigma^2}{n},$$

concluimos

$$\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

que es lo que buscábamos.

(b) Pruebe que

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i - \mu)^2 - n(\overline{Y} - \mu)^2\right].$$

(Solución)

En primer lugar, desarrollando y agrupando términos

$$\begin{aligned}
\frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i - \mu)^2 - n(\overline{Y} - \mu)^2\right] &= \frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2 - 2\mu\sum_{i=1}^{n}Y_i + n\mu^2 - n(\overline{Y}^2 - 2\mu\overline{Y} + \mu^2)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2 - 2n\mu\overline{Y} - n\overline{Y}^2 + 2n\mu\overline{Y}\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2 - n\overline{Y}^2\right].
\end{aligned}$$

De lo anterior se sigue que

$$\frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2-n(\overline{Y}-\mu)^2\right]=\frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-n\overline{Y}^2\right]. \tag{1.5}$$

Por otro lado desarrollando el binomio al cuadrado en s^2 y simplificando obtenemos que

$$\begin{aligned} s^2 =& \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i^2-2Y_i\overline{Y}+\overline{Y}^2)\right] \\ =& \frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-2\overline{Y}\sum_{i=1}^{n}Y_i+n\overline{Y}^2\right] \\ =& \frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-2n\overline{Y}^2+n\overline{Y}^2\right] \\ =& \frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-n\overline{Y}^2\right]. \end{aligned}$$

Por lo tanto

$$s^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-n\overline{Y}^2\right], \tag{1.6}$$

de (1.5) y (1.6) concluimos que

$$s^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2-n(\overline{Y}-\mu)^2\right],$$

que es lo que se quería probar.

(c) Del inciso (b) se sigue que

$$\sum_{i=1}^{n}\frac{(Y_i-\mu)^2}{\sigma^2}=\frac{(n-1)s^2}{\sigma^2}+\left[\frac{(\overline{Y}-\mu)^2 n}{\sigma^2}\right]. \tag{1.7}$$

¿Como se sigue de aquí que $\overline{Y}$ y s^2 son independientes?

(Solución)
Sabemos del inciso (a) que $\overline{Y}=\mathbf{u}^TY$.

Sea $\{e_i\}$ la base canónica de $\mathbb{R}^n$. Notamos que

$$Y_i-\overline{Y}=e_i^TY-\mathbf{u}^TY=(e_i-\mathbf{u})^TY$$

Por lo tanto

$$\begin{pmatrix}\overline{Y}\\ Y_i-\overline{Y}\end{pmatrix}=\begin{pmatrix}\mathbf{u}^T\\ e_i-\mathbf{u}^T\end{pmatrix}Y.$$

Y como $Y \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{I})$ si $A = \begin{pmatrix} \mathbf{u}^T \\ e_i - \mathbf{u}^T \end{pmatrix}$ entonces

$$\begin{pmatrix} \overline{Y} \\ Y_i - \overline{Y} \end{pmatrix} \sim N_2(\mu A\mathbf{1}, \sigma^2 AA^T),$$

luego para probar la independencia de $\overline{Y}$ y $Y_i - \overline{Y}$ basta con verificar que su covarianza es 0.

Por la bilinealidad de la covarianza y desarrollando todas las expresiones tenemos

$$\begin{aligned} \mathrm{Cov}(Y_i - \overline{Y}, \overline{Y}) =& \mathrm{Cov}((e_i - \mathbf{u})^T Y, \mathbf{u}^T Y) \\ =&(e_i - \mathbf{u})^T \mathrm{Var}(Y)\mathbf{u} \\ =&\sigma^2 (e_i - \mathbf{u})^T \mathbf{u} \\ =&\sigma^2 \left[e_i^T \mathbf{u} - \mathbf{u}^T \mathbf{u}\right] \\ =&\sigma^2 \left[\frac{1}{n} - \frac{n}{n^2}\right] \\ =&0. \end{aligned}$$

Concluimos de lo anterior que $Y_i - \overline{Y}$ es independiente de $\overline{Y}$. Como s^2 es una función continua de las variables $Y_i - \overline{Y}$ entonces s^2 es independiente de $\overline{Y}$ que es lo que queríamos ver.

(d) ¿Cuál es la distribución de $\frac{(n-1)s^2}{\sigma^2}$?

Para cada $i = 1, 2, \ldots, n$ sea $Z_i = \frac{Y_i - \mu}{\sigma}$.

Como $\{Y_i\}$ es una familia de variables aleatorias independientes con $Y_i \sim N(\mu, \sigma^2)$ entonces $\{Z_i\}$ es una familia de variables aleatorias independientes con $Z_i \sim N(0, 1)$.

Por lo tanto el lado izquierdo de (1.7) es la variable aleatoria

$$Z = \sum_{i=1}^{n} Z_i^2,$$

por lo que $Z \sim \chi^2_{(n)}$, si M es la función generadora de momentos de Z entonces

$$M(t) = (1 - 2t)^{-\frac{n}{2}} \quad \text{para } 2t < 1.$$

Sea $U = \frac{(n-1)s^2}{\sigma^2}$ y $W = \frac{(\overline{Y}-\mu)^2 n}{\sigma^2}$, sean M_U y M_W sus funciones generadoras de momentos respectivas.

Del inciso (c) se tiene que U y W son independientes, de esto y (1.7) podemos factorizar la función generadora de momentos M de la siguiente forma

$$M(t) = M_U(t) M_W(t). \tag{1.8}$$

Del inciso (a) tenemos que $\overline{Y} \sim N(\mu, \sigma^2/n)$, por lo que $\frac{(\overline{Y}-\mu)\sqrt{n}}{\sigma} \sim N(0, 1)$, lo que implica $W \sim \chi^2_{(1)}$, luego

$$M_W(t) = (1 - 2t)^{-\frac{1}{2}}.$$

De esto y (1.8) se sigue

$$M_U(t) = (1-2t)^{-\frac{(n-1)}{2}} \quad \text{para } 2t < 1.$$

Concluimos así que $U \sim \chi^2_{(n-1)}$.

(e) ¿Cuál es la distribución de $\frac{\overline{Y}-\mu}{s/\sqrt{n}}$?
Sea

$$T = \frac{(\overline{Y}-\mu)}{s/\sqrt{n}}.$$

Multiplicando y dividiendo por σ a la expresión anterior y reacomodando los términos tenemos

$$T = \frac{(\overline{Y}-\mu)/(\sigma/\sqrt{n})}{s/\sigma}.$$

En primer lugar notamos que

$$\frac{s}{\sigma} = \sqrt{\frac{U}{(n-1)}},$$

donde U es como en el inciso anterior.
Sea

$$Z = \frac{(\overline{Y}-\mu)}{\sigma/\sqrt{n}}.$$

por el inciso (a) se tiene $Z \sim N(0,1)$.
Notamos que con estos cambios de variable

$$T = \frac{Z}{\sqrt{U/((n-1)}},$$

donde Z y U son independientes por el inciso (c), $Z \sim N(0,1)$ y $U \sim \chi^2_{(n-1)}$.
Concluimos de lo anterior que $T \sim t_{(n-1)}$ donde $t_{(n-1)}$ denota a la distribución t-Student con $n-1$ grados de libertad, y esto es lo que buscábamos determinar.

$\square$

```
# (Tarea 3 Modelos Estadisticos)
import numpy as np
import sympy as sym

##############
# PROBLEMA 1 #
##############
X=np.array([[1,0],[2,-1],[1,2]])
y1, y2, y3 = sym.symbols('y1,y2,y3')
Y=np.array([y1,y2,y3])
XTX_inv=np.linalg.inv(X.T@X)
print("Los estimadores de theta y phi son: ")
print(XTX_inv@ X.T@ Y)

##############
# PROBLEMA 2 #
##############
X=np.array([[1,-1,1],[1,0,-2],[1,1,1]])
XTX_inv=np.linalg.inv(X.T@X)
print("\n\nLos estimadores de b0, b1 y b2 son: ")
print(XTX_inv@ X.T@ Y)
```

Listing 1: Código Tarea 3