Universidad Nacional Autónoma de México

Facultad de Ciencias

Tarea - Examen 3

Modelos No Paramétricos y de Regresión
Semestre 2023 - 2

Ruth Selene Fuentes García
Jorge Iván Reyes Hernández

Alumno: Olvera Trejo Alberto

---

1. Ejercicios

Ejercicio 1. Para el modelo $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ con las hipótesis usuales: $\mathbb{E}\left[\mathbf{e}\right] = 0$ y $\mathbb{E}\left[\mathbf{e} \cdot \mathbf{e}^t\right] = \sigma^2\mathbf{I}$, con una distribución normal para los errores.

a) Muestre que: $Var(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H}$

b) Muestre que: $SS_{(R)} = \mathbf{y^t H y}$

c) Obtenga un intervalo de confianza para el valor esperado de $\mathbf{Y}$ dado que el vector $\mathbf{X} = \mathbf{X_0}$

Demostración. a) Recordemos que $\hat{\beta} = (X^tX)^{-1} X^t y$, entonces se sigue que:

$$\hat{y} = X\hat{\beta} = X \left(X^tX\right)^{-1} X^t y$$

Procedemos a calcular la varianza

$$\begin{aligned}
Var(\hat{Y}) &= Var(X\hat{\beta}) \\
&= Var(X \left(X^tX\right)^{-1} X^t y) \\
&= [X \left(X^tX\right)^{-1} X^t] Var(y) [X \left(X^tX\right)^{-1} X^t]^t \\
&= \sigma^2 [X \left(X^tX\right)^{-1} X^t][X \left(X^tX\right)^{-1} X^t] \\
&= \sigma^2 X \left(X^tX\right) X^t \\
&= \sigma^2 H
\end{aligned}$$

Por lo tanto $Var(y) = \sigma^2 H$

b) Recordemos que $SS_R = r^t r$, con $r = y - \hat{y} = (I - H)\, y$. Susitutyendo en la formula de SS_R tenemos que:

$$\begin{aligned} SS_R &= r^t r \\ &= [(I - H)\, y]^t [(I - H)\, y] \\ &= [y^t\, (I - H)^t][(I - H)\, y] \end{aligned}$$

Usando que H es simétrica y además es idempotente:

$$\begin{aligned} SS_R &= [y^t\, (I - H)^t][(I - H)\, y] \\ &= [y^t \left(I - H^t\right)][(I - H)\, y] \\ &= y^t \left(I^2 - IH - HI + H^2\right) y \\ &= y^t\, (I - 2H + H)\, y \\ &= y^t\, (I - H)\, y \end{aligned}$$

Por lo tanto $SS_R = y^t\, (I - H)\, y$

c) Por lo visto en clase, sabemos que $\frac{n\hat{\sigma^2}}{\sigma^2} \sim \chi^2_{n-k}$ y que $\hat{\beta} \sim N\left(\beta, \sigma^2\, (X'X)^{-1}\right)$. Por otro lado, $\hat{Y}_0 = X_0\hat{\beta}$ de distribuye de manera normal debido a que es combinación lineal de normales. Vamos a calcular sus parámetros.

$$\mathbb{E}\left[\hat{Y}_0\right] = \mathbb{E}\left[X_0\hat{\beta}\right] = X_0\mathbb{E}\left[\hat{\beta}\right] = X_0\beta$$

$$Var(\hat{Y}_0) = Var(X_0\hat{\beta}) = X_0 Var(\hat{\beta}) X_0' = \sigma^2 X_0\, (X'X)^{-1} X_0'$$

Por lo tanto $\hat{Y}_0 \sim N\left(X_0\beta, \sigma^2 X_0\, (X'X)^{-1} X_0'\right)$. Al estandarizar tenemos que:

$$Z = \frac{\hat{Y}_0 - X_0\beta}{\sqrt{\sigma^2 X_0\, (X'X)^{-1} X_0'}}$$

Luego, usando que $U = \frac{n\hat{\sigma^2}}{\sigma^2} \sim \chi^2_{n-k}$, entonces tenemos que:

$$\frac{Z}{\sqrt{U/n-k}} \sim t_{(n-k)}$$

Entonces, el intervalo de confianza está dado por:

$$\left(\hat{Y}_0 - t_{\alpha/2,n-k}\sqrt{\sigma^2 X_0\, (X'X)^{-1} X_0'}, \quad \hat{Y}_0 + t_{\alpha/2,n-k}\sqrt{\sigma^2 X_0\, (X'X)^{-1} X_0'}\right)$$

QED

Ejercicio 2. Considere el modelo lineal de la forma $\mathbf{y} = \mathbf{Z}\beta + \mathbf{e}$; donde $\mathbb{E}[\mathbf{e}] = 0$ y $\mathbb{E}[\mathbf{e} \cdot \mathbf{e}^t] = \sigma^2\mathbf{V}$, donde $\mathbf{V}$ es una matriz conocida, positiva definida, de rango completo. Muestre que el estimador de mínimos cuadrados es $\widehat{\beta_W} = (\mathbf{Z^t V^{-1} Z})^{-1} \mathbf{Z^t V^{-1} y}$. Utilice que $\mathbf{V^{-1/2} y} = \mathbf{V^{-1/2} Z}\beta + \mathbf{V^{-1/2} e}$

Demostración. Dado que V es una matriz positiva definida, entonces existe una única matriz, denotada por $V^{1/2}$, tal que $\left(V^{1/2}\right)' V^{1/2} = V^{1/2}V^{1/2} = V$. Tomemos la matriz inversa de $V^{1/2}$, denotada por $V^{-1/2}$. Por la información proporcionada sabemos que:

$$V^{-1/2}y = V^{-1/2}Z\beta + V^{-1/2}e$$

Sean $y^* = V^{-1/2}y$, $X = V^{-1/2}Z$ y $e^* = V^{-1/2}e$, entonces la ecuación anterior se transforma a:

$$y^* = X\beta + e^*$$

Calculemos $\mathbb{E}[e]$ y $Var(e)$

$$\begin{aligned}\mathbb{E}[e] &= \mathbb{E}\left[V^{-1/2}e\right] \\ &= V^{-1/2}\mathbb{E}[e] \\ &= V^{-1/2} \cdot 0 = 0\end{aligned}$$

Por otro lado:

$$\begin{aligned}Var(e) &= Var(V^{-1/2}e) \\ &= V^{-1/2}Var(e)\left(V^{-1/2}\right) \\ &= V^{-1/2}\sigma^2 V\left(V^{-1/2}\right) \\ &= \sigma^2 V^{-1/2}\left(V^{1/2}\right)V^{1/2}\left(V^{-1/2}\right) \\ &= \sigma^2 I\end{aligned}$$

Por lo tanto $y^* = X\beta + e^*$ es un problema de regresión lineal múltiple con los supuestos usuales, y por lo visto en clase el estimador de mínimos cuadrados está dado por:

$$\hat{\beta} = \left(X^t X\right)^{-1} X^t y^*$$

Sustituyendo $X = V^{-1/2}Z$, $y^* = V^{-1/2}y$ obtenemos:

$$\begin{aligned}\hat{\beta} &= \left(\left(V^{-1/2}Z\right)^t V^{-1/2}Z\right)^{-1}\left(V^{-1/2}Z\right)^t y^* \\ &= \left(Z^t V^{-1/2}V^{-1/2}Z\right)^{-1} Z^t\left(V^{-1/2}\right)^t V^{-1/2}y \\ &= \left(Z^t V^{-1}Z\right)^{-1} Z^t V^{-1}y\end{aligned}$$

Por lo tanto, el estimador de mínimos cuadrados es:

$$\hat{\beta} = \left(Z^t V^{-1}Z\right)^{-1} Z^t V^{-1}y$$

QED

Ejercicio 3. Un investigador está interesado en determinar el log(10) de conteos de microbios obtenidos de un cupó contaminado de $2.3cm^2$ a diferentes temperaturas y diferentes medios. Su hipótesis sostiene que las variaciones de temperatura de 208řC a 408řC y la concentración del medio afectarían los conteos. Puede apoyar la hipótesis con un modelo lineal? (mic.txt: y $= \log_{10} conteos$, $x_1 = temperatura$, $x_2 = medio$, $x_3 = x_1 \cdot x_2$)