Práctico 2
Regresión Lineal Múltiple

Modelos Lineales - 2024

EJERCICIO 1

En el modelo lineal de la forma:

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

Donde la *matriz de diseño* X tiene n filas y $k+1$ columnas (con $n < k+1$), se define la matriz de proyección ortogonal:

$$H = X\left(X^{'}X\right)^{-1}X^{'}$$

Demuestre las siguientes propiedades:

1. $H^j = H \quad \forall j = 1, 2, 3, \ldots$
2. Si H es idempotente, entonces $I_{n\times n} - H$ también lo es
3. $H(I_{n\times n} - H) = \mathbf{0}_{n\times n}$

EJERCICIO 2

Demuestre que la matriz identidad es una matriz de proyección ortogonal y especifique a qué modelo lineal corresponde.

EJERCICIO 3

Los elementos de la diagonal de la matriz H poseen ciertas caracteríticas de interés, sobre todo los elementos de su diagonal (que denotaremos por h_{ii}).

- Pruebe que $\frac{\partial \hat{y}_i}{\partial y_i} = h_{ii}$
- Pruebe que $\mathbb{V}ar(\hat{y}_i) = \sigma^2 h_{ii}$
- Pruebe que $\mathbb{V}ar(\hat{\epsilon}_i) = \sigma^2(1 - h_{ii})$
- Pruebe que $\mathbb{C}orr(y_i, \hat{y}_i) = \sqrt{h_{ii}}$
- Prueba que $0 \leq h_{ii} \leq 1$

Nota: Para demostrar el último punto debe considerar las propiedades de simetría e idempotencia y plantear una expresión de h_{ii} como producto de la i-ésima fila de H por su i-ésima columna.

EJERCICIO 4

Considere el modelo $Y = X\beta + \epsilon$, con los siguientes supuestos:

- $rg(X) = k + 1$
- $\mathbb{E}(\epsilon) = \mathbf{0}$
- $\mathbb{V}ar(\epsilon) = \sigma^2 V$

Siendo V una matriz conocida **distinta** de la identidad.

1. Demuestre que el estimador $\hat{\beta}_{MCO}$ tiene por matriz de covarianzas:

$$\mathbb{V}ar(\hat{\beta}_{MCO}) = \sigma^2 X^{'} \left(X^{'}X\right)^{-1} VX \left(X^{'}X\right)^{-1} X^{'}$$

2. ¿Por qué el estimador del item anterior no es $MELI$?
3. Para obtener un estimador $MELI$ considere la siguiente descomposición de la matriz V.

$$V = PP^{'}$$

Luego, premultiplique el modelo por P^{-1}, es decir:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\epsilon$$

Y denote $Y^* = P^{-1}Y$, $X^* = P^{-1}X$ y $\epsilon^* = P^{-1}\epsilon$.
¿Qué efecto tiene esta operación sobre los errores del modelo?

4. Demuestre que el estimador de β en este modelo corregido es:

$$\hat{\beta}_{MCG} = \left(X^{'}V^{-1}X\right)^{-1} X^{'}V^{-1}Y$$

y que **sí** es $MELI$.

5. Obtenga la matriz de covarianzas de $\hat{\beta}_{MCG}$

EJERCICIO 5

Suponga que en una aplicación económica donde se quiere modelar el ahorro de un conjunto de hogares en función de sus ingresos, el modelo correcto a aplicar es:

$$ahorro_i = \beta_0 + \beta_1 ingreso_i + \epsilon_i$$

Pero que usted decide elimintar la constante del modelo debido a que *sin engresos es imposible ahorrar*. Por ende, ajusta el siguiente modelo:

$$ahorro_i = \beta_1^* ingreso_i + \epsilon_i^*$$

- Demuestre que el estimador de mínimos cuadrados es:

$$\hat{\beta}_1 = \frac{\sum_i ingreso_i \times ahorro_i}{\sum_i ingreso_i^2}$$

- Demuestre que (si el modelo correcto incluye la constante) este estimador es sesgado.
- Encuentre una expresión para dicho sesgo.
- ¿Qué idea se le ocurre para solucionar el problema del sesgo y continuar trabajando con el modelo sin constante?

EJERCICIO 6

En el archivo *ageandheight.xls* hay mediciones de la altura (en cm), edad (en meses) y cantidad de hermanos de un conjunto infantes. Con estos datos, y valiéndose de R, se pide lo siguiente:

1) Utilice la librería *readxl* para cargar los datos guardándolos en un objeto llamado *edadaltura* (también es recomendable cambiar el directorio de trabajo previamente).

2) Realice un breve análisis descriptivo de los datos (visualización de datos, medidas de resumen y correlaciones)

3) Utilizando la funcion *lm*, plantee un modelo de *RLS* simple que relacione la altura con la edad. Represéntelo gráficamente.

4) Corrobore que el cuadrado de la correlación entre los valores predichos y los originales coincide con el R^2 del modelo.

5) ¿Qué ocurre con el modelo si cambiamos la escala la variable edad, de meses a años?

Por otro lado, utilizando los datos de sobre presión y temperatura del archivo *pressure.xlsx* considere los siguientes puntos:

6) Realice un modelo lineal que permita predecir la presión a partir de la temperatura, ¿Existe algún problema con este modelo? ¿Qué ocurre con sus residuos?

7) Realice un nuevo modelo que incorpore el cuadrado de la temperatura como una nueva variable explicativa. ¿Cuál de los dos modelos prefiere?

EJERCICIO 7

En el archivo *pgc.txt* se encuentran las mediciones de 252 personas. Se desea construir una fórmula que sea capaz de estimar el porcentaje de grasa corporal en función de algunas mediciones (circunferencia abdominal, circunferencia de cintura, de cuello, entre otras).

1. Visualice la relación de cada una de las medidas corporales con respecto a la variable porc_grasa y estime las correlaciones.

2. Estime la matriz de correlación de las mediciones corporales.

3. Describa con sus palabras las conclusiones que haya podido obtener a partir de los dos puntos anteriores.

4. Realice un conjunto de modelos de *RLS* que expliquen la relación de cada medida corporal con respecto al porcentaje de grasa corporal. ¿Qué observa?

5. Ajuste la siguiente serie de modelo:

   - $porc\_grasa_i = \beta_0 + \beta_1 abdom_i + \epsilon_i$.
   - $porc\_grasa_i = \beta_0 + \beta_1 abdom_i + \beta_2 cintura_i + \epsilon_i$.
   - $porc\_grasa_i = \beta_0 + \beta_1 abdom_i + \beta_2 cintura_i + \beta_3 cuello_i + \epsilon_i$.
   - $porc\_grasa_i = \beta_0 + \beta_1 abdom_i + \beta_2 cintura_i + \beta_3 cuello_i + \beta_4 muñeca + \epsilon_i$.

   A partir de las estimaciones de estos modelo:

   - Compare la evolución del R^2.
   - Compare la evolución de la estimación de σ^2.
   - Compare la evolución del coeficiente asociado a la circunferencia abdominal.