Primera Revisión
Jueves 11 de mayo

Modelos Lineales - 2023

EJERCICIO 1

Considere el siguiente modelo de regresión lineal múltiple:.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

Indique la matriz C y el vector b que necesitaría emplear para poner a prueba la siguiente hipótesis lineal

$$\begin{array}{ll} H_0) & \beta_2 = 0, \beta_3 + \beta_4 = 1 \\ H_1) & \text{no } H_0) \end{array}$$

EJERCICIO 2

En grupo liceal conformado por 15 estudiantes se quiere determinar si es posible inferir la nota de un examen de matemática a partir del resultado del examen de química. Se cuenta con los siguientes datos:

- $\sum_i x_i = 115,82$.
- $\sum_i y_i = 94,86$.
- $\sum_i x_i y_i = 762,97$.
- $\sum_i x_i^2 = 970,88$.
- $\sum_i y_i^2 = 614,74$.

Siendo X la variable "Nota de química" y Y la variable "Nota de Matemática". A partir de estos datos:

1. Plantee el modelo de regresión lineal simple en el que se pretende explicar la nota del exámen de matemática en función de la nota del exámen de química.
2. Obtenga el valor de la correlación entre X e Y.
3. Obtenga el valor del R^2.
4. Estime los componentes del vector β.
5. Obtenga la predicción de la nota del exámen de matemática para un alumno que en el examen de química obtuvo 7 puntos.

EJERCICIO 3

Una empresa crediticia está interesada en conocer cuáles son los factores que afectan la deuda en tarjetas de crédito en una cartera de clientes. Con este objetivo se ajusta un modelo de regresión lineal múltiple (RLM) empleando como variable dependiente el monto de la deuda en tarjetas de crédito como variable de respuesta. Las variables explicativas empleadas fueron el límite de crédito medido en u.m. ($Limit$), el ingreso de los clientes medido en miles de u.m. ($Income$), los años de educación (*Education*) y el número de tarjetas que posee cada cliente ($Cards$).

A continuación se presenta le *summary* del modelo de RLM correspondiente a esta situación.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -466.43      44.69  -10.44   <0.001 ***
Limit           0.26       0.01   45.37   ???
Income         -7.60       0.38  -19.98   <0.001 ***
Education       1.49       2.62    ???     0.569
Cards            ???       5.98    3.52   <0.001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 163.2 on 396 degrees of freedom
Multiple R-squared:  0.875,     Adjusted R-squared:  0.874
F-statistic: 923.8 on 3 and 396 DF,  p-value: < 0.001
```

A partir de esta información responda los siguientes puntos.

1. Complete la tabla
2. ¿Con cuantas observaciones se contó para estimar el modelo?
3. ¿Qué porcentaje de la variabilidad del rendimiento pudo ser explicada por el modelo?
4. Determine la significación global del modelo justificando sus afirmaciones.
5. Considerando un nivel de significación del 5 %, ¿es significativo el aporte de la variable "*Limit*"?
6. ¿Podría contestar la pregunta de respecto de cuáles son los factores que inciden sobre el monto de la deuda?
7. Interprete el valor estimado del parámetro asociado a la variable $Cards$.
8. Realice la predicción correspondiente a un cliente que posee 2 tarjetas de crédito, tiene un ingreso de 60, un límite de crédito de 3000 y 12 años de educación.

EJERCICIO 4

Considere el modelo lineal general en su forma matricial $Y = X\beta + \epsilon$. Suponga que las variables explicativas son sometidas a un cambio de escala posmultiplicando la matriz X por una matriz cuadrada y diagonal Λ obteniendo $Z = X\Lambda$.
Demuestre que los residuos obtenidos en la regresión de Y sobre X son los mismos que se obtienen al regresar Y sobre Z.

Nota: Recuerde que, si A, B y C son matrices invertibles (y de la misma dimensión) se cumple que $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

EJERCICIO 5

En base al modelo de regresión lineal simple dado por la ecuación

$$y = \beta_0 + \beta_1 x + \epsilon$$

se observó que las estimaciones de los coeficientes estaban dadas por:

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \tfrac{s_y}{s_x} r_{xy}\end{aligned}$$

Determine

1. ¿Cuál es la estimación de β_0 cuando se “centra” la variable x?
2. ¿Y cuándo se centran ambas variables?
3. ¿Qué modificación debería realizar a x y/o a y para que la estimación de β_1 coincida con el coeficiente de correlación?