

Modelos Lineales con R

Práctica 3: Factores y selección de modelos

Profesor: Andrés García Medina
andres.garcia.medina@uabc.edu.mx

Fecha de entrega: jueves 14 de marzo, 2024 (12pm).

Instrucciones: Subir un documento pdf a classroom con las respuestas de cada uno de los ejercicios solicitados. Adjunta el código fuente en formato `.r` o `.ipynb`. Justificar detalladamente cada una de sus respuestas.

Ejercicio 1

Resuelva los siguientes problemas conceptuales:

1. Dadas n observaciones x_i, y_i , encuentre la estimación de mínimos cuadrados de β en el modelo lineal: $y_i = \mu_i + \epsilon_i$; $\mu_i = \beta$.
2. Escriba el siguiente modelo en la forma $y = X\beta + \epsilon$, asegurándose de que sea identificable. Un modelo con dos variables explicativas: una variable del tipo factor y la otra variable continua:

$$y_i = \alpha + \beta_j + \gamma x_i + \epsilon_i \quad \text{si la observación } i \text{ está asociada al factor } j \quad (1)$$

asuma $i = 1, \dots, 6$, y que las primeras dos observaciones están asociadas al factor 1, mientras que las restantes 4 al factor 2. Además suponga que las x_i 's son 0.1, 0.4, 0.5, 0.3, 0.4, y 0.7

3. Considere los valores de deformación (mm) y_i de 3 tipo de aleaciones distintas bajo diferentes cargas (kg) x_i . Cuando no hay carga no hay deformación, y la deformación se espera que cambie linealmente con la carga de la misma manera (exactamente) para las 3 aleaciones. Sin embargo, a medida que la carga aumenta las aleaciones se desvían de su comportamiento lineal (ideal) de distintas formas, volviéndose la relación ligeramente curva (cuadrática). Las cargas se conocen de manera precisa, por lo que los errores en x_i 's se pueden ignorar, pero las deformaciones están sometidas a errores de medición. Defina un modelo lineal de la forma $y = X\beta + \epsilon$ capaz de describir estos datos, asumiendo que las mismas 6 cargas se aplican a cada aleación.

Ejercicio 2

Considere los datos en R denominados `InsectSprays`, los cuales contienen recuentos de insectos en cada una de varias parcelas. Cada una de las parcelas ha sido rociada con un insecticida distinto. Un modelo para estos datos podría ser:

$$y_i = \mu + \beta_j, \quad \text{si la parcela } i \text{ ha sido rociada con el spray } j \quad (2)$$

- (a) a) ¿Cuántos factores y niveles tiene el conjunto de datos? Se puede apoyar en el comando `as.factor`
- (b) b) Ajuste un modelo lineal propuesta para determinar si el tipo de insecticida determina el conteo de insectos, y en dado caso en que medida. ¿Cual insecticida tiene mayor impacto en la disminución del numero de mosquitos?
- (c) Determine el tipo impacto que tiene el spray A sobre el conteo de mosquitos usando como base el nivel F. Hint: Use la función `relevel`.
- (d) d) Analice el ajuste del modelo a través de los **cuantiles de los residuales** con apoyo de la función `qnorm`. Compare su gráfico con los que se obtienen por defecto mediante la función `lm`.
- (e) e) Omita los posibles datos atípicos y vuelva a ajustar el modelo. Considere como datos atípicos a las observaciones sugeridas por los gráficos de diagnostico de la función `lm`.
- (f) ¿Cual de los dos modelos es mejor en términos de R^2 ajustada y el estadístico F ?
- (g) Después de analizar los resultados desarrolle sus conclusiones y sugerencias para un hipotético agricultor.

Ejercicio 3

En relación al ejercicio 2, comparar ambos modelos a través del criterio de Akaike. Utilice la función `AIC` y una implementación propia. ¿Es justa su comparación?