

Práctico 4

ANOVA

Modelos Lineales - 2024

EJERCICIO 1

Sea el modelo de análisis de varianza a una vía dado por:

$$y_{ij} = \mu + \tau_j + \epsilon_{ij} \quad \forall i = 1, \dots, n_j$$

Donde el vector y_{ij} es la j -ésima réplica (valor de la variable Y) observada en el grupo j . Como se vió en clase, la tanto la interpretación como la estimación de los parámetros cambia dependiendo de la restricción que implemente para estimar los parámetros del modelo.

Trabajando bajo la restricción $\tau_1 = 0$ en un experimento con $J = 3$ se observó que la matriz de diseño correspondiente a este modelo es:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix}$$

Teniendo en cuenta que $n_1 + n_2 + n_3 = n$ y llamando $S_y = \sum_{j=1}^3 \sum_i^{n_j} y_{ij}$ a la suma de todas las observaciones,

y $S_y^j = \sum_i^{n_j} y_{ij}$ a la suma de las observaciones del j -ésimo grupo, demuestre que:

$$\begin{aligned} \hat{\mu} &= \frac{S^1}{n_1} \\ \hat{\tau}_2 &= \frac{S^2}{n_2} - \frac{S^1}{n_1} \\ \hat{\tau}_3 &= \frac{S^3}{n_3} - \frac{S^1}{n_1} \end{aligned}$$

EJERCICIO 2


En la primera clase de análisis de varianza se mencionó que una posible restricción para estimar los parámetros del modelo a una vía consistía en:

$$\sum_j \tau_j = 0$$

Empleando los datos de la vida útil de los neumáticos (en el archivo *anova_1_via.xlsx* disponible en EVA) estime los parámetros del modelo bajo esta especificación y compare los resultados con los obtenidos bajo la restricción $\tau_1 = 0$.

A la hora de llevar a cabo la comparación tenga en cuenta:

-
- R^2
 - Prueba de significación global.
 - Interpretación de los parámetros.

Nota: A la hora de especificar esta restricción en  utilice alguna de las siguientes maneras:

- Bien puede incluir el argumento `contrasts = list(marca = 'contr.sum')` dentro de la llamada a la función `lm`.
- Bien puede especificarlo antes de ajustar el modelo con `contrasts(datos$marca) <- contr.sum(4)` donde el 4 hace referencia al número de categorías de la variable `marca`.
- Existen otras alternativas, si lo prefiere, siéntase libre de explorarlas.