Soluciones a la hoja de ejercicios 8

Ejercicio 1

Supongamos que las variables X, Y y Z indican las respuestas en los grupos. Cada una de estas variables sigue una función de densidad normal, con parámetros que dependen del modelo. La función de log-verosimilitud de es

$$l(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{i=1}^{4} \log f(x_i) + \sum_{j=1}^{4} \log f(y_i) + \sum_{k=1}^{4} \log f(z_i)$$

Los parámetros de la función de log-verosimilitud dependen del modelo. Para el modelo 1, eliminando todas las constantes innecesarias, la función es

$$\begin{aligned} l_1(\boldsymbol{\mu}, \boldsymbol{\sigma}) = & -4 \log \sigma - \frac{1}{2} \sum_{i=1}^{4} \frac{(x_i - \mu)^2}{\sigma^2} \\ & -4 \log \sigma - \frac{1}{2} \sum_{j=1}^{4} \frac{(y_i - \mu)^2}{\sigma^2} \\ & -4 \log \sigma - \frac{1}{2} \sum_{j=1}^{4} \frac{(z_i - \mu)^2}{\sigma^2} \end{aligned}$$

donde se han escrito en líneas distintas los términos procedentes de cada una de las tres muestras por mayor claridad.

La función de log-verosimilitud del modelo 2 es

$$\begin{aligned} l_2(\boldsymbol{\mu}, \boldsymbol{\sigma}) = & -4 \log \sigma - \frac{1}{2} \sum_{i=1}^{4} \frac{(x_i - \mu_1)^2}{\sigma^2} \\ & -4 \log \sigma - \frac{1}{2} \sum_{j=1}^{4} \frac{(y_i - \mu_2)^2}{\sigma^2} \\ & -4 \log \sigma - \frac{1}{2} \sum_{j=1}^{4} \frac{(z_i - \mu_3)^2}{\sigma^2} \end{aligned}$$

Para el modelo 3 la función es

$$\begin{aligned} l_3(\boldsymbol{\mu}, \boldsymbol{\sigma}) = & -4 \log \sigma_1 - \frac{1}{2} \sum_{i=1}^{4} \frac{(x_i - \mu)^2}{\sigma_1^2} \\ & -4 \log \sigma_2 - \frac{1}{2} \sum_{j=1}^{4} \frac{(y_i - \mu)^2}{\sigma_2^2} \\ & -4 \log \sigma_3 - \frac{1}{2} \sum_{j=1}^{4} \frac{(z_i - \mu)^2}{\sigma_3^2} \end{aligned}$$

Modelo 4:

$$
\begin{aligned}
l_4(\boldsymbol{\mu}, \boldsymbol{\sigma}) = & -4 \log \sigma_1 - \frac{1}{2} \sum_{i=1}^{4} \frac{(x_i - \mu_1)^2}{\sigma_1^2} \\
& -4 \log \sigma_2 - \frac{1}{2} \sum_{j=1}^{4} \frac{(y_i - \mu_2)^2}{\sigma_2^2} \\
& -4 \log \sigma_3 - \frac{1}{2} \sum_{j=1}^{4} \frac{(z_i - \mu_3)^2}{\sigma_3^2}
\end{aligned}
$$

Siguiendo el proceso habitual para encontrar los estimadores máximo-verosímiles, llegamos al siguiente resultado para el modelo 1:

$$
\begin{aligned}
\hat{\mu} &= \frac{\sum_{i=1}^4 x_i + \sum_{j=1}^4 y_j + \sum_{k=1}^4 z_k}{12} \\
\hat{\sigma}^2 &= \frac{\sum_{i=1}^4 (x_i - \mu)^2 + \sum_{j=1}^4 (y_j - \mu)^2 + \sum_{k=1}^4 (z_k - \mu)^2}{12}
\end{aligned}
$$

Para el modelo 2:

$$
\begin{aligned}
\hat{\mu}_1 &= \frac{\sum_{i=1}^4 x_i}{4}, \qquad \hat{\mu}_2 = \frac{\sum_{j=1}^4 y_j}{4}, \qquad \hat{\mu}_3 = \frac{\sum_{k=1}^4 z_k}{4} \\
\hat{\sigma}^2 &= \frac{\sum_{i=1}^4 (x_i - \mu_1)^2 + \sum_{j=1}^4 (y_j - \mu_2)^2 + \sum_{k=1}^4 (z_k - \mu_3)^2}{12}
\end{aligned}
$$

Modelo 3:

$$
\begin{aligned}
\hat{\mu} &= \frac{\sum_{i=1}^4 x_i + \sum_{j=1}^4 y_j + \sum_{k=1}^4 z_k}{12} \\
\hat{\sigma}_1^2 &= \frac{\sum_{i=1}^4 (x_i - \mu)^2}{4}, \qquad \hat{\sigma}_2^2 = \frac{\sum_{j=1}^4 (y_j - \mu)^2}{4}, \qquad \hat{\sigma}_3^2 = \frac{\sum_{k=1}^4 (z_k - \mu)^2}{4}
\end{aligned}
$$

Modelo 4:

$$
\begin{aligned}
\hat{\mu}_1 &= \frac{\sum_{i=1}^4 x_i}{4}, \qquad \hat{\mu}_2 = \frac{\sum_{j=1}^4 y_j}{4}, \qquad \hat{\mu}_3 = \frac{\sum_{k=1}^4 z_k}{4} \\
\hat{\sigma}_1^2 &= \frac{\sum_{i=1}^4 (x_i - \mu_1)^2}{4}, \qquad \hat{\sigma}_2^2 = \frac{\sum_{j=1}^4 (y_j - \mu_2)^2}{4}, \qquad \hat{\sigma}_3^2 = \frac{\sum_{k=1}^4 (z_k - \mu_3)^2}{4}
\end{aligned}
$$

Aplicando estas fórmulas, encontramos los siguientes estimadores:

- Modelo 1. $\hat{\mu} = 6$, $\hat{\sigma}^2 = 6,17$.
- Modelo 2. $\hat{\mu}_1 = 3$, $\hat{\mu}_2 = 7$, $\hat{\mu}_3 = 8$, $\hat{\sigma}^2 = 1,50$.
- Modelo 3. $\hat{\mu} = 6$, $\hat{\sigma}_1^2 = 9,50$, $\hat{\sigma}_2^2 = 4,50$, $\hat{\sigma}_3^2 = 4,50$.
- Modelo 4. $\hat{\mu}_1 = 3$, $\hat{\mu}_2 = 7$, $\hat{\mu}_3 = 8$, $\hat{\sigma}_1^2 = 0,50$, $\hat{\sigma}_2^2 = 3,50$, $\hat{\sigma}_3^2 = 0,50$.

El máximo de la función de verosimilitud y el AIC es

- Modelo 1. $l_1 = -27,94$, $\text{AIC}_1 = 59,88$.
- Modelo 2. $l_2 = -19,46$, $\text{AIC}_2 = 46,92$.
- Modelo 3. $l_3 = -27,55$, $\text{AIC}_3 = 63,09$.
- Modelo 4. $l_4 = -16,76$, $\text{AIC}_4 = 45,52$.

Comparamos los modelos por pares utilizando el G^2. El modelo más general es el 4, en el cual están anidados el 2 y el 3. Las comparaciones son

$$G^2_{24} = 5,40, \quad gl = 2, \quad p_{24} = 0,07$$
$$G^2_{34} = 21,57, \quad gl = 2, \quad p_{34} < 0,01$$

Por tanto, mantenemos $H_{0(24)}$ por lo que no podemos descartar que el modelo 2 ajuste igual que el 4. En cambio rechazamos la hipótesis $H_{0(34)}$ y concluimos que el modelo 3 ajusta peor que el 4.

Al comparar el modelo 1 con el 2, el resultado es

$$G^2_{12} = 16,96, \quad gl = 2, \quad p_{12} = 0,00$$

En definitiva, descartamos los modelos 1 y 3. Entre los otros dos, el G^2 permite mantener el modelo 2 frente al 4, aunque el AIC favorece al modelo 4.

En conclusión, la evidencia nos permite rechazar que las medias poblacionales de los tres grupos sean iguales. Con respecto a las varianzas, podemos mantener la hipótesis de homocedasticidad.