

Técnicas de Validación en Análisis Cluster

- **Criterio del investigador.**
- **Método gráfico:** eje OX n° de clusters, eje OY niveles de fusión.
- **Mójena:** $\bar{\alpha} + 1.25S_{\alpha}$ $\begin{cases} \alpha = \text{Niveles de fusión} \\ \bar{\alpha} = \text{Media de los valores de fusión.} \\ S_{\alpha} = \text{Cuasidesviación típica de } \alpha \end{cases}$
- **Beale:** Contraste basado en la distribución F de Snedecor.

$$H_1 : c_2 > c_1 \implies c_2 \text{ mejora } c_1$$

$$DC_1 = \frac{1}{n - c_1} \sum_{i=1}^{c_1} \sum_{j=1}^{n_i} \|x_i^j - m^j\|^2$$

$$DC_2 = \frac{1}{n - c_2} \sum_{i=1}^{c_2} \sum_{j=1}^{n_i} \|x_i^j - m^j\|^2$$

$$F(p(c_2 - c_1), p(n - c_2)) = \frac{\frac{DC_1 - DC_2}{DC_2}}{\left[\left(\frac{n - c_1}{n - c_2} \right) \left(\frac{c_2}{c_1} \right)^{\frac{2}{p}} - 1 \right]}$$

- **Métodos basados en procedimientos del análisis multivariante paramétrico.**

$$\text{Desviación total} \quad T = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_j^i - m)(x_j^i - m)'$$

$$\text{Desviación intra clusters} \quad W = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_j^i - m^i)(x_j^i - m^i)'$$

$$\text{Desviación entre clusters} \quad B = \sum_{i=1}^k n_i (m^i - m)(m^i - m)'$$

$$k = \text{n° de Clusters} \quad T = W + B$$

1. Minimización de la traza de W .

$$T = W + B \implies \text{tr}[T] = \text{tr}[W] + \text{tr}[B]$$

$$\text{Minimizar } \text{tr}[W] \iff \text{Maximizar } \text{tr}[B]$$

2. **Minimización de $k^2|W|$.**

3. **Minimización del determinante de W .**

$$\text{Criterio de Wilks en MANOVA} \quad \Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W + B|}$$

$$\text{Minimizar } \frac{|W|}{|T|} \iff \text{Maximizar } \frac{|T|}{|W|}$$

H_1 : medias poblacionales distintas

4. **Maximización de la traza de BW^{-1} .**

Criterio de Lawley y Hotelling en MANOVA

$$H_1 : \text{tr}[BW^{-1}] \text{Máxima}$$

5. **Criterio de Calinski y Harabasz**

$$C = \frac{\frac{\text{tr}[B]}{k-1}}{\frac{\text{tr}[W]}{n-k}} \implies \text{Maximizar } C$$