Universidad Politécnica de Madrid
ETS de Ingenieros Industriales

ESTUDIO DE SIMULACIÓN PARA UNA MEJOR COMPRENSIÓN DEL MANOVA FRENTE AL ANOVA

TRABAJO FIN DE GRADO

Grado en Ingeniería en Tecnologías Industriales

Tutor: José Manuel Mira McWilliams

Roberto Delgado Ferrezuelo

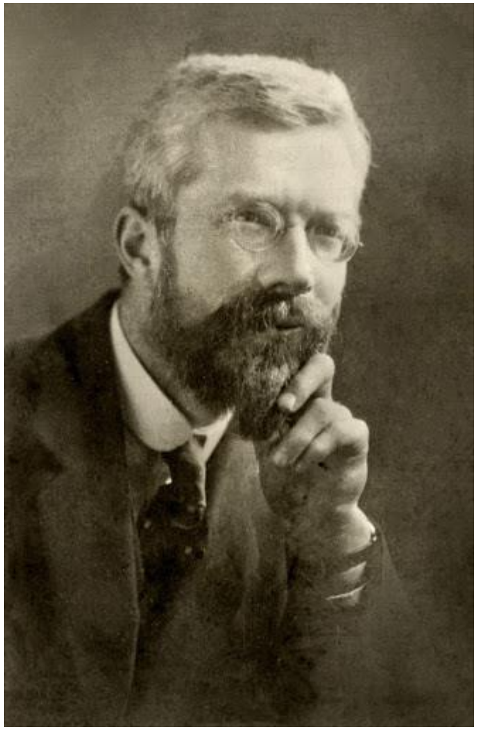
Febrero de 2022

ÍNDICE

Introducción y objetivos

Diseños experimentales

- Son una rama de la estadística aplicada
- Ronald Aylmer Fisher (1980−1962) es considerado como el pionero
  - Definición de la varianza y relación con la media poblacional
  - Uso generalizado de ANOVA a raíz de la publicación “Statistical methods for research workers”
- Aparecen los diseños factoriales en la publicación “The Design of Experiments”



Ronald Fisher. Adaptado de *R.A. FISHER*, por Indian Statistical Institute, Google Arts & Culture

Objetivos

- Adquirir un conocimiento más profundo de ANOVA y MANOVA mediante una serie de experimentos computacionales

Metodología

- Simulación estocástica haciendo uso de los métodos de Monte Carlo con RStudio
  - La simulación nos permite resolver las expresiones que subyacen el modelo de MANOVA de forma más sencilla
  - No se emplean datos reales, nos interesa entender la herramienta
  - Fijando los valores de los parámetros (α_i, β_j, $\alpha\beta_{ij}$, σ^2) obtenemos el p-valor y el valor de F, para medir el rendimiento de los análisis
  - Se comparan p ANOVAs con el MANOVA correspondiente

MANOVA vs ANOVA

- MANOVA es un análisis multivariante, ANOVA es univariante
- MANOVA es más completo que ANOVA, al tener en cuenta las relaciones que existen entre las variables que intervienen

ANOVA

$$y_{1ijk} = \mu_1 + \alpha_1 + \varepsilon_{1ijk}$$

$$y_{2ijk} = \mu_2 + \alpha_2 + \varepsilon_{2ijk}$$

p-valor del contraste

MANOVA

$$\begin{bmatrix} y_{1ijk} \\ y_{2ijk} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1ijk} \\ \varepsilon_{2ijk} \end{bmatrix}$$

H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_I$
H_1: algún α_i es distinto

Términos de error relacionados
$\rho \neq 0$

ANOVA

ANOVA de un factor

- Una sola variable independiente

- Modelo

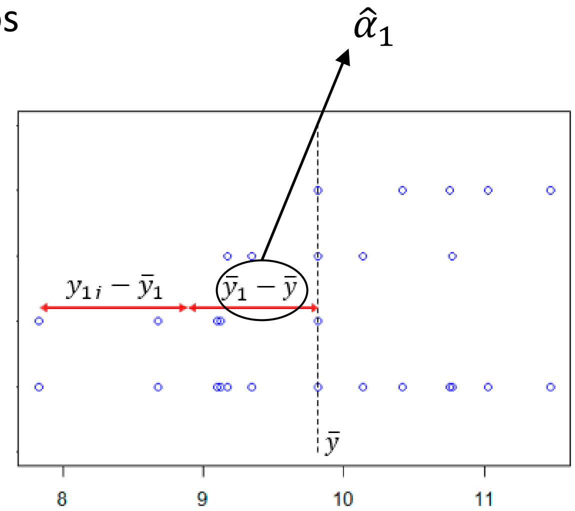
$$y_{ij} = \mu_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \to N(0, \sigma^2)$$

Para calcular los parámetros se utilizan sus estimadores máximos verosímiles:

$$\mu_i \to \bar{y}_{i\bullet} = \sum_{j=1}^{n_i} \frac{y_{ij}}{n_i}, \qquad \sigma^2 \to \hat{s}_R{}^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} e^2{}_{ij}}{n-k}$$

Estimación del término aleatorio:

$$\varepsilon_{ij} \to e_{ij} = y_{ij} - \bar{y}_{i\bullet}, \qquad \sum_{j=1}^{n_i} e_{ij} = 0, \qquad \forall i$$



ANOVA de un factor

- Contraste de hipótesis

$$H_0:\ \mu_1 = \mu_2 \cdots = \mu_k$$

$$H_1:\ alguna\ media\ es\ distinta\ de\ otra$$

- Descomposición de la variabilidad

$$y_{ij} - \bar{y}_{\bullet\bullet} = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet} + y_{ij} - \bar{y}_{i\bullet}, \quad \bar{y}_{\bullet\bullet} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{y_{ij}}{n_i}$$

Elevando al cuadrado y sumando para todas las observaciones se obtiene:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2$$

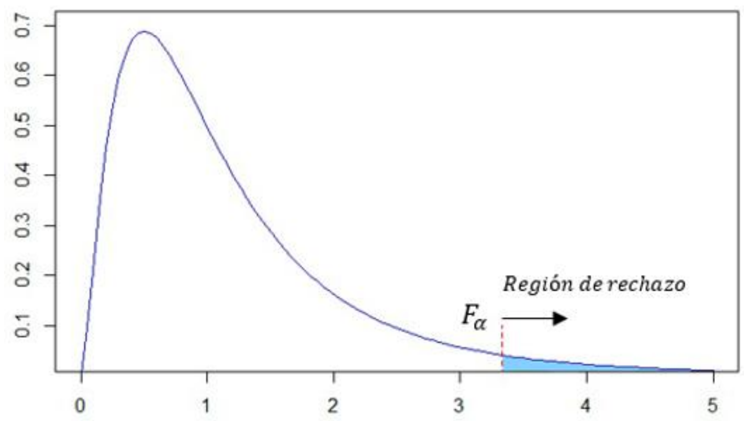
$$\boxed{\text{VT}} = \boxed{\text{VE}} + \boxed{\text{VNE}}$$

$$\hat{s}_T^2 = \frac{VT}{n-1} \qquad \hat{s}_E^2 = \frac{VE}{k-1} \qquad \hat{s}_R^2 = \frac{VNE}{n-k}$$

ANOVA de un factor

Si H_0 es cierta

$$F_0 = \frac{\hat{s}_E^2}{\hat{s}_R^2} \sim F_{k-1,n-k},\ P(F \geq F_0) = \text{p-valor} \rightarrow \text{Si p-valor} < \alpha\text{, se rechaza } H_0$$



Potencia del ANOVA

	H0 es cierta	H1 es cierta
Se acepta H0	$1-\alpha$	Error tipo β
Se acepta H1	Error tipo α	$1-\beta$

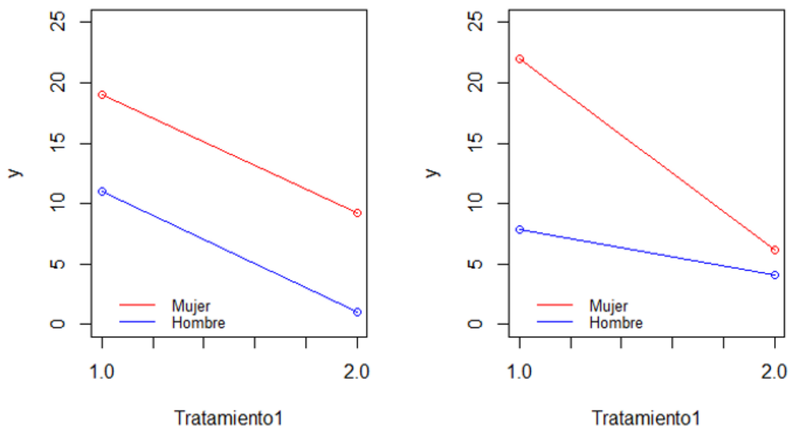
ANOVA de dos factores

- Dos variables independientes

- Modelo

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \to N(0, \sigma^2)$$

Interacción entre los factores



Representa la desviación de la media del nivel formado por la combinación de los dos factores con respecto a sus efectos principales

$$\alpha\beta_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$$

ANOVA de dos factores

Estimación de los parámetros

$$\mu \to \bar{y}_{\bullet\bullet\bullet} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}\frac{y_{ijk}}{n}$$

$$\alpha_i \to \hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \qquad \bar{y}_{i\bullet\bullet} = \sum_{j=1}^{J}\sum_{k=1}^{m}\frac{y_{ijk}}{mJ}$$

$$\beta_j \to \hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}, \qquad \bar{y}_{\bullet j\bullet} = \sum_{i=1}^{I}\sum_{k=1}^{m}\frac{y_{ijk}}{mI}$$

$$\alpha\beta_{ij} \to \widehat{\alpha\beta}_{ij} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}, \qquad \bar{y}_{ij\bullet} = \sum_{k=1}^{m}\frac{y_{ijk}}{n}$$

$$\sigma^2 \to \hat{s}^2{}_R = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m} e_{ijk}^2}{IJ(m-1)}$$

$$u_{ijk} \to e_{ijk} = y_{ijk} - (\bar{y}_{\bullet\bullet\bullet} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{\alpha\beta}_{ij}) = y_{ijk} - \bar{y}_{ij\bullet}$$

$$\boxed{\begin{gathered}\sum_{i=1}^{I}\hat{\alpha}_i = 0 \\ \sum_{j=1}^{J}\hat{\beta}_j = 0 \\ \sum_{i=1}^{I}\widehat{\alpha\beta}_{ij} = 0, \forall j, \sum_{j=1}^{J}\widehat{\alpha\beta}_{ij} = 0, \forall i \\ \sum_{k=1}^{m} e_{ijk} = 0, \qquad \forall i, j.\end{gathered}}$$

ANOVA de dos factores

Descomposición de la variabilidad

$$y_{ijk} - \bar{y}_{\bullet\bullet\bullet} = \hat{\alpha}_i + \hat{\beta}_j + \widehat{\alpha\beta}_{ij} + e_{ijk}$$

Elevando al cuadrado y sumando para todas las observaciones se obtiene:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2 = mJ\sum_{i=1}^{I}\hat{\alpha}_i^{\,2} + mI\sum_{j=1}^{J}\hat{\beta}_j^{\,2} + m\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\alpha\beta}_{ij}^{\,2} + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}e_{ijk}^{\,2}$$

$$\boxed{\text{VT}} = \boxed{\text{VE(F1)}} + \boxed{\text{VE(F2)}} + \boxed{\text{VE(F1XF2)}} + \boxed{\text{VNE}}$$

$$\hat{s}_T^2 = \frac{VT}{n-1} \qquad \hat{s}_{F1}^2 = \frac{VE(F1)}{I-1} \qquad \hat{s}_{F2}^2 = \frac{VE(F2)}{J-1} \qquad \hat{s}_{F1F2}^2 = \frac{VE(F1XF2)}{(I-1)(J-1)} \qquad \hat{s}_R^2 = \frac{VNE}{IJ(m-1)}$$

MANOVA

MANOVA de un factor con dos niveles

Test de la T^2 de Hotelling

$$T^2 = \left(\frac{n_1 \times n_2}{n_1 + n_2}\right) MD^2 \longrightarrow F_{p,(n_1+n_2-p-1)} = \frac{(n_1 + n_2 - p - 1)}{2(n_1 + n_2 - p)} T^2$$

Distancia de Mahalanobis entre los centroides de los dos niveles

$$MD_{12} = (x_1 - x_2)^T \Sigma^{-1} (x_1 - x_2)$$

$$x_1 - x_2 = \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{21} - \mu_{22} \\ \vdots \\ \mu_{p1} - \mu_{p2} \end{bmatrix}$$

Medida del **tamaño del efecto del factor**

- Eta-squared

$$\eta^2 = 1 - \Lambda = \frac{|SSCP_b|}{|SSCP_t|}$$

- Distancia de Mahalanobis
- Diferencia de medias

Σ: matriz de covarianzas, dimensión $p \times p$

MANOVA de un factor

- Modelo

$$y_{uij} = \mu_{ui} + \varepsilon_{uij}, \quad \varepsilon_{uij} \rightarrow N(0, \sigma_u^2),$$

$$u = 1,2 \ldots, p, \qquad i = 1,2, \ldots k, \qquad j = 1,2, \ldots, n_i$$

- Contraste de hipótesis

$$H_0: \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{1k} \\ \mu_{2k} \\ \vdots \\ \mu_{pk} \end{bmatrix}$$

H_1: algún vector de medias es distinto

- Descomposición de la variabilidad (simplificado al caso de $p = 2$)

$$\begin{bmatrix} y_{111} - \bar{y}_{1\bullet\bullet} & y_{211} - \bar{y}_{2\bullet\bullet} \\ y_{112} - \bar{y}_{1\bullet\bullet} & y_{212} - \bar{y}_{2\bullet\bullet} \\ \vdots & \vdots \\ y_{11n_1} - \bar{y}_{1\bullet\bullet} & y_{21n_1} - \bar{y}_{2\bullet\bullet} \\ y_{121} - \bar{y}_{1\bullet\bullet} & y_{221} - \bar{y}_{2\bullet\bullet} \\ \vdots & \vdots \\ y_{12n_2} - \bar{y}_{1\bullet\bullet} & y_{222} - \bar{y}_{2\bullet\bullet} \\ \vdots & \vdots \\ y_{1kn_k} - \bar{y}_{1\bullet\bullet} & y_{2kn_k} - \bar{y}_{2\bullet\bullet} \end{bmatrix} = \begin{bmatrix} \bar{y}_{11\bullet} - \bar{y}_{1\bullet\bullet} & \bar{y}_{21\bullet} - \bar{y}_{2\bullet\bullet} \\ \vdots & \vdots \\ \vdots & \vdots \\ \bar{y}_{12\bullet} - \bar{y}_{1\bullet\bullet} & \bar{y}_{22\bullet} - \bar{y}_{2\bullet\bullet} \\ \vdots & \vdots \\ \vdots & \vdots \\ \bar{y}_{13\bullet} - \bar{y}_{1\bullet\bullet} & \bar{y}_{23\bullet} - \bar{y}_{2\bullet\bullet} \\ \vdots & \vdots \\ \bar{y}_{1k\bullet} - \bar{y}_{1\bullet\bullet} & \bar{y}_{2k\bullet} - \bar{y}_{2\bullet\bullet} \end{bmatrix} + \begin{bmatrix} y_{111} - \bar{y}_{11\bullet} & y_{211} - \bar{y}_{21\bullet} \\ y_{112} - \bar{y}_{11\bullet} & y_{212} - \bar{y}_{21\bullet} \\ \vdots & \vdots \\ y_{11n_1} - \bar{y}_{12\bullet} & y_{21n_2} - \bar{y}_{22\bullet} \\ y_{121} - \bar{y}_{12\bullet} & y_{221} - \bar{y}_{22\bullet} \\ \vdots & \vdots \\ y_{12n_1} - \bar{y}_{12\bullet} & y_{22n_2} - \bar{y}_{22\bullet} \\ \vdots & \vdots \\ y_{1kn_1} - \bar{y}_{1k\bullet} & y_{21k} - \bar{y}_{2k\bullet} \end{bmatrix}$$

T (VT) = B (VE) + W (VNE)

MANOVA de un factor

Multiplicando por T^T a ambos lados de la expresión:

$$T^T \mathrm{T} = (\mathrm{B}+\mathrm{W})^T(\mathrm{B}+\mathrm{W}) = B^T\mathrm{B} + W^T W + \cancelto{0}{B^T W} + \cancelto{0}{W^T B}$$

$$\boxed{SSCP_T} = \boxed{SSCP_B} + \boxed{SSCP_W}$$

$$SSCP_W = \begin{bmatrix} \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{1ij}-\bar{y}_{1i\bullet})^2 & \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{1ij}-\bar{y}_{1i\bullet})(y_{2ij}-\bar{y}_{2i\bullet}) \\ \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{1ij}-\bar{y}_{1i\bullet})(y_{2ij}-\bar{y}_{2i\bullet}) & \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{2ij}-\bar{y}_{2i\bullet})^2 \end{bmatrix}$$

$$SSCP_B = \begin{bmatrix} \sum_{i=1}^{k} n_i(\bar{y}_{1i\bullet}-\bar{y}_{1\bullet\bullet})^2 & \sum_{i=1}^{k} n_i(\bar{y}_{1i\bullet}-\bar{y}_{1\bullet\bullet})(\bar{y}_{2i\bullet}-\bar{y}_{2\bullet\bullet}) \\ \sum_{i=1}^{k} n_i(\bar{y}_{1i\bullet}-\bar{y}_{1\bullet\bullet})(\bar{y}_{2i\bullet}-\bar{y}_{2\bullet\bullet}) & \sum_{i=1}^{k} n_i(\bar{y}_{2i\bullet}-\bar{y}_{2\bullet\bullet})^2 \end{bmatrix}$$

$$SSCP_T = \begin{bmatrix} \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{1ij}-\bar{y}_{1\bullet\bullet})^2 & \sum_{i=1}^{k}(y_{1ij}-\bar{y}_{1\bullet\bullet})(y_{2ij}-\bar{y}_{2\bullet\bullet}) \\ \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{1ij}-\bar{y}_{1\bullet\bullet})(y_{2ij}-\bar{y}_{2\bullet\bullet}) & \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{2ij}-\bar{y}_{2\bullet\bullet})^2 \end{bmatrix}$$

Suma de productos cruzados

Suma de cuadrados

MANOVA de un factor

Existen diferentes parametrizaciones para calcular el estadístico F a partir de las matrices anteriores:

- Lambda de Wilks

$$\Lambda = \frac{|SSCP_W|}{|SSCP_W + SSCP_B|}$$

$\Lambda = \prod_{j=1}^{p}(1-\theta_j),\quad \theta_j$: autovalor j de la matriz $SSCP_B(SSCP_W + SSCP_B)^{-1}$

$$F_{ph,ft-g} \sim \frac{(ft-g)(1-\Lambda^{\frac{1}{t}})}{ph\Lambda^{\frac{1}{t}}}$$

$$f = e - \frac{1}{2}(p-h+1), \qquad g = \frac{ph-2}{2}, \qquad t = \begin{cases} \sqrt{\frac{p^2h^2-4}{p^2+h^2-5}} & si\ p^2+h^2-5 > 0 \\ 1 & si\ p^2+h^2-5 \leq 0 \end{cases}$$

h: grados de libertad de la matriz $SSCP_B$, e: grados de libertad de la matriz $SSCP_W$

MANOVA de dos factores

- Modelo

$$y_{uijk} = \mu_u + \alpha_{ui} + \beta_{uj} + \alpha\beta_{uij} + \varepsilon_{uijk}, \qquad \varepsilon_{uijk} \to N(0, \sigma_u^2).$$

$$u = 1,2,\dots,p, \qquad i = 1,2\dots,I, \qquad j = 1,2,\dots,J, \qquad k = 1,2,\dots,m$$

- Descomposición de la variabilidad

$$y_{uijk} - \bar{y}_{u\bullet\bullet\bullet} = \hat{\alpha}_{ui} + \hat{\beta}_{uj} + \widehat{\alpha\beta}_{uij} + (y_{uijk} - \bar{y}_{uij\bullet})$$

$$\begin{bmatrix} y_{1111} - \bar{y}_{1\bullet\bullet\bullet} & y_{2111} - \bar{y}_{2\bullet\bullet\bullet} \\ \vdots & \vdots \\ y_{111m} - \bar{y}_{1\bullet\bullet\bullet} & y_{211m} - \bar{y}_{2\bullet\bullet\bullet} \\ \vdots & \vdots \\ y_{1IJ1} - \bar{y}_{1\bullet\bullet\bullet} & y_{2IJ1} - \bar{y}_{2\bullet\bullet\bullet} \\ \vdots & \vdots \\ y_{1IJm} - \bar{y}_{1\bullet\bullet\bullet} & y_{2IJm} - \bar{y}_{2\bullet\bullet\bullet} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{11} & \hat{\alpha}_{21} \\ \hat{\alpha}_{12} & \hat{\alpha}_{22} \\ \hat{\alpha}_{13} & \hat{\alpha}_{23} \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \hat{\alpha}_{1I} & \hat{\alpha}_{2I} \end{bmatrix} + \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{21} \\ \hat{\beta}_{12} & \hat{\beta}_{22} \\ \hat{\beta}_{13} & \hat{\beta}_{23} \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \hat{\beta}_{1J} & \hat{\beta}_{2J} \end{bmatrix} + \begin{bmatrix} \widehat{\alpha\beta}_{111} & \widehat{\alpha\beta}_{211} \\ \vdots & \vdots \\ \widehat{\alpha\beta}_{11J} & \widehat{\alpha\beta}_{21J} \\ \vdots & \vdots \\ \widehat{\alpha\beta}_{1I1} & \widehat{\alpha\beta}_{2I1} \\ \vdots & \vdots \\ \widehat{\alpha\beta}_{1IJ} & \widehat{\alpha\beta}_{2IJ} \end{bmatrix} + \begin{bmatrix} y_{1111} - \bar{y}_{111\bullet} & y_{2111} - \bar{y}_{211\bullet} \\ \vdots & \vdots \\ y_{111m} - \bar{y}_{111\bullet} & y_{211m} - \bar{y}_{211\bullet} \\ \vdots & \vdots \\ y_{1IJ1} - \bar{y}_{111\bullet} & y_{2IJ1} - \bar{y}_{211\bullet} \\ \vdots & \vdots \\ y_{1IJm} - \bar{y}_{111\bullet} & y_{2IJm} - \bar{y}_{211\bullet} \end{bmatrix}$$

T = B(F1) + B(F2) + B(F1XF2) + W

(VT) (VE(F1)) (VE(F2)) (VE(F1XF2)) (VNE)

MANOVA de dos factores

Multiplicando en ambos lados de la igualdad por T^T

$$T^T T = (B(F1) + B(F2) + B(F1XF2) + W)^T (B(F1) + B(F2) + B(F1XF2) + W)$$
$$= \mathrm{B(F1)}^T \mathrm{B(F1)} + \mathrm{B(F2)}^T \mathrm{B(F2)} + \mathrm{B(F1XF2)}^T \mathrm{B(F1XF2)} + W^T \mathrm{W}$$

$$\boxed{SSCP_T = SSCP_{F1} + SSCP_{F2} + SSCP_{F1F2} + SSCP_W}$$

$$SSCP_T = \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{1ijk} - \bar{y}_{1\bullet\bullet\bullet})^2 & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{1ijk} - \bar{y}_{1\bullet\bullet\bullet})(y_{2ijk} - \bar{y}_{2\bullet\bullet\bullet}) \\ \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{1ijk} - \bar{y}_{1\bullet\bullet\bullet})(y_{2ijk} - \bar{y}_{2\bullet\bullet\bullet}) & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{2ijk} - \bar{y}_{2\bullet\bullet\bullet})^2 \end{bmatrix}$$

$$SSCP_{F1} = \begin{bmatrix} mJ\sum_{i=1}^{I}(\bar{y}_{1i\bullet\bullet} - \bar{y}_{1\bullet\bullet\bullet})^2 & mJ\sum_{i=1}^{I}(\bar{y}_{1i\bullet\bullet} - \bar{y}_{1\bullet\bullet\bullet})(\bar{y}_{2i\bullet\bullet} - \bar{y}_{2\bullet\bullet\bullet}) \\ mJ\sum_{i=1}^{I}(\bar{y}_{1i\bullet\bullet} - \bar{y}_{1\bullet\bullet\bullet})(\bar{y}_{2i\bullet\bullet} - \bar{y}_{2\bullet\bullet\bullet}) & mJ\sum_{i=1}^{I}(\bar{y}_{2i\bullet\bullet} - \bar{y}_{2\bullet\bullet\bullet})^2 \end{bmatrix}$$

$$SSCP_{F2} = \begin{bmatrix} mI\sum_{j=1}^{J}(\bar{y}_{1\bullet j\bullet} - \bar{y}_{1\bullet\bullet\bullet})^2 & mI\sum_{j=1}^{J}(\bar{y}_{1\bullet j\bullet} - \bar{y}_{1\bullet\bullet\bullet})(\bar{y}_{2\bullet j\bullet} - \bar{y}_{2\bullet\bullet\bullet}) \\ mI\sum_{j=1}^{J}(\bar{y}_{1\bullet j\bullet} - \bar{y}_{1\bullet\bullet\bullet})(\bar{y}_{2\bullet j\bullet} - \bar{y}_{2\bullet\bullet\bullet}) & mI\sum_{j=1}^{J}(\bar{y}_{2\bullet j\bullet} - \bar{y}_{2\bullet\bullet\bullet})^2 \end{bmatrix}$$

MANOVA de dos factores

$$SSCP_{F1F2} = \begin{bmatrix} m\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\alpha\beta}_{1ij}^{\;2} & m\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\alpha\beta}_{1ij}\widehat{\alpha\beta}_{2ij} \\ m\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\alpha\beta}_{1ij}\widehat{\alpha\beta}_{2ij} & m\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\alpha\beta}_{2ij}^{\;2} \end{bmatrix}$$

$$SSCP_W = \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{1ijk}-\bar{y}_{1ij\bullet})^2 & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{1ijk}-\bar{y}_{1ij\bullet})(y_{2ijk}-\bar{y}_{2ij\bullet}) \\ \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{1ijk}-\bar{y}_{1ij\bullet})(y_{2ijk}-\bar{y}_{2ij\bullet}) & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{2ijk}-\bar{y}_{2ij\bullet})^2 \end{bmatrix}$$

MANOVA **NO** debe entenderse como varios ANOVAs independientes

- MANOVA permite un mayor control de la probabilidad de error tipo I
- En MANOVA se tiene en cuenta la relación de dependencia existente entre la combinación de variables respuesta recopiladas

EXPERIMENTOS COMPUTACIONALES

Cuadro resumen de los experimentos

Experimento 1	
Número de variables respuesta	2
Número de factores	1
Número de niveles	2
Parámetros	$n = 100, 1000, \quad \rho \in [-0.9, 0.9],$ $\sigma^2 = 100, \quad \Delta\mu_{ui}$ grande-grande, pequeño-pequeño, grande-pequeño
Experimento 2	
Número de variables respuesta	3
Número de factores	2
Número de niveles por factor	3
Parámetros	$n = 180, 90, 135, \quad \rho \in [-0.7, 0.7]$ $\sigma^2 = 100, \quad \Delta\mu_{ui}, \quad \Delta\mu_{uj}, \quad \Delta\mu_{uij}$

Primer experimento

- Se estudian tres casos dependiendo de los tamaños del efecto
  - Las dos variables dependientes con efecto del factor pequeño
  - Las dos variables dependientes con efecto del factor grande
  - Una variable con efecto del factor grande y la otra pequeño
- Se estudia un cuarto caso en el que se modifica el tamaño muestral
- Por simplificación, se consideran los valores poblacionales idénticos a los muestrales

Primera simulación

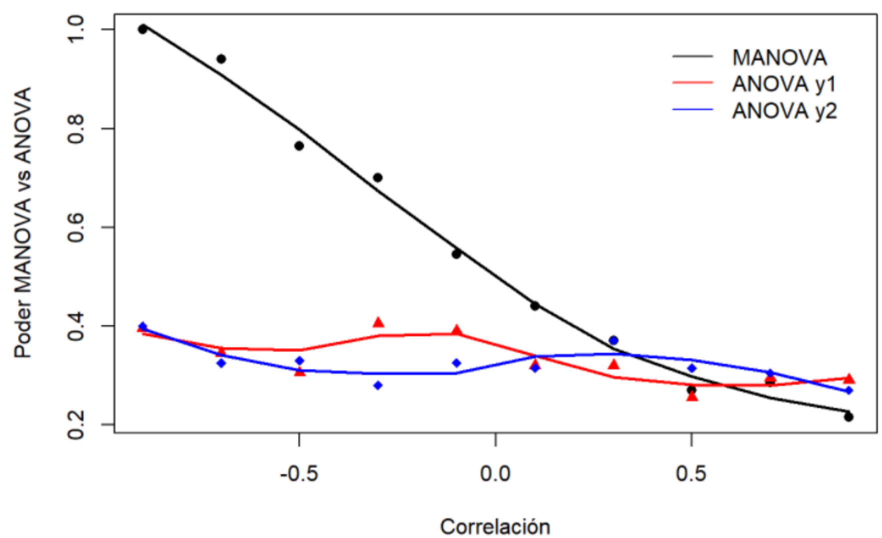
Se simula el modelo:

$$\Delta\mu_1 = \Delta\mu_2 = 2 \qquad \rho = cor(\varepsilon_{1ij}, \varepsilon_{2ij})$$

$$y_{uij} = \mu_{ui} + \varepsilon_{uij}, \qquad \varepsilon_{uij} \to N(0, \sigma_u^2),$$

$$u = 1,2\ldots,p, \qquad i = 1,2,\ldots k, \qquad j = 1,2,\ldots,n_i$$

		r= −0.9	0.5	r= 0.7	r= 0.9
ANOVAy1	F	2.114	16	2.041	1.978
	p	0.379	88	0.34	0.364
ANOVAy2	F	2.153	45	2.172	1.973
	p	0.337	49	0.335	0.367
MANOVA	F	11.751	15	1.701	1.461
	p	0.004	73	0.346	0.391



Aumento de la probabilidad de error tipo II

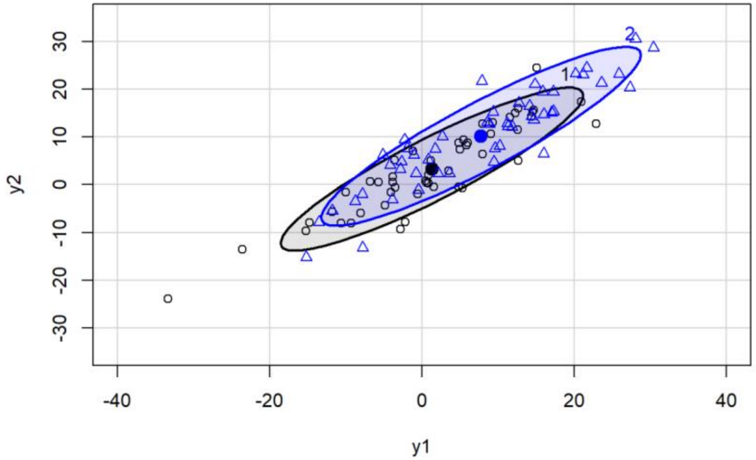
Primera simulación

$$MD^2 = [\Delta\bar{y}_1\ \Delta\bar{y}_2]S^{-1}\begin{bmatrix}\Delta\bar{y}_1\\ \Delta\bar{y}_2\end{bmatrix} = \Delta\bar{y}_1{}^2(s_{11}{}^{-1} + 2s_{12}{}^{-1} + s_{22}{}^{-1})$$

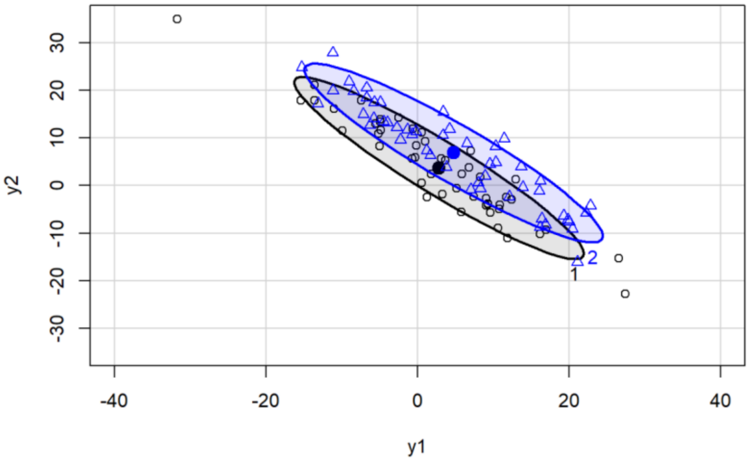
$$s_{11}{}^{-1} = {s_{11}^2}/{|S|} = s_{22}{}^{-1}, \qquad s_{12}{}^{-1} = {-s_{12}^2}/{|S|}, \qquad |S| = s_{11}{}^4 - s_{12}{}^4$$

$$\boxed{MD^2 = \frac{2\Delta\bar{y}_1{}^2}{s_{11}^2(1+r)}} \longrightarrow F = \frac{(n_1+n_2-p-1)}{(n_1+n_2-p)2}\left(\frac{n_1\times n_2}{n_1+n_2}\right)MD^2$$

$\rho = 0{,}9$



$\rho = -0{,}9$



Segunda simulación

Las dos variables dependientes con efecto del factor grande, $\Delta\mu_1 = \Delta\mu_2 = 6$

		r= −0.9	r= −0.7	r= −0.5	r= −0.3	r= −0.1	r= 0.1	r= 0.3	r= 0.5	r= 0.7	r= 0.9
ANOVA y1	F	10.64	10.69	9.279	10.6	9.369	9.923	9.949	10.93	10.89	10.37
	p	0.042	0.026	0.046	0.028	0.03	0.036	0.039	0.027	0.032	0.031
ANOVA y2	F	9.722	10.88	10.99	9.755	10.60	10.83	9.373	10.45	10.01	10.28
	p	0.038	0.033	0.023	0.043	0.025	0.039	0.038	0.033	0.033	0.031
MANOVA	F	90.82	33.3	19.63	14.09	11.01	9.608	7.67	7.52	6.543	5.84
	p	0	0	0	0.001	0.003	0.011	0.021	0.024	0.039	0.048

-0.5 0.0 0.5

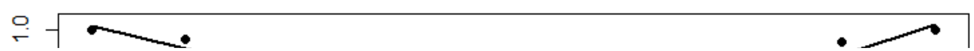
Correlación

Tercera simulación

Una variable con efecto del factor grande y la otra pequeño, $\Delta\mu_1 = 0, \;\; \Delta\mu_2 = 6$

$$MD^2 = {\Delta\bar{y}_1}^2 {s_{11}}^{-1} + 2\Delta\bar{y}_1\Delta\bar{y}_2 {s_{12}}^{-1} + {\Delta\bar{y}_2}^2 {s_{22}}^{-1}, \quad s_{11}^2 \approx s_{22}^2, \;\; \Delta\bar{y}_1 \approx 0$$

$$MD^2 = \frac{{\Delta\bar{y}_1}^2 + {\Delta\bar{y}_2}^2 - 2r\Delta\bar{y}_1\Delta\bar{y}_2}{s_{11}^2(1-r^2)} = \boxed{\frac{{\Delta\bar{y}_2}^2}{s_{11}^2(1-\mathrm{r}^2)}}$$



		r= −0.9	r= −0.7	r= −0.5	r= −0.3	r= −0.1	r= 0.1	r= 0.3	r= 0.5	r= 0.7	r= 0.9
ANOVA y1	F	0.922	1.009	1.053	0.836	0.999	1.11	1.34	1.231	1.038	1.113
	p	0.513	0.499	0.496	0.509	0.5	0.491	0.461	0.43	0.502	0.505
ANOVA y2	F	9.591	9.626	10.5	10.094	10.277	10.551	9.847	10.373	10.585	9.49
	p	0.04	0.051	0.029	0.033	0.029	0.043	0.031	0.036	0.034	0.041
MANOVA	F	24.533	9.49	7.083	5.921	5.7	5.877	6.243	7.069	10.607	25.555
	p	0	0.009	0.022	0.049	0.05	0.052	0.044	0.029	0.006	0

Correlación

Cuarta simulación

Aumento del tamaño muestral

- Aumenta la correlación entre las variables dependientes
- Mayor evidencia de diferencias entre los grupos formados en la variable independiente al tener una muestra más representativa → Aumento del tamaño del efecto

Se generan dos conjuntos de datos, ambos con:

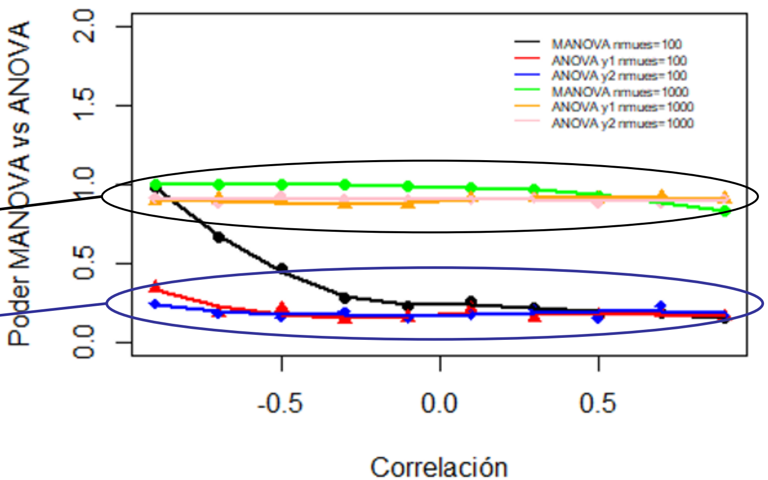
$$y_1: \quad \Delta\mu_1 = 2 \quad \sigma_1^2 = 100$$

$$y_2: \quad \Delta\mu_2 = 2 \quad \sigma_2^2 = 100$$

El resultado final es un aumento de la potencia con el tamaño muestral

n=1000

n=100



Segundo experimento

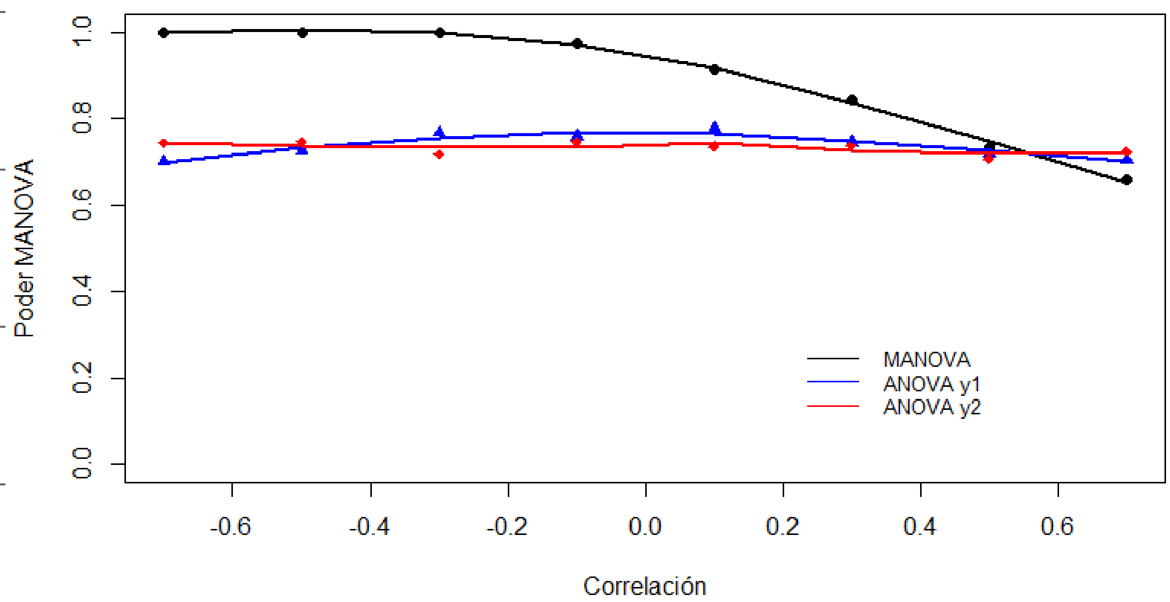
- Se estudian tres casos distintos, en cada uno se modifica la correlación entre dos de las variables de la misma forma que en el primer experimento, fijando un valor en la tercera variable dependiente con respecto a las otras dos

- Se estudia un cuarto caso en el que se modifica el tamaño muestral

- Se siguen considerando valores muestrales y poblacionales como idénticos

Primera simulación

Se varía la correlación entre las variables y_1 e y_2, ambas con efecto del factor grande



	.5	$r = 0.7$
ANOVA y1		5.463
		0.058
ANOVA y2		5.388
		0.056
ANOVA y3		1.006
		0.501
MANOVA		2.847
		0.073

Primera simulación

Expresiones analíticas

$$MD^2 = [\Delta\bar{y}_1 \quad \Delta\bar{y}_2 \quad \Delta\bar{y}_3] \begin{bmatrix} {s_{11}}^{-1} & {s_{12}}^{-1} & {s_{13}}^{-1} \\ {s_{21}}^{-1} & {s_{22}}^{-1} & {s_{23}}^{-1} \\ {s_{31}}^{-1} & {s_{32}}^{-1} & {s_{33}}^{-1} \end{bmatrix} \begin{bmatrix} \Delta\bar{y}_1 \\ \Delta\bar{y}_2 \\ \Delta\bar{y}_3 \end{bmatrix}$$

Simplificaciones:

$$\Delta\bar{y}_1 \approx \Delta\bar{y}_2 = \Delta\bar{y}, \quad \Delta\bar{y}_3 \approx 0, \quad s_{11}^2 \approx s_{22}^2 \approx s_{33}^2 = s^2, \quad r_{y1y3} \approx r_{y2y3} = r_{gp}$$

$$MD^2 = \Delta\bar{y}^2({s_{11}}^{-1} + {s_{12}}^{-1} + {s_{21}}^{-1} + {s_{22}}^{-1})$$

$${s_{11}}^{-1} = s^4(1 - {r^2}_{gp})/|S|,$$

$${s_{12}}^{-1} = s^4({r^2}_{gp} - r_{y1y2})/|S|,$$

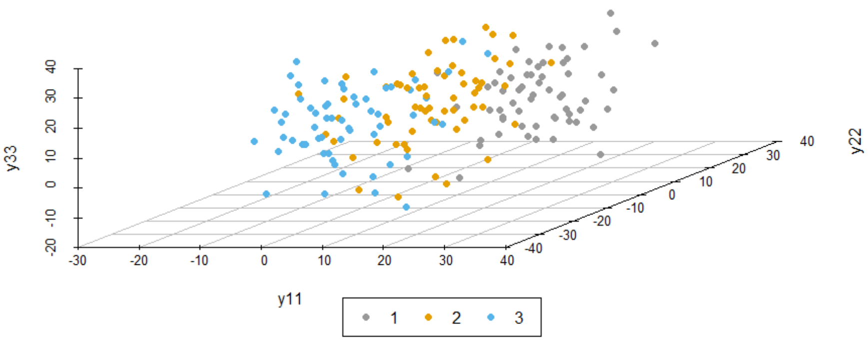
$${s_{22}}^{-1} = s^4(1 - {r^2}_{gp})/|S|,$$

$$|S| = s^6(1 - r_{y1y2})(1 + r_{y1y2} - 2{r^2}_{gp})$$

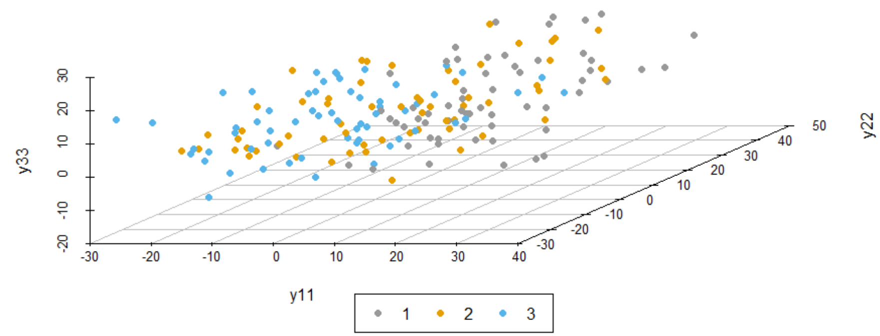
$$\boxed{MD^2 = \frac{2\Delta\bar{y}^2}{s^2\left(1 + r_{y1y2} - 2{r^2}_{gp}\right)}}$$

Primera simulación

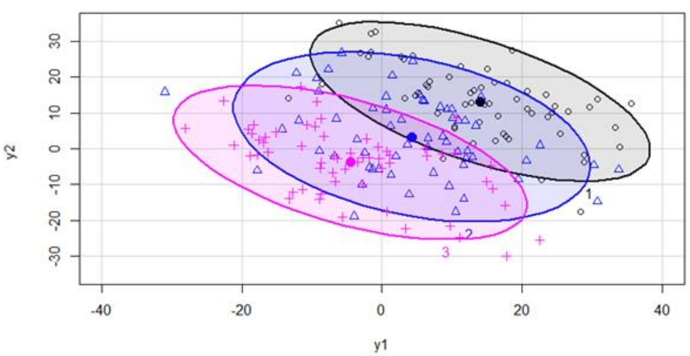
$\rho = -0{,}7$

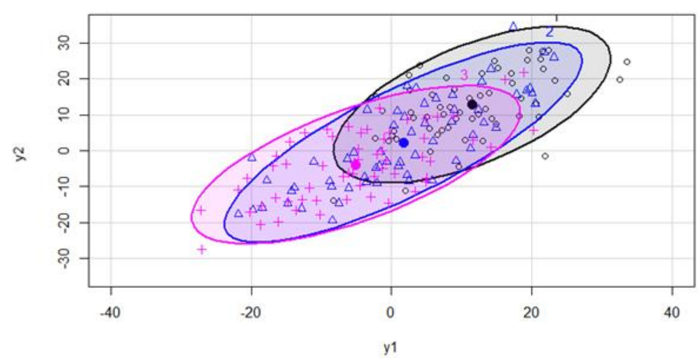


$\rho = 0{,}7$



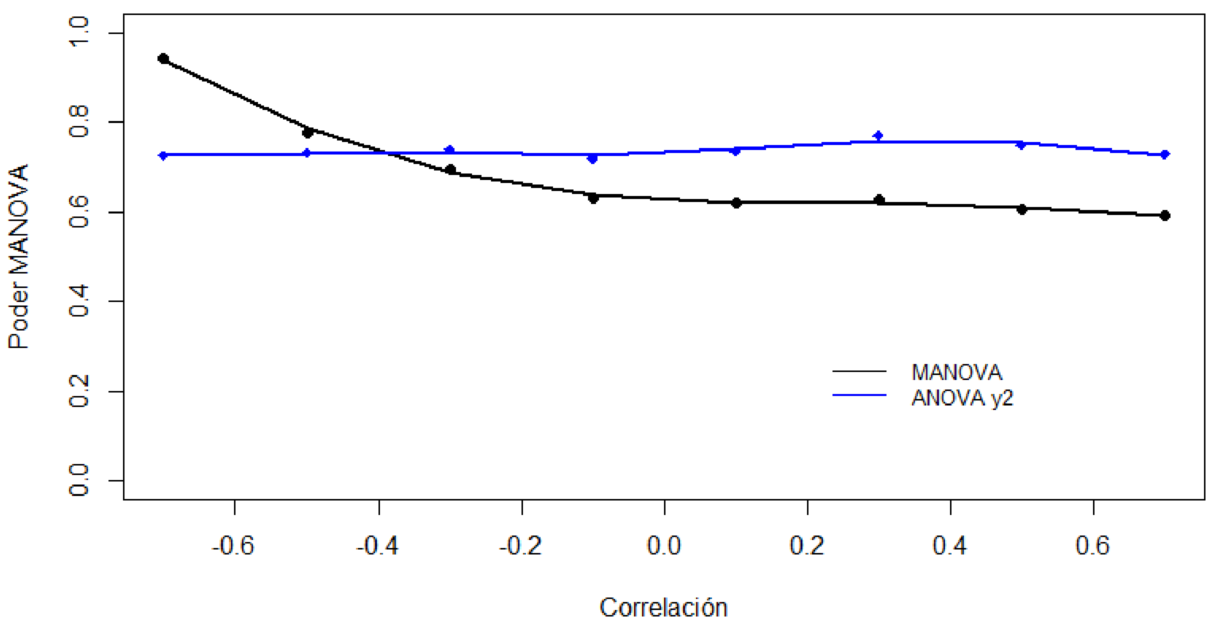
Proyección sobre los ejes de y_1 e y_2





Segunda simulación

Se varía la correlación entre las variables y_1 e y_3, ambas con efecto del factor pequeño



	$r = 0.7$
ANOVA y	0.942
	0.518
ANOVA y	5.096
	0.06
ANOVA y	0.94
	0.524
MANOV	2.464
	0.107

Segunda simulación

Expresiones analíticas:

- Simplificaciones

$$\Delta\bar{y}_1 \approx \Delta\bar{y}_3 \approx 0, \quad s_{11}^2 \approx s_{22}^2 \approx s_{33}^2 = s^2, \quad r_{y1y2} \approx r_{y2y3} = r_{gp}$$

$$MD^2 = \Delta\bar{y}^2 {s_{22}}^{-1}$$

$${s_{22}}^{-1} = s^4(1 - r_{y1y3}^2)/|S|$$

$$\boxed{MD^2 = \frac{\Delta\bar{y}^2(1 + r_{y1y3})}{s^2(1 + r_{y1y3} - 2{r_{gp}}^2)}}$$

$$MD^2 - MD^{2*} = \frac{\Delta\bar{y}^2(1 + r_{y1y3})}{s^2(1 + r_{y1y3} - 2{r_{gp}}^2)} - \frac{\Delta\bar{y}^2(1 + {r_{y1y3}}^*)}{s^2(1 + {r_{y1y3}}^* - 2{r_{gp}}^2)}$$

$$MD^2 - MD^{2*} = \frac{-2{r_{gp}}^2(r_{y1y3} - {r_{y1y3}}^*)}{s^2(1 + r_{y1y3} - 2{r_{gp}}^2)(1 + {r_{y1y3}}^* - 2{r_{gp}}^2)}$$

Si $r_{y1y3} < {r_{y1y3}}^*$

$\Downarrow$

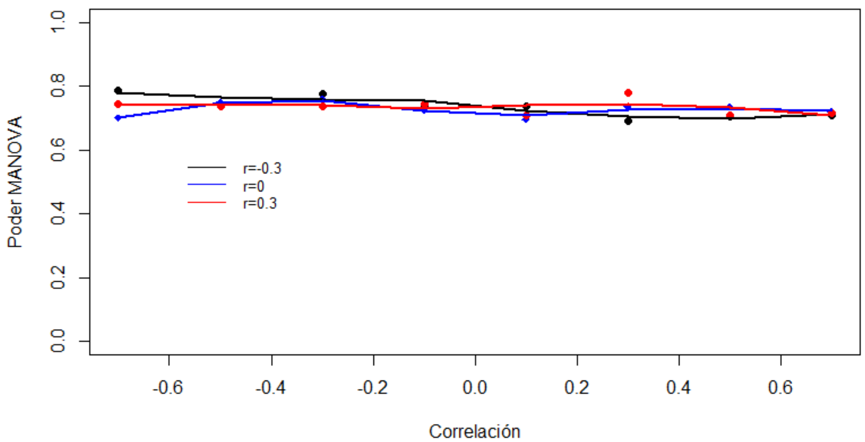
$MD^2 > MD^{2*}$

Tercera simulación

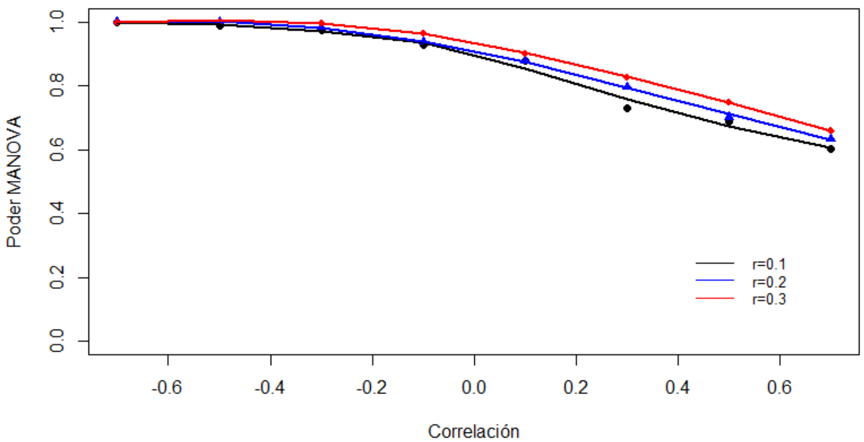
Cambio en la correlación de una variable con efecto del factor grande y la otra pequeño

$$MD^2 = \frac{\Delta\bar{y}^2\left(1 + r_{y1y3}\right)}{s^2\left(1 + r_{y1y3} - 2r_{gp}{}^2\right)} \qquad MD^2 = \frac{2\Delta\bar{y}^2}{s^2\left(1 + r_{y1y2} - 2r_{gp}{}^2\right)}$$

ANOVA



MANOVA

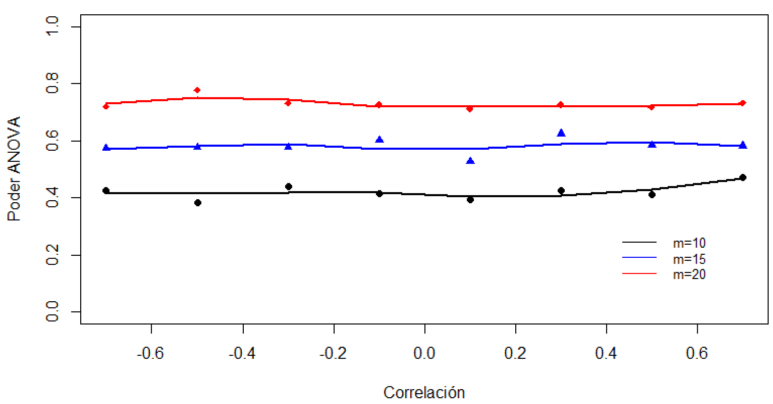


Cuarta simulación

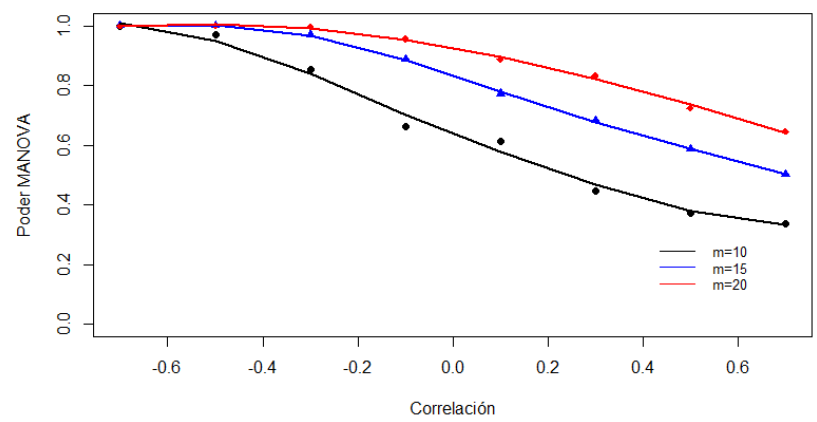
Tamaño muestral

- Se emplean los parámetros definidos en la segunda simulación, variando el número de observaciones por grupo, desde 10 hasta 20

ANOVA



MANOVA



CONCLUSIONES Y LÍNEAS FUTURAS

Conclusiones

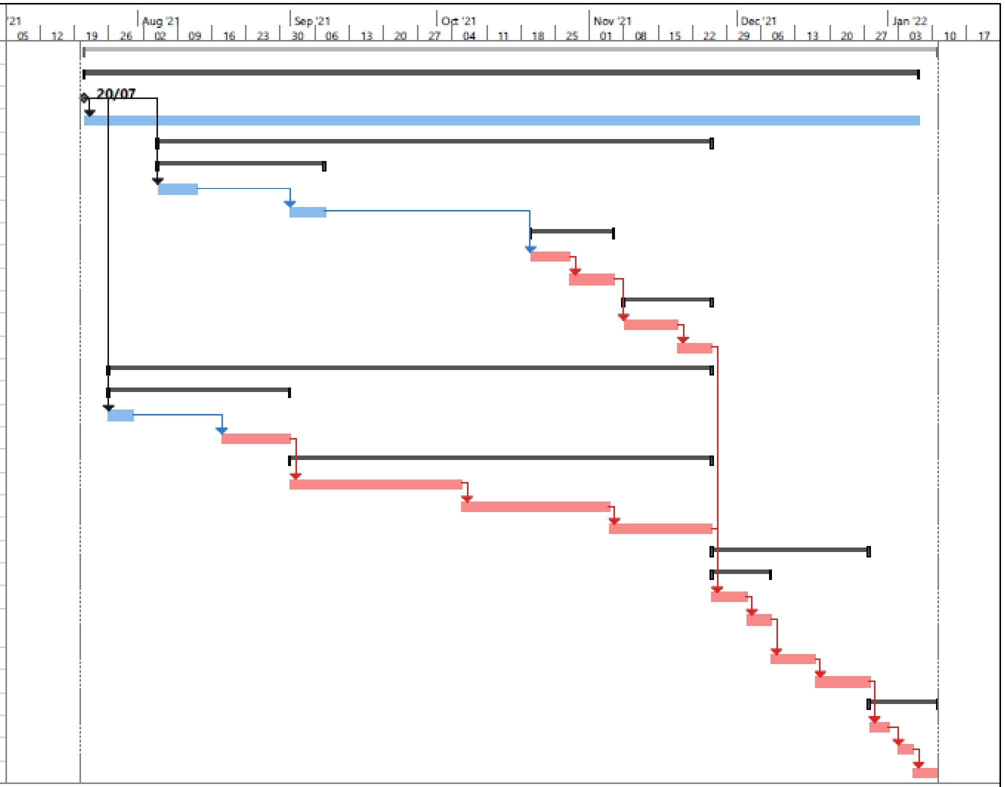
- Se ha comprobado que el MANOVA no es equivalente a varios ANOVAs independientes
- En experimentos en los que dos de las variables recopiladas presenten tamaño del efecto grande o pequeño, correlaciones negativas entre estas dos dan lugar a aumentos en la potencia del MANOVA
- Pude ser interesante para el investigador recopilar datos de variables relacionadas al fenómeno que se está estudiando, en las cuales no se esperan diferencias significativas entre grupos si están altamente correlacionadas con otras en las que si se observan claras diferencias.
- A mayor cantidad de datos recopilados mayor será la potencia del análisis

Líneas futuras

- Ampliar el número de variables dependientes
- Emplear otros tamaños del efecto
  - ¿A partir de qué valor de la diferencia de medias se considera un efecto del factor/es como “grande” o “pequeño”?
  - Se ha considerado nulo el tamaño del efecto de las variables con efecto pequeño
- Ampliar número de niveles formados en las variables independientes
- Desarrollo de nuevas expresiones analíticas empleando otros teoremas
  - Test de la intersección de la unión

PLANIFICACIÓN TEMPORAL Y PRESUPUESTO

ID	Task Name	Duration	Start	Finish
0	**Estudio de simulación**	**124 days**	**Tue 20/07/21**	**Mon 10/01/22**
1	**Gestión del proyecto**	**122 days**	**Tue 20/07/21**	**Thu 06/01/22**
2	Asignación tema	0 days	Tue 20/07/21	Tue 20/07/21
3	Tutorías	122 days	Wed 21/07/21	Thu 06/01/22
4	**Conocimientos RStudio**	**81 days**	**Thu 05/08/21**	**Thu 25/11/21**
5	**Repaso de conocimientos previos**	**24 days**	**Thu 05/08/21**	**Tue 07/09/21**
6	Repaso general	6 days	Thu 05/08/21	Thu 12/08/21
7	Instalación y pruebas con R Markdowr	5 days	Wed 01/09/21	Tue 07/09/21
8	**Sentencias ANOVA**	**13 days**	**Wed 20/10/21**	**Fri 05/11/21**
9	ANOVA de un factor	6 days	Wed 20/10/21	Wed 27/10/21
10	ANOVA de dos factores	7 days	Thu 28/10/21	Fri 05/11/21
11	**Sentencias MANOVA**	**14 days**	**Mon 08/11/21**	**Thu 25/11/21**
12	MANOVA de un factor	9 days	Mon 08/11/21	Thu 18/11/21
13	MANOVA de dos factores	5 days	Fri 19/11/21	Thu 25/11/21
14	**Fundamentos teóricos**	**89 days**	**Mon 26/07/21**	**Thu 25/11/21**
15	**ANOVA**	**27 days**	**Mon 26/07/21**	**Tue 31/08/21**
16	Repaso	5 days	Mon 26/07/21	Fri 30/07/21
17	Profundización en aspectos de ANOVA	10 days	Wed 18/08/21	Tue 31/08/21
18	**Estudio de MANOVA**	**62 days**	**Wed 01/09/21**	**Thu 25/11/21**
19	Lectura libro Sharma	25 days	Wed 01/09/21	Tue 05/10/21
20	Lectura libro Multivariate Data Analys	22 days	Wed 06/10/21	Thu 04/11/21
21	Lectura libro Mardia	15 days	Fri 05/11/21	Thu 25/11/21
22	**Simulaciones**	**22 days**	**Fri 26/11/21**	**Mon 27/12/21**
23	**Primeras pruebas de simulación**	**8 days**	**Fri 26/11/21**	**Tue 07/12/21**
24	Creación de tablas	5 days	Fri 26/11/21	Thu 02/12/21
25	Generación de números aleatorios y bucles	3 days	Fri 03/12/21	Tue 07/12/21
26	MANOVA de un factor	7 days	Wed 08/12/21	Thu 16/12/21
27	MANOVA de dos factores	7 days	Fri 17/12/21	Mon 27/12/21
28	**Análisis resultados**	**10 days**	**Tue 28/12/21**	**Mon 10/01/22**
29	Expresiones analíticas	4 days	Tue 28/12/21	Fri 31/12/21
30	Contraste con simulaciones	3 days	Mon 03/01/22	Wed 05/01/22
31	Conclusiones	3 days	Thu 06/01/22	Mon 10/01/22



Project: Estudio de simulación
Date: Sun 06/02/22

Task · Split · Milestone · Summary · Project Summary · Inactive Task · Inactive Milestone · Inactive Summary · Manual Task · Duration-only · Manual Summary Rollup · Manual Summary · Start-only · Finish-only · External Tasks · External Milestone · Deadline · Critical · Critical Split · Progress · Manual Progress

Page 1

Presupuesto

- Costes directos

  Se estiman los sueldos de las dos personas que intervienen en el proyecto (tutor y alumno)

- Costes indirectos

  Coste del consumo de electricidad del portátil en el que se han ejecutado las simulaciones, se estima a partir del precio medio de la luz en los meses correspondientes y del tiempo medio diario usado el ordenador

- Costes de equipos

  Amortización del portátil, plazo de 10 años, amortización lineal

Concepto	Coste (€)
Salarios	3.848
Suministros	9
Equipos	31
Total	3.888


THANK
YOU
GRACIAS
ARIGATO
SHUKURIA
DANKSCHEEN
TASHAKKUR ATU
YAQHANYELAY
SUKSAMA
EKHMET
GOZAIMASHITA
EFCHARISTO
JUSPAXAR
KOMAPSUMNIDA
MAAKE
GRAZIE
MEHRBANI
PALDIES
BOLZÏN
MERCI
BİYAN
SHUKRIA
TINGKI

Objetivos AGENDA 2030

Se han llevado a cabo estudios para evaluar la calidad de la educación superior, con objeto de mejorarla, haciendo uso de estas técnicas

⇓

“Analysis of Psychometric Properties of the Quality and Satisfaction Questionnaire Focused on Sustainability in Higher Education”

Evaluación de la eficacia de la concienciación a jóvenes de los problemas del cambio climático, tratar de cambiar los hábitos de consumo para frenarlo

⇓

“Environmental Education to Change the Consumption Model and Curb Climate Change”

ANOVA de un factor

Hipótesis del modelo

1. Normalidad

- Método gráfico

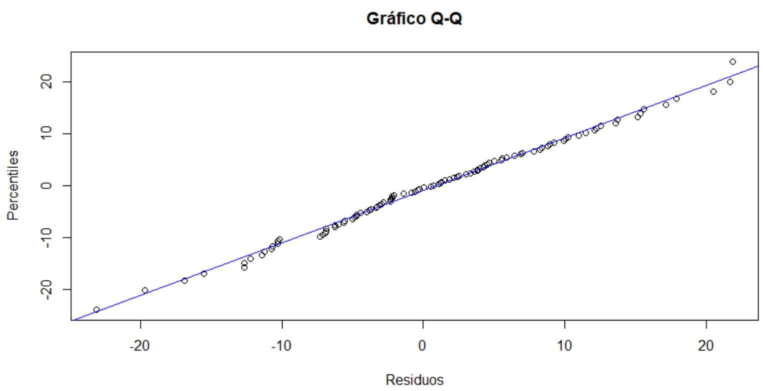
- Pruebas formales
  - Kolmogorov-Smirnov
  - Shapiro-Wilks

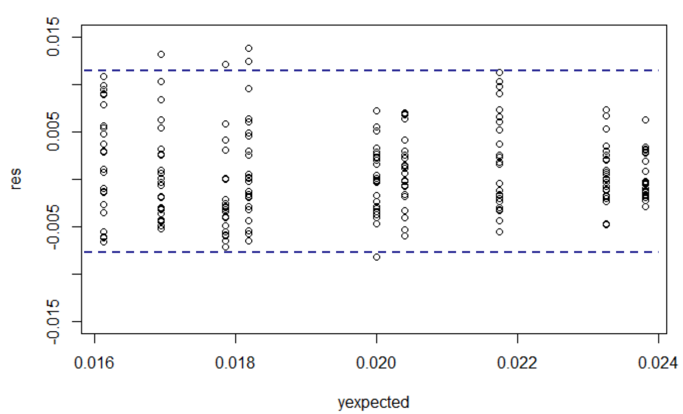
2. Homocedasticidad

- Método gráfico

- Pruebas formales
  - Bartlett
  - Brown-Forsythe
  - Levene

3. Independencia



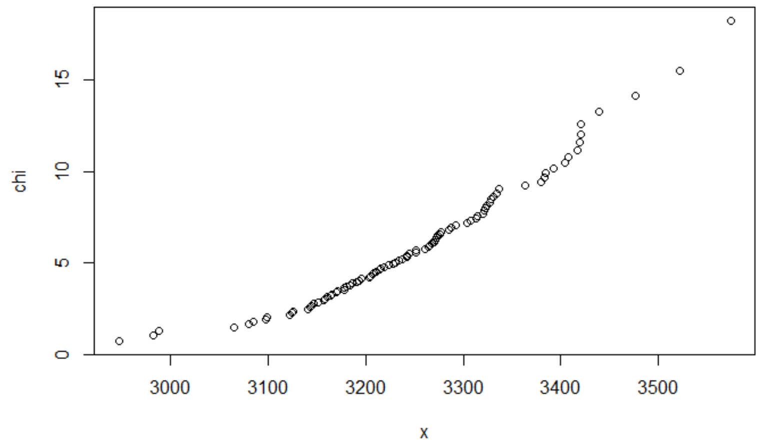


MANOVA de un factor

Hipótesis del modelo

1. Normalidad multivariante

- Método gráfico



2. Homocedasticidad

- Pruebas formales
  - M de Box

3. Independencia

Dos variables correspondientes a una misma observación pueden estar relacionadas entre sí, pero las que correspondan a distintas observaciones no

MANOVA de un factor

Existen diferentes parametrizaciones para calcular el estadístico F a partir de las matrices anteriores:

- Lambda de Wilks
- La traza de Hotelling
- La traza de Pillai
- La mayor raíz de Roy

¿Para qué se usan ANOVA y MANOVA?

Diferencias entre dos grupos de compañías a partir de datos financieros como el ROA y el ROIC

Grupo 1: Empresas más admiradas

Grupo 2: Empresas menos admiradas

$$\left.\begin{array}{l} \text{ROA} = y_1 \\ \text{ROIC} = y_2 \end{array}\right\} \text{Variables dependientes}$$

MANOVA (y_1, y_2 ~ Grupo) | ANOVA (y_1~ Grupo), ANOVA (y_2~ Grupo)

Introducción al análisis multivariante

Se parte del principio de la estimación de la máxima verosimilitud

$$\underbrace{X = [X_1, X_2, \dots, X_n]}_{\text{Variable aleatoria}} \qquad \underbrace{x = [x_1, x_2, \dots, x_n]}_{\text{Vector de observaciones}} \qquad \underbrace{\theta = [\theta_1, \theta_2, \dots, \theta_n]}_{\text{Parámetros de las distribuciones}}$$

$$\underbrace{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n}_{\text{Estimadores de los parámetros}} \rightarrow f(x_1, x_2, \dots, x_n; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n) = \max f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_n)$$

$$\left.\begin{array}{l} L_0 \rightarrow \text{Probabilidad bajo } H_0 \\ L_1 \rightarrow \text{Probabilidad bajo } H_1 \end{array}\right\} \begin{array}{l} \text{Si } \frac{L_0}{L_1} \gg 1 \text{ se acepta } H_0 \\ \text{Si } \frac{L_0}{L_1} \sim 0 \text{ se acepta } H_1 \end{array}$$

→ Test Lambda de Wilks