

MAT-269: Sesión 12

Test de hipótesis III

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Análisis multivariado usando la distribución t de Student

- ▶ Osorio, F., Galea, M., Arellano-Valle, R.
Using the multivariate t distribution for robust modelling in multivariate analysis.
En preparación.

Características del problema:

- ▶ El análisis multivariado ha sido desarrollado esencialmente bajo el normalidad.
- ▶ Evidencia empírica indica que en muchas situaciones el supuesto de normalidad es violado.
- ▶ En general, la inferencia estadística está basada en resultados asintóticos.

Aplicación:

- ▶ Test de hipótesis sobre la estructura de covarianza con aplicación a datos de AFP.
- ▶ Implementación computacional disponible en biblioteca MVT.



Aplicación:

Rentabilidades mensuales de AFPs:
Cuprum, Habitat, PlanVital y ProVida
entre Agosto/2005 y Diciembre/2013.

Base de datos con 101 observaciones y
4 variables (sólo datos del Fondo D).

Obs. 28, 30, 35, 38, 39, 42, 66, 73 y 75
son identificadas como outliers.

QQ-plots de distancias transformadas
revelan la presencia de colas pesadas.

Definición del problema:

Suponga x_1, \dots, x_n muestra aleatoria
desde $N_p(\mu, \Sigma)$, donde $\mu \in \mathbb{R}^p$ y $\Sigma \geq 0$
son parámetros desconocidos.

Nuestro interés es llevar a cabo test de
hipótesis del tipo

$$H_0 : G\theta = g,$$

con $\theta = (\mu^\top, \text{vech}^\top \Sigma)^\top$.

Substituyendo el supuesto de normalidad
por el de la distribución t de Student.



Estimación de parámetros:

Para x_1, \dots, x_n muestra aleatoria desde $N_p(\mu, \Sigma)$, es fácil notar que los MLE de μ y Σ están dados por:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^\top$$

Además \bar{x} y $S = \frac{n}{n-1} \hat{\Sigma}$ son estimadores consistentes para μ y Σ .

Test de hipótesis:

Hipótesis usuales en análisis multivariado son:

Test sobre el vector de medias:

$$H_0 : \mu = \mu_0, \Sigma \text{ desconocido}$$

esto lleva al estadístico T^2 de Hotelling.

Test sobre la matriz de covarianza:

$$H_0 : \Sigma = \Sigma_0, \mu \text{ desconocido}$$

cuyo estadístico de prueba es basado en el LRT.



Distancias de Mahalanobis: Datos de AFP

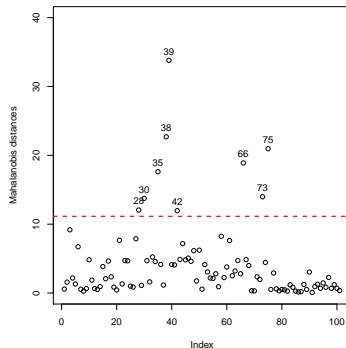
Para $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, tenemos

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2(p)$$

Permite identificar **observaciones atípicas (outliers)** usando, por ejemplo, algún **valor cuantil** desde la chi-cuadrado con p grados de libertad.

En particular, usamos el punto de corte (línea segmentada):

$$\chi_{0.975}^2(4) = 11.14329.$$



Distancias de Mahalanobis: Datos de AFP

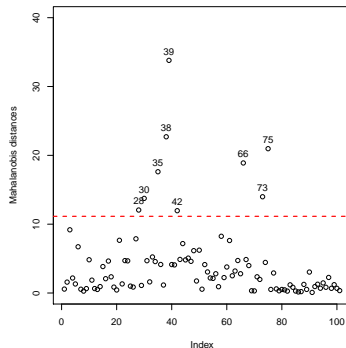
Para $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, tenemos

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2(p)$$

Permite identificar **observaciones atípicas (outliers)** usando, por ejemplo, algún **valor cuantil** desde la chi-cuadrado con p grados de libertad.

En particular, usamos el punto de corte (**línea segmentada**):

$$\chi_{0.975}^2(4) = 11.14329.$$



Coficiente de curtosis multivariado: Datos de AFP

Sea \mathbf{x} vector aleatorio con $E(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$. Mardia (1970) definió el **coeficiente de curtosis multivariado** como:

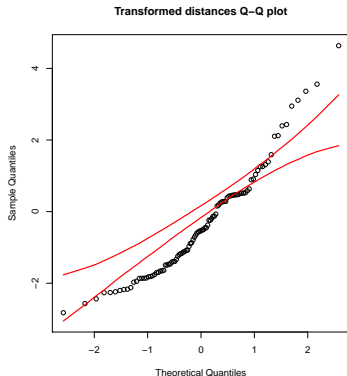
$$\beta_{2,p} = E[\{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^2],$$

bajo normalidad, tenemos $\beta_{2,p} = p(p+2)$.

Sea $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde \mathbf{x} . Un **estimador consistente** de $\beta_{2,p}$ es:

$$b_2 = \frac{1}{n} \sum_{i=1}^n \{(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})\}^2.$$

En nuestro caso tenemos $b_2 = 44.26018$.



Coficiente de curtosis multivariado: Datos de AFP

Sea \mathbf{x} vector aleatorio con $E(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$. Mardia (1970) definió el **coeficiente de curtosis multivariado** como:

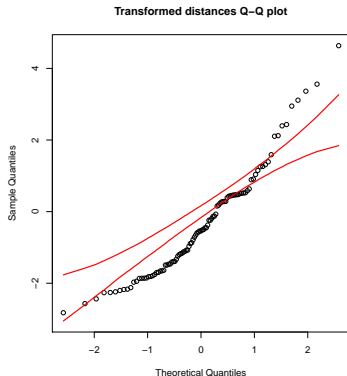
$$\beta_{2,p} = E[\{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^2],$$

bajo normalidad, tenemos $\beta_{2,p} = p(p+2)$.

Sea $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde \mathbf{x} . Un **estimador consistente** de $\beta_{2,p}$ es:

$$b_2 = \frac{1}{n} \sum_{i=1}^n \{(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})\}^2.$$

En nuestro caso tenemos $b_2 = 44.26018$.



Procedimientos para estimación robusta

- ▶ **M-estimadores:** Maronna (1976), Kent y Tyler (1991).
- ▶ **S-estimadores:** Davies (1987), Rousseeuw y Leroy (1987), Rousseeuw y van Zomeren (1990).
 - ▶ Elipsoide de volumen mínimo (MVE).
 - ▶ Covarianza de determinante mínimo (MCD).

Software: `cov.rob` (MASS), `CovMcd` (rrcov), `covRob` (robust).

- ▶ Estimación ML usando la **distribución t de Student**: Lange et al. (1989).
 - ▶ **Algoritmo EM**: Arslan et al. (1995), Liu y Rubin (1995), Liu (1997).

Software: `cov.trob` (MASS), `MVTMLE` (fastM), `heavyFit` (heavy).



Procedimientos para estimación robusta

- ▶ **M-estimadores:** Maronna (1976), Kent y Tyler (1991).
- ▶ **S-estimadores:** Davies (1987), Rousseeuw y Leroy (1987), Rousseeuw y van Zomeren (1990).
 - ▶ Elipsoide de volumen mínimo (MVE).
 - ▶ Covarianza de determinante mínimo (MCD).

Software: `cov.rob` (MASS), `CovMcd` (rrcov), `covRob` (robust).

- ▶ Estimación ML usando la **distribución t de Student**: Lange et al. (1989).
 - ▶ **Algoritmo EM:** Arslan et al. (1995), Liu y Rubin (1995), Liu (1997).

Software: `cov.trob` (MASS), `MVTMLE` (fastM), `heavyFit` (heavy).



- ▶ Versiones robustas del T^2 de Hotelling (1931): Anderson (1992), Srivastava y Mudholkar (2003) y Van Aelst y Willems (2013).

Software:

- ▶ bajo normalidad: `hotelling.test` (Hotelling), `HotellingT2` (ICSNP).
 - ▶ versión robusta: `FRBhotellingMM/S` (FRB), `center.test` (MVT).
-
- ▶ Hipótesis sobre la matriz de covarianza $H_0 : \Sigma = \Sigma_0$: Sutradhar (1993).

Software: `equicorrelation.test`, `homogeneity.test` (MVT).



- Versiones robustas del T^2 de Hotelling (1931): Anderson (1992), Srivastava y Mudholkar (2003) y Van Aelst y Willems (2013).

Software:

- bajo normalidad: `hotelling.test` (Hotelling), `HotellingT2` (ICSNP).
 - versión robusta: `FRBhotellingMM/S` (FRB), `center.test` (MVT).
- Hipótesis sobre la matriz de covarianza $H_0 : \Sigma = \Sigma_0$: Sutradhar (1993).

Software: `equicorrelation.test`, `homogeneity.test` (MVT).



Distribución t multivariada (Cornish, 1954)

El vector aleatorio $\mathbf{y} = (y_1, \dots, y_p)^\top$ tiene **distribución t de Student** si su densidad es de la forma:

$$f(\mathbf{y}) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{p/2}} |\mathbf{\Lambda}|^{-1/2} \left(1 + \frac{D^2}{\nu}\right)^{-\frac{\nu+p}{2}}, \quad \nu > 0,$$

con $D^2 = (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{\Lambda}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, y anotamos $\mathbf{y} \sim t_p(\boldsymbol{\mu}, \mathbf{\Lambda}, \nu)$.

Propiedades:

- ▶ Mezcla de escala normal: $\mathbf{y}|\tau \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma}/\tau)$ y $\tau \sim \text{Gama}(\frac{\nu}{2}, \frac{\nu}{2})$.
- ▶ $E(\mathbf{y}) = \boldsymbol{\mu}$, para $\nu > 1$, y $\text{Cov}(\mathbf{y}) = \frac{\nu}{\nu-2} \mathbf{\Lambda}$ ($= \mathbf{\Sigma}$), para $\nu > 2$.
- ▶ $D^2/p \sim F(p, \nu)$ (Box y Tiao, 1973).



Distribución t multivariada (Sutradhar, 1993)

El objetivo es **reparametrizar la densidad** de un vector aleatorio \mathbf{x} con **distribución t de Student**, tal que $E(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$.

Es decir, consideramos la siguiente **función de densidad**:

$$f(\mathbf{x}) = K_p(\eta) |\boldsymbol{\Sigma}|^{-1/2} (1 + c(\eta) D^2)^{-\frac{1}{2\eta}(1+\eta p)}, \quad \mathbf{x} \in \mathbb{R}^p, \quad (1)$$

donde

$$K_p(\eta) = \left(\frac{c(\eta)}{\pi} \right)^{p/2} \frac{\Gamma(\frac{1+\eta p}{2\eta})}{\Gamma(\frac{1}{2\eta})}, \quad 0 \leq \eta < \frac{1}{2},$$

con $c(\eta) = \eta/(1 - 2\eta)$ y $D^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$.

Cuando \mathbf{x} tiene la densidad dada en (1) anotamos $\mathbf{x} \sim T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)^1$.

¹ $\eta = 0$ corresponde a la distribución $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.



Propiedades de la distribución t multivariada

Sea $\mathbf{x} \sim T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$. Entonces,

(i) Tenemos la representación

$$\mathbf{x}|\omega \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\omega), \quad \omega \sim \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right)$$

(ii) Para $\omega \sim \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right)$, sigue que

$$E(\omega^{-r}) = \left(\frac{1}{2c(\eta)}\right)^r \frac{\Gamma(\frac{1}{2\eta} - r)}{\Gamma(\frac{1}{2\eta})}, \quad r < \frac{1}{\eta}.$$

(iii) Desde (i), sigue que $\omega|\mathbf{x} \sim \text{Gama}\left(\frac{1/\eta+p}{2}, \frac{1/c(\eta)+D^2}{2}\right)$.



(continuación)

(iv) $E(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$.

(v) Sea $D^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$. Entonces

$$F = \left(\frac{1}{1 - 2\eta} \right) \frac{D^2}{p} \sim F(p, 1/\eta).$$

(vi) El **coeficiente de curtosis multivariado** (Mardia, 1970) es dado por

$$\beta_{2,p} = E[\{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^2] = p(p+2)(\kappa + 1),$$

donde $\kappa = 2\eta/(1 - 4\eta)$ representa el **exceso de curtosis** ($\eta < \frac{1}{4}$).

Además la curtosis para la distribución $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ es **siempre mayor** que para la distribución normal (Yamaguchi, 2000).



(continuación)

(iv) $E(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$.

(v) Sea $D^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$. Entonces

$$F = \left(\frac{1}{1 - 2\eta} \right) \frac{D^2}{p} \sim F(p, 1/\eta).$$

(vi) El **coeficiente de curtosis multivariado** (Mardia, 1970) es dado por

$$\beta_{2,p} = E[\{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^2] = p(p+2)(\kappa + 1),$$

donde $\kappa = 2\eta/(1 - 4\eta)$ representa el **exceso de curtosis** ($\eta < \frac{1}{4}$).

Además la curtosis para la distribución $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ es **siempre mayor** que para la distribución normal (Yamaguchi, 2000).



Estimación usando el algoritmo EM

Para obtener los MLE de $\theta = (\mu^\top, \phi^\top, \eta)^\top$ con $\phi = \text{vech}^\top(\Sigma)$ usamos el **algoritmo EM** considerando la siguiente **representación jerárquica**:

$$x_i | \omega_i \stackrel{\text{ind}}{\sim} N_p(\mu, \Sigma / \omega_i), \quad \omega_i \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n,$$

En este caso, $\omega_i | x_i \sim \text{Gama}\left(\frac{1/\eta + p}{2}, \frac{1/c(\eta) + D_i^2}{2}\right)$. De este modo:

$$Q(\theta | \theta^{(k)}) = Q(\mu, \phi | \theta^{(k)}) + Q(\eta | \theta^{(k)}),$$

donde

$$Q(\mu, \phi | \theta^{(k)}) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu)^\top \Sigma^{-1} (x_i - \mu),$$

$$Q(\eta | \theta^{(k)}) = n \left\{ \frac{1}{2\eta} \log \left(\frac{1}{2c(\eta)} \right) - \log \Gamma \left(\frac{1}{2\eta} \right) + \frac{1}{2c(\eta)} \left[\psi \left(\frac{1/\eta^{(k)} + p}{2} \right) - \log \left(\frac{1/\eta^{(k)} + p}{2} \right) + \frac{1}{n} \sum_{i=1}^n (\log \omega_i^{(k)} - \omega_i^{(k)}) \right] \right\},$$

(2)



Estimación usando el algoritmo EM

Para obtener los MLE de $\theta = (\mu^\top, \phi^\top, \eta)^\top$ con $\phi = \text{vech}^\top(\Sigma)$ usamos el algoritmo EM considerando la siguiente representación jerárquica:

$$x_i | \omega_i \stackrel{\text{ind}}{\sim} N_p(\mu, \Sigma / \omega_i), \quad \omega_i \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n,$$

En este caso, $\omega_i | x_i \sim \text{Gama}\left(\frac{1/\eta + p}{2}, \frac{1/c(\eta) + D_i^2}{2}\right)$. De este modo:

$$Q(\theta | \theta^{(k)}) = Q(\mu, \phi | \theta^{(k)}) + Q(\eta | \theta^{(k)}),$$

donde

$$Q(\mu, \phi | \theta^{(k)}) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu)^\top \Sigma^{-1} (x_i - \mu),$$

$$Q(\eta | \theta^{(k)}) = n \left\{ \frac{1}{2\eta} \log \left(\frac{1}{2c(\eta)} \right) - \log \Gamma \left(\frac{1}{2\eta} \right) + \frac{1}{2c(\eta)} \left[\psi \left(\frac{1/\eta^{(k)} + p}{2} \right) - \log \left(\frac{1/\eta^{(k)} + p}{2} \right) + \frac{1}{n} \sum_{i=1}^n (\log \omega_i^{(k)} - \omega_i^{(k)}) \right] \right\},$$

(2)



Estimación usando el algoritmo EM

Para obtener los MLE de $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top$ con $\boldsymbol{\phi} = \text{vech}^\top(\boldsymbol{\Sigma})$ usamos el algoritmo EM considerando la siguiente representación jerárquica:

$$\mathbf{x}_i | \omega_i \stackrel{\text{ind}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\omega_i), \quad \omega_i \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n,$$

En este caso, $\omega_i | \mathbf{x}_i \sim \text{Gama}\left(\frac{1/\eta + p}{2}, \frac{1/c(\eta) + D_i^2}{2}\right)$. De este modo:

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)}) = Q(\boldsymbol{\mu}, \boldsymbol{\phi} | \boldsymbol{\theta}^{(k)}) + Q(\eta | \boldsymbol{\theta}^{(k)}),$$

donde

$$Q(\boldsymbol{\mu}, \boldsymbol{\phi} | \boldsymbol{\theta}^{(k)}) = -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n \omega_i^{(k)} (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}),$$

$$Q(\eta | \boldsymbol{\theta}^{(k)}) = n \left\{ \frac{1}{2\eta} \log \left(\frac{1}{2c(\eta)} \right) - \log \Gamma \left(\frac{1}{2\eta} \right) + \frac{1}{2c(\eta)} \left[\psi \left(\frac{1/\eta^{(k)} + p}{2} \right) - \log \left(\frac{1/\eta^{(k)} + p}{2} \right) + \frac{1}{n} \sum_{i=1}^n (\log \omega_i^{(k)} - \omega_i^{(k)}) \right] \right\},$$

(2)



Estimación usando el algoritmo EM

La **etapa-M del algoritmo**, con relación a $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ asume la forma:

$$\boldsymbol{\mu}^{(k+1)} = \frac{1}{\sum_{j=1}^n \omega_j^{(k)}} \sum_{i=1}^n \omega_i^{(k)} \mathbf{x}_i$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} (\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})(\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})^\top,$$

Actualizamos $\eta^{(k+1)}$ optimizando $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ con relación a η . Es decir,

$$\eta^{(k+1)} = \arg \max_{\eta} Q(\eta|\boldsymbol{\theta}^{(k)}),$$

La **función de pesos** adopta la forma:

$$\omega_i^{(k)} = \frac{1/\eta^{(k)} + p}{1/c(\eta^{(k)}) + D_i^2(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)})}.$$



Una “curiosa propiedad” de la distribución t

- ▶ Basados en Kent et al. (1994), sigue desde **condición de primer orden** para Σ que:

$$\frac{1}{n} \sum_{i=1}^n \hat{\omega}_i \hat{D}_i^2 = p$$

que es **válido para ambas parametrizaciones** de la t de Student.

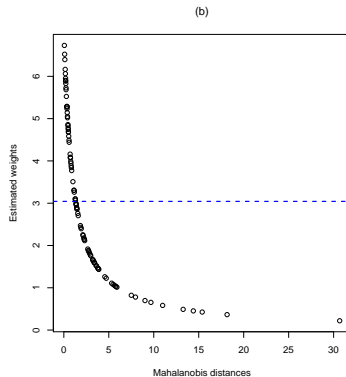
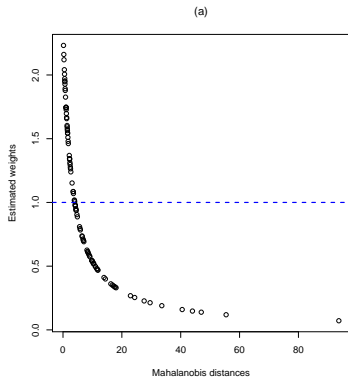
- ▶ Desde donde obtenemos la **“curiosa” relación**:

$$\frac{1}{n} \sum_{i=1}^n \hat{\omega}_i = \frac{1}{1 - 2\eta}.$$

- ▶ La propiedad anterior tiene la siguiente utilidad:
 - ▶ Permite **“acelerar” la convergencia** del algoritmo EM.
 - ▶ ¿Cuando **declarar pequeño** a un peso ω_i estimado?



“Curiosa propiedad” de los pesos estimados: Datos de AFP



Parametrización: (a) original y (b) de Sutradhar (1993).



Algoritmo PX-EM (η conocido)

Una **versión PX** (parameter-expanded) **del algoritmo EM** se basa en el siguiente modelo de datos incompletos:

$$\mathbf{X}_i | \omega_i \stackrel{\text{ind}}{\sim} \mathcal{N}_p(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_* / \omega_i), \quad \frac{\omega_i}{\alpha} \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n.$$

En este caso $\boldsymbol{\Theta} = (\boldsymbol{\mu}_*^\top, \text{vech}^\top(\boldsymbol{\Sigma}_*), \alpha)^\top$ y existe una **función de reducción**, tal que $\boldsymbol{\theta} = R(\boldsymbol{\Theta}) = (\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_* / \alpha)$.

Así $\boldsymbol{\mu}^{(k+1)}$ debe ser actualizado tal como se definió en el algoritmo EM original, mientras que

$$\boldsymbol{\Sigma}^{(k+1)} = \left(\frac{1}{1-2\eta}\right) \frac{1}{\sum_{j=1}^n \omega_j^{(k)}} \sum_{i=1}^n \omega_i^{(k)} (\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})(\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})^\top.$$



Desde la propiedad (v) tenemos que:

$$q = 1 + c(\eta)D^2 \stackrel{d}{=} 1 + p\eta F(p, 1/\eta).$$

De este modo, los pesos definidos por la distribución T de Student,

$$\omega = c(\eta) \frac{1/\eta + p}{q},$$

tienen la función de densidad:

$$f(\omega) = \frac{m^{1-(1/\eta+p)/2}}{B\left(\frac{1}{2\eta}, \frac{p}{2}\right)} \omega^{1/(2\eta)-1} (m-\omega)^{p/2-1}, \quad 0 < \omega < m,$$

con $m = (1 + p\eta)/(1 - 2\eta)$, $0 \leq \eta < \frac{1}{2}$ y $p > 0$.



Es fácil notar que

$$E(\omega) = \frac{1}{1 - 2\eta}, \quad \text{var}(\omega) = c^2(\eta) \left\{ \frac{2p}{1 + (p+2)\eta} \right\}.$$

Mientras que la cdf es dada por:

$$F_{\omega}(x) = I_{x/m} \left(\frac{1}{2\eta}, \frac{p}{2} \right),$$

donde $I_z(x, y)$ representa la función beta incompleta regularizada.

Observación: esto permite juzgar si un **peso estimado es pequeño** comparándolo con algún valor cuantil, p.ej. $F_{\omega}^{-1}(0.025)$.



Una manera de hacer **comparables los pesos estimados** de distintos modelos estadísticos $T_p(\mu, \Sigma, \eta)$, es definir los **pesos escalados**:

$$u = (1 + c(\eta)D^2)^{-1} \sim \text{Beta}\left(\frac{1}{2\eta}, \frac{p}{2}\right).$$

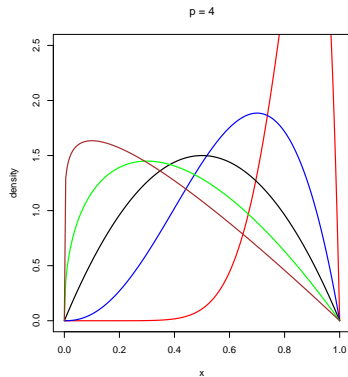
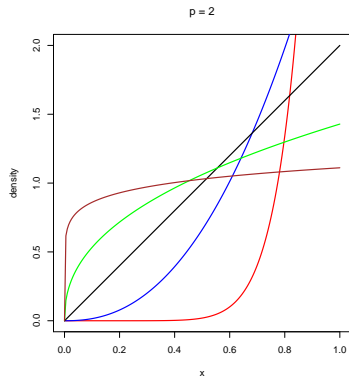
En cuyo caso tenemos

$$\frac{1}{n} \sum_{i=1}^n \hat{u}_i = \frac{1}{1 + \eta p} \quad (= E(u)).$$

Observación: Etapa-M del algoritmo EM se modifica de forma sencilla para introducir los pesos escalados (**versión PX-EM es idéntica** a la delineada anteriormente).



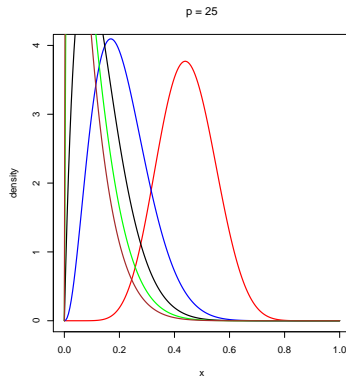
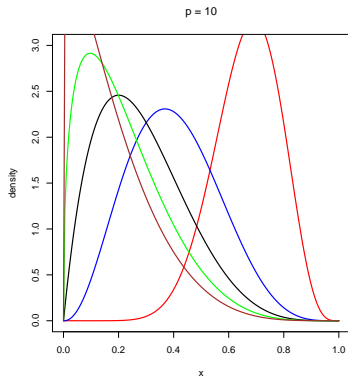
Densidad de los pesos escalados



Función de densidad para $\eta = 0.05$ (rojo), 0.15 (azul), 0.25 (negro), 0.35 (verde) y 0.45 (café).



Densidad de los pesos escalados



Función de densidad para $\eta = 0.05$ (rojo), 0.15 (azul), 0.25 (negro), 0.35 (verde) y 0.45 (café).



Evaluando la calidad del ajuste

Desde la propiedad (v), tenemos

$$F = \frac{1}{1 - 2\eta} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / p \sim F(p, 1/\eta).$$

Usamos la **transformación Wilson-Hilferty** (1931) para construir **QQ-plots con “envelopes”**,

$$Z = \frac{\left(1 - \frac{2\eta}{9}\right) F^{1/3} - \left(1 - \frac{2}{9p}\right)}{\left\{\frac{2\eta}{9} F^{2/3} + \frac{2}{9p}\right\}^{1/2}} \dot{\sim} N(0, 1).$$

Para $\eta = 0$, obtenemos

$$Z = \frac{F^{1/3} - \left(1 - \frac{2}{9p}\right)}{\sqrt{\frac{2}{9p}}} \dot{\sim} N(0, 1),$$

que corresponde a la transformación Wilson-Hilferty para **variables chi-cuadrado**.



Evaluando la calidad del ajuste

Desde la propiedad (v), tenemos

$$F = \frac{1}{1 - 2\eta} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / p \sim F(p, 1/\eta).$$

Usamos la **transformación Wilson-Hilferty** (1931) para construir **QQ-plots con “envelopes”**,

$$Z = \frac{\left(1 - \frac{2\eta}{9}\right) F^{1/3} - \left(1 - \frac{2}{9p}\right)}{\left\{\frac{2\eta}{9} F^{2/3} + \frac{2}{9p}\right\}^{1/2}} \dot{\sim} N(0, 1).$$

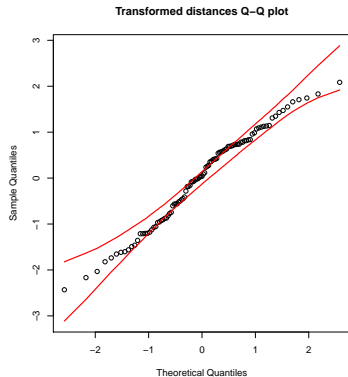
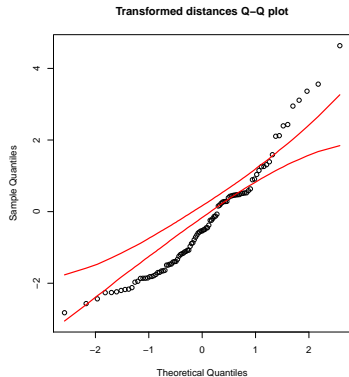
Para $\eta = 0$, obtenemos

$$Z = \frac{F^{1/3} - \left(1 - \frac{2}{9p}\right)}{\sqrt{\frac{2}{9p}}} \dot{\sim} N(0, 1),$$

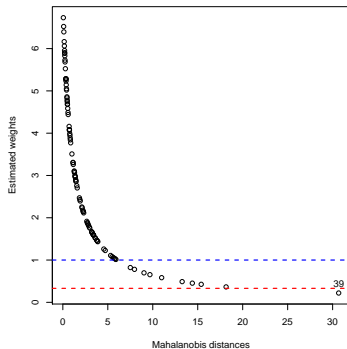
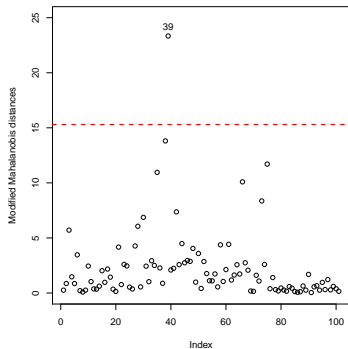
que corresponde a la transformación Wilson-Hilferty para **variables chi-cuadrado**.



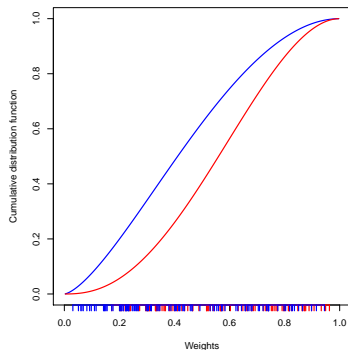
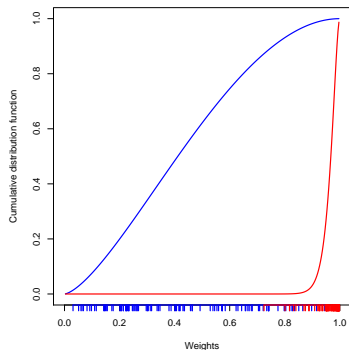
QQ-plots con 'envelopes': Datos de AFP



Distancias de Mahalanobis y pesos estimados: Datos de AFP



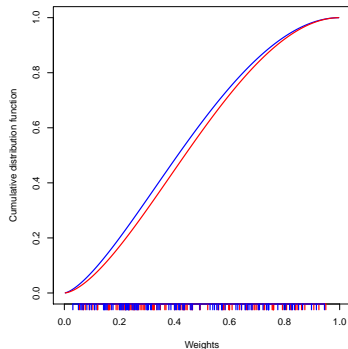
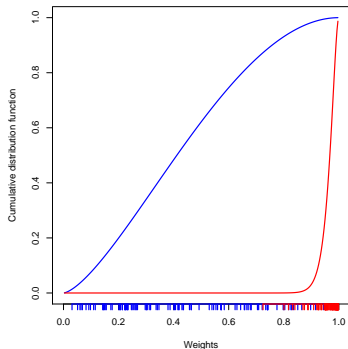
Comparando las cdf de los pesos escalados



Modelo T con $\hat{\eta} = 0.3357$ (azul) vs. $\eta = 0.01$ (panel izq.) y $\hat{\eta} = 0.2030$ retirando el conjunto de observaciones $I = \{28, 30, 35, 38, 39, 42, 66, 73, 75\}$ (panel der.).



Comparando las cdf de los pesos escalados



Modelo T con $\hat{\eta} = 0.3357$ (azul) vs. $\eta = 0.01$ (panel izq.) y $\hat{\eta} = 0.3083$ eliminando la observación 39 (panel der.).



Evaluando el supuesto de normalidad

- ▶ Un test para probar el **supuesto de Normalidad vs. T de Student** no es trivial.
- ▶ Deseamos probar la hipótesis:

$$H_0 : \eta = 0, \quad \text{vs.} \quad H_1 : \eta > 0,$$

note que la hipótesis nula está en la frontera del espacio paramétrico.

Usamos el **test de mezclas de chi-cuadrado** propuesto por Self y Liang (1987),

$$LR = 2\{\ell_o(\hat{\boldsymbol{\theta}}) - \ell_o(\tilde{\boldsymbol{\theta}})\} \xrightarrow{D} \frac{1}{2} \chi^2(0) + \frac{1}{2} \chi^2(1).$$

- ▶ En nuestro caso tenemos

$$LR = 2(1775.394 - 1736.386) = 78.0151,$$

que es mayor al valor crítico 2.7055 obtenido desde la mezcla 50 : 50 entre una $\chi^2(0)$ y $\chi^2(1)$ y por tanto **los datos no siguen una distribución normal**.



Estadístico T^2 de Hotelling generalizado

Considere $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$. Entonces

$$\sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$\sqrt{n}(\text{vech}(\mathbf{S}) - \text{vech}(\boldsymbol{\Sigma})) \xrightarrow{D} N_{p^*}(\mathbf{0}, \boldsymbol{\Omega}),$$

con $p^* = p(p+1)/2$ y

$$\boldsymbol{\Omega} = (\kappa + 1) \mathbf{D}_p^\top \{(\mathbf{I}_{p^2} + \mathbf{K}_p)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \kappa(\text{vec } \boldsymbol{\Sigma})(\text{vec } \boldsymbol{\Sigma})^\top\} \mathbf{D}_p,$$

donde $\kappa = 2\eta/(1 - 4\eta)$ y \mathbf{D}_p , \mathbf{K}_p son matrices de duplicación y conmutación, respectivamente.

De este modo, para el estadístico T^2 de Hotelling generalizado, tenemos

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \xrightarrow{D} \chi^2(p).$$

Observación: Este resultado es válido para toda la clase de distribuciones elípticas.



Test de hipótesis

Considere $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ con $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top \in \Theta \subset \mathbb{R}^k$, donde $k = p(p+3)/2 + 1$.

Se desea **probar hipótesis del tipo**

$$H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{g}, \quad \text{vs.} \quad H_1 : \mathbf{G}\boldsymbol{\theta} \neq \mathbf{g},$$

con $\mathbf{G} \in \mathbb{R}^{r \times k}$ matriz de contrastes tal que $\text{rg}(\mathbf{G}) = r$ y $\mathbf{g} \in \mathbb{R}^r$.

Utilizamos los **estadísticos de prueba**

$$\begin{aligned} LR &= 2\{\ell_o(\hat{\boldsymbol{\theta}}) - \ell_o(\tilde{\boldsymbol{\theta}})\}, \\ W &= n(\mathbf{G}\hat{\boldsymbol{\theta}} - \mathbf{g})^\top \{\mathbf{G}\mathcal{F}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{G}^\top\}^{-1}(\mathbf{G}\hat{\boldsymbol{\theta}} - \mathbf{g}), \\ S &= \frac{1}{n}\mathbf{U}^\top(\tilde{\boldsymbol{\theta}})\mathcal{F}^{-1}(\tilde{\boldsymbol{\theta}})\mathbf{U}(\tilde{\boldsymbol{\theta}}), \\ G &= \mathbf{U}^\top(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \end{aligned}$$

donde $\tilde{\boldsymbol{\theta}}$ y $\hat{\boldsymbol{\theta}}$ son los MLE bajo H_0 y H_1 , respectivamente, y $\mathbf{U}(\boldsymbol{\theta})$, $\mathcal{F}(\boldsymbol{\theta})$, representan la función score y matriz de información de Fisher.

Bajo $H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{g}$ los estadísticos anteriores tienen **distribución asintótica chi-cuadrado** con r grados de libertad.



Test de hipótesis

Considere $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ con $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top \in \Theta \subset \mathbb{R}^k$, donde $k = p(p+3)/2 + 1$.

Se desea **probar hipótesis del tipo**

$$H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{g}, \quad \text{vs.} \quad H_1 : \mathbf{G}\boldsymbol{\theta} \neq \mathbf{g},$$

con $\mathbf{G} \in \mathbb{R}^{r \times k}$ matriz de contrastes tal que $\text{rg}(\mathbf{G}) = r$ y $\mathbf{g} \in \mathbb{R}^r$.

Utilizamos los **estadísticos de prueba**

$$\begin{aligned} LR &= 2\{\ell_o(\hat{\boldsymbol{\theta}}) - \ell_o(\tilde{\boldsymbol{\theta}})\}, \\ W &= n(\mathbf{G}\hat{\boldsymbol{\theta}} - \mathbf{g})^\top \{\mathbf{G}\mathcal{F}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{G}^\top\}^{-1}(\mathbf{G}\hat{\boldsymbol{\theta}} - \mathbf{g}), \\ S &= \frac{1}{n}\mathbf{U}^\top(\tilde{\boldsymbol{\theta}})\mathcal{F}^{-1}(\tilde{\boldsymbol{\theta}})\mathbf{U}(\tilde{\boldsymbol{\theta}}), \\ G &= \mathbf{U}^\top(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \end{aligned}$$

donde $\tilde{\boldsymbol{\theta}}$ y $\hat{\boldsymbol{\theta}}$ son los MLE bajo H_0 y H_1 , respectivamente, y $\mathbf{U}(\boldsymbol{\theta})$, $\mathcal{F}(\boldsymbol{\theta})$, representan la función score y matriz de información de Fisher.

Bajo $H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{g}$ los estadísticos anteriores tienen **distribución asintótica chi-cuadrado** con r grados de libertad.



Test de hipótesis

Considere $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ con $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top \in \Theta \subset \mathbb{R}^k$, donde $k = p(p+3)/2 + 1$.

Se desea **probar hipótesis del tipo**

$$H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{g}, \quad \text{vs.} \quad H_1 : \mathbf{G}\boldsymbol{\theta} \neq \mathbf{g},$$

con $\mathbf{G} \in \mathbb{R}^{r \times k}$ matriz de contrastes tal que $\text{rg}(\mathbf{G}) = r$ y $\mathbf{g} \in \mathbb{R}^r$.

Utilizamos los **estadísticos de prueba**

$$\begin{aligned} LR &= 2\{\ell_o(\hat{\boldsymbol{\theta}}) - \ell_o(\tilde{\boldsymbol{\theta}})\}, \\ W &= n(\mathbf{G}\hat{\boldsymbol{\theta}} - \mathbf{g})^\top \{\mathbf{G}\mathcal{F}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{G}^\top\}^{-1}(\mathbf{G}\hat{\boldsymbol{\theta}} - \mathbf{g}), \\ S &= \frac{1}{n} \mathbf{U}^\top(\tilde{\boldsymbol{\theta}}) \mathcal{F}^{-1}(\tilde{\boldsymbol{\theta}}) \mathbf{U}(\tilde{\boldsymbol{\theta}}), \\ G &= \mathbf{U}^\top(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \end{aligned}$$

donde $\tilde{\boldsymbol{\theta}}$ y $\hat{\boldsymbol{\theta}}$ son los MLE bajo H_0 y H_1 , respectivamente, y $\mathbf{U}(\boldsymbol{\theta})$, $\mathcal{F}(\boldsymbol{\theta})$, representan la función score y matriz de información de Fisher.

Bajo $H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{g}$ los estadísticos anteriores tienen **distribución asintótica chi-cuadrado con r grados de libertad**.



Matriz de información de Fisher

Para $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$, tenemos

$$\mathcal{F}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{F}_{\mu\mu}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{F}_{\phi\phi}(\boldsymbol{\theta}) & \mathcal{F}_{\phi\eta}(\boldsymbol{\theta}) \\ \mathbf{0} & \mathcal{F}_{\phi\eta}^\top(\boldsymbol{\theta}) & \mathcal{F}_{\eta\eta}(\boldsymbol{\theta}) \end{pmatrix},$$

donde

$$\mathcal{F}_{\mu\mu}(\boldsymbol{\theta}) = c_\mu(\eta) \boldsymbol{\Sigma}^{-1},$$

$$\mathcal{F}_{\phi\phi}(\boldsymbol{\theta}) = \frac{1}{4} \mathbf{D}_p^\top \{ 2c_\phi(\eta)(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{N}_p + (c_\phi(\eta) - 1)(\text{vec } \boldsymbol{\Sigma}^{-1})(\text{vec } \boldsymbol{\Sigma}^{-1})^\top \} \mathbf{D}_p,$$

$$\mathcal{F}_{\phi\eta}(\boldsymbol{\theta}) = -\frac{c(\eta)(p+2)}{(1+p\eta)(1+(p+2)\eta)} \mathbf{D}_p^\top \text{vec } \boldsymbol{\Sigma}^{-1},$$

$$\mathcal{F}_{\eta\eta}(\boldsymbol{\theta}) = -\frac{1}{2\eta^2} \left\{ \frac{p}{(1-2\eta)^2} \left(\frac{1+p\eta(1-4\eta)-8\eta^2}{(1+p\eta)(1+(p+2)\eta)} \right) - \beta'(\eta) \right\},$$

con $\mathbf{N}_p = \frac{1}{2}(\mathbf{I}_{p^2} + \mathbf{K}_p)$, $c_\mu(\eta) = c_\phi(\eta)/(1-2\eta)$, $c_\phi(\eta) = (1+p\eta)/(1+(p+2)\eta)$ y

$$\beta'(\eta) = -\frac{1}{2\eta^2} \left\{ \psi' \left(\frac{1+p\eta}{2\eta} \right) - \psi' \left(\frac{1}{2\eta} \right) \right\},$$

donde $\psi'(z)$ denota la función trigama (Abramowitz y Stegun, 1970).



Matriz de información de Fisher

Para x_1, \dots, x_n muestra aleatoria desde $T_p(\mu, \Sigma, \eta)$, tenemos

$$\mathcal{F}(\theta) = \begin{pmatrix} \mathcal{F}_{\mu\mu}(\theta) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{F}_{\phi\phi}(\theta) & \mathcal{F}_{\phi\eta}(\theta) \\ \mathbf{0} & \mathcal{F}_{\phi\eta}^\top(\theta) & \mathcal{F}_{\eta\eta}(\theta) \end{pmatrix},$$

donde

$$\mathcal{F}_{\mu\mu}(\theta) = c_\mu(\eta) \Sigma^{-1},$$

$$\mathcal{F}_{\phi\phi}(\theta) = \frac{1}{4} \mathbf{D}_p^\top \{ 2c_\phi(\eta)(\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{N}_p + (c_\phi(\eta) - 1)(\text{vec } \Sigma^{-1})(\text{vec } \Sigma^{-1})^\top \} \mathbf{D}_p,$$

$$\mathcal{F}_{\phi\eta}(\theta) = -\frac{c(\eta)(p+2)}{(1+p\eta)(1+(p+2)\eta)} \mathbf{D}_p^\top \text{vec } \Sigma^{-1},$$

$$\mathcal{F}_{\eta\eta}(\theta) = -\frac{1}{2\eta^2} \left\{ \frac{p}{(1-2\eta)^2} \left(\frac{1+p\eta(1-4\eta)-8\eta^2}{(1+p\eta)(1+(p+2)\eta)} \right) - \beta'(\eta) \right\},$$

con $\mathbf{N}_p = \frac{1}{2}(\mathbf{I}_{p^2} + \mathbf{K}_p)$, $c_\mu(\eta) = c_\phi(\eta)/(1-2\eta)$, $c_\phi(\eta) = (1+p\eta)/(1+(p+2)\eta)$ y

$$\beta'(\eta) = -\frac{1}{2\eta^2} \left\{ \psi' \left(\frac{1+p\eta}{2\eta} \right) - \psi' \left(\frac{1}{2\eta} \right) \right\},$$

donde $\psi'(z)$ denota la función trigama (Abramowitz y Stegun, 1970).



Test sobre la media

Considere el test de hipótesis:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$

con $\boldsymbol{\mu}_0$ vector $p \times 1$ conocido.

En este caso es directo que

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0),$$

$$W = \frac{c_\phi(\hat{\boldsymbol{\eta}})}{1 - 2\hat{\boldsymbol{\eta}}} n(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^\top \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0).$$

Usando la definición del MLE para $\boldsymbol{\mu}$ y la curiosa relación de los pesos estimados, es posible notar la [equivalencia entre los estadísticos de Wald, score y gradiente](#). En efecto,

$$S = \frac{1 - 2\hat{\boldsymbol{\eta}}}{c_\phi^2(\hat{\boldsymbol{\eta}})} W \quad \text{y} \quad G = \frac{1}{c_\phi(\hat{\boldsymbol{\eta}})} W.$$



Test en forma explícita

Considere hipótesis en forma explícita:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\lambda}) \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{g}(\boldsymbol{\lambda}),$$

donde $\boldsymbol{g} : \mathbb{R}^m \rightarrow \mathbb{R}^k$ es función diferenciable cuyo subespacio paramétrico nulo es dado por

$$\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \Lambda \subset \mathbb{R}^m\},$$

y $\partial \boldsymbol{g}(\boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}^\top$ es matriz de rango completo.

En este caso tenemos

$$W = n(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^\top \mathcal{F}(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \quad S = \frac{1}{n} \boldsymbol{U}^\top(\tilde{\boldsymbol{\theta}}) \mathcal{F}^{-1}(\tilde{\boldsymbol{\theta}}) \boldsymbol{U}(\tilde{\boldsymbol{\theta}}),$$

y análogamente para los estadísticos LR y G . En este contexto, $\tilde{\boldsymbol{\theta}} = \boldsymbol{g}(\tilde{\boldsymbol{\lambda}})$ con $\tilde{\boldsymbol{\lambda}}$ el MLE de $\boldsymbol{\lambda}$.



Test sobre la matriz de covarianza

En particular estamos interesados en probar hipótesis de la forma:

$$H_0 : \phi = g_2(\lambda) \quad (\theta = (\mu^\top, \phi^\top, \eta)^\top).$$

- **Test de homogeneidad de varianzas:** En este caso deseamos probar

$$H_0 : \sigma_{11} = \sigma_{22} = \cdots = \sigma_{pp},$$

que puede ser escrito de forma alternativa como

$$H_0 : \Sigma = \sigma^2 \Phi,$$

donde $\sigma^2 > 0$ y Φ es matriz de correlación $p \times p$.

- **Test de equicorrelación:** En cuyo caso el objetivo es probar la hipótesis

$$H_0 : \Sigma = \sigma^2 \mathbf{R},$$

donde $\sigma^2 > 0$, $\mathbf{R} = (1 - \rho)\mathbf{I}_p + \rho\mathbf{J}_p$ con $-1/(p - 1) < \rho < 1$.



- Sea $\mathcal{J}(\theta) = \mathcal{F}^{-1}(\theta)$. Entonces, podemos aprovechar la **estructura en bloques** de la matriz de información de Fisher para obtener

$$W = n(\hat{\phi} - \tilde{\phi})^\top \mathcal{J}_{\phi\phi}^{-1}(\hat{\theta})(\hat{\phi} - \tilde{\phi}), \quad \text{and} \quad S = \frac{1}{n} U^\top(\tilde{\phi}) \mathcal{J}_{\phi\phi}(\tilde{\theta}) U(\tilde{\phi}),$$

con

$$\mathcal{J}_{\phi\phi}^{-1}(\theta) = \mathcal{F}_{\phi\phi}(\theta) - \mathcal{F}_{\phi\eta}(\theta) \mathcal{F}_{\phi\eta}^\top(\theta) / \mathcal{F}_{\eta\eta}(\theta),$$

- Obtenemos los MLEs bajo $H_0 : \Sigma = \sigma^2 \Phi$ y $H_0 : \Sigma = \sigma^2 R$ usando un **algoritmo EM condicional** (ECM).
- Implementación de estos test de hipótesis están disponibles en el **paquete MVT para R**.



Estimación restringida bajo homogeneidad de varianzas

Usamos un **algoritmo ECM** (Meng, 1993) para obtener los **MLE** de μ , $\Sigma = \sigma^2 \Phi$ y η , donde Φ representa una matriz de correlación.

En este caso tenemos que

$$x_i | \omega_i \stackrel{\text{ind}}{\sim} N_p(\mu, \sigma^2 \Phi / \omega_i), \quad \omega_i \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n.$$

Sea $\Delta = \Phi^{-1}$, de este modo es fácil notar que:

$$Q(\theta | \theta^{(k)}) = Q(\tau | \theta^{(k)}) + Q(\eta | \theta^{(k)}),$$

donde $\tau = (\mu^\top, \sigma^2, (\text{vecp } \Phi)^\top)^\top$, $Q(\eta | \theta^{(k)})$ fue definido en (2) y

$$Q(\tau | \theta^{(k)}) = -\frac{np}{2} \log \sigma^2 + \frac{n}{2} \log |\Delta| - \frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu)^\top \Delta (x_i - \mu),$$

Observación: Este algoritmo representa una **alternativa más simple** que la propuesta de Modarres (1993).



Estimación restringida bajo homogeneidad de varianzas

Usamos un **algoritmo ECM** (Meng, 1993) para obtener los **MLE** de μ , $\Sigma = \sigma^2 \Phi$ y η , donde Φ representa una matriz de correlación.

En este caso tenemos que

$$x_i | \omega_i \stackrel{\text{ind}}{\sim} N_p(\mu, \sigma^2 \Phi / \omega_i), \quad \omega_i \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n.$$

Sea $\Delta = \Phi^{-1}$, de este modo es fácil notar que:

$$Q(\theta | \theta^{(k)}) = Q(\tau | \theta^{(k)}) + Q(\eta | \theta^{(k)}),$$

donde $\tau = (\mu^\top, \sigma^2, (\text{vecp } \Phi)^\top)^\top$, $Q(\eta | \theta^{(k)})$ fue definido en (2) y

$$Q(\tau | \theta^{(k)}) = -\frac{np}{2} \log \sigma^2 + \frac{n}{2} \log |\Delta| - \frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu)^\top \Delta (x_i - \mu),$$

Observación: Este algoritmo representa una **alternativa más simple** que la propuesta de Modarres (1993).



Estimación restringida bajo homogeneidad de varianzas

Usamos un **algoritmo ECM** (Meng, 1993) para obtener los **MLE** de μ , $\Sigma = \sigma^2 \Phi$ y η , donde Φ representa una matriz de correlación.

En este caso tenemos que

$$x_i | \omega_i \stackrel{\text{ind}}{\sim} N_p(\mu, \sigma^2 \Phi / \omega_i), \quad \omega_i \stackrel{\text{ind}}{\sim} \text{Gama}\left(\frac{1}{2\eta}, \frac{1}{2c(\eta)}\right), \quad i = 1, \dots, n.$$

Sea $\Delta = \Phi^{-1}$, de este modo es fácil notar que:

$$Q(\theta | \theta^{(k)}) = Q(\tau | \theta^{(k)}) + Q(\eta | \theta^{(k)}),$$

donde $\tau = (\mu^\top, \sigma^2, (\text{vecp } \Phi)^\top)^\top$, $Q(\eta | \theta^{(k)})$ fue definido en (2) y

$$Q(\tau | \theta^{(k)}) = -\frac{np}{2} \log \sigma^2 + \frac{n}{2} \log |\Delta| - \frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu)^\top \Delta (x_i - \mu),$$

Observación: Este algoritmo representa una **alternativa más simple** que la propuesta de Modarres (1993).



Estimación restringida bajo homogeneidad de varianzas

Paso CM-1: Actualizar $\mu^{(k+1)}$ como:

$$\mu^{(k+1)} = \frac{1}{\sum_{j=1}^n \omega_j^{(k)}} \sum_{i=1}^n \omega_i^{(k)} x_i.$$

Paso CM-2: Sea:

$$\Sigma^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu^{(k+1)})(x_i - \mu^{(k+1)})^\top,$$

reorganizando los cálculos podemos escribir $Q(\tau|\theta^{(k)})$ para $\mu = \mu^{(k+1)}$ como

$$Q(\tau|\theta^{(k)}) = -\frac{np}{2} \log \sigma^2 + \frac{n}{2} \log |\Delta| - \frac{n}{2\sigma^2} \text{tr}(\Sigma^{(k+1)} \Delta).$$

Así, actualizamos $\sigma^{2(k+1)}$ para $\Delta = \Delta^{(k)}$ como

$$\sigma^{2(k+1)} = \frac{1}{p} \text{tr}(\Sigma^{(k+1)} \Delta^{(k)}).$$



Estimación restringida bajo homogeneidad de varianzas

Paso CM-1: Actualizar $\mu^{(k+1)}$ como:

$$\mu^{(k+1)} = \frac{1}{\sum_{j=1}^n \omega_j^{(k)}} \sum_{i=1}^n \omega_i^{(k)} x_i.$$

Paso CM-2: Sea:

$$\Sigma^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} (x_i - \mu^{(k+1)})(x_i - \mu^{(k+1)})^\top,$$

reorganizando los cálculos podemos escribir $Q(\tau|\theta^{(k)})$ para $\mu = \mu^{(k+1)}$ como

$$Q(\tau|\theta^{(k)}) = -\frac{np}{2} \log \sigma^2 + \frac{n}{2} \log |\Delta| - \frac{n}{2\sigma^2} \text{tr}(\Sigma^{(k+1)} \Delta).$$

Así, actualizamos $\sigma^{2(k+1)}$ para $\Delta = \Delta^{(k)}$ como

$$\sigma^{2(k+1)} = \frac{1}{p} \text{tr}(\Sigma^{(k+1)} \Delta^{(k)}).$$



Estimación restringida bajo homogeneidad de varianzas

Paso CM-3: Para $\sigma^2 = \sigma^{2^{(k+1)}}$, tenemos

$$Q(\boldsymbol{\tau}|\boldsymbol{\theta}^{(k)}) = -\frac{np}{2} \log \sigma^{2^{(k+1)}} + \frac{n}{2} \log |\boldsymbol{\Delta}| - \frac{n}{2} \text{tr}(\boldsymbol{\Delta} \boldsymbol{\Sigma}^{(k+1)} / \sigma^{2^{(k+1)}}).$$

Usando el Lema 3.2.2 de Anderson (2003), sigue que podemos actualizar $\boldsymbol{\Phi}^{(k+1)}$ como

$$\boldsymbol{\Phi}^{(k+1)} = \boldsymbol{\Sigma}^{(k+1)} / \sigma^{2^{(k+1)}}.$$

Paso CM-4: Actualizar $\eta^{(k+1)}$ maximizando $Q(\eta|\boldsymbol{\theta}^{(k)})$ con relación a η , como:

$$\eta^{(k+1)} = \arg \max_{\eta} Q(\eta|\boldsymbol{\theta}^{(k)}),$$

utilizando algún procedimiento de optimización uni-dimensional (nosotros usamos el método de Brent, 1973).



Estimación restringida bajo homogeneidad de varianzas

Paso CM-3: Para $\sigma^2 = \sigma^{2^{(k+1)}}$, tenemos

$$Q(\boldsymbol{\tau}|\boldsymbol{\theta}^{(k)}) = -\frac{np}{2} \log \sigma^{2^{(k+1)}} + \frac{n}{2} \log |\boldsymbol{\Delta}| - \frac{n}{2} \text{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}^{(k+1)}/\sigma^{2^{(k+1)}}).$$

Usando el Lema 3.2.2 de Anderson (2003), sigue que podemos actualizar $\boldsymbol{\Phi}^{(k+1)}$ como

$$\boldsymbol{\Phi}^{(k+1)} = \boldsymbol{\Sigma}^{(k+1)}/\sigma^{2^{(k+1)}}.$$

Paso CM-4: Actualizar $\eta^{(k+1)}$ maximizando $Q(\eta|\boldsymbol{\theta}^{(k)})$ con relación a η , como:

$$\eta^{(k+1)} = \arg \max_{\eta} Q(\eta|\boldsymbol{\theta}^{(k)}),$$

utilizando algún procedimiento de **optimización uni-dimensional** (nosotros usamos el **método de Brent**, 1973).



Se consideró un estudio de simulación siguiendo el propuesto por [Li, Fang y Zhu \(1997, JCGS 6, 435-450\)](#).

Se simuló $M = 10\,000$ conjuntos de datos de tamaño $n = 60, 180, 960$ desde:

$$1) N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{y} \quad 2) T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta),$$

para $p = 5, 15, 45$ con

$$\boldsymbol{\mu} = \mathbf{1}, \quad \boldsymbol{\Sigma} = 0.5 \mathbf{I}_p + 0.5 \mathbf{1}\mathbf{1}^\top, \quad \eta = 0.25,$$

y se probó la [hipótesis de equicorrelación](#) asumiendo un modelo [normal](#) y [Student- \$t\$](#)

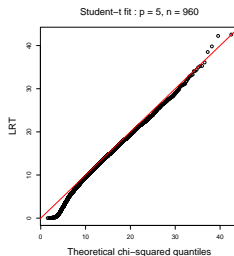
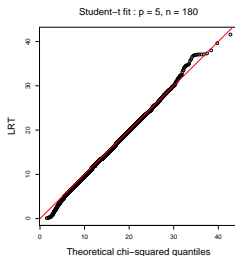
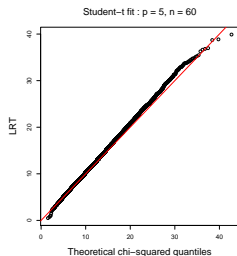
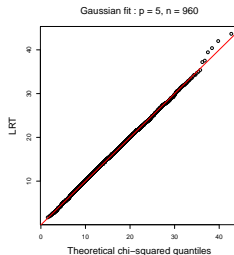
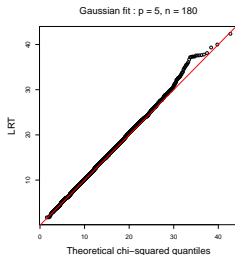
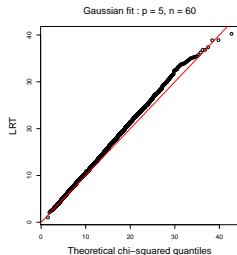
$$H_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{R}, \quad \mathbf{R} = (1 - \rho) \mathbf{I} + \rho \mathbf{1}\mathbf{1}^\top,$$

usando la [biblioteca MVT para R](#) (rutina [implementada en C](#), usando [API de R 3.1.1](#)).

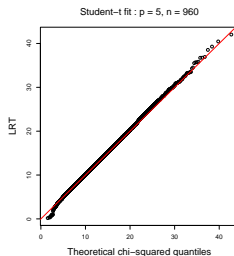
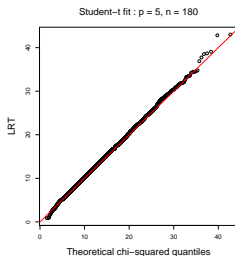
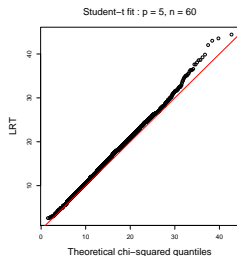
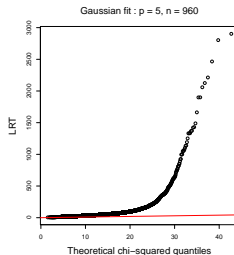
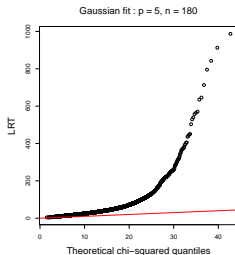
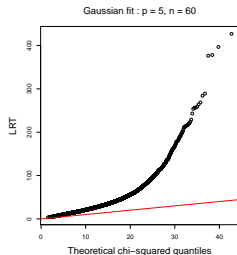
Simulaciones fueron ejecutadas en un [Servidor IBM x3650 M4](#), con 2 procesadores Intel Xeon E5-2670 y 264 GB de RAM Quadro 6000 (tiempo: 14 hrs, 29 min, 15 seg).



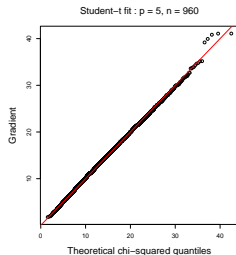
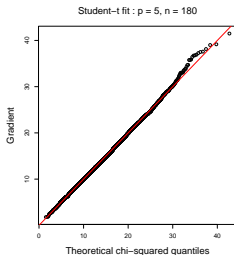
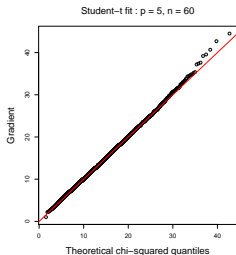
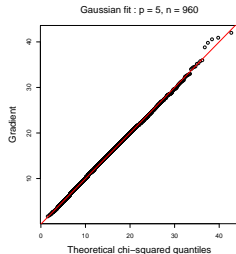
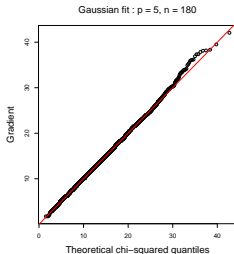
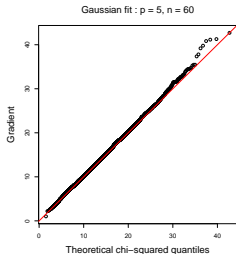
Resultados: LRT bajo supuesto gaussiano, $p = 5$



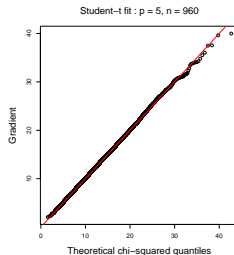
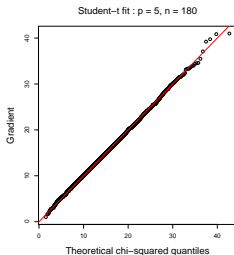
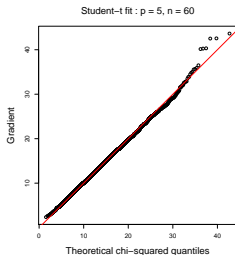
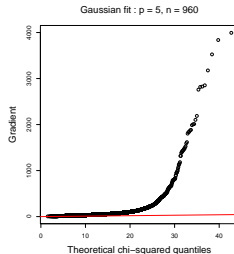
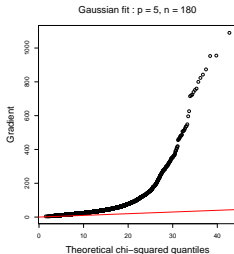
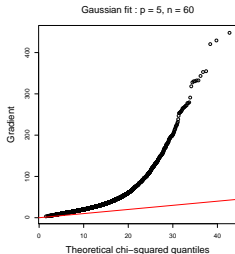
Resultados: LRT bajo supuesto Student- t , $p = 5$



Resultados: test gradiente bajo supuesto gaussiano, $p = 5$



Resultados: test gradiente bajo supuesto Student- t , $p = 5$



Resultados: Tamaños empíricos, datos generados bajo normalidad²

Test	<i>n</i>	normal fit			multivariate <i>t</i> fit		
		<i>p</i>			<i>p</i>		
		5	15	45	5	15	45
LR	60	7.28	22.81	100.00	6.40	21.69	100.00
	180	5.63	8.83	73.07	4.86	7.51	71.16
	960	4.84	5.78	10.83	3.87	3.61	7.31
Wald	60	31.16	99.98	100.00	28.19	99.96	100.00
	180	12.76	79.47	100.00	10.86	75.29	100.00
	960	6.05	15.69	98.78	5.64	12.20	97.20
score	60	5.45	6.57	10.78	4.96	5.76	10.52
	180	5.07	5.73	6.57	4.53	5.08	5.89
	960	4.72	5.31	5.39	4.60	4.69	4.39
gradient	60	5.45	6.57	10.78	5.16	6.03	10.70
	180	5.07	5.73	6.57	4.78	5.33	6.24
	960	4.72	5.31	5.39	4.82	4.87	4.82

²5% de significancia

Resultados: Tamaños empíricos bajo el supuesto de la t multivariada³

Test	n	normal fit			multivariate t fit		
		p			p		
		5	15	45	5	15	45
LR	60	64.33	99.44	100.00	6.48	23.91	100.00
	180	75.88	99.92	100.00	5.58	8.84	73.05
	960	88.23	100.00	100.00	5.69	5.18	9.50
Wald	60	83.83	100.00	100.00	31.70	99.86	100.00
	180	82.35	100.00	100.00	12.58	72.55	100.00
	960	88.78	100.00	100.00	6.51	13.78	95.14
score	60	61.30	98.44	100.00	4.70	6.31	9.76
	180	75.31	99.90	100.00	4.81	5.49	6.48
	960	88.19	100.00	100.00	5.13	4.88	5.28
gradient	60	61.30	98.44	100.00	4.55	6.24	18.30
	180	75.31	99.90	100.00	4.71	5.01	6.42
	960	88.19	100.00	100.00	5.12	4.70	4.25

³5% de significancia

Test de homogeneidad de varianzas: Datos de AFP

Test de hipótesis $H_0 : \Sigma = \sigma^2 \Phi$, para los modelos normal y Student- t

Estadístico de prueba	normal		Student- t	
	test	valor- p	test	valor- p
LRT	6.9882	0.0723	3.5075	0.3198
Wald	7.7109	0.0524	4.0586	0.2552
Score	6.9873	0.0723	3.3771	0.3371
Gradiente	6.9873	0.0723	3.4043	0.3334



Test de equicorrelación: Datos de AFP

Test de hipótesis $H_0 : \Sigma = \sigma^2\{\rho\mathbf{I} + (1 - \rho)\mathbf{J}\}$, para los modelos normal y Student- t

Estadístico de prueba	normal		Student- t	
	test	valor- p	test	valor- p
LRT	25.2786	0.0014	14.1053	0.0791
Wald	34.8397	0.0000	22.8326	0.0036
Score	25.4298	0.0013	14.0520	0.0804
Gradiente	25.4298	0.0013	11.9989	0.1513

Estimación bajo H_0 : $\tilde{\rho} = 0.9896$, $\tilde{\eta} = 0.3639$ (Student- t) y $\tilde{\rho} = 0.9792$ (normal).



MVT: Estimation and testing for the multivariate t-distribution

Biblioteca MVT desarrolla la inferencia en el modelo multivariado asumiendo la distribución t de Student.

Cálculos son realizados eficientemente usando código compilado en C y Fortran (llamando rutinas en BLAS, LINPACK y Mathlib).

- ▶ Estimación: `studentFit`
- ▶ Hipótesis: `center.test`, `equicorrelation.test`, `homogeneity.test`.
- ▶ RNG: `rmt` (también para el caso $\eta = 0$).
- ▶ Calidad del modelo: `envelope`, `[d,p,q]weights`, `wilson.hilferty`.
- ▶ Matrices: `duplication`, `commutation`.
- ▶ Bases de datos: `companies`, `cork`, `examScor`.



Usando biblioteca MVT: Ajuste via la distribución $\mathcal{T}_p(\mu, \Sigma, \eta)$

```
# Carga biblioteca e invoca función para estimación de parámetros
> library(MVT)
> fit <- studentFit(~ cuprum.D + habitat.D + planvital.D + provida.D,
+                 data = PFA, family = Student(eta = .25))
```

```
# Resultados del ajuste
```

```
> fit
```

```
Call:
```

```
studentFit(x = ~cuprum.D + habitat.D + planvital.D + provida.D,
           data = PFA, family = Student(eta = 0.3357))
```

```
Converged in 129 iterations
```

```
Center:
```

cuprum.D	habitat.D	planvital.D	provida.D
0.0041	0.0043	0.0040	0.0040

```
Scatter matrix estimate:
```

	cuprum.D	habitat.D	planvital.D	provida.D
cuprum.D	0.0002065546			
habitat.D	0.0002061143	0.0002137712		
planvital.D	0.0002011461	0.0002064906	0.0002071240	
provida.D	0.0001993436	0.0002036495	0.0001990968	0.0001993050

```
Number of Observations: 101
```



Usando biblioteca MVT: Elementos del objeto ajustado

```
# Elementos en un objeto de clase 'StudentFit'
> attributes(fit)
$names
[1] "call"          "x"              "dims"           "family"         "settings"
[6] "start"         "center"         "Scatter"        "logLik"         "numIter"
[11] "control"       "weights"        "distances"      "speed"          "converged"
[16] "eta"

$class
[1] "studentFit"

# Explorando algunos elementos
> fit$logLik
[1] 1775.394

> fit$speed
  user  system elapsed
0.007   0.000   0.031

> fit$weights
[1] 5.2884469 3.3019543 0.8227232 2.3968817 3.3093736 1.2604426
[7] 5.5246361 6.3957585 5.2881307 1.6644460 2.9792357 4.7457558
...
[99] 3.9460360 4.6913872 5.8861983
```



Usando biblioteca MVT: Test de hipótesis

Para realizar el **test de equicorrelación**, $H_0 : \Sigma = \sigma^2 \mathbf{R}$, usamos:

```
> z <- equicorrelation.test(fit, test = "gradient")  
> z
```

Gradient test **for** equicorrelation

data: PFA

Gradient statistic = 11.9989, **df** = 8, p-value = 0.1513
alternative hypothesis: true **scale matrix** not follows an
equicorrelation **structure**.

sample estimate:

	cuprum.D	habitat.D	planvital.D	provida.D
cuprum.D	0.0002065546			
habitat.D	0.0002061143	0.0002137712		
planvital.D	0.0002011461	0.0002064906	0.0002071240	
provida.D	0.0001993436	0.0002036495	0.0001990968	0.0001993050

otras opciones para el argumento **test**, son "LRT", "Wald" y "score".



Usando biblioteca MVT: Elementos del objeto

```
# Elementos en un objeto de clase 'equicorrelation.test'
> attributes(z)
$names
[1] "statistic" "parameter" "p.value" "estimate" "null.value"
[6] "method" "null.fit" "data" "family"

$class
[1] "equicorrelation.test"

# Explorando algunos elementos
> attributes(z$null.fit)
$names
[1] "call" "center" "Scatter" "sigma2" "rho"
[6] "eta" "distances" "weights" "logLik" "numIter"

> z$null.fit$sigma2
[1] 0.0002362117

> z$null.fit$rho
[1] 0.9806336

> z$null.fit$eta
[1] 0.3638998
```



Referencias



Fiorentini, G., Sentana, E., and Calzolari, G. (2003).

Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations.

Journal of Business & Economic Statistics **21**, 532-546.



Harris, P. (1985).

Testing the variance homogeneity of correlated variables.

Biometrika **72**, 103-107.



Kent, J.T., Tyler, D.E., and Vardi, Y. (1994).

A curious likelihood identity for the multivariate t -distribution.

Communications in Statistics - Simulation and Computation **23**, 441-453.



Lange, K., Little, R.J.A., and Taylor, J.M.G. (1989).

Robust statistical modeling using the t distribution.

Journal of the American Statistical Association **84**, 881-896.



Modarres, R. (1993).

Testing the equality of dependent variables.

Biometrical Journal **7**, 785-790.



Sutradhar, B.C. (1993).

Score test for the covariance matrix of elliptical t -distribution.

Journal of Multivariate Analysis **46**, 1-12.

