Análisis Multivariado I - Práctica 3 - Parte 2

Inferencia para matrices de covarianza

1. Consideremos el modelo de regresión lineal

$$y_i = \theta_0 + \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\theta} + \varepsilon_i$$

donde $\mathbf{x}_i$ m.a. $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ independientes de ε_i m.a. $N(0, \sigma^2)$.

   (a) ¿Qué distribución tiene $\mathbf{z}_i = \left(\mathbf{x}_i^{\mathrm{T}}, y_i\right)^{\mathrm{T}}$?

   (b) ¿Qué expresión tiene (en términos de $\mathbf{x}_i^{\mathrm{T}}$ e y_i) el estadístico del test del cociente de verosimilitud para la independencia de bloques aplicado a las $\mathbf{z}_i$?

Comentario: el test de F usual para $H_0 : \boldsymbol{\theta} = \mathbf{0}$ es equivalente al test del cociente de verosimilitud anterior.

2. Dada un muestra $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mostrar que el estadístico del cociente de verosimilitud para testear $H_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ cumple

$$\ell^{2/(nd)} = \frac{|\mathbf{Q}|^{1/d}}{\operatorname{tr}(\mathbf{Q})/d}.$$

Escribirlo en función de los autovalores e interpretar. ¿Cuál es la distribución asintótica de $-2\ln\ell$ bajo H_0?

3. Consideremos una muestra $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Hallar el estadístico del test del cociente de verosimilitud para las hipótesis

$$H_0 : \boldsymbol{\Sigma} = \sigma^2 (1-\rho) \mathbf{I} + \sigma^2 \rho \mathbf{1}_d \mathbf{1}_d^{\mathrm{T}}$$

versus

$$H_1 : \nexists\, \sigma^2 \text{ ó } \rho \in (0,1) \text{ tal que } \boldsymbol{\Sigma} = \sigma^2 (1-\rho) \mathbf{I} + \sigma^2 \rho \mathbf{1}_d \mathbf{1}_d^{\mathrm{T}}.$$

4. Sea $\mathbf{x}_1, \ldots, \mathbf{x}_n$ una muestra de una $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Encontrar el test basado en el principio de unión–intersección para testear $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$. Explicar cómo se podrían calcular los valores críticos del test.

5. A partir de una muestra $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ encontrar intervalos de confianza para $\mathbf{v}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{v}$ de nivel simultáneo $1-\alpha$ para todo $\mathbf{v} \neq \mathbf{0}$.

6. La tabla 3.5 contiene información fisonómica de hermanos. Las variables son

X_1 = longitud de la cabeza del primer hijo,

X_2 = ancho de la cabeza del primer hijo,

X_3 = longitud de la cabeza del segundo hijo,

X_4 = ancho de la cabeza del segundo hijo;

referidas a 25 familias distintas. Consideremos el siguiente vector aleatorio:

$$\mathbf{y} = (X_1 + X_3\,,\, X_2 + X_4\,,\, X_1 - X_3\,,\, X_2 - X_4)^{\mathrm{T}}$$

Testear si las primeras dos coordenadas de $\mathbf{y}$ son independientes de las segundas dos, a nivel 0.05 con el test de máxima verosimilitud y con el test basado en el principio de unión-intersección. Comparar las conclusiones.

7. Sea $\mathbf{x}_1, \ldots, \mathbf{x}_n$ una m.a. $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Se desea testear que las d variables son independientes, es decir

$$H_0 : \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{dd})\,.$$

Consideremos la matriz de correlación muestral $\mathbf{R} = (R_{j\ell})_{1\le j,\ell\le d}$ cuyos elementos son

$$R_{jk} = \frac{\widehat{\sigma}_{jk}}{(\widehat{\sigma}_{jj}\,\widehat{\sigma}_{kk})^{1/2}} = \frac{Q_{jk}}{(Q_{jj}\,Q_{kk})^{1/2}}\,,$$

donde

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\mathbf{Q} = (\widehat{\sigma}_{j\ell})_{1\le j,\ell\le d} \qquad \mathbf{Q} = (Q_{j\ell})_{1\le j,\ell\le d}$$

(a) Probar que el estadístico del test del cociente de verosimilitud es $\Lambda = |\mathbf{R}|^{n/2}$.

(b) ¿Cuál es la distribución asintótica de $-2\ln\Lambda$ bajo H_0?

Test para igualdad de covarianza entre dos poblaciones

8. En el ejercicio 1 de la Parte 1 de la Práctica 3 se estudiaba el costo de transporte de la leche desde las granjas hasta las lecherías para $n_1 = 36$ camiones nafteros y $n_2 = 23$ camiones a diesel.

(a) Para testear si había diferencias entre los vectores de costos medios, se supuso que las dos poblaciones tenían igual matriz de covarianza. ¿Es este supuesto razonable? Tomar $\alpha = 0.01$.

(b) Si la hipótesis de igualdad de matrices de covarianzas es rechazada en la parte (a), ¿cómo testearía la igualdad de vectores medios?

9. En el ejercicio 4 de la Parte 1 de la Práctica 3 se estudiaban los resultados de tomar un test de habilidad sicolingual a dos grupos de 27 chicos de edades 8-9 años. Nos interesaba estudiar las siguientes hipótesis:

H_{01}: los dos perfiles son similares

H_{02}: los dos perfiles están al mismo nivel

H_{03}: no hay diferencias entre las medias de los tests

Sean

x_1 = recepción auditiva

$x_2 =$ recepción visual

$x_3 =$ memoria visual

$x_4 =$ asociación auditiva

$x_5 =$ memoria auditiva

$x_6 =$ asociación visual

$x_7 =$ oclusión visual

$x_8 =$expresión oral

$x_9 =$ oclusión gramatical

$x_{10} =$ destreza manual

(a) Para cada una de las hipótesis anteriores, ¿es necesario suponer que las matrices de covarianza del vector $\mathbf{x} = (x_1, \ldots, x_{10})^{\mathrm{T}}$ en las dos poblaciones sea la misma o puede hacer un supuesto más débil? Exprese matemáticamente la hipótesis sobre la igualdad de covarianzas que sea necesaria en cada caso.

(b) Para cada una de las situaciones de interés estudie si el supuesto de igualdad de covarianzas es razonable. Tomar $\alpha = 0.01$.