1 Análisis Multivariado I - Práctica 2 - Parte 1

Test de Hotelling para una muestra

Definición: Se define el elipsoide como el conjunto de puntos $\mathbf{x} \in \mathbb{R}^d$ tales que

$$(\mathbf{x} - \mathbf{b})^{\mathrm{T}} \mathbf{A} (\mathbf{x} - \mathbf{b}) = c^2 \tag{1}$$

donde $\mathbf{A}$ es una matriz definida positica y $\mathbf{b}$ es el centro del elipsoide.
Si $d = 2$ se llama elipse.

1. En este ejercicio graficaremos elipses. Consideremos la definición dada en (1).

   (a) Supongamos que $\mathbf{b} = (0, 0)^{\mathrm{T}}$ y $\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Desarrollar (1) en este caso y graficar en `R` el caso $\lambda_1 = 2$, $\lambda_2 = 3$ y $c = 1$ usando dos funciones (función superior y función inferior) o coordenadas polares.

   (b) Considerar el caso $\mathbf{A}$ simétrica y definida positiva. Usar la descomposición espectral para transformar el problema $\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} = c^2$ en

   $$\mathbf{y}^{\mathrm{T}} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \mathbf{y} = c^2$$

   Usar la transformación para graficar en `R` la elipse que tiene $\mathbf{A} = \begin{pmatrix} 2.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix}$ y $c = 2$.

   (c) ¿Cómo se modifica el ítem anterior si ahora tenemos $\mathbf{b}$ genérico? Graficar en `R` la elipse anterior pero ahora centrada en $\mathbf{b} = (1, 3)^{\mathrm{T}}$.

2. Supongamos que los datos de la tabla 2.1 son una muestra aleatoria normal multivariada.

   (a) Testear la hipótesis de que el peso medio es 63 kg y la altura media es 1.60m.

   (b) Hallar un elipsoide de confianza de nivel 95% para el peso y la altura medios de los indios peruanos.

   (c) Usando el test de normalidad, ¿le parece razonable este supuesto? ¿A qué se lo atribuiría? ¿Qué haría?

3. Dada una muestra aleatoria $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, sea $T^2_{\mathbf{x}}$ el estadístico de Hotelling para testear $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

   Consideremos la siguiente transformación de los datos. Sean

   $$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b}$$

   con $\mathbf{A} \in \mathbb{R}^{d \times d}$ inversible fija y $\mathbf{b} \in \mathbb{R}^{d \times 1}$, sea $T^2_{\mathbf{y}}$ el estadístico de Hotelling para testear $H_0 : \boldsymbol{\mu}_{\mathbf{y}} = \mathbf{A}\boldsymbol{\mu}_0 + \mathbf{b}$, donde $\boldsymbol{\mu}_{\mathbf{y}} = \mathbb{E}(\mathbf{y}_1)$. Probar que $T^2_{\mathbf{x}} = T^2_{\mathbf{y}}$.

4. Sea $\mathbf{x}_1, \ldots, \mathbf{x}_n$ una m.a. con distribución $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, dada $\mathbf{C} \in \mathbb{R}^{q\times d}$ con rango $(\mathbf{C}) = q$ y $\mathbf{b} \in \mathbb{R}^{q\times 1}$, encontrar un test de Hotelling para testear $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{b}$.

5. Sea $\mathbf{x}_1, \ldots, \mathbf{x}_n$ una m.a. con distribución $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y sea $\mathbf{K} \in \mathbb{R}^{d\times q}$ con $q < d$ y rango $(\mathbf{K}) = q$. Se quiere testear la hipótesis

$$H_0 : \exists \boldsymbol{\beta} \in \mathbb{R}^{q\times 1} \quad \text{tal que} \quad \boldsymbol{\mu} = \mathbf{K}\boldsymbol{\beta}\,.$$

   (a) Interpretar el significado de esta hipótesis.
   (b) Mostrar que H_0 puede escribirse como $\mathbf{A}\boldsymbol{\mu} = \mathbf{0}$ para cierta matriz $\mathbf{A}$ y por lo tanto, puede testearse usando un test de Hotelling.

6. Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ i.i.d. $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu}^{\mathrm{T}} = (\mu_1, \mu_2)$ y sea $y_i = x_{i1} - x_{i2}$ $(1 \le i \le n)$. Mostrar que el test T^2 de Hotelling para testear $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \ne \mu_2$ es equivalente al test usual para muestras apareadas basado en el estadístico $\bar{y}\sqrt{n}/s_y$.

7. Sean $\phi_i = \mu_i - \mu_d$ $(1 \le i \le d-1)$. Mostrar que el conjunto de todas las combinaciones lineales $\sum_{i=1}^{d-1} h_i\phi_i$ es equivalente al conjunto de contrastes $\sum_{i=1}^{d} c_i\mu_i$ $\left(\sum_{i=1}^{d} c_i = 0\right)$.

8. Consideremos los 14 datos de la tabla 2.2. Suponiendo que son una muestra normal multivariada:

   (a) Realizar un test para estudiar el supuesto de normalidad multivariada de los datos.
   (b) Testear $H_0 : \boldsymbol{\mu} = (45, 42, 45, 42)^{\mathrm{T}}$.
   (c) Testear la hipótesis de que las medias de los pesos son iguales en las cuatro direcciones.
   (d) Testear la hipótesis de que las medias de los pesos en las direcciones norte y sur son iguales y en las direcciones este y oeste también. Comparar con el resultado obtenido en el inciso (a).
   (e) Encontrar intervalos de confianza simultáneos de nivel 95% para $\mu_1 - \mu_3$ y $\mu_2 - \mu_4$. ¿Qué método conviene usar?
   (f) Construir intervalos de confianza simultáneos de nivel 95% para todos los contrastes $\mathbf{c}^{\mathrm{T}}\boldsymbol{\mu}$ donde $\sum_{i=1}^{4} c_i = 0$. Probar que existe $\mathbf{c}$ tal que 0 no pertenece al intervalo de confianza para $\mathbf{c}^{\mathrm{T}}\boldsymbol{\mu}$. Intentar hallarlo explícitamente.