

Tarea 2 de estadística multivariada

J.M. Ponciano

Entregar el martes 23 de Septiembre del 2008

1. método de Box-Muller para general v.a. normales

a) Sean U_1 y U_2 dos v.a. uniformes independientes. Sean

$$X = \sqrt{-2\ln U_1} \cos(2\pi U_2)$$

y

$$Y = \sqrt{-2\ln U_1} \sin(2\pi U_2).$$

Demuestre que la distribución conjunta $\mathbf{Z} = [X, Y]'$ es bivariada normal.

- b) Use el resultado de la respuesta anterior para crear una función en $\mathbf{R}$ que entregue n muestras de una distribución Normal univariada con media μ y varianza σ^2 . Tome $n = 10000$ y verifique que su función sirve graficando un histograma de las muestras para $\mu = 3$ y $\sigma^2 = 1,5$ y superponiendo al histograma la función de densidad de la distribución normal.
- c) Escriba una función para muestrear de una distribución normal multivariada usando la función de la pregunta anterior y cheque que funciona.

2. Sea $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\mu} = [-3, 1, 4]'$ y

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- a) ¿Cual de las siguientes variables aleatorias son independientes? Justifique su respuesta
- 1) X_1 y X_2
 - 2) X_2 y X_3

- 3) $(X_1, X_2)/2$ y X_3
- 4) X_2 y $X_2 - (5/2)X_1 - X_3$
- b) Especifique la distribución de $X_2|(X_1 = x_1, X_3 = x_3)$
3. Demuestre que si $|\Sigma| \neq 0$, $|\Sigma| = |\Sigma_{22}| \cdot |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|$
4. Demuestre que la T cuadrada de Hotelling es invariante a cambios en las unidades de medida de $\mathbf{X}_{p \times 1}$ de la forma

$$\mathbf{Y}_{p \times 1} = \mathbf{C}_{p \times p} \mathbf{X}_{p \times 1} + \mathbf{d}_{p \times 1},$$

dado que $\mathbf{C}$ es no-singular. Investigue y describa una situación en la que este resultado sea útil.

5. Sea

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Si $\mathbf{A}$ se distribuye Wishart $W(\Sigma, n)$, cual es la distribución de $\mathbf{A}_{11}$?

6. Calcular los intervalos de confianza individuales y simultáneos de los datos de las pruebas pre-universitarias (están en la página web de la clase, en la casilla del Jueves 11 de Septiembre). Estos datos son las calificaciones de 87 estudiantes en 3 pruebas: X_1 = Historia y ciencias sociales, X_2 = habilidad verbal y escrita, X_3 = ciencias. Analice y explore los datos con cualquier herramienta gráfica y analítica que quiera (sea creativo). Por último, escoja dos variables cualesquiera y haga una gráfica de la elipse de confianza simultánea, con sus ejes correspondientes. Como producto, quiero que entreguen un análisis ordenado de estos datos. No quiero que me entreguen sólo el código de R y las gráficas. Quiero que expliquen paso a paso lo que hacen, porque lo hacen y que den su interpretación de los resultados. Quiero que sea una respuesta bien escrita, voy a ser muy estricto con la redacción y presentación de los resultados. Las gráficas tienen que estar rotuladas y bien explicadas, con una buena leyenda explicativa.
7. Un educador de música examinó a miles de estudiantes Finlandeses para evaluar sus habilidades musicales naturales. Se midieron las siguientes variables de manera cuantitativa: habilidad melódica (X_1), habilidad para las armonías (X_2), tempo (X_3), seguimiento del metrónomo (X_4), fraseo (X_5), balance (X_6) y estilo (X_7). Con los resultados, se establecieron normas educativas y programas de música a nivel nacional. A

Cuadro 1: Medias y desviaciones estándar en siete pruebas realizadas para evaluar la aptitud musical de 96 estudiantes finlandeses de sexto de primaria.

Variable	Media	desviación estándar
(X_1)	28.1	5.76
(X_2)	26.6	5.85
(X_3)	35.4	3.82
(X_4)	34.2	5.12
(X_5)	23.6	3.76
(X_6)	22.0	3.93
(X_7)	22.7	4.03

continuación, se presentan los datos para una muestra de $n = 96$ niños de sexto primaria.

Para estos datos de aptitud musical, calcular los intervalos de confianza individuales, simultáneos y de Bonferroni para cada una de las 6 variables. Interprete los resultados, compare los intervalos y explíquelos claramente.