

UNIVERSIDAD NACIONAL DE COLOMBIA
DEPARTAMENTO DE ESTADÍSTICA
ANÁLISIS MULTIVARIADO
Docente: Rubén Darío Guevara G.

Taller 1

Instrucciones

- Elabore un informe que de respuesta a los interrogantes que se presentan a continuación.
- Además de realizar la parte procedimental, tenga claro un marco teórico que soporte dicho procedimiento.
- El informe puede ser grupal, con un número máximo de tres estudiantes.
- El informe debe ser sustentado en un horario que se acordará posteriormente.

Del dataset Paper-Quality Measurements presentado por Johnson & Wichern (2014) en la Tabla 1.2, página 15, encuentre e interprete:

1. El vector de medias muestrales usando los vectores observados y la matriz de datos
2. La matriz de observaciones centradas
3. La matriz de covarianza muestral usando los vectores observados y la matriz de observaciones centradas
4. Varianza total
5. Varianza generalizada
6. La matriz de correlación muestral
7. La matriz de correlación muestral a partir de la matriz de covarianza muestral.
8. Los datos estandarizados (estandarización univariada y multivariada)
9. El vector de media muestral de los datos estandarizados
10. La matriz de covarianza de los datos estandarizados
11. Determine gráficamente si los datos siguen una distribución normal multivariada.
12. Aplique una prueba formal para determinar si los datos siguen una distribución normal p-variada.
13. Encuentre la mediana
14. Realice un bagplot e interprete sus resultados.
15. Encuentre observaciones outliers
 - a. Usando el procedimiento clásico sugerido en el libro de Johnson, R. A., & Wichern, D. W. (2014)
 - b. Usando la distancia euclidiana robusta
 - c. Usando el concepto de profundidad
16. Realice e interprete el gráfico DD plot

REFERENCIAS

- Engel, J., Buydens, L., & Blanchet, L. (2017). An overview of large-dimensional covariance and precision matrix estimators with applications in chemometrics. *Journal of chemometrics*, 31(4), e2880.
- Hubert, M., Debruyne, M., & Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(3), e1421.
- Johnson, R. A., & Wichern, D. W. (2014). *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Prentice hall.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: an R package for assessing multivariate normality. *R J.*, 6(2), 151.
- Peña, D. (2013). *Análisis de datos multivariantes*. Cambridge: McGraw-Hill España.
- Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4), 382-387
- Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics*, 461-482.
- Van Aelst, S., & Rousseeuw, P. (2009). Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 71-82.