

UNIVERSIDAD NACIONAL DE COLOMBIA
DEPARTAMENTO DE ESTADÍSTICA
TEORÍA ESTADÍSTICA MULTIVARIADA, 2025^a.
TALLER 1, PRIMERA PARTE

1. Estudiar el siguiente paper
Hubert, M., & Debruyne, M. (2010). Minimum covariance determinant. *Wiley interdisciplinary reviews: Computational statistics*, 2(1), 36-43.
2. Explique:
 - a. Qué es el efecto “masking” ?
 - b. Qué es el efecto “swamping”
 - c. Por qué los estimadores MCD de localización y dispersión son estimadores robustos.?
 - d. Los supuestos que requieren los estimadores MCD.
 - e. Elipse de tolerancia
 - f. Equivarianza afin de un estimador
 - g. Punto de roptura de un estimador
 - h. Eficiencia del estimador
 - i. El efecto de las observaciones outliers en los valores propios de la matriz de covarianza clásica.
 - j. El papel de los valores propios de la matriz de covarianza clásica en la construcción de las elipses de tolerancia.
 - k. El papel de los valores propios de la matriz de covarianza MCD en la construcción de las elipses de tolerancia.
3. Estudie el artículo de Zuo, Y., & Serfling, R. (2000).
4. Explique las siguientes profundidades: Mahalanobis, Half Space de Tukey y espacial (Spatial Depth).
5. Utilice el data set “wine” usado en Hubert, M., & Debruyne, M. (2010) y encuentre:
 - a. Las observaciones outliers usando los estimadores MCD. Explique sus resultados
 - b. Las observaciones outliers usando las profundidades de Mahalanobis, Half Space y Spatial. Explique sus resultados y compare con los resultados del punto anterior.
6. Del libro de Ojeda Martínez de Castilla, I., & Gago Vargas, J. (2008) estudiar la **Proposición VI.1.2.** (Página 161).
7. Demuestre que la matriz de covarianza muestral es singular en escenarios de alta dimensionalidad.
8. Basado en el artículo de Engel, J., Buydens, L., & Blanchet, L. (2017):
 - a. Explique por qué la matriz de covarianzas muestral no es un buen estimador de la dispersión en ambientes de alta dimensionalidad?

- b. Explique algunas propuestas para estimar la matriz de covarianza muestral en alta dimensionalidad.
9. Basado en el paper de Boudt, K., Rousseeuw, P. J., Vanduffel, S., & Verdonck, T. (2020) use la metodología MRCD sobre el data set VDP, y encuentre:
- a. Las distancias T^2_{MRCD} para 3 observaciones seleccionadas arbitrariamente. Explique sus resultados.
 - b. Las observaciones outliers de este dataset.
10. Utilice la table 1.2 (Paper Quality) del libro de Johnson & Wichern (2014).
- a. Verifique si en este conjunto de datos hay observaciones outliers
 - b. Encuentre estimadores apropiados de localización y dispersión para el vector aleatorio que explica la calidad del papel.

REFERENCIAS

- Boudt, K., Rousseeuw, P. J., Vanduffel, S., & Verdonck, T. (2020). The minimum regularized covariance determinant estimator. *Statistics and Computing*, 30(1), 113-128.
- 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. (2010). Minimum covariance determinant. *Wiley interdisciplinary reviews: Computational statistics*, 2(1), 36-43.
- Johnson, R. A., & Wichern, D. W. (2014). Applied multivariate statistical analysis. Pearson Educacion Limited.
- Ojeda Martínez de Castilla, I., & Gago Vargas, J. (2008). *Métodos matemáticos para estadística*. Universidad de Extremadura, Servicio de Publicaciones.
- Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics*, 461-482.