

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

**FUNDAMENTO TEÓRICO 1ª PARTE**

Las siguientes preguntas tienen el objetivo de ayudarlo a construir un marco teórico básico en análisis multivariado. Si desea puede realizar un informe para Ud, que le ayude a resumir y precisar los conceptos relacionados en las preguntas. No debe entregar informe.

1. Suponga que tiene  $p$  características. Por qué no es adecuado realizar  $p$  análisis univariados independientes?
2. Explique en qué consisten el “masking” y el “swamping”, por qué se presentan y cómo se puede reducir su impacto?
3. Qué ventajas tiene la distancia de Mahalanobis Vs la distancia euclidiana en un análisis estadístico en un análisis estadístico?
4. Demuestre cuál es la representación geométrica de la distancia de Mahalanobis
5. Determine las distribuciones que presenta la distancia de Mahalanobis y en qué escenarios se presentan?
6. Qué es un outlier?
7. Qué deficiencias presentan los estimadores media muestral y matriz de covarianza muestral tradicionales en presencia de outliers? Cómo se pueden superar estas deficiencias?
8. Presente algunos estimadores alternos a los de media muestra y matriz de covarianza que sean robustos a outliers?
9. Qué significa la expresión “La maldición de la dimensionalidad”
10. Qué efectos negativos tiene la alta dimensionalidad en un análisis estadístico?
11. Presente algunos estimadores de localización y dispersión multivariados que presenten mejor desempeño en alta dimensionalidad que los estimadores tradicionales?
12. Con relación a la distribución normal multivariada:
  - a. Indique cuál es su función de densidad y qué papel cumplen la varianza generalizada y la expresión que va en la función exponencial.
  - b. Indique cuáles son sus propiedades
  - c. Explique qué son los contornos, cómo se construyen, qué forma geométrica presentan, y qué distribución presentan?
13. Suponga que se tiene una muestra aleatoria de una distribución normal multivariada:
  - a. Cuáles son los estimadores de máxima verosimilitud?

- b. Cuál es la distribución muestral del vector de media muestral y la matriz de covarianza muestral?
- c. Qué distribución sigue la matriz de covarianza muestral tradicional?
- d. Explique el concepto de estadísticas suficientes y presente algunos ejemplos de ellas.
- e. Qué sucede si utiliza estadísticos basados en la media muestral y en la matriz de covarianza muestral clásicos cuando la distribución de los datos no sigue una distribución normal multivariada?

14. Explique

- a. La ley de los grandes números
- b. El teorema del límite central

15. Explique:

- a. El concepto de profundidad
- b. Qué condiciones debe cumplir una función para que sea una función de profundidad?
- c. Qué es el DD-plot y cómo se interpreta?

## 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.