

UNIVERSIDAD NACIONAL DE COLOMBIA
DEPARTAMENTO DE ESTADÍSTICA
ANÁLISIS DE DATOS FUNCIONALES
Docente: Rubén Darío Guevara G.
TALLER 1

1. El proceso de producción de azúcar proveniente de la remolacha es muy importante en ciertas regiones del mundo como en Escandinavia. En este proceso se emplea la espectrometría para detectar impurezas y controlar la calidad del azúcar. En los siguientes links encuentran la página principal del grupo de Quimiometría del Departamento de Ciencias de los Alimentos de la Universidad de Copenhague, Dinamarca, los datos y su descripción.

http://www.models.kvl.dk/Sugar_Process

<https://ucphchemometrics.com/datasets/>

<https://ucphchemometrics.com/sugar-process-data/>

En particular las curvas espectrométricas corresponden a 7 longitudes de onda de excitación ($Y_1=230$, $Y_2=240$, $Y_3=255$, $Y_4=290$, $Y_5=305$, $Y_6=325$ y $Y_7=340$ nm), donde cada una de estas longitudes representa un proceso funcional continuo (Y_i), así que el proceso multivariado funcional \mathbf{Y} está conformado por $\mathbf{Y}=(Y_1, Y_2, \dots, Y_7)$.

- 1.1. Realice un suavizado spline para cada una de las 268 muestras
- 1.2. Seleccione un proceso cualesquiera (Y_j , $j=1, \dots, 7$) correspondiente a una de las 7 longitudes de onda. Para este proceso seleccionado encuentre
 - La función media
 - La función media recortada al 10%
 - La función varianza
 - La función covarianza
 - La función mediana
 - Siguiendo a López-Pintado, S., & Romo, J. (2009):
 - Ordene las funciones del centro hacia afuera usando las profundidades de banda (BD) y banda modificada (MBD).
 - Siguiendo a Sun, Y., & Genton, M. G. (2011).
 - Encuentre las Funciones Quantiles 0.90 y 0.95
 - Encuentre la Región central 0.75
 - Grafique el boxplot funcional y encuentre las observaciones outliers
 - Siguiendo a Sun, Y., & Genton, M. G. (2012).
 - Grafique el boxplot funcional ajustado e identifique las observaciones outliers.
 - Siguiendo a Huang, H., & Sun, Y. (2019) encuentre las observaciones outliers. ¿Cuáles de ellas corresponden a outliers de forma y cuáles a outliers de magnitud?

- 1.3. De los 7 procesos que conforman el dataset, seleccione 3 procesos cualesquiera
 - Siguiendo a Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014), encuentre la mediana funcional multivariada.
 - Siguiendo a Dai, W., & Genton, M. G. (2019), encuentre los outliers del proceso funcional multivariado
2. Acceda al conjunto de 416 imágenes de resonancia magnética del cerebro (MRI) de personas entre los 18 y 96 años, almacenadas en el sitio www.oasis-brains.org y replique la sección 4 del artículo de Rousseeuw, P. J., Raymaekers, J., & Hubert, M. (2018).

REFERENCIAS

- Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109(505), 411-423.
- Dai, W., & Genton, M. G. (2018). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27(4), 923-934.
- Huang, H., & Sun, Y. (2019). A decomposition of total variation depth for understanding functional outliers. *Technometrics*.
- López-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486), 718-734.
- Rousseeuw, P. J., Raymaekers, J., & Hubert, M. (2018). A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics*, 27(2), 345-359.
- Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of computational and graphical statistics*, 20(2), 316-334.
- Sun, Y., & Genton, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, 23(1), 54-64.

NOTA: En la siguiente referencia se han implementado algunas de las metodologías requeridas para detectar outliers.

- Ojo, O., Lillo, R. E., & Anta, A. F. (2021). Outlier Detection for Functional Data with R Package fdaoutlier. *arXiv preprint arXiv:2105.05213*.
- Pueden encontrar paquetes de R que implementan algunas metodologías escribiendo en el navegador "R-project" nombre del artículo.

- Algunos artículos incorporan links con el código para implementar las metodologías propuestas.
- El taller puede ser desarrollado en cualquier software (Matlab, R, Python).