

UNIVERSIDAD NACIONAL DE COLOMBIA
DEPARTAMENTO DE ESTADÍSTICA
ANÁLISIS DE DATOS FUNCIONALES
TALLER 1

Fecha de entrega y sustentación: Abril 6/2021

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 el siguiente link Ud encuentra los datos y su descripción.

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

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 Y está conformado por $Y=(Y_1, Y_2, \dots, Y_7)$.

1. Usando este dataset seleccione un proceso cualesquiera y para este proceso encuentre
 - La función media
 - La función media recortada al 10%
 - La función varianza
 - La función covarianza
 - La función mediana
 - Funciones Quantiles 90 y 95
 - Región central 0.75
 - Outliers
 - El factor de expansión de 0.5 central región, F , tal que la tasa de falsos outliers sea de 0.007.
2. Usando este mismo dataset, encuentre:
 - La función mediana multivariada
 - Los outliers multivariados.
 - Encuentre la correlación entre los diferentes procesos usando los coeficientes de correlación de Kendall y de Spearman para datos funcionales.
3. No existe una única manera de determinar la profundidad de objetos funcionales, de hecho, en la literatura existe una gran cantidad de propuestas. De acuerdo a la siguiente distribución, estudie la propuesta que le corresponde e impleméntela usando el conjunto de datos asociado a la producción de azúcar:
 - Extremal depth: Estudiantes que reportaron grupo de trabajo
 - Total variation depth: Estudiantes mujeres que no reportaron grupo de trabajo
 - Curve depth: Estudiantes hombres que no reportaron grupo de trabajo
4. De acuerdo a la siguiente distribución, estudie la propuesta que le corresponde e impleméntela usando el conjunto de datos asociado a la producción de azúcar:
 - Dai, W., & Genton, M. G. (2018). Functional boxplots for multivariate curves. *Stat*, 7(1), e190. (Estudiantes hombres que no reportaron grupo de trabajo)

- Dai, W., & Genton, M. G. (2018). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27(4), 923-934. (Estudiantes mujeres que no reportaron grupo de trabajo)
- Dai, W., & Genton, M. G. (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131, 50-65. (Estudiantes que reportaron grupo de trabajo)

REFERENCIAS:

Dai, W., & Genton, M. G. (2018). Functional boxplots for multivariate curves. *Stat*, 7(1), e190.

Dai, W., & Genton, M. G. (2018). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27(4), 923-934.

Dai, W., & Genton, M. G. (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131, 50-65.

de Micheaux, P. L., Mozharovskyi, P., & Vimond, M. (2020). Depth for Curve Data and Applications. *Journal of the American Statistical Association*, 1-17.

Huang, H., & Sun, Y. (2019). A decomposition of total variation depth for understanding functional outliers. *Technometrics*, 61(4), 445-458.

Narisetty, N. N., & Nair, V. N. (2016). Extremal depth for functional data and applications. *Journal of the American Statistical Association*, 111(516), 1705-1714.