

Parcial 1 - Series de tiempo univariadas

IMPORTANTE: Tenga en cuenta lo siguiente:

- El archivo con la solución de este parcial debe entregarse en un único archivo PDF, el cual puede obtenerse en RMarkdown o en Word copiando y pegando imágenes o fotos de la solución a cada punto.
- El archivo debe tener los nombres de los intergantes del grupo de trabajo junto con sus cédulas. Deben marcar el archivo con el nombre parcial_1 seguido de las cédulas de los integrantes separadas por raya al piso. Como ejemplo, **parcial_1_8273827_83784.pdf**.
- El documento entregado debe tener un **máximo de 20 páginas** y por cada página adicional se penalizará reduciendo la nota en un 0.2. El objetivo de esto es forzar a resumir el trabajo elaborado de manera concisa sin incluir códigos, tablas o gráficos innecesarios.
- **LA FECHA MÁXIMA DE ENTREGA ES EL MARTES 13 DE ABRIL DE 2022 A LAS 12:00 DE LA MEDIA NOCHE EN EL LINK QUE APARECE EN MOODLE. NO MODIFICARÉ LA FECHA DE ENTREGA, POR LO CUAL TENGAN EN CUENTA QUE EL CRONÓMETRO QUE APARECE EN MOODLE ES INMODIFICABLE Y LUEGO DE LA FECHA ESTIPULADA NO LOS DEJARÁ SUBIR EL PARCIAL Y LA NOTA SERÁ CERO.**

Plantemiento del parcial:

1. (25%) Considere el proceso estocástico dado por:

$$X_t = w_{t-2} + 0.5w_{t-1} + 2w_t + 0.5w_{t+1} + w_{t+2}$$

donde los w_t son independientes con media cero y varianza $\sigma_w^2 = 4.8$.

- a. Encuentre la media y la varianza del proceso.

- b. Encuentre y grafique las funciones ACF y PACF del proceso.
 - c. Simule y grafique una realización del proceso estocástico X_t de tamaño 200, dando los valores que usted considere apropiados a cada uno de los parámetros.
 - d. Obtenga y grafique la ACF y la PACF muestral de la realización obtenida en (c) y compárela con las gráficas obtenidas en (b). ¿Qué observa?
2. (25%) Considere que el siguiente proceso estocástico es estacionario:

$$X_t = 3.1 + 0.9X_{t-1} - 0.6X_{t-2} + w_t$$

donde w_t es un ruido blanco Gaussiano con media 0 y varianza $\sigma_w^2 = 6.2$.

- a. Encuentre la media del proceso.
 - b. Encuentre la varianza del proceso.
 - c. Encuentre y grafique las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para 20 lags.
 - d. Simule y grafique una realización del proceso estocástico de tamaño 180. Obtenga y grafique la ACF y la PACF muestral de esta realización y compárela con las gráficas obtenidas en (c). ¿Qué observa?
3. (50%) Considere la BD relacionada con la Afluencia de pasajeros en el sistema Metro de Medellín en los años 2019, 2020 y 2021 y que puede descargar en el link del [SITIO WEB](#). Tenga en cuenta que debe descargar los tres archivos .xlsx y en el archivo “Diccionario de datos” explican cada una de las variables.
- a. Lea cuidadosamente en RStudio las 3 bases de datos verificando que no aparezca ningún error o advertencia. ¿Cuáles son las dimensiones de cada data frame?
 - b. Una las tres bases de datos en un solo data frame. Nombrelo “datos_juntos”. ¿Cuáles son las dimensiones del data frame?
 - c. Con base en el dataframe del item (b) elabore otro data frame que contenga las variables (columnas): fecha, hora, día, día de la semana (lunes, martes, ...), semana, mes, año, número de pasajeros por hora, total de pasajeros por día, línea del metro. ¿Cuáles son las dimensiones del data frame?
 - d. Elabore dos data frames: uno con datos de la línea A (nombrelo “dat_lin_A”) y otro con datos de la línea B (nombrelo “dat_lin_B”).

- Ordene los dos data frames de acuerdo a la fecha y hora. ¿Cuáles son las dimensiones de cada data frame?
- e. Para cada línea (A y B), calcule el promedio de pasajeros por hora los lunes, los martes, los miércoles, los jueves, los viernes, los sábados y los domingos antes del 23 de marzo de 2020 (¿qué pasó en esta fecha?) y luego del 23 de marzo de 2020. ¿Qué observa? Incluya gráficos y/o tablas que ayuden a argumentar sus observaciones y comentarios.
 - f. Obtenga dos nuevos data frames (uno para la línea A y otro para la línea B) que resuman el número total de pasajeros por día, que contengan las variables: fecha, día, día de la semana, semana, mes, año y el total de pasajeros por día.
 - g. Para cada data frame del item (f) grafique las series del total de pasajeros por día a lo largo del tiempo. ¿Qué observa? Incluya gráficos y/o tablas que ayuden a argumentar sus observaciones y comentarios.
 - h. Divida cada uno de los dos data frames obtenidos en el item (f) en antes y después del 23 de marzo de 2020. Gráfique y contaste ambos conjuntos para cada línea. ¿Observa algún comportamiento estacional o tendencia? Argumente con gráficos explicando cada uno.
 - i. Usando la función “lm” del RStudio, ajuste un modelo para los datos antes y otro para los datos después en cada línea que explique el número de pasajeros. Argumente por qué selecciona cada covariable y explique los resultados del “summary” de cada uno de los cuatro modelos contrastando los resultados del antes y el después en cada línea.