Modelos de supervivencia y de series de tiempo

Tarea 4

Fecha de entrega: 4 de diciembre

Instrucciones. La tarea se puede realizar en equipos de máximo 3 integrantes. Se entregará de forma física (o digital) a la hora de la clase y no habrá prórroga. El código utilizado para realizar los análisis no deberá incluirse en el reporte y se entregará en un archivo separado. El código deberá permitir reproducir los resultados numéricos y las gráficas presentadas en la tarea. Las actividades de DataCamp son individuales y podrán terminarse hasta el final de la fecha de entrega de la tarea.

1. (2 puntos) Considera el proceso estocástico X_t definido como

$$X_t = a + bt + Y_t,$$

donde Y_t es un ruido blanco de media cero y varianza σ^2.

(a) Si $m_t = a + bt$, demuestra que el filtro de promedios móviles centrados de orden q, con pesos $(2q+1)^{-1}$, deja intacto a m_t, esto es,

$$m_t = \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t-j}.$$

(b) Si ahora se define

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j},$$

obtén la media y la función de autocovarianza de W_t. ¿Es estacionario el proceso W_t? Justifica tu respuesta.

(c) Construye una serie de tiempo (fijando una semilla para efectos de reproducibilidad) de tamaño 100, con un componente de tendencia dado por $m_t = 1 + 2t$ y $Y_t \sim \mathcal{N}(0, 100^2)$. Realiza lo siguiente:

i. Grafica la serie y calcula su media muestral y su función de autocorrelación muestral.

ii. Aplica filtros de promedios móviles centrados con $q = 1, 2, 3$, utilizando únicamente las observaciones para las cuales cada filtro está completamente definido. Grafica las series filtradas y calcula sus medias muestrales. Compáralas con las medias muestrales de las observaciones correspondientes de la serie original. ¿Qué puedes observar? ¿Es consistente con los resultados teóricos obtenidos anteriormente?

iii. Obtén los residuales $\widehat{Y}_t = X_t - W_t$ para cada valor de q. Grafica los residuales y sus funciones de autocorrelación muestrales. ¿Existe evidencia de que sean estacionarios? ¿Podemos considerarlos ruido blanco? Justifica tus respuestas.

2. (3.5 puntos) Considera la base de datos *tempNottingham*, que contiene las temperaturas promedio mensuales de la ciudad de Nottingham de 1920 a 1939.

   (a) Grafica los datos, así como la ACF y PACF de la serie. ¿Consideras que la serie es estacionaria? Si no la consideras estacionaria, explica por qué.

   (b) Grafica el periodograma de la serie e identifica las frecuencias dominantes. Determina el periodo estacional s y justifica tu elección.

   (c) Considera el proceso $Y_t = (1 - B^s)X_t$ y grafica su ACF y PACF. ¿Consideras que el proceso resultante es estacionario? ¿Qué modelos ARMA podrían ser adecuados para Y_t? Justifica tu respuesta.

   (d) Ajusta un modelo SARIMA $(1, 0, 1) \times (1, 1, 1)_s$ a los datos originales y realiza el análisis de residuales. ¿Qué puedes decir del ajuste?

   (e) Ajusta el modelo propuesto por la función `auto.arima` y realiza el análisis de residuales. Compara este modelo con el SARIMA ajustado en el inciso anterior, utilizando los criterios AIC y BIC, así como el análisis de residuales. ¿Con cuál de los dos modelos te quedarías? Justifica tu respuesta.

   (f) A partir del modelo elegido, realiza las predicciones para los siguientes 12 meses. Grafica los pronósticos junto con sus intervalos de predicción al 95%. ¿Qué puedes decir de los resultados?

   (g) Finalmente, utiliza la función `ets` de la paquetería `forecast` de R. Analiza los residuales resultantes y realiza las predicciones para los siguientes 12 meses. Compara los pronósticos obtenidos con los del modelo seleccionado anteriormente. ¿Qué diferencias observas?

3. (3.5 puntos) Considera la base de datos *co2*, que contiene los niveles de dióxido de carbono en la ciudad de Alert, Canadá, de enero de 1994 a diciembre de 2004.

Divide la serie en un conjunto de entrenamiento (las primeras 120 observaciones) y uno de prueba (las últimas 12 observaciones). Para los siguientes incisos, utiliza únicamente el conjunto de entrenamiento, salvo que se indique lo contrario.

(a) Grafica la serie de entrenamiento, así como sus funciones de autocorrelación (ACF) y autocorrelación parcial (PACF). ¿Consideras que la serie es estacionaria? Justifica tu respuesta.

(b) Grafica el periodograma de la serie e identifica las frecuencias dominantes. ¿La frecuencia correspondiente al máximo del periodograma representa un ciclo estacional interpretable? En caso contrario, identifica la frecuencia asociada al ciclo estacional de la serie y determina su periodo s.

(c) Considera el proceso $Y_t = (1 - B^s)X_t$ y grafica su ACF y PACF. ¿Consideras que el proceso resultante es estacionario? ¿Qué modelos ARMA podrían ser adecuados para Y_t? Justifica tu respuesta.

(d) Ajusta un modelo SARIMA $(1, 1, 1) \times (1, 1, 1)_s$ a la serie de entrenamiento y realiza el análisis de residuales. ¿Qué puedes decir del ajuste?

(e) Utiliza la función `auto.arima` para seleccionar y ajustar un modelo alternativo a la serie de entrenamiento. Realiza el análisis de residuales y compara ambos modelos utilizando los criterios AIC y BIC. ¿Cuál elegirías con base en estos criterios y el análisis de residuales?

(f) Utilizando los dos modelos ajustados, obtén las predicciones para las 12 observaciones del conjunto de prueba y calcula sus respectivos RMSE. ¿Cuál presenta un mejor desempeño predictivo? ¿Coincide con tu elección del inciso anterior?

(g) Reajusta el modelo con menor RMSE utilizando la serie completa, conservando los órdenes del modelo seleccionado. Realiza las predicciones para los siguientes 12 meses e incluye sus intervalos de predicción al 95%. Interpreta los resultados.

(h) Finalmente, utiliza la función `ets` de la paquetería `forecast` de R para ajustar un modelo a la serie completa. Analiza sus residuales y realiza las predicciones para los siguientes 12 meses. Compara estos pronósticos con los obtenidos mediante el modelo SARIMA seleccionado.

Actividades de DataCamp

1. (0.5 puntos) ARIMA Models in R.
2. (0.5 puntos) Forecasting in R.