

Modelado y estudio de la serie temporal de los datos de las CCAA de España.

Instrucciones

El fichero `DatosCCAA.csv` que tienes como recurso descargable contiene los datos del número de infectados (casos), fallecidos, hospitalizados y hospitalizados en las UCI de las 17 comunidades autónomas de España excepto Ceuta y Melilla.

Cárgalo en la variable `datos` y échale un vistazo con la función `View`.

Descripción de la tarea

En esta tarea vamos a modelar los infectados, fallecidos y hospitalizados de coronavirus de las comunidades autónomas de España.

Pregunta 1

Usando el paquete `tidyverse`, filtra los datos del fichero que has cargado seleccionando sólo las variables **fecha, CCAA, total y tipo**.

A partir de la nueva tabla de datos, vuelve a filtrar los datos seleccionando sólo los datos de la variable total de la comunidad de Madrid sólo del tipo “casos”.

Pregunta 2

Modeliza el número de infectados de la comunidad de Madrid como una serie temporal indicando el día de la semana que empieza la serie.

Haz también un gráfico de la serie temporal usando la función `autoplot` del paquete `forecast`.

Pregunta 3

Suponiendo que la serie que nos da el número de infectados de la comunidad de Madrid es estacional y suponiendo el modelo aditivo, calcula las tres componentes de la serie temporal, componente estacional, tendencia y aleatoria. Comenta los días de la semana en que según la componente estacional hay más infectados y los días de la semana en que hay menos.

Haz un gráfico de las componentes de la serie comentando los resultados.

Pregunta 4

Ajusta la serie temporal que nos da el número de infectados de la comunidad de Madrid eliminando la componente estacional y haz un gráfico de la misma.

Usando el doble suavizado exponencial del modelo de Holt-Winters, calcula la serie predicha comentando los valores de los parámetros α y β .

Calcula el error SSE cometido y su raíz cuadrada de la serie predicha y haz un gráfico de la serie original ajustada y la predicha.

Pregunta 5

A partir de la serie ajustada que nos da el número de infectados de la comunidad de Madrid, calcula la predicción del número de infectados a una semana vista junto con los intervalos de confianza al 80 y al 95%. Haz un gráfico de la predicción anterior mostrando los intervalos de confianza anteriores.

Pregunta 6

Testea las hipótesis siguientes sobre los residuos de la serie predicha a una semana vista de la serie ajustada que nos da el número de infectados de la comunidad de Madrid:

- los residuos no presentan autocorrelaciones a partir del gráfico del correlograma y del test de Ljung-Box.
- los residuos son normales.

Comenta los resultados obtenidos.