

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”.

Solución

Primero cargamos la librería `tidyverse`:

```
library(tidyverse)
```

Los datos se cargarían con:

```
datos = read.csv("DatosCCAA.csv")
```

El primer filtrado pedido sería:

```
datos.filtrados = data_long %>% select(fecha, CCAA, total, tipo)
```

Los datos de la comunidad de Madrid pedidos serían:

```
casos.Madrid = datos.filtrados %>%  
  filter(tipo=="casos") %>%  
  filter(CCAA == "Madrid") %>%  
  select(total)
```

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`.

Solución

Las funciones a aplicar serían:

```
ts.casos.Madrid = ts(casos.Madrid, frequency = 7, start=c(1,4))  
# ya que la serie empieza el 27 de febrero, un jueves.  
  
library(forecast)  
  
autoplot(ts.casos.Madrid)
```

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.

Solución

```
components = decompose(ts.casos.Madrid, type="additive")  
  
autoplot(components)  
  
components$seasonal
```

Vemos que los días según la componente estacional con más infectados serían los jueves, viernes y sábados y los días con menos infectados, los domingos, lunes y martes.

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 alpha y beta.

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.

Solución

La serie ajustado y el gráfico pedido sería:

```
casos.ajustados = ts.casos.Madrid-components$seasonal  
  
autoplot(casos.ajustados)
```

La predicción según el doble suavizado exponencial del modelo de Holt-Winters sería:

```
predicción.casos=HoltWinters(casos.ajustados, gamma=FALSE)
```

El valor de $\alpha = 1$, lo que significa que la tendencia sólo usa el valor actual de la serie ajustada y el valor de beta es aproximadamente 0.5, lo que significa que la predicción la pendiente de la tendencia con un peso de 0.56.

El gráfico de la serie original y predicha es:

```
plot(predicción.casos)
```

El valor de SSE vale:

```
predicción.casos$SSE
```

```
sqrt(predicción.casos$SSE)
```

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.

Solución

La predicción pedida es la siguiente:

```
predicción.casos.semana = forecast::forecast.HoltWinters(predicción.casos, h=7)
```

El gráfico es el siguiente:

```
forecast::autoplot(predicción.casos.semana)
```

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.

Solución

El correlograma es el siguiente:

```
ggAcf(predicción.casos.semana$residuals[3:75], lag.max=7)
```

donde vemos la autocorrelación de los residuos con un lag=7 es significativa.

Si hacemos un test de Ljung-Box, obtenemos un p-valor en la zona de penumbra:

```
Box.test(predicción.casos.semana$residuals, lag=7, type="Ljung-Box")
```

Por último, observamos que los residuos no son normales, indicándonos que el modelo no es correcto:

```
shapiro.test(predicción.casos.semana$residuals)
```