PROG8430

Fei Yun

8680643

Assignment 3

Data Transformation

1. Read in the Welland data and transform it into an appropriate time

series datatype.

#read data

Temperatue<-read.csv('Welland20F.csv',header = TRUE)

#convert data to time series

tsWelland\_FY<-ts(Temperatue$Temp,frequency = 12,start = c(1985,1))

head(tsWelland\_FY)

[1] -6.616129 -4.328571 1.367742 8.926667 14.151613 16.450000

#Descriptive Data Analysis

1. Summarize the precipitation information (mean, std dev, etc.)

summary(tsWelland\_FY)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-9.71290 -0.07903 8.93269 8.46305 17.00032 22.63871

2. Plot the time series data and make note of anything significant you observe.

plot.ts(tsWelland\_FY,main='Averange Tempreature of Welland',ylim=c(-10,25))

A picture containing chart

Description automatically generated

3. Decompose the times series data in to the constituent components.Comment on each (any trends you observe, etc.)

decoWelland\_FY<-decompose(tsWelland\_FY,type = 'additive')

decoWelland\_FY

plot(decoWelland\_FY)

A picture containing diagram

Description automatically generated

we can see the seasonal data amplitude doesn't change with times(year) change

the seasonal data amplitude is almost same every year so additive model is the best

4. Determine if the time series is stationary.

library(tseries)

adf.test(tsWelland\_FY)

Augmented Dickey-Fuller Test

data: tsWelland\_FY

Dickey-Fuller = -13.342, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

5. Deseasonalize the information and plot the result.

seasonal\_adj\_Welland\_FY<-tsWelland\_FY-decoWelland\_FY$seasonal

plot.ts(seasonal\_adj\_Welland\_FY,main='Deseasonal Temperature of Welland',ylim=c(-10,25))

Chart

Description automatically generated

6. Add any comments about what you observe: seasonality of precipitation, trends, etc.

I observe trend at 1991-1992，the temperature was going to a big increase

Data Transformation

1. Read in the Waterloo data and transform it into an appropriate time

series datatype.

#read data

Precip<-read.csv('Waterloo20F.csv',header = TRUE)

head(Precip)

#convert data to time series

tsWaterloo\_FY<-ts(Precip$Precip,frequency = 1,start = c(1970))

head(tsWaterloo\_FY)

[1] 731.3 707.7 899.5 847.9 846.4 999.8

Descriptive Data Analysis

1. Summarize the information (mean, std dev, etc.)

library(pastecs)

stat.desc(tsWaterloo\_FY)

nbr.val 26.00000

nbr.null 0.00000

nbr.na 0.00000

min 674.80000

max 1186.40000

range 511.60000

sum 23648.20000

median 936.10000

mean 909.54615

SE.mean 24.39122

CI.mean.0.95 50.23467

var 15468.22738

std.dev 124.37133

coef.var 0.13674

3. Smooth the precipitation chart using a moving average. Try 3 different

values for the moving average and choose the one you think best shows the trend (if any).

library(TTR)

WaterlooSMA10\_FY<-SMA(tsWaterloo\_FY,n=10)

plot.ts(WaterlooSMA10,main='Total Precipitation of Waterloo',ylim=c(670,1200))

Chart, line chart

Description automatically generated

this lost too much details

Chart, line chart, histogram

Description automatically generated

this has details and can clear its trends and moving

Chart, line chart

Description automatically generated

this can not identify its trends

4. Determine if the time series is stationary.

library(tseries)

adf.test(tsWaterloo\_FY)

Augmented Dickey-Fuller Test

data: tsWaterloo\_FY

Dickey-Fuller = -2.5257, Lag order = 2, p-value = 0.3721

alternative hypothesis: stationary

5. Create an autocorrelation chart (using acf) and comment on which lags are significant.

#Do previous values seem to influence current values?

acf(tsWaterloo\_FY)

Chart

Description automatically generated

lag3,lag8,lag11 are significant,seems previous values does not influence current values

3

1. Create a simple moving average forecast of precipitation in Waterloo

for five years beyond the data provided. Graph your results along with a 75% prediction interval.

library(smooth)

smaWaterloo\_FY<-sma(tsWaterloo\_FY)

smaWaterloo\_FY

smaWaterloo\_FY<-forecast(smaWaterloo\_FY,h=5,level=0.75)

plot(smaWaterloo\_FY)

Time elapsed: 0.04 seconds

Model estimated: SMA(10)

Initial values were produced using backcasting.

Loss function type: MSE; Loss function value: 15397.7461

Error standard deviation: 129.1545

Sample size: 26

Number of estimated parameters: 2

Number of degrees of freedom: 24

Information criteria:

AIC AICc BIC BICc

328.4762 328.9979 330.9924 331.8423

Chart, line chart, histogram

Description automatically generated

2. Create an exponentially smoothed forecast of precipitation in Waterloo

for five years beyond the data provided. Graph your results along with a 75% prediction interval.

exsWaterloo\_FY<-es(tsWaterloo\_FY)

exsWaterloo\_FY

exsWaterloo\_FY<-forecast(exsWaterloo\_FY,h=5,level = 0.75)

plot(exsWaterloo\_FY)

Time elapsed: 0.27 seconds

Model estimated: ETS(ANN)

Persistence vector g:

alpha

0

Initial values were optimised.

Loss function type: MSE; Loss function value: 14873.2956

Error standard deviation: 126.936

Sample size: 26

Number of estimated parameters: 2

Number of provided parameters: 1

Number of degrees of freedom: 24

Information criteria:

AIC AICc BIC BICc

327.5752 328.0969 330.0914 330.9413

Chart, histogram

Description automatically generated

3. Compare the two forecasts you created in steps 1 and 2 above.

smaWaterloo\_FY

Time Series:

Start = 1996

End = 2000

Frequency = 1

Point forecast Lower bound (12.5%) Upper bound (87.5%)

1996 913.6000 761.3417 1065.858

1997 898.5000 745.4823 1051.518

1998 893.9700 740.0385 1047.902

1999 888.8970 733.8669 1043.927

2000 910.3067 753.9577 1066.656

exsWaterloo\_FY

Time Series:

Start = 1996

End = 2000

Frequency = 1

Point forecast Lower bound (12.5%) Upper bound (87.5%)

1996 909.5462 763.072 1056.02

1997 909.5462 763.072 1056.02

1998 909.5462 763.072 1056.02

1999 909.5462 763.072 1056.02

2000 909.5462 763.072 1056.02

We can see the forecast based on simple moving ,the precipitation was between 889 to 914 and based on exponetially smoothed forecast the precipitation was just at 909

the actual precipitation of 1996-2000 is 1043.0,865.7,656.5,811.4,933.9

the simple moving forecast has higher accuracy,which is better model,because simple moving calculate the average of data,but exponential gives more weight to recent data, here the precipitation is not influenced by previous year's data,so simple moving is better .