MODERN COLLEGE OF ARTS, SCI & COMM. PUNE-05.

DEPARTMENT OF STATISTICS.

EXPT.NO. 1

Title: Smoothing the series using various Filters and Estimation of trend and seasonal component.

Q.1 Consider the following data related to monthly accidental death in India for the year 2005-2010.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| Jan | 9007 | 7750 | 8162 | 7717 | 7792 | 7836 |
| Feb | 8106 | 6981 | 7306 | 7461 | 6957 | 6892 |
| Mar | 8928 | 8038 | 8124 | 7776 | 7726 | 7791 |
| Apr | 9137 | 8422 | 7870 | 7925 | 8106 | 8129 |
| May | 10017 | 8714 | 9387 | 8634 | 8890 | 9115 |
| June | 10826 | 9512 | 9556 | 8945 | 9299 | 9434 |
| July | 11317 | 10120 | 10093 | 10078 | 10625 | 10484 |
| Aug | 10744 | 9823 | 9620 | 9179 | 9302 | 9827 |
| Sept | 9713 | 8743 | 8285 | 8037 | 8314 | 9110 |
| Oct | 9938 | 9129 | 8433 | 8488 | 8850 | 9070 |
| Nov | 9161 | 8710 | 8160 | 7874 | 8265 | 8633 |
| Dec | 8927 | 8680 | 8034 | 8647 | 8796 | 9240 |

1. Plot the group of monthly accidental deaths in India for 2005-2010.
2. Estimate trend.
3. Plot the graph of monthly accidental deaths after subtracting the trend estimated by moving average.
4. Estimate seasonal component of monthly accidental deaths.
5. Plot the graph of the detrended and deseasonlised monthly accidental deaths.
6. Plot the graph of differential series { , Xt t= 13... 72} derived from monthly accidental deaths.

Q.2 Consider the following time series observations (30 time points in rows)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 486 | 474 | 434 | 441 | 435 | 401 | 414 | 414 | 386 | 405 |
| 411 | 389 | 414 | 426 | 410 | 441 | 459 | 449 | 486 | 510 |
| 506 | 549 | 579 | 581 | 630 | 666 | 674 | 729 | 771 | 785 |

1. Identify the nature of trend and seasonal component by plotting the data.
2. Apply the filter [a-2, a-1, a0, a1, a2] = [-1, 4, 3, 4, -1] \* 1/9 and discuss the results.
3. Calculate the mean absolute deviation (MAD) and mean square deviation (MSD) for the fitted model.

Q.3 Consider the data AIRPASS.TSM from ITSM and check smoothing using

Box-COX transformation, check stationary by difference of lag 1 and check

Normality.

Q.1) Q.1 Consider the following data related to monthly accidental death in India for the year 2005-2010.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| Jan | 9007 | 7750 | 8162 | 7717 | 7792 | 7836 |
| Feb | 8106 | 6981 | 7306 | 7461 | 6957 | 6892 |
| Mar | 8928 | 8038 | 8124 | 7776 | 7726 | 7791 |
| Apr | 9137 | 8422 | 7870 | 7925 | 8106 | 8129 |
| May | 10017 | 8714 | 9387 | 8634 | 8890 | 9115 |
| June | 10826 | 9512 | 9556 | 8945 | 9299 | 9434 |
| July | 11317 | 10120 | 10093 | 10078 | 10625 | 10484 |
| Aug | 10744 | 9823 | 9620 | 9179 | 9302 | 9827 |
| Sept | 9713 | 8743 | 8285 | 8037 | 8314 | 9110 |
| Oct | 9938 | 9129 | 8433 | 8488 | 8850 | 9070 |
| Nov | 9161 | 8710 | 8160 | 7874 | 8265 | 8633 |
| Dec | 8927 | 8680 | 8034 | 8647 | 8796 | 9240 |

1. Plot the group of monthly accidental deaths in India for 2005-2010.
2. Plot the group of monthly accidental deaths in India for 2005-2010.

> library(readxl)

>library(forecast)

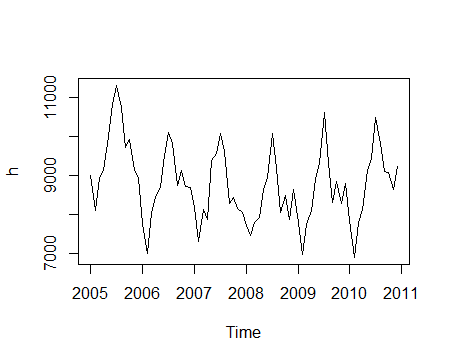
>library(tseries)

>

> t=scan(‘clipboard’)

> deaths<-ts(data=t,start=c(2005,1), frequency=12)

> plot.ts(deaths)



Interpretation: Plot shows a decreasing trend initially then it increases slowly. Also seasonality can be seen from the plot.

1. Estimate trend.

> library(forecast)

> death\_trends<-ma(deaths,order=12,centre=TRUE) ####trend estimated by moving averages

> death\_trends

Jan Feb Mar Apr May Jun Jul Aug Sep Oct

2005 NA NA NA NA NA NA 9599.375 9500.125 9416.167 9349.292

2006 9051.542 8963.292 8884.500 8810.375 8757.875 8728.792 8735.667 8766.375 8783.500 8764.083

2007 8799.708 8790.125 8762.583 8714.500 8662.583 8612.750 8567.292 8555.208 8547.167 8534.958

2008 8422.958 8403.958 8375.250 8367.208 8357.583 8371.208 8399.875 8382.000 8358.917 8364.375

2009 8445.542 8473.458 8490.125 8516.750 8548.125 8570.625 8578.667 8577.792 8577.792 8581.458

2010 8606.542 8622.542 8677.583 8719.917 8744.417 8778.250 NA NA NA NA

Nov Dec

2005 9265.208 9156.167

2006 8769.125 8799.000

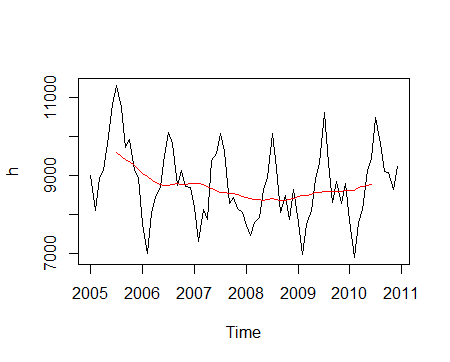
2007 8505.875 8449.042

2008 8382.583 8408.000

2009 8591.792 8606.792

2010 NA NA

> lines(death\_trends,col="red")



Interpretation:

1. Plot the graph of monthly accidental deaths after subtracting the trend estimated by moving average.

> detrend\_deaths<-deaths-death\_trends

> detrend\_deaths

Jan Feb Mar Apr May Jun Jul Aug

2005 NA NA NA NA NA NA 1717.6250 1243.8750

2006 -1301.5417 -1982.2917 -846.5000 -388.3750 -43.8750 783.2083 1384.3333 1056.6250

2007 -637.7083 -1484.1250 -638.5833 -844.5000 724.4167 943.2500 1525.7083 1064.7917

2008 -705.9583 -942.9583 -599.2500 -442.2083 276.4167 573.7917 1678.1250 797.0000

2009 -653.5417 -1516.4583 -764.1250 -410.7500 341.8750 728.3750 2046.3333 724.2083

2010 -770.5417 -1730.5417 -886.5833 -590.9167 370.5833 655.7500 NA NA

Sep Oct Nov Dec

2005 296.8333 588.7083 -104.2083 -229.1667

2006 -40.5000 364.9167 -59.1250 -119.0000

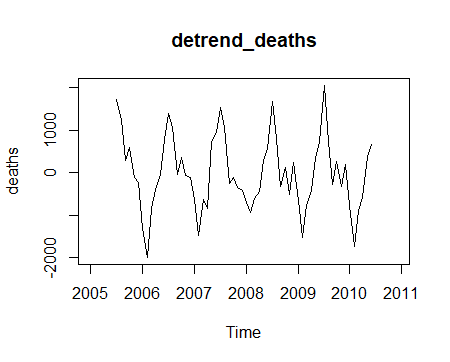
2007 -262.1667 -101.9583 -345.8750 -415.0417

2008 -321.9167 123.6250 -508.5833 239.0000

2009 -263.7917 268.5417 -326.7917 189.2083

2010 NA NA NA NA

> plot.ts(detrend\_deaths)



Interpretation: We can see that there no trend effect in above plot. Thus we have successfully deleted the trend component. But still we can say seasonality present in the dat4.

1. Estimate seasonal component of monthly accidental deaths.

> d<-decompose(deaths)

> ### classical seasonal decomposition by moving average

> est\_seas<-d$figure

> est\_seas

[1] -804.31944 -1521.73611 -737.46944 -525.81111 343.42222 746.41389 1679.96389 986.83889

[9] -108.76944 258.30556 -259.37778 -57.46111

1. Plot the graph of the detrended and deseasonlised monthly accidental deaths.

> deseas\_deaths<-deaths-d$seasonal-d$trend

> deseas\_deaths

Jan Feb Mar Apr May Jun Jul Aug

2005 NA NA NA NA NA NA 37.661111 257.036111

2006 -497.222222 -460.555556 -109.030556 137.436111 -387.297222 36.794444 -295.630556 69.786111

2007 166.611111 37.611111 98.886111 -318.688889 380.994444 196.836111 -154.255556 77.952778

2008 98.361111 578.777778 138.219444 83.602778 -67.005556 -172.622222 -1.838889 -189.838889

2009 150.777778 5.277778 -26.655556 115.061111 -1.547222 -18.038889 366.369444 -262.630556

2010 33.777778 -208.805556 -149.113889 -65.105556 27.161111 -90.663889 NA NA

Sep Oct Nov Dec

2005 405.602778 330.402778 155.169444 -171.705556

2006 68.269444 106.611111 200.252778 -61.538889

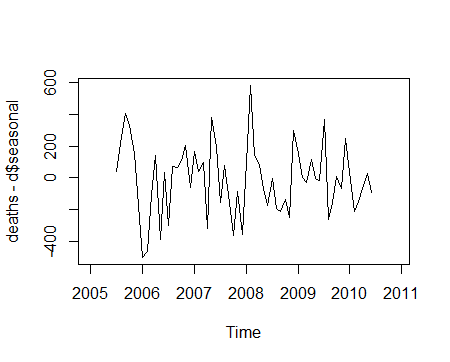
2007 -153.397222 -360.263889 -86.497222 -357.580556

2008 -213.147222 -134.680556 -249.205556 296.461111

2009 -155.022222 10.236111 -67.413889 246.669444

2010 NA NA NA NA

> plot.ts(deseas\_deaths)

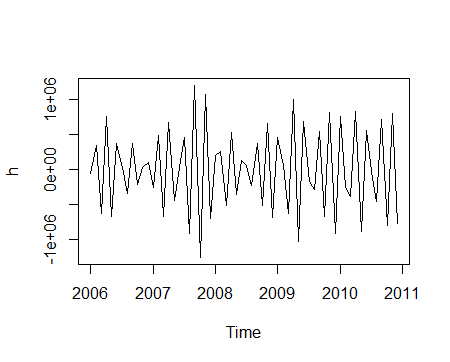


Interpretation: We can see we have successfully removed the seasonal components from the time series.

1. Plot the graph of differential series { , Xt t= 13... 72} derived from monthly accidental deaths.

> diff\_deaths<-diff(deaths,difference=12)

> plot.ts(diff\_deaths)



Interpretation: Plot represents a fully decomposed time series without seasonal and trend components.

> Box.test(diff\_deaths,type="Ljung-Box")

Box-Ljung test

data: diff\_deaths

X-squared = 46.805, df = 1, p-value = 7.84e-12

H0: There is no autocorrelation present ρi = 0

V/s H1: There is autocorrelation present for at least one i ρi ≠ 0

Here p-value=0.044 < 0.05 we reject H0 that means conclude that, there is autocorrelation present for at least one i

> adf.test(diff\_deaths,alternative="stationary")

Augmented Dickey-Fuller Test

data: diff\_deaths

Dickey-Fuller = -19.949, Lag order = 3, p-value = 0.01

alternative hypothesis: stationary

Warning message:

In adf.test(diff\_deaths, alternative = "stationary") :

p-value smaller than printed p-value

H0: Process is not stationary

V/s H1: Process is stationary

Here p-value = 0.01 Hence reject H0 and conclude that process is stationary.

> kpss.test(diff\_deaths)

KPSS Test for Level Stationarity

data: diff\_deaths

KPSS Level = 0.036309, Truncation lag parameter = 3, p-value = 0.1

Warning message:

In kpss.test(diff\_deaths) : p-value greater than printed p-value

H0: Process is not stationary

V/s H1: Process is stationary

Here p-value = 0.01 Hence reject H0 and conclude that process is stationary.

Q.2) Q.2 Consider the following time series observations (30 time points in rows)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 486 | 474 | 434 | 441 | 435 | 401 | 414 | 414 | 386 | 405 |
| 411 | 389 | 414 | 426 | 410 | 441 | 459 | 449 | 486 | 510 |
| 506 | 549 | 579 | 581 | 630 | 666 | 674 | 729 | 771 | 785 |

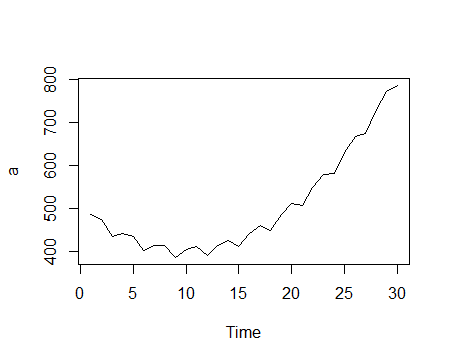
1. Identify the nature of trend and seasonal component by plotting the data.
2. Apply the filter [a-2, a-1, a0, a1, a2] = [-1, 4, 3, 4, -1] \* 1/9 and discuss the results.
3. Calculate the mean absolute deviation (MAD) and mean square deviation (MSD) for the fitted model.
4. Identify the nature of trend and seasonal component by plotting the data.

> library(readxl)

> v=scan(‘clipboard’)

> t=ts(v)

> plot(t)

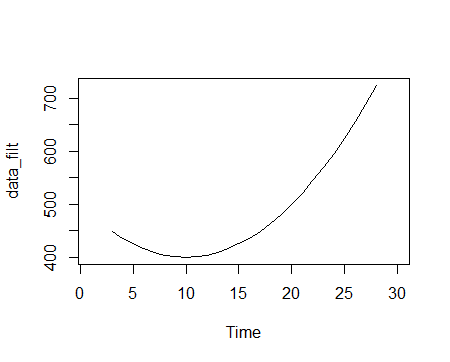


Interpretation: We can see there is an increasing trend and absence of seasonal components.

1. Apply the filter [a-2, a-1, a0, a1, a2] = [-1, 4, 3, 4, -1] \* 1/9 and discuss the results.

> data\_filter=filter(t,filter=c(-1/9,4/9,3/9,4/9,-1/9),method="convolution")

> plot.ts(data\_filter)



Interpretation: As we can see a perfectly smooth increasing curve. We can conclude that data is filtered.

1. Calculate the mean absolute deviation (MAD) and mean square deviation (MSD) for the fitted model.

> double\_exp<- HoltWinters(t,gamma=FALSE)

####to fit the double exp (absence of seasonal effects)

> data1=cbind(t,double\_exp$fitted[,1])

> MAD=mean(abs(data1[,1]-data1[,2]),na.rm=TRUE) # Mean Absolute Deviation

> MAD

[1] 15.83333

> MSD=mean(abs((data1[,1]-data1[,2])^2),na.rm=TRUE) # Mean Square Deviation

> MSD

[1] 271.9022

Q.3) Q.3 Consider the data AIRPASS.TSM from ITSM and check smoothing using

Box-COX transformation, check stationary by difference of lag 1 and check

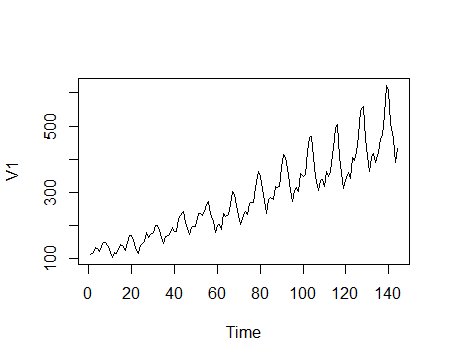
Normality.

library(tseries)

library(forecast)

> a=(airpass)

> plot.ts(a)



Interpretation: We can see a clear increasing trend with non-constant variance.

> A=ts(a)

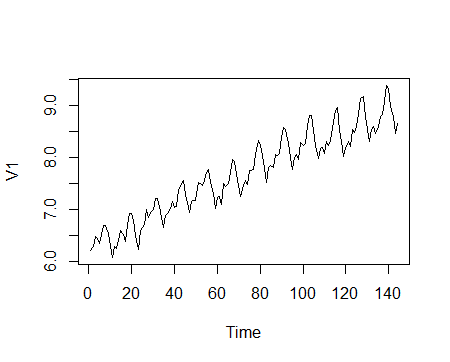
> lambda<-BoxCox.lambda(A)

> lambda

[1] 0.1107583

> trans\_data<-BoxCox(A,lambda)

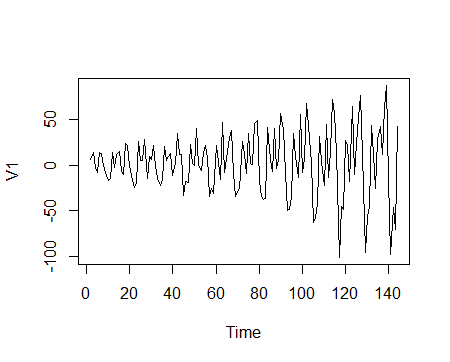
> plot.ts(trans\_data)



Interpretation: Here we can see a series generated with constant variance.

> diff\_airpass<-diff(A,differenc=1) #### lag differencing

> plot.ts(diff\_airpass)



Interpretation: Given plot shows we have obtained a stationary Time series.

> kpss.test(diff\_airpass)

KPSS Test for Level Stationarity

data: diff\_airpass

KPSS Level = 0.014626, Truncation lag parameter = 4, p-value = 0.1

Warning message:

In kpss.test(diff\_airpass) : p-value greater than printed p-value

H0: Process is not stationary

V/s H1: Process is stationary

Here p-value = 0.01 Hence reject H0 and conclude that process is stationary.

> Box.test(diff\_airpass)

Box-Pierce test

data: diff\_airpass

X-squared = 13.116, df = 1, p-value = 0.0002928

H0: There is no autocorrelation present ρi = 0

V/s H1: There is autocorrelation present for at least one i ρi ≠ 0

Here p-value < 0.01 we reject H0 that means conclude that, there is autocorrelation present for at least one i

> adf.test(diff\_airpass, alternative="stationary")

Augmented Dickey-Fuller Test

data: diff\_airpass

Dickey-Fuller = -7.0177, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

Warning message:

In adf.test(diff\_airpass, alternative = "stationary") :

p-value smaller than printed p-value

H0: Process is not stationary

V/s H1: Process is stationary

Here p-value = 0.01 Hence reject H0 and conclude that process is stationary.