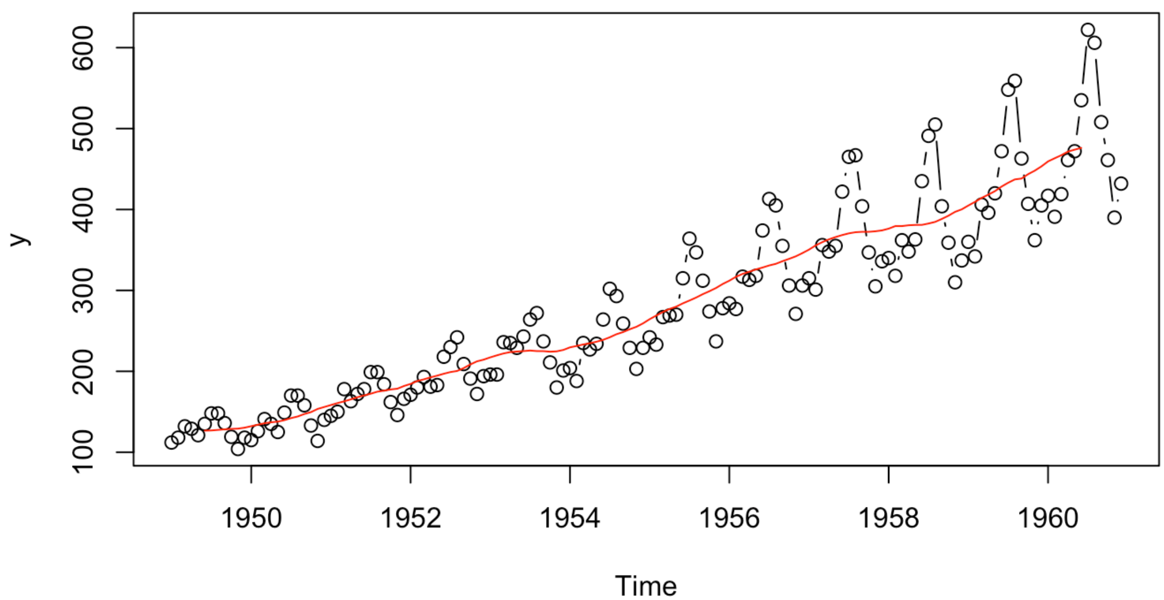
Predictive Analytics II: Assignment #4

Question 1

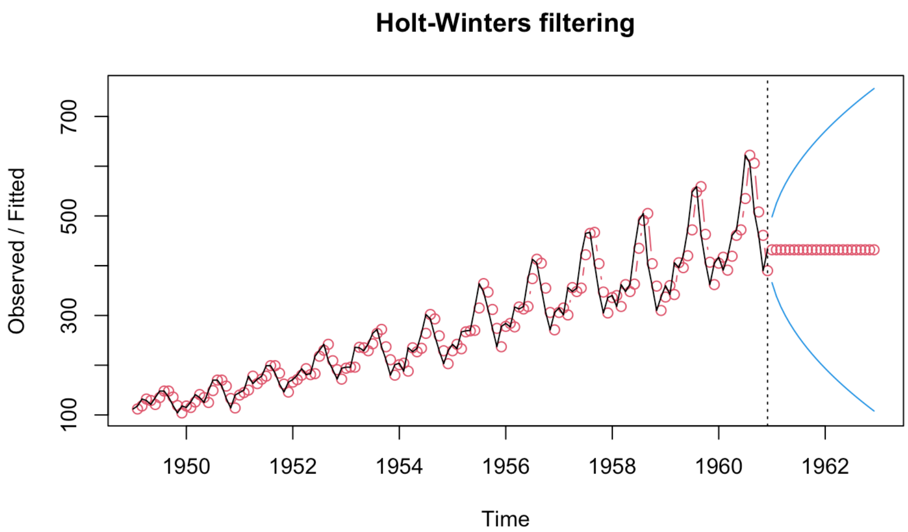
a) I used a centered moving average filter of 12 months to smooth out seasonality in the data since it is a monthly data of airline passengers.

Below is the plot visualizing the original time series and the red line indicates the time series after smoothening out seasonal components.



b) Below is the plot visualizing the EWMA smoothened curve. The black line is the original time series and red line is EWMA smoothened time series. The extrapolated plot beyond the dotted line is the 2-year forecasted value.

The optimal α = 1.00. From the EWMA equations, for α = 1.00, almost all the weight is given to just the previous time step with very little or no weight to all the past time steps. That is, the best forecast of future values is the last observed time series value $\hat{y}_{t+k|t} = L_t = y_t$ (Current Level)

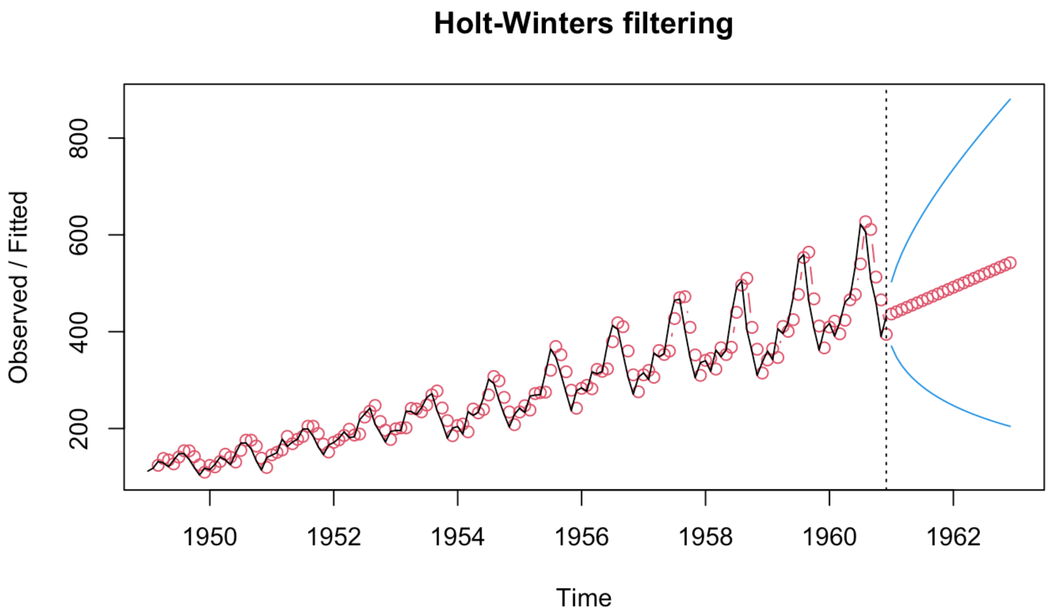


c) Below are the outputs and plots of the Holts model with alpha and beta considered.

```
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = y, gamma = FALSE, seasonal = "additive")

Smoothing parameters:
 alpha: 1
 beta : 0.003218516
 gamma: FALSE

Coefficients:
        [,1]
a 432.000000
b   4.597605
```

The observed alpha coefficient is 1 and beta is 0.032. This means that in predicting the level of the time series at t, the complete weight is given to t-1 only. The lower weight to trend component means high weight is given to previous trends as well and hence capture the trend well.

The model is therefore –
$\hat{y}_{t+k|t} = L_t + k \times T_t$
$L_t = y_t$ (Current Level)
$T_t = 0.0032(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$ (Current Trend)



In addition to the current level, we can now estimate the trend (slope) of the data. As β has a slightly positive value, there is an upward trend in the data, which is confirmed by the plot above. However, the trend is only slightly adjusted over time. Upon comparing this to the Exponential Weighted Moving Average (EWMA) forecasts, it is evident that the Holt Method includes a trend component. Although the α values are nearly identical for both models, indicating that for this model, the forecasts place significant weight on the most recent observation.

d) Below is the output of additive Holts' model considering alpha, beta and gamma parameters for next 24 months.

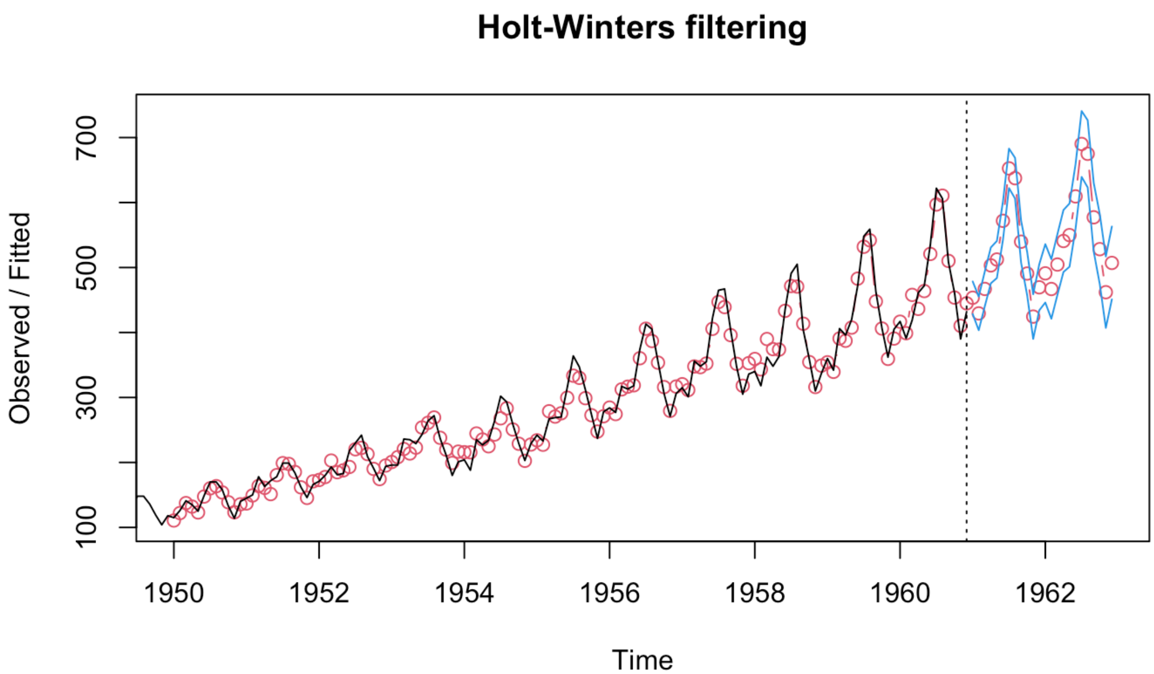
Smoothing parameters:

- alpha: 0.2479595
- beta: 0.03453373
- gamma: 1

Coefficients:

- a 477.827781
- b 3.127627
- s1 -27.457685
- s2 -54.692464
- s3 -20.174608

- s4   12.919120
- s5   18.873607
- s6   75.294426
- s7 152.888368
- s8 134.613464
- s9   33.778349
- s10 -18.379060
- s11 -87.772408
- s12 -45.827781



From the summary output of the Holt Winter model, we can see the optimum $\propto = \mathbf{0.2479}, \boldsymbol{\beta} = \mathbf{0.0345}, \boldsymbol{\gamma} = \mathbf{1}$

The model is therefore –

$\hat{y}_{t+k|t} = L_t + k \times T_t + \ S_{t+k-s}$

$L_t \ = 0.2479(y_t - S_{t-s}) + \ 0.7521(L_{t-1} + T_{t-1})$ (Current Level)

$T_t \ = 0.0345(L_t - L_{t-1}) + \ 0.9655(T_{t-1})$ (Current Trend)

$S_t \ = (y_t - L_t)$ (Current Seasonality)

Here we can see that alpha is lower than the previous model which means that this model considers previous time steps as well and hence cause more smoothening effect of seasoning. The beta value is quite low, therefore capture the trends well. Here the seasonality coefficient (gamma) is 1 which means it considers only current seasonality i.e. for each month only. The seasonality coefficients suggest that passenger numbers are at their peak in the seventh month, closely followed by the eighth month, and at their lowest in November and February. December and January see a slight increase, possibly due to holiday travel.

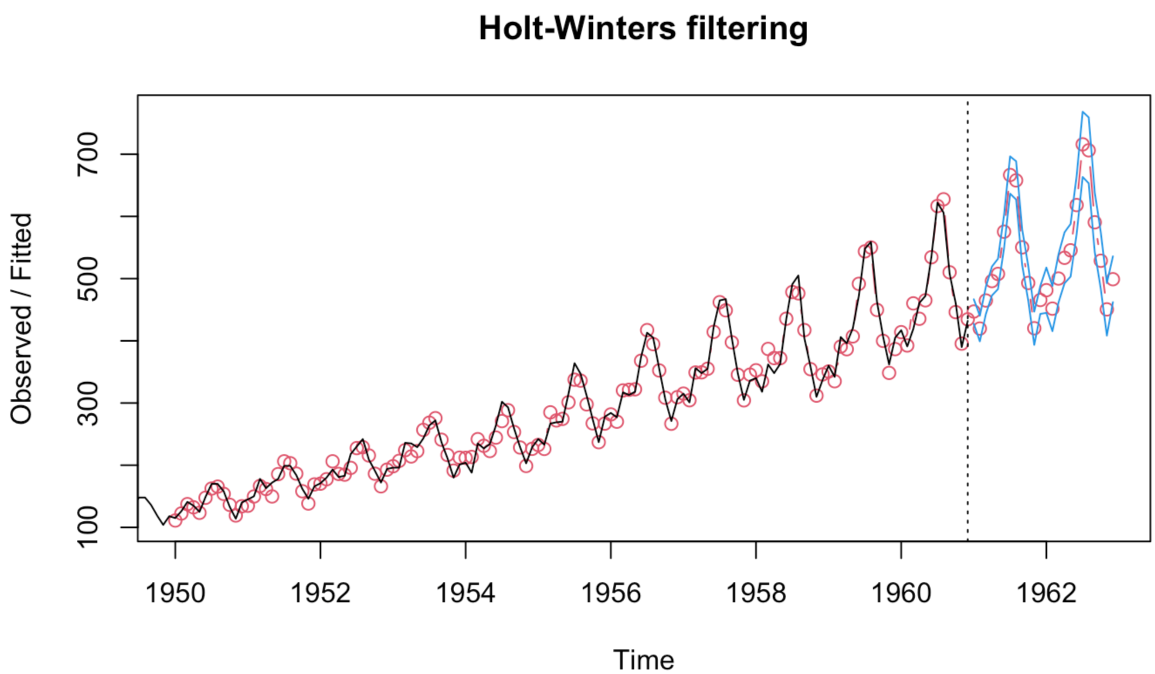
e) Below is the output of multiplicative Holts' model considering alpha, beta, and gamma parameters for next 24 months.

Smoothing parameters:
alpha: 0.2755925
beta: 0.03269295
gamma: 0.8707292

Coefficients:

a	469.3232206
b	3.0215391
s1	0.9464611
s2	0.8829239
s3	0.9717369
s4	1.0304825
s5	1.0476884
s6	1.1805272
s7	1.3590778
s8	1.3331706
s9	1.1083381
s10	0.9868813
s11	0.8361333
s12	0.9209877

From the summary output of the Holt Winter model, we can see the optimum.
$\propto = \mathbf{0.2755}, \boldsymbol{\beta} = \mathbf{0.0327}, \boldsymbol{\gamma} = \mathbf{0.8707}$
The model is therefore –
$\hat{y}_{t+k|t} = (L_t + k \times T_t) \times S_{t+k-s}$
$L_t = 0.2755(y_t/S_{t-s}) + 0.7245(L_{t-1} + T_{t-1})$ (Current Level)
$T_t = 0.0326(L_t - L_{t-1}) + 0.9674(T_{t-1})$ (Current Trend)
$S_t = 0.8707(y_t - L_t) + 0.1293S_{t-s}$ (Current Seasonality)

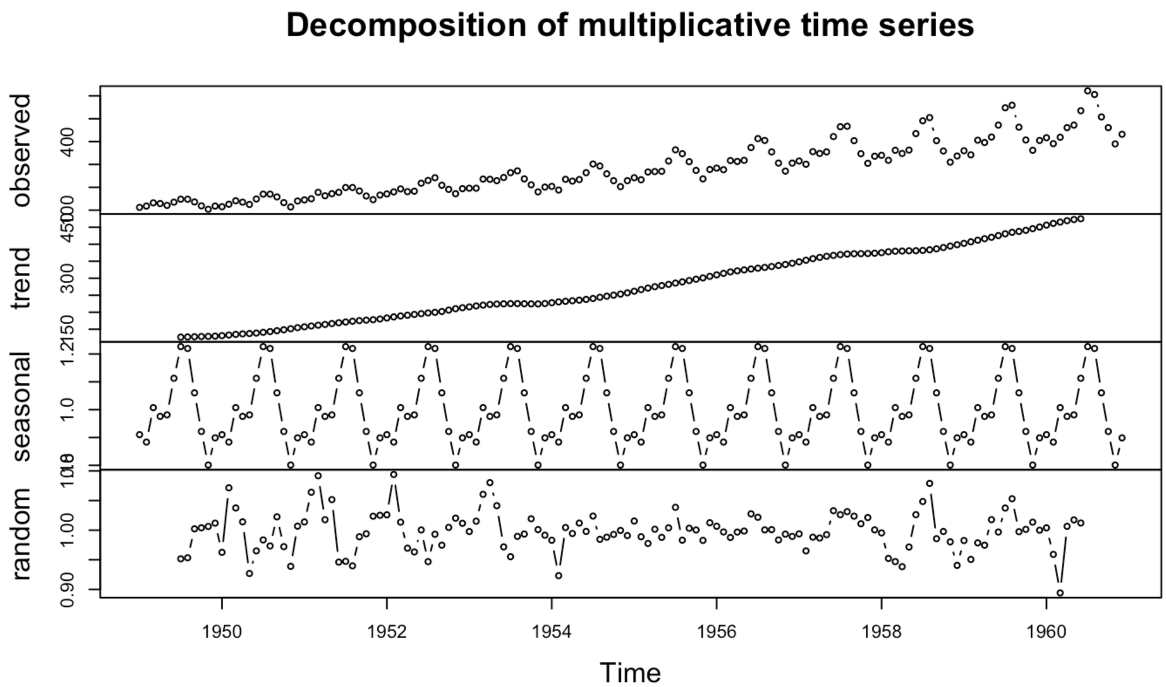


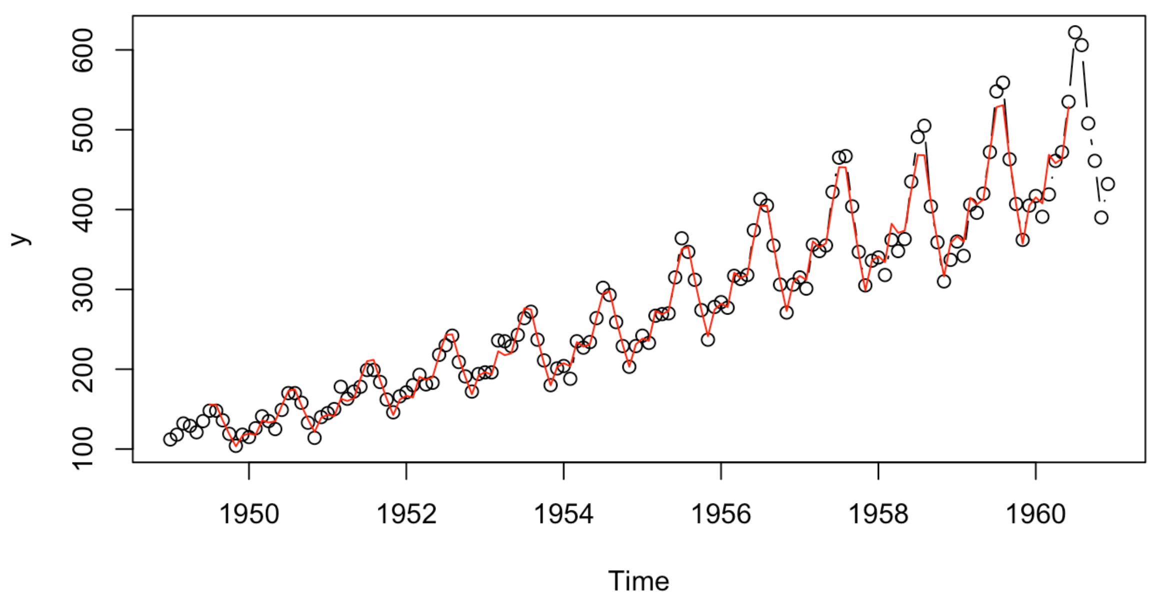
The seasonality coefficients in the multiplicative model have the same interpretation as those in the additive model discussed in part (d), but with a slight difference. In the multiplicative model, the seasonalities multiply the level estimates, and they are adjusted so that their product equals 1. This interpretation applies to both the most and least active months for travel.

f) The given time series data clearly depicts an upward trend and strong yearly seasonality. Also, as the trend increases the amplitude of seasonal component increases. This clearly lays the foundation of using some kind of multiplicative model. There upon comparison, Multiplicative Holt-Winters model is the most suitable choice than the additive one. This is because the amplitude of the seasonal fluctuations increases as the level rises.

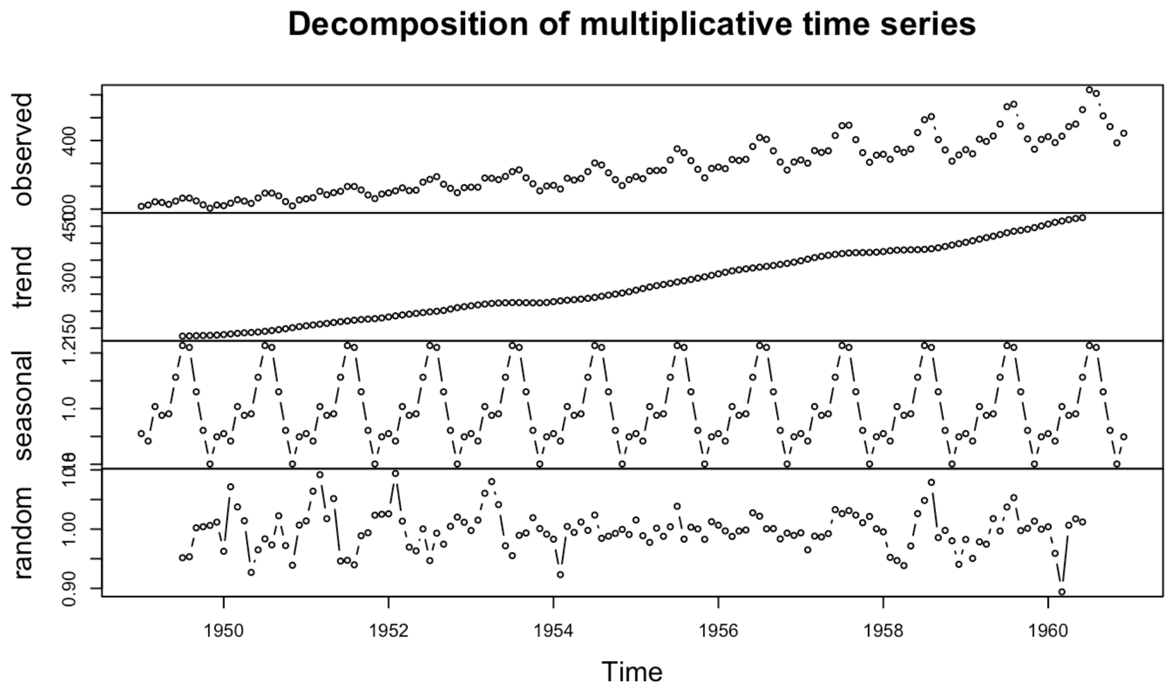
Question 2

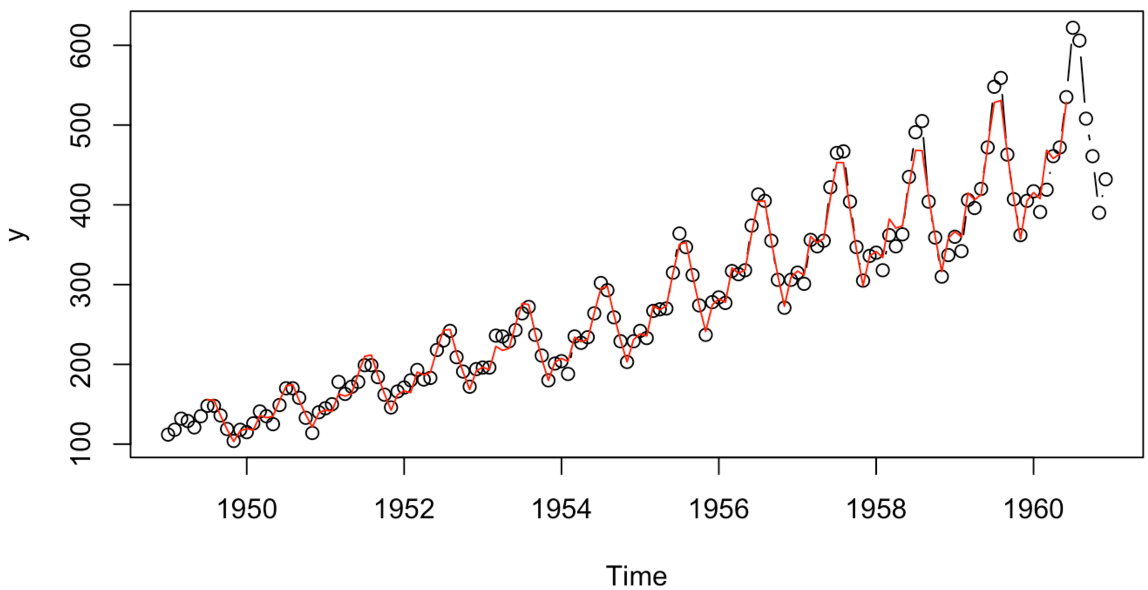
a) The variance of the residuals is significantly smaller than that of the original time series, as evident in the provided plots. However, the residuals still exhibit a distinct seasonality, with overestimation in the early years and underestimation in the later years. The data from 1949 to 1960 shows a consistent upward trend, indicating continuous growth over time. This growth is steady and gradual, increasing each month. There's a noticeable seasonal pattern that repeats annually. February and November consistently show decreased activity, while June, July, and August show increased activity. This suggests a predictable seasonal effect with peaks in the summer and dips in late winter and late autumn. Upon examining the residual variance, it's clear that the seasonality in the residuals differs between the early and later years. This suggests that an additive model may not fully capture the data's characteristics, and a multiplicative model might be a better fit.





b) From the plots below (especially the second plot), the multiplicative model clearly fits better than the additive model.





The multiplicative model reveals that the seasonal variations in the time series are dependent on the series' level. For example, January consistently has a factor below 1 (0.9102304), suggesting less activity than the trend, while July has factors above 1

(1.2265555), indicating higher activity levels. This consistent pattern year after year confirms a stable seasonal effect. The trend component shows a steady increase over time, which matches the observed rise in the time series data. The consistent growth from 1949 through the late 1950s reflects the underlying long-term trends in the data, separate from seasonal fluctuations. There are some variations, but they don't show a clear pattern or systematic deviation, suggesting that the trend and seasonal components have successfully captured most of the systematic information in the data.

Therefore multiplicative model is a choice to go!

PA2 HW4

Ayush Agarwal

2024-03-02

Question 1

```
#Load and Inspect the data
data = read_excel("HW4_data.xls", col_names = "y")
head(data)
```

Part a)

```
# Converting Data into time series
y<-ts(data[[1]], start=c(1949, 1), end=c(1960, 12), frequency=12)
m=12;n=length(y) # m = 12 is the window size of 12 months
MAair<-filter(y, filter=rep(1/m,m), method = "convolution", sides = 2)
plot(y,type="b")
lines(MAair,col="red")
```

Part b)

```
k=24;n=length(y) #k = prediction horizon
EWMA_air<-HoltWinters(y, seasonal = "additive", beta = FALSE, gamma = FALSE)
EWMA_airPred<-predict(EWMA_air, n.ahead=k, prediction.interval = T, level = 0.95)
plot(EWMA_air,EWMA_airPred,type="b")
EWMA_air
```

Part c)

```
k=24;n=length(y) #k = prediction horizon
holt_air<-HoltWinters(y, seasonal = "additive", gamma = FALSE)
holt_air_Pred<-predict(holt_air, n.ahead=k, prediction.interval = T, level = 0.95)
plot(holt_air,holt_air_Pred,type="b")
holt_air
```

Part d)

```
k=24;n=length(y) #k = prediction horizon
Hold_winter_air<-HoltWinters(y, seasonal = "additive")
Hold_winter_air_pred<-predict(Hold_winter_air, n.ahead=k, prediction.interval = T, level = 0.95)
plot(Hold_winter_air,Hold_winter_air_pred,type="b")
Hold_winter_air
```

Part e)

```
k=24;n=length(y) #k = prediction horizon
Hold_winter_air_mul<-HoltWinters(y, seasonal = "multiplicative")
Hold_winter_air_mul_Pred<-predict(Hold_winter_air_mul, n.ahead=k, prediction.interval = T, level = 0.95
plot(Hold_winter_air_mul,Hold_winter_air_mul_Pred,type="b")
Hold_winter_air_mul
```

Question 2

Part a)

```
k=24;n=length(y) #k = prediction horizon
dec_air<-decompose(y, type = "additive")
plot(dec_air,type="b")
y_hat<-dec_air$trend+dec_air$seasonal
plot(y,type="b")
lines(y_hat,col="red")
```

Part b)

```
k=24;n=length(y) #k = prediction horizon
dec_air<-decompose(y, type = "multiplicative")
plot(dec_air,type="b")
y_hat<-dec_air$trend*dec_air$seasonal
plot(y,type="b")
lines(y_hat,col="red")
```