INTRODUCTION:

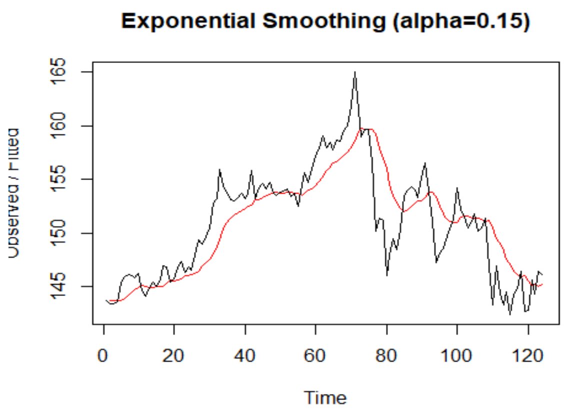
Assignment 3 gives us a good exposure on Time-Series Forecasting and regression analysis, by performing certain techniques in time series such as holtwinters with different alpha and beta values and observing the MSE of the respective ones and forecasting. Regression analysis helps us understand the R2 and the residuals.

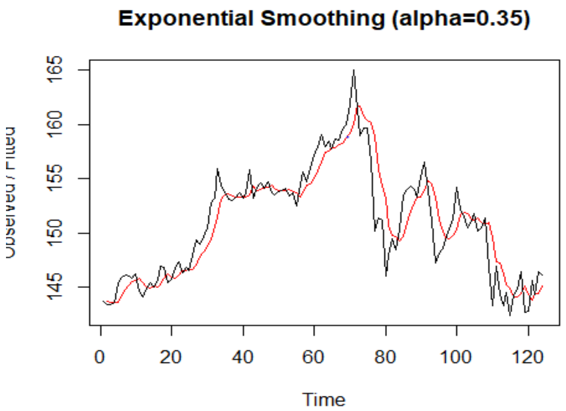
ANALYSIS:

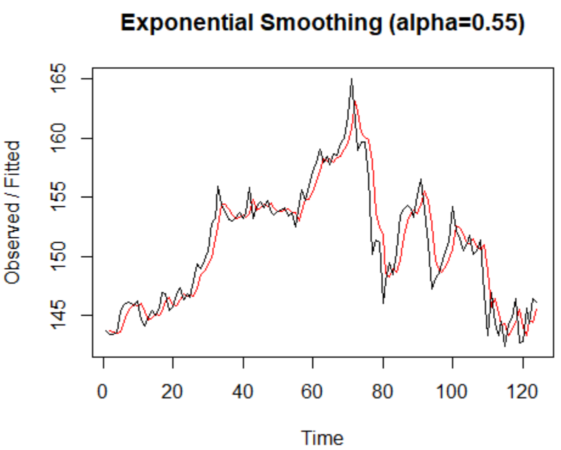
Problem 1: A]

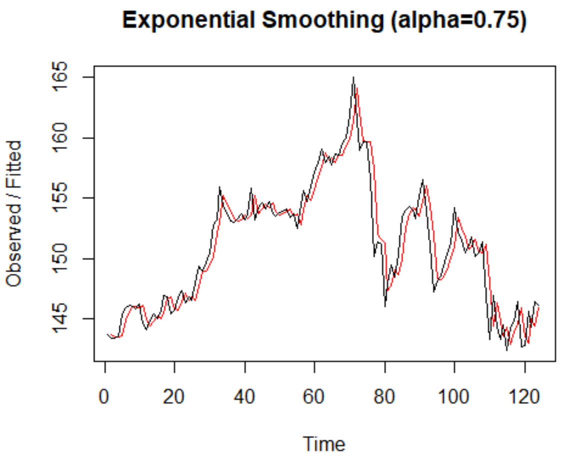
This problem contained data on the on the historical stock prices of the Honeywell International Incorporated, which is an American multinational company that produces a variety of commercial and consumer products for both private and government use.

The price forecasted for the **4/16/2018** by using four different values where α was **0.15, 0.35, 0.55, 0.75** and the values were **145.38, 145.45, 145.86, 146.08** respectively. Then it was asked to calculate the MSE values which is the calculated by the difference between the actual and the expected value taking alpha into consideration and is squared with the difference. The MSE calculated for all the 4 alpha values are **7.907, 4.505, 3.38 and 2.953.**







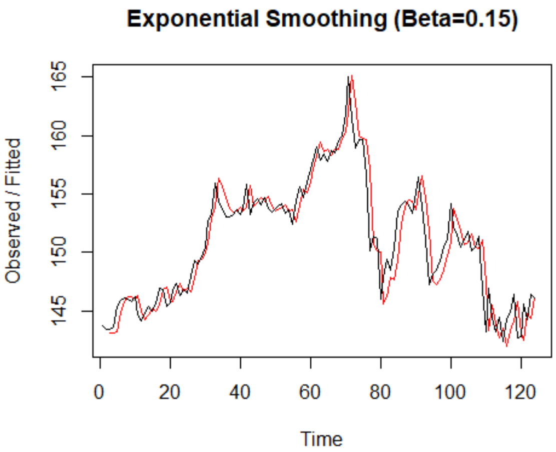


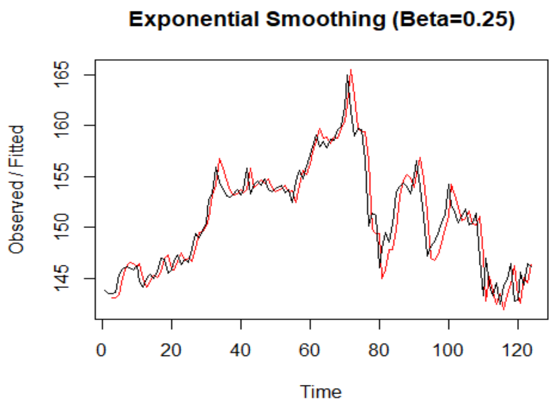
As we see the exponential smoothing plot retrieved after performing holtwinters on the time-series data we can see that the alpha=0.75 is more accurate with the actual value and the MSE obtained for the same is also less than all the others which is **2.952.**

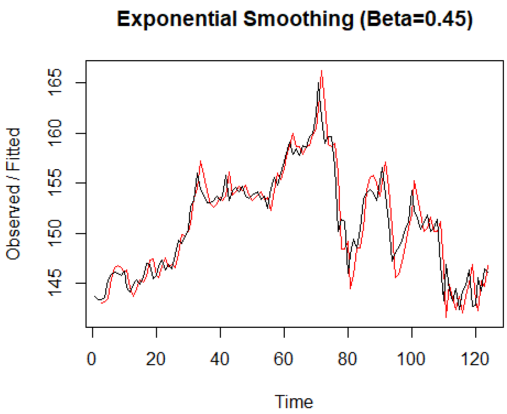
Problem 1: B]

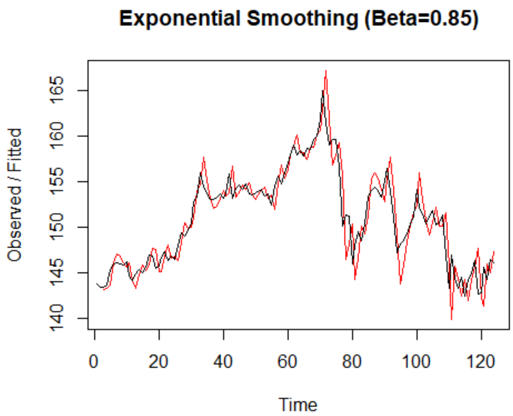
In the first part of the problem I have kept the Beta value as false , now in this part we are considering and exponential smoothing technique by taking alpha as a constant value which we obtained from part 1 with the best accuracy and low MSE value which Is $\boldsymbol{\alpha}$**=0.75**. Beta is used to calculate the trend over time in this part we are calculating beta in four different levels which are **0.15, 0.25, 0.45, 0.85**. I then calculated the forecast values by adding the alpha obtained in 0.75 which each individual Beta value to get the individual result for all beta values.

Then calculated the MSE values for beta **0.15, 0.25, 0.45, 0.85** and obtained values **3.56, 3.41, 3.22, 3.30** respectively.









From the above figure we can see that the Beta value **0.45** perfectly fits the actual values with slight deviation and the MSE obtained for that particular beta value is very low compared to others.

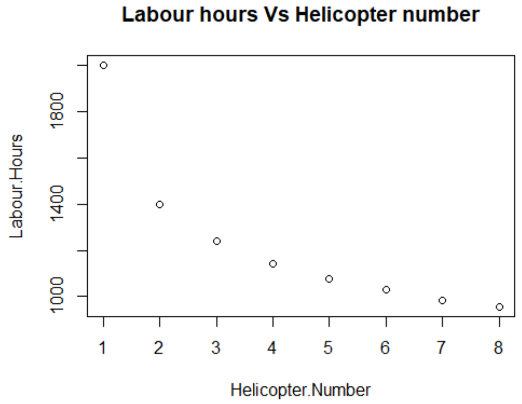
Problem 2:

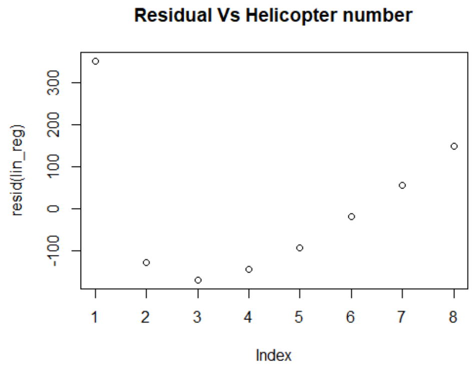
The problem given in problem 2 contains data on Helicopter Division of Aerospatiale who are studying the assembly cost at its Marseilles plant. I performed a linear regression and received a p-value of 0.00702 which is way below the significance level. Then retrieved the residuals in R using the

```
> summary(lin_reg)

Call:
lm(formula = Labour.Hours ~ ., data = heli_dicision)

Residuals:
    Min      1Q  Median      3Q     Max 
-170.07 -132.51  -56.37   79.42  352.17 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       1767.71     150.88  11.716 2.33e-05 ***
Helicopter.Number -119.88      29.88  -4.012  0.00702 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' :

Residual standard error: 193.6 on 6 degrees of freedom
Multiple R-squared:  0.7285,	Adjusted R-squared:  0.6832 
F-statistic:  16.1 on 1 and 6 DF,  p-value: 0.00702
```

Resid command for the applied linear regression. The R2 obtained from the regression is show in the figure below.

The below plot shows the result of the residuals obtained after applying the linear regression for that we can conclude that the obtained graph looks like polynomial.





Problem 3:

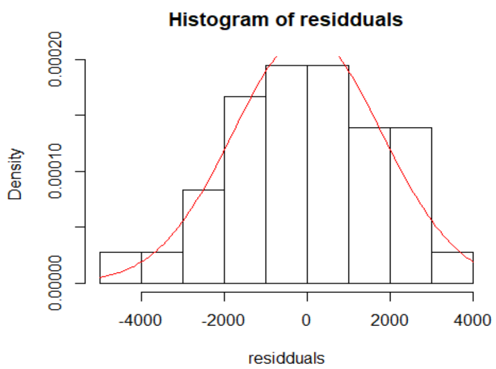
I performed multiple linear regression to check the seasonality for last three years of data for New car sales dataset.

In multiple regression you strive to get result with variables with significant p-values, the process follows in such a way that, firstly you run the entire model, then you remove value that has highest p-value. You do this process removing one variable at a time to check whether removing that p-value may adjust another variable which has the highest p-value if not then remove that variable in another iteration this process continues until you get all values that are significant.

After getting the significant result for all the variables we predicted the model to get the forecast sale values. Then applied anova to check the variances and we can see that variance is not very large from the figure.

```
> anova(mul_reg)
Analysis of Variance Table

Response: Units
          Df    Sum Sq   Mean Sq F value    Pr(>F)
Month     11 463007522  42091593  8.2665 1.185e-05 ***
Year       1 494732881 494732881 97.1621 1.002e-09 ***
Residuals 23 117112097   5091830
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finally obtained the residual plot to check which kind of distribution it follows, it looks like, it follows normal distribution as shown in figure below:



Furthermore, analysis done by preforming the chi-square test to check whether We can accept the null hypothesis: Follows normal distribution or alternate hypothesis which is its opposite. After the result of chi square test, we can conclude that the p-value obtained from it is large than the level of significance. Therefore, null hypothesis is accepted: It follows normal distribution.

```
> pearson.test(residduals)

        Pearson chi-square normality test

data:  residduals
P = 2, p-value = 0.9197
```