**MSBA444 EXAM – I Due: Nov. 22nd , 2022**

Note: Please upload your final report on Canvas by Tuesday, Nov. 22nd. Your report should include not just the final answer but should show the process you went through to derive it. Upload all supporting material (R/EXCEL files) to Canvas.

In preparing your report, you may follow following outline:

1. Brief Problem description
2. Your answers and conclusions (Managerial Summary)
3. Details of the Methodology/Analysis you used to derive your answers and conclusions.
4. Attach relevant printout of your model (EXCEL or R file) and outputs as an Appendix.

Your report should be complete so that I need to go to your uploaded files only as an exception.

Note: **This is an individual exam!!** You should not discuss any aspect of the exam with fellow students, or share any computer input or output files. **Receiving or offering help** in an exam is against the school’s academic integrity policies – any violation can result in serious consequences. Also note that searching internet for the solution to the problem is also violates school’s academic integrity policies. If you need any clarification about integrity policies or about the exam, feel free to contact me. You can stop by my office or send me an email.

In taking this examination, I agree to abide by the honesty and academic integrity values of the Weatherhead School of Management and Case Western Reserve University.

Signature:\_\_\_\_\_\_\_\_\_Ziming Cui\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name: \_\_\_\_\_\_\_\_\_\_Ziming Cui\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Include signed copy of this page as the cover sheets for your exam report**

1. 15

2. 15

30

# Predicting Gross Margin for the national Pharmacy

**1 Introduction**

**1.1 Problem Description**

The national Pharmacy came up with a pricing model to make the repricing process more efficient. The company think that the two key factors which influence the Gross Margin are COST and whether the product needs refrigeration not, and develop a linear or nonlinear regression model (try stepwise regression) to estimate GM as a function of COST and whether product needs refrigeration or not.

**1.2 Dataset Description**

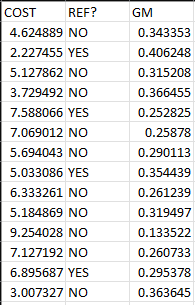
3 variables, 100 observations

1) Dependent Variable (Y) :

GM

2) Independent Variables (X) :

COST, REF(YES or NO)



**2 Predicting pricing model**

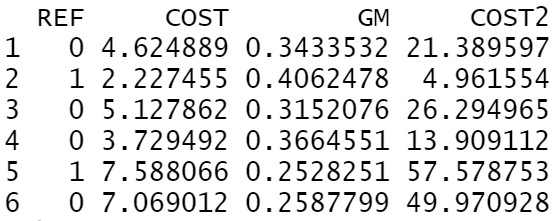
Using the stepwise method, the predicting regression line is :

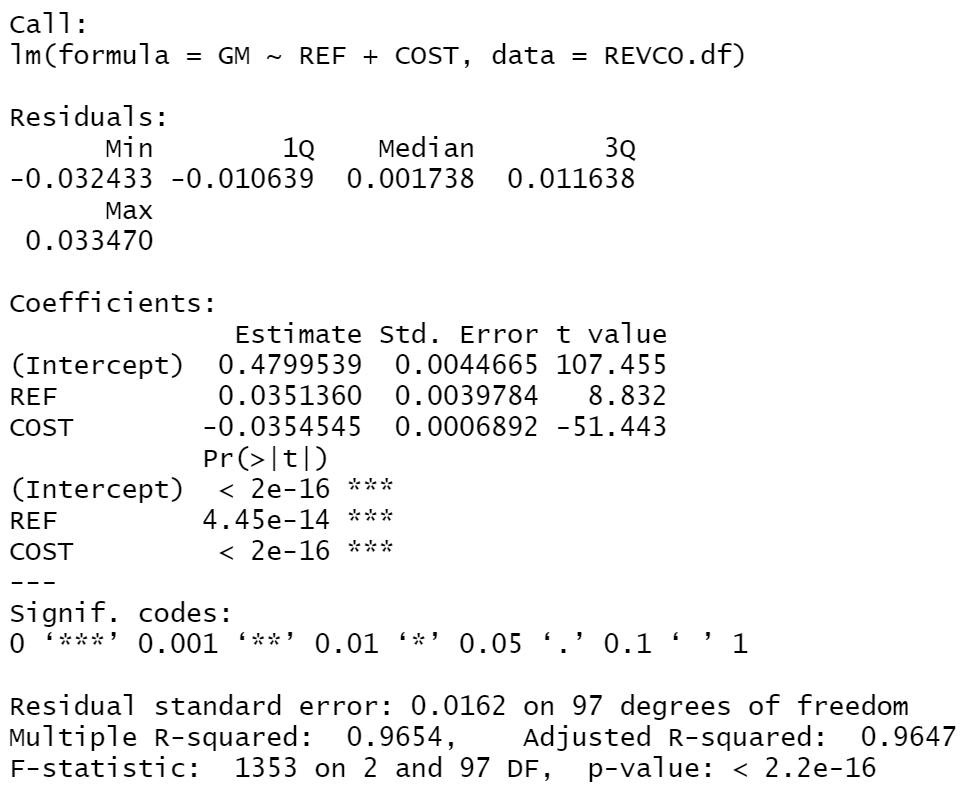
***GM = 0.392 – 0.003 \* COST^2 + 0.0326 \* REF***

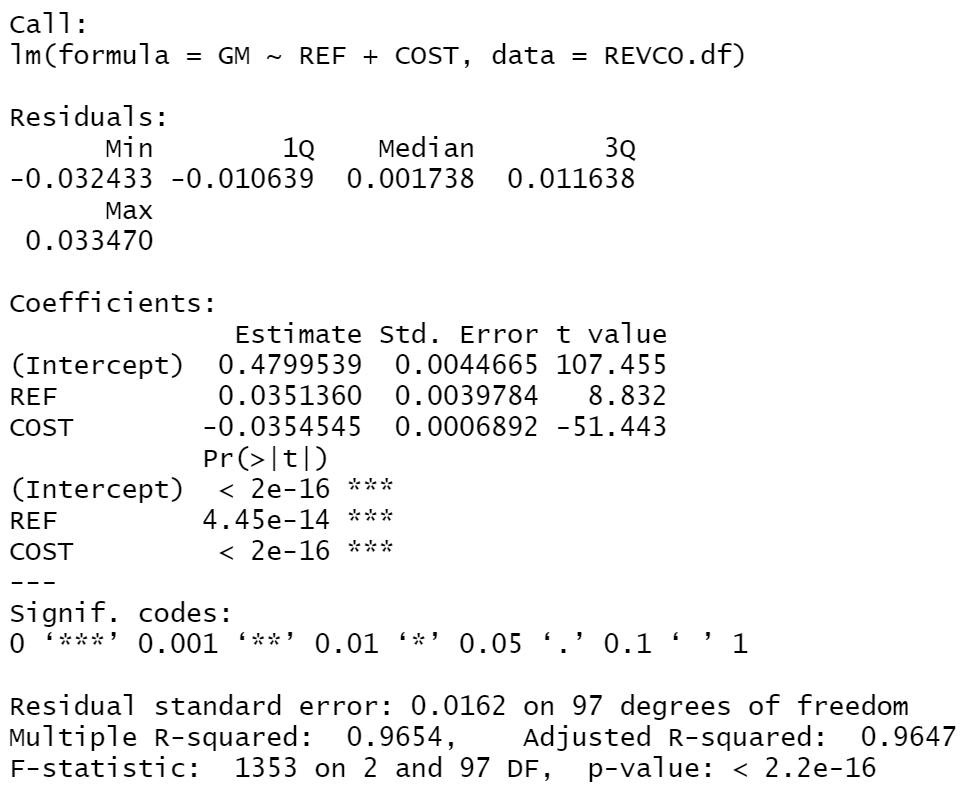
**3 Details of** **Analysis**

**Step 1**

Using the stepwise method, we can get the Linear Regression model.



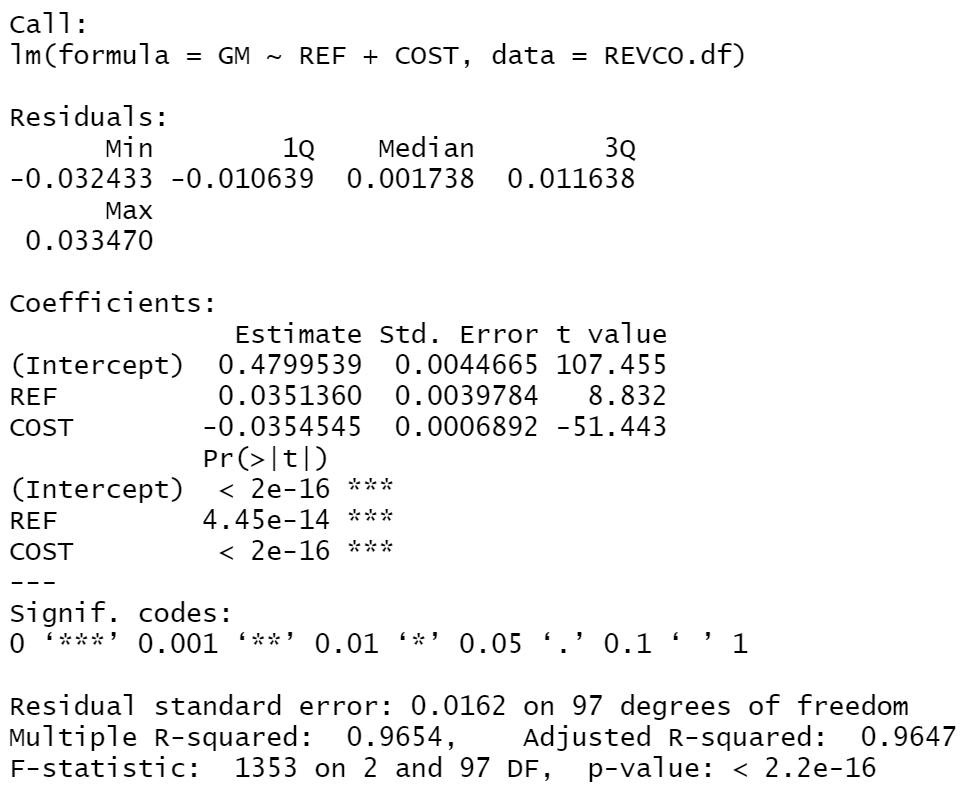




According to Regression Analysis, we can know that the initial value of GM is 0.4800, for the first coefficient (0.48). The GM will decrease , when the cost increase, not only for the second coefficient (0.00355) and the third (0.0351).

**Step 2**

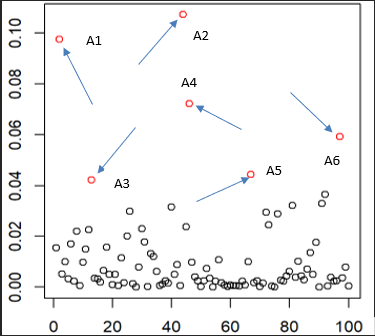
From regression analysis, we can also get Standard Error (SE) and R2.



SE = 0.0162 and R2 = 0.9654 show that with 98% probability, the data is between -0.0324 and 0.0324, which shows that the model fits the price well.

**Step 3**

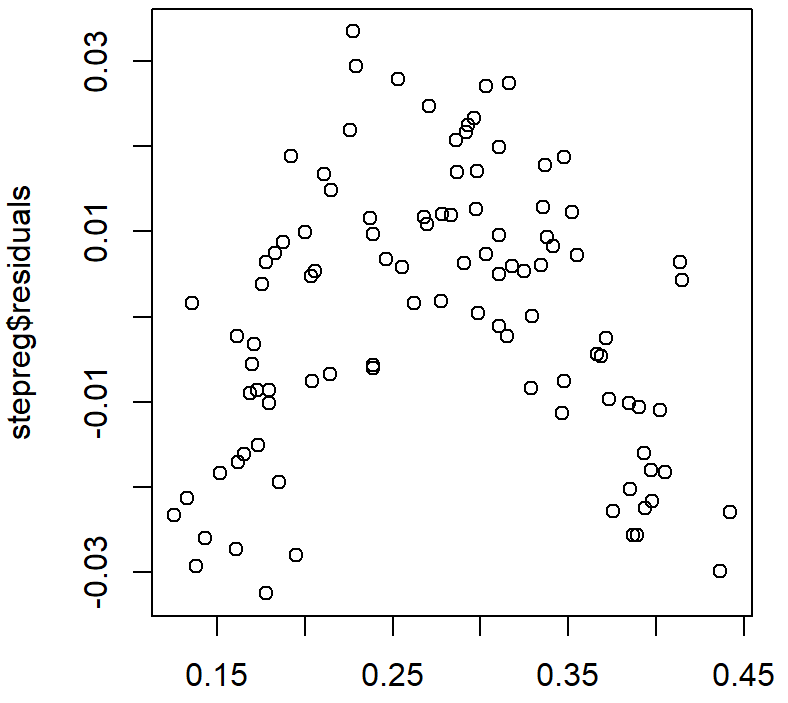
Check outliers

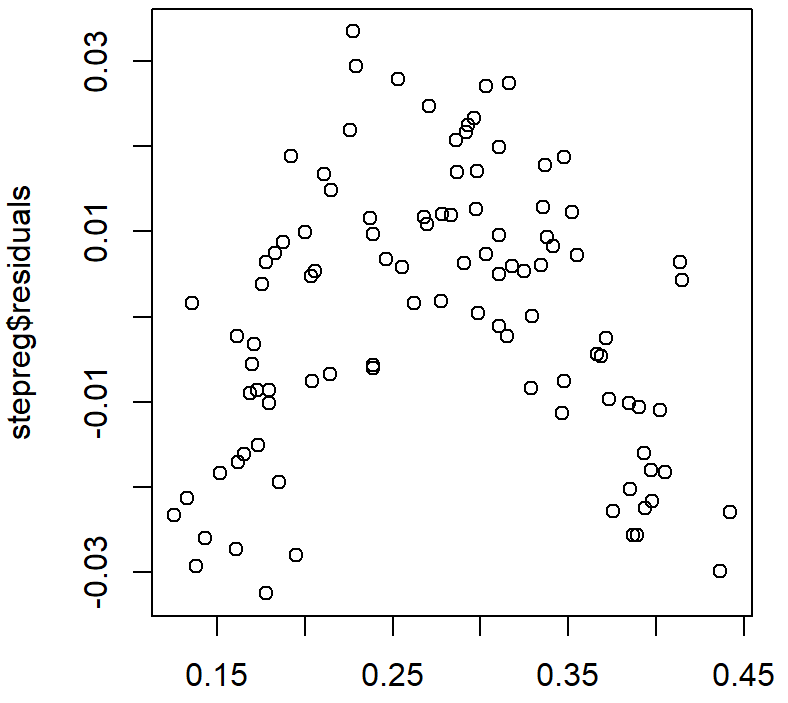


We can clearly find outliers (A1, A2, A3, A4, A5, A6) from the figure. These data are very important for our subsequent analysis, we cannot ignore them, otherwise the accuracy of predictive analysis will be reduced.

**Step 4**

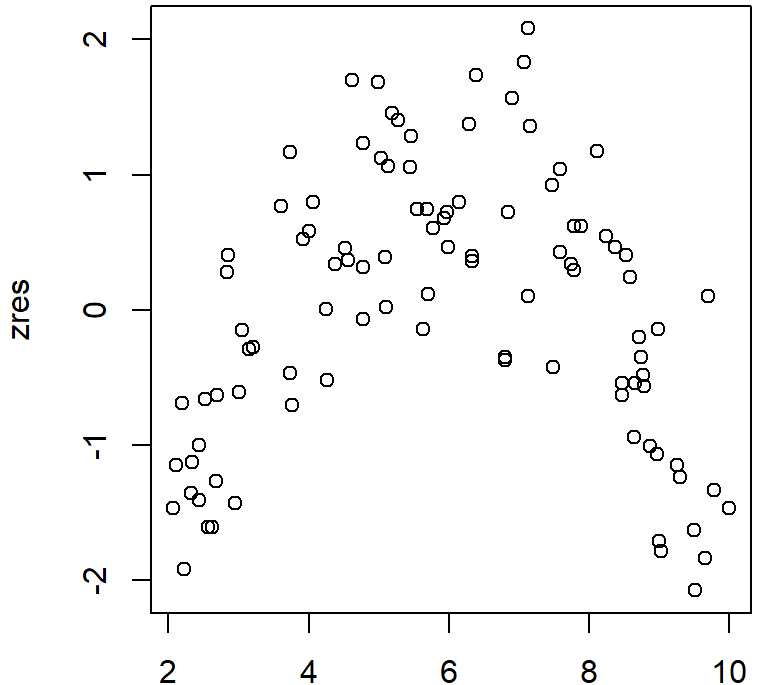
* 1. **Testing Homoscedasticity**





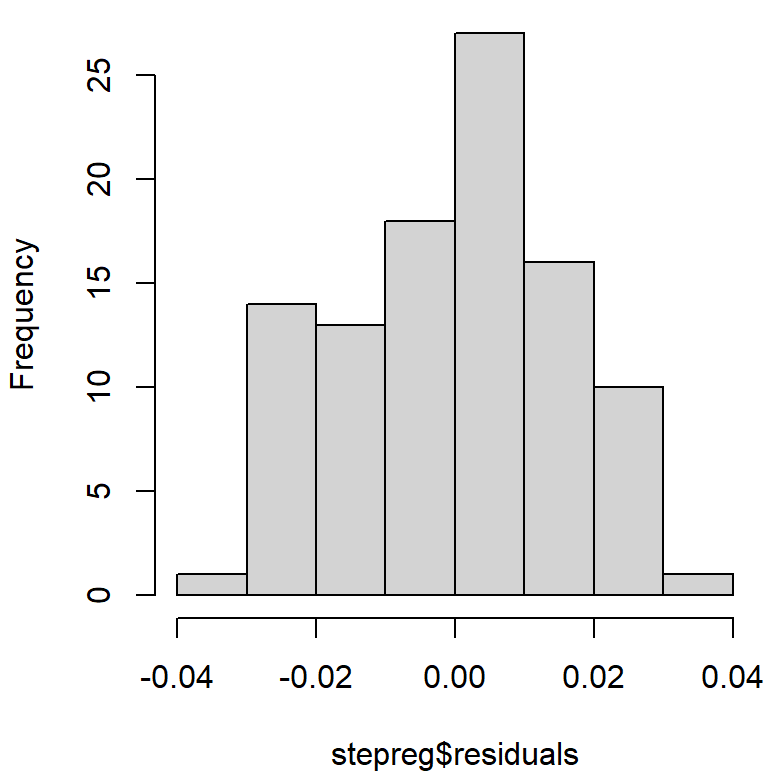
According to the homoscedasticity test plot, we can deduce that the Homoscedasticity Test obviously failed.

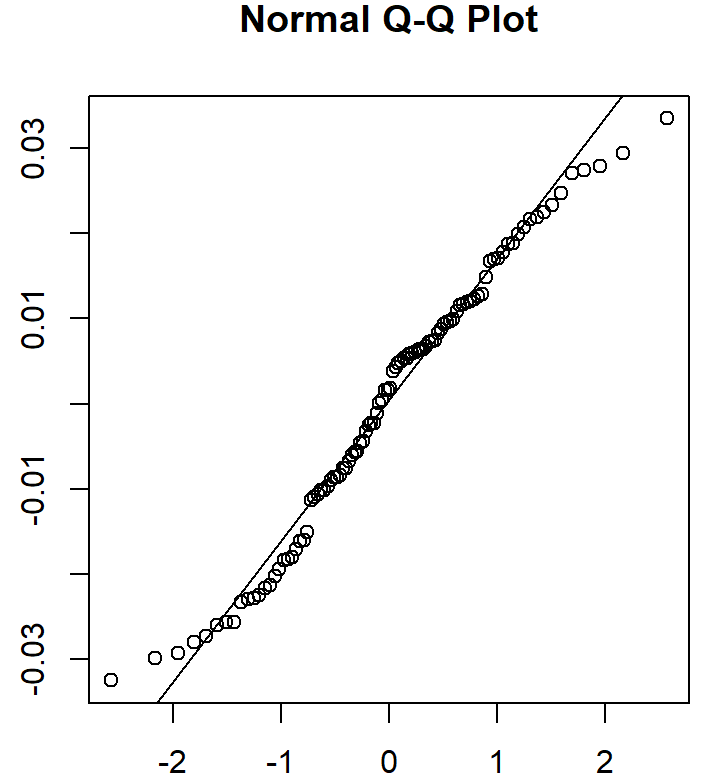
* 1. **Testing Linearity**



We can also know from the distribution of points in the plot that the Linearity Test also failed.

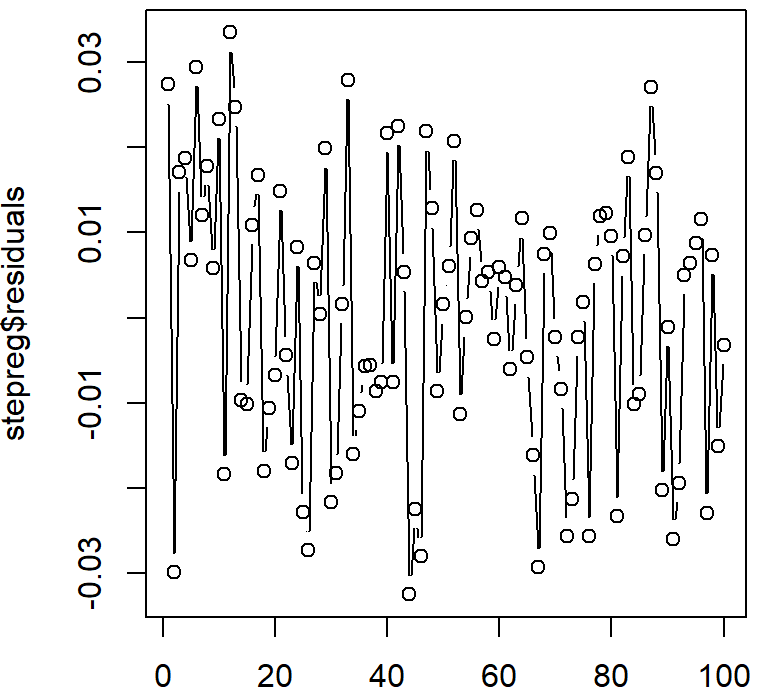
* 1. **Testing Normality**





p-value = 0.08043, it is hard to prove the effectiveness of Normality.

* 1. **Testing Independence**



The scatter plot of the independence test is randomly distributed, so the test passes.

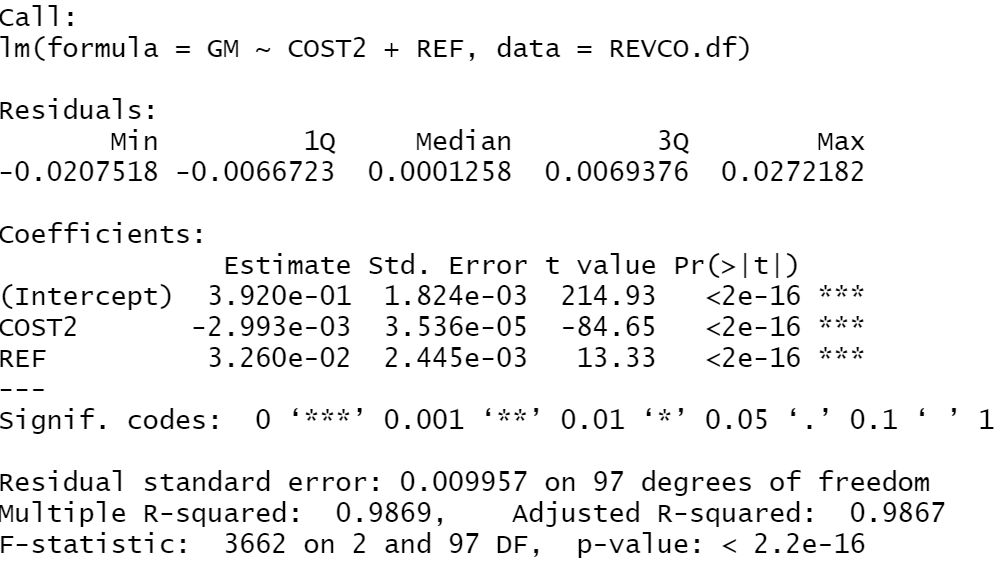
* 1. **Conclusion**

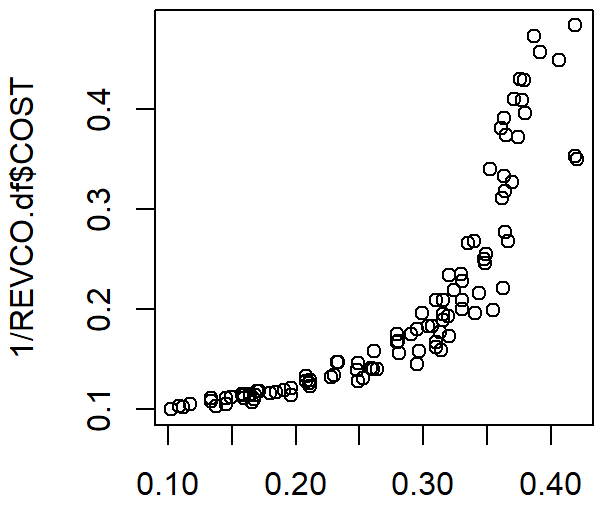
The Linear Regression assumption fails, so we need to rebuild a new regression model and test it.

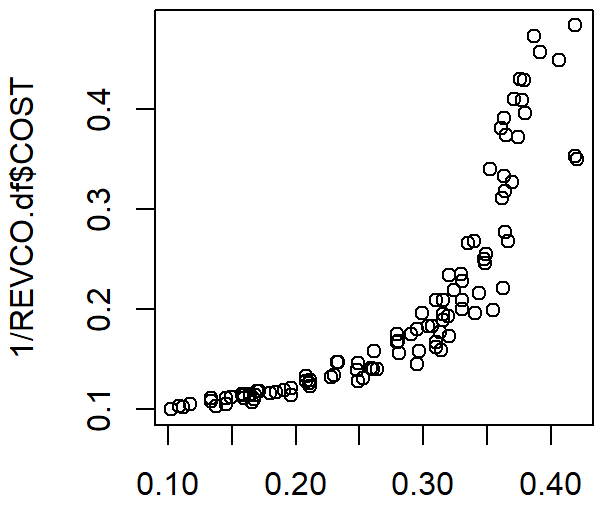
**4 Nonlinear regression model**

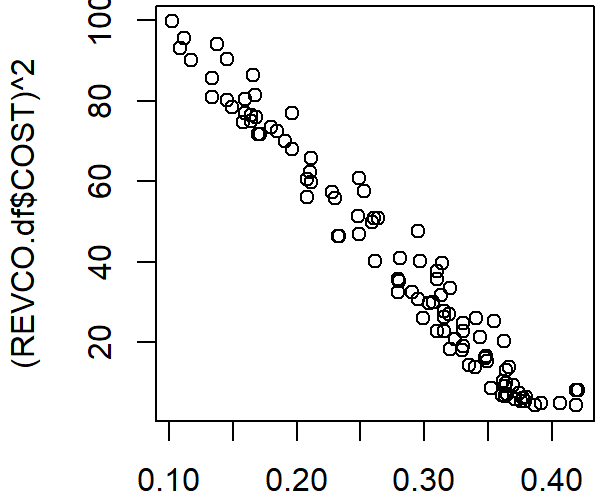
**Step 1**

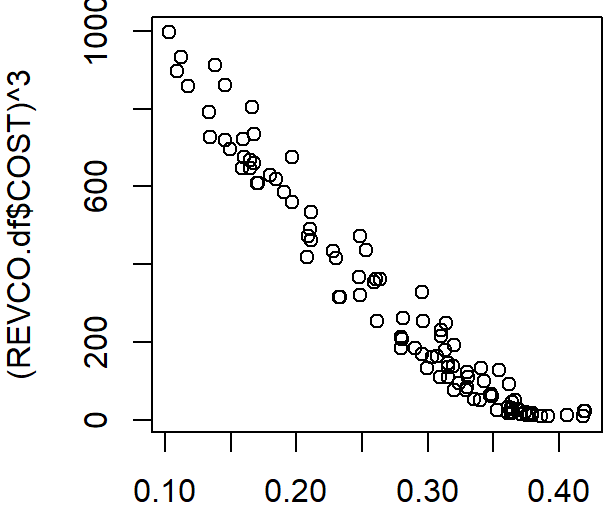
From the existing nonlinear regression models, find the best matching model, using Stepwise ways.



 plot(Data.df$GM,1/Data.df$COST)

 plot(Data.df$GM,log(Data.df$COST))

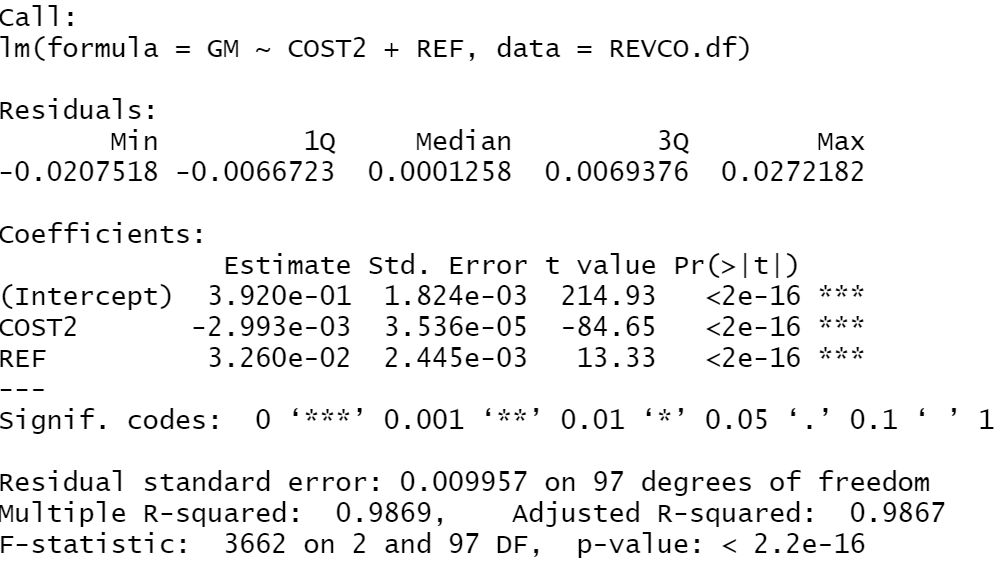
plot(Data.df$GM,( Data.df$COST)\*\*2)

plot(Data.df$GM,( Data.df$COST)\*\*3)

We can see from the figure that cost^2 is the best fit.

**Step 2**

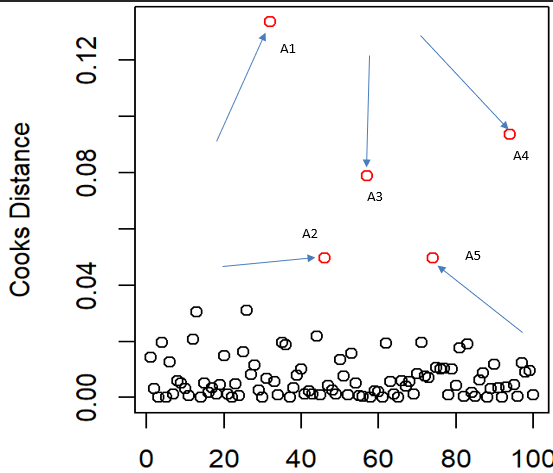
From regression analysis, we can also get Standard Error (SE) and R2.



SE = 0.01 and R2 = 0.9869 show that with 98.69% probability, the data is between -0.02 and 0.02, which shows that the model fits the price better.

**Step 3**

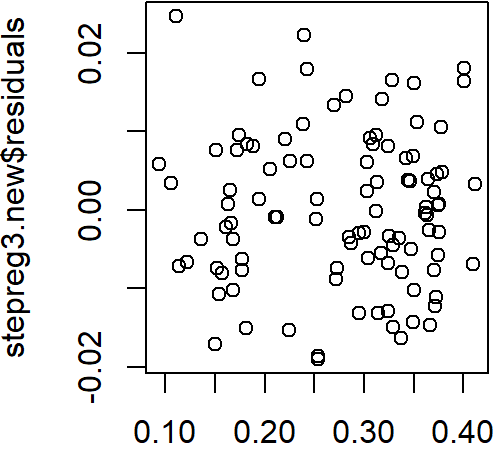
**Check outliers**

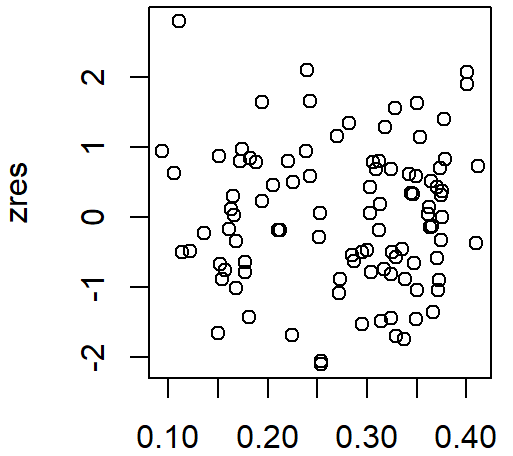


We can clearly find outliers (A1, A2, A3, A4, A5) from the figure. These data are very important for our subsequent analysis, we cannot ignore them, otherwise the accuracy of predictive analysis will be reduced.

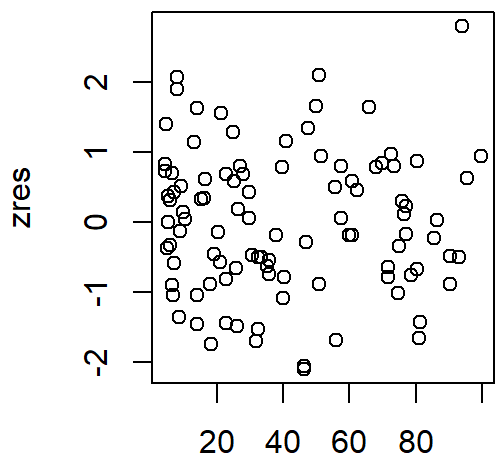
**Step 4**

* 1. **Testing Homoscedasticity**

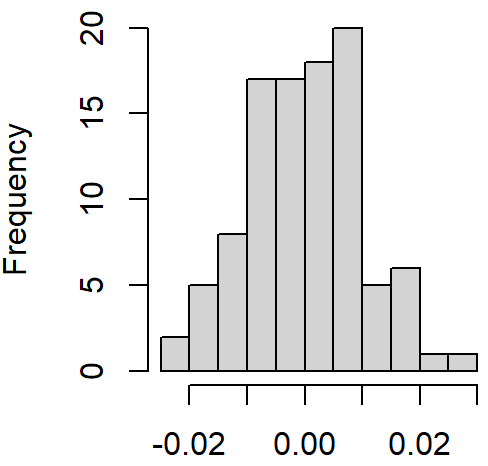


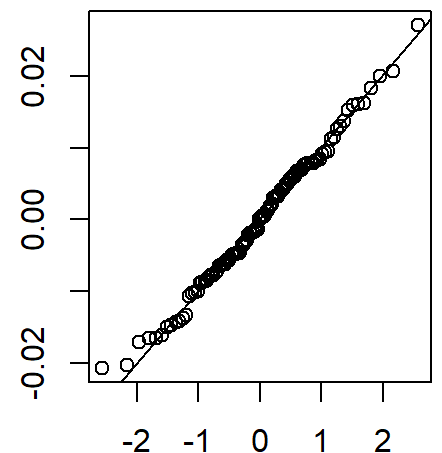


* 1. **Testing Linearity**

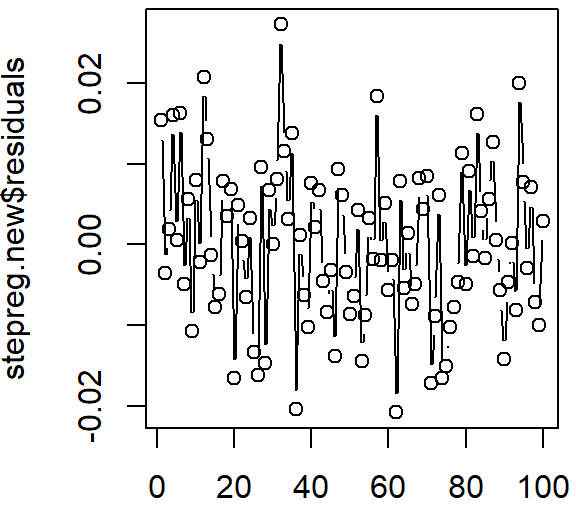


* 1. **Test for Normality**





* 1. **Test of Independence**



P-value = 0.756, all four tests passed.

**5 Conclusion**

GM = 0.392-0.003\*COST^2+0.033\*REF = 0.392-0.003\*6.50\*\*2+0.033 = 0.29825

Because COST = 6.50, and Price = COST/(1-GM), GM = 0.29825

# Predicting Monthly Occupancy rate

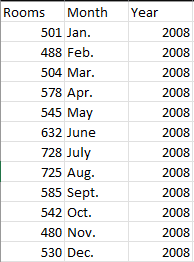
**1 Introduction**

**1.1 Problem Description**

The client (Hotel) wished to create a model that could be applied to get short-term projections (up to a year) of how many hotel rooms would be filled. These projections are helpful when ordering materials and supplies with lengthy lead periods, adding on extra staff, etc.

**1.2 Dataset Description**

The dataset provides the monthly occupancy rate over the previous 14 years (daily average rooms throughout a month) (Jan. 2008 to Dec. 2021).



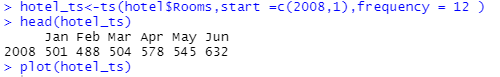
168 observations ( 14\*12 months )

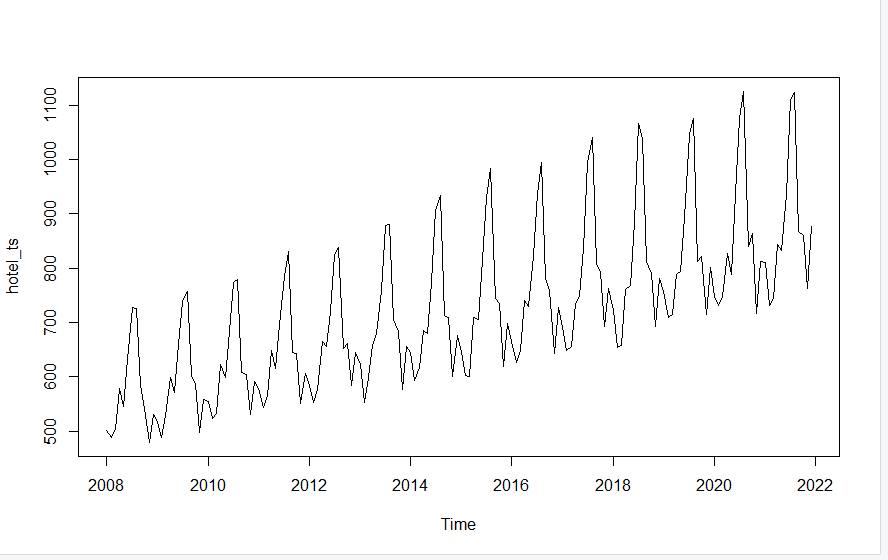
3 variables ( columns )

**2 Details of Analysis**

**2.1 Exponential Smoothing Algorithm**

1) Plot the data and choose a model.





There is a Seasonal Behavior in this image, I think it maybe use the AAA or MAM models after timeseries definition.

**2.2 AAA and MAM Model**

This is a model about AAA.

形状, 箭头

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

This is a model about MAM.

图片包含 图表

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

Since predicting is one of my project's primary goals, I divide the data into a training set and a validation set. Next, I create the models using my training data. Since I don't want the data to be overly sensitive to the new data, I set the coefficient's range to be between 0.1 and 0.5. The performance of these two models' predicting is then assessed.

Comparing the AIC and RMSE values of the AAA model and the MAM model, we can know that these values ​​of the AAA model are smaller, and alpha = 0.1157, beta = 0.1, gamma = 0.2289, so we choose the AAA model.

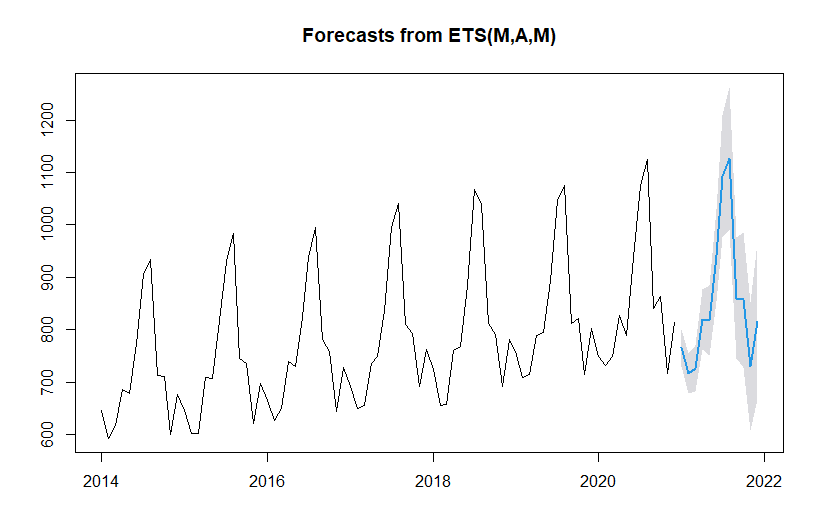
**2.3 Forecast the next year (12 months)**

一些文字和图案

中度可信度描述已自动生成

手机屏幕截图

描述已自动生成

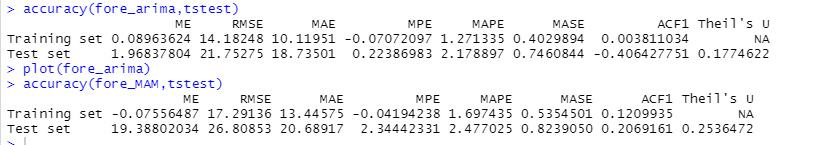


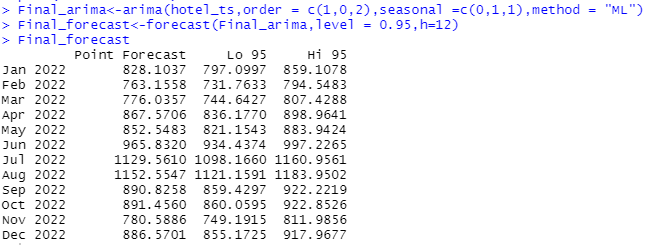
The following forecasts for the upcoming 12 months are provided (with a 95% confidence interval).

**2.4 Best Arima model**

手机屏幕截图

描述已自动生成





My Arima model is ARIMA (1,0,2) (0,1,1) [12]. I built the model using the training set and set the interval of the model's coefficient as the problem's instruction. Non-seasonality coefficient values are φ1=-0.957, θ1=1.2392 and θ2=0.3957. D=1.9381, θ1s=-0.5949, for the seasonality coefficient. The performance of forecasting is then used to compare the ARIMA model and the MAM exponential model. In predicting the validation set, our ARIMA model's ME, MAE, and RMSE are all lower than those of the Multi trend-seasonality exponential model. As a result, I advise using the Arima model and forecasting using it.

**3 Conclusion**

The 12-month forecast in its final form.

