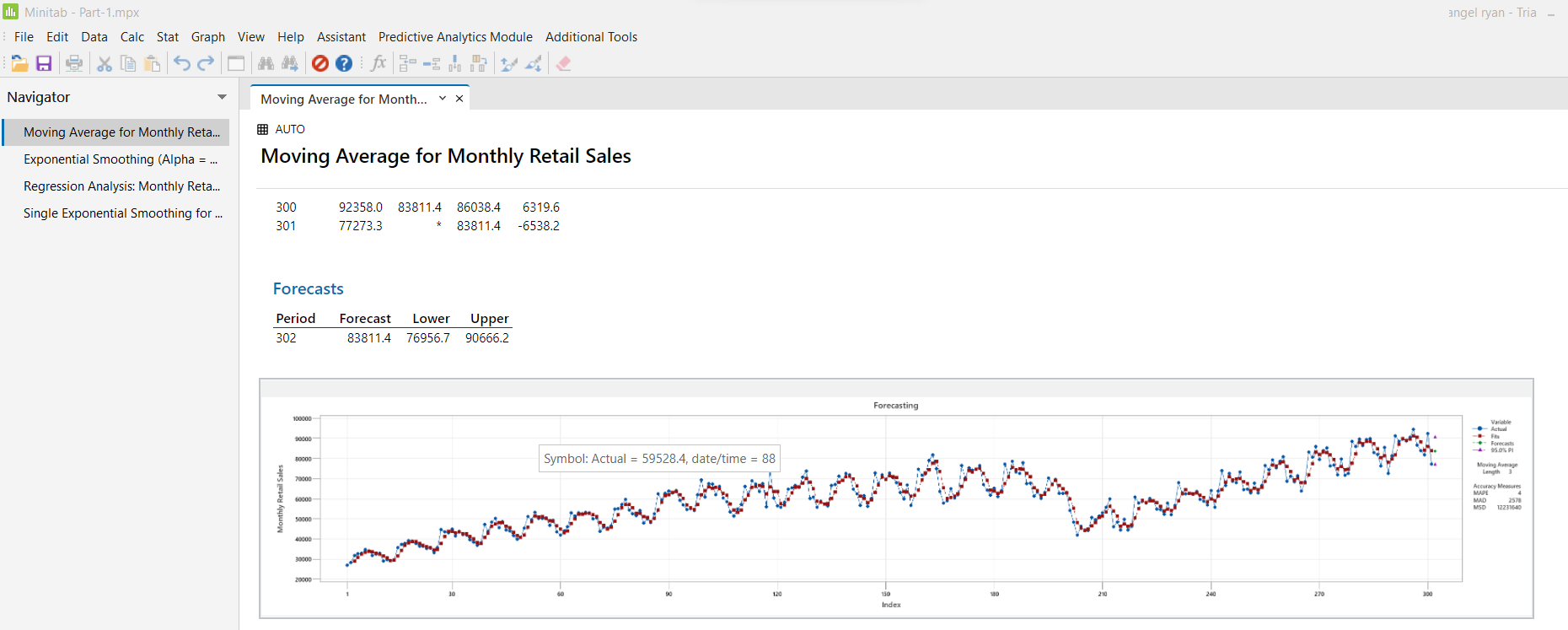
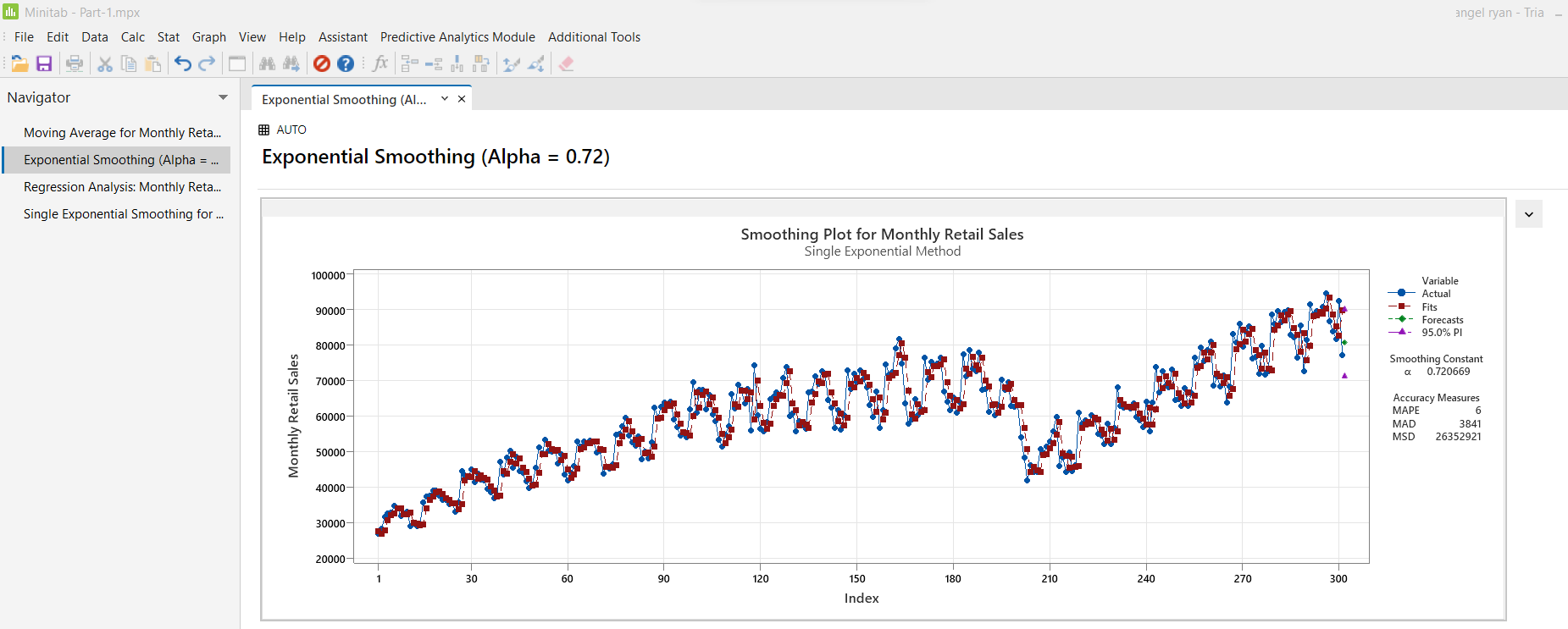
**Time Series Data Analysis**

**Part-1: Auto time series dataset selected.**

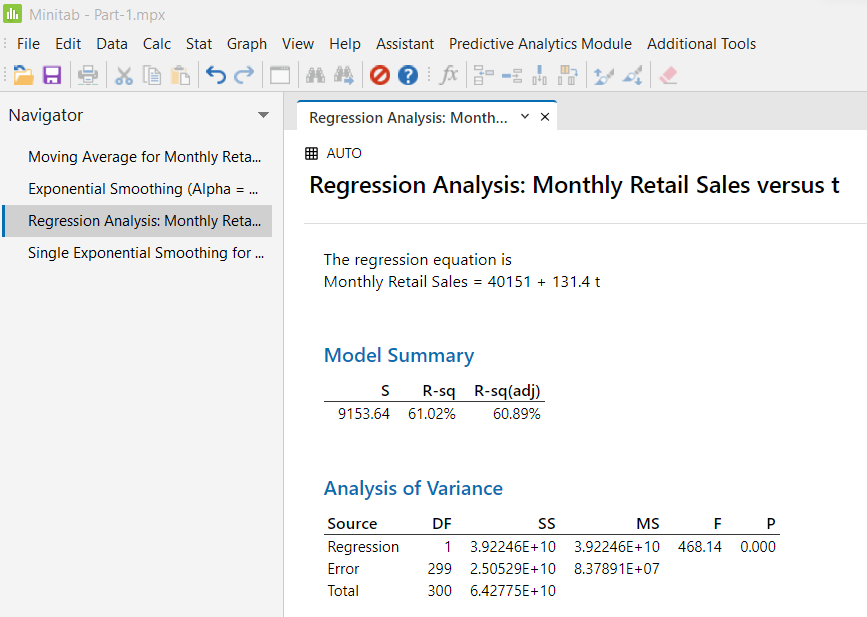
* **Moving Average (the number of periods = 3)**

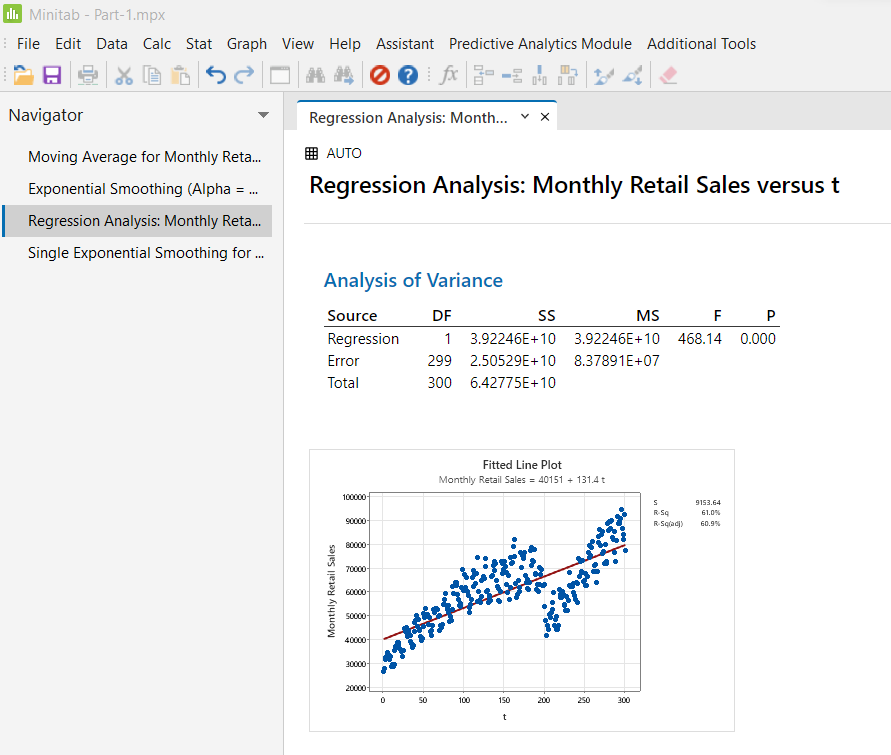


* **Exponential Smoothing (Alpha = 0.72)**



* **Simple Regression (where x represents the month)**





For calculating MAPE,

* Monthly Retail Sales = 40151 + 131.4(302)

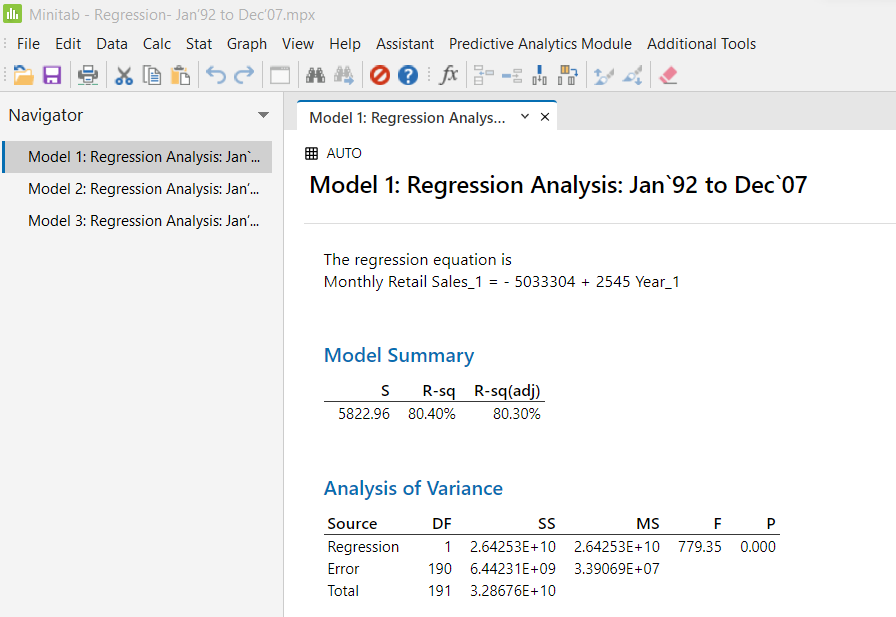
Therefore, predicted Monthly Retail Sales = 79833.8

MAPE = (|(77273.3 – 79833.8)|)/ 77273.3 = 2600.5/77273.3 = 0.0337 = 3.37%.

Explanation: In case of Moving average, we get the result in which include MAPE equals to 4% and the forecasted value of period 302 is 83811.4 (or error is -6538.2). After that, the next method comes, exponential smoothing, we get the result in which MAPE is 6%, the optimal alpha value is 0.72, and the forecasted value of period 302 is 80738.2 (or error = -12404.4). On the other hand, the R-square of the regression model is 61.02% and the MAPE is 3.37%.

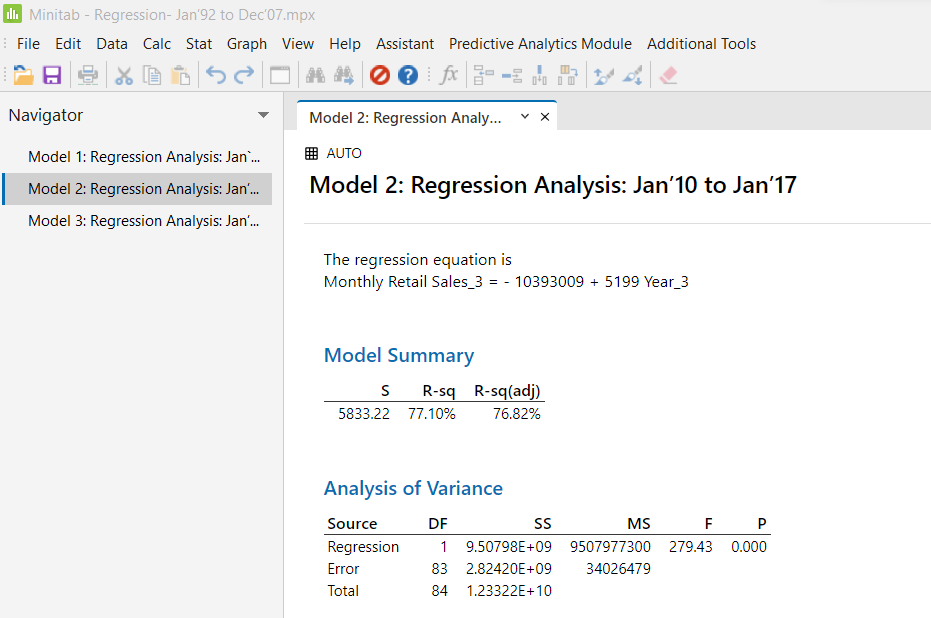
The rule of MAPE (mean or average of the absolute percentage errors of forecasts) is a lower MAPE is better. So, we can easily conclude that the regression model is better or more accurate model because its MAPE is lower as compared to other models.

* Now, we are going to spilt the data set into two from Jan’92 to Dec’07 and from Jan’10 to Jan’17.
* First, we run a regression (in Minitab) for the data from Jan’92 to Dec’07 (see output below).



The regression equation is  
Monthly Retail Sales\_1 = - 5033304 + 2545 Year\_1

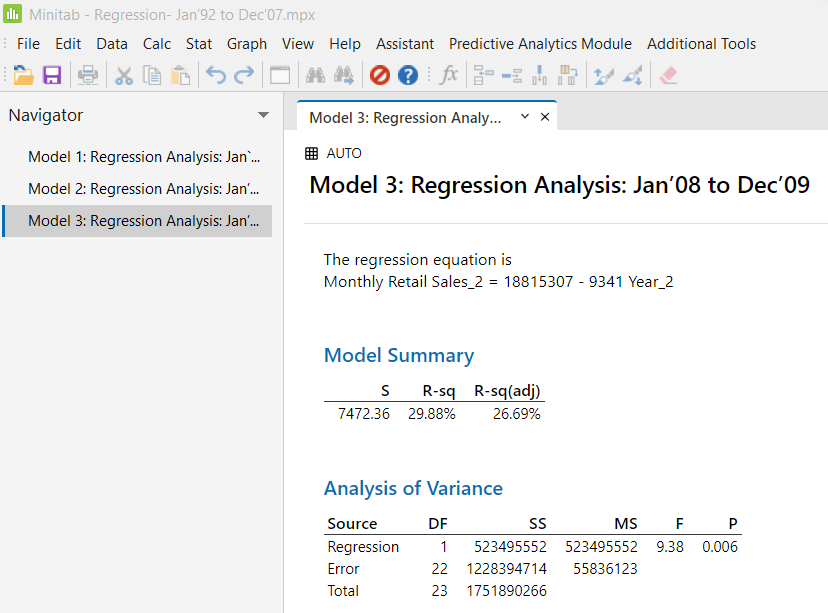
* Again, we run regression (in Minitab) for the data from Jan’10 to Jan’17 (see output below).



The regression equation is  
Monthly Retail Sales\_3 = **–** 10393009 + 5199 Year\_3

After performing both the regressions, we get R-square in both the regression, so in the regression model 1, the R-square is 80.40% whereas the R-square in model 2 is 77.10%. The interpretation of R-square is that 80% of the variation in the dependent variable can be explained by the variation in the independent variable. Therefore, higher the R-square better the model. So, we can easily conclude that the model 1 is better or more accurate as compare to model 2.

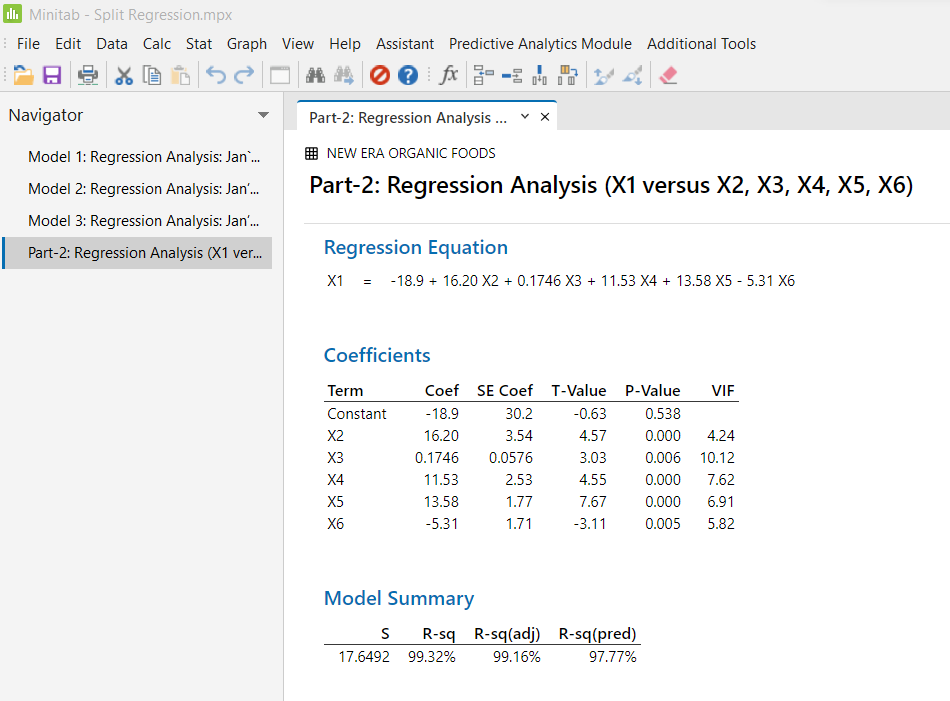
* Now, we again run the regression for the data from Jan’08 to Dec’09 (see output below).

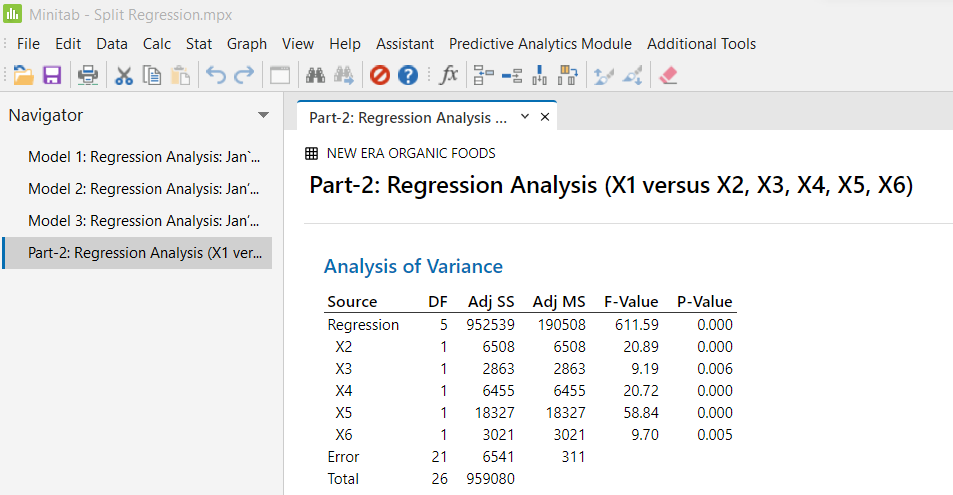


The regression equation is  
Monthly Retail Sales\_2 = 18815307 - 9341 Year\_2

In this case, the coefficient parameter is changed from positive to negative. Also, the R-square of the model is too lower as compared to the other model.

**Part-2: New Era Organic Foods**





**Regression Equation**

X1 = -18.9 + 16.20 X2 + 0.1746 X3 + 11.53 X4 + 13.58 X5 - 5.31 X6

Where, X1 = annual net sales/$1000

X2 = number sq. ft./1000

X3 = inventory/$1000

X4 = amount spent on advertising/$1000

X5 = size of sales district/1000 families

X6 = number of competing stores in district

In this model, our dependent variable is annual net sales (X1) whereas the independent variables are X2, X3, X4, X5, and X6. The regression equation is shown above.

Interpretation of all the slopes:

* The slope, X2 = 16.20 implies that each increase of 1000 square feet, then the value of annual net sales is estimated to increases by $16.20.
* The Slope, X3 = 0.17, implies that for each increases of inventory per 1000, then the value of net sales is estimated to increases by $0.17.
* The slope, X4 = 11.53 implies that each increase of 1000 amount spent on advertising, then the value of annual net sales is estimated to increases by $11.53.
* The slope, X5 = 13.58 implies that each increase of 1000 size of sales district, then the value of annual net sales is estimated to increases by $13.58.
* The slope, X6 = 5.31 implies that each increase of 1 number of competing stores in district, then the value of annual net sales is estimated to decrease by $5.31.

One interesting thing can be observed in the result is that, all the variance of inflation (VIF) is less than the 10 but the VIF of independent variable is inventory (X3) which implies that it indicates that multicollinearity exist, so it is better to remove X3 variable.

**Model Diagnosis:**

As we see the p-value of all the independent variables is close to zero and the rule of p-value is that when the p-value is less than the level of significance (or alpha level) which implies that the variable is statistically significant, otherwise insignificant. Thus, all the variables are statistically significant as 0.000 is less than 0.05.

On the other hand, the *R-square* of the model is close to 100% which implies that the 99% of the variation in the annual net sales (i.e., dependent variable) can be explained by the variation in the independent variable.