**RETAIL ANALYSIS BY AJAY**

**1.Import the required dataset**

**Code:**

**/\* Retail Analysis \*/**

**/\* Step 1: Import data by name of Retail\_Analysis \*/**

**FILENAME REFFILE '/home/ajaykasanna0/projects/Project 04\_Retail Analysis\_Dataset.xlsx';**

**PROC IMPORT DATAFILE=REFFILE**

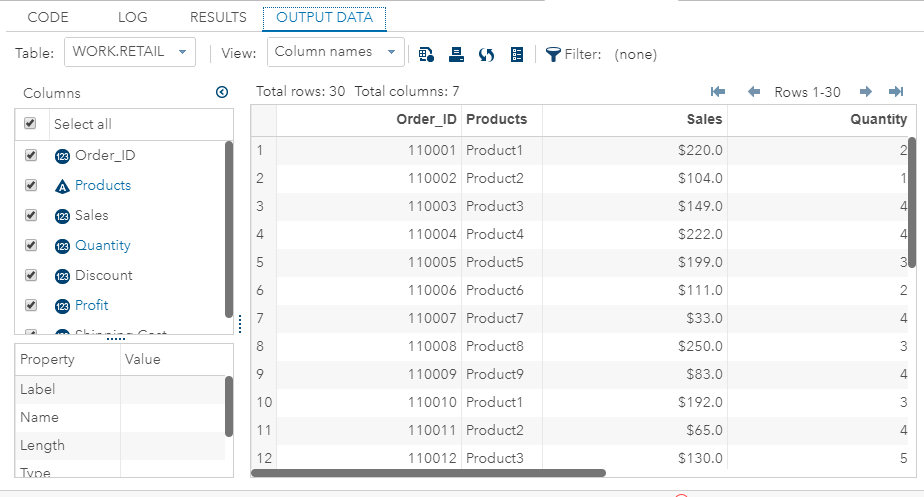
**DBMS=XLSX**

**OUT=WORK.Retail;**

**GETNAMES=YES;**

**RUN;**

**OUTPUT:**



**Step2**:perform descriptive statistics for dataset

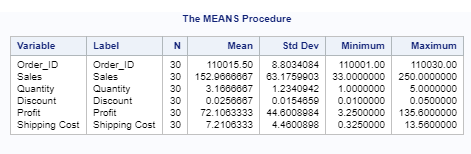
**Code**:

/\*step 2: descriptive statistics \*/

proc means data=work.retail;

run;

OUTPUT:



Step 3:check the significance of variables

Code:

/\* Create a new variable Total\_Sales = sales\*quantity as sales $ value is just per unit price \*/

data Retail;

set Retail;

Total\_Sales = sales\*quantity;

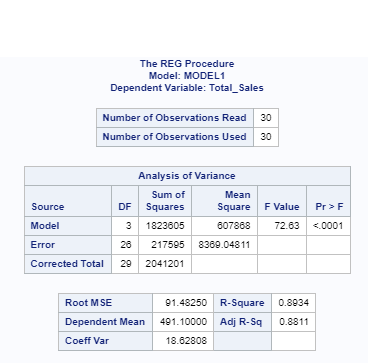
run;

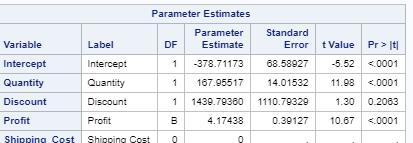
PROC REG DATA=Retail;

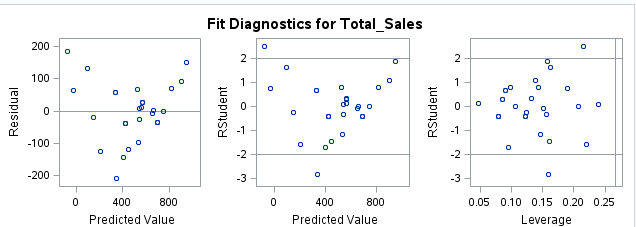
MODEL Total\_Sales= Quantity Discount Profit Shipping\_Cost;

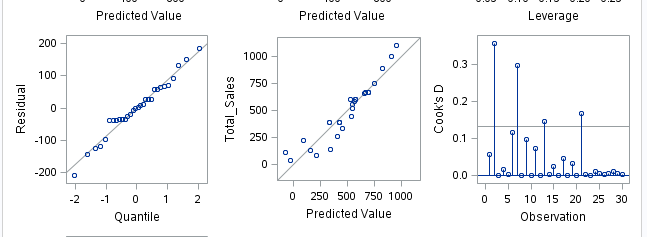
run;

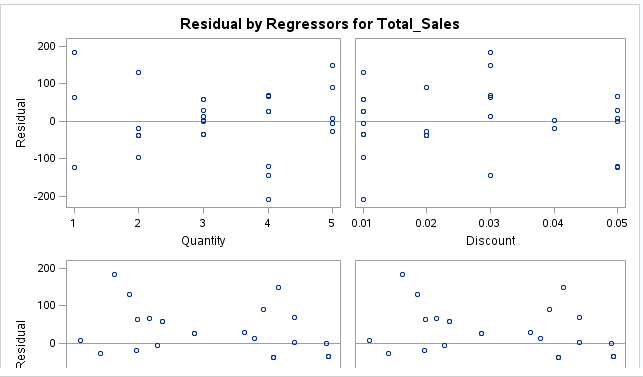
OUTPUT:











->significant variables are p<0.5 are known as significant variables.

**Step4:** Create a new data set with exponential, cube, squared, and log values for each variable

**CODE:**

/\* step3:Create a new data set with exponential, cube, squared, and log values for each variable \*/

DATA sales\_computation;

SET work.retail;

Sales\_log = LOG(sales); /\* log base e \*/

Sales\_exp = EXP(sales); /\* e raised to the power \*/

pow = 2;

Sales\_sq = sales \*\* pow ; /\* square of value\*/

pow = 3;

Sales\_cube = sales \*\* pow; /\*cube of value\*/

RUN;

PROC PRINT DATA=sales\_computation;

VAR sales sales\_log sales\_exp sales\_sq sales\_cube;

RUN;

/\* Step 4: As per result output: Since shipping cost is straight away 10% of profit across all

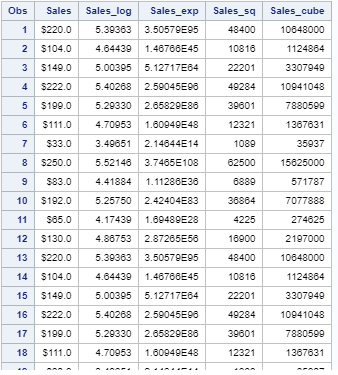
observations this variable (Shipping cost) should be dropped \*/

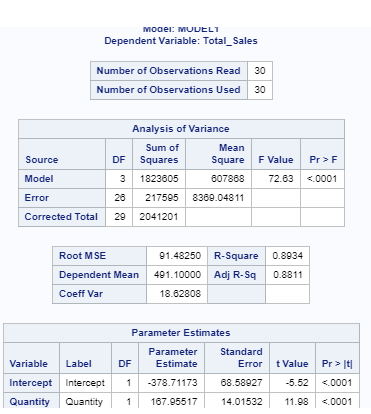
PROC REG DATA=Retail;

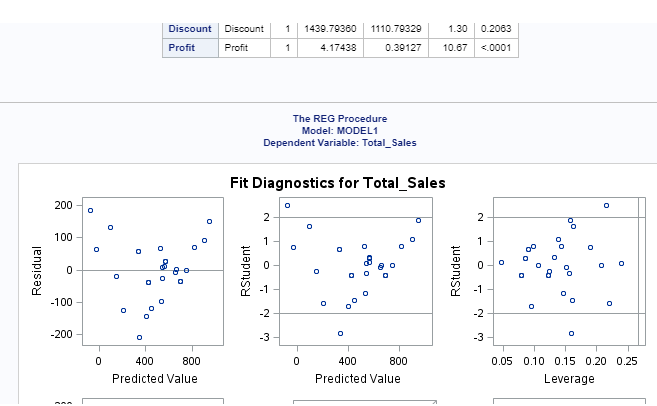
MODEL Total\_Sales= Quantity Discount Profit ;

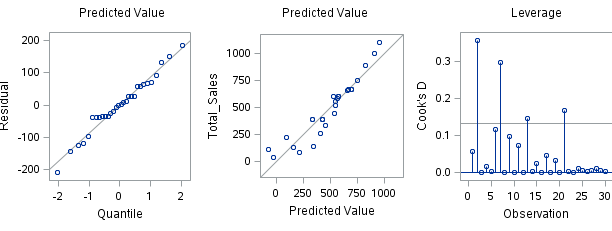
run;

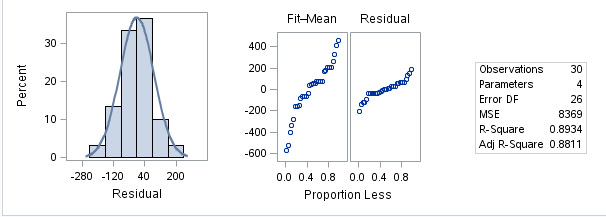
OUTPUT:











STEP5: Perform regression test

CODE:

/\* Now predict total\_sales and not sales \*/

PROC REG DATA=Retail;

MODEL Total\_Sales= Quantity Discount Profit Shipping\_Cost;

run;

/\* As per result output: Since shipping cost is straight away 10% of profit across all

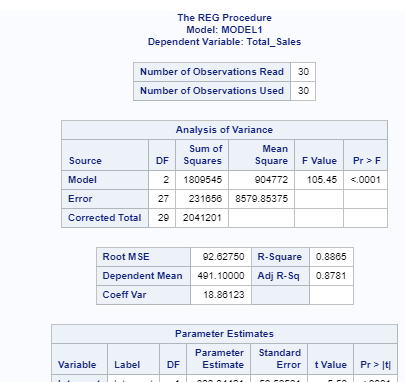
observations this variable (Shipping cost) should be dropped \*/

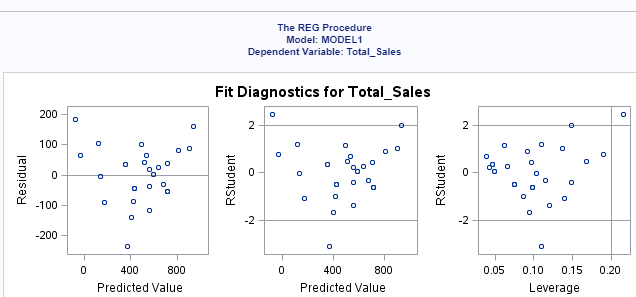
PROC REG DATA=Retail\_Analysis;

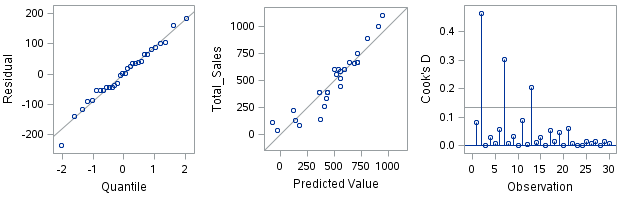
MODEL Total\_Sales= Quantity Discount Profit ;

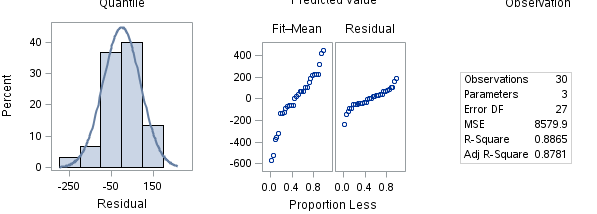
run;

OUTPUT:









STEP6:print the output of dataset

Code:

/\* You can also see the prediction in output dataset "\_SalesPrediction" and compare it with

Actual Total\_Sales and gauge the accuracy of model\*/

PROC REG DATA=Retail;

MODEL Total\_Sales= Quantity Profit ;

output out = \_SalesPrediction p=Predicted\_Total\_Sales;

run;