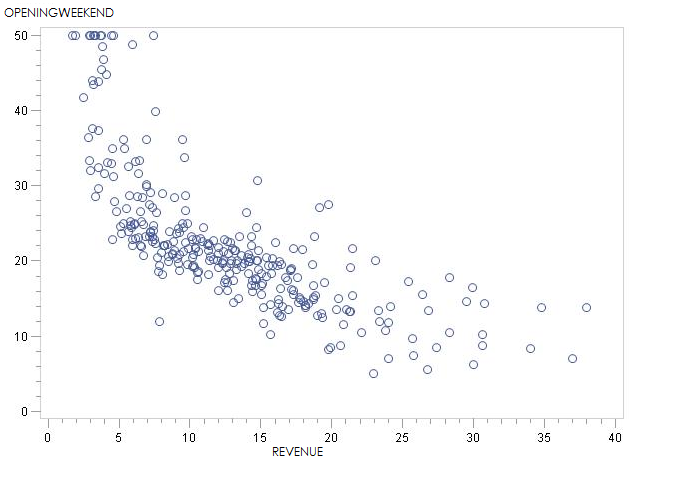
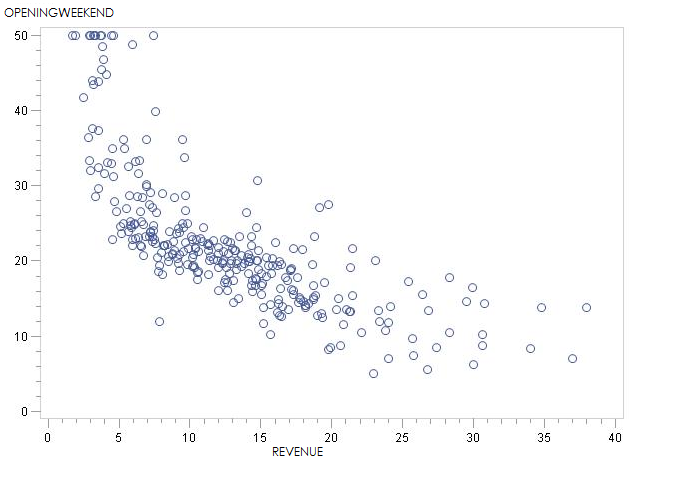
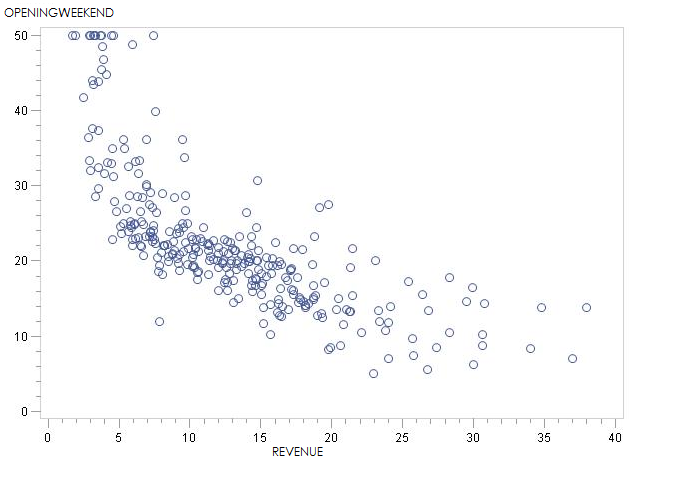
**REGRESSION ANALYSIS:**

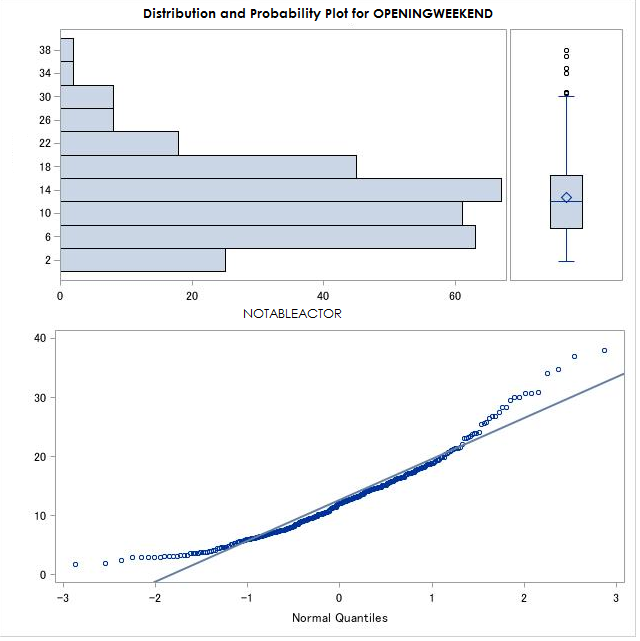
|  |  |
| --- | --- |
| **OPENING WEEKEND Vs REVENUE SCATTER PLOT** |  |
| Input Data Set | Unicorns |
| Random Number Seed | 49201 |
| Sampling Rate | 0.59 |
| Sample Size | 299 |
| Selection Probability | 0.590909 |
| Sampling Weight | 0 |
| Output Data Set | MOVIE |
|  |  |







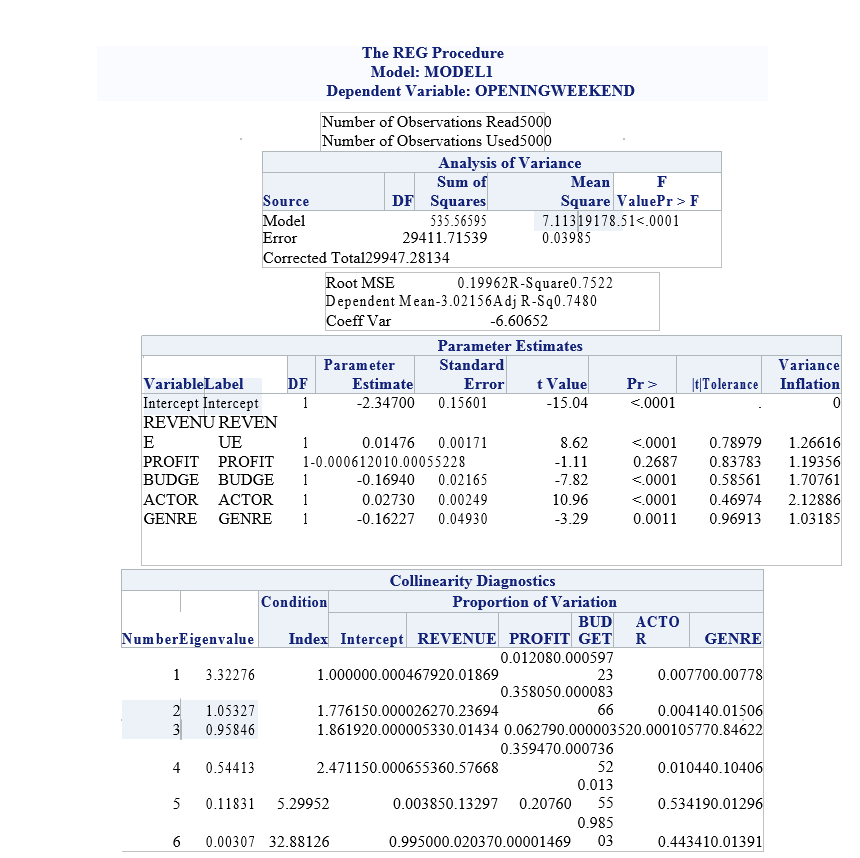
We would conclude that REVENUE seems to have a linear negative impact on OPENINGWEEKEND. However, the outliers seem to muddle this observation, since several outliers (large) values of OPENINGWEEKEND are bunched for small values of REVENUE. The example probably shows how cleaning of outliers might be a good idea before proceeding with regression.



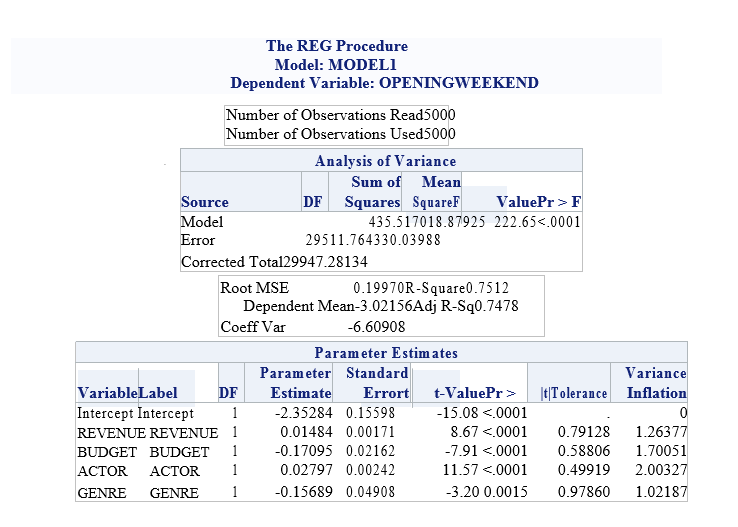
**Extreme Observations**

Lowest Highest

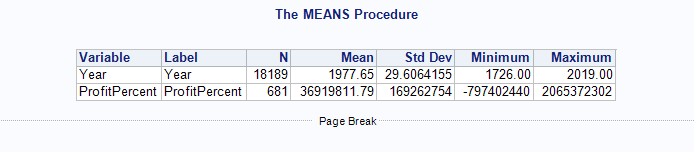
|  |  |  |  |
| --- | --- | --- | --- |
| Value | Obs | Value | Obs |
| 1.73 | 99 | 30.81 | 28 |
| 1.92 | 100 | 34.02 | 254 |
| 2.47 | 141 | 34.77 | 221 |
| 2.87 | 117 | 36.98 | 242 |
| 2.88 | 123 | 37.97 | 222 |



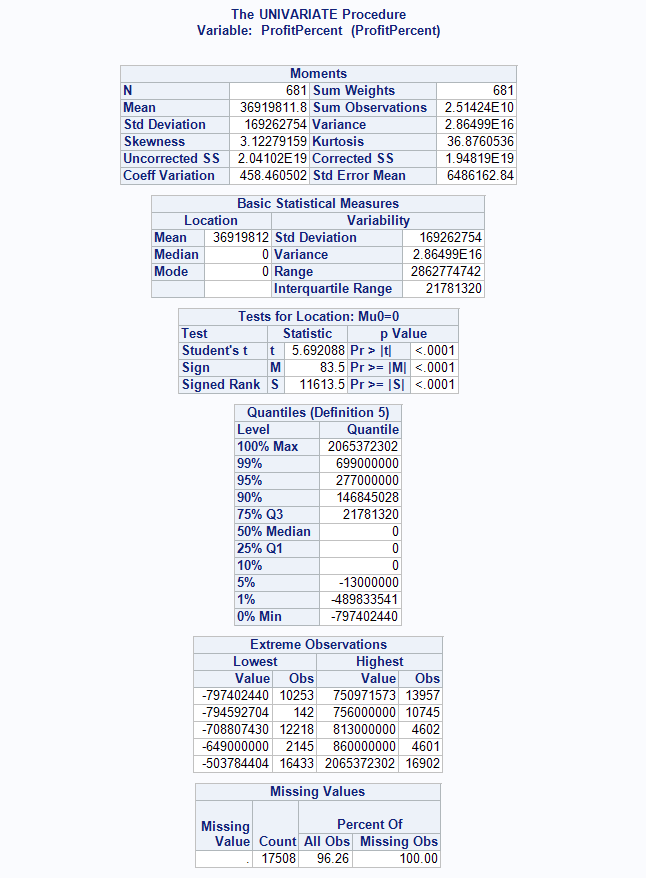
Based on the t-value we expect that the PROFIT variable would not explain the variance in the dependent variable OPENINGWEEKEND and therefore could be dropped the MLR model. The VIF and collinearity diagnostics looks good.



Means procedure:



Univariate:



**SAS CODE:**

**proc** **import** datafile="C:\Users\nshaik1\Downloads\Unicorn.xls" dbms=excel out=MOVIE;

**run**;

**proc** **means** data=MOVIE;

var year ProfitPercent;

**run**;

**proc** **univariate** data=movie;

var ProfitPercent Actor;

**run**;

**proc** **corr** data=movie;

var ProfitPercent Revenue NotabaleActors;

**run**;

**proc reg** data=Movie;

Model OPENINGWEEKEND = REVENUE PROFIT BUDGET ACTOR GENRE / tol vif collin;

plot r.\*p. student.\*nqq.;

**run**;