Monthly Product Sales Descriptive Statistics

Lawrence Wilson

MIS543 – Enterprise Performance Management

Colorado State University – Global Campus

Dr. Richard Morse

June 23, 2019

Monthly Product Sales Descriptive Statistics

For this assignment, SAS University Edition was used to conduct basic descriptive analysis on the sample dataset prod\_sales.csv. Not much metadata was provided for this dataset except for the variable names which consist of the total collected monthly sales and the promotional expenses in dollars for radio, newspaper, point of sale, and television advertisements within each respective sample month. In order to complete this assignment, SAS procedures were used to create table outputs of the dataset and its respective general descriptive statistics, discover correlation by generating scatter plots between the promotional expenses and total monthly collected sales, and test for normality relevant to correlation tests in order to discover if the assumption held true. The following SAS code was written to complete this assignment.



\*Module 2 Critical Thinking Assignment - Lawrence Wilson 6.23.2019;

DATA MONTHLY\_PRODUCT\_SALES;

INFILE '/folders/myfolders/prod\_sales\_mod2.CSV' DLM=',' DSD FIRSTOBS=2;

INPUT MONTH $ TOTAL\_MONTH\_SALES RADIO\_AD\_EXPENSES PAPER\_AD\_EXPENSES TV\_AD\_EXPENSES POS\_AD\_EXPENSES;

TOTAL\_EXPENSES = RADIO\_AD\_EXPENSES+PAPER\_AD\_EXPENSES+TV\_AD\_EXPENSES+POS\_AD\_EXPENSES;

NET\_SALES = TOTAL\_MONTH\_SALES-TOTAL\_EXPENSES;

PROC PRINT DATA = MONTHLY\_PRODUCT\_SALES;

TITLE "MONTHLY PRODUCT SALES DATA";

RUN;

TITLE "SIMPLE DESCRIPTIVE STATISTICS";

PROC MEANS MAXDEC=2 DATA = MONTHLY\_PRODUCT\_SALES N MEAN MEDIAN STD MIN MAX;

RUN;

PROC CORR DATA=MONTHLY\_PRODUCT\_SALES PLOTS=MATRIX;

VAR TOTAL\_MONTH\_SALES;

WITH RADIO\_AD\_EXPENSES PAPER\_AD\_EXPENSES TV\_AD\_EXPENSES POS\_AD\_EXPENSES;

TITLE "CORRELATED EXPENSES DATA";

RUN;

PROC CORR DATA=MONTHLY\_PRODUCT\_SALES PLOTS=MATRIX;

VAR TOTAL\_MONTH\_SALES;

WITH TOTAL\_EXPENSES;

TITLE "CORRELATED SALES DATA";

RUN;

PROC UNIVARIATE NORMAL PLOT DATA=MONTHLY\_PRODUCT\_SALES; VAR PAPER\_AD\_EXPENSES TV\_AD\_EXPENSES POS\_AD\_EXPENSES RADIO\_AD\_EXPENSES; HISTOGRAM PAPER\_AD\_EXPENSES TV\_AD\_EXPENSES POS\_AD\_EXPENSES RADIO\_AD\_EXPENSES/NORMAL;

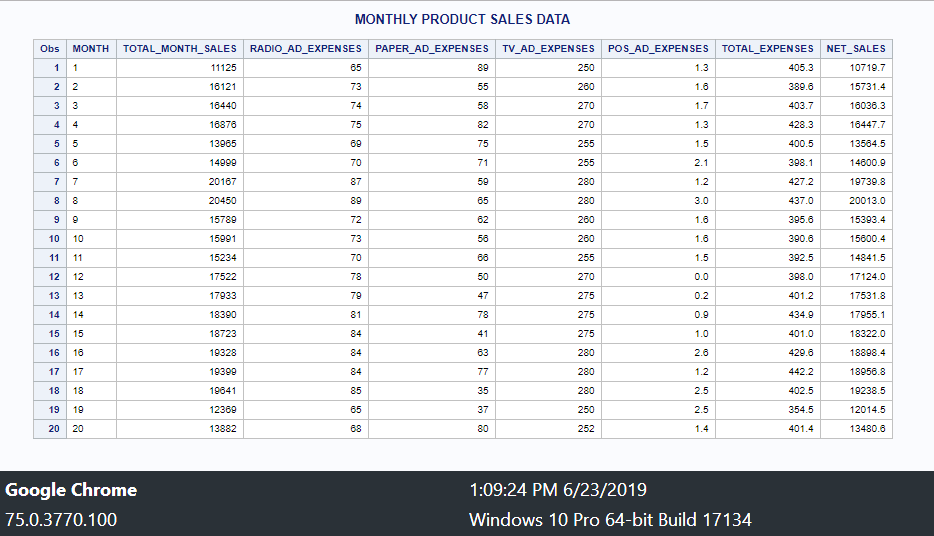
TITLE "NORMALITY TEST";

RUN;

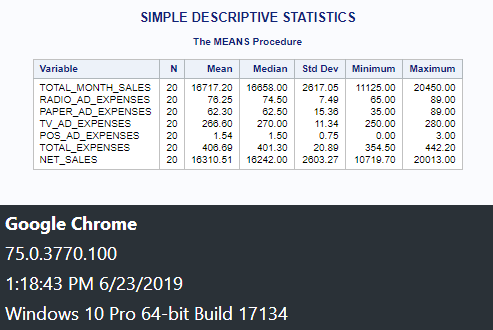
TITLE;FOOTNOTE;

**Description of PROC Statements**

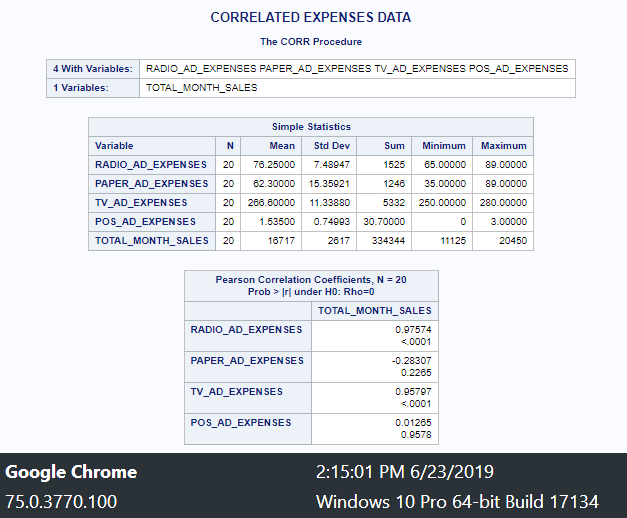
Firstly, two new derived variables named TOTAL\_EXPENSES and NET\_SALES were created by summing the promotional expenses and then subtracting the amount from TOTAL\_MONTH\_SALES. The PROC PRINT statement was used to make sure that data from prod\_sales.csv along with the newly created variables were correctly imported into SAS.

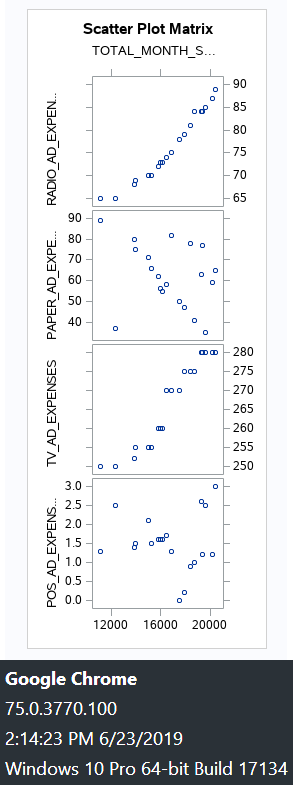


Secondly, the PROC MEANS statement was used to generate descriptive statistics of the dataset; the table consists of the total number of observations, mean, median, standard deviation, minimum, and maximum values.

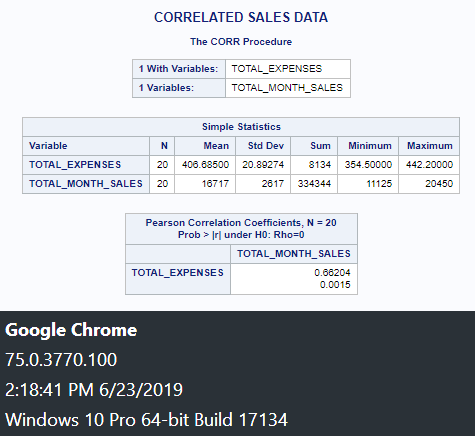


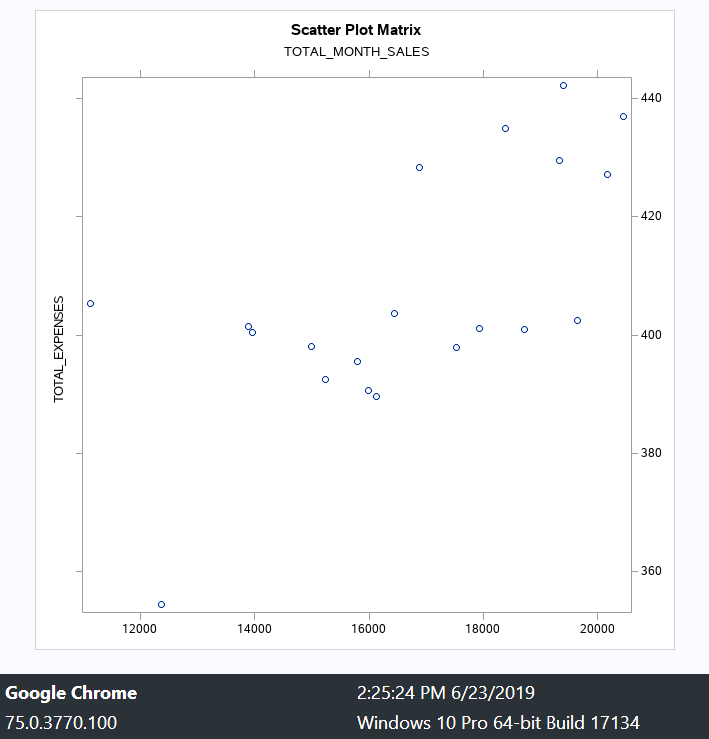
Thirdly, the PROC CORR statement was used to generate Pearson correlation coefficients and scatter plot matrices between the RADIO\_AD\_EXPENSES, PAPER\_AD\_EXPENSES, TV\_AD\_EXPENSES, and POS\_AD\_EXPENSES independent variables with the TOTAL\_MONTH\_SALES dependent variable.



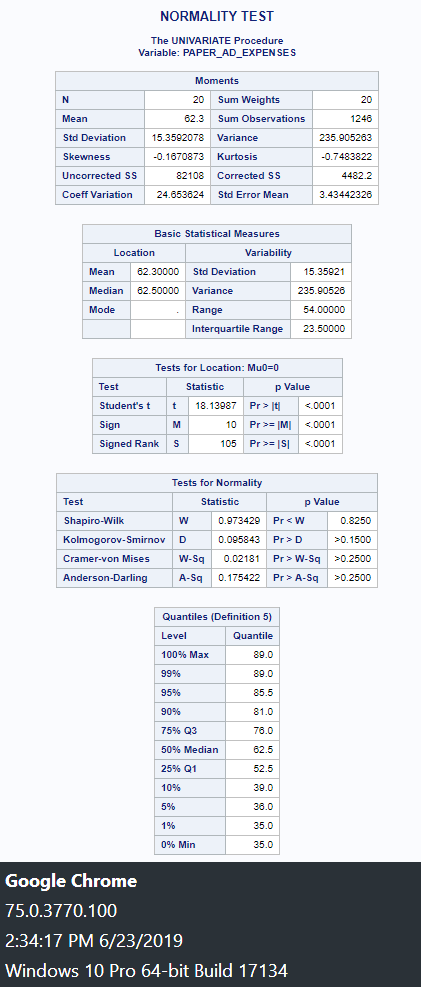


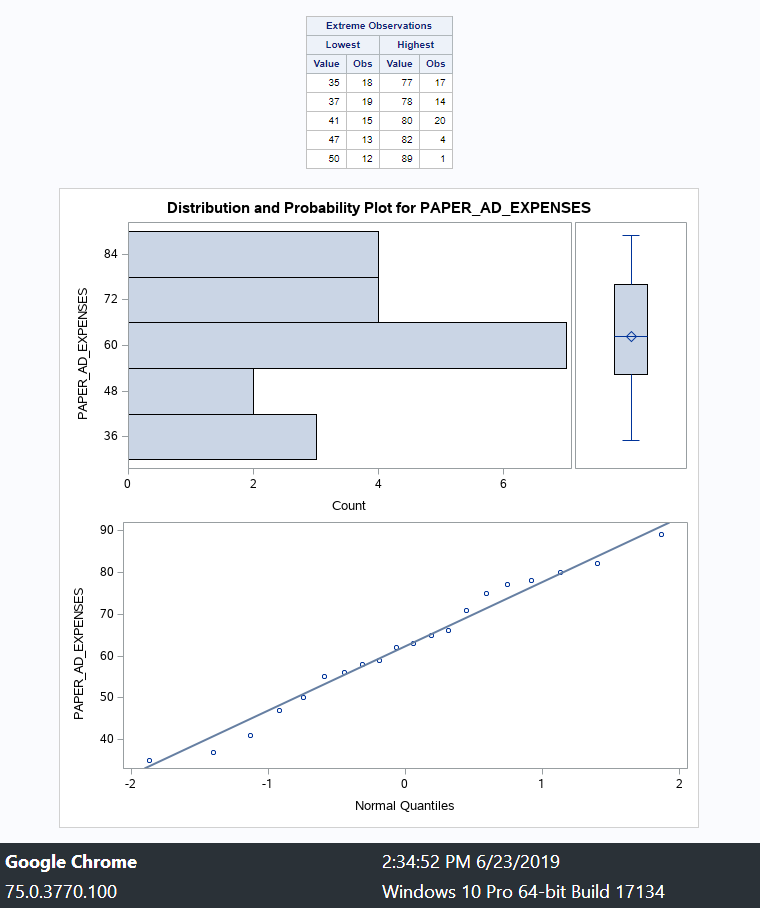
Next, another PROC CORR statement was used to generate Pearson correlation coefficients and scatter plot matrices between the derived TOTAL\_EXPENSES independent variable and the TOTAL\_MONTH\_SALES dependent variable.

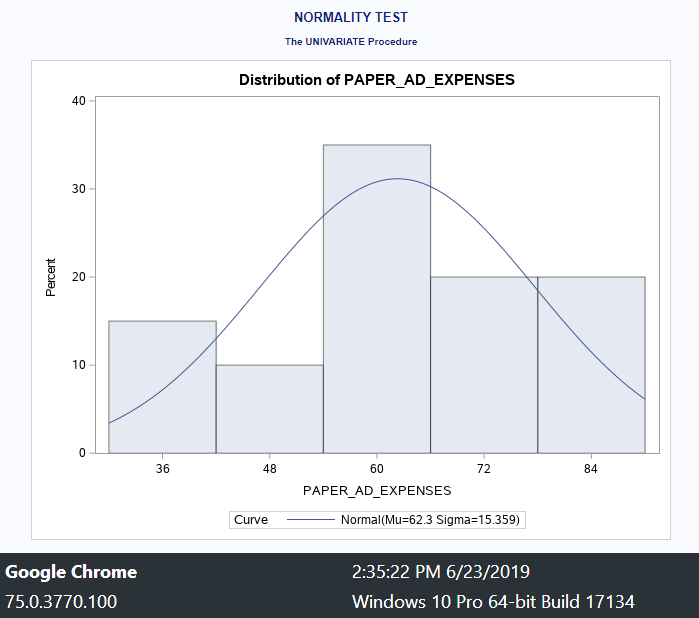


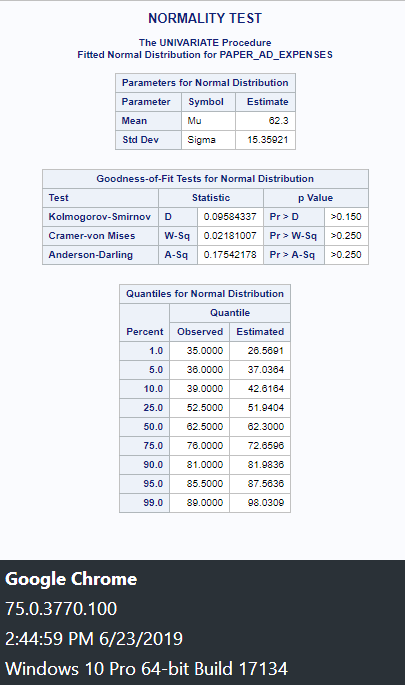


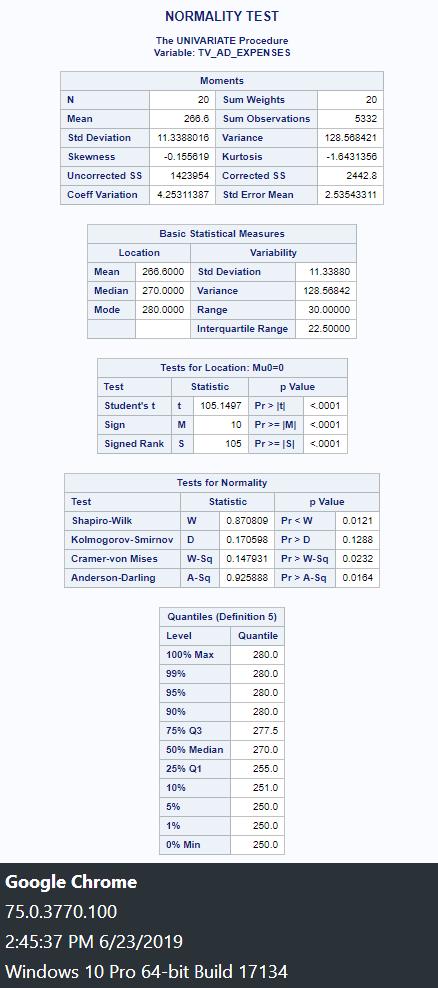
Finally, the PROC UNIVARIATE statement was used to complete normality tests on each of the initial dependent variables RADIO\_AD\_EXPENSES, PAPER\_AD\_EXPENSES, TV\_AD\_EXPENSES, and POS\_AD\_EXPENSES; histograms with a normalized curve were incorporated in the output as a data visualization of normality.

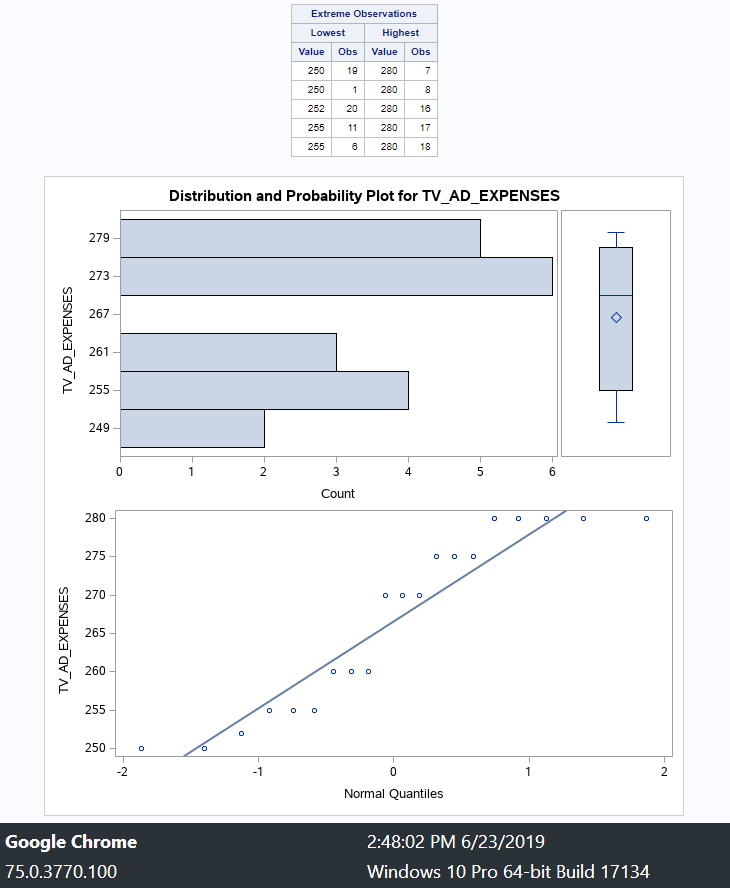


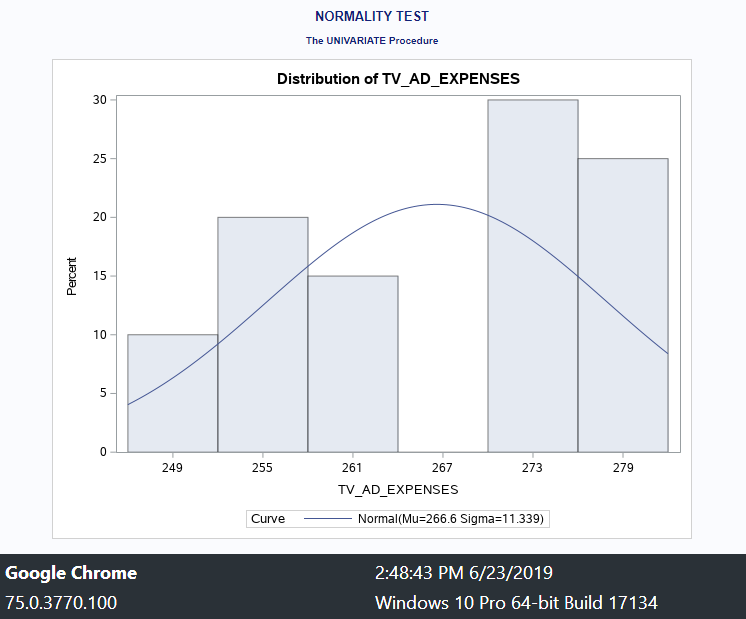


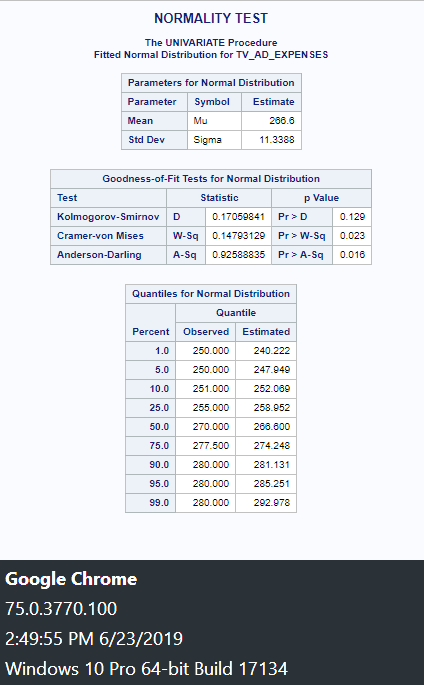


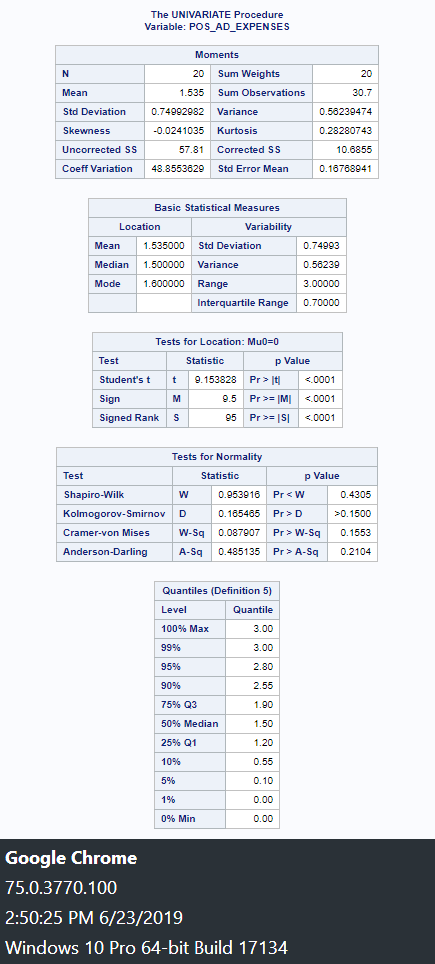


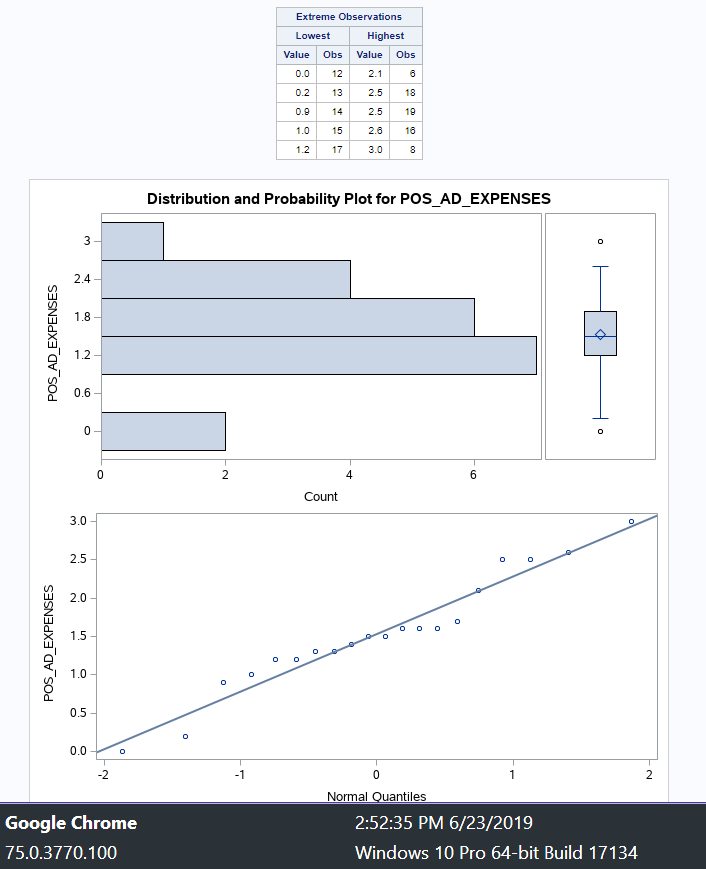


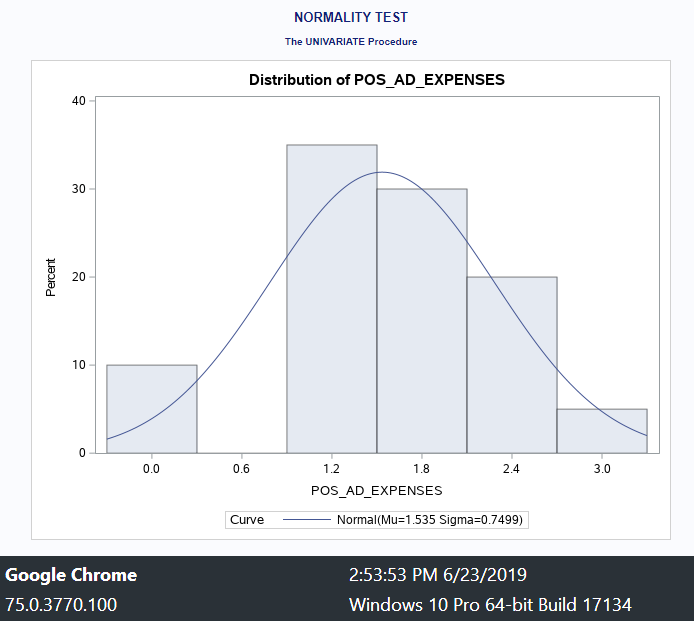


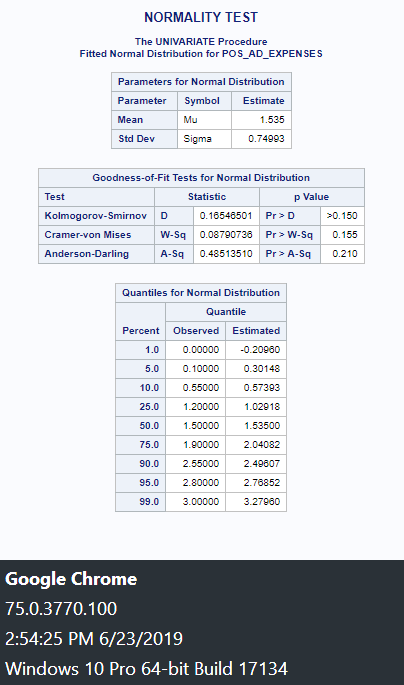


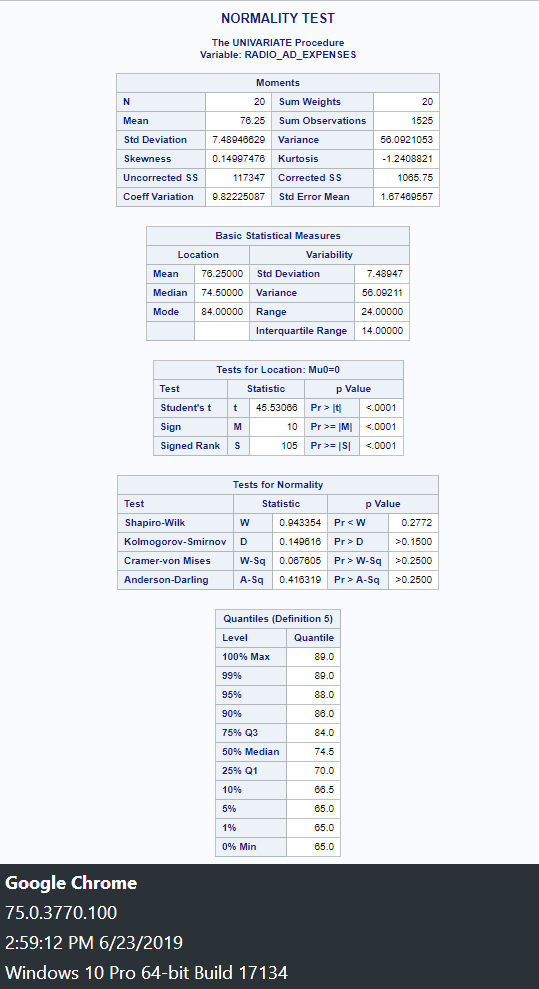


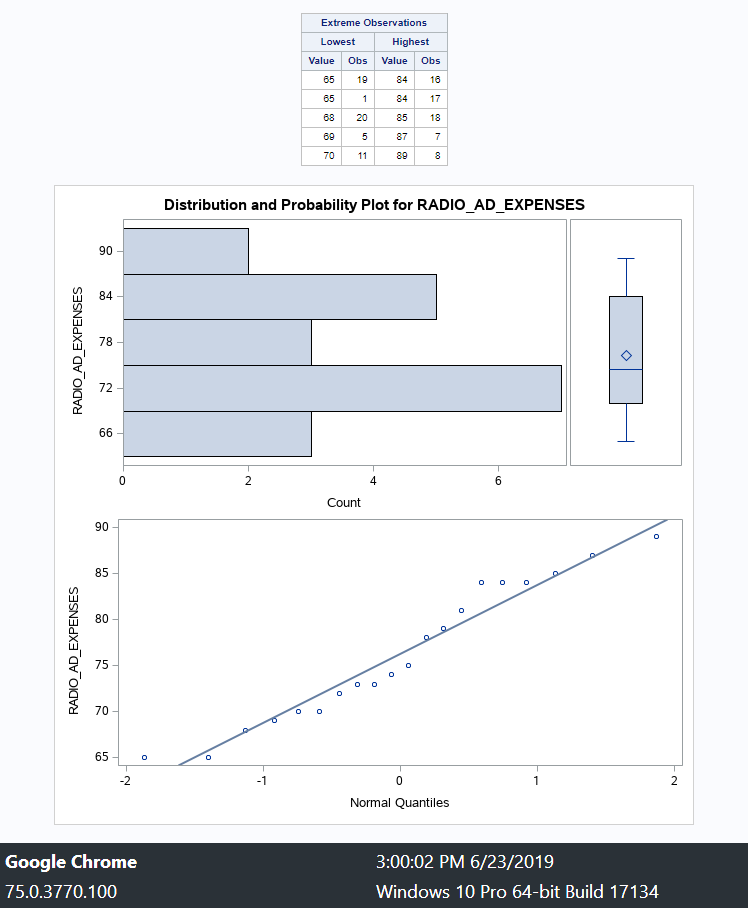


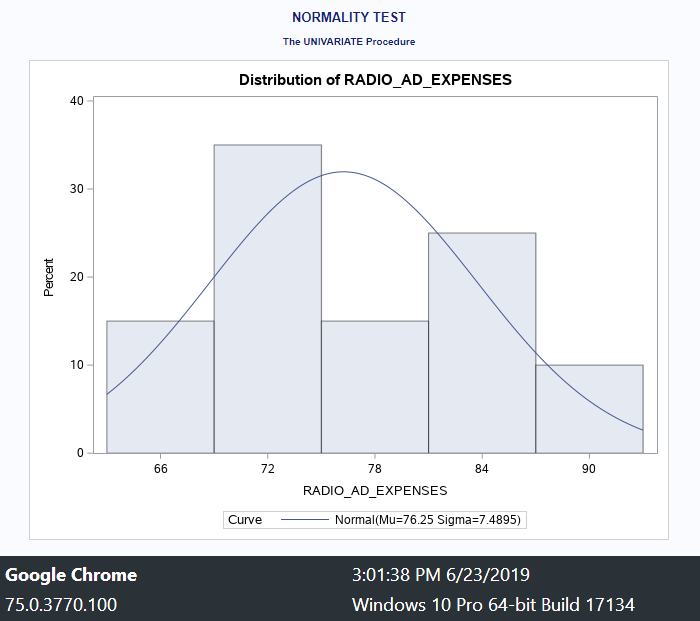


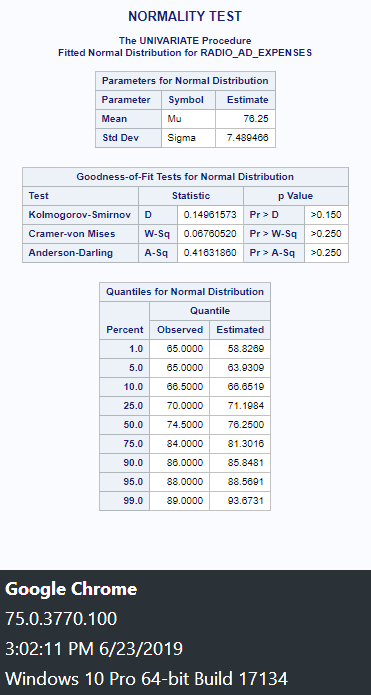












**Output Review and Business Questions**

In continuation, there are several ways that the organization may use the information obtained from the descriptive analyses. After all, descriptive statistics are graphical and numerical techniques that organize the dataset and describe its factors and characteristics (Fisher & Marshall, 2009). Much of descriptive statistics have to do with measuring central tendency and the spread of the respective scores such as variation and deviance (Fisher & Marshall, 2009). For example, one could see that the mean and median for each of the quantitative variables are relatively close together. If they were rather far apart, they may have indicated the presence of extreme outliers (Bartlett, 2013). However, it appears that the spending is rather controlled. In fact, the organization typically spends between 354.50 and 442.20 per month with a median of 401.30 and mean 406.69. Even if the minimum and maximum values of the dataset were unknown, one could simply see that by comparing the median and mean, very few outliers could be influencing the sample. One question that the organization can ask that may be answered from this data is whether there is a positive correlation between TOTAL\_EXPENSE and TOTAL\_MONTH\_SALES. Some people may believe in the old adage that one has to spend money to make money. Therefore, the question can be posed as a null hypothesis.

* H0: ρ = 0. Promotional expenses do not impact total monthly collected sales; there is no linear relationship.
* H*a*: ρ ≠ 0. Promotional expenses impact total monthly collected sales; there is a linear relationship.

By reviewing the result of the PROC CORR procedure between the TOTAL\_EXPENSES independent variable and TOTAL\_MONTH\_SALES dependent variable, Pearson’s *r* = .66204. Therefore, the null hypothesis can be rejected and there is a strong positive correlation between TOTAL\_EXPENSES and TOTAL\_MONTH\_SALES. In fact, this can be taken a step further to see exactly which of the respective independent promotional expense variables positively correlate with the total monthly sales collected. Firstly, the result of correlation test between the independent variable RADIO\_AD\_EXPENSES and the dependent variable TOTAL\_MONTH\_SALES is *r* = .97574. Secondly, the result of correlation test between the independent variable PAPER\_AD\_EXPENSES and the dependent variable TOTAL\_MONTH\_SALES is *r* = - .28307. Thirdly, the result of correlation test between the independent variable TV\_AD\_EXPENSES and the dependent variable TOTAL\_MONTH\_SALES is *r* = .95797. Finally, the result of correlation test between the independent variable POS\_AD\_EXPENSES and the dependent variable TOTAL\_MONTH\_SALES is *r* = .01265. Therefore, the only null hypothesis we cannot reject is that point of sale ads do not impact total monthly collected sales. In addition, it appears that spending money on paper advertisements results in a moderate negative correlation with TOTAL\_MONTH\_SALES. It can be argued that spending money on paper advertisements results in lower monthly sales. One may question how this is possible. Perhaps the dataset is not representative of a normal distribution in regards to PAPER\_AD\_EXPENSES. One of the assumptions of correlation tests is that errors are normally distributed (Ishiyama & Breuning, 2011). Although researchers may argue that this is the least important assumption, it is relevant in smaller datasets with less than 50 observations (Elliot & Woodward, 2015). Therefore, a normality test may be conducted on PAPER\_AD\_EXPENSES to eliminate the possibility that it does not hold true to a key statistical assumption.

* H0: ρ = 0. The sample does not reflect a normal distribution of paper advertising expenses.
* H*a*: ρ ≠ 0. The sample reflects a normal distribution of paper advertising expenses.

Each of the various tests for normality reflect that the null hypothesis may safely be rejected and that the distribution for PAPER\_AD\_EXPENSES is relatively normal. Therefore, it can be argued that the sample is indeed representative and for some reason it appears that spending money on paper advertisements results in less collected total monthly sales.

**Conclusion**

Although the results of the analyses reflect that the monthly total promotional expenses had a relatively strong positive correlation with the total collected sales each month, there is always the chance of a false positive. Factors such as the plausibility of the hypothesis must be taken prior to an experiment being done so that the false positive risk may be calculated and the *P* value is not misinterpreted (Nuijten, 2017). It can be assumed that there is a high possibility that total promotional expenses will be positively correlated with total collected sales as such a hypothesis is grounded in traditional microeconomic and marketing theory. However, one may question how plausible that paper advertisements would actually result in a negative correlation and less total collected sales. It is possible that the paper advertisements were atrocious and resulted in negative brand perception for the organization. Yet, is also plausible that it is a false positive and there may have been missing relevant variables. After all, one may readily believe that either paper advertisements had no effect on total collected sales for the month or there was a positive correlation; one would probably not have predicted that paper advertisements would result in a negative correlation. Furthermore, there is much about the dataset that is unknown as there is a lack of metadata. For example, it is unknown whether the 20 months of collected data reflect a year and eight months of continuously collected monthly data or whether the 20 months were randomly selected over a five-year period. This may be relevant in that the data could have been taken during different seasons. If all of the higher paper advertisement spending happened during low sale seasons, it may greatly influence the result. In fact, the organization may have decided to spend more heavily on television and radio advertising during high sale seasonal months such as Christmas season months while simply deciding to spend less on paper advertising. One of the most important assumptions in correlation testing is that all relevant dependent variables are included in the model and that irrelevant variables were excluded (Ishiyama & Breuning, 2011). In continuation, another issue is that there is no knowledge of how the data was collected. If the paper advertisements consisted of several thousands of flyers that were produced yet were never distributed due to lack of boots on the ground, one may question whether such data should be included in the analysis. In fact, there is no way to distinguish what constitutes paper advertising or how such data was collected without simply asking for the relevant metadata. Thus, although there appears to be a strong positive correlation between spending on advertising and total collected sales, there are still questions about variability, false positive risk, excluded relevant variables, and the metadata.

References

Bartlett, R. (2013). A practitioner’s guide to business analytics: Using data analysis tools to improve your organization’s decision making and strategy. New York, NY: McGraw- Hill.

Elliot, A.C., & Woodward, W.A. (2015) SAS essentials: Mastering SAS for analytics (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Fisher, & Marshall. (2009). Understanding descriptive statistics. *Australian Critical Care,* *22*(2), 93-97

Ishiyama, J., & Breuning, M. (2011). *21st century political science a reference handbook* (21st century reference series (Thousand Oaks, Calif.)). Los Angeles, [Calif.] ; London: SAGE.

Nuijten, M. (2017). Five ways to fix statistics: Share analysis plans and results. *Nature,* *551*, 557-559.