

**BACHELOR OF SCIENCE (HONS) STATISTICS**

**FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES**

**STA610**

**SAS PROGRAMMING**

**SALES RECORDED IN SEVERAL REGIONS**

**PREPARED BY**

|  |  |
| --- | --- |
| MIFTAKUL HUDA BIN MUCHAMAD IMRON | 2019329007 |
| MUHAMMAD FITRI BIN SALIM | 2019314589 |
| NORFAIZATULAH BINTI ABDULLAH | 2019314473 |

**PREPARED FOR**

MADAM NOOREZATTY BINTI MOHD YUSOP

**DATE OF SUBMISSION**

29th JUNE 2020

**TABLE OF CONTENTS**

[CHAPTER 1 : INTRODUCTION 3](#_Toc44191522)

[1.1 Background of Study 3](#_Toc44191523)

[1.2 Research Objectives 4](#_Toc44191524)

[1.3 Data Description 4](#_Toc44191525)

[CHAPTER 2 : METHODOLOGY 5](#_Toc44191526)

[2.1 Introduction 5](#_Toc44191527)

[2.2 Data Management 5](#_Toc44191528)

[2.2.1 SAS Data Library 5](#_Toc44191529)

[2.2.2 Import Data 5](#_Toc44191530)

[2.2.3 Creating New Data Set 6](#_Toc44191531)

[2.2.4 Reading Data from Existing Data Set 6](#_Toc44191532)

[2.2.5 Selecting Variables 6](#_Toc44191533)

[2.2.6 Sub-setting Data 7](#_Toc44191534)

[2.2.7 Formatting Data Values 7](#_Toc44191535)

[2.2.8 The Print Procedure 8](#_Toc44191536)

[2.2.9 Print Selected Variable 8](#_Toc44191537)

[2.2.10 Defining Title 8](#_Toc44191538)

[2.3 Descriptive Statistics 9](#_Toc44191539)

[2.3.1 Summary Statistics 9](#_Toc44191540)

[2.3.2 Summary Reports 9](#_Toc44191541)

[2.3.3 Tabulate Procedure 10](#_Toc44191542)

[2.3.4 Bar and Pie Chart 10](#_Toc44191543)

[2.4 Inferential Statistics 11](#_Toc44191544)

[2.4.1 Test of Normality 11](#_Toc44191545)

[2.4.2 Non-parametric Tests 11](#_Toc44191546)

[2.4.2.1 Wilcoxon Rank Sum Test (Mann-Whitney U Test) 11](#_Toc44191547)

[2.4.2.2 Kruskal-Wallis Test 12](#_Toc44191548)

[2.4.2.3 Chi-square Test 12](#_Toc44191549)

[CHAPTER 3 : RESULTS AND ANALYSIS 13](#_Toc44191550)

[3.1 Data Conversion 13](#_Toc44191551)

[3.2 Data Management 13](#_Toc44191552)

[3.3 Descriptive Analysis 15](#_Toc44191553)

[3.3.1 Frequency Report 15](#_Toc44191554)

[3.4 Summary Statistic 18](#_Toc44191555)

[3.4.1 Pie Chart 21](#_Toc44191556)

[3.4.2 Horizontal Bar Chart 22](#_Toc44191557)

[3.4.3 Vertical Bar Chart 24](#_Toc44191558)

[3.5 Inferential Statistic 25](#_Toc44191559)

[3.5.1 Normality Test on Total Revenue for Asia Region and Europe Region 25](#_Toc44191560)

[3.5.2 Mann-Whitney U Test to Compare the Differences Between Asia Region and Europe Region on Total Revenue 26](#_Toc44191561)

[3.5.3 Normality Test for Number of Items Sold 28](#_Toc44191562)

[3.5.4 Kruskal-Wallis Test to Determine Whether There is a Significant Mean Difference in the Number of Items Sold Among Item Types 29](#_Toc44191563)

[3.5.5 Chi-Square Test to Determine the Association Between Regions and Three Total Profit Levels. 31](#_Toc44191564)

[CHAPTER 4 : CONCLUSION 33](#_Toc44191565)

[REFERENCES 34](#_Toc44191566)

[APPENDICES 35](#_Toc44191569)

[Appendix A : Data Management 35](#_Toc44191570)

[Appendix B : Descriptor Portion 36](#_Toc44191571)

[Appendix C: Printing Partial Dataset 40](#_Toc44191572)

[Appendix D : Normality Test for Total Revenue between Asia and Europe 42](#_Toc44191573)

[Appendix E : Normality Test for Number of Items Sold 43](#_Toc44191574)

[Appendix F : Box Plot for Total Revenue between Asia and Europe 44](#_Toc44191575)

[Appendix G : Box Plot for Number of Items Sold among Item Types 45](#_Toc44191576)

# INTRODUCTION

## Background of Study

A sale is defined as a transaction that occurs between two or more parties in which the buyer receives tangible or intangible goods, services or assets in exchange for money. In other word, a sale is essentially a contract between the buyer and the seller of a particular good or service in question. Sales can also be completed between businesses such as when one raw materials provider sells available materials to a business that uses the materials to produce consumer goods. This led to a further action into recording a sale that happened as a reference for business purposes in the future. Sales records can be defined as the information that are collected on customers, including but not limited to their contact information, how often they purchase and what they purchase. Recording a sale is important since it is hard to remember key personal and professional details about the customers that have been purchasing goods and services from a company. The information and details gained from each occurring sale may help a company to make better decision in generating more profit in the future.

The secondary dataset in this study is about transactions that have been successfully completed in selling distinct item types to other parties by different countries from different regions. The original dataset consists of fourteen variables that were recorded in a Microsoft Excel sheet, but the researchers only choose five variables that are related to the objectives that are to be achieved which are region, item type, number of items sold, total profit and total revenue. The researchers also decided to use 500 observations out of 1000 observations as the sample for this study. The main criteria of the records are the type of items sold by different countries and regions, the number of items sold, and the total profit and revenue obtained from the transactions. These variables are crucial in helping one country or region in the dataset to take appropriate measures in increasing their revenue and profit while considering on the type of item they are selling and the number of items sold. Therefore, by using these selected variables, a further study on the dataset should be done to determine the impact of selling different type of products and identify the profit standings for each region.

## Research Objectives

1. To compare the differences between Asia region and Europe region towards the total revenue made.
2. To identify whether there is a significant difference in total number of units sold among item types.
3. To determine the association between regions and three total profit levels.

## Data Description

The data used for this study is a secondary dataset that was obtained from <http://eforexcel.com/wp/downloads-18-sample-csv-files-data-sets-for-testing-sales/>. The dataset is about sales recorded in several regions. The original dataset consists of 14 variables but only 3 quantitative variables (Units Sold, Total Revenue, Total Profit) and 2 qualitative variables (Region, Item Type) were used in this study prior to the objectives of interest as given below.

Table 1.1 : Variable Descriptions

|  |  |  |
| --- | --- | --- |
| VARIABLES | DESCRIPTION | TYPE |
| Region | Regions in which sales had occurred | Nominal |
| Item\_Type | The type of items sold | Nominal |
| Units\_Sold | The total number of units sold | Numeric |
| Total\_Revenue | The total income received from the sales | Numeric |
| Total\_Profit | The total profit after deducting the total cost from the total revenue obtained | Numeric |

# METHODOLOGY

## Introduction

In this chapter, we will be discussing about all the method that will be applied in order to achieve to the success of this project. Methods discussed in methodology help to achieve the objective of this project.

## Data Management

### SAS Data Library

Libname statements is used to assign a libref to SAS library.

General form of the LIBNAME statement:

LIBNAME libret ‘SAS-library’;

*\*refer to appendix 1*

### Import Data

The data is import from excel document into SAS by using proc import.

General form of the IMPORT statement:

PROC IMPORT OUT= <libref> ‘SAS-data-set’

DATAFILE= ‘external-file-name’

DBMS=’identifier’;

RUN;

*\*refer to appendix 2*

### Creating New Data Set

Data statement will be used to create SAS data set name and the data created can be temporary or permanent data sets.

The general form of the DATA statement:

DATA ‘output-SAS-data-set’;

RUN;

### Reading Data from Existing Data Set

SET statement reads all observations and all variable from an existing SAS data set.

General form of the SET statement:

DATA ‘output-SAS-data-set’;

SET ’input-SAS-data-set’;

RUN;

### Selecting Variables

KEEP statement can be used to control the variable that want to attain in the new SAS dataset

General form of the KEEP statement:

DATA ‘output-SAS-data-set’;

SET ’input-SAS-data-set’;

KEEP variables;

RUN;

### Sub-setting Data

WHERE statement is a sequence of operators and operands. This statement used to select observations that meet condition needed.

General form of WHERE statement:

DATA ‘output-SAS-data-set’;

SET ’input-SAS-data-set’;

KEEP variables;

WHERE where-expression

RUN;

### Formatting Data Values

There are two methods can be use in assigning format to the data values which are formatted values and user-defined formats. For formatted values, FORMAT is used as instruction that SAS uses to write data values while user-defined formats used PROC FORMAT statement to change the attribute name, data format and so on. In this study, user-defined formats will be used.

The general form of PROC FORMAT;

PROC FORMAT;

VALUE format name range1= ‘label’

range2= ‘label’;

RUN;

### The Print Procedure

PRINT PROCEDURE will be applied to display titles, footnotes, description column headings, formatted data values, column totals, column subtotal and page break for each subgroup.

The general form of PROC PRINT:

PROC PRINT DATA = SAS-data-set;

RUN;

### Print Selected Variable

VAR statement will be used to select the variables that will be included in the report and it will arrange follow by its sequence.

The general form of VAR statement:

PROC PRINT DATA = SAS-data-set;

VAR variable(s);

RUN;

### Defining Title

TITLE statement used to as title of the report when printing.

The general form of TITLE statement:

TITLE “title”;

## Descriptive Statistics

### Summary Statistics

The MEAN PROCEDURE used to display simple descriptive statistic like N, MEAN, STD, MIN and MAX of the variable.

The general form of simple PROC MEANS;

PROC MEANS DATA = SAS-data-set;

RUN;

### Summary Reports

The PROC FREQ statement used to provide output like frequency, percent and cumulative percent.

The General form of PROC FREQ;

PROC FREQ DATA = SAS-data-set;

TABLES variable1\*variable2;

RUN;

### Tabulate Procedure

One and two-dimensional tabular reports can be created by using TABULATE procedure. The report will include control of table construction, specifying statistics and formatting values.

The general form of PROC TABULATE:

PROC TABULATE

DATA = SAS-data-set<option>;

CLASS classvariables;

VAR analysis-variables;

TABLES pageexpression;

RUN;

### Bar and Pie Chart

Chart such as bar chat and pie chart can be display in SAS Programming by using GCHART PROCEDURE. The type of chart needed can be specify by HBAR3D, VBAR or PIE STATEMENT.

General form of the PROC GCHART:

PROC GCHART DATA= SAS-DATA-SET;

Statements to specify the desired type of chart

HBAR3D chart-variable…</options>;

VBAR chart-variable…</options>;

PIE3D chart-variable…</options>;

## Inferential Statistics

### Test of Normality

Normality test is needed to determine either the data is normally distributed or not normally distributed. By obtaining the result, the variable can be classified into parametric or non-parametric test.

The general form of PROC UNIVARIATE:

PROC UNIVARIATE DATA = SAS-data-set;

VAR variable(s);

PROBPLOT/NORMAL (MU=EST SIGMA=EST);

RUN;

### Non-parametric Tests

#### Wilcoxon Rank Sum Test (Mann-Whitney U Test)

The Mann-Whitney Test is a useful nonparametric alternative to the independent T Test. It is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous but not normally distributed.

The general form of PROC NPAR1WAY:

PROC NPAR1WAY DATA = SAS-data-set WILCOXON;

CLASS variable(s);

VAR variable(s);

EXACT Wilcoxon;

RUN;

#### Kruskal-Wallis Test

Kruskal-Wallis Test used to determine if there are any significant differences between two or more groups of independent variables on a continuous or ordinal dependent variable. It is ANOVA for non-parametric data.

The general form of PROC NPAR1WAY:

PROC NPAR1WAY DATA = SAS-data-set WILCOXON;

CLASS variable(s);

VAR variable(s);

RUN;

#### Chi-square Test

Chi-Square test is used to examine the differences between categorical variables in the data set.

General form of the PROC FREQ:

PROC FREQ DATA = SAS-data-set;

TABLES variable\*variable/EXPECTED CHISQ FISHER;

RUN;

# RESULTS AND ANALYSIS

## Data Conversion

The original dataset is created in an Excel file that were retrieved from an online website, therefore the dataset is converted into a SAS dataset to be further analyze using the proc import step as given below.

**proc** **import** out = project.all

datafile = "H:\Users\Asus\Documents\Degree UiTM Seremban\PART 5\STA 610 (SAS)\Project\SalesRecords.xlsx"

dbms = excel;

**run**;

## Data Management

The dataset which consist of 1000 sales recorded as the population is partitioned into new dataset which only consist of 500 sales recorded since the researchers are interested in selecting 500 observations as the sample of this study. The first 500 observations are selected using the OBS statement. Then, the new sample dataset is partitioned into several datasets named project.AE, project.items and project.regions that are related to each objective of this study that the researchers are interested in achieving. The total profit is classified into three levels which are low, moderate and high for each region to be used on the third objective of this study. All the datasets are coded as below :

1. **data** project.sales;

set project.all (obs=**500**);

keep Region Item\_Type Units\_Sold Total\_Revenue Total\_Profit;

**run**;

1. **data** project.AE;

set project.sales;

keep Region Total\_Revenue;

where Region in ('Asia','Europe');

**run**;

1. **data** project.items;

set project.sales;

keep item\_type units\_sold;

**run**;

1. **proc** **format**;

value profitfmt

low -<**500000** = 'Low'

**500000**-<**1000000** = 'Moderate'

**1000000**-high = 'High';

**run**;

**data** project.regions;

set project.sales;

keep region total\_profit;

format total\_profit profitfmt.;

**run**;

## Descriptive Analysis

### Frequency Report

title 'Distribution of Regions';

**proc** **freq** data=project.sales nlevels;

tables Region;

**run**;

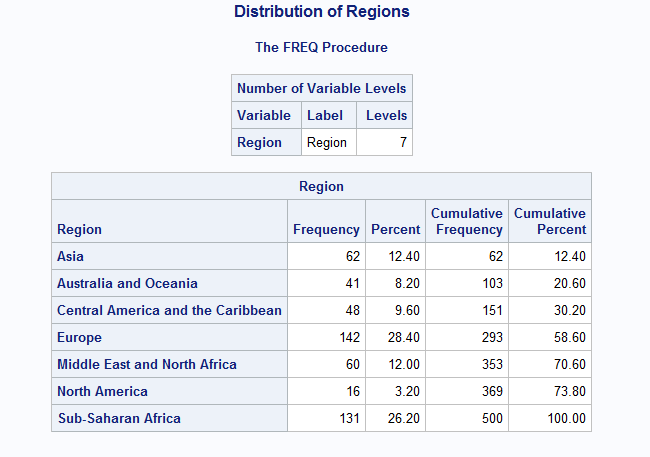


Figure 3.1 : The frequency report on distribution of regions

Interpretation :

The above frequency report shows that there exist seven different regions that were recorded in the sales dataset. It can be seen that Europe region has the highest number of sales occurred followed by Sub-Saharan Africa region. In contrast, North America region has lowest number of sales recorded.

title 'Distribution of Item Types';

**proc** **freq** data=project.sales nlevels;

tables Item\_Type;

**run**;

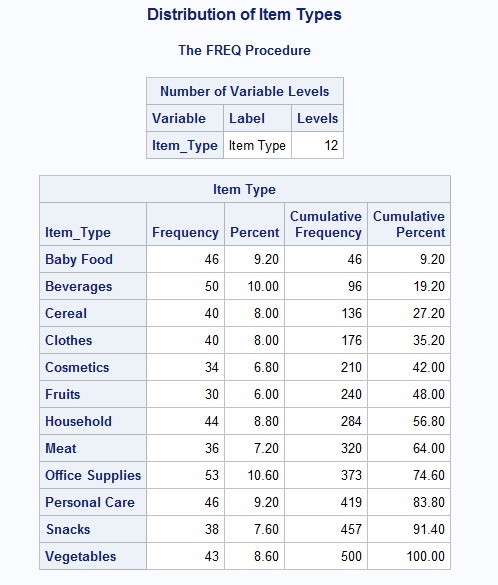


Figure 3.2 : Frequency report on distribution of item types

Interpretation :

Figure 3.2 indicates that there are 12 different item types that were managed to be sold by all regions. According to the figure above, office supplies is the most favourable type of item to be sold. In opposition, fruits have the lowest number of sales compared to other item types.

**proc** **format**;

value profitfmt

low -<**500000** = 'Low'

**500000**-<**1000000** = 'Moderate'

**1000000**-high = 'High';

**run**;

title 'The Distribution of Total Profit Levels by Regions';

**proc** **freq** data=project.sales;

tables Region\*Total\_Profit;

format Total\_Profit profitfmt.;

**run**;

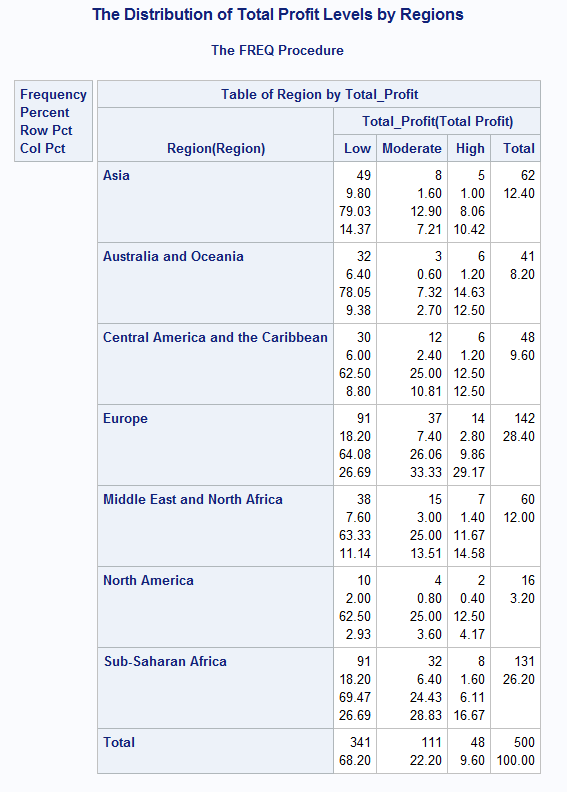


Figure 3.3 : Frequency report on the distribution of total profit levels by regions

Interpretation :

The above figure shows the distribution of regions towards 3 levels of total profit which are low, moderate and high. When a region successfully makes a profit between $0 until $499,000, it will be categorised in the low level. The profit that lies between $500,000 until $999,999 will be categorised in the moderate level and any total profit above that will be categorised in the high level. The result indicates that all the regions have the majority number of sales categorised in the low level.

## Summary Statistic

title 'Total Profit by Regions';

**proc** **means** data=project.sales;

var Total\_Profit;

class Region;

**run**;

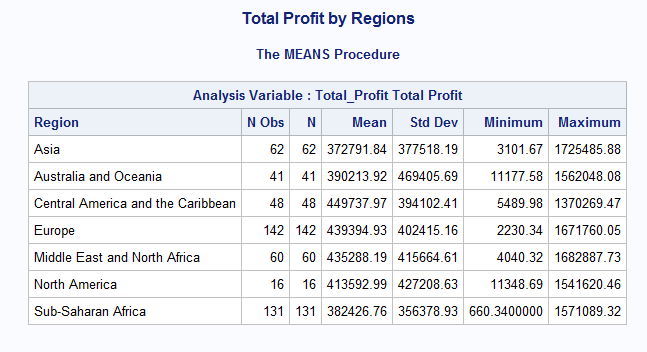


Figure 3.4 : Summary statistic on total profit by regions

Interpretation :

Figure 3.4 shows that the Central America and the Caribbean region has the highest average total profit of $449,737.97 followed by Europe region of $439,394.93. In contrast, Asia region has the lowest average total profit of $372,791.84.

Title ‘Total revenue by Asia Region and Europe Region’;

**Proc** **means** data=project.sales maxdec = **2**;

Var Total\_Revenue;

Class Region;

Where Region in (‘Asia’,’Europe’);

**Run**;

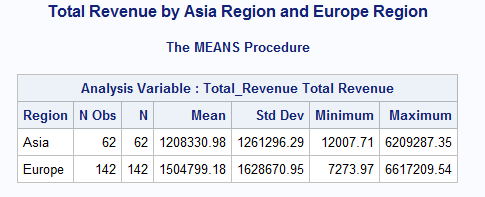


Figure 3.5 : Summary statistic on total revenue by Asia region and Europe region

Interpretation :

Figure 3.5 shows the comparison on summary statistic of total revenue obtained by Asia region and Europe region. It is revealed that the average revenue for Asia region is $1,208,330.98. Meanwhile, Europe region has a higher average of revenue which is $1504,799.18. The minimum and maximum revenue for Asia region is $12,007.71 and $6,209,287.35 respectively. Meanwhile, the minimum and maximum revenue for Europe region is $7273.97 and $6,617,209.54 respectively.

title1 'The Average of Total profit made by Asia Region and Europe Region';

title2 'For Each Item Type';

**proc** **tabulate** data=project.sales;

where region in ('Asia','Europe');

class region item\_type;

var total\_profit;

table item\_type, region\*total\_profit;

**run**;

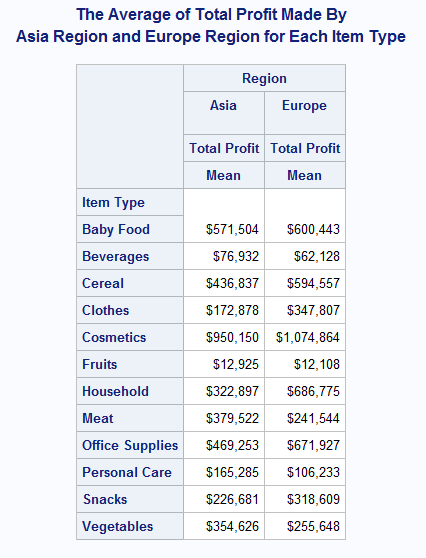


Figure 3.6 : The average of total profit made by Asia and Europe for each item type

Interpretation :

The above figure shows the average of total profit made by Asia and Europe for each item type. It can be indicated that both regions have the highest average of total profit on cosmetics item compared to other item types which are $950,150 and $1,074,864 respectively. The lowest average of total profit made by both regions is still on the same item type which is fruits valued $12,925 and $12,108 respectively.

### Pie Chart

title h=**2** f=broadway 'The Percentage Distribution of Sales between Asia and Europe';

**proc** **gchart** data=project.sales;

pie3d region / type=percent noheading;

where region in ('Asia','Europe');

**run**;

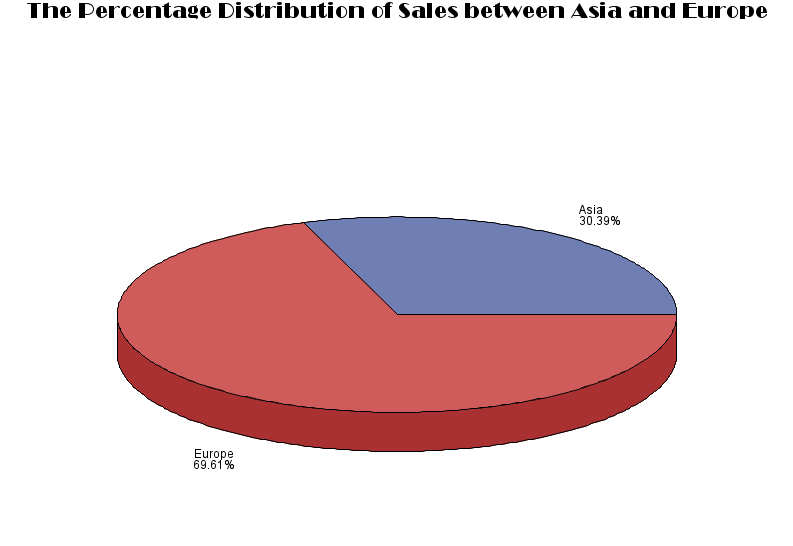


Figure 3.7 : Pie chart on the percentage distribution of sales between Asia and Europe

Interpretation :

The pie chart above describes the percentage distribution of sales made between Asia region and Europe region. By comparing these two regions, it can be indicated that Europe has a higher percentage of 69.61% on sales recorded compared to Asia which is only 30.39%.

### Horizontal Bar Chart

title h=**2** f=broadway 'The Total Number of Units Sold by Each Item Type';

**proc** **gchart** data=project.sales;

hbar3d item\_type / sumvar=units\_sold nostats;

**run**;

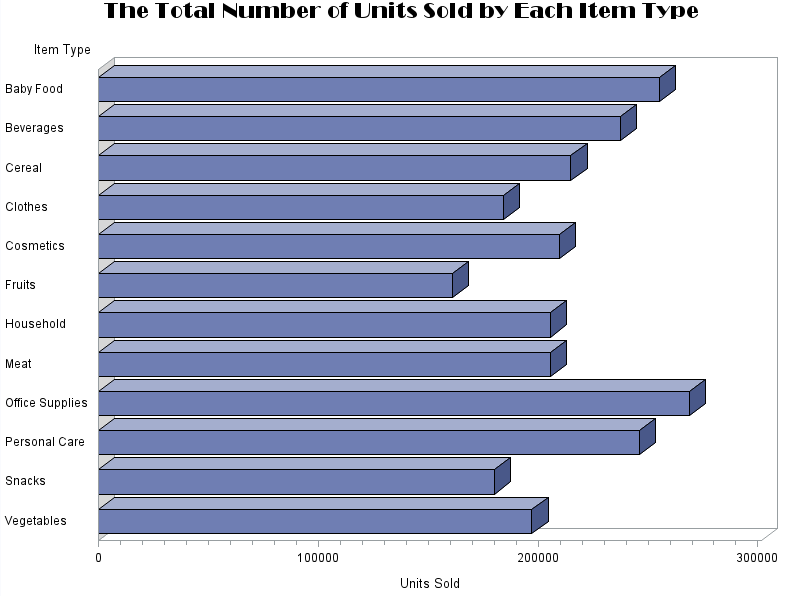


Figure 3.8 : Horizontal bar chart on the total number of units sold by each item type

Interpretation :

The above figure illustrates on the total number of units sold by each item type. The horizontal bar chart above indicates that the highest number of units sold by all regions is office supplies followed by baby food and personal care items. Meanwhile, the lowest item to be sold is fruits.

title h=**2** f=broadway 'The Total Profit Made by Each Region';

**proc** **gchart** data=project.sales;

hbar3d region / sumvar=total\_profit nostats;

format total\_profit dollar12.;

**run**;

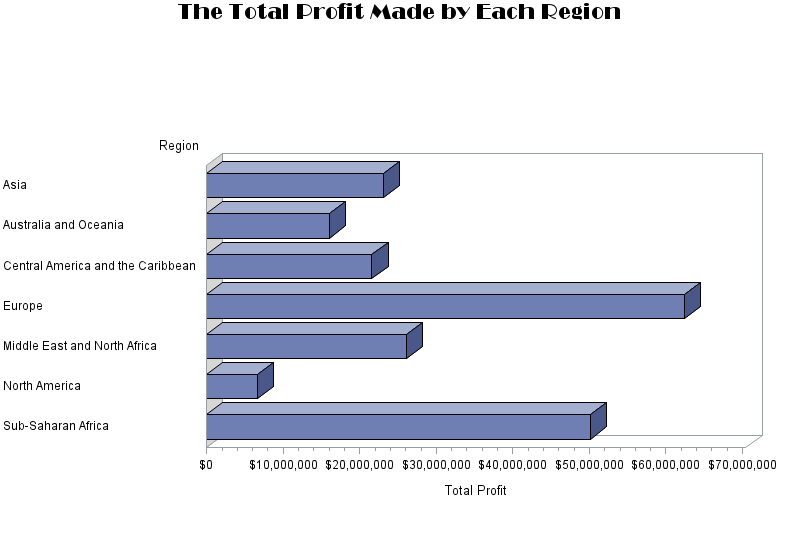


Figure 3.9 : Horizontal bar chart on the total profit made by each region

Intepretation :

The above diagram illustrates on the total profit made by each region. The horizontal bar chart above shows that Europe has the highest total profit made compared to other regions followed by Sub-Saharan African. Meanwhile, North America has the lowest total profit made.

### Vertical Bar Chart

title h=**2** f=broadway c=brown 'The Total Number of Sales Made by Each Region';

axis1 stagger label=none;

axis2 label=(a=**90** 'Frequency');

**proc** **gchart** data=project.sales;

vbar region / patternid=midpoint width=**10** maxis=axis1 raxis=axis2;

**run**;

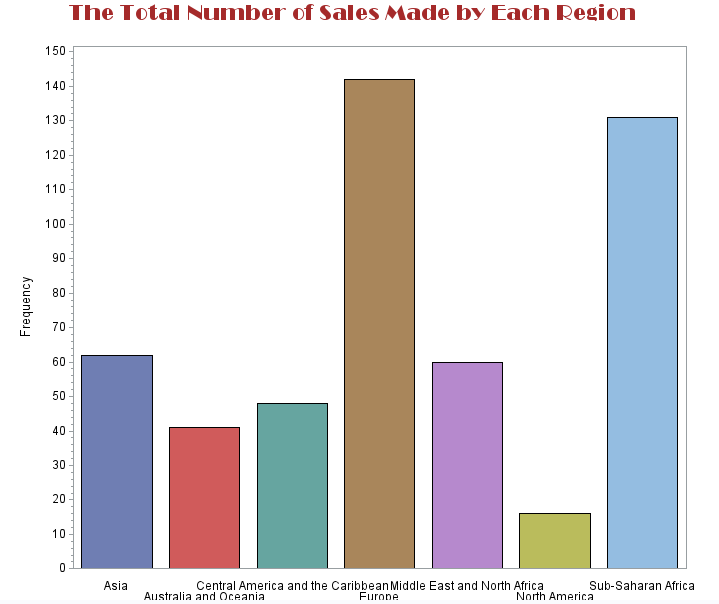


Figure 3.10 : Vertical bar chart on the total number of sales made by each region

Interpretation :

The above diagram illustrates on the total number of sales recorded for each particular region. From the above vertical bar chart, it can be concluded that Europe is the region with the highest number of sales frequencies followed by Sub-Saharan Africa region. In contrary, North America has the lowest number of sales occurred.

## Inferential Statistic

### Normality Test on Total Revenue for Asia Region and Europe Region

**proc** **univariate** data=project.AE normal;

var total\_revenue;

probplot / normal (mu=est sigma=est);

**run**;

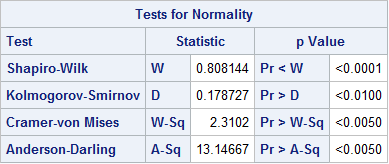


Figure 3.11 : Test for Normality on Total Revenue

Hypothesis Testing :

: The data is normally distributed.

: The data is not normally distributed.

α : 0.05

p-value : 0.01

Decision Rule : Reject since p-value = 0.01 < α = 0.05

Conclusion : The data is not normally distributed.

Interpretation :

Figure 3.12 shows that there is not enough evidence to conclude that the total revenue for Asia and Europe follows a normal distribution since the p-value of the Kolmogorov-Smirnov of 0.01 is less than the alpha value of 0.05. Therefore, the null hypothesis is rejected and hence, the data is not normally distributed. The above findings can also be proven by Figure C.1 in Appendix C which indicates that the probability plot of total revenue does not follow a normal distribution since the points does not lie approximately in a straight line.

### Mann-Whitney U Test to Compare the Differences Between Asia Region and Europe Region on Total Revenue

title 'Mann\_Whitney U Test';

**proc** **npar1way** data=project.AE wilcoxon;

class region;

var total\_revenue;

exact Wilcoxon;

**run**;

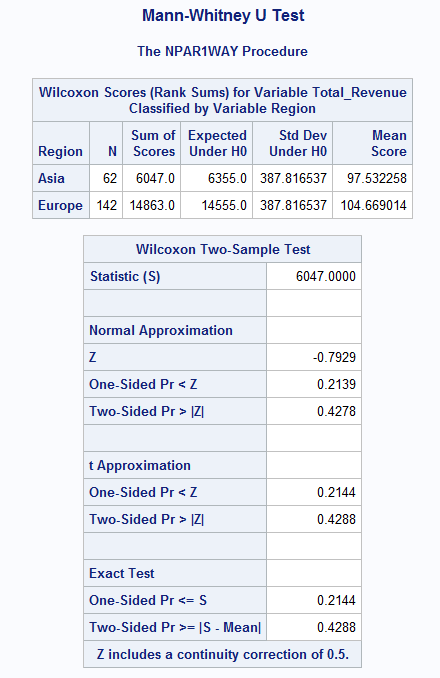


Figure 3.12 : Mann-Whitney U Test to compare mean differences in total revenue between Asia and Europe

Hypothesis Testing :

: =

: ≠

α : 0.05

p-value : 0.4288

Decision rule : Failed to reject since p-value = 0.4288 > α = 0.05

Conclusion : There is no significant mean difference on the total revenue between Asia region and Europe region.

Interpretation :

Figure 3.13 shows the output of a non-parametric Mann-Whitney U test on total revenue among Asia and Europe. Based on the figure, it can be concluded that there is not enough evidence to indicate that there is a significant difference in the mean of total revenue between Asia region and Europe region. This is because the p-value of 0.4288 is greater than 5% significant level and therefore, the null hypothesis failed to be rejected.

### Normality Test for Number of Items Sold

**proc** **univariate** data=project.items normal;

var units\_sold;

probplot / normal (mu=est sigma=est);

**run**;

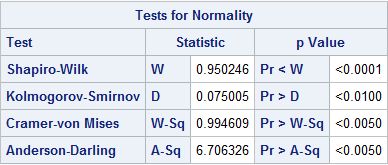


Figure 3.13: Normality Test for number of units sold

Hypothesis Testing :

: The data is normally distributed.

: The data is not normally distributed.

α : 0.05

p-value : 0.01

Decision Rule : Reject since p-value = 0.01 < α = 0.05

Conclusion : There is not enough evidence to conclude that the data follows a normal distribution.

Interpretation :

Figure 3.14 shows that there is not enough evidence to conclude that the total number of units sold for each item type follows a normal distribution since the p-value of the Kolmogorov-Smirnov of 0.01 is less than the alpha value of 0.05. Therefore, the null hypothesis is rejected and hence, the data is not normally distributed. The above findings can also be proven by Figure D.1 in Appendix D which indicates that the probability plot of number of units sold does not follow a normal distribution since the points does not lie approximately in a straight line.

### Kruskal-Wallis Test to Determine Whether There is a Significant Mean Difference in the Number of Items Sold Among Item Types

title 'Kruskal-Wallis Test';

**proc** **npar1way** data=project.items wilcoxon;

class item\_type;

var units\_sold;

**run**;

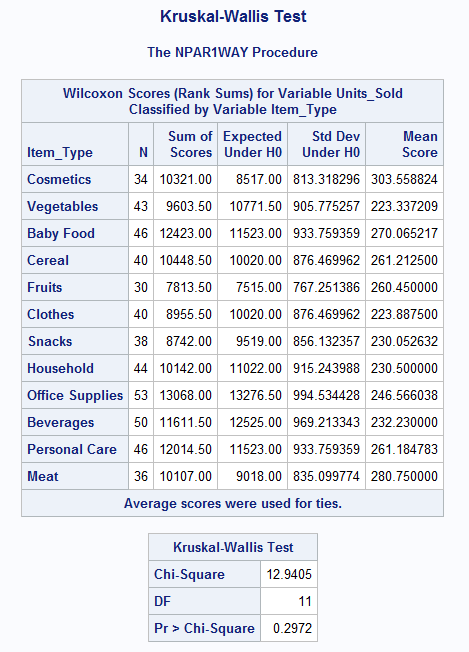


Figure 3.14 : Kruskal-Wallis Test on number of items sold for item types

Hypothesis Testing :

:

:

α : 0.05

p-value : 0.2972

Decision rule : Failed to reject since p-value = 0.2972 > α = 0.05

Conclusion : There is no significant mean difference on the number of items sold among the 12 item types.

Interpretation :

A non-parametric Kruskal-Wallis Test is used to compare the differences between the number of items sold and item types since the dependent variable does not follow a normal distribution. Figure 3.13 shows the output of Kruskal-Wallis Test on total number of units sold among 12 different item types. Based on the figure, it can be concluded that there is not enough evidence to indicate that there are statistically significant differences in the mean of number of units sold among the item types. This is because the p-value of the test which is 0.2972 is greater than 5% significant level and therefore, the null hypothesis is to be accepted and therefore the number of items sold are identical for every item type.

### Chi-Square Test to Determine the Association Between Regions and Three Total Profit Levels.

**proc** **freq** data = project.regions;

tables region\*total\_profit / expected chisq fisher;

**run**;

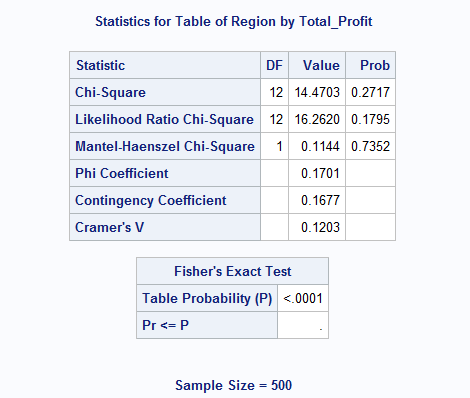


Figure 3.15 : Chi-Square Test on the association between regions and three total profit levels

Hypothesis Testing :

: There is no association between regions and total profit levels.

: There is an association between regions and total profit levels.

α : 0.05

p-value : 0.2717

Decision rule : Failed to reject since p-value = 0.2717 > α = 0.05

Conclusion : There is no association between regions and total profit levels.

Interpretation :

The above figure and hypothesis testing shows that there is not enough evidence to indicate that there is a relationship between regions and total profit levels since the p-value of the Chi-Square test is 0.2717 which is greater than the significant value of 0.05. Therefore, both regions and total profit levels are independent to each other.

# CONCLUSION

As a conclusion for this study regarding the sales recorded by several regions, descriptive statistics procedures such as summary statistic, bar charts and pie charts have been used to produce a better interactive and understanding results. Furthermore, the researchers have successfully applied inferential statistics such as normality test to determine whether the data follows a normal distribution and advanced statistical analysis have also been deployed to answer the objectives of this study.

It was found that the total revenue for both Asia and Europe region does not follow a normal distribution and therefore, a non-parametric Man-Whitney U Test was used. The result indicates that the total revenue is identical for both Asia and Europe region. It can be said that both regions are very competitive in obtaining revenue from the items sold.

A non-parametric was also used to determine whether there is a significant difference in the number of items sold among different item types since there is no sufficient evidence to indicate that the data is normal. As a result, the mean number of items sold by 12 different item types are the same. This shows that all of the items sold recorded in this study are important in determining the profit and revenue obtained from all the regions.

Lastly, a Chi-Square Test was used to observe the association between all regions and three total profit levels. The profit levels are categorised into low, moderate and high. It can be seen that there is no obvious relationship between regions and the total profit obtained and hence, regions and total profit are independent to each other. It may come to a conclusion that regions are not the main factor for the total profit obtained in this study since they did not show any relationship.

As a recommendation for this study, future researchers should consider on using a primary data instead of a secondary data since it will be more accurate for analysis and decision making. Not only that, more observations should be used to fulfil the normality assumption of an analysis so that a parametric test can be conducted.

# REFERENCE

Charlot Bennett, M. E. (2016). *SAS Programming 1 : Essentials Course Notes.* North Carolina: SAS Institute Inc.

Charlot Bennett, M. E. (2016). *SAS Programming 2: Data Manipulation Techniques.* North Carolina: SAS Institute Inc.

*E for Excel*. (n.d.). Retrieved from Sample Data Sets for Testing: http://eforexcel.com/wp/downloads-18-sample-csv-files-data-sets-for-testing-sales/

Law, V. (2004). SAS Programming; The One Day Course. *Journal of Statistical Software*.

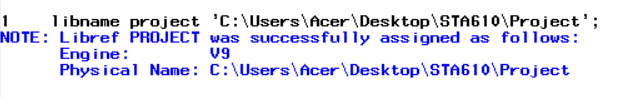
Twin, A. (17 February, 2020). *Sale*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/sale.asp

# APPENDICES

## Appendix A : Data Management

SAS Data Library

libname project 'C:\Users\Acer\Desktop\STA610\Project';



*Appendix 1*

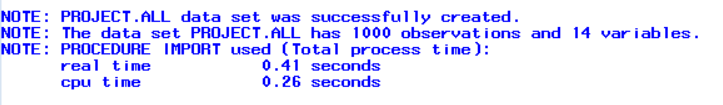
Proc Import

**proc** **import** out = project.all

datafile = "C:\Users\Acer\Desktop\STA610\Project\Sales Records.xlsx"

dbms = excel;

**run**;



*Appendix 2*

## Appendix B : Descriptor Portion

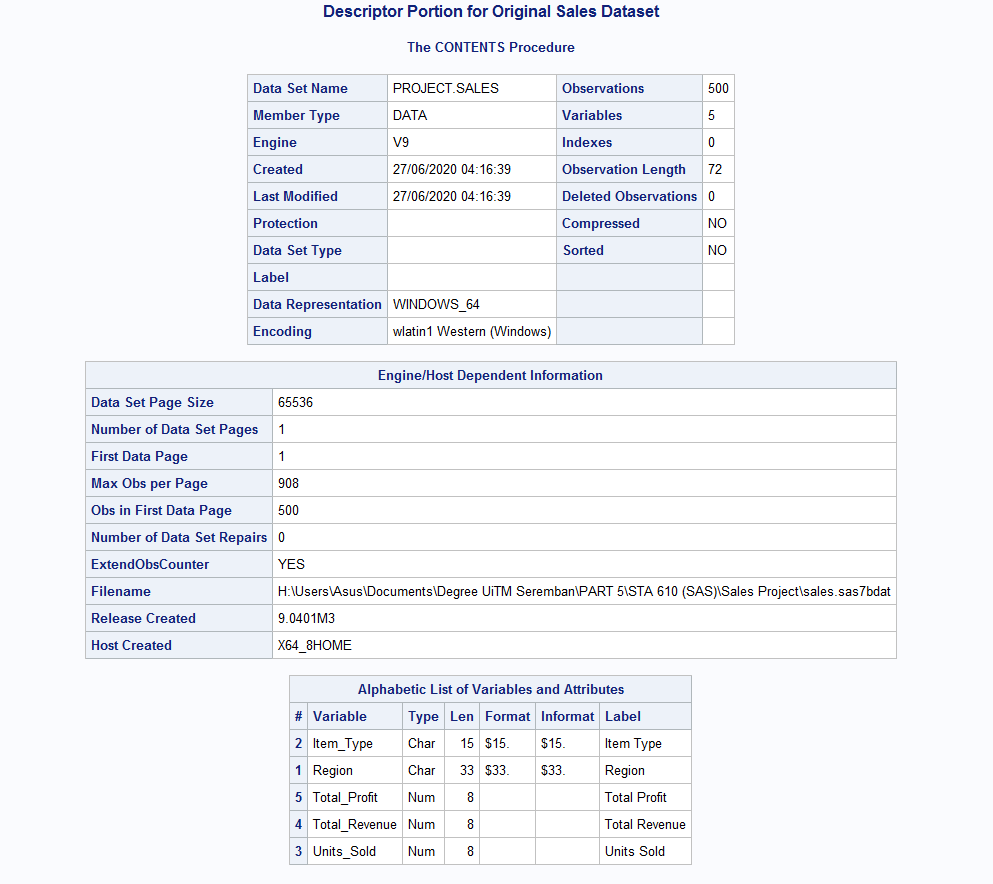


Figure A.1 : Descriptor Portion for Original Sales Dataset

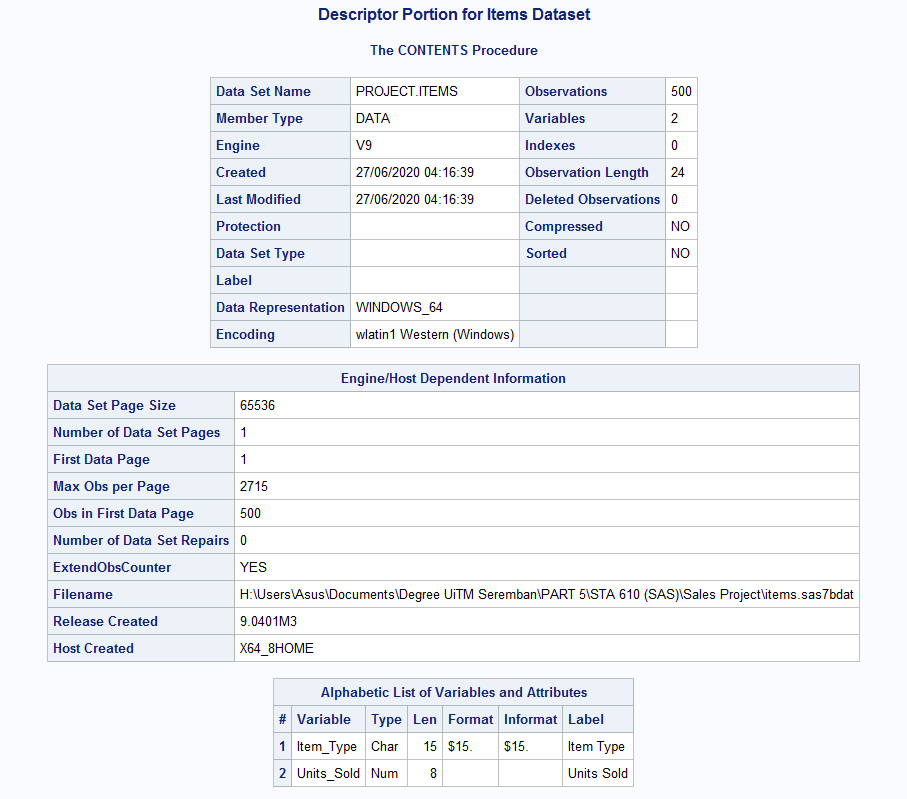


Figure A.2 : Descriptor Portion for Items Dataset

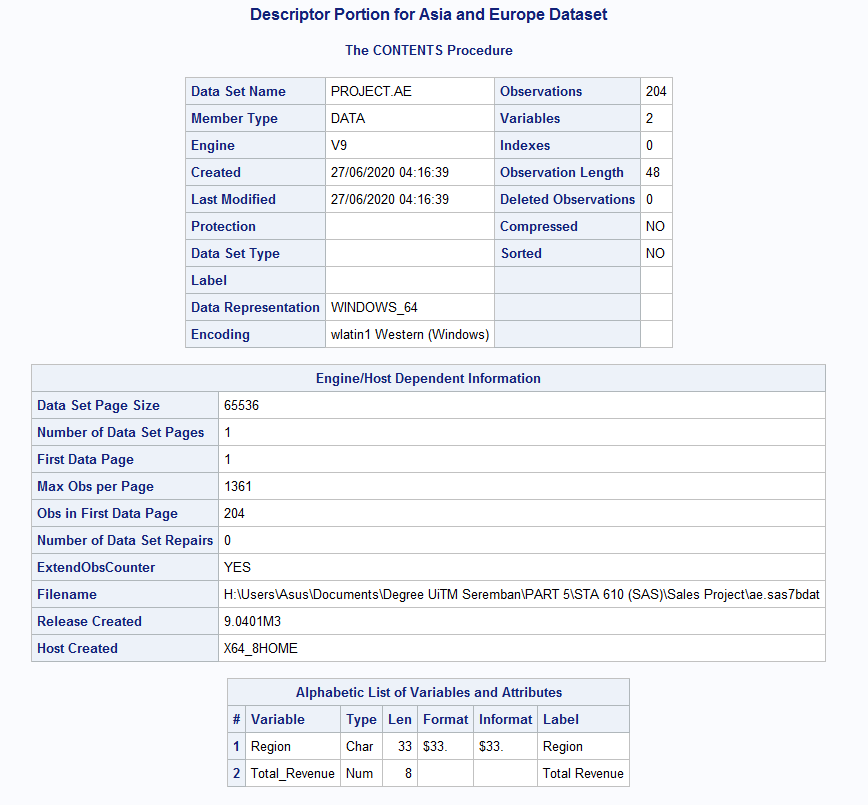


Figure A.3 : Descriptor Portion for Asia and Europe Dataset

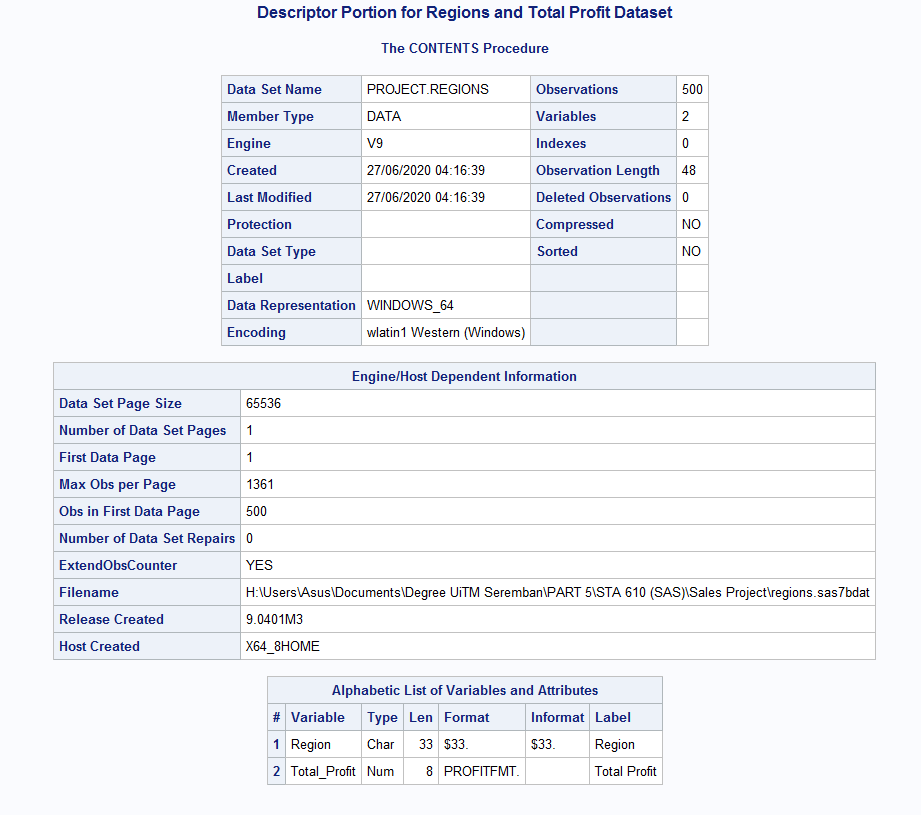


Figure A.4 : Descriptor Portion for Regions and Total Profit Dataset

## Appendix C: Printing Partial Dataset

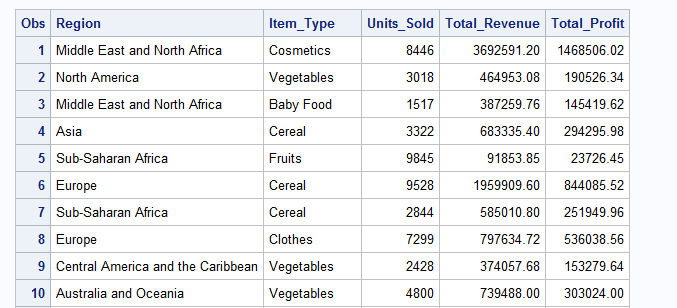


Figure B.1 : Partial Data for Original Sales Dataset

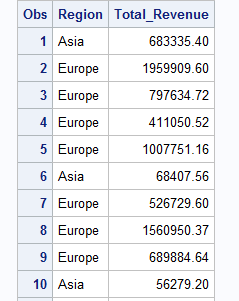


Figure B.2 : Partial Data for Asia and Europe Dataset

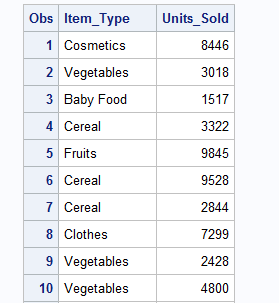


Figure B.3 : Partial Data for Items Dataset

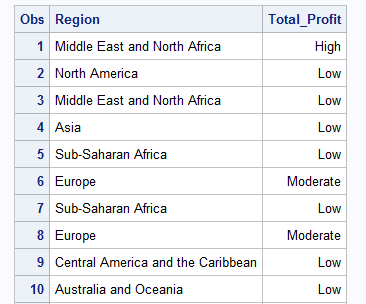


Figure B.4 : Partial Data for Regions and Total Profit Dataset

## Appendix D : Normality Test for Total Revenue between Asia and Europe

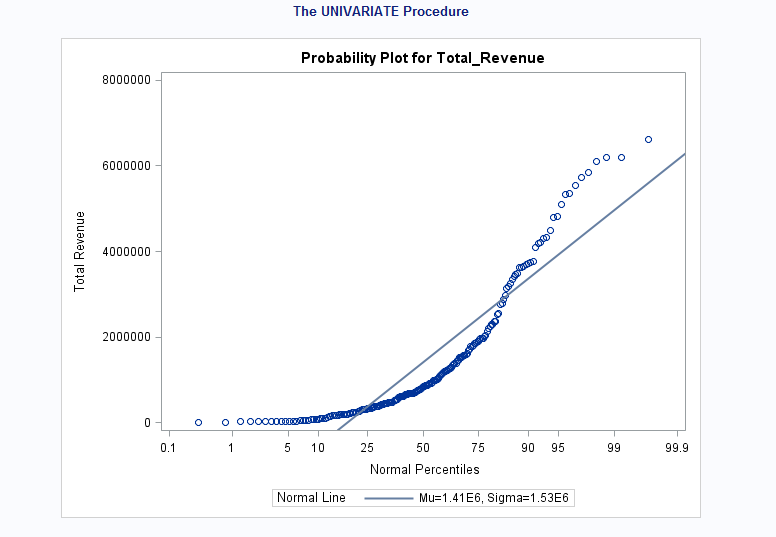


Figure C.1 : Normal P-P Plot for Total Revenue between Asia and Europe Region

## Appendix E : Normality Test for Number of Items Sold

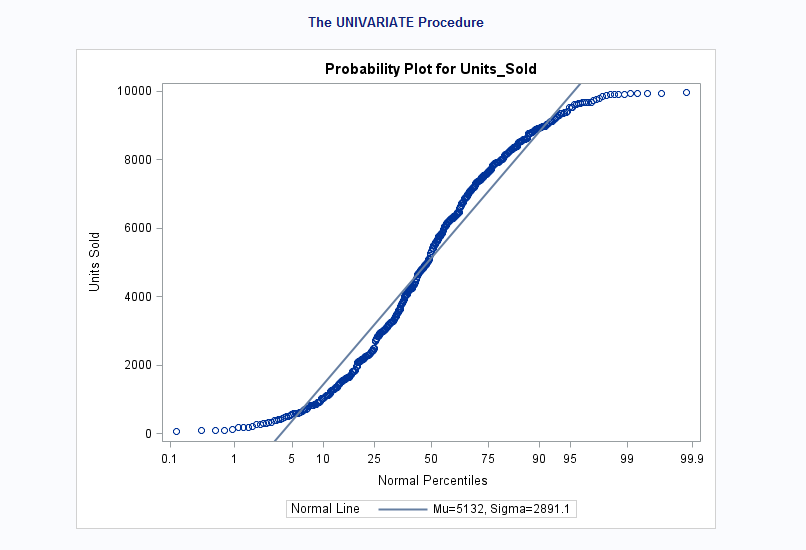


Figure D.1 : Normal P-P Plot for Number of Items Sold

## Appendix F : Box Plot for Total Revenue between Asia and Europe

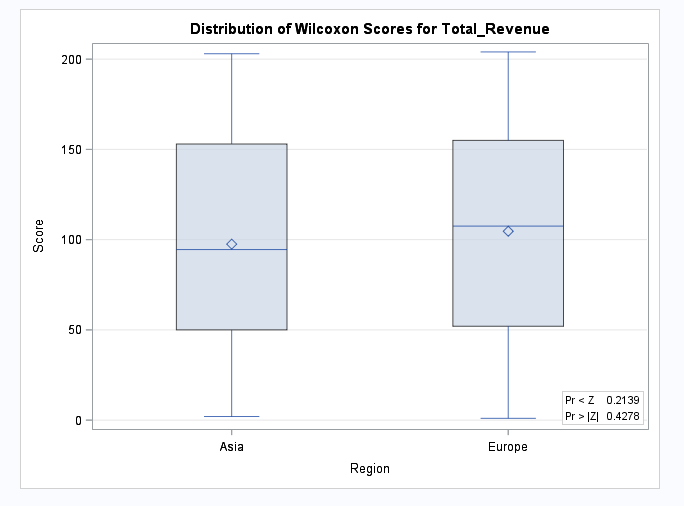


Figure E.1 : Box Plot of Wilcoxon Scores for Total Revenue Among Asia and Europe

## Appendix G : Box Plot for Number of Items Sold among Item Types

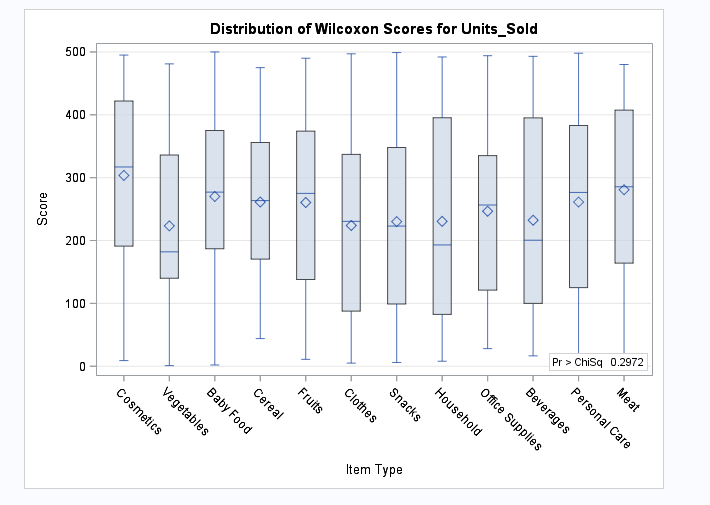


Figure F.1 : Box Plot of Wilcoxon Scores for Number of Units Sold Among Item Types