

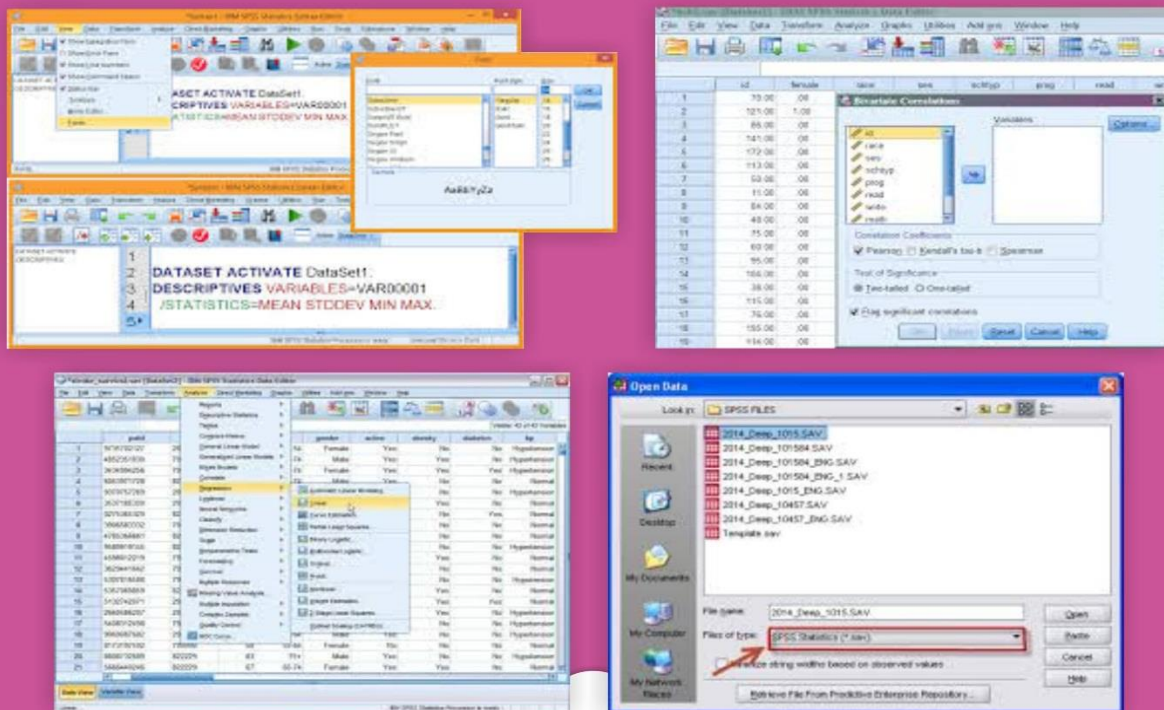
Monograph On

SPSS

BASICS - Handbook

2019

by **DR. SHINEY CHIB**
shinychib@gmail.com



Publishing Partner : **IJSRP Inc.**

PREFACE

Research is very vital part in Academics. Academician publish their research work in form of research papers, articles, case studies, Monograph etc. Data analysis has a major role to play in research and researchers use various software and packages to do Data Analysis. Statistical Package for Social Sciences is one such package which researchers are using for data analysis. This Monograph demonstrates how one can use the statistical techniques to analyse the data using SPSS. This Monograph is an effort to present the concepts and techniques in a very simple manner.

Shiney Chib, Ph.D.

ABOUT THE AUTHOR:



Dr.Shiney Chib is working as Director & Research Head with Datta Meghe Institute of Management Studies, Nagpur. She is appointed as adjunct faculty with SEGI-University, Malaysia. She has been conferred with '**Distinguished Educator Award**' by 'Discovery Education Media', '**Rajiv Gandhi Excellence Award in the field of Education**' and '**Mahila Ratan Gold Medal for Academic Excellence**' and '**Eminent Educationist Award**'. She has also received '**Academic Excellence Award**' from NIPM Nagpur chapter. Eight students have been awarded PhD under her guidance. She has industrial experience of 11 years and 15 years of teaching experience. She has more than 100 publications, in the reputed journals, National and International conferences. She has guided 12 Ph.Ds and authored 4 books. She has participated and chaired the conferences held at Hamburg, Germany, Tokyo Japan and Hong Kong. She is a life member of NIPM, NHRD, All India Commerce Association, Lion's Club International.

CONTENTS

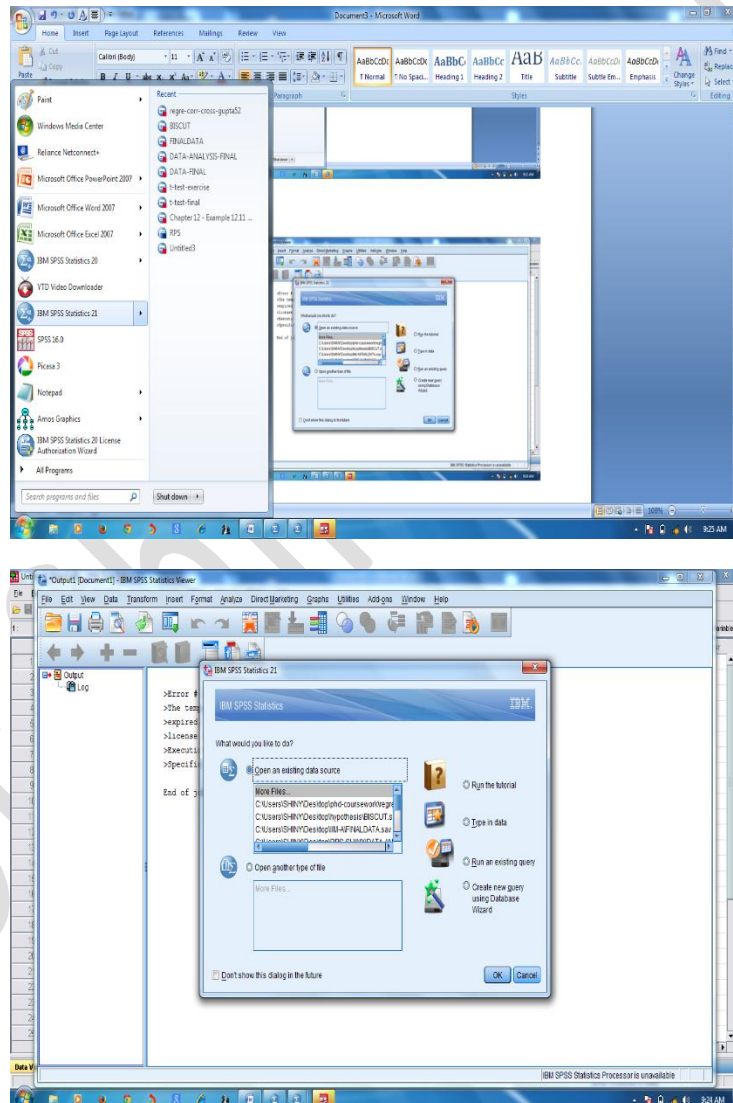
1. Chapter 1: Introduction to SPSS
2. Chapter 2: Descriptive Statistics
3. Chapter 3: Transforming Data
4. Chapter 4: T –TEST
5. Chapter 5: Correlation And Regression Analysis
6. Chapter 6: ANOVA (F –Test)
7. Chapter 7: Factor Analysis

Chapter 1

Introduction to SPSS

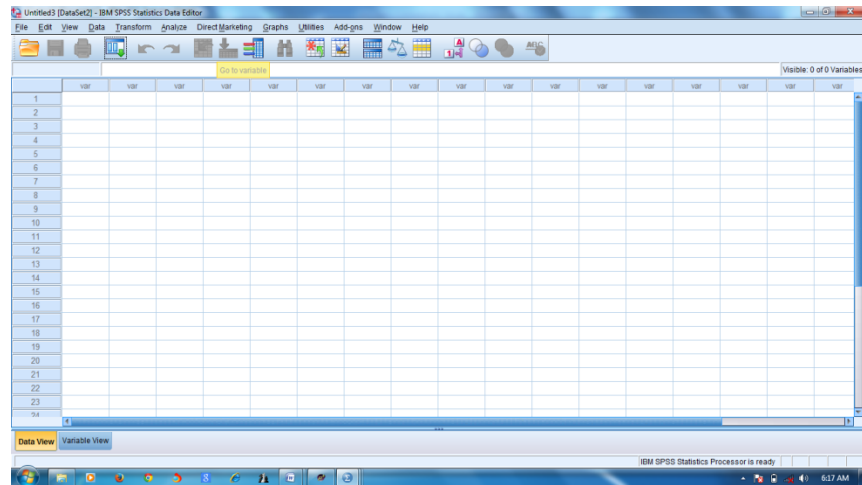
SPSS (Statistical Package for Social Science) is a powerful and easy-to-use software program for data manipulation and data analyses. The advantages of SPSS is that it allows the researcher to manipulate and analyze data via syntax commands or through a visual data editor.

Click **START ->ALL PROGRAMS->SPSS Inc.->SPSS 20.0**



For New File creation:

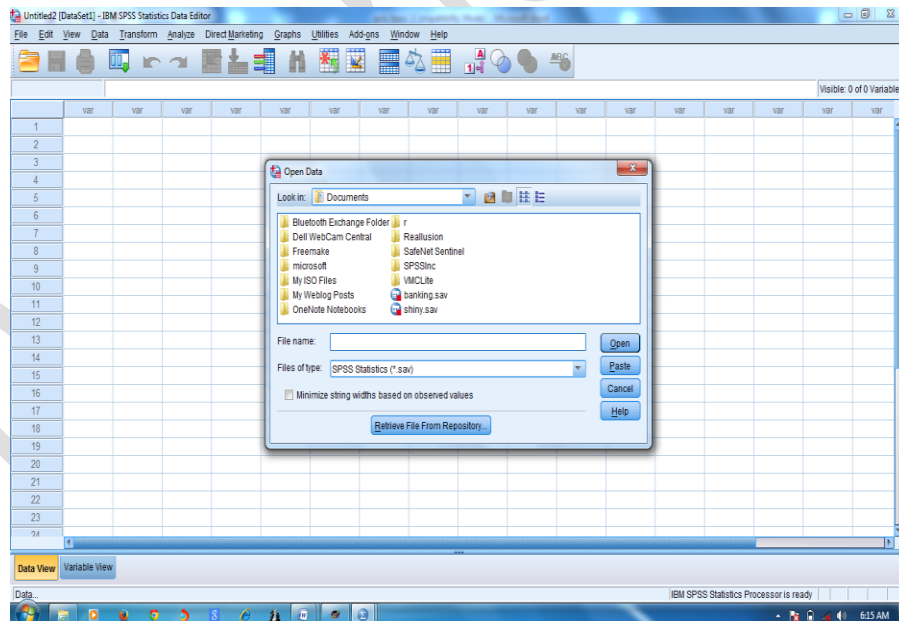
Click FILE →NEW→DATA



For Existing File:

Click FILE →OPEN→DATA

You will get the results as shown below:

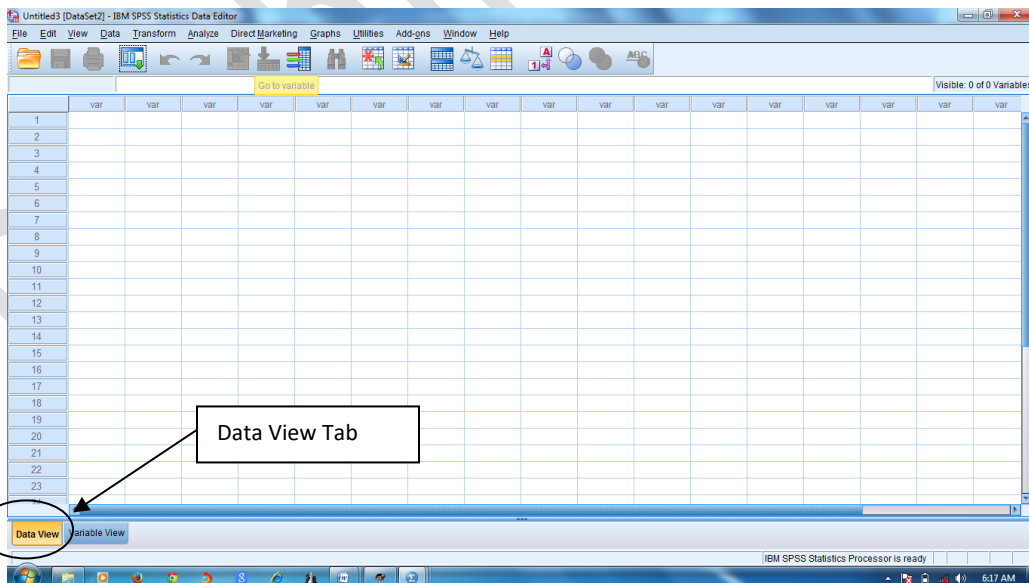
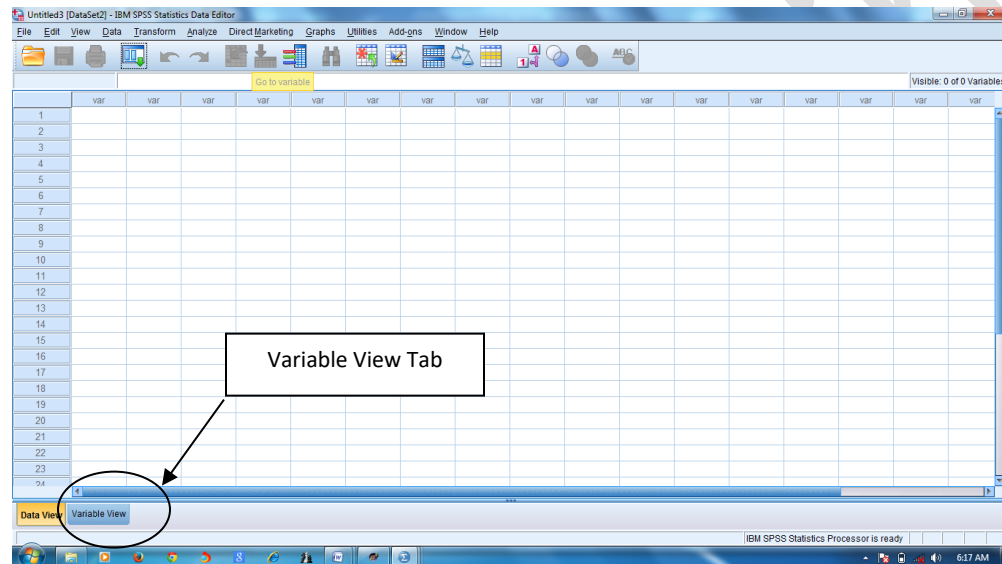


GETTING FAMILIAR WITH SPSS

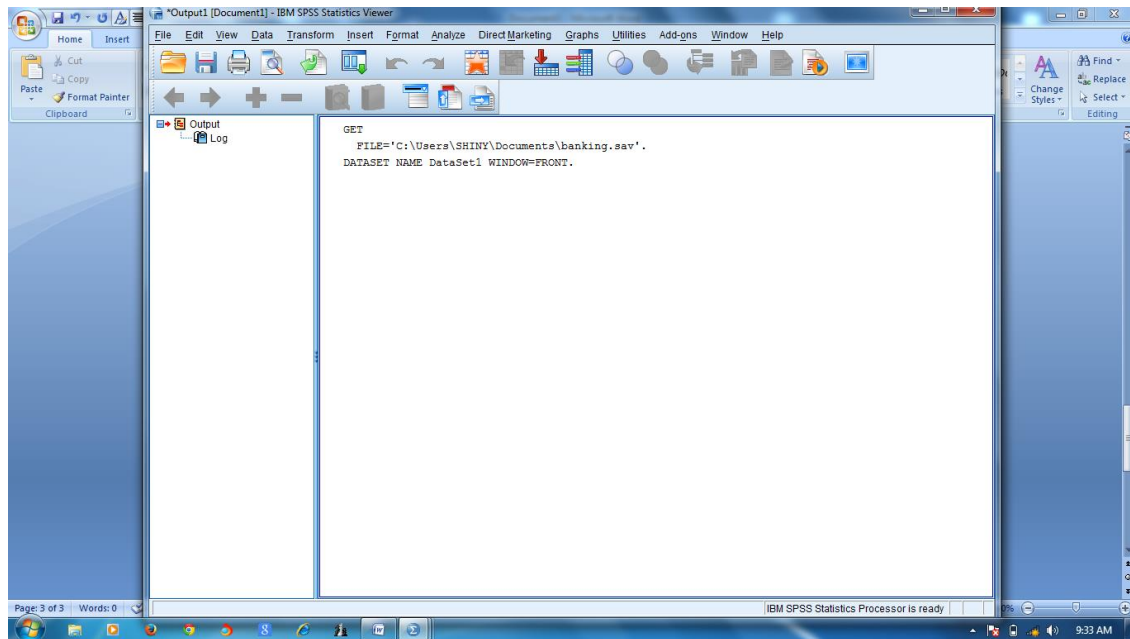
➤ DATA AND VARIABLE EDITOR:

For New File creation:

Click FILE → NEW → DATA



➤ OUT PUT WINDOW



The screenshot shows the IBM SPSS Statistics Viewer window. The Output window displays the following command:

```
GET
FILE='C:\Users\Shiny\Desktop\E-BANKING_1.sav'.
FREQUENCIES VARIABLES=GENDER AGE EDUCATION INCOME PROFESSION PROF_TYPE BANK
ACCOUNT_TYPE CARD E_BANKING YEARS TYPE CHANNEL FREQUENCY
INFORMATION SECURITY THREAT CONTINUE REFER TRANSACTION CHARGES SERVICES
SATISFACTION SECURITY_MEASURES EXPERIENCE
/STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE
/ORDER=ANALYSIS.
```

Frequencies

[DataSet1] C:\Users\Shiny\Desktop\E-BANKING_1.sav

		Statistics												
		GENDER	AGE	EDUCATION	INCOME	PROFESSION	PROF_TYPE	BANK	ACCOUNT_TYPE	CARD	E_BANKING	YEARS	TYPE	CHAN
N	Valid	517	500	517	500	517	500	517	517	517	517	500	517	
	Missing	0	17	0	17	0	17	0	0	0	0	17	0	
Mean			2.3800		2.1000		4.5580					1.8320		
Median			2.0000		2.0000		6.0000					1.0000		
Mode			3.00		1.00		6.00					1.00		
Std. Deviation			1.09086		1.13692		1.83559					1.05736		
Variance			1.190		1.293		3.369					1.118		
Minimum			1.00		1.00		1.00					1.00		
Maximum			4.00		4.00		6.00					5.00		

Frequency Table

GENDER

➤ TOOLBAR

The TOOLBAR below Menu bar provides a quick and easy access to the different features.

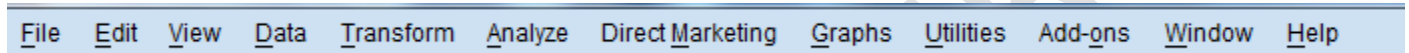


- **Open data document**-→to open particular data file
- **Save the document**-→to save particular data file
- **Print**-→to print a particular data file
- **Recall recently used dialogues**-→shows a list of recently opened dialogue boxes, for easy and handy access. **Undo a user action**-→to undo user action
- **Redo a user action**-→to redo user action
- **Goto case**-→to go to a particular variable by directly typing the name of variable or selecting it from a drop down list.
- **Goto variable**-→to go to a particular variable by directly typing the name of variable or selecting it from a drop down list.
- **Variables**-→provide data definition for all the variable in the current data file.
- **Find** -→to search particular data in the existing data editor.
- **Insert cases**-→to insert a case above the active cell case.
- **Insert variable**-→to insert a variable to the left of active cell variable.
- **Split File**-→to split a file into separate groups.
- **Weigh Cases**-→to give different weighs to cases for statistical Analysis.
- **Select Cases**-→to select varied criteria to be used in case selection. It provides several methods for selecting groups/sub groups of cases based on some criterion.
- **Value Label**-→It allows toggling between value labels and actual values in the Data Editor.
- **Use Variable Sets**-→used to select particular sets of variables in the current data file.
- **Show all variable**-→used to see all variables in the current data file.

- **Spell Check**-→used to check spelling in the current data file.

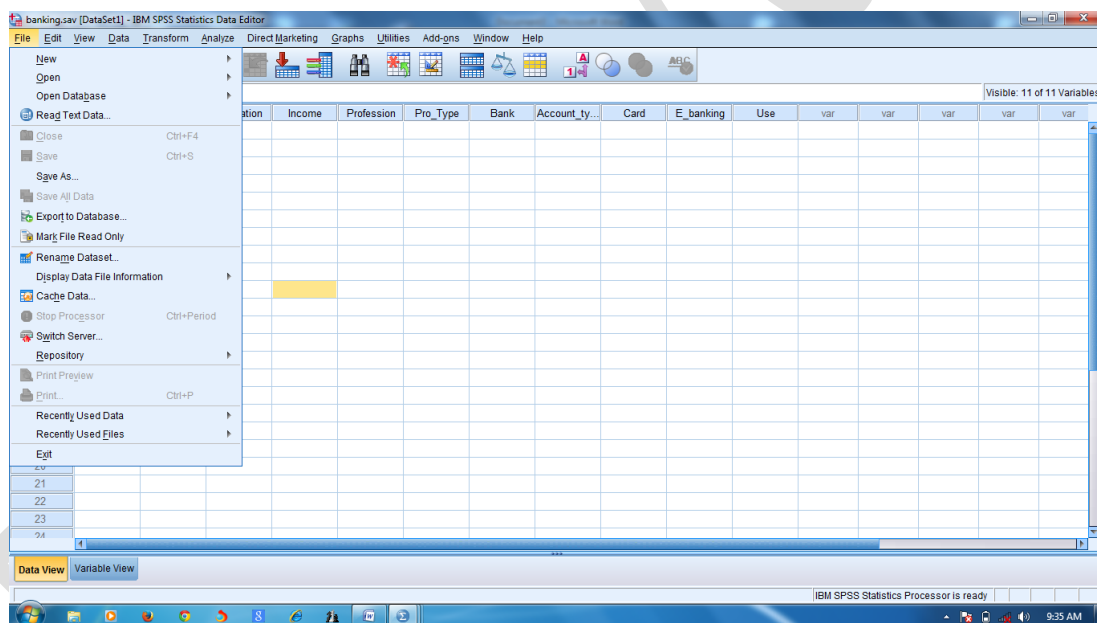
➤ MENU BAR

The features of main MENU bar are as follows:



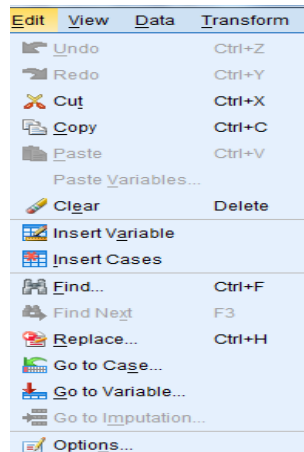
FILE MENU

The FILE MENU allows creating new files, opening existing files, read text data, save files, print and other basic starting or output activities. You can exit SPSS through this menu. This menu resembles the FILE MENU features of Ms-word.



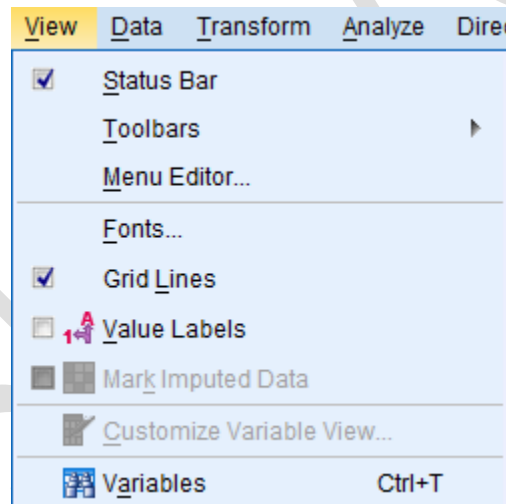
EDIT MENU

The EDIT MENU allows to cut,copy, paste from the output or windows clipboard. It also allows to redo/undo any action. You can also inert/goto variable or case through this Menu.



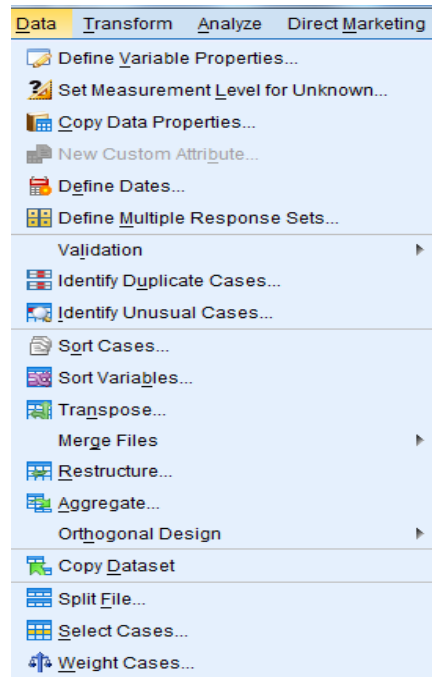
VIEW MENU

The VIEW menu allows to activate/deactivate the status bar and other toolbars. You can change fonts, grid lines, value label.



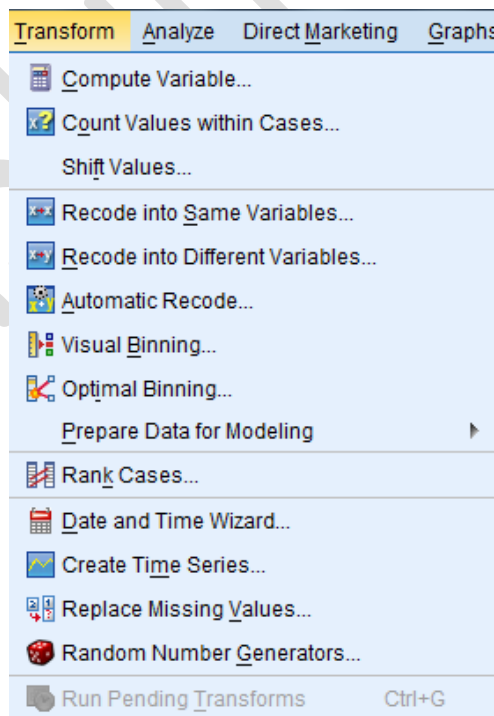
DATA MENU

The DATA Menu allows you to define variable properties, dates, multiple response sets. It also allows in identifying duplicate cases, sorting cases/variables, transposing, restructuring and aggregating data along with other functions. You can also merge files, split a file, select and weigh cases through this menu.



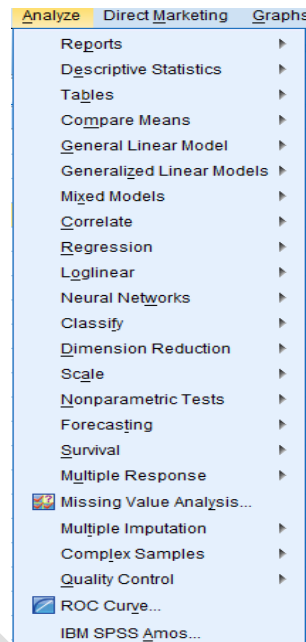
TRANSFORM MENU

TRANSFORM menu allows creating new variables using compute variable and helps in changing variables in the data file through recode, rank case and many others.



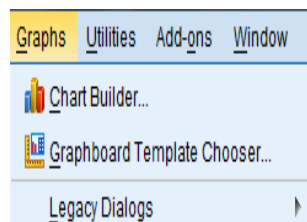
ANALYZE MENU

ANALYZE menu allows the analysis of data label based on various statistical tools and techniques.



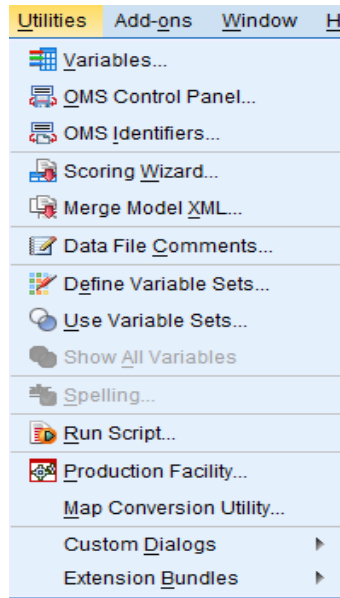
GRAPHS MENU

The GRAPH menu allows you to create bar chart, line chart area chart, pie chart, histogram, scatter plots along with many other variations.



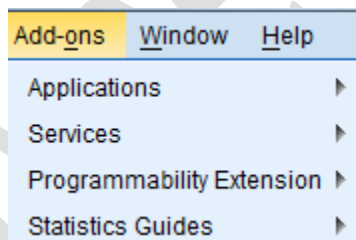
UTILITIES MENU

The UTILITIES menu helps in using various other utilities, which may or may not be required in normal usage. This menu shows data file comments, allows define or use different variable sets and spell check.



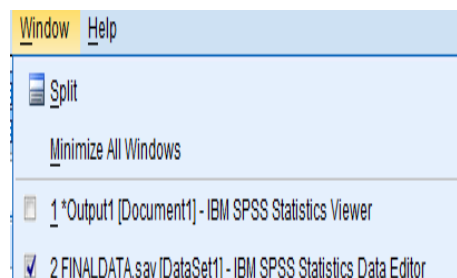
ADD-ONS MENU

The ADD-ONS menu provides a list of various add-on features requisite for special requirements and are mostly meant for advanced level users.



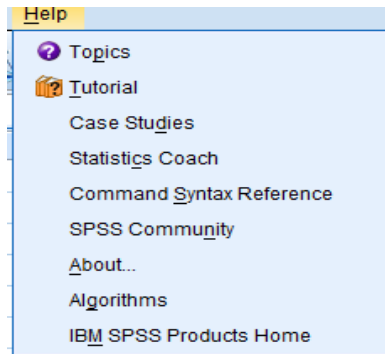
WINDOWS MENU

The WINDOWS Menu allows you to arrange, select and control the attributes of various windows. Through this menu, you can easily split a window, minimize all windows or switch over between the various open windows.



HELP MENU

The HELP Menu allows you to access the information related to usage of various features of SPSS.



➤ CREATING A DATA FILE

EXAMPLE DEMONSTRATION.

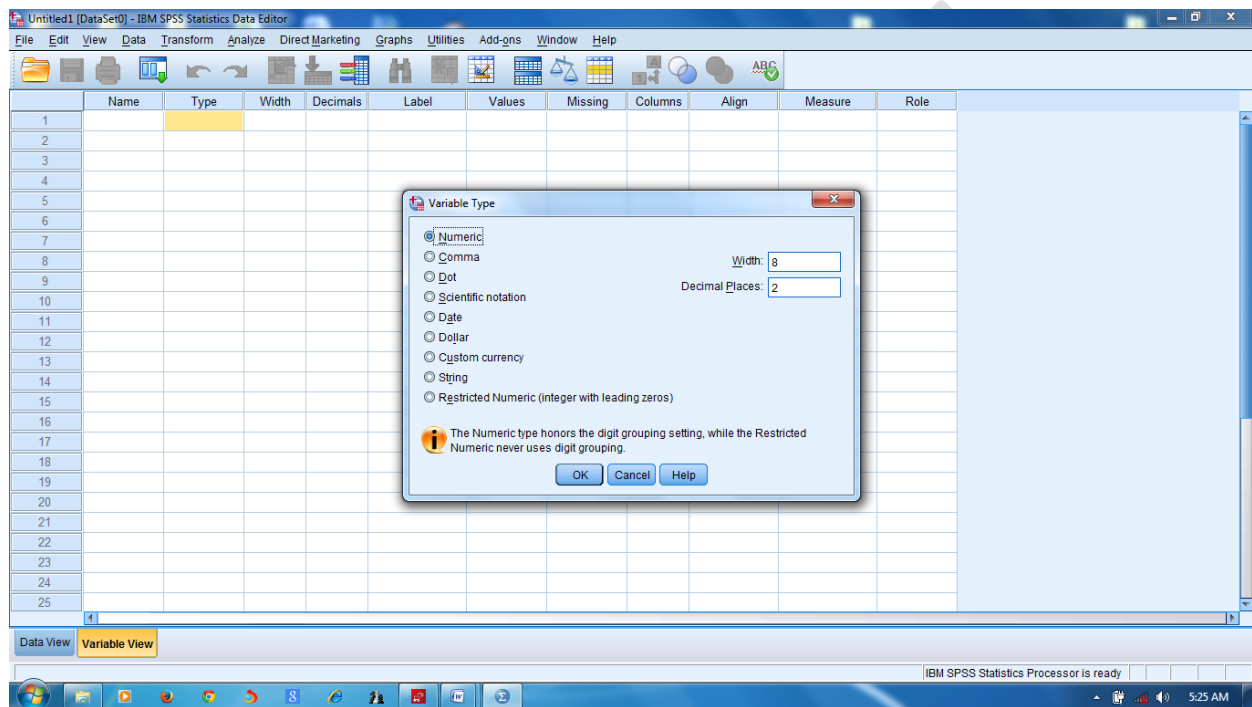
Lets discuss the data file creation. A close ended questionnaire is defined for knowing the consumer preference with respect to mobile phones. The questions were: Age , Gender, Qualification, Occupation, Monthly Income, Residence, Mobile Brand Preference.

DEFINING VARAIBLES:

In order to define variables based on questionnaire, following points need to be taken into consideration:

- *Must begin with a letter.*
- *Cannot end with full stop or underscore.*
- *Must be unique; duplication not allowed.*
- *Blanks and special characters cannot be used.*
- *Length of name cannot exceed 64 bytes.*
- *Reserve key words (like AND,ALL,NOT,OR TO, WITH) cannot be used as variable names.*
- *Names are not case sensitive. They can be written in upper or lower case.*
- *Long variable names need to wrap to multiple lines in the output, so SPSS attempts to break lines at underscore, periods and a change from lower to upper case.*

You can select variable type according to your requirement. By default numeric variable type with eight width and 2 decimal is chosen by SPSS. You can also change the width and decimal places.



1. AGE : 1)Below 30 2)30-35 3)35-40 4)40-45
5)45-50 6)Above 50
2. GENDER : 1) Male 2)Female
3. QUALIFICATION:1)Under graduate 2)Graduate 3)Post Graduate 4)MPhil
5)PhD.
4. OCCUPATION : Student() Business () Service () Others ()
5. MONTHLY INCOME : Less than 20000() 20000-30000() 30000-40000()
Above 50000()
6. RESIDENCE : Metro () Non-Metro() Rural ()

7. MOBILE BRAND PREFERENCE : IPHONE () SAMSUNG () BLACKBERRY ()
SONY () OTHERS()

CODE BOOK

SR.NO.	VARIABLE NAME	VALUE/CODE
1	AGE	1=<30 2=30-35 3=35-40 4=40-45 5=45-50 6=>50
2	GENDER	1=MALE 2=FEMALE
3	QUALIFICATION	1=UNDER GRADUATE 2=GRADUATE 3=POST GRADUATE 4=MPHIL 5=PHD
4	OCCUPATION	1=STUDENT 2=BUSINESS 3=SERVICE 4=OTHER
5	MONTHLY INCOME	1=<20000 2=20000-30000 3=30000-40000 4=40000-50000

		5=>50000
6	RESIDENCE	1=METRO 2=NON METRO 3=RURAL
7	MOBILE BRAND PREFERENCE	1=IPHONE 2=SAMSUNG 3=BLACKBERRY 4=KARBON 5=SONY 6=OTHERS

SUMMARY :

- *For New File creation: Click FILE →NEW →DATA*
- *In case of defining variables you must take care of naming, labels, variable type, column format and measurement level.*
- *Once entries are made, you can save the data file. Just click FILE Menu- →Save and click will open the save data as dialogue box. Enter the name of file, choose the location to save the file and press **save**.*
- *For Existing File:Click FILE →OPEN →DATA*

EXERCISE

Create a data file with the following variable and their information.

Sr.No.	Variable Name	Information
1.	Gender	Male Female
2.	Age	Below 20 20-30

		30-40 40-50 Above 50
3.	Education	Under graduate Graduate Post graduate Doctorate
4.	occupation	Student Business Service others
5.	Monthly salary	Less than 30000 30000-40000 40000-50000 Above 50000
6.	Usage of Plastic Bags should be banned.	Strongly Disagree Disagree Neutral Agree Strongly agree

Chapter 2

Descriptive Statistics

Descriptive statistics are used to have a basic view of the data, i.e. to describe the data, like how many respondents were there, how many male, how many female, the average age range, what is the median salary etc.

Frequency Distributions:

There are display of the frequency of occurrence of each value. The frequency distribution can be represented in tabular or graphical form

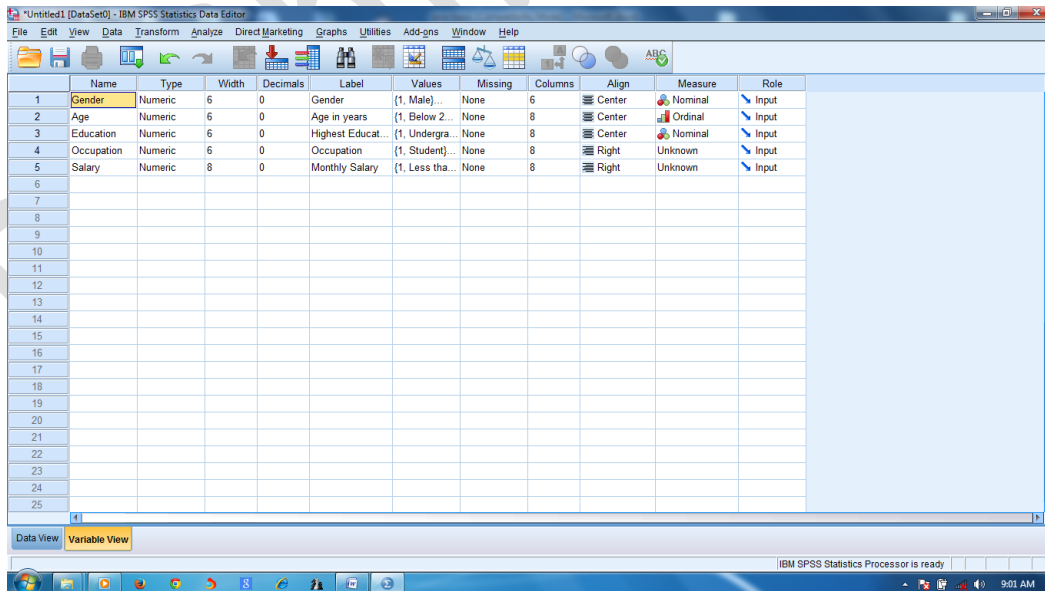
- For continuous variable with ratio/interval scales, histogram or frequency polygons may be used.
- For categorical variables with nominal/ordinal scales, bar charts can be used.
- Measures of Central Tendency
- The measures are mainly useful with interval or ratio scales.
- There are three main measures of central tendency, that is mean, median and mode.
- The measures of variability include range, interquartile range, standard deviation and variance.

WORKING EXAMPLE:

Lets discuss descriptive statistics with a working example.

Sr.No.	Variable Name	Information
1.	Gender	Male Female
2.	Age	Below 20 20-30 30-40 40-50

		Above 50
3.	Education	Under graduate Graduate Post graduate Doctorate
4.	occupation	Student Business Service others
5.	Monthly salary	Less than 30000 30000-40000 40000-50000 Above 50000



Form the data file and enter the data as shown below

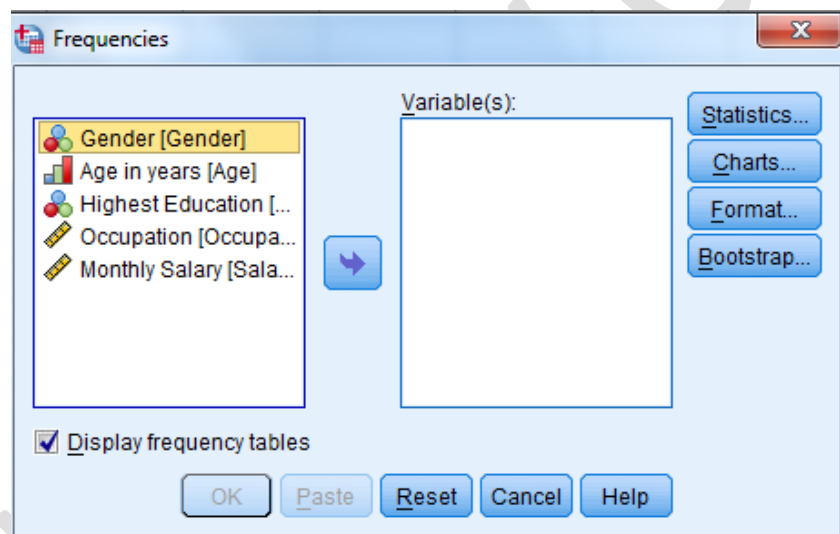
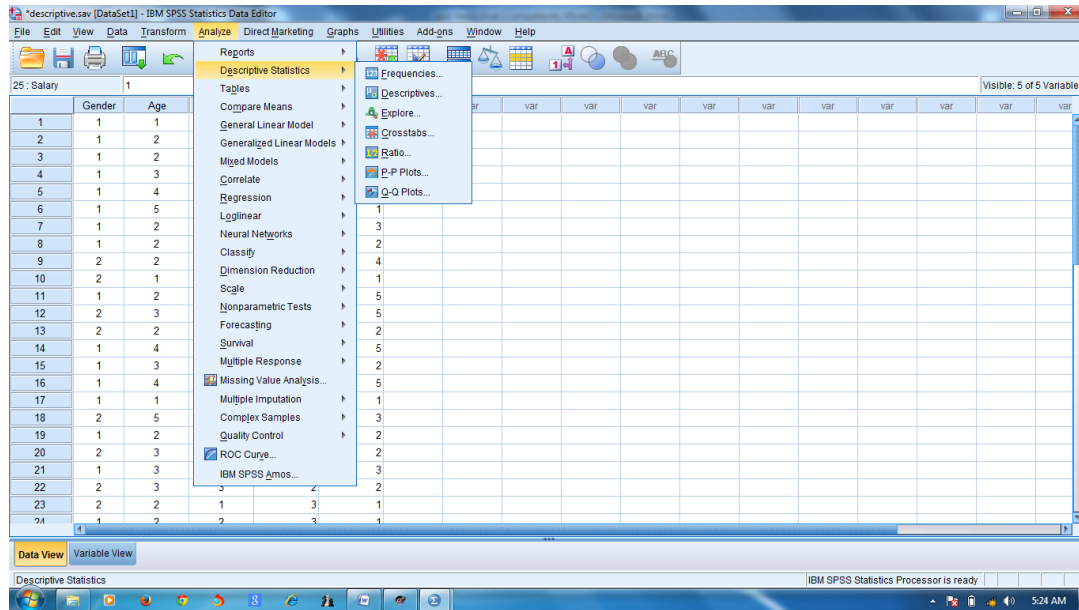
	Gender	Age	Education	Occupation	Salary	var	var	var	var	var	var	var	var	var	var
1	1	1	1	1	2										
2	1	2	2	1	2										
3	1	2	2	2	1										
4	1	3	2	2	2										
5	1	4	3	2	2										
6	1	5	4	2	1										
7	1	2	2	3	3										
8	1	2	2	3	2										
9	2	2	2	2	4										
10	2	1	1	1	1										
11	1	2	3	2	5										
12	2	3	3	2	5										
13	2	2	1	2	2										
14	1	4	3	2	5										
15	1	3	2	3	2										
16	1	4	3	2	5										
17	1	1	1	1	1										
18	2	5	4	2	3										
19	1	2	3	2	2										
20	2	3	3	2	2										
21	1	3	3	2	3										
22	2	3	3	2	2										
23	2	2	1	3	1										
24	1	2	2	3	1										

We are ready to analyze the data and see the descriptive statistics. The frequency distribution display the frequency of occurrence of each score value. This can be represented in tabular and graphical form. In case of continuous variables (measured on ratio or interval scale) histogram or frequency polygon are appropriate. In case of categorical variables (measured on nominal or ordinal scale) bar charts are suitable. Main measures of central tendency are mean, median and mode. The measures of variability include range, interquartile range, standard deviation and variance. These measures of variability are more suitable for interval or ratio data.

In order to the analysis:

Click Analyze → Descriptive Statistics → Frequencies

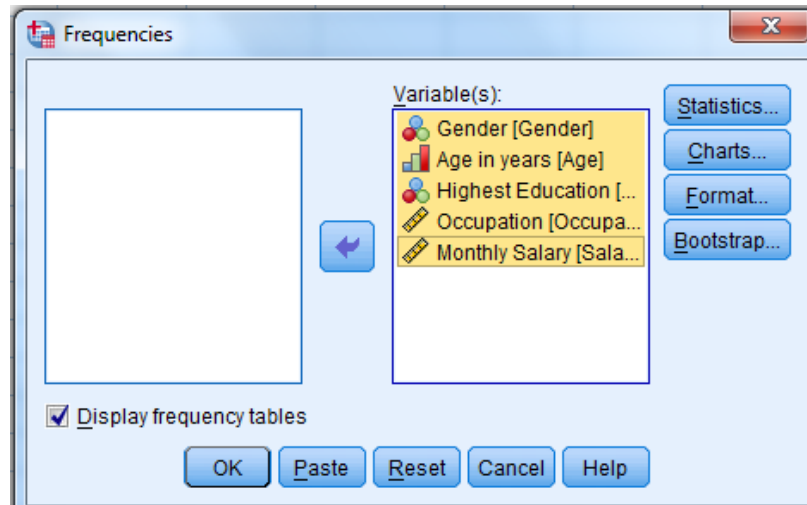
This will open **Frequencies** dialogue box.



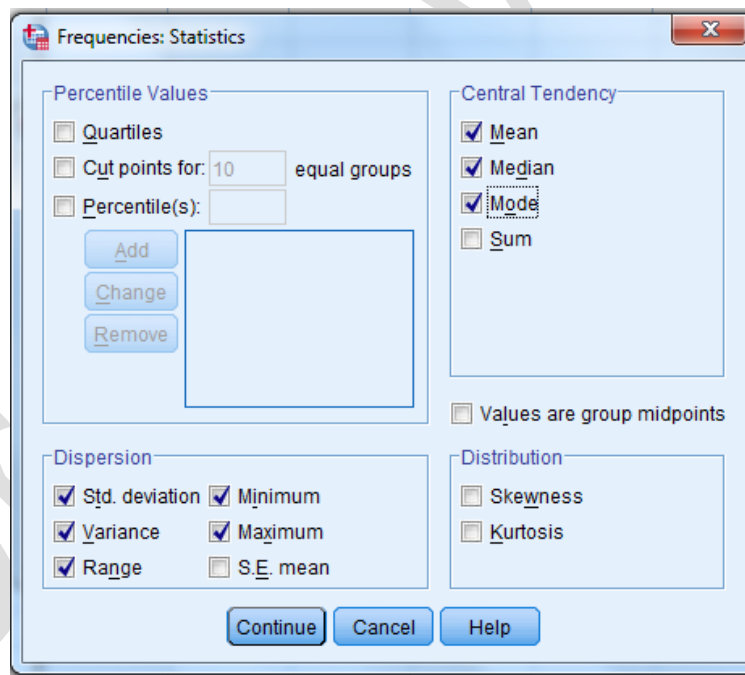
Select the variable you wish to compute. In order to select a variable,

click on that variable in left side box → click the **right arrow** button between the two boxes → the variable will shift to the right side **in Variables** list box

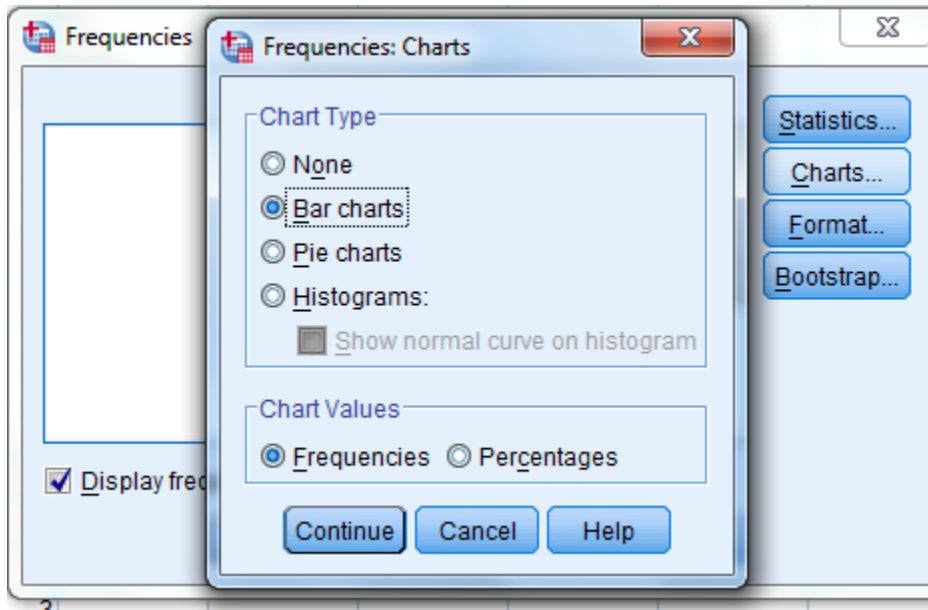
Similarly do this for all the variables.



Now click **Statistics** button: a sub dialogue box will appear on the screen. Click **check box**, as desired. In this working example , we have selected various statistics based on the following figure. After selecting click **continue**. The sub dialogue box will be closed and previous dialogue box will reappear.



Click **Charts** button to open its sub dialogue box. Click the **radio buttons**, as desired. In this working example we have selected radio button based on the following figure.



After selection click continue. The sub dialogue box will be closed and previous dialogue box will reappear. In this dialogue box click OK to open the Output viewer.

➤ OUTPUT

Frequencies

[DataSet1] C:\Users\SHINY\Desktop\phd-coursework\SPSS-SHINY\descriptive.sav

Statistics

	Gender	Age in years	Highest Education	Occupation	Monthly Salary
N Valid	25	25	25	25	25
Missing	0	0	0	0	0
Mean	1.32	2.56	2.28	2.00	2.40
Median	1.00	2.00	2.00	2.00	2.00
Mode	1	2	3	2	2
Std. Deviation	.476	1.158	.936	.645	1.384
Variance	.227	1.340	.877	.417	1.917
Range	1	4	3	2	4
Minimum	1	1	1	1	1
Maximum	2	5	4	3	5

Frequency Table**Gender**

	Frequency	Percent	Valid Percent	Cumulative Percent
Male	17	68.0	68.0	68.0
Valid Female	8	32.0	32.0	100.0
Total	25	100.0	100.0	

Age in years

	Frequency	Percent	Valid Percent	Cumulative Percent
Below 25	4	16.0	16.0	16.0
25-35	10	40.0	40.0	56.0
Valid 35-45	6	24.0	24.0	80.0
45-55	3	12.0	12.0	92.0
Above 55	2	8.0	8.0	100.0
Total	25	100.0	100.0	

Highest Education

	Frequency	Percent	Valid Percent	Cumulative Percent
Undergraduate	6	24.0	24.0	24.0
Graduate	8	32.0	32.0	56.0
Valid Post graduate	9	36.0	36.0	92.0
Doctrate	2	8.0	8.0	100.0
Total	25	100.0	100.0	

Occupation

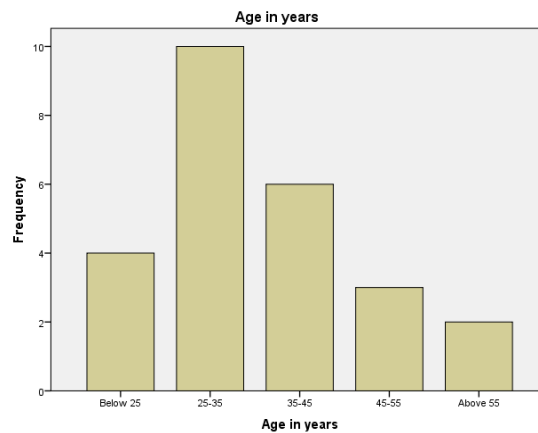
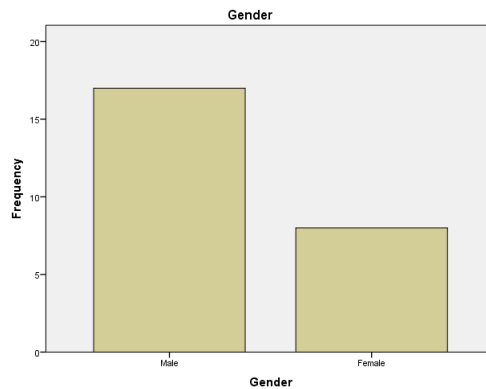
	Frequency	Percent	Valid Percent	Cumulative Percent
Student	5	20.0	20.0	20.0
Valid Business	15	60.0	60.0	80.0
Service	5	20.0	20.0	100.0

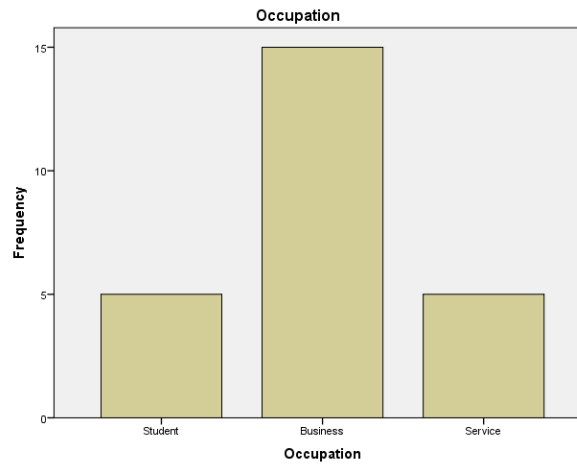
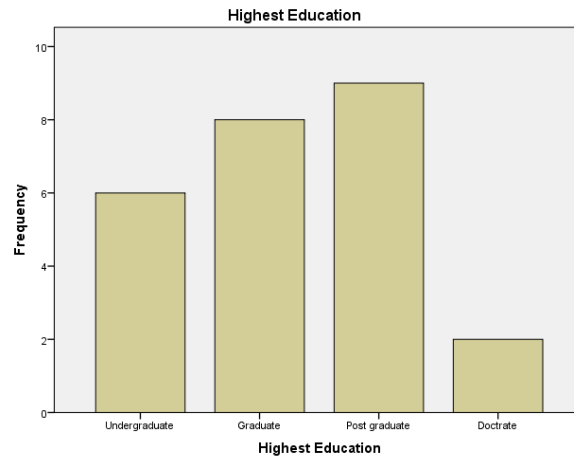
Total	25	100.0	100.0	
-------	----	-------	-------	--

Monthly Salary

	Frequency	Percent	Valid Percent	Cumulative Percent
Less than 20000	7	28.0	28.0	28.0
20000-30000	10	40.0	40.0	68.0
Valid 30000-40000	3	12.0	12.0	80.0
40000-50000	1	4.0	4.0	84.0
Above 50000	4	16.0	16.0	100.0
Total	25	100.0	100.0	

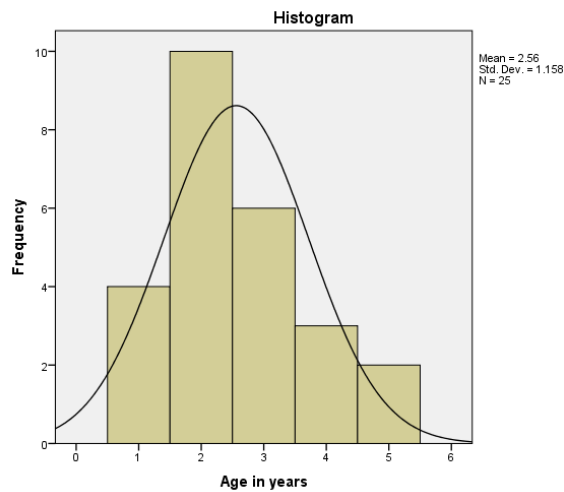
Bar Chart





For normal curve:

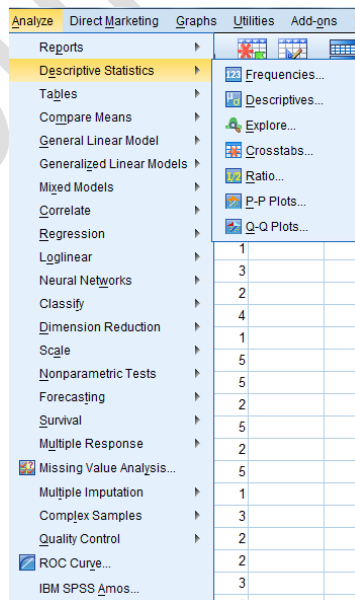
Click histograms → click normal curve.

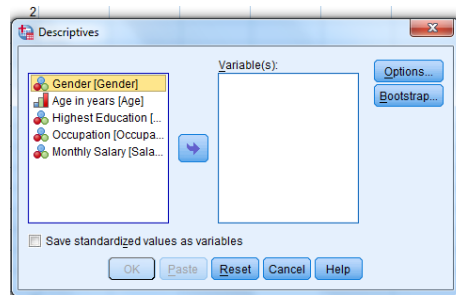


In order to start further analysis through descriptive,

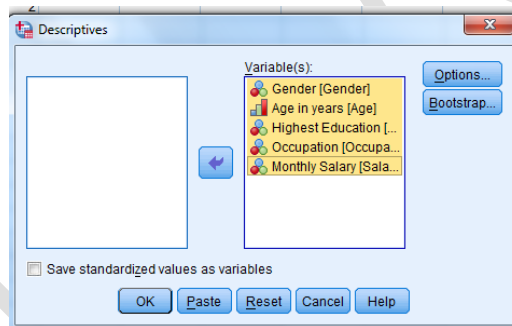
click Analyze menu → Descriptive Statistics → Descriptive

This will open Descriptive dialogue box.

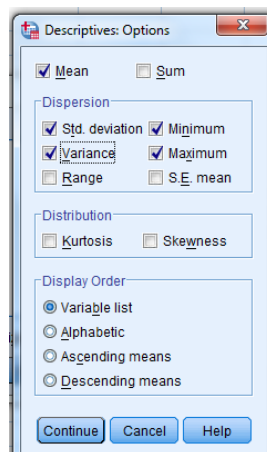




Select the variables, you wish to compute. In this working example we have selected all the five variables. In order to select a variable, click on that variable in left side box → click the **Right Arrow** button between the two boxes. The variable will shift to the right side in **Variables** list. Similarly, do this for all other variables. Now click '**Save Standardized values as variable**' check box to save the standardized z-scores for further computation (like interaction terms in multiple regression) or in comparing samples from different population.



Now click **Options** button to open its sub dialogue box. Click the check boxes, as desired. After selection click **continue**.



In current dialogue box click **OK** to open the Output Viewer.

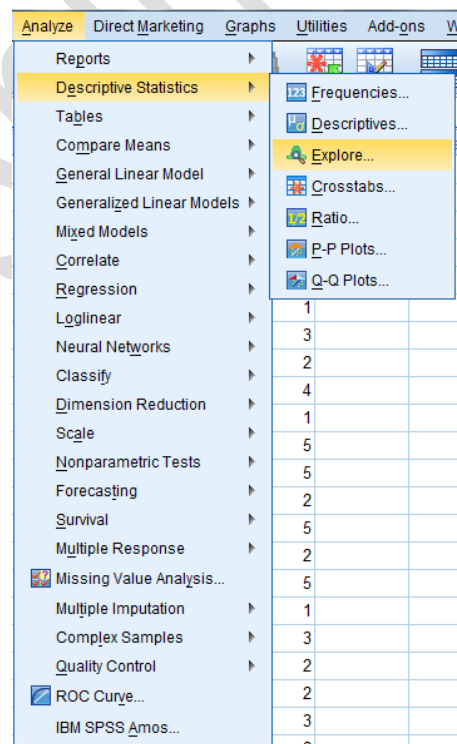
Descriptives

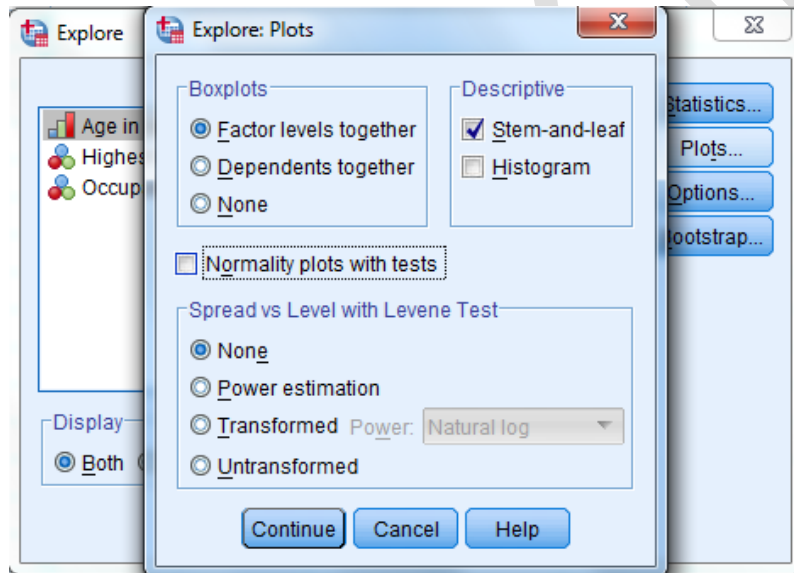
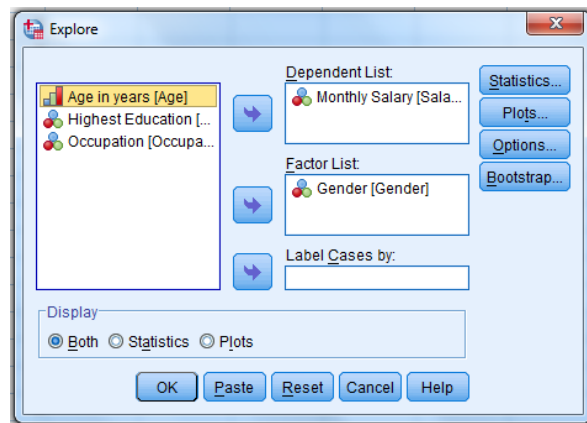
Descriptive Statistics

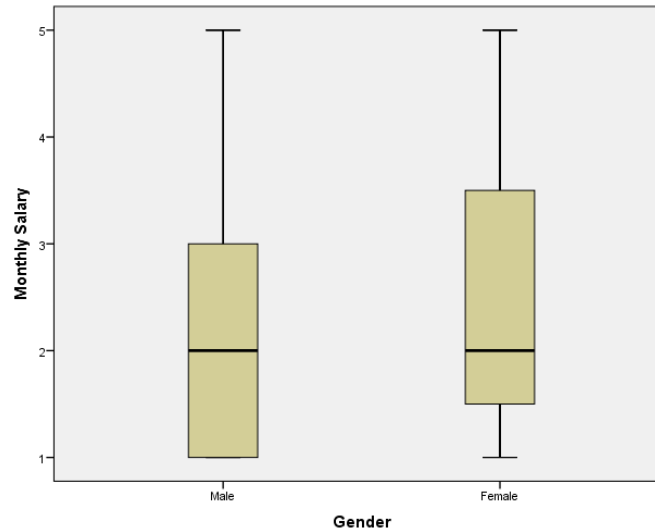
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Gender	25	1	2	1.32	.476	.227
Age in years	25	1	5	2.56	1.158	1.340
Highest Education	25	1	4	2.28	.936	.877
Occupation	25	1	3	2.00	.645	.417
Monthly Salary	25	1	5	2.40	1.384	1.917
Valid N (listwise)	25					

SPSS provides graphical and semigraphical techniques collectively referred to as exploratory data analysis (EDA). EDA is useful for describing the characteristics of a dataset, identifying outliers and providing summary descriptions. Most widely used EDA techniques are box plots and stem and leaf displays.

Click Analyze→Descriptive Statistics→Explore







EXERCISE

Gender	Age group	Education	Work experience group
Female	25-30	Under graduate	Below 3 years
Male	30-35	Graduate	3-5 years
	35-40	Post graduate	5-10 years
	40-45		Above 10 years
	45-50		

Calculate descriptive statistics. Draw pie chart for each observation. Test the Normality, using Histogram.

Chapter 3

Transforming Data

Windows can transform the values in several ways. SPSS for Windows can:

- Combine values of a variable into several categories,
- Create new variables out of old variables,
- Select particular cases and analyze only these cases,
- Weight cases so that some cases count more heavily than others.




SPSS provides many ways to transform data: recode, compute, if, and weight.

Recoding Variables

Recoding is a way of combining the values of a variable into fewer categories. Let me give a hypothetical example. Let's say you have conducted a survey and one of your demographic questions was the age of the respondent. Entering the actual age in years would be the simplest way of working with the data. But let's also say that you want to compare people of different age groupings. In other words, your data would be more useful if it was organized into collapsed categories, like "Young", "Middle age", and "Older". Using SPSS for Windows you could reorganize the data so that you had these three groupings. There are two things you need to know before you recode the values.

- *First, you need to decide the number of categories you want to end up with.*
- *Generally this will be determined by the way you plan to use the information.*

If you are going to analyze the data using a table where you cross-tabulate two variables , you probably want to limit the number of new categories to three or four. The second thing you need to know is which of the old values are going to be combined into a new category. For example, you might do something like this.

The actual age of the respondent as originally recorded in the data file.	The new, collapsed, category.
 18 years through 35 years	Young
 36 years through 55 years	Middle age
 56 years of age or over	Older

Another example might be if respondents were asked how often they prayed, and the original responses were several times a day, once a day, several times a week, once a week, less than once a week, once a month, once a year, or never. With recode we can combine the people who said "several times a day" with the people who said "once a day" and put all these respondents into a new category which we could call "a lot". Similarly, we could combine the people who said "several times a week" with those who said "once a week" and call this category "sometimes" and combine those who said "less than once a week" and "once a month" and call this category "infrequently". Those who said "once a year" or "never" could be combined into a fourth category called "hardly ever". Recoding is the process in SPSS that will do the above examples.

➤ WORKING EXAMPLE :

Please note that in the below table,

1	First column indicates the respondent number
2	Second column indicates the preference for fast food. The respondents were asked to state their preference for fast food on a 5-point scale where 1 = not at all preferred, 2 = not preferred, 3 = neutral, 4 = preferred, 5 = very much preferred.
3	Third column in the table indicates the actual age of the respondent
4	Fourth column of the table states the household monthly income coded as: 1 = Household income less than Rs.25,000/- per month (low income) 2 = Household income of Rs.25,000/- per month but less than Rs.50,000/- per month (middle income) 3 = Household income of Rs.50,000/- & above (High income)
5	Fifth column 1=Male 2=Female

Data Table:

Resp No.	Preference	Age	Income	Gender
1	1	46	2	2

2	2	24	1	1
3	4	22	3	1
4	4	18	3	1
5	2	46	1	1
6	2	38	1	1
7	1	47	1	2
8	2	54	2	2
9	5	50	3	2
10	4	46	3	1
11	3	29	1	1
12	2	32	2	2
13	4	26	2	1
14	5	19	2	1
15	4	41	3	1
16	2	20	1	1
17	3	36	2	2
18	4	31	3	1
19	5	28	1	2
20	2	54	1	1
21	4	30	1	2
22	2	46	2	1
23	3	37	3	1
24	4	22	3	2
25	3	26	1	2
26	4	47	3	2
27	3	45	1	2
28	2	50	1	1
29	2	54	2	2
30	5	26	3	2
31	3	41	1	1
32	5	42	2	1
33	1	61	3	1
34	2	31	1	1
35	1	19	3	1
36	3	20	1	1
37	4	29	3	2
38	3	26	2	1
39	4	31	3	1
40	4	28	3	2
41	3	41	2	2
42	2	51	1	1
43	1	49	1	1

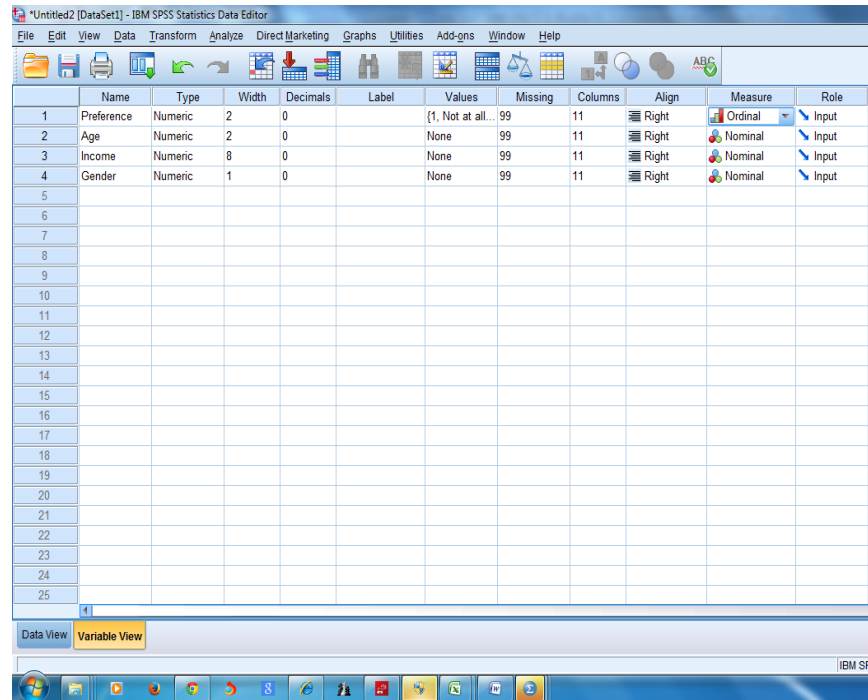
44	5	31	3	1
45	3	46	2	1
46	4	26	1	1
47	5	31	3	1
48	4	35	2	1
49	5	32	3	2
50	4	39	3	2
51	1	52	2	1
52	3	46	1	2
53	2	21	2	1
54	4	21	3	2
55	5	18	3	1
56	5	29	2	2
57	4	51	3	1
58	1	52	2	2
59	2	46	2	2
60	2	31	1	2
61	3	34	3	2
62	3	46	3	1
63	4	60	3	2
64	4	18	3	2
65	5	27	2	1
66	5	25	3	2
67	2	31	1	2
68	1	32	3	1
69	3	47	3	1
70	5	42	2	1
71	1	59	3	1
72	2	50	3	2
73	3	26	1	2
74	4	28	3	1
75	5	31	2	2
76	5	52	2	1
77	4	41	3	2
78	3	38	1	1
79	2	46	2	2
80	1	41	3	1
81	3	46	2	1
82	5	24	3	2
83	3	44	1	2
84	4	27	2	1
85	2	58	2	1

86	1	56	1	1
87	4	29	3	2
88	3	52	3	2
89	4	26	3	2
90	3	24	2	2
91	5	42	3	1
92	5	34	3	1
93	4	22	3	2
94	2	22	3	2
95	3	26	2	1
96	2	38	2	2
97	4	33	3	1
98	4	33	3	2
99	5	28	1	2
100	1	19	3	2

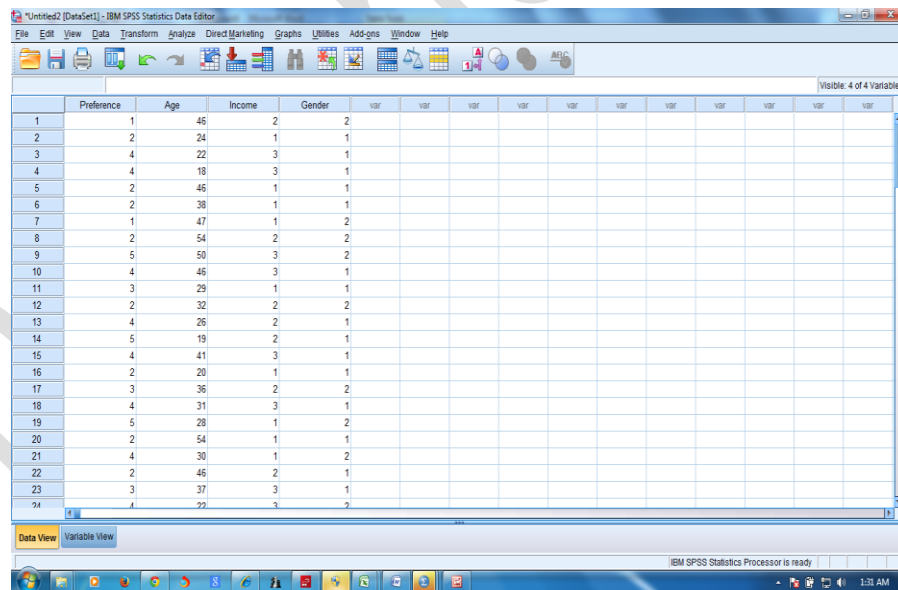
Another example might be if respondents were asked how often they prayed, and the original responses were several times a day, once a day, several times a week, once a week, less than once a week, once a month, once a year, or never. With recode we can combine the people who said "several times a day" with the people who said "once a day" and put all these respondents into a new category which we could call "a lot". Similarly, we could combine the people who said "several times a week" with those who said "once a week" and call this category "sometimes" and combine those who said "less than once a week" and "once a month" and call this category "infrequently". Those who said "once a year" or "never" could be combined into a fourth category called "hardly ever". Recoding is the process in SPSS that will do the above examples.

Create a SPSS file for the above data as shown below.

➤ VARIABLE VIEW

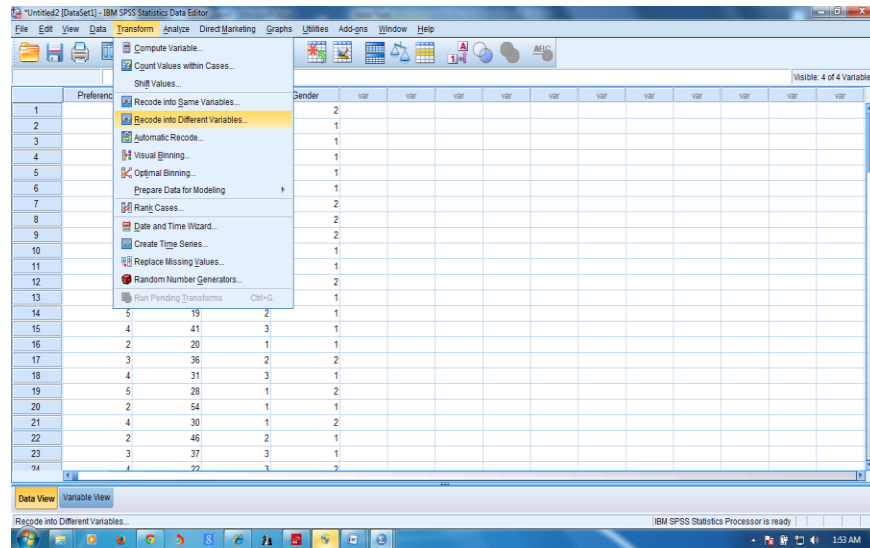


➤ DATA VIEW



Click on "Transform" → "Recode".

Your screen will look like:



Now we have two options: "Recode Into Different Variables" and "Recode Into Same Variables". It is strongly suggested to use the "Recording Into Different Variables" option. If you make an error, your original variable is still in the file and you can try again. If you make an error using "Recode Into the Same Variables", you have changed the original variable. If you also saved the file after doing this, and you did not have another copy of the file, you have just eliminated any chance of correcting your error.

Recoding Into Different Variables

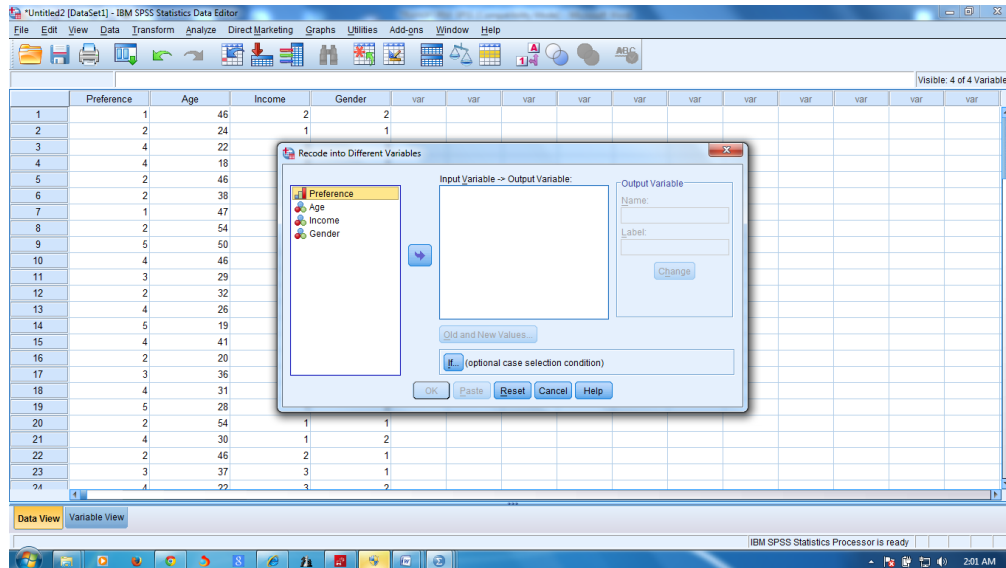
The recoding into a different variable starts with giving the new variable a variable name. For example, Let's take 'Preference'

If we recode into different variables we could combine the preference like

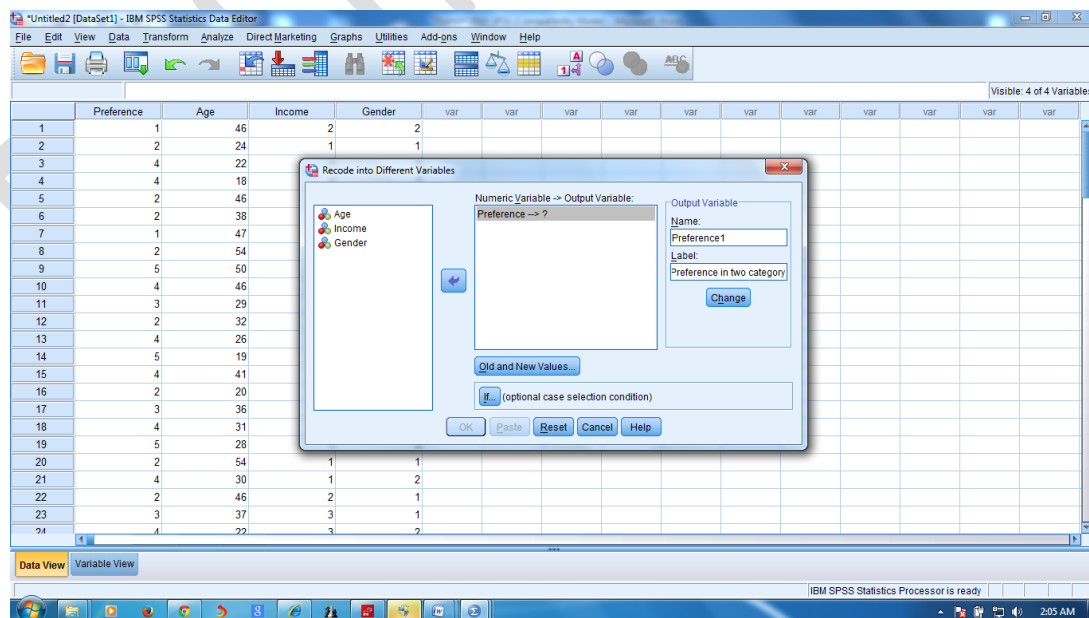
Not at all preferred=code 1, Not preferred=code 2, Neutral = code 3 as New variable-Not preferred as code=1.

Preferred=code 4, Very much preferred=code 5 preferred as code=2.

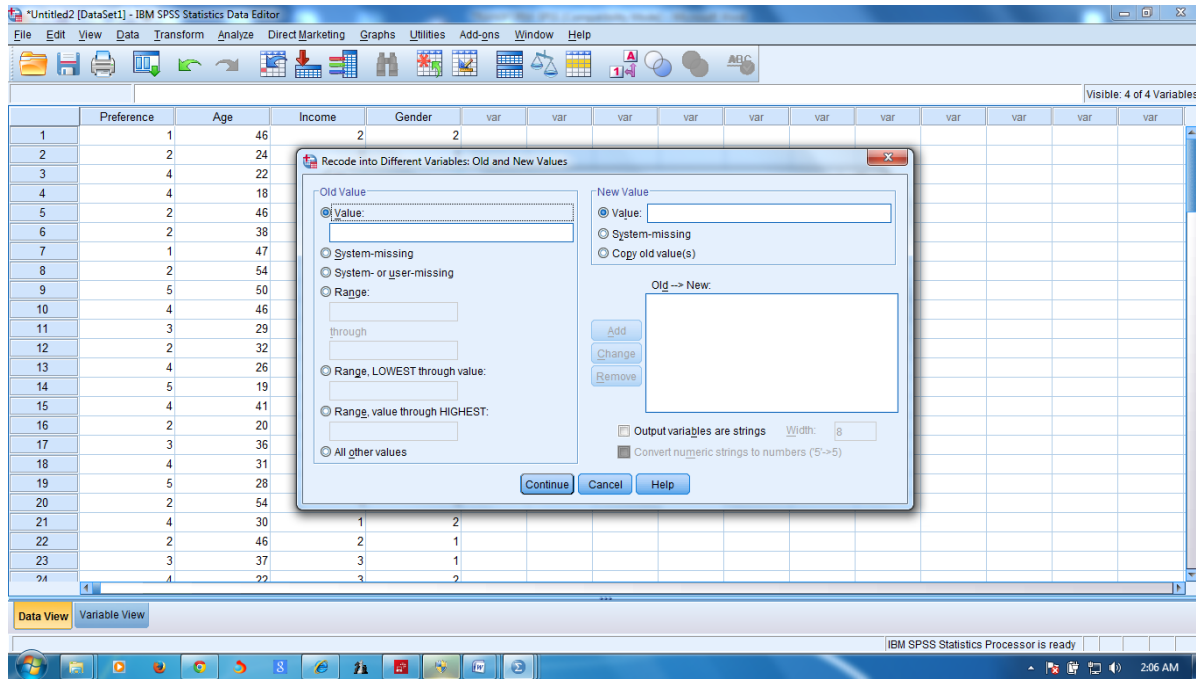
To do that, click on "Into Different Variables". Your screen will look like:



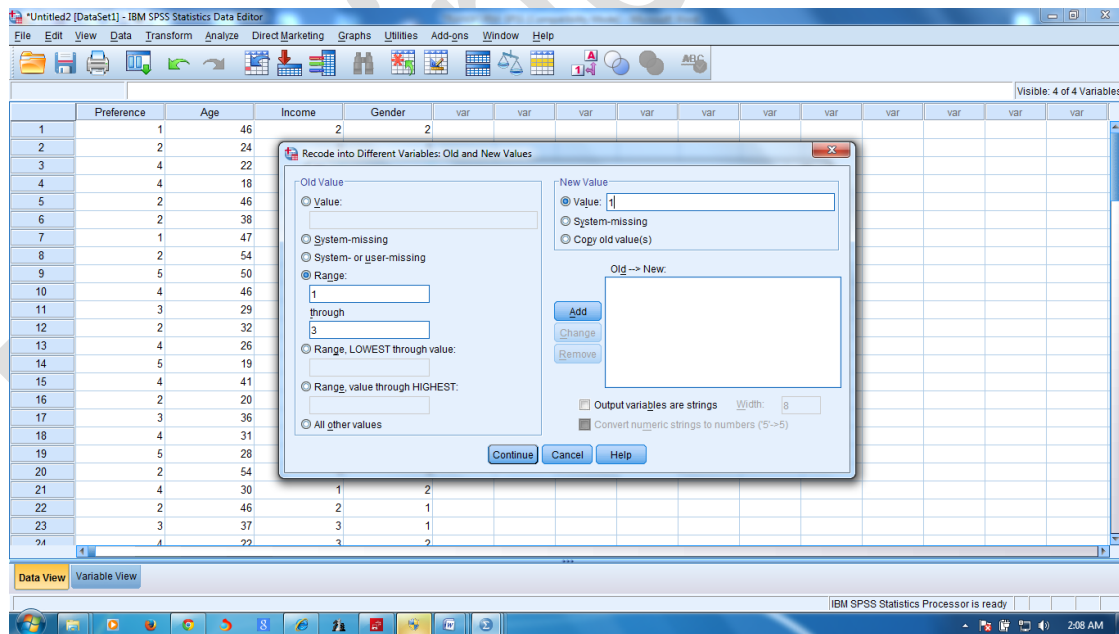
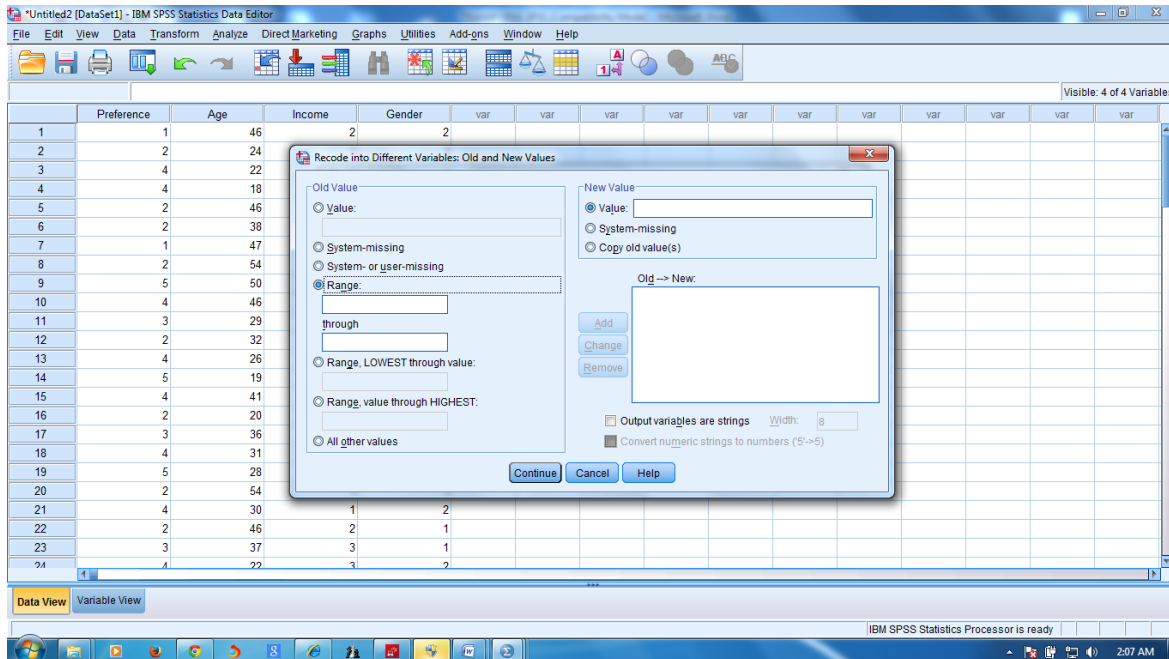
Find '**Preference**' in the list of variables on the left and click on it to highlight it, and then click on the arrow just to the left of the big box in the middle of the window. This will move **Preference** into the list of variables to recode. You want to give a name to this new variable so click in the "**Name**" box under Output Variable and type the name "**Preference1**" in this box. You can even type a variable label for this new variable in the Label box just below the Name box. Try typing "Age in Four Categories" as your label. Click on the "Change" button to tell SPSS to make these changes. Your screen will look like as below.

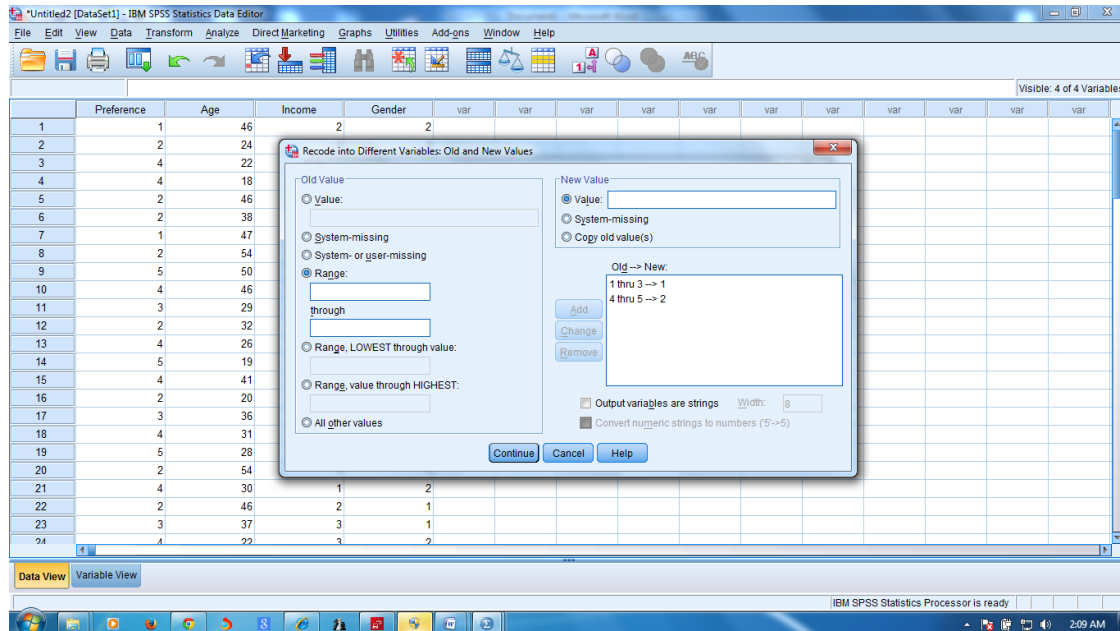


Now we have to tell SPSS how to create these categories. Click on the "Old and New Values" button at the bottom of the window. The screen will look like :



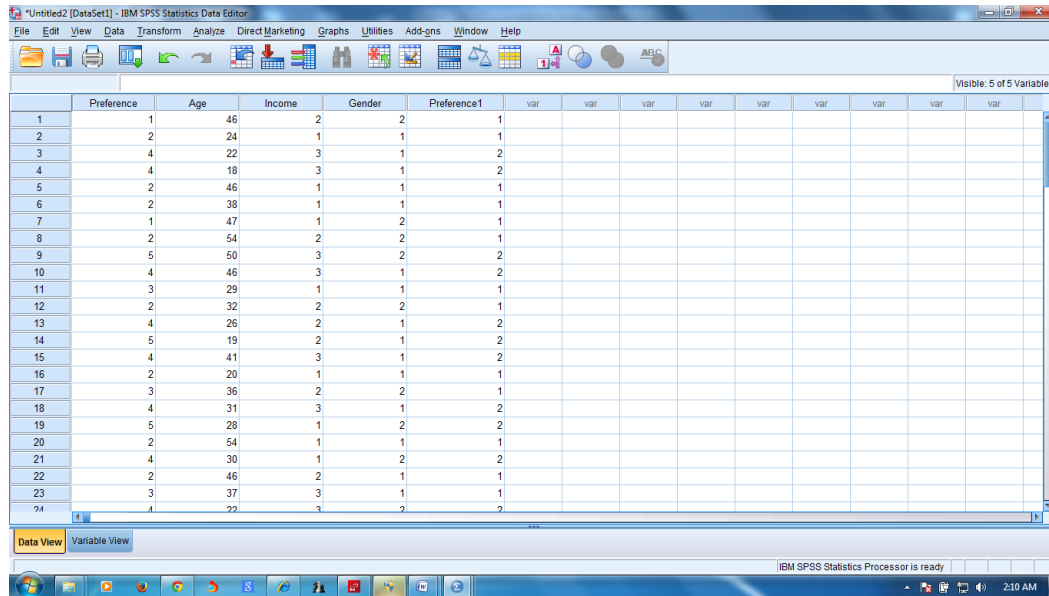
There are several options. You can change a particular value into a new value by entering the value to be changed into the Old Value box and the new value into the New Value box and then clicking on Add. You will usually change one "real" value to another "real" value. For example, change 1 thru 3 into value 1. There are also other options. You can also change a range of values into a new value and that is what we are going to do. Click on the fourth bubble from the top labeled **"Range"**. Notice how this marks this choice by filling in the bubble. Then type "1" (the youngest value in the data set) in the box to the left of "through", click on the box to the right of through, and type "3" in that box. Then click on "Value" just below New Value and type "1" in that box. This will have SPSS combine all preferences from 1 through 3 into a single category and give it the value of 1. Then click on "Add". Now do the same thing for the other categories. Click on the box under Range and type "4" in the box to the left of "through", click on the box to the right of through, and type "5" in that box. Click on "Value" just below New Value and type "2" in that box and click on Add.





To change one of your categories, highlight that category in the Old?> New box and make the changes, then click on **"Change"**. The new category should appear in the Old?>New box. To remove a category, highlight it and click on **"Remove"**.

Now we want SPSS to carry out the recoding. Click on **"Continue"** at the bottom of the window. This will take you back to the **"Recode into Different Variables"** box. Click on **"OK"** and SPSS will take a few seconds to carry out your commands. The data matrix should appear on the screen. When it says that the SPSS Processor is Ready at the bottom of the window you know that SPSS has finished with the recoding.



Click on "**Analyze**", then point your mouse at "**Descriptive Statistics**", and then click on "**Frequencies**". Notice that **Preference1** has appeared in the list of variables on the left.

Chapter 3

T-TEST

BASIC ASSUMPTIONS FOR ALL T TESTS:

1. *Scale of measurement-interval or ratio scale.*
2. *Random Sampling-data should be randomly sampled.*
3. *Population normality-data is numerical data representing samples from normally distributed populations.*

WORKING EXAMPLE:

Bharat Petroleum stocks has two kinds of petrol. 'Normal Petrol' and 'Speed Petrol'. 30 two wheelers were tested with Normal Petrol and with Speed Petrol. The two wheelers were having either 'Self-start' or 'Kick-start'.

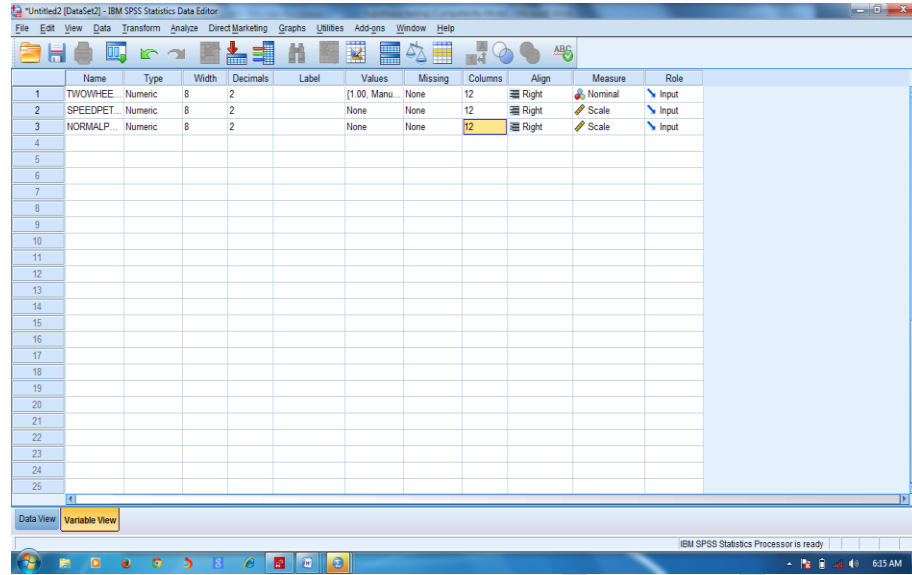
Two wheeler coding: **1(Manual), 2(Automatic)**

The earlier trail shows that mean number of Kilometer per litre was **14**.

Questions:

1. *Whether second trail efficiency of two wheeler is better than the previous trail?*
2. *Whether efficiency of engine improves with 'Speed Petrol'?*
3. *Whether efficiency of engine with speed petrol and normal petrol differs between self start and kick start?*

T Test using SPSS.



Enter the data as shown in the table below.

Two-wheeler Type	Speed- petrol	Normal- Petrol
2	17	17
2	18	17
1	22	21
1	24	20
2	20	17
1	22	20
2	12	13
1	21	22
2	11	10.5
2	10	12
2	20	15
1	15	14
1	22	17
2	18	19
1	21	19
2	16	21
1	20	18
1	25	17
2	16	16
1	18	19
2	26	18

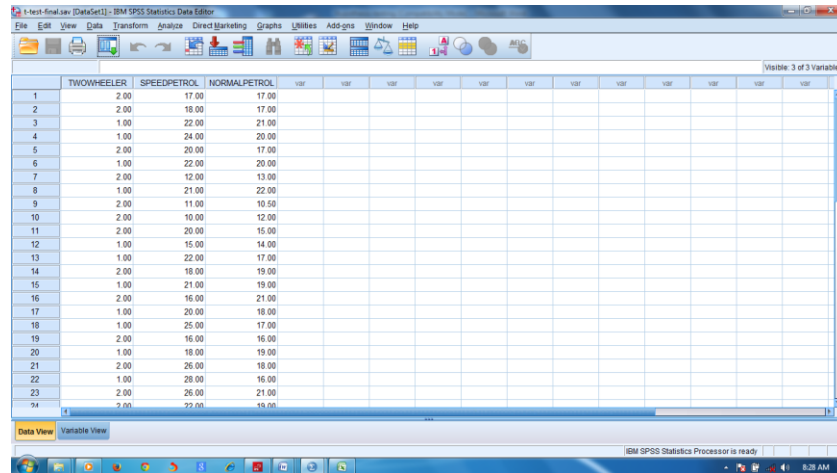
1	28	16
2	26	21
2	22	19
2	23	20
1	18	17
1	19	15
2	17	18
2	18	21
1	16	20
1	19	16
1	10	19
1	17	18
2	18	15
2	22	20
1	23	18
2	22	19
2	18	17
2	17	16
1	26	15

➤ T TEST WITH ONE SAMPLE

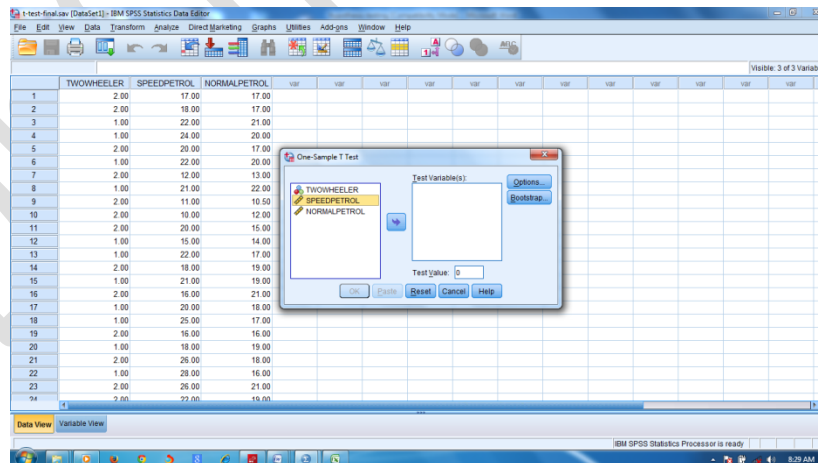
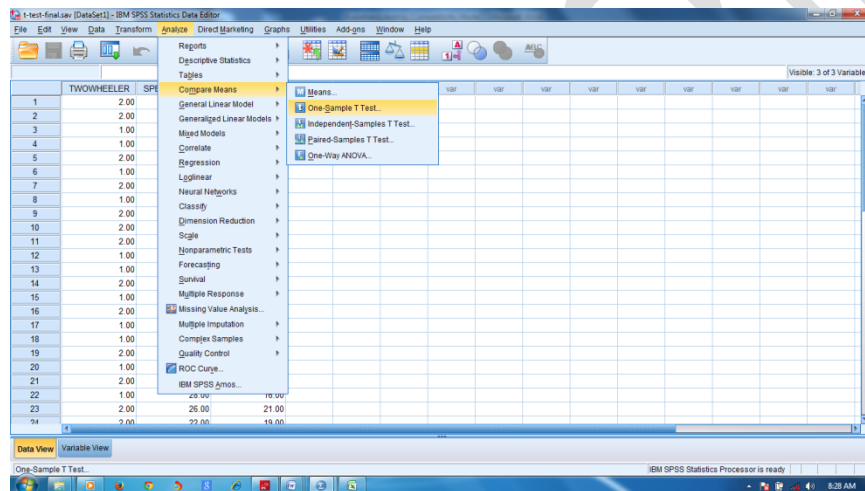
This test is used when data from a single sample is there and when you want to know whether the mean of population from which sample is drawn is same as the hypothesised mean.

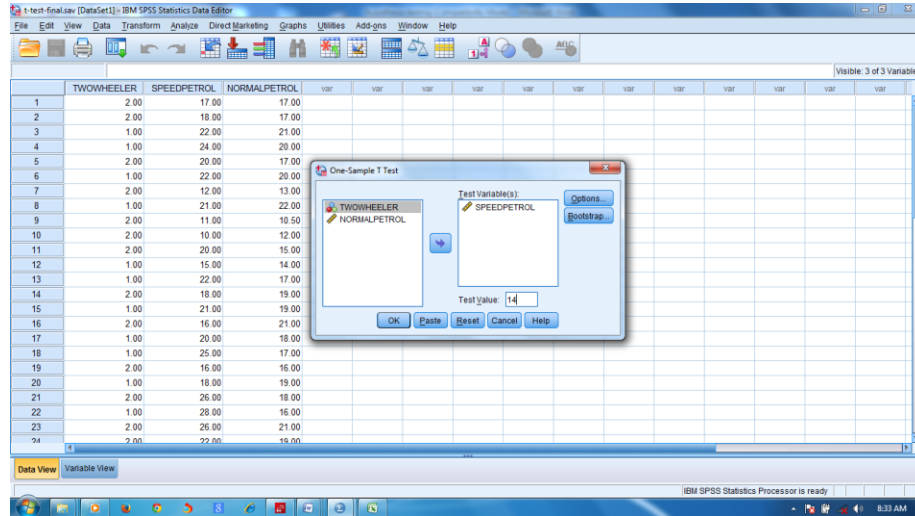
Click **Analyze**→ **Compare Means**→**One sample T test**.....This will open **One Sample T Test** dialogue box.

Select the test variable and click **Right Arrow** button to shift the variable in **Test Variable** list. Change the **Test Value** to 14, as in the working example it is 12. Click **OK** to see the Output Viewer.



	TWOWHEELER	SPEEDPETROL	NORMALPETROL	V10F	V10F	V10F	V10F	V10F	V10F	V10F	V10F	V10F	V10F
1	2.00	17.00	17.00										
2	2.00	18.00	17.00										
3	1.00	22.00	21.00										
4	1.00	24.00	20.00										
5	2.00	20.00	17.00										
6	1.00	22.00	20.00										
7	2.00	12.00	13.00										
8	1.00	21.00	22.00										
9	2.00	11.00	10.50										
10	2.00	10.00	12.00										
11	2.00	20.00	15.00										
12	1.00	15.00	14.00										
13	1.00	22.00	17.00										
14	2.00	18.00	19.00										
15	1.00	21.00	19.00										
16	2.00	16.00	21.00										
17	1.00	20.00	18.00										
18	1.00	25.00	17.00										
19	2.00	16.00	16.00										
20	1.00	18.00	19.00										
21	2.00	26.00	18.00										
22	1.00	28.00	16.00										
23	2.00	26.00	21.00										
24	2.00	22.00	19.00										





➤ OUT PUT

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
SPEEDPETROL	30	19.3667	4.35085	.79435

The above table shows the mean and standard deviation values of one sample t test along with standard error mean.

One-Sample Test

	Test Value = 14					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SPEEDPETROL	6.756	29	.000	5.36667	3.7420	6.9913

The value of two tail significance is less than 0.05 ($p < 0.05$), as such the difference between means is significant. The output data indicates that there is a significant difference in engine efficiency between the previous and current trial. The two wheeler in the current trial have more engine efficiency than those in earlier trial with $t(29) = 6.756, p < 0.05$.

➤ T TEST WITH MORE THAN ONE SAMPLE:

This test is used to ask whether two sets of values are random samples from same or different populations.

- If they are random samples from same population, then any differences across conditions or groups can be attributed to random sampling variability.

- *If two sets of values are random samples from different populations, then you can attribute any difference between means across condition to the independent variable.*

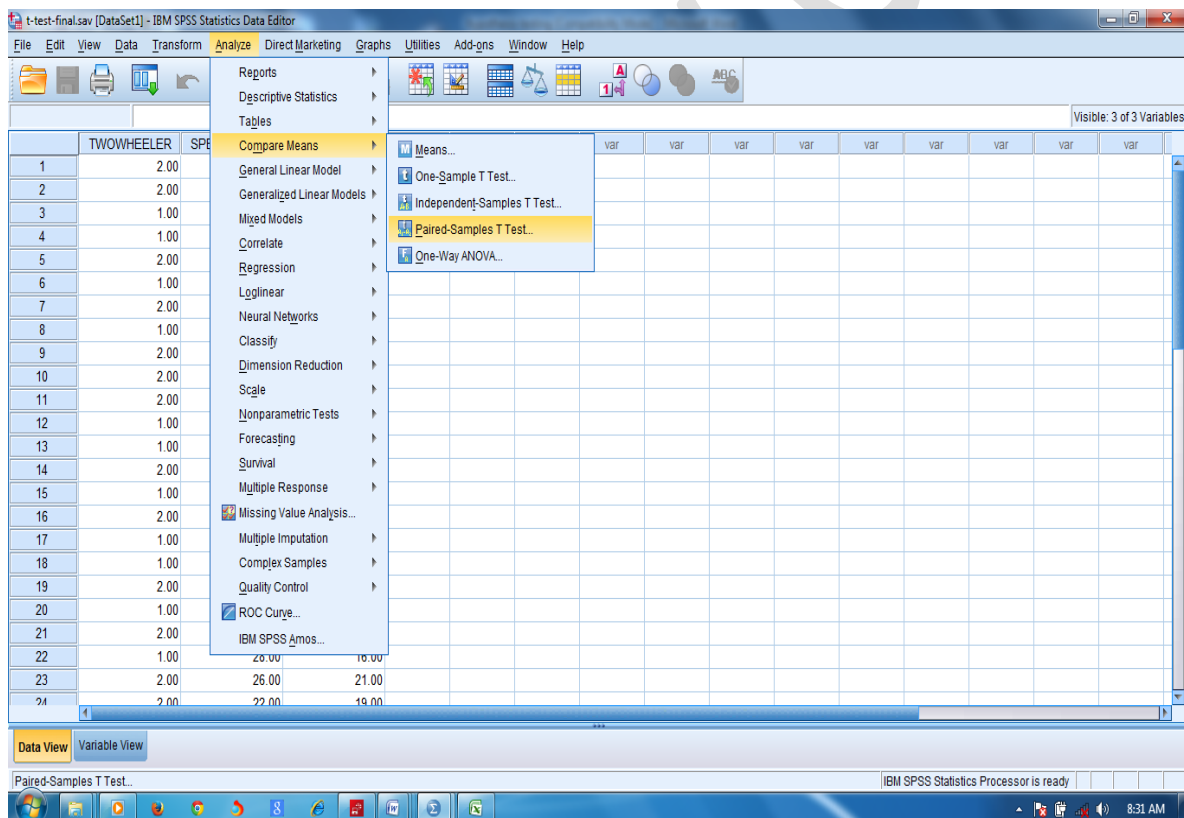
➤ REPEATED MEASURES T-TEST(DEPENDENT SAMPLES/ PAIRED T-TESTS)

This test is used when you have data from one group of participants that individual obtains two values under different levels of the independent variable.

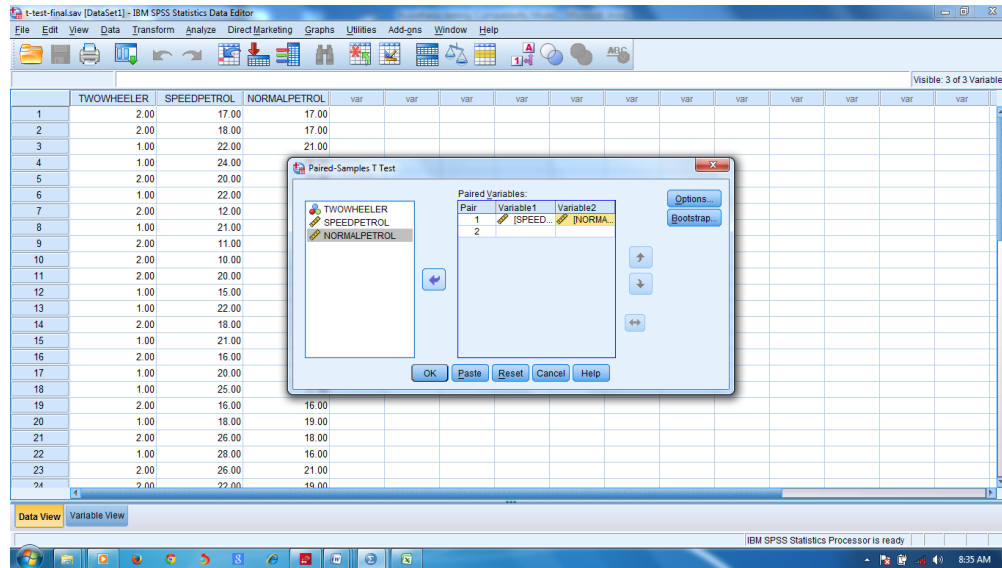
Additional Assumptions:

1. **Normality of population difference value**-the difference between the values for each respondent should be normally distributed. The sample size is not too small (30+).

Click **Analyze → Compare Means → Paired Sample T Tests.....**This will open **Paired Sample T Test** dialogue box.



Select the paired variables according to your requirement. In this example Normal Petrol and Speed Petrol are paired variables. Send the paired variables in the **Paired Variables** list box as shown in the figure below. Click **OK** to see the output viewer.



➤ OUT PUT

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SPEEDPETROL	19.3667	30	4.35085	.79435
	NORMALPETROL	17.6500	30	2.83801	.51815

The table shows paired sample statistics of with Normal Petrol and Speed Petrol responses.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	SPEEDPETROL & NORMALPETROL	30	.544	.002

The table above shows the paired sample correlations with correlation value of 0.544, $p < 0.05$.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	SPEEDPETROL - NORMALPETROL	1.71667	3.68053	.67197	.34234	3.09100	2.555	29	.016

The value of two tail significance is less than 0.05($p < 0.05$), as such the difference between mean is significant. The output indicates there is a significant difference in the engine efficiency between Speed Petrol and Normal Petrol trail. The two wheelers with speed petrol have more engine efficiency than those with, Normal Petrol with $t(29)=2.555$, $p < 0.05$.

➤ INDEPENDENT GROUPS T-TEST

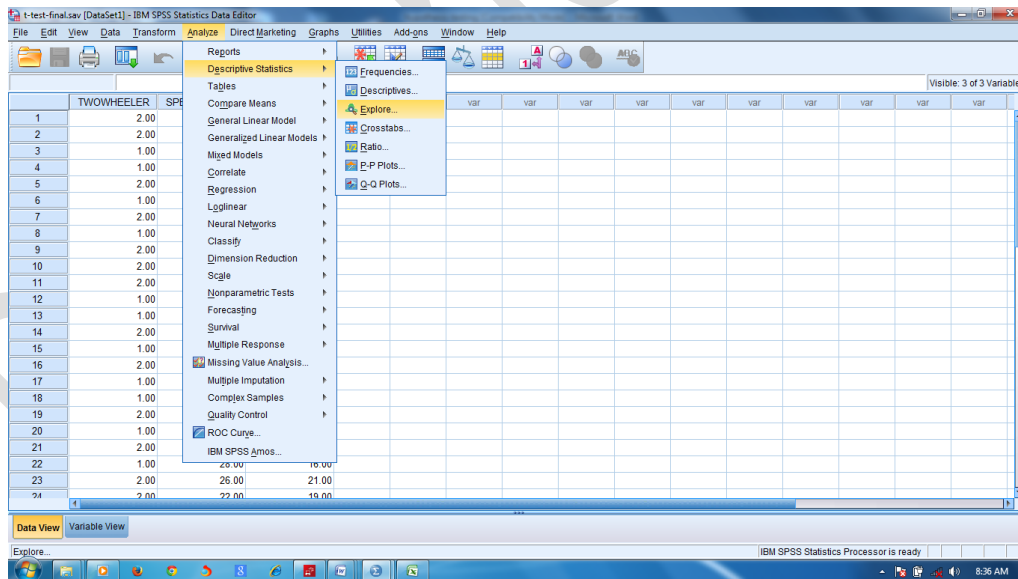
This test is appropriate when respondent in one condition are different from respondents in other condition. This is commonly referred to as between subject design.

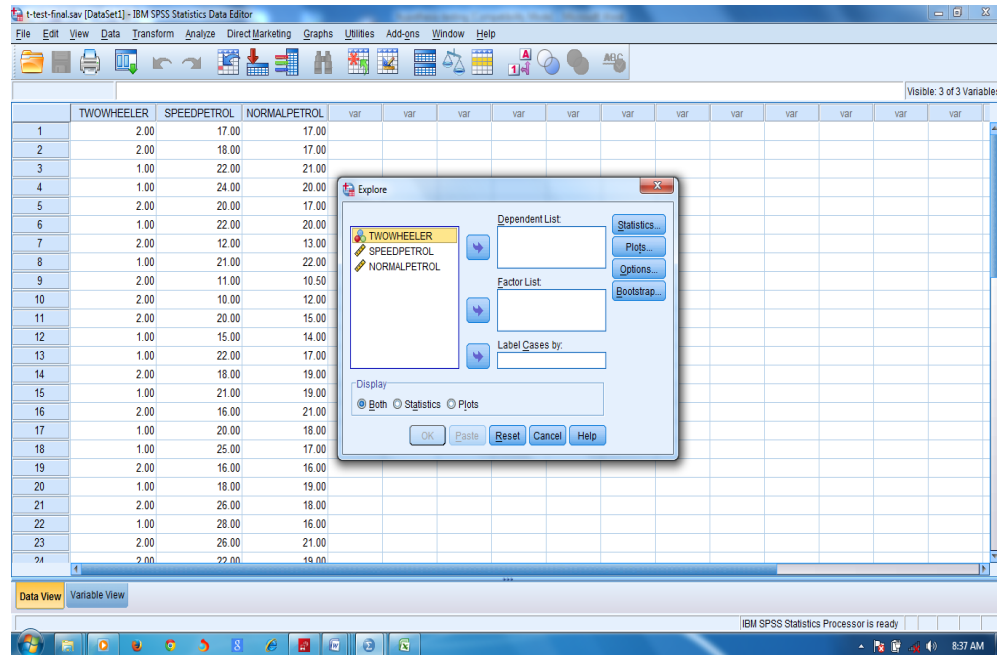
Additional Assumptions:

- **Independence of Groups**-respondent should appear in only one group and these groups should be unrelated.
- **Homogeneity of variance**-The variance of groups should be similar.

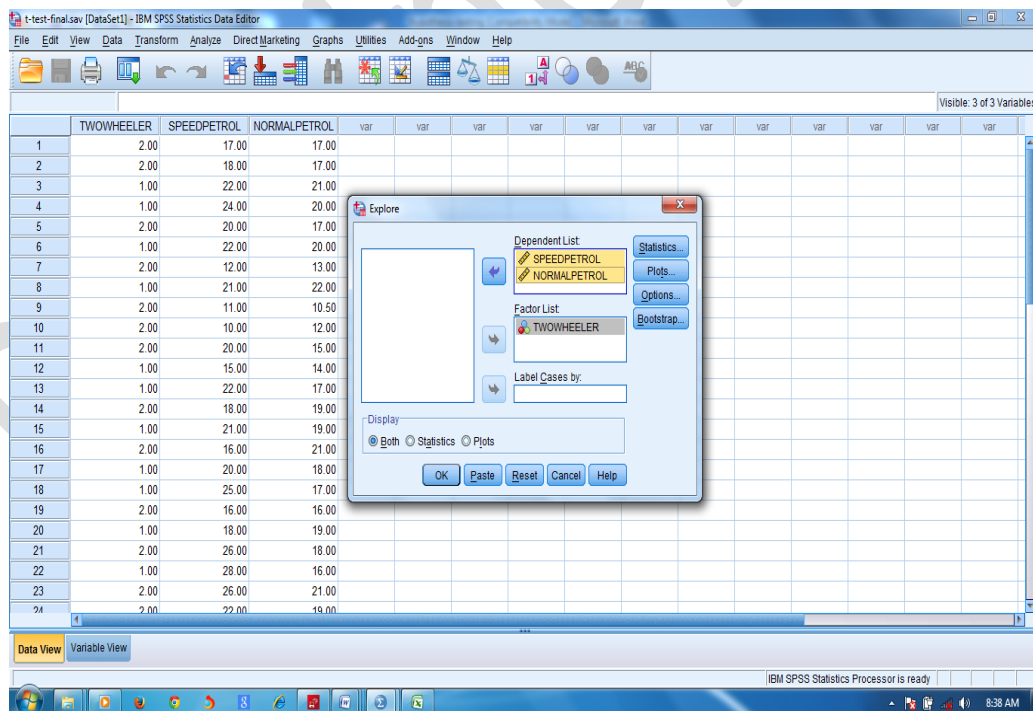
To check normality assumption:

Click **Analyze ⇒ Descriptive Statistics ⇒ Explore.....** This will open **Explore dialogue box**.





Select the dependent variable and click **Upper Right Arrow** button to shift the variable into **Dependent List**. Similarly do this for other variable(s). Now, shift the independent variable into **Factor List** as shown below. Click **OK** to see the output viewer.



➤ **SPSS OUT PUT****Case Processing Summary**

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
SPEEDPETROL	Manual	14	100.0%	0	0.0%	14	100.0%
	Automatic	16	100.0%	0	0.0%	16	100.0%
NORMALPETROL	Manual	14	100.0%	0	0.0%	14	100.0%
	Automatic	16	100.0%	0	0.0%	16	100.0%

Descriptives

		Statistic	Std. Error
SPEEDPETROL	Mean	20.7857	.93866
	95% Confidence Interval for Mean	Lower Bound 22.8136	
		Upper Bound 20.7063	
	5% Trimmed Mean	20.7063	
	Median	21.0000	
	Variance	12.335	
	Std. Deviation	3.51215	
	Minimum	15.00	
	Maximum	28.00	
	Range	13.00	
	Interquartile Range	4.50	
	Skewness	.282	.597
	Kurtosis	.111	1.154
	Mean	18.1250	1.18278
	95% Confidence Interval for Mean	Lower Bound 20.6460	
		Upper Bound 18.1389	
Automatic	5% Trimmed Mean	18.1389	
	Median	18.0000	
	Variance	22.383	
	Std. Deviation	4.73110	
	Minimum	10.00	

NORMALPETROL	Manual	Maximum	26.00	
		Range	16.00	
		Interquartile Range	5.50	
		Skewness	.010	.564
		Kurtosis	-.349	1.091
		Mean	18.2143	.62175
		95% Confidence Interval for Mean	Lower Bound 16.8711 Upper Bound 19.5575	
		5% Trimmed Mean	18.2381	
		Median	18.5000	
		Variance	5.412	
		Std. Deviation	2.32639	
		Minimum	14.00	
		Maximum	22.00	
	Automatic	Range	8.00	
		Interquartile Range	3.25	
		Skewness	-.216	.597
		Kurtosis	-.709	1.154
		Mean	17.1563	.80327
		95% Confidence Interval for Mean	Lower Bound 15.4441 Upper Bound 18.8684	
		5% Trimmed Mean	17.3125	
		Median	17.5000	
		Variance	10.324	
		Std. Deviation	3.21309	
		Minimum	10.50	
		Maximum	21.00	
		Range	10.50	
		Interquartile Range	4.50	
		Skewness	-.708	.564
		Kurtosis	-.216	1.091

➤ SPSS OUT PUT

SPEEDPETROL Stem-and-Leaf Plot for

TWOWHEELER= Manual

Frequency Stem & Leaf

.00	1 .
5.00	1 . 56889
7.00	2 . 0112224
2.00	2 . 58

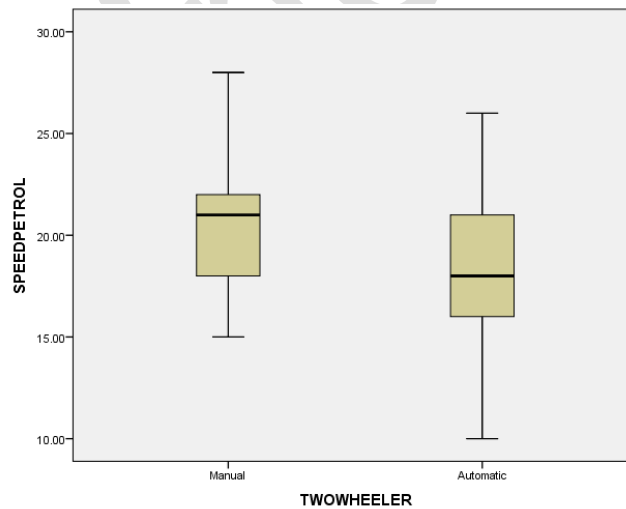
Stem width: 10.00
Each leaf: 1 case(s)

SPEEDPETROL Stem-and-Leaf Plot for
TWOWHEELER= Automatic

Frequency Stem & Leaf

3.00	1 . 012
7.00	1 . 6677888
4.00	2 . 0023
2.00	2 . 66

Stem width: 10.00
Each leaf: 1 case(s)



NORMALPETROL Stem-and-Leaf Plot for
TWO Wheeler= Manual

Frequency Stem & Leaf

1.00 1 . 4
8.00 1 . 56777899
5.00 2 . 00012

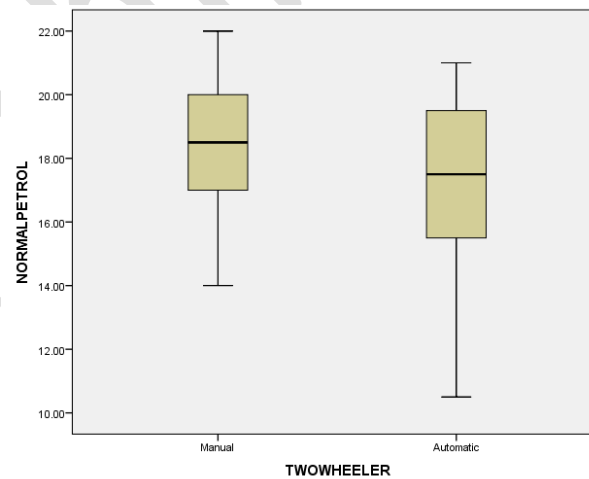
Stem width: 10.00
Each leaf: 1 case(s)

NORMALPETROL Stem-and-Leaf Plot for
TWO Wheeler= Automatic

Frequency Stem & Leaf

3.00 1 . 023
9.00 1 . 567778899
4.00 2 . 0111

Stem width: 10.00
Each leaf: 1 case(s)

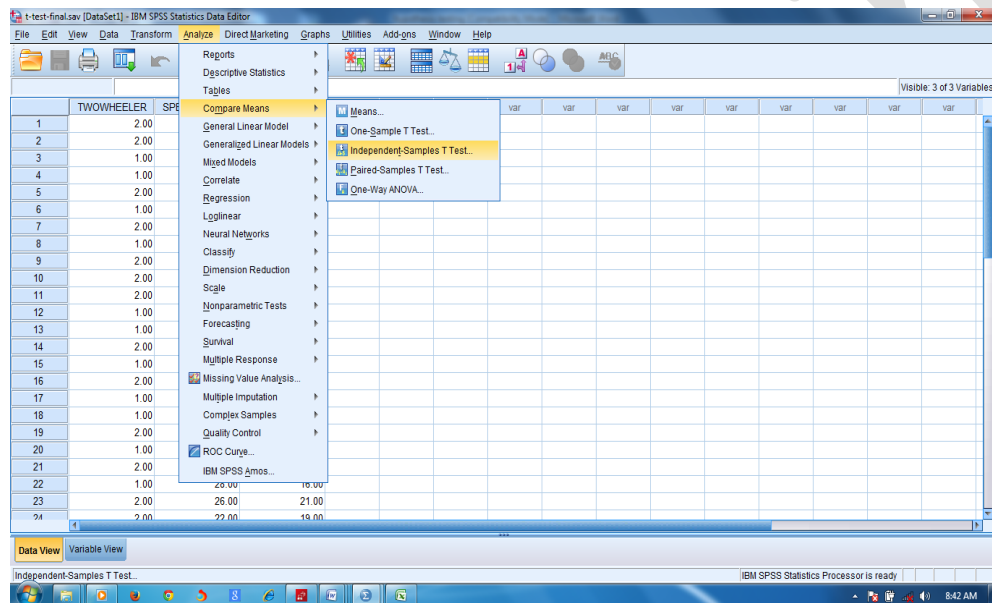


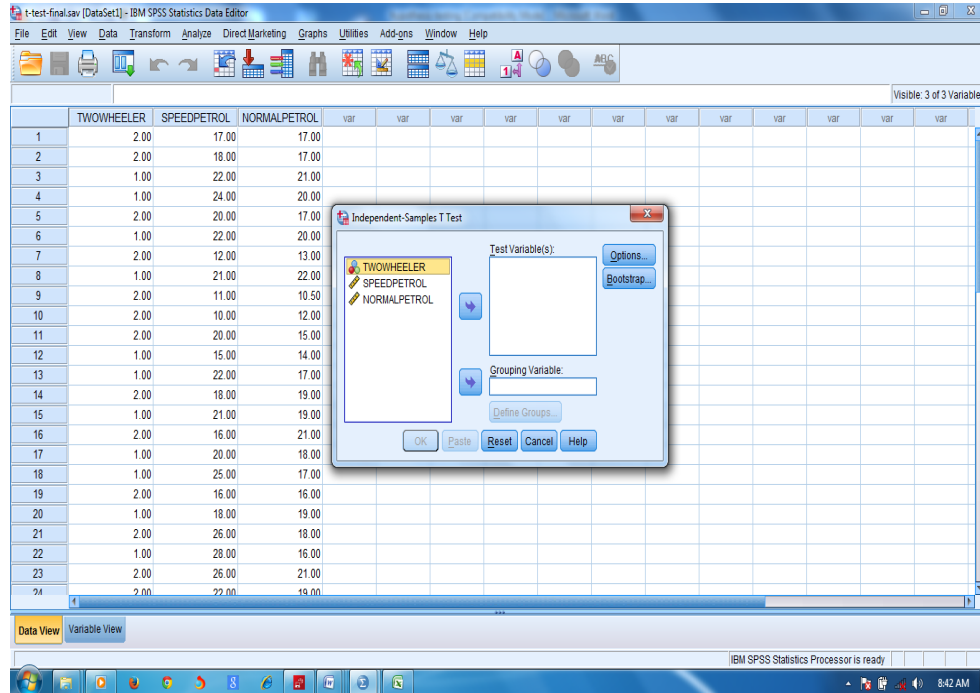
The above analysis shows that normality assumption is satisfied. We can proceed further with analysis.

➤ **TO CONDUCT INDEPENDENT GROUP T-TEST:**

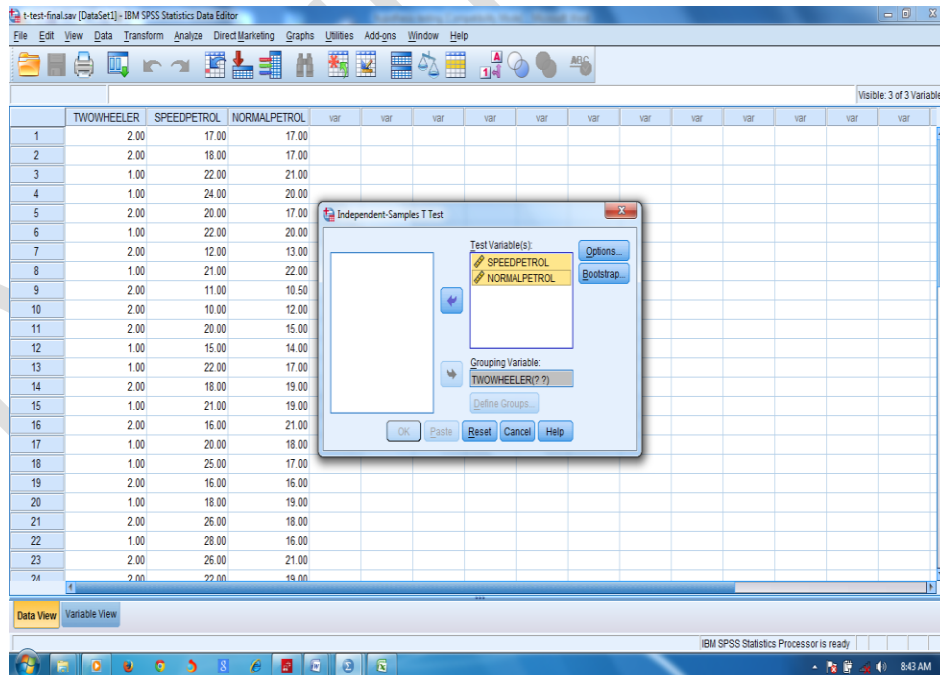
Click Analyze → **Compare Means** → **Independent- Sample T test**

This will open **Independent Samples T test** dialogue box.

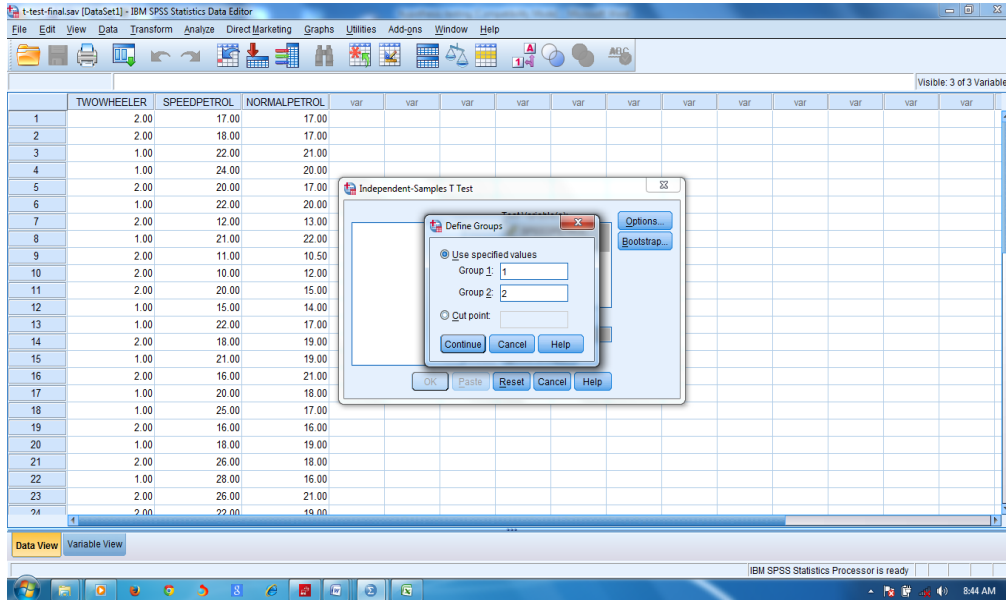




Select the variable and click upper right arrow button to shift the variable in Test variable(s) list box. Similarly, do this for another variable to be tested. Now, click the grouping variable and end it in grouping Variable list box as shown in the figure below.



Click define group button below Grouping Variable. This will open Define Groups sub dialog box. Enter the values as desired for analysis. In this case enter values as shown and click OK.



➤ OUT PUT

Group Statistics

	TWOWHEELER	N	Mean	Std. Deviation	Std. Error Mean
SPEEDPETROL	Manual	14	20.7857	3.51215	.93866
	Automatic	16	18.1250	4.73110	1.18278
NORMALPETROL	Manual	14	18.2143	2.32639	.62175
	Automatic	16	17.1563	3.21309	.80327

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper

SPEEDPETROL	Equal variances assumed	.783	.384	1.727	28	.095	2.66071	1.54044	-.49474	5.81617
	Equal variances not assumed			1.762	27.334	.089	2.66071	1.50998	-.43574	5.75717
NORMALPETROL	Equal variances assumed	.807	.377	1.019	28	.317	1.05804	1.03791	1.06802	3.18409
	Equal variances not assumed			1.042	27.124	.307	1.05804	1.01579	1.02574	3.14181

The Levene's test in Speed Petrol shows probability greater than 0.05, that is 0.383, which says that population variance is relatively equal. The two tail significance for speed petrol shows that $p > 0.05$ and therefore is not significant. We can say that two groups come from same population since no significant difference exist $-t(28)=1.727, p > 0.05$. This can also said as Null Hypothesis. No Significant difference exist in two wheelers types in case of speed petrol.

The Levene's test in Normal Petrol shows probability greater than 0.05, that is 0.377, which says that population variance are relatively equal. The two tail significance for speed petrol shows that $p > 0.05$ and therefore is not significant. We can say that two groups come from same population since no significant difference exist $-t(28)=1.019, p > 0.05$. This can also said as Null Hypothesis. No Significant difference exist in two wheelers types in case of Normal petrol.

➤ EXERCISE:

'Talent Hunt' Soft skill training institute has conducted QUIZ1 and QUIZ2 on General awareness. In total 25 students appeared for the same.

Student -GENDER coding: **1(MALE), 2(FEMALE)**

The earlier mean was 36.

Questions:

1. Whether Quiz 2 was better than QUIZ 1.

2. *Whether the score has improved in QUIZ2?*

3. *Whether there is difference in score between MALE & FEMALE Gender.*

GENDE R	1	1	1	2	2	1	2	1	2	2	1	1	2	1	2	2	2	2	2	1	1	2	1	2	2
QUIZ1	4	4	3	4	3	3	4	4	4	3	4	3	3	3	4	4	4	3	3	3	4	4	3	3	3
	5	6	4	5	6	7	3	4	3	0	1	2	6	2	1	7	7	6	1	9	1	9	8	3	2
QUIZ2	3	4	4	4	4	3	3	4	4	4	3	2	4	3	4	4	4	3	2	3	3	4	4	3	4
	4	1	8	3	1	6	7	8	7	2	3	7	1	8	5	9	9	2	9	5	8	7	1	4	1

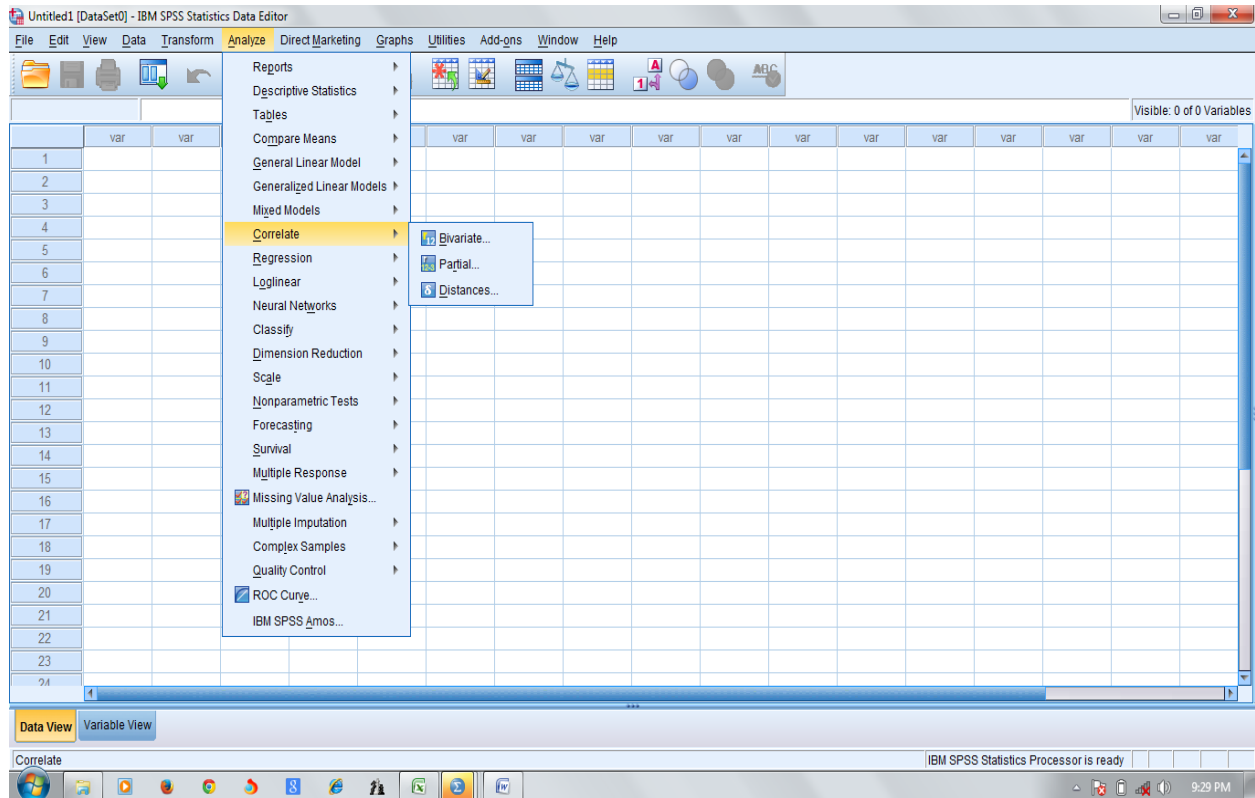
Chapter 5

Correlation and Regression Analysis

Correlation coefficient summarizes and quantifies the relationship between two variables in a single number and it is denoted by small letter 'r'. This number range from -1 to +1. -1 represents a perfect negative or inverse relationship to 0 representing no relationship or complete independence to +1 representing a perfect positive or direct relationship. When we calculate a coefficient correlation from sample data, we need to see whether it is having positive, negative and no relation. Scatter plot helps to examine the nature of relationship. Sometimes correlation is low not because of a lack of relationship but because of a lack of linear relationship. In such case, examining the scatter plot will assist in finding whether the relationship may be non linear.

Steps for calculating correlation:

Analyze → correlate → Bivariate.....



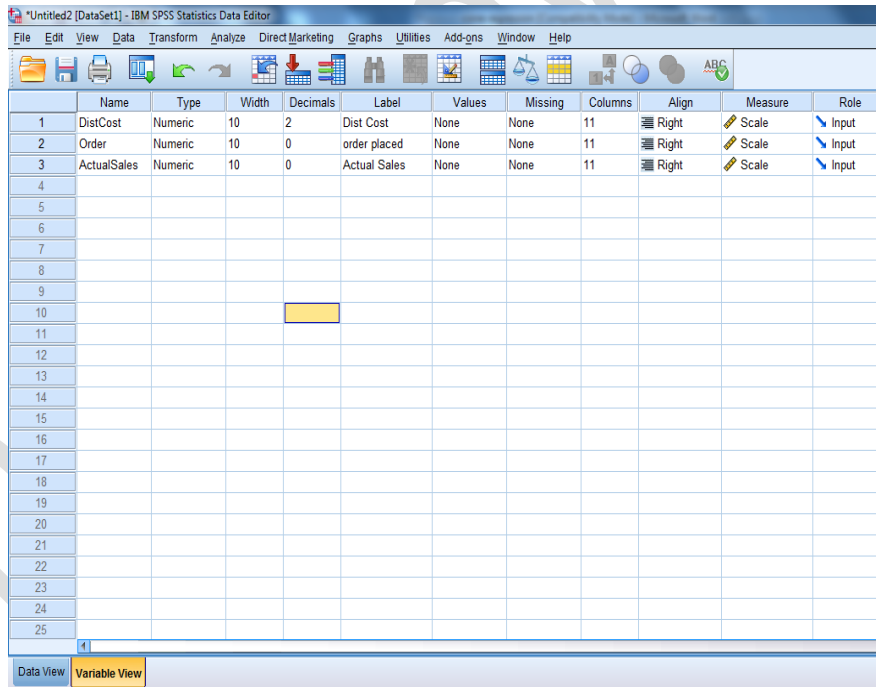
➤ **WORKING EXAMPLE:**

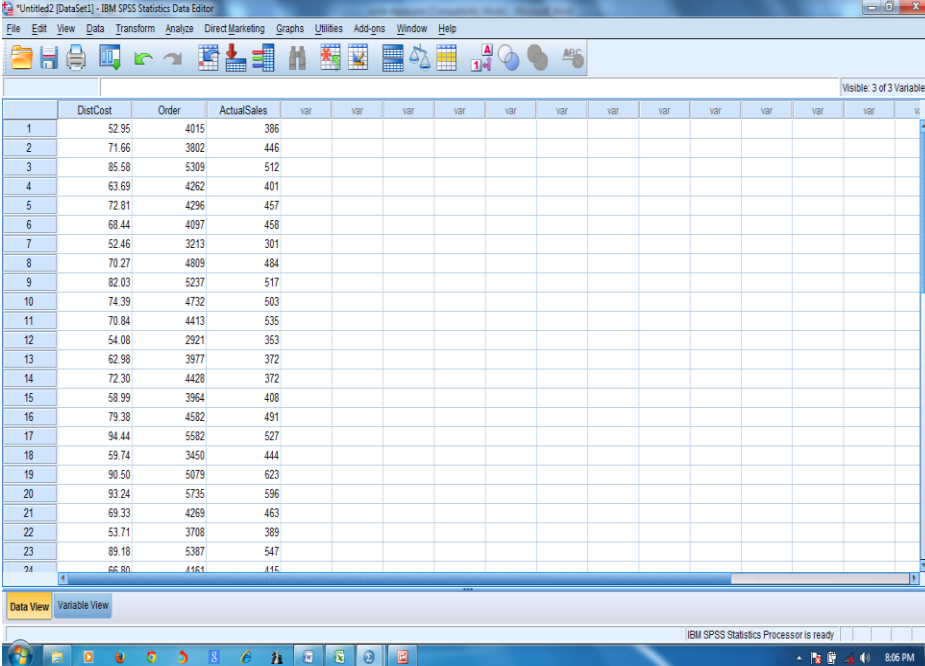
Table shows the data of sales department. Distribution cost is given in thousands. Check whether any correlation exist between distribution cost , order placed and actual sales.

Sr.No.	Distribution Cost	Orders placed	Actual Sales
1.	52.95	4015	386
2.	71.66	3802	446
3.	85.58	5309	512
4.	63.69	4262	401
5.	72.81	4296	457
6.	68.44	4097	458
7.	52.46	3213	301
8.	70.27	4809	484
9.	82.03	5237	517
10.	74.39	4732	503
11.	70.84	4413	535

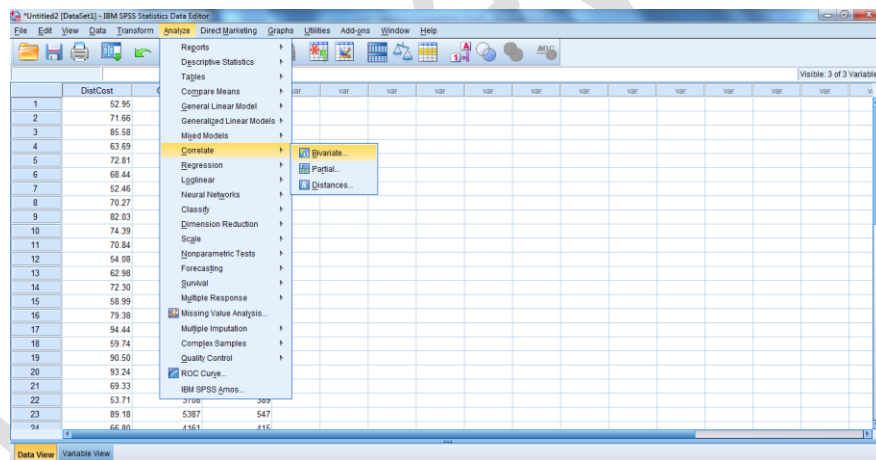
12.	54.08	2921	353
13.	62.98	3977	372
14.	72.3	4428	372
15.	58.99	3964	408
16.	79.38	4582	491
17.	94.44	5582	527
18.	59.744	3450	444
19.	90.5	5079	623
20.	93.24	5735	596
21.	69.33	4269	463
22.	53.71	3708	389
23.	89.18	5387	547
24.	66.8	4161	415

Create a SPSS file for the above data.

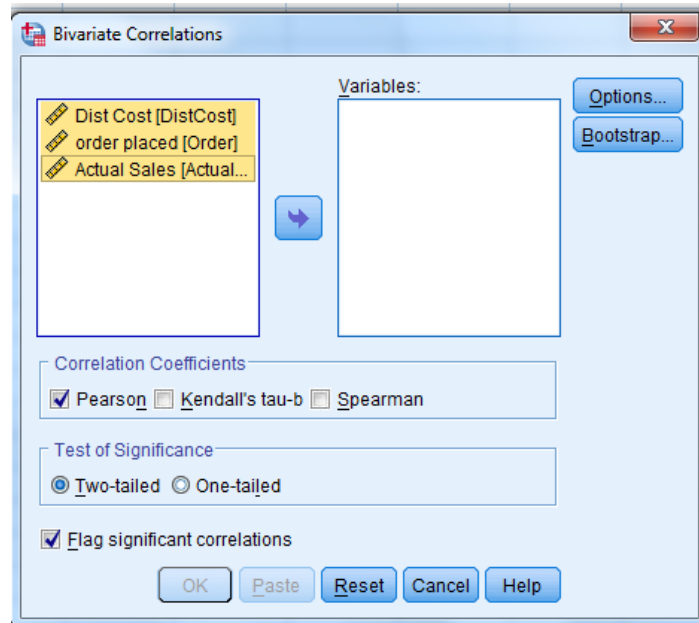




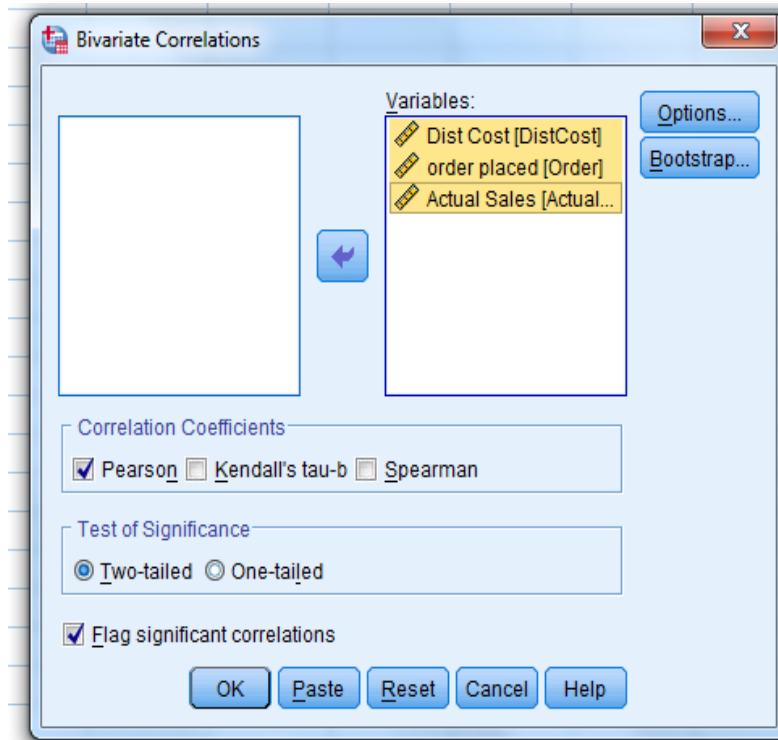
	DistCost	Order	ActualSales	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14
1	52.95	4015	386														
2	71.66	3802	446														
3	85.58	5309	512														
4	63.69	4262	401														
5	72.81	4296	457														
6	68.44	4097	458														
7	52.46	3213	301														
8	70.27	4809	484														
9	82.03	5237	517														
10	74.39	4732	503														
11	70.84	4413	535														
12	54.08	2921	353														
13	62.98	3977	372														
14	72.30	4428	372														
15	58.99	3964	408														
16	79.38	4582	491														
17	94.44	5582	527														
18	59.74	3450	444														
19	90.50	5079	623														
20	93.24	5735	596														
21	69.33	4269	463														
22	53.71	3708	389														
23	89.18	5387	547														
24	66.80	4161	416														



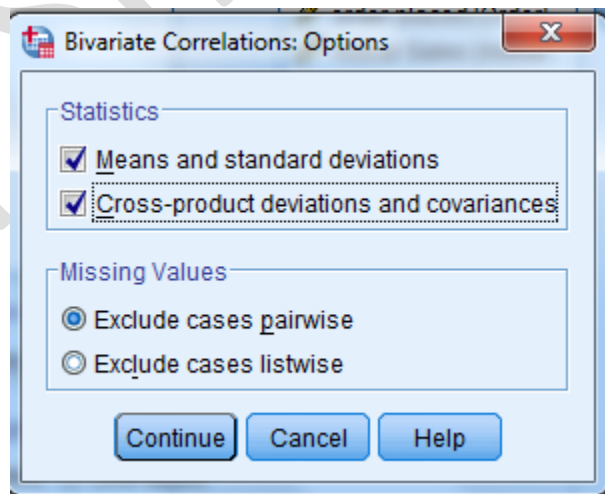
Open the **Bivariate Correlations** dialogue box . following dialogue box will open.



Select the variables and send it in **variables** list box by clicking **right arrow** button. Click **Pearson** check box and **Flag significant correlations** check box as shown in the figure below. See the test of significance is set according to the requirement. In our case it is **two tailed**.



Click **Options...** button to open its sub dialogue box. Select the **Statistics** and **Missing values** option . Click **Continue** to close the sub dialogue box. The previous dialogue box will appear again. Click **OK** to get the output.



Descriptive Statistics

	Mean	Std. Deviation	N
Dist Cost	71.2414	12.93087	24
order placed	4392.83	737.220	24
Actual Sales	458.33	78.971	24

Correlations

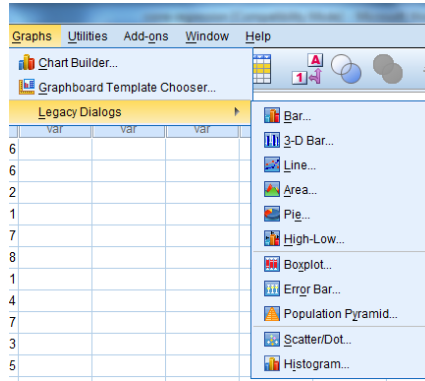
		Dist Cost	order placed	Actual Sales
Dist Cost	Pearson Correlation	1	.918**	.871**
	Sig. (2-tailed)		.000	.000
	Sum of Squares and Cross-products	3845.770	201185.604	20451.133
	Covariance	167.207	8747.200	889.180
	N	24	24	24
order placed	Pearson Correlation	.918**	1	.827**
	Sig. (2-tailed)	.000		.000
	Sum of Squares and Cross-products	201185.604	12500337.333	1107733.333
	Covariance	8747.200	543492.928	48162.319
	N	24	24	24
Actual Sales	Pearson Correlation	.871**	.827**	1
	Sig. (2-tailed)	.000	.000	
	Sum of Squares and Cross-products	20451.133	1107733.333	143439.333
	Covariance	889.180	48162.319	6236.493
	N	24	24	24

** . Correlation is significant at the 0.01 level (2-tailed).

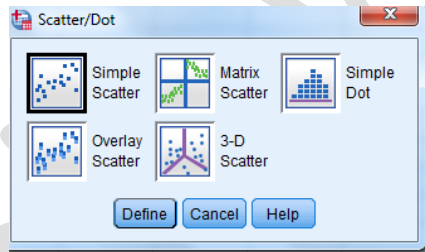
The bivariate correlation is undertaken between the respondents. it was hypothesized that a relationship exists between Distribution cost and order placed, distribution cost and actual sales. The above result shows that there exist the correlation. Distribution cost and Orders placed the correlation ratio ' r ' is 0.918, which shows that both are highly correlated. Distribution cost and Actual Sales correlation ratio ' r ' is 0.871, which shows that both are highly correlated. And $p < 0.05$ in both the cases.

➤ CONSTRUCTING A SCATTER PLOT

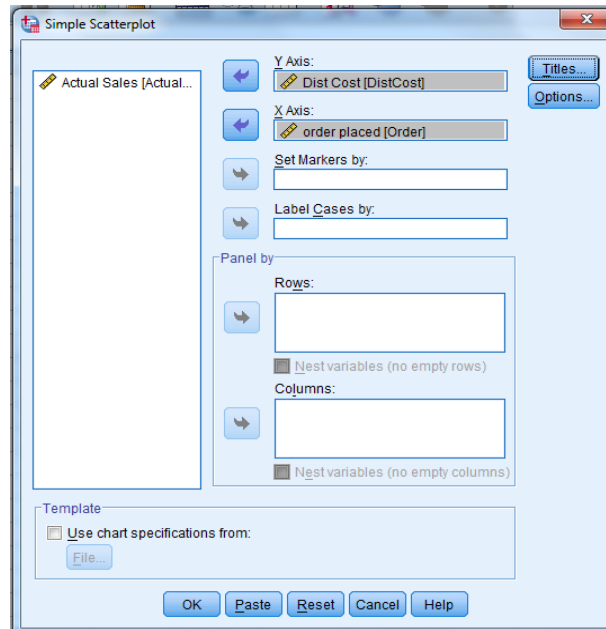
Select **Graph**→**Legacy Dialogs**→**Scatter/Dot**.....This will open **Scatter/Dot** dialogue box



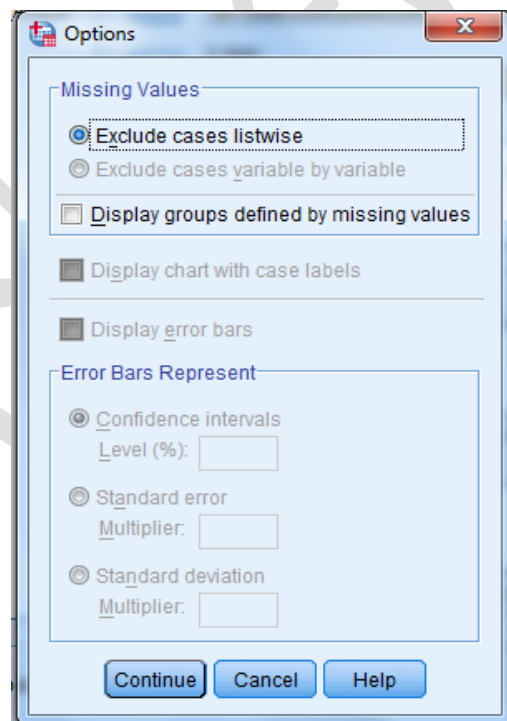
Click **Simple Scatter** button and click **Define** to open **Simple Scatter** plot dialogue box.

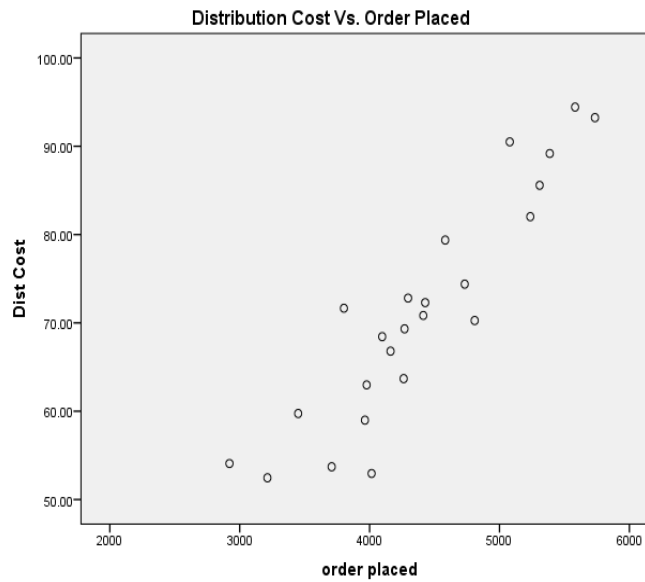


Select the variables of X axis and Y axis .



Click **Options...** button to open **Options** sub dialogue box. Select the options accordingly. Click **continue** to get the output.





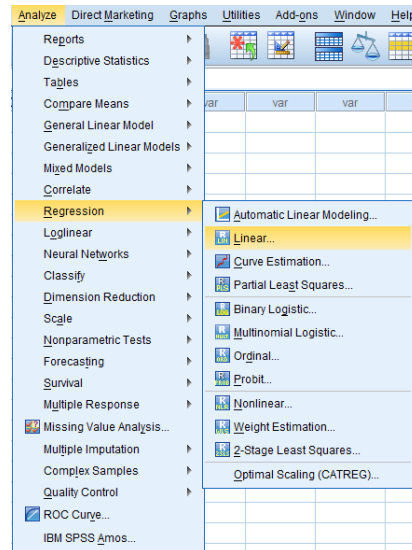
➤ REGRESSION

In Linear regression, we are finding the equation of a straight line that best fits the observation. We can use the line to predict a value of Y for any value of X. This line will have an intercept term and a slope coefficient and is represented as:

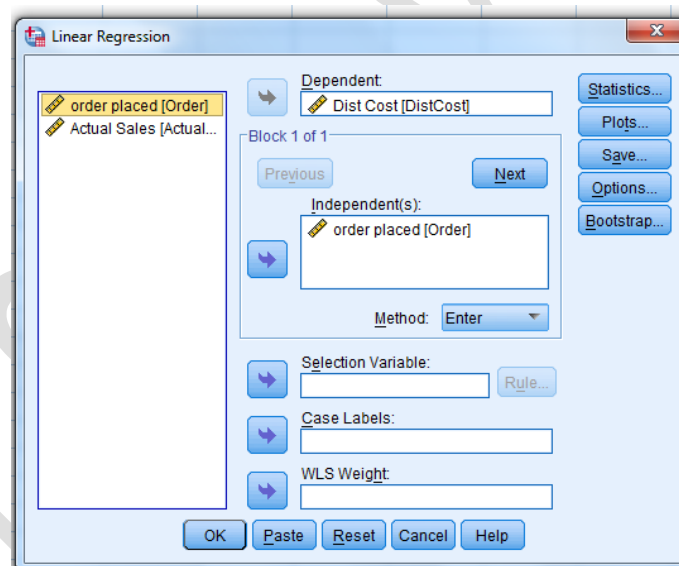
$$Y=a+bX$$

Lets use the same example for regression also.

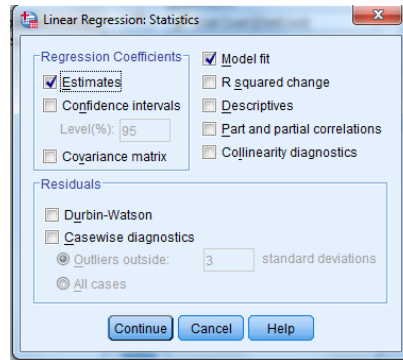
Click **Analyze**→ **Regression**→**Linear**..... this will open **Linear Regression** Dialouge box.



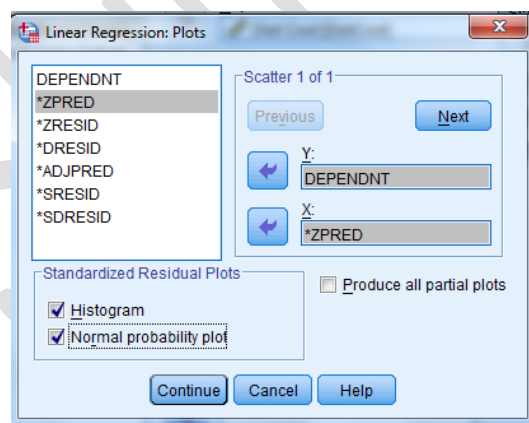
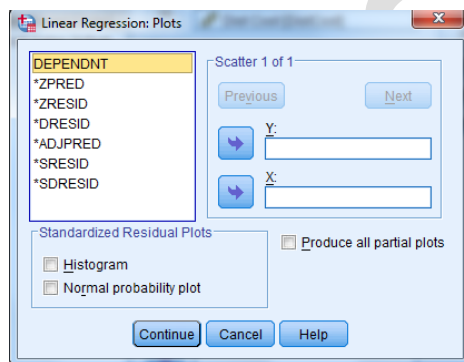
Select the dependent variable as 'Distribution cost' and send it to Dependent list box. Select 'orders placed' as independent variable and send it to independent variable.



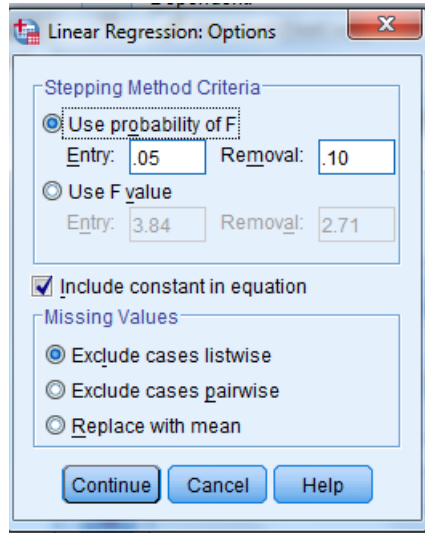
Click **Statistics...** button to open its sub dialogue box. Select the required statistics as shown in the figure below. Click **Continue** to close sub dialogue box. Previous dialogue box will reappear.



Click plots....to open Plots sub dialogue box. Select the plots as required by you. Click continue to close the sub dialogue box. Previous dialogue box will reappear.



Click **Options**....to open **Options** sub dialogue box. Select the options as required by you. Click **Continue** to close the sub dialogue box. Previous dialogue box will reappear. Click **OK** to get the output.



Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	order placed ^b	.	Enter

a. Dependent Variable: Dist Cost

b. All requested variables entered.

The table above shows the variable entered or removed in the model.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.918 ^a	.842	.835	5.25619

a. Predictors: (Constant), order placed

b. Dependent Variable: Dist Cost

The table above shows the model summary.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3237.964	1	3237.964	117.201	.000 ^b
	Residual	607.806	22	27.628		
	Total	3845.770	23			

a. Dependent Variable: Dist Cost

b. Predictors: (Constant), order placed

The table above shows the ANOVA values with sum of squares, degree of freedom, mean square, F statistic and level of significance.

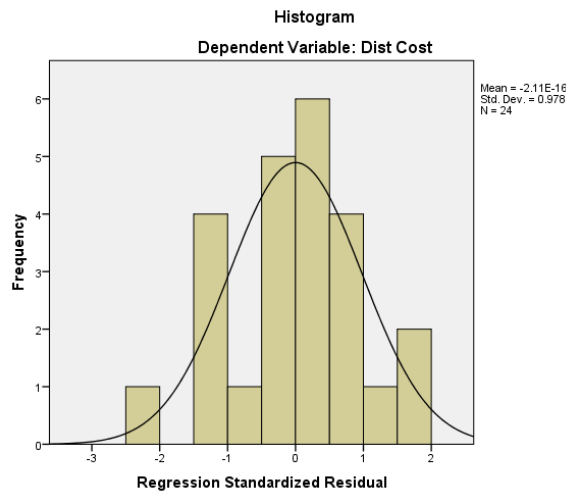
Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.541	6.618		.082	.936
	order placed	.016	.001	.918	10.826	.000

a. Dependent Variable: Dist Cost

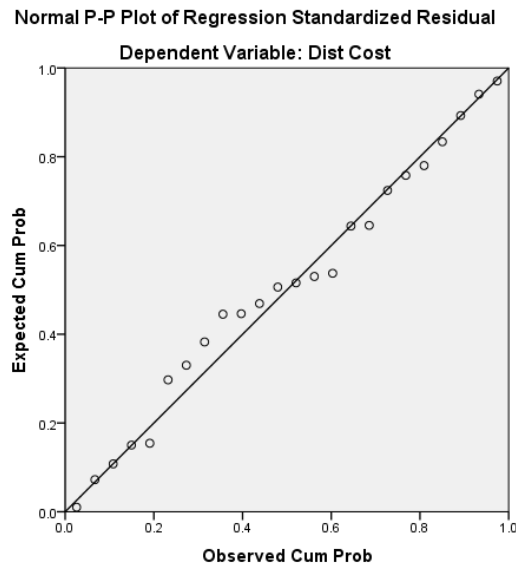
Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	47.5531	92.8428	71.2414	11.86512	24
Residual	-12.21041	9.92770	.00000	5.14066	24
Std. Predicted Value	-1.996	1.821	.000	1.000	24
Std. Residual	-2.323	1.889	.000	.978	24

a. Dependent Variable: Dist Cost

The table above shows the residual statistics with predicted value, residual, standard predicted values and standard residual values.



The figure above shows the Histogram of dependent variable, Distribution cost.



The figure above values shows the Normal P-P plot of Regression Standard Residual values. The above all steps can be repeated, by taking 'Distribution cost ' as dependent variable and 'Actual Sales' as the other independent variable.

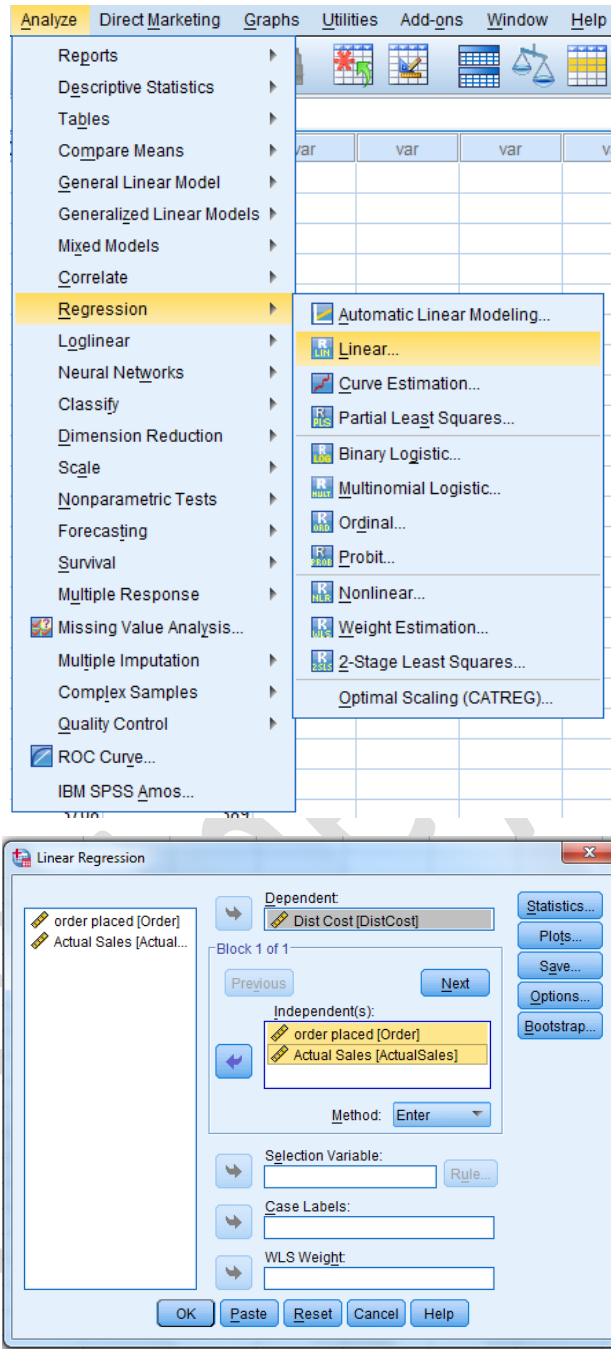
➤ **MULTIPLE REGRESSION MODEL**

Multiple regression involves one dependent variable and two or more independent variables. The multiple regression equation is represented as:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_n X_n$$

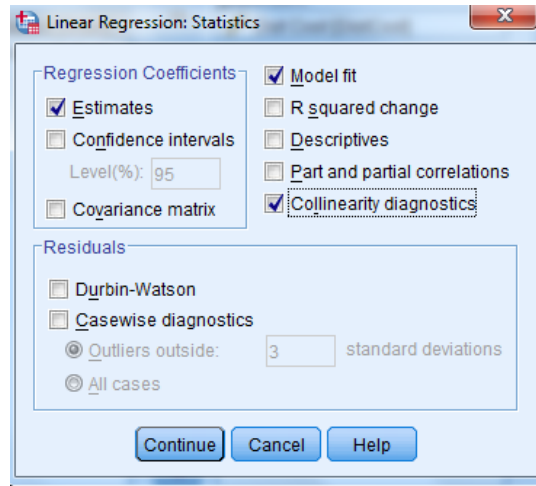
Lets use the same example for regression also.

Click **Analyze** → **Regression** → **Linear**..... this will open **Linear Regression** Dialouge box.



Take 'Distribution cost' as dependent variable and 'orders placed' and 'Actual sales' as two independent variables.

Click statistics button and check in front of collinearity diagnostics.



Select Continue and then click on Plots to request standardized residual plots and also to request scatter diagrams. Select histogram and normal distribution plot of the standardized residuals.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Actual Sales, order placed ^b	.	Enter

a. Dependent Variable: Dist Cost

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.939 ^a	.881	.870	4.65918

a. Predictors: (Constant), Actual Sales, order placed

Value of R^2 is 0.881 and it indicates 88.1 % of variation.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3389.902	2	1694.951	78.080	.000 ^b
	Residual	455.868	21	21.708		
	Total	3845.770	23			

a. Dependent Variable: Dist Cost

b. Predictors: (Constant), Actual Sales, order placed

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-3.459	6.058		-.571	.574
1 order placed	.011	.002	.625	4.673	.000
Actual Sales	.058	.022	.354	2.646	.015

a. Dependent Variable: Dist Cost

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	order placed	Actual Sales
1	1	2.979	1.000	.00	.00	.00
1	2	.017	13.413	.99	.07	.10
	3	.005	25.302	.00	.93	.90

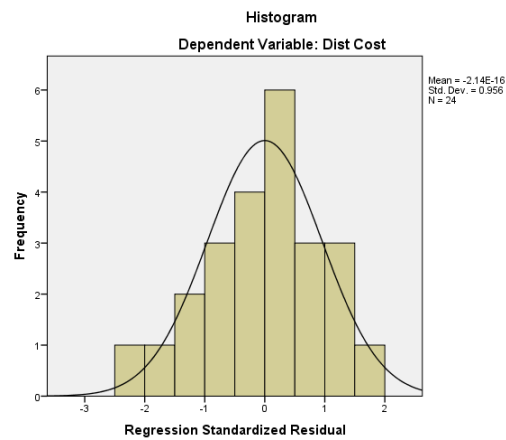
a. Dependent Variable: Dist Cost

The collinearity diagnostics indicate a low degree of overlap between the predictors. If two predictor variables were orthogonal (uncorrelated), the variance inflation factor for each would be 1. Thus we can say that there is no problem of collinearity in this case.

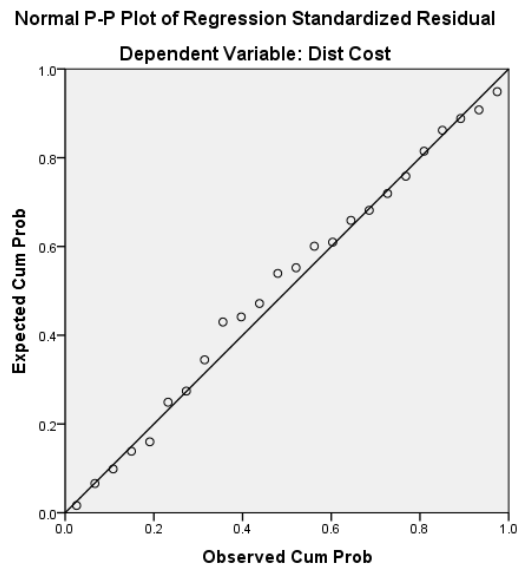
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	49.0069	93.9277	71.2414	12.14031	24
Residual	-9.95979	7.60913	.00000	4.45200	24
Std. Predicted Value	-1.831	1.869	.000	1.000	24
Std. Residual	-2.138	1.633	.000	.956	24

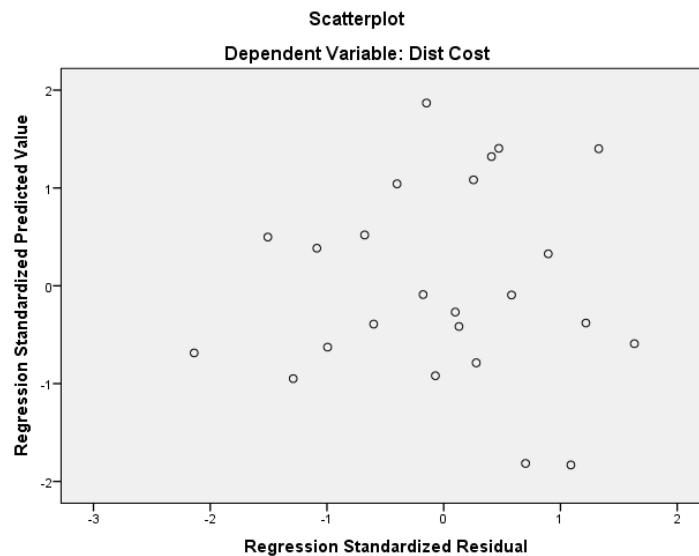
a. Dependent Variable: Dist Cost



The histogram of the standardized residual shows that the departure from normality is not too severe.



The normal P-P plot indicates some departure from normality and may suggest a curvilinear relationship between the predictors and the criterion.



The plot of the standardized predicted values against the standardized residuals indicates a large degree of heteroscedasticity.

➤ **WORKING EXAMPLE:**

A survey is conducted in order to find the preference criteria of a popular brand of ice cream. Responses were collected in 7 point Likert scale. Preferences were measured with respect to its nutrition value, flavor and presentation.

preference	Nutrition value	flavour	presentation
7	5	6	5
6	4	6	6
5	5	7	4
6	6	7	5
4	3	2	4
2	2	1	2
3	3	2	3
6	5	6	5
7	7	7	6
5	6	5	4
4	4	3	2
3	6	2	3
1	1	2	1
2	2	3	1
4	5	4	3
4	4	5	4
3	2	1	3
6	7	5	4
6	5	5	6

7	6	4	5
7	5	6	6
3	2	3	4
2	2	1	1
5	5	4	4
6	5	6	4
7	6	5	7
2	1	1	2
4	2	1	2
6	4	5	5
7	6	5	5
6	3	6	5
5	4	4	4
2	1	1	2
3	2	1	1
4	3	2	2
6	5	7	6
7	6	7	6
7	5	6	7
4	3	2	3
5	3	4	3

Find whether there is any relation between preferences and its nutrition value, flavor and presentation. Perform Single regression and multiple regression analysis and interpret your answers.

Chapter 6

ANOVA

ANOVA is used when you want to compare means of more than two groups or level of an independent variable. ANOVA test comes under parametric and it satisfies the following assumptions.

- *Data is numerical data representing **normally distributed populations**.*
- *The variance of groups are similar.*
- *Sizes of group are similar.*
- *The groups should be independent.*

ANOVA tests the null hypothesis that means of all the groups being compared are equal and produces a statistics called F. If the mean of all the groups tested by ANOVA are equal, then null hypothesis is proved. In case of the rejection of null hypothesis, we perform Brown-Forsythe and Welch test in SPSS.

Homogeneity of variance :If significance value is less than 0.05, variances of groups are significantly different.

Brown-Forsythe and Welch test : :If significance value is less than 0.05, reject null hypothesis.

ANOVA: :If significance value is less than 0.05, reject null hypothesis.

Post-Hoc analysis are used to protect against type I errors, given that all the possible comparisons are going to be made. Some of the Post-Hoc tests are:

- **Scheffe test**-allows every possible comparison to be made but is tough on rejecting the null hypothesis.
- **Tukey test/honestly significant difference(HSD)** test-lenient but the types of comparison that can be made are restricted.

➤ **One Way ANOVA:**

WORKING EXAMPLE:

Excellence coaching centre want to compare the competitive exam score of its registered students from 5 different centers namely, Ramnagar, Gandhinagar, SBI square, Azad Nagar and Civil Lines . 10 score based on random sampling from each centre has been taken as sample for study. Note that the sample is independent sample as they belong to the different study centre.

1 =Ramnagar, 2=Gandhinagar, 3= SBI square, 4=Azad Nagar and 5=Civil Lines.

Null Hypothesis: There is no significant difference in scores from different study centers.

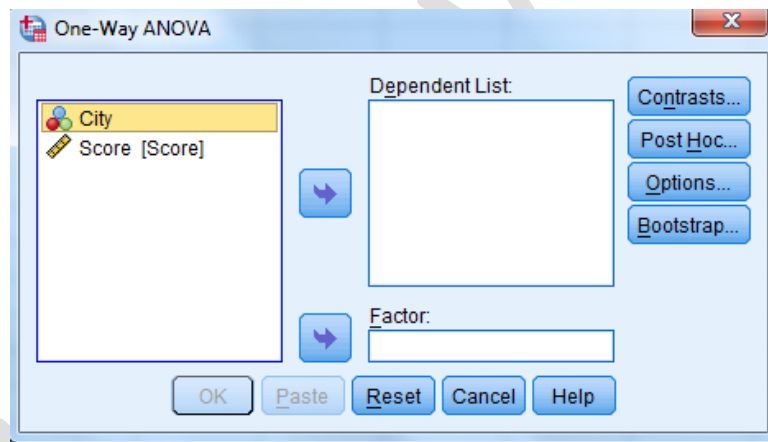
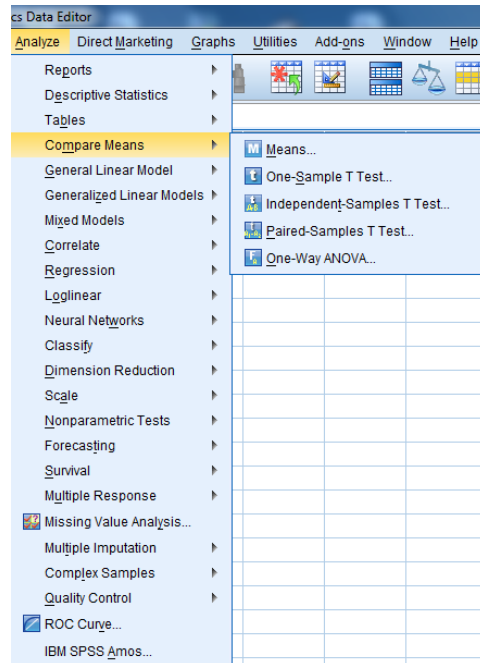
Alternate Hypothesis: There is significant difference in scores from the different study centers.

Sr.no.	City	Score		Sr.No.	City	Score
1.	1	400		26.	3	385
2.	1	450		27.	3	358
3.	1	499		28.	3	299
4.	1	480		29.	3	298
5.	1	495		30.	3	389
6.	1	300		31.	4	450
7.	1	350		32.	4	400
8.	1	356		33.	4	450
9.	1	269		34.	4	428
10.	1	298		35.	4	398
11.	2	389		36.	4	359
12.	2	398		37.	4	360
13.	2	399		38.	4	302
14.	2	599		39.	4	310
15.	2	598		40.	4	295
16.	2	457		41.	5	299
17.	2	498		42.	5	599
18.	2	400		43.	5	466
19.	2	300		44.	5	591
20.	2	369		45.	5	502
21.	3	488		46.	5	598
22.	3	469		47.	5	548

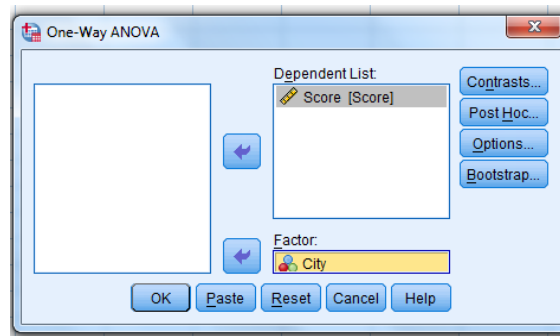
23.	3	425		48.	5	459
24.	3	450		49.	5	489
25.	3	399		50.	5	499

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	City	Numeric	11	0	City	None	None	11	Right	Nominal	Input
2	Score	Numeric	11	0	Score	None	None	11	Right	Scale	Input
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

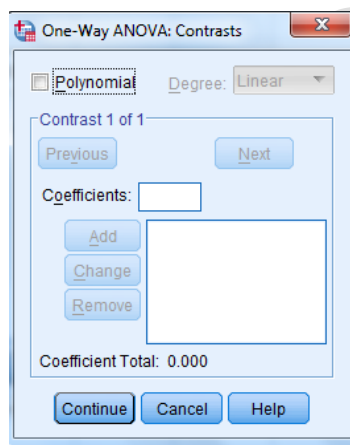
	City	Score	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	1	300														
2	1	350														
3	1	399														
4	1	380														
5	1	395														
6	1	200														
7	1	250														
8	1	256														
9	1	169														
10	1	198														
11	2	198														
12	2	199														
13	2	499														
14	2	366														
15	2	491														
16	2	402														
17	2	498														
18	2	448														
19	2	359														
20	2	399														
21	3	289														
22	3	298														
23	3	499														
24	3	357														



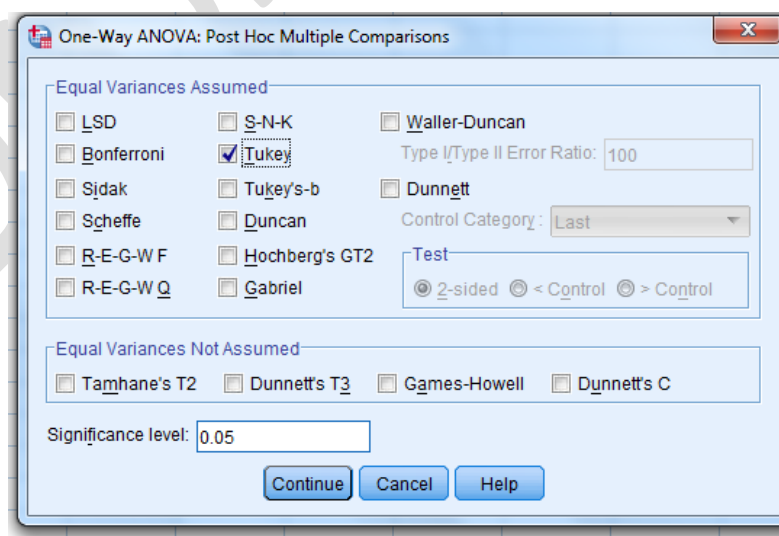
Transfer '**Score**' to dependent list and '**City**' to Factor.



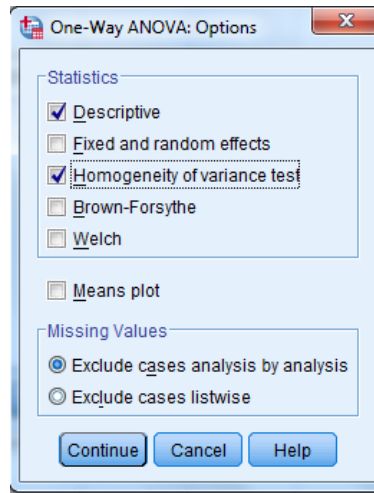
Click contrasts.....push button. Contrasts sub dialogue box will open. See all the settings remain as shown in the figure below. Click continue to close and come back to One-Way ANOVA dialogue box.



Click Post-Hoc. Click Tukey test and click continue. Come back to One-way ANOVA dialogue box. Also note significant level in this sub dialogue box is 0.05, which can be changed according to the need.



Click options. Select Descriptive and Homogeneity of variance test and see all the settings remain as shown in the figure below. Click continue to close this sub dialogue box and come back to One way ANOVA dialogue box. Click OK to see the output.



OUTPUT:

Descriptives

Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	389.70	87.347	27.622	327.22	452.18	269	499
2	10	440.70	97.855	30.944	370.70	510.70	300	599
3	10	396.00	65.137	20.598	349.40	442.60	298	488
4	10	375.20	59.299	18.752	332.78	417.62	295	450
5	10	505.00	90.126	28.500	440.53	569.47	299	599
Total	50	421.32	91.389	12.924	395.35	447.29	269	599

Test of Homogeneity of Variances

Score

Levene Statistic	df1	df2	Sig.
.868	4	45	.491

Since homogeneity of variance should not be there for conducting ANOVA tests, which is one of the assumption of ANOVA, we see that Levene's test shows that homogeneity of variance is not significant ($p > 0.05$). the population variance for each group are approximately equal.

ANOVA

Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	111459.080	4	27864.770	4.211	.006
Within Groups	297783.800	45	6617.418		
Total	409242.880	49			

Table shows the F test value along with the degrees of freedom and significance is 0.006, you can reject the null hypothesis and accept the alternate hypothesis. That is there is significant difference in scores from the different study centers. $F=4.211$, $p < 0.05$.

Multiple Comparisons

Dependent Variable: Score

Tukey HSD

(I) City	(J) City	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Ramnagar	Gandhinagar	-51.000	36.380	.630	-154.37	52.37
	SBI square	-6.300	36.380	1.000	-109.67	97.07
	Azad Nagar	14.500	36.380	.994	-88.87	117.87
	Civil Lines	-115.300*	36.380	.022	-218.67	-11.93
Gandhinagar	Ramnagar	51.000	36.380	.630	-52.37	154.37
	SBI square	44.700	36.380	.735	-58.67	148.07
	Azad Nagar	65.500	36.380	.386	-37.87	168.87
	Civil Lines	-64.300	36.380	.405	-167.67	39.07
SBI square	Ramnagar	6.300	36.380	1.000	-97.07	109.67
	Gandhinagar	-44.700	36.380	.735	-148.07	58.67
	Azad Nagar	20.800	36.380	.979	-82.57	124.17
	Civil Lines	-109.000*	36.380	.034	-212.37	-5.63
Azad Nagar	Ramnagar	-14.500	36.380	.994	-117.87	88.87
	Gandhinagar	-65.500	36.380	.386	-168.87	37.87
	SBI square	-20.800	36.380	.979	-124.17	82.57
	Civil Lines	-129.800*	36.380	.007	-233.17	-26.43
Civil Lines	Ramnagar	115.300*	36.380	.022	11.93	218.67
	Gandhinagar	64.300	36.380	.405	-39.07	167.67
	SBI square	109.000*	36.380	.034	5.63	212.37

	Azad Nagar	129.800*	36.380	.007	26.43	233.17
*. The mean difference is significant at the 0.05 level.						

Using Tukey HSD further, we can conclude that Gandhinagar and Civil Lines have significant difference in their scores. This can be concluded from figure above and figure below.

Score

Tukey HSD

City	N	Subset for alpha = 0.05	
		1	2
Azad Nagar	10	375.20	
Ramnagar	10	389.70	
SBI square	10	396.00	
Gandhinagar	10	440.70	440.70
Civil Lines	10		505.00
Sig.		.386	.405

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.000.

➤ TWO WAY ANOVA

Two way ANOVA is similar to one way ANOVA , except in this case additional independent variable is introduced and each independent variable includes two or more variants.

WORKING EXAMPLE:

Pinnacle Marketing Research firm wants to conduct a study to find whether Sales of the respondent depends upon territory and qualification. 13 sales people were considered from each territory and each respondent can select three education levels.

Territory : Gandhinagar=1,Cosmos Town=2 and Sadar=3.

Qualification: Under graduate=1, Graduate=2 and post graduate=3.

A total of $3 \times 3 \times 13 = 117$ responses were collected.

Research Questions:

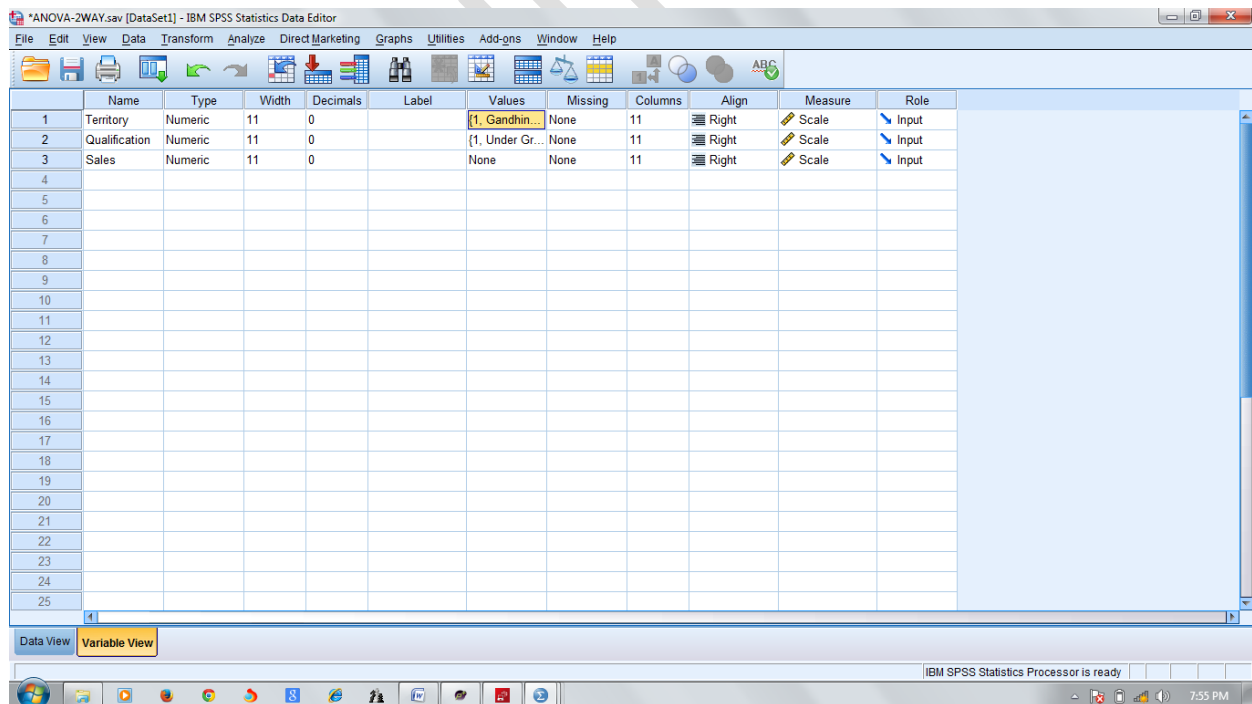
Whether territory influences sales?

Education influences sales?

Sr.No.	Territory	Qualification	Sales		Sr.No.	Territory	Qualification	Sales
1.	1	1	20		21.	3	3	75

2.	1	3	50	22.	3	2	30
3.	1	2	45	23.	3	1	35
4.	1	2	50	24.	3	2	43
5.	1	3	64	25.	3	2	40
6.	1	1	35	26.	3	2	40
7.	1	3	54	27.	3	3	55
8.	1	3	60	28.	3	2	30
9.	1	3	55	29.	3	2	75
10.	1	2	35	30.	3	3	60
11.	2	3	65	31.	2	3	50
12.	2	3	65	32.	2	1	20
13.	2	2	60	33.	2	3	75
14.	2	2	50	34.	3	1	21
15.	2	3	69	35.	3	2	25
16.	2	1	25	36.	3	3	60
17.	2	3	75	37.	1	3	55
18.	2	3	66	38.	1	2	35
19.	2	3	65	39.	1	1	15
20.	2	2	75				

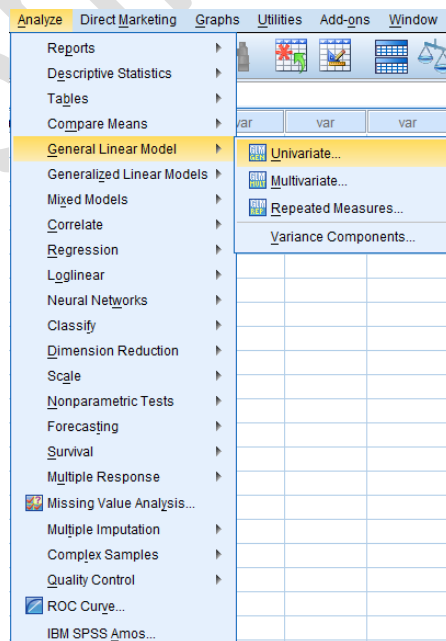
Create a SPSS file as below:

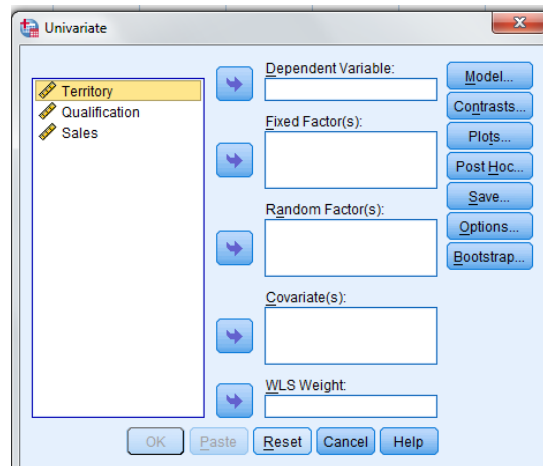


	Territory	Qualification	Sales	var	var	var	var	var	var	var	var	var	var	var	var
1	1	1	20												
2	1	3	50												
3	1	2	45												
4	1	2	50												
5	1	3	64												
6	1	1	35												
7	1	3	54												
8	1	3	60												
9	1	3	55												
10	1	2	35												
11	2	3	65												
12	2	3	65												
13	2	2	60												
14	2	2	50												
15	2	3	69												
16	2	1	25												
17	2	3	75												
18	2	3	66												
19	2	3	65												
20	2	2	75												
21	3	3	75												
22	3	2	30												
23	3	1	35												
24	3	2	43												

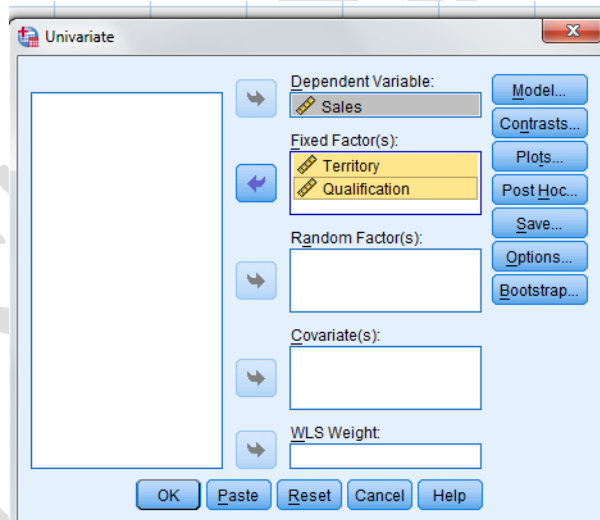
Click **Analyze**→ **General Linear Model**→ **Univariate...**

This will open **Univariate** dialogue box.

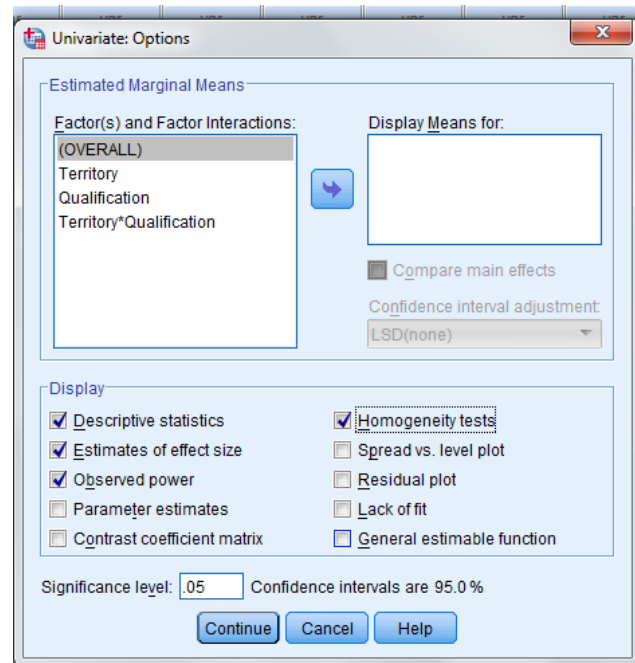




Choose **Sales** and send it in **dependent variable** box. Similarly, choose **territory** and **qualification** and send to fixed factors list box.



Click **Options** button to open its sub dialogue box.



Click **Descriptive Statistics, Estimates of effect size, Observed power and Homogeneity tests** check boxes in the **Display** box and click **continue**. Previous dialouge box will open. Click OK to see the output.

Between-Subjects Factors

		Value Label	N
Territory	1	Gandhinagar	13
	2	Cosmos Town	13
	3	Sadar	13
Qualification	1	Under Graduate	7
	2	Graduate	14
	3	Post graduate	18

Descriptive Statistics

Dependent Variable: Sales

Territory	Qualification	Mean	Std. Deviation	N
Gandhinagar	Under Graduate	23.33	10.408	3
	Graduate	41.25	7.500	4
	Post graduate	56.33	4.926	6
	Total	44.08	15.086	13

Cosmos Town	Under Graduate	22.50	3.536	2
	Graduate	61.67	12.583	3
	Post graduate	66.25	7.833	8
	Total	58.46	17.938	13
Sadar	Under Graduate	28.00	9.899	2
	Graduate	40.43	16.622	7
	Post graduate	62.50	8.660	4
	Total	45.31	18.094	13
Total	Under Graduate	24.43	7.786	7
	Graduate	45.21	15.636	14
	Post graduate	62.11	8.094	18
	Total	49.28	17.899	39

Levene's Test of Equality of Error Variances^a

Dependent Variable: Sales

F	df1	df2	Sig.
.662	8	30	.720

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Territory + Qualification + Territory * Qualification

The Levene's test shows that homogeneity of assumption has not been violated, since $p > 0.05$, i.e. main effects of territory and qualification are not significant. Therefore neither territory nor qualification influences the sales by respondents.

Tests of Between-Subjects Effects

Dependent Variable: Sales

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	8927.766 ^a	8	1115.971	10.314	.000	.733	82.508	1.000
Intercept	62207.917	1	62207.917	574.911	.000	.950	574.911	1.000
Territory	515.992	2	257.996	2.384	.109	.137	4.769	.444
Qualification	6683.744	2	3341.872	30.885	.000	.673	61.770	1.000
Territory * Qualification	635.915	4	158.979	1.469	.236	.164	5.877	.399
Error	3246.131	30	108.204					
Total	106894.000	39						
Corrected Total	12173.897	38						

a. R Squared = .733 (Adjusted R Squared = .662)

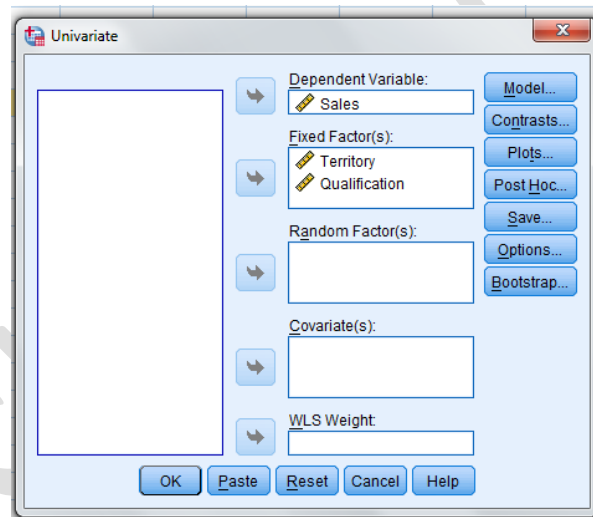
b. Computed using alpha = .05

The above table shows that, qualification level shows significant effect on sales with $p < 0.05$ $F=30.885$, $p < 0.05$. The output also shows that neither territory nor territory and qualification combined have significant interaction effect on sales with both having $p > 0.05$. We can also say that, territory does not influence sales, $F=2.384$, $p > 0.05$. Also territory and qualification both does not collectively influence sales, $F=158.979$, $p > 0.05$.

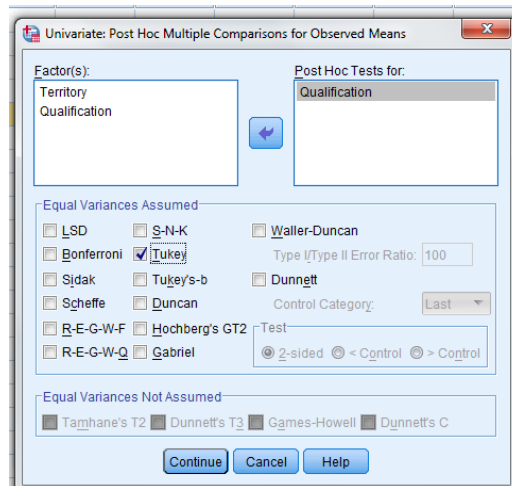
➤ POST-HOC ANALYSIS

Click Analyze → General Linear → univariate

This will open Univariate dialogue box.



Click **Post-Hoc** for opening **Post-Hoc** sub dialogue box. Select **Qualification** and send it in **Post Hoc Tests** for list box. Click **Tukey** check box and click continue. Click **OK** to see the output.



Multiple Comparisons

Dependent Variable: Sales

Tukey HSD

(I) Qualification	(J) Qualification	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Under Graduate	Graduate	-20.79*	4.815	.000	-32.66	-8.91
	Post graduate	-37.68*	4.633	.000	-49.11	-26.26
Graduate	Under Graduate	20.79*	4.815	.000	8.91	32.66
	Post graduate	-16.90*	3.707	.000	-26.04	-7.76
Post graduate	Under Graduate	37.68*	4.633	.000	26.26	49.11
	Graduate	16.90*	3.707	.000	7.76	26.04

Based on observed means.

The error term is Mean Square(Error) = 108.204.

*. The mean difference is significant at the .05 level.

The analysis shows that Post Graduate sales differ significantly with under graduate and graduate sales i.e. Post Graduate and other two education levels have significant different in sales. You can see the same results in the table below also.

Sales

Tukey HSD

Qualification	N	Subset		
		1	2	3
Under Graduate	7	24.43		
Graduate	14		45.21	
Post graduate	18			62.11

Sig.		1.000	1.000	1.000
------	--	-------	-------	-------

Means for groups in homogeneous subsets are displayed.

Based on observed means.

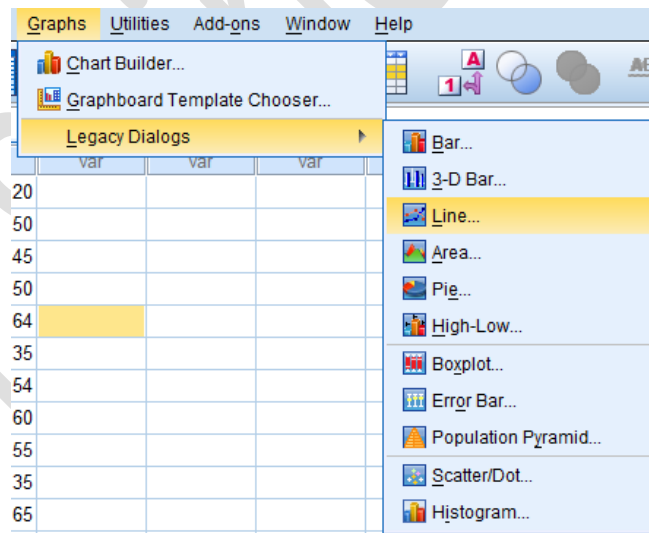
The error term is Mean Square(Error) = 108.204.

- Uses Harmonic Mean Sample Size = 11.118.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

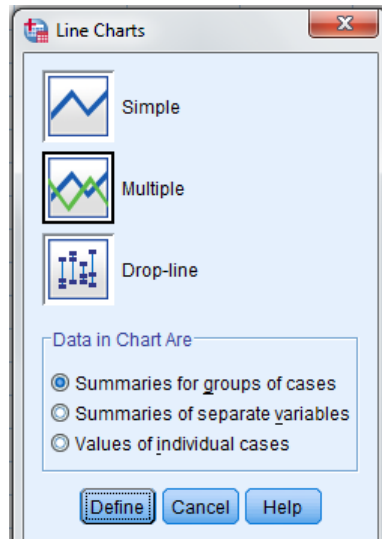
With significant interaction effects, you may see simple effects and comparison based on graphs. The discussion ahead provides ahead provide brief about line graph results.

Click Graphs menu → Legacy Dialogs → Line.....

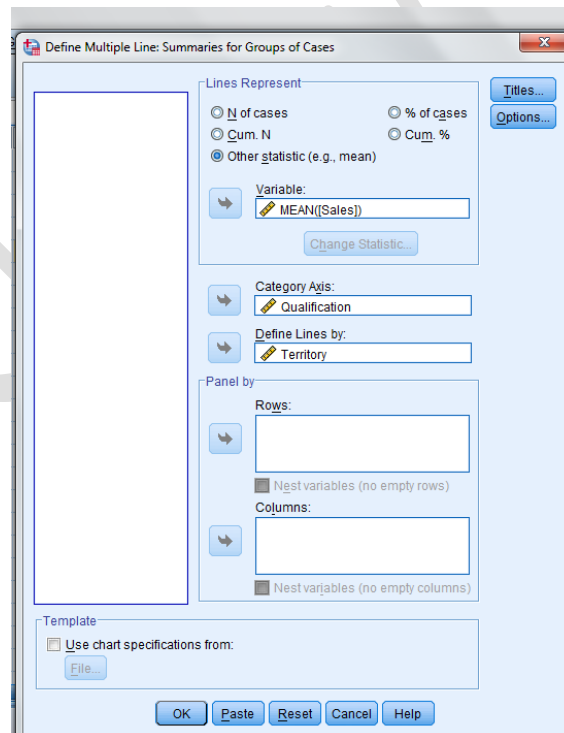
This will open Line Charts dialogue box

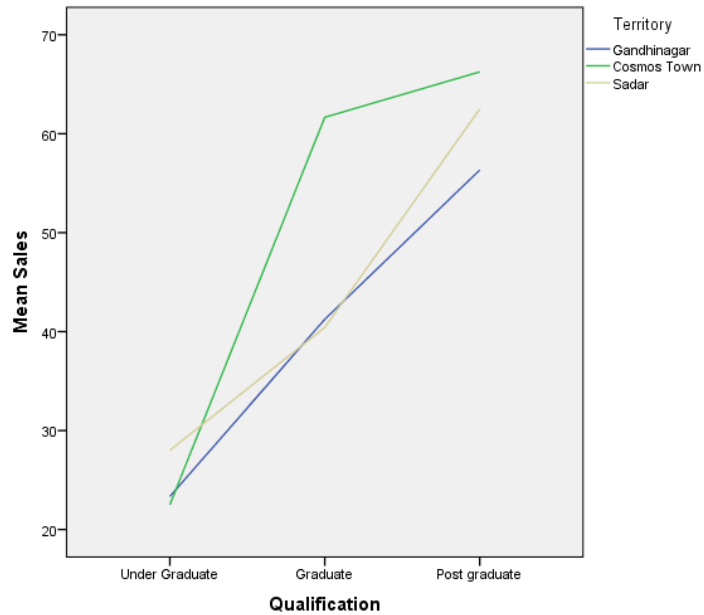


Click Multiple and Define to open Define sub dialogue box.



Select **other Statistics** radio button. By default mean will be shown , when we send **Sales** in box. In **Category Axis** send **Qualification** variable and put **Territory** in **define lines**.





The output above shows that sales increases with increase in education in all the three places.

Chapter 7

Factor Analysis

Factor Analysis is a statistical process used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying factors. The objective is to condense the information contained in a larger number of variable into a smaller set of variables. Factor analysis is a generic name given to a class of multivariate statistical methods whose primary purpose is to define the underlying structure in a data matrix. Broadly speaking it addresses the problem of analyzing the structure of interrelationships (correlations) among a large number of variables. Once these dimensions and the explanation of each variable are determined, the two primary uses for factor analysis-summarization and data reduction can be achieved (Hair et al.,1998)

Step 1: objectives of Factor Analysis

The first step involves

➤ Identifying structure through data summarization

Factor analysis can identify the structure of relationship among either variables or respondents by examining either the correlation between the variables or the correlation between the respondents. The most common type of factor analysis is referred to as R factor analysis. R factor analysis analyses a set of variables to identify the dimension that are latent (not easily observed). But if factor analysis be applied to a correlation matrix of the individual respondents based on their characteristics then it would be referred to as Q factor analysis, which is a method of converting a large number of people into distinctly different groups within a large population.

➤ Data reduction

Factor analysis can also :Identify representative variables from a much larger set of variables for use in subsequent multivariate analyses, Or, Create an entirely new set of variables, much smaller in number, to partially or completely replace the original set of variables for inclusion in subsequent techniques. In both the cases the objective is to retain the nature and character of the original variables, but reduce their number to simplify the subsequent multivariate analysis(Hair et al.,1998)

Step 2:Designing Factor Analysis.

Hair et al., (1998) mentioned that the design of a factor analysis involves three basic decisions:

- Calculation of input data (a correlation matrix) to meet the specified objectives of grouping variables or respondents.

- The design of the study in terms of number of variables, measurement properties of variables, and the types of allowable variables; and
- Sample size necessary, both in absolute terms and as a function of number of variables in the analysis.

Step 3: Assumptions in Factor Analysis :

Sample size : a sample of 100 subjects is acceptable.

Normality : variables should be normally distributed.

Linearity :linearity should be present since correlation are required.

Overall significance of correlation matrix should be checked with Bartlett test of sphericity and KMO measure of Sampling Adequacy.

Step 4: Deriving Factors and Assessing Overall Fit

Hair et.al.(1998) mention that once the variables are specified and the correlation matrix is prepared, then factor analysis can be applied to identify the underlying structure of relationships. In doing so, decisions must be made concerning

- The method of extracting the factors ; common factors analysis versus component analysis)
- The number of factors selected to represent the underlying structure in the data.

Component analysis is used when the objective is to summarize most of the original information in minimum number of factors for prediction purpose. When a large set of variables is factored, the method first extracts the combination of variables explaining the greatest amount of variance and then proceed to combinations that account for smaller amounts of variance.

For extracting the number of factors the latent root criterion has been used. According to Hair et al.,(1998), the latent root criterion is most commonly used technique and its rationale is that any individual factor should account for the variance of at least single variable if it is to be retained for interpretation. Each variable contributes a value of 1 to the total eigenvalue. Only the factors having latent root or eigenvalues greater than 1 are considered significant ;all factors with latent root less than 1 are considered insignificant and are disregarded. Another criteria for the number of factor to extract is the percentage of variance criterion which according to Hair et al.,(1998) is an approach based on achieving a specified cumulative percentage of total variance extracted by successive factors. The purpose is to ensure practical significance for the derived factors by ensuring that they explain at least a specified amount of variance. The scree test is derived by plotting the latent roots against the number of factors in their order of extraction and the shape of the resulting curve is used to evaluate the cutoff point. Therefore, single criterion

named latent root as a guideline for determining how many factors to extract as the first attempt at interpretation and also a scree test to complete determination.

Step 5: Interpreting the Factors

Hair et al.(1998) explain the three steps are involved in the interpretation of the factors and the selection of the final factor solution. The initial unrotated factor matrix is computed to assist in obtaining a preliminary indication of the number of factors extract. The factor matrix contains factor loadings for each variable on each factor. In computing the unrotated factor matrix, the best linear combination of variables is expected. Therefore, the first factor may be viewed as the single best summary of linear relationship exhibited in the data. The second factor is defined as the second-best linear combination of the variables; subject to the constraint it is orthogonal to the first factor. To be orthogonal to the first factor, the second factor must be derived from the variance remaining after the first factor has been extracted.

Factor loading is the means of interpreting the role each variable plays in defining each factor. Factor loadings are the correlation of each variable and the factor. Loadings indicate the degree of correspondence between the variable and the factor, with higher loadings making the variable representative of the factor. The unrotated factor solution may not provide a meaningful pattern of variable loadings. Generally, rotation will be desirable because it simplifies the factor structure and it is usually difficult to determine whether unrotated factors will be meaningful. Therefore the second step employs a rotation method to achieve simpler and theoretically more meaningful factor solutions (Hair et al.,1998). In the third step, we assess the need to specify the factor model owing to;

- Deletion of variables from the analysis
- The desire to employ a different rotational method for interpretation
- The need to extract a different number of factors
- The desire to change from one extraction method to another.

Re-specification of a factor model is accomplished by returning to the extraction stage, extracting factors, and interpreting them again(Hair et al.,1998)

According to Hair et al. (1998) three major orthogonal approaches have been developed

QUARTIMAX: the ultimate goal of a QUARTIMAX rotation is to simplify the rows of a factor matrix ; that is QUATRIMAX focuses on rotating the initial factor so that a variable loads high on one factor and as low as possible on all other factors.

VARIMAX: VARIMAX criterion centres on simplifying columns of the factor matrix. With the VARIMAX rotational approach, the maximum possible simplification is reached if there are only 1s and 0s in a column.

EQUIMAX: the EQUIMAX approach is a compromise between the QUARIMAX and VARIMAX approaches. Rather than concentrating either on simplification of the rows or on simplification of the columns, it tries to accomplish some of each.

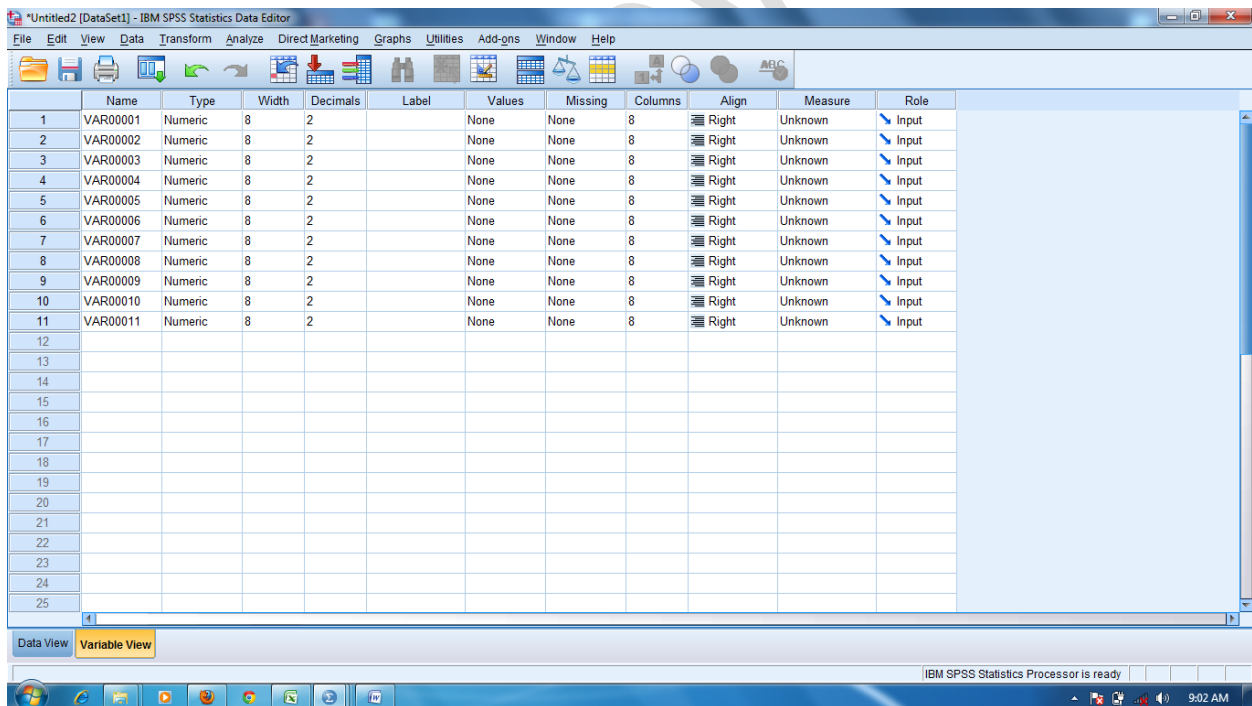
In selecting among the above mentioned rotational methods for the research, there is a point to mention that no specific rules have been developed to guide the researcher in selecting a particular orthogonal or oblique rotational technique. In the working example we intend to utilize the rotational technique VARIMAX as the orthogonal solution for the analysis of existing variable.

➤ WORKING EXAMPLE:

The researcher wants to conduct factor analysis on the factors affecting the employee satisfaction.

Eleven variables were considered for conducting factor analysis. He conducted analysis based on the rating: strongly disagree(1), disagree(2), Neutral (3), Agree (4), Strongly agree (5)

Make a data file based on the eleven factors as shown in the figure below:



Respondents	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10	Var 11
1.	4	4	4	4	4	5	3	5	5	5	5
2.	5	5	5	4	5	5	4	4	5	5	4
3.	3	4	4	5	4	5	4	5	5	4	4

4.	5	5	4	5	4	4	4	4	5	4	4
5.	5	5	4	4	3	5	2	4	4	3	3
6.	4	3	3	3	3	4	3	3	5	3	4
7.	1	3	3	3	2	4	3	4	3	2	3
8.	1	3	3	3	3	4	4	4	2	2	4
9.	1	2	2	2	1	4	4	4	3	3	3
10.	3	4	4	4	4	5	4	4	5	4	4
11.	3	4	4	4	4	2	4	2	5	4	4
12.	4	4	4	4	4	2	4	2	5	4	4
13.	4	5	5	5	5	4	3	3	5	5	4
14.	2	4	2	2	3	2	4	4	4	4	5
15.	3	4	3	4	5	5	5	4	5	5	5
16.	3	1	1	3	3	5	3	3	5	2	2
17.	3	4	2	4	3	2	3	2	2	3	4
18.	1	1	1	1	1	5	1	1	1	1	1
19.	4	4	2	4	4	2	4	2	4	2	4
20.	5	4	4	4	5	5	3	4	5	5	5
21.	3	4	3	3	4	4	4	3	5	4	4
22.	4	4	4	4	3	4	3	4	3	2	3
23.	5	4	4	5	4	4	4	4	2	4	4
24.	4	4	3	3	3	4	4	3	4	3	4
25.	4	3	1	3	2	5	5	4	5	2	4
26.	4	4	4	4	4	5	4	4	5	4	4
27.	3	2	5	5	2	5	4	1	2	1	5
28.	4	3	3	4	3	4	4	3	5	3	4
29.	2	2	1	1	1	5	2	3	5	1	2
30.	1	1	1	1	3	3	3	3	1	1	3
31.	4	4	4	4	5	5	2	4	5	4	2
32.	4	4	3	4	4	4	4	4	4	3	4
33.	3	4	4	4	4	4	1	4	5	4	2
34.	4	4	4	4	4	5	3	5	5	5	5
35.	5	5	5	4	5	5	4	4	5	5	4
36.	3	4	4	5	4	5	4	5	5	4	4
37.	5	5	4	5	4	4	4	4	5	4	4
38.	5	5	4	4	3	5	2	4	4	3	3
39.	4	3	3	3	3	4	3	3	5	3	4
40.	1	3	3	3	2	4	3	4	3	2	3
41.	1	3	3	3	3	4	4	4	2	2	4
42.	1	2	2	2	1	4	4	4	3	3	3

43.	3	4	4	4	4	5	4	4	5	4	4
44.	4	4	4	4	4	2	4	2	5	4	4
45.	4	4	4	4	4	2	4	2	5	4	4
46.	4	5	5	5	5	4	3	3	5	5	4
47.	2	4	2	2	3	2	4	4	4	4	5
48.	3	4	3	4	5	5	5	4	5	5	5
49.	3	1	1	3	3	5	3	3	5	2	2
50.	3	4	2	4	3	2	3	2	2	3	4
51.	1	1	1	1	1	5	1	1	1	1	1
52.	4	4	2	4	4	2	4	2	4	2	4
53.	5	4	4	4	5	5	3	4	5	5	5
54.	3	4	3	3	4	4	4	3	5	4	4
55.	4	4		4	3	4	3	4	3	2	3
56.	5	4	4	5	4	4	4	4	2	4	4
57.	4	4	3	3	3	4	4	3	4	3	4
58.	4	3	1	3	2	5	5	4	5	2	4
59.	4	4	4	4	4	5	4	4	5	4	4
60.	3	2	5	5	2	5	4	1	2	1	5
61.	4	3	3	4	3	4	4	3	5	3	4
62.	2	2	1	1	1	5	2	3	5	1	2
63.	1	1	1	1	3	3	3	3	1	1	3
64.	4	4	4	4	5	5	2	4	5	4	2
65.	4	4	3	4	4	4	4	4	4	3	4
66.	3	4	4	4	4	4	1	4	5	4	2
67.	5	5	5	5	5	4	3	4	5	5	4
68.	2	4	4	3	2	4	2	2	2	2	3
69.	4	5	5	3	4	5	4	4	5	5	4
70.	4	4	4	4	5	4	4	3	5	4	4
71.	4	5	5	4	5	5	4	4	4	4	5
72.	1	1	4	5	4	5	5	1	4	3	4
73.	4	4	4	4	5	4	4	3	5	4	4
74.	3	4	3	5	4	5	5	5	3	4	4
75.	1	4	4	4	2	4	2	3	5	2	2
76.	3	3	3	3	4	3	2	2	4	3	3
77.	4	4	3	4	4	4	4	3	4	3	4
78.	4	4	3	4	4	4	4	3	4	3	4
79.	4	5	4	3	3	4	3	3	4	3	3
80.	4	3	4	4	2	5	5	2	4	2	5
81.	4	5	5	5	4	4	4	5	5	5	5

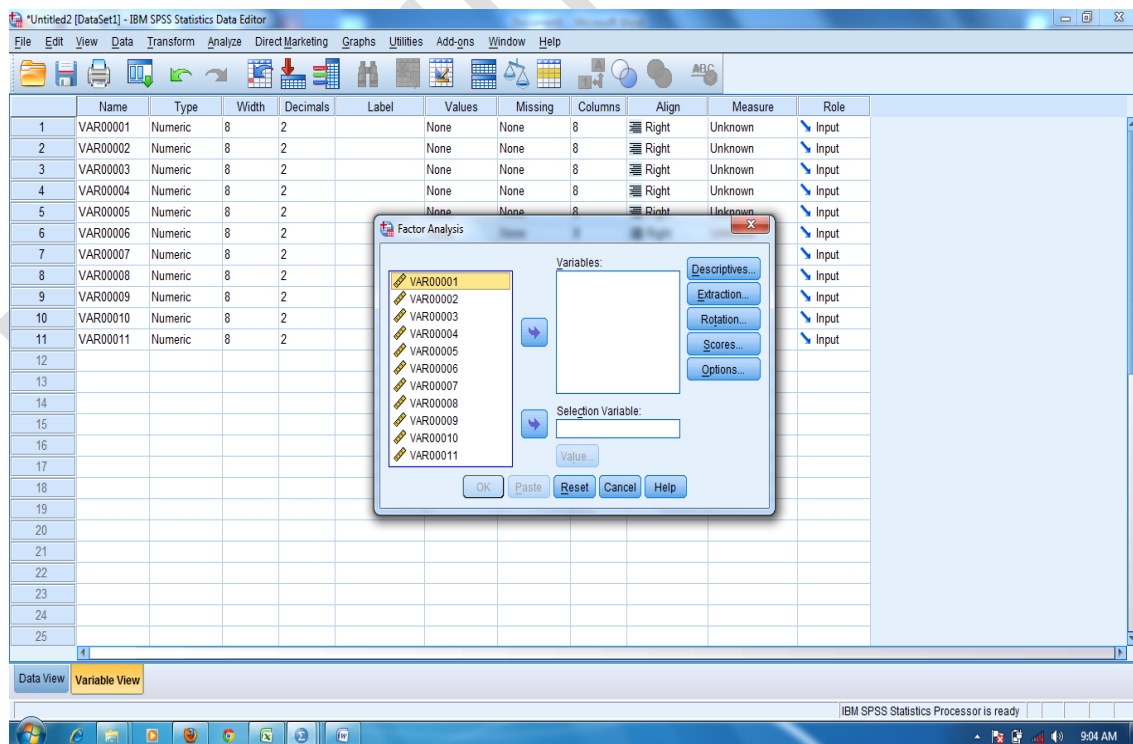
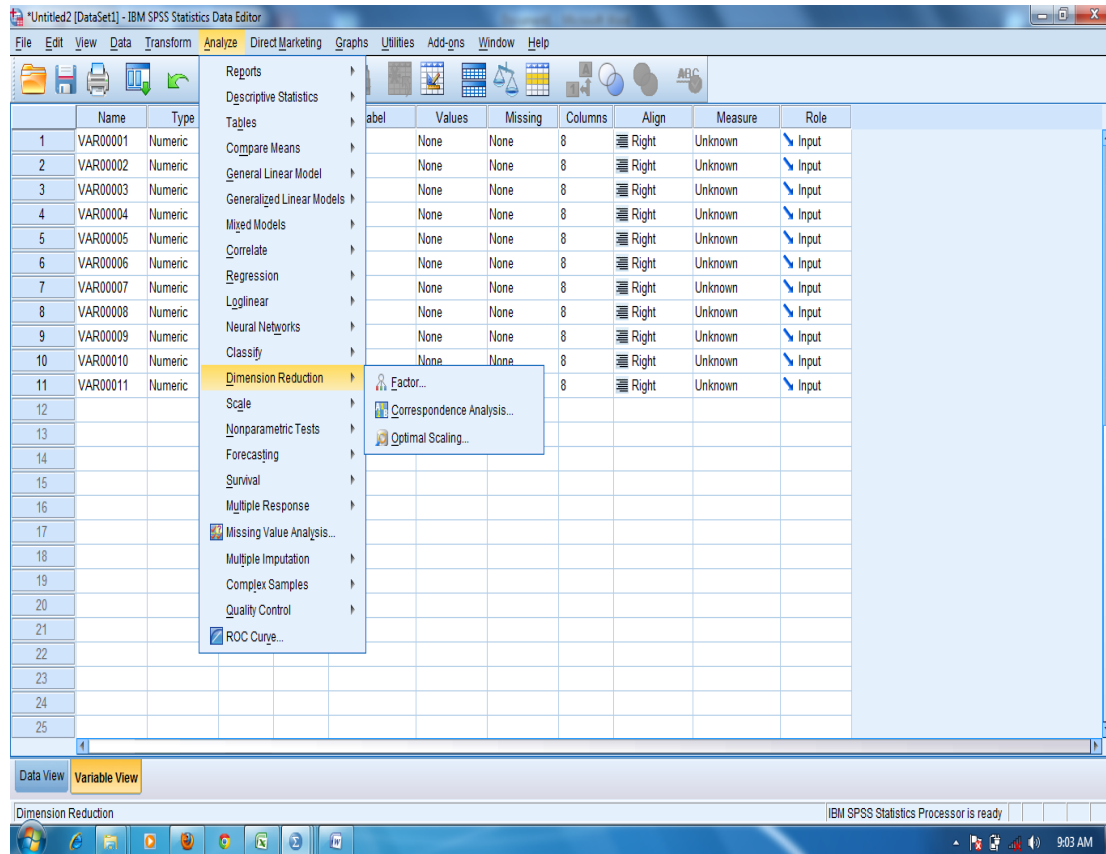
82.	4	4	4	4	4	4	4	4	4	4	4
83.	4	5	5	4	5	5	3	4	4	5	3
84.	2	3	1	5	3	5	2	2	5	3	1
85.	4	4	2	4	3	2	2	3	3	2	2
86.	5	5	5	5	5	5	4	4	4	4	4
87.	4	4	4	4	4	4	4	4	4	4	4
88.	4	4	4	4	4	2	4	2	5	4	3
89.	2	2	2	2	1	4	2	4	4	2	4
90.	1	1	1	1	5	5	5	1	4	1	1
91.	3	4	2	2	4	5	2	5	4	2	2
92.	4	4	4	5	5	5	5	5	5	5	5
93.	4	5	5	5	5	4	5	5	5	4	5
94.	4	4	4	4	4	4	3	4	4	3	3
95.	4	4	3	4	4	4	4	4	3	4	4
96.	1	3	4	4	5	4	4	1	2	4	4
97.	4	4	4	4	4	4	4	3	4	5	5
98.	4	4	4	4	4	4	4	3	4	5	5
99.	4	4	4	4	4	4	4	4	2	1	2
100.	4	5	5	5	5	5	5	5	5	4	4
101.	3	2	1	1	1	4	1	1	3	1	2
102.	1	3	1	2	1	4	3	4	4	1	3
103.	2	1	1	1	1	4	1	2	2	3	2
104.	2	4	4	5	5	4	2	4	4	5	4
105.	2	4	4	5	5	4	2	4	4	5	4
106.	4	2	3	1	2	4	2	3	4	2	1
107.	3	4	5	3	2	4	2	2	5	3	3
108.	4	4	4	4	4	4	4	4	4	5	3
109.	4	4	4	4	4	4	4	4	4	5	3
110.	3	3	3	3	3	4	2	4	5	2	2
111.	4	4	4	4	4	5	3	4	4	4	4
112.	1	1	1	1	1	1	2	1	2	1	1
113.	1	4	1	4	1	4	2	4	4	1	3
114.	5	5	5	5	5	5	5	5	5	5	5
115.	4	4	4	4	4	5	2	4	4	4	3
116.	3	4	5	4	4	4	3	4	4	4	4
117.	2	3	3	5	5	5	5	5	5	3	4
118.	3	5	4	4	3	4	4	4	3	4	4
119.	3	2	2	2	2	2	2	2	2	2	2
120.	4	4	4	5	4	4	4	3	4	4	4

121.	4	5	5	5	5	5	5	5	5	5	5
122.	1	1	1	1	2	1	1	1	1	1	1
123.	4	4	4	3	2	4	3	2	4	4	4
124.	4	2	3	3	4	3	3	4	5	5	3
125.	4	4	4	5	5	5	5	3	5	4	5
126.	2	2	4	2	2	1	3	3	4	2	3
127.	2	1	2	2	2	1	1	1	1	1	1
128.	2	3	3	4	5	4	4	4	4	3	3
129.	4	4	4	5	5	3	4	3	4	4	4
130.	3	3	1	3	1	3	1	1	4	3	4
131.	2	4	4	4	5	5	2	4	4	5	2
132.	1	4	3	3	1	5	3	4	5	1	4
133.	2	3	2	1	3	4	3	3	4	3	3
134.	5	5	5	5	4	5	3	4	3	4	4
135.	3	4	3	3	3	4	3	3	5	2	2
136.	4	3	2	1	3	4	3	4	4	2	3
137.	5	3	2	4	4	4	4	4	5	3	4
138.	2	2	2	2	4	3	5	2	5	3	4
139.	2	2	2	2	4	3	5	2	5	3	4
140.	1	1	1	1	1	1	1	1	1	1	1
141.	4	5	2	5	5	5	5	4	4	5	5
142.	4	4	4	4	4	4	4	4	4	4	4
143.	4	3	2	4	3	5	5	3	5	4	4
144.	2	2	3	1	1	3	2	2	3	4	2
145.	4	5	4	4	5	5	4	5	5	5	
146.	3	4	4	3	3	3	2	3	2	4	2
147.	4	4	4	4	4	4	3	4	3	3	3
148.	1	1	2	2	1	1	2	2	1	1	2
149.	5	5	5	5	5	4	3	4	5	5	4
150.	4	5	4	4	5	4	3	3	5	5	4
151.	3	3	3	2	2	4	4	3	5	2	4
152.	4	3	3	3	4	4	4	4	3	4	4
153.	4	4	3	3	4	4	3	4	4	4	4
154.	2	3		2	4	4	2	2	4	2	2
155.	4	5	4	4	4	4	4	4	4	4	4
156.	4	4	4	4	5	4	5	4	5	5	5
157.	5	5	5	5	5	5	5	5	5	5	5
158.	3	4	5	4	3	4	3	2	4	3	4
159.	3	2	2	3	3	4	2	3	4	2	2

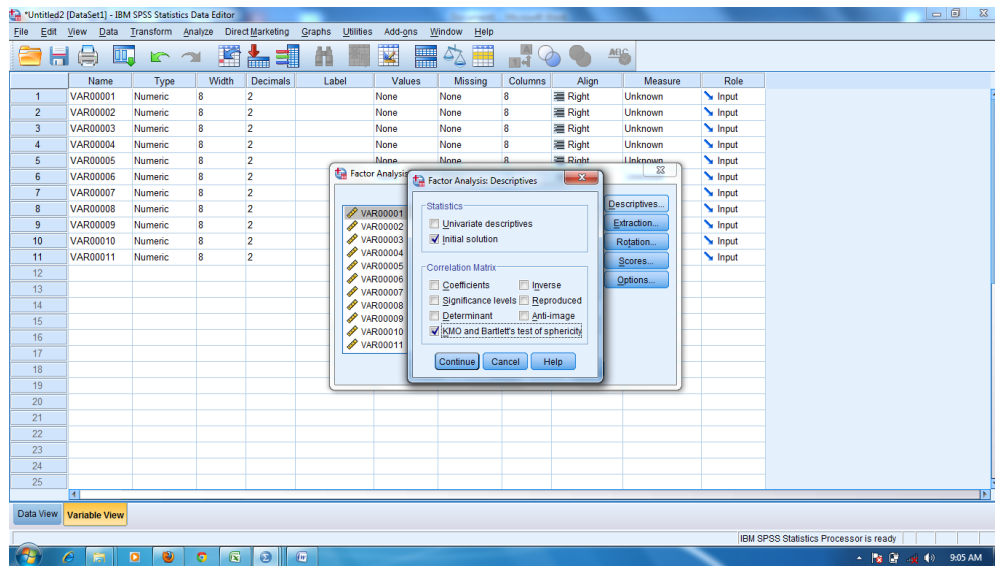
160.	5	5	5	5	4	4	4	4	5	4	4
161.	1	1	1	1	1	1	1	1	1	1	1
162.	1	1	1	1	1	1	1	1	1	1	1
163.	1	1	1	1	1	1	1	1	1	1	1
164.	2	4	4	5	1	4	4	3	4	2	4
165.	3	3	2	2	3	5	4	3	4	3	4
166.	3	4	5	5	4	3	5	5	2	2	5
167.	4	4	4	4	5	5	4	3	4	4	4
168.	4	4	4	4	5	5	4	3	4	4	4
169.	4	4	4	4	5	5	4	3	4	4	4
170.	2	2	2	3	5	5	1	1	5	1	3
171.	2	3	3	3	3	3	2	1	5	2	4
172.	3	2	2	3	4	1	3	2	1	2	3
173.	4	5	5	5	5	4	5	4	4	5	5
174.	5	5	3	4	5	4	5	4	5	3	5
175.	5	3	4	5	3	4	5	5	4	3	5

Click Analyze – Dimension Reduction Factor.....

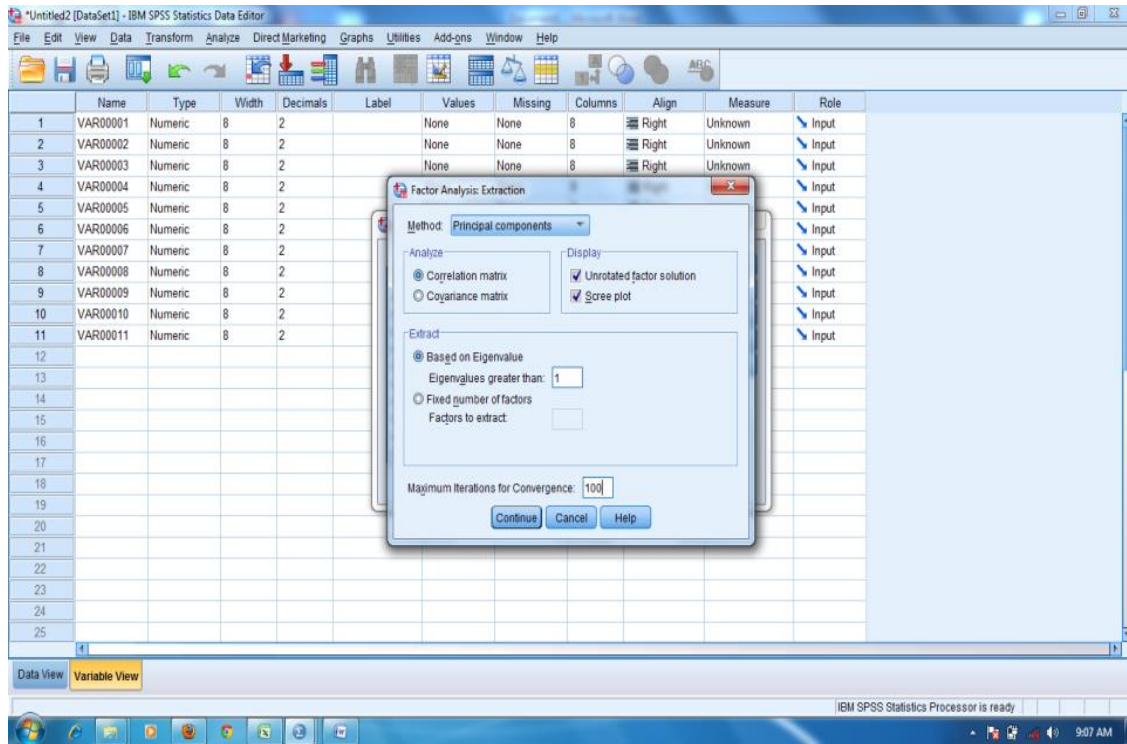
This will open **Factor Analysis** dialogue box.



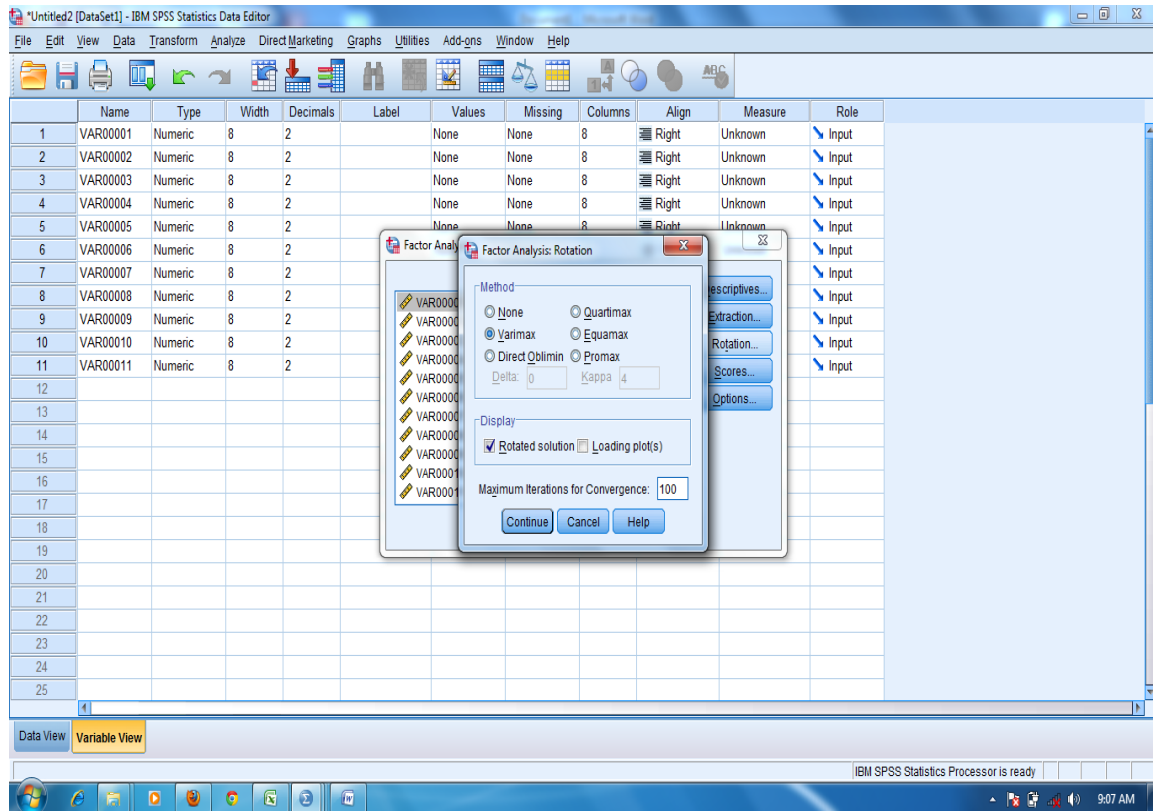
Click the variable in the left list box and send it in **Variables** list box by clicking upper **right arrow** button. Similarly do this for other variables. Click **Descriptives...** button to open its sub dialogue box and select the desired statistics as well as correlation matrix properties according to your need. In the working example, we have selected **KMO and Barlett's test of sphericity** as well as **Anti-image** in the correlation matrix. Click **Continue** button to close the sub dialogue box. The previous dialogue box will reappear.



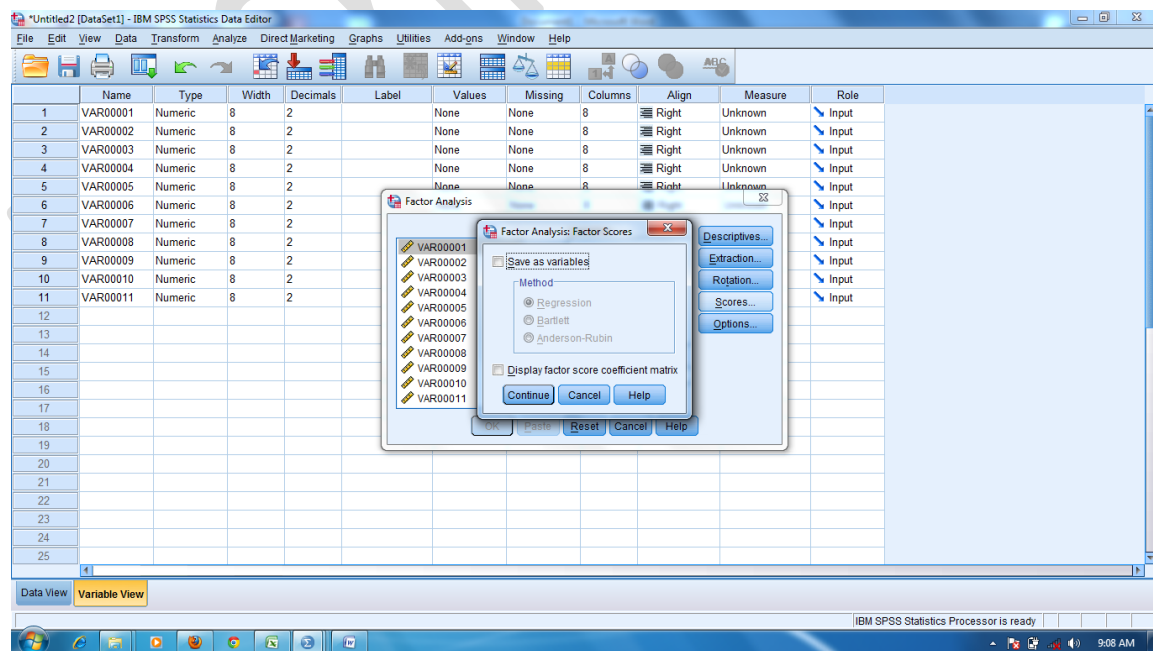
Click **Extraction....** Button to open its sub dialogue box. Select the **method** as **principal components**. You can also change other extraction methods and options as per you need. In the working example, we have selected the following extraction options. Click **Continue** button to close the sub dialogue box. The previous dialogue box will reappear.



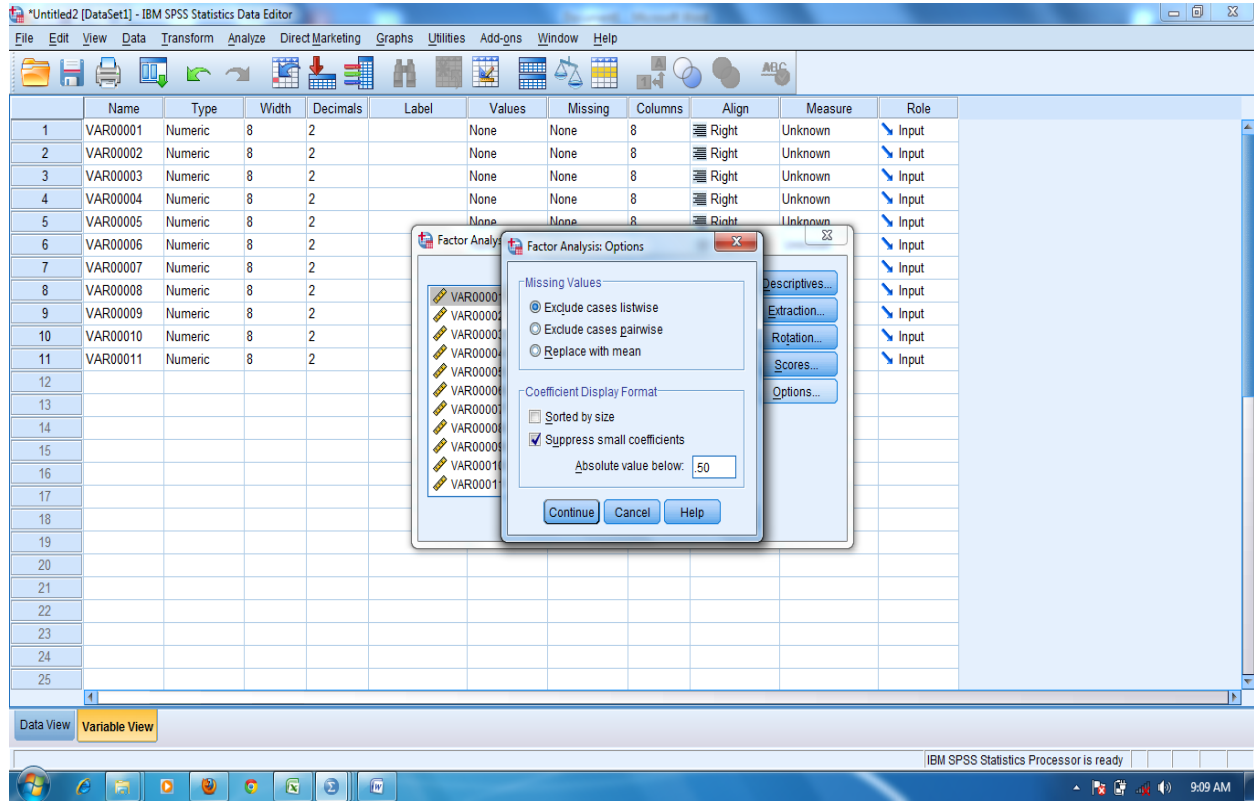
Click **Rotation ...** button to open its sub dialogue box. Select the **method** as **Varimax** and in **Display** select the check box of **Rotated solution**. Click **Continue** button to close the sub dialogue box. The previous dialogue box will reappear.



Click **Scores...** button to open its sub dialogue box. select the scores according to your need. Click **Continue** button to close the sub dialogue box. The previous dialogue box will reappear.



Click **Options...** button to open its sub dialogue box. Select the options according to your need. Click **Continue** button to close the sub dialogue box. The previous dialogue box will reappear. Click **Ok** to see the **output viewer**.



KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.882
Approx. Chi-Square		1179.049
Bartlett's Test of Sphericity	df	55
	Sig.	.000

Communalities

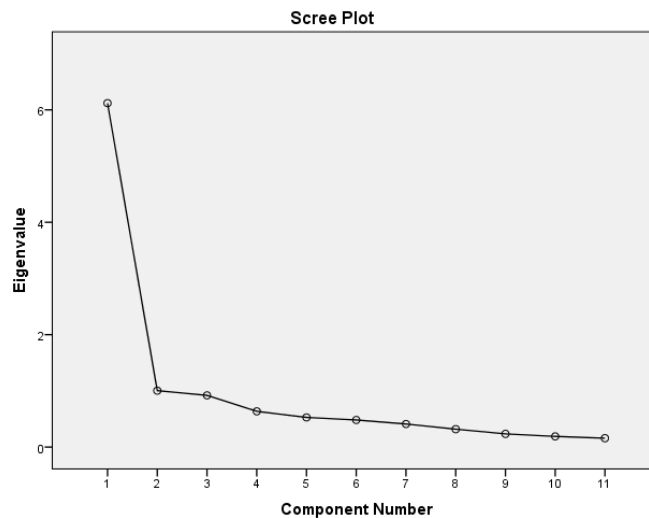
	Initial	Extraction
VAR0001	1.000	.595
VAR0002	1.000	.759
VAR0003	1.000	.704
VAR0004	1.000	.737
VAR0005	1.000	.658
VAR0006	1.000	.774
VAR0007	1.000	.425

VAR0008	1.000	.594
VAR0009	1.000	.587
VAR00010	1.000	.704
VAR00011	1.000	.587

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.119	55.630	55.630	6.119	55.630	55.630	4.811	43.734	43.734
2	1.004	9.126	64.755	1.004	9.126	64.755	2.312	21.021	64.755
3	.920	8.364	73.119						
4	.636	5.782	78.901						
5	.527	4.795	83.695						
6	.482	4.381	88.077						
7	.411	3.733	91.810						
8	.318	2.894	94.704						
9	.235	2.135	96.839						
10	.190	1.726	98.565						
11	.158	1.435	100.000						

Extraction Method: Principal Component Analysis.



Component Matrix^a

	Component	
	1	2

VAR0001	.759	
VAR0002	.860	
VAR0003	.789	
VAR0004	.836	
VAR0005	.796	
VAR0006	.533	.699
VAR0007	.649	
VAR0008	.679	
VAR0009	.648	
VAR00010	.825	
VAR00011	.762	

Rotated Component Matrix^a

	Component	
	1	2
VAR0001	.724	
VAR0002	.813	
VAR0003	.825	
VAR0004	.821	
VAR0005	.765	
VAR0006		.873
VAR0007	.531	
VAR0008		.658
VAR0009		.680
VAR00010	.788	
VAR00011	.697	

➤ **EXERCISE**

80 Respondents were surveyed based on 15 variables with a rating scale of :

Strongly disagree-1, Disagree-2, Neutral-3, Agree-4, Strongly Agree-5.

Conduct a factor analysis based on the following data table and conclude the results of output.

Respondents	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10	Var 11	Var 12	Var 13	Var 14	Var 15
1.	5	5	5	4	5	3	3	3	3	4	4	4	4	4	5
2.	5	4	5	3	5	4	4	4	5	5	5	5	4	5	5
3.	4	4	4	3	4	3	5	2	3	3	4	4	5	4	5
4.	4	4	5	4	4	4	4	3	4	5	5	4	5	4	4
5.	5	5	5	2	5	4	4	2	4	5	5	4	4	3	5
6.	5	4	4	4	3	3	4	3	3	4	3	3	3	3	4
7.	5	3	4	1	3	1	2	2	4	1	3	3	3	2	4
8.	5	5	5	5	3	3	3	3	2	1	3	3	3	3	4
9.	3	3	3	1	3	2	2	3	3	1	2	2	2	1	4
10.	4	5	4	2	4	3	4	4	4	3	4	4	4	4	5
11.	5	5	5	5	3	1	5	1	4	3	4	4	4	4	2
12.	5	5	5	5	3	1	5	1	4	4	4	4	4	4	2
13.	5	5	5	2	5	3	5	5	5	4	5	5	5	5	4
14.	4	3	4	3	3	4	3	3	4	2	4	2	2	3	2
15.	5	5	5	3	5	4	5	4	5	3	4	3	4	5	5
16.	5	5	5	3	2	3	3	3	1	3	1	1	3	3	5
17.	4	4	4	4	4	3	3	2	3	3	4	2	4	3	2
18.	5	1	5	5	1	1	1	1	1	1	1	1	1	1	5
19.	4	2	1	1	4	2	2	4	4	4	4	2	4	4	2
20.	5	3	3	3	4	4	3	3	4	5	4	4	4	5	5
21.	5	5	5	4	4	3	4	4	4	3	4	3	3	4	4
22.	5	5	5	2	4	4	4	4	4	4	4	4	4	3	4
23.	5	4	5	3	4	3	3	3	4	5	4	4	5	4	4
24.	4	5	4	3	5	4	4	3	3	4	4	3	3	3	4
25.	5	4	5	4	4	4	4	4	2	4	3	1	3	2	5
26.	5	4	4	3	4	3	4	4	5	4	4	4	4	4	5
27.	3	1	5	5	4	2	3	4	3	3	2	5	5	2	5
28.	5	4	5	1	5	4	5	4	4	4	3	3	4	3	4
29.	1	5	4	4	2	1	2	2	1	2	2	1	1	1	5

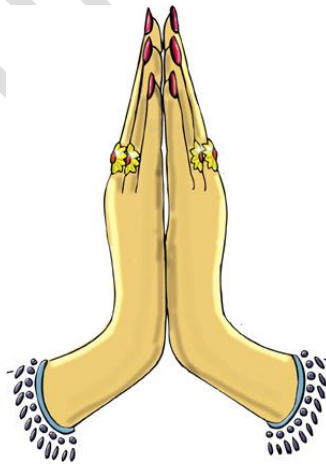
30.	3	1	1	1	4	2	2	2	1	1	1	1	1	3	3
31.	4	5	4	2	4	2	2	4	4	4	4	4	4	5	5
32.	4	5	5	4	3	3	2	2	2	4	4	3	4	4	4
33.	5	5	5	4	4	3	2	3	4	3	4	4	4	4	4
34.	5	5	5	4	5	3	3	3	3	4	4	4	4	4	5
35.	5	4	5	4	5	4	4	5	5	5	5	5	5	4	5
36.	4	4	4	3	4	3	5	2	3	3	4	4	4	5	4
37.	4	4	5	4	4	4	4	3	4	5	5	4	5	4	4
38.	5	5	5	2	5	4	4	2	4	5	5	4	4	3	5
39.	5	4	4	4	3	3	4	3	3	4	3	3	3	3	4
40.	5	3	4	1	3	1	2	2	4	1	3	3	3	2	4
41.	5	5	5	5	3	3	3	3	2	1	3	3	3	3	4
42.	3	3	3	1	3	2	2	3	3	1	2	2	2	1	4
43.	4	5	4	2	4	3	4	4	4	3	4	4	4	4	5
44.	5	5	5	5	3	1	5	1	4	4	4	4	4	4	2
45.	5	5	5	5	3	1	5	1	4	4	4	4	4	4	2
46.	5	5	5	2	5	3	5	5	5	4	5	5	5	5	4
47.	4	3	4	3	3	4	3	3	4	2	4	2	2	3	2
48.	5	5	5	3	5	4	5	4	5	3	4	3	4	5	5
49.	5	5	5	3	2	3	3	3	1	3	1	1	3	3	5
50.	4	4	4	4	4	3	3	2	3	3	4	2	4	3	2
51.	5	1	5	5	1	1	1	1	1	1	1	1	1	1	5
52.	4	2	1	1	4	2	2	4	4	4	4	2	4	4	2
53.	5	3	3	3	4	4	3	3	4	5	4	4	4	5	5
54.	5	5	5	4	4	3	4	4	4	3	4	3	3	4	4
55.	5	5	5	2	4	4	4	4	4	4	4		4	3	4
56.	5	4	5	3	4	3	3	3	4	5	4	4	5	4	4
57.	4	5	4	3	5	4	4	3	3	4	4	3	3	3	4
58.	5	4	5	4	4	4	4	4	2	4	3	1	3	2	5
59.	5	4	4	3	4	3	4	4	5	4	4	4	4	4	5
60.	3	1	5	5	4	2	3	4	3	3	2	5	5	2	5
61.	5	4	5	1	5	4	5	4	4	4	3	3	4	3	4
62.	1	5	4	4	2	1	2	2	1	2	2	1	1	1	5
63.	3	1	1	1	4	2	2	2	1	1	1	1	1	3	3
64.	4	5	4	2	4	2	2	4	4	4	4	4	4	5	5
65.	4	5	5	4	3	3	2	2	2	4	4	3	4	4	4
66.	5	5	5	4	4	3	2	3	4	3	4	4	4	4	4
67.	5	5	5	4	4	2	5	4	4	5	5	5	5	5	4
68.	4	5	5	2	4	2	2	2	3	2	4	4	3	2	4

69.	5	5	5	5	5	5	5	3	4	4	5	5	3	4	5
70.	5	5	5	4	5	2	2	4	3	4	4	4	4	5	4
71.	5	5	5	4	5	5	4	5	5	4	5	5	4	5	5
72.	5	5	5	3	3	1	2	1	2	1	1	4	5	4	5
73.	5	5	5	4	5	2	2	4	3	4	4	4	4	5	4
74.	1	4	3	5	2	2	4	3	4	3	4	3	5	4	5
75.	1	5	5	4	3	2	4	1	2	1	4	4	4	2	4
76.	4	5	5	4	1	1	1		3	3	3	3	3	4	3
77.	5	3	4	2	5	2	4	4	4	4	4	3	4	4	4
78.	5	3	4	2	5	2	4	4	4	4	4	3	4	4	4
79.	5	5	4	3	4	2	1	1	4	4	5	4	3	3	4
80.	5	4	5	5	1	2	3	3	5	4	3	4	4	2	5

Reference:

1. Andy Field, Discovering Statistics using SPSS, Sage Publications (2005).
2. Dr. Deepak Chawla, Dr. Neena Sondhi, Research Methodology, Vikas Publishing House Pvt. Ltd. (2011).
3. Darren George, Paul Mallery (2009). SPSS for Windows Step by Step. 8th edition, Pearson
4. Dr. S L Gupta, Hitesh Gupta, SPSS 17.0 for Researchers, International Book House Pvt. Ltd. (2011)
5. Darren George, Paul Mallery, SPSS for Windows Step by Step, Pearson (8th Edition)

THANK YOU.....& HAPPY RESEARCHING.....



DR.SHINEY CHIB

Publication Partner:

International Journal of Scientific and Research Publications (ISSN: 2250-3153)

shinychib@gmail.com

Dr. Shiney Chib