PRACTICAL FILE

COURSE : PROBABILITY AND STATISTICS

SUBJECT CODE : MC-205

(Based on SPSS)

MATHEMATICS AND COMPUTING

B.TECH SEMESTER-3

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(FORMELY DELHI COLLEGE OF ENGINEERING)

BAWANA ROAD , DELHI – 110042

SUBMITED BY : NAMAN SAGAR

ROLL NO . : 23/MC/093

SUBMITED TO : Dr. KANICA GOEL

Dr. NILAM

INDEX

Experiment - 1

Objective: Transporation of Data Set to SPSS Editor

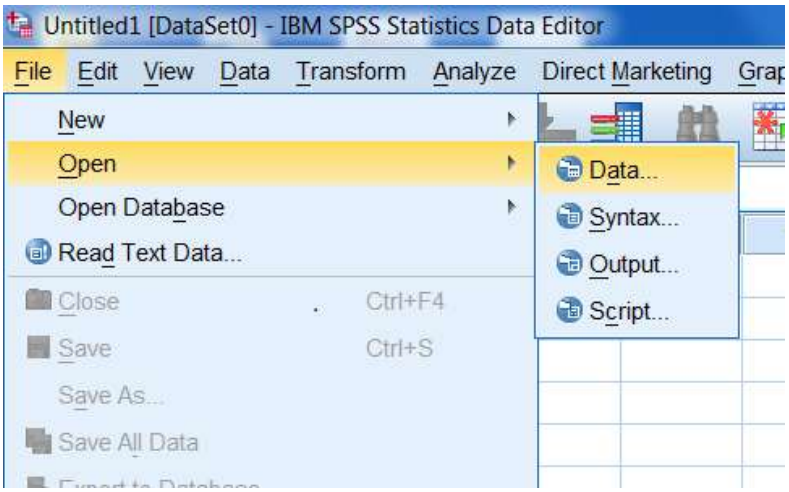
Resources: SPSS editor, Source file

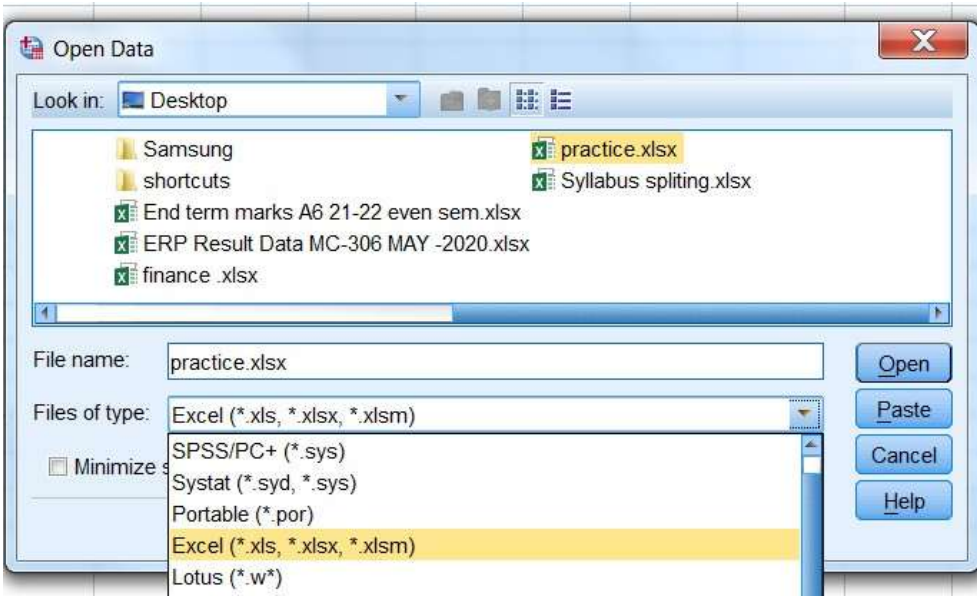
Proceedure And Methodology:

Open the spss editor window and follow the steps as shown below:

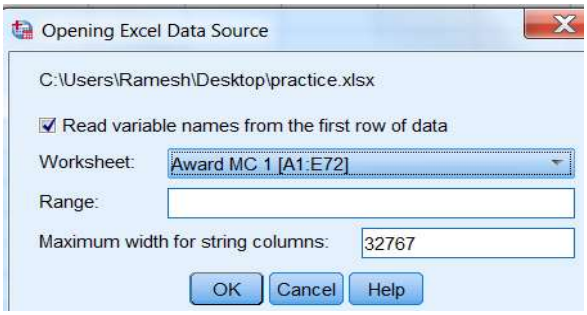
Case 1. Data from an excel file

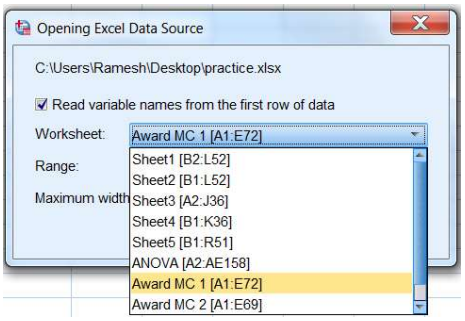
Commands: File > Open > Data> give file path



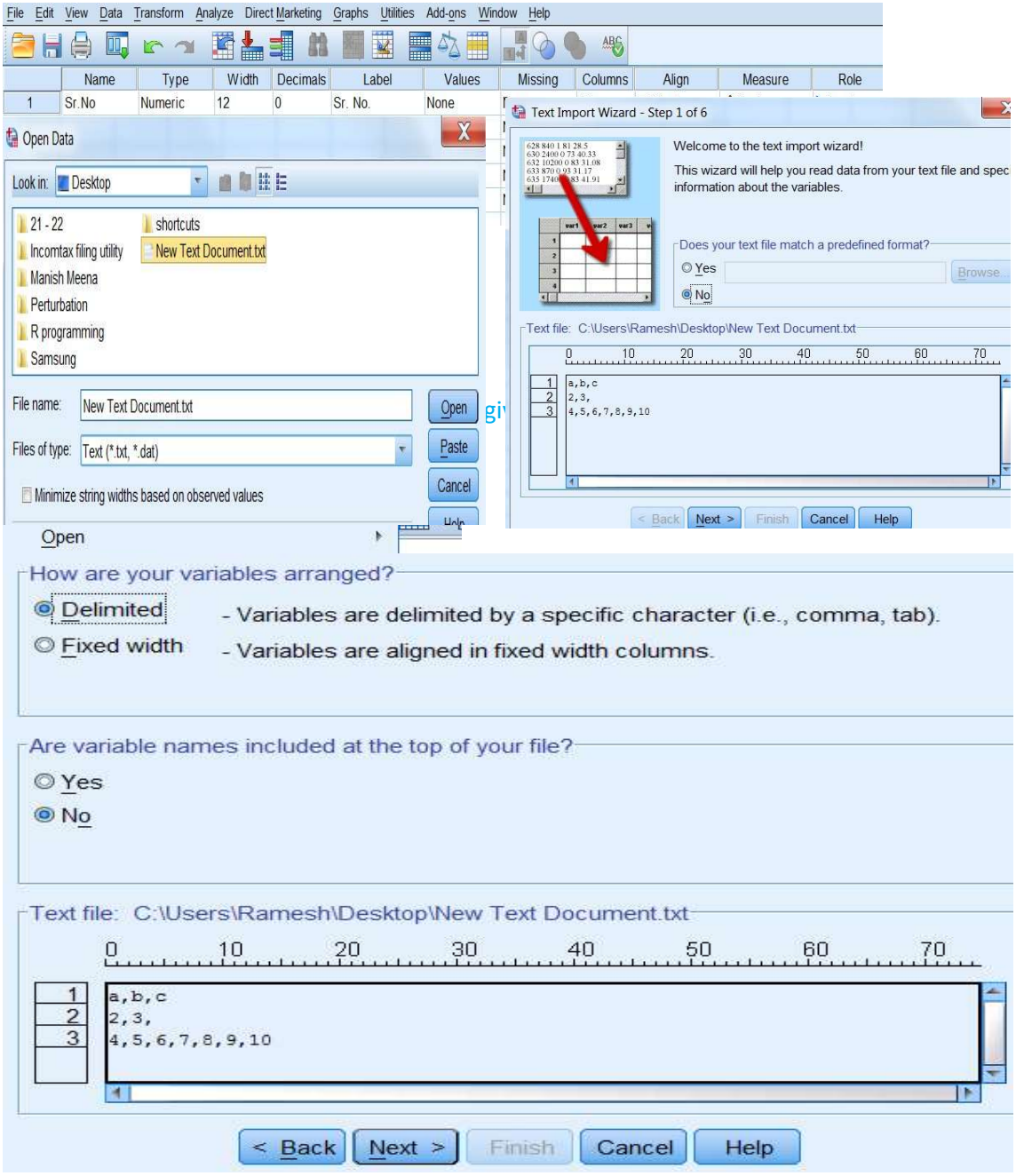


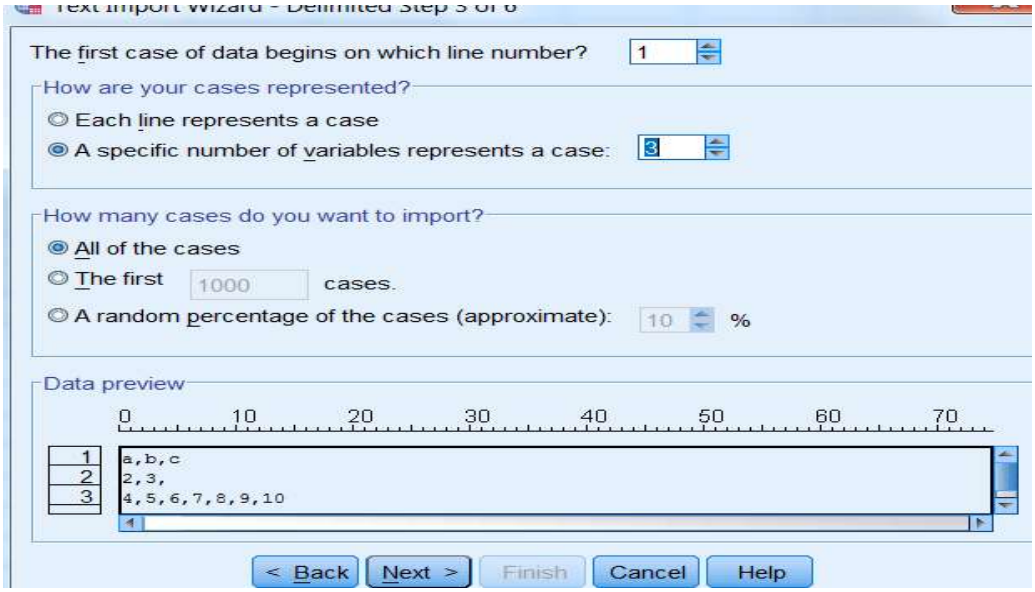
(select the requisite file extension)



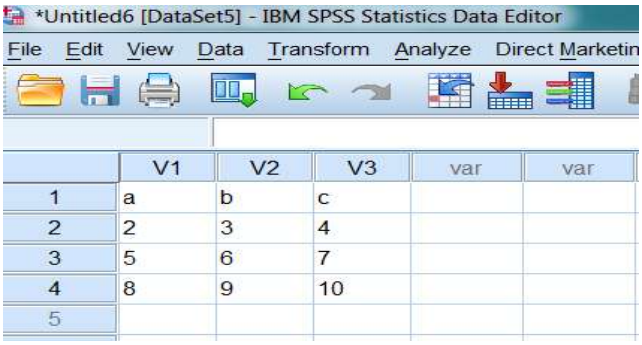


Output Window





Output window



Conclusion:

Data from files in excel or text format can be transported to SPSS editor window.

Precautions:

1. There should be proper delimiter between data entry in text file.
2. Extensions of the files should be strictly taken care of.

Experiment - 2

Objective: Merging of Data set

Resources: SPSS editor, Source files with .sav extension

Proceedure And Methodology:

Case 1. Merging data as add cases.

Open the spss editor window and read the data of 1st file follow the steps as shown below:

Sample files are origanally shown below

Sam 1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing G

7 : RollNo

	Name	RollNo	Marksmath	marksphy
1	A	1	60	65
2	B	2	70	75
3	C	3	80	85
4	D	4	90	95
5				

Sam 2.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing G

	Name	RollNo	Marksmath	marksphy
1	E	5	62	66
2	F	6	72	77
3	G	7	82	88
4	H	8	92	99
5				

Sam 3.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing G

	Name	RollNo	Markseng	markschm
1	A	1	61	63
2	B	2	71	73
3	C	3	81	83
4	D	4	91	93
5				

Sam 2 - Copy.sav [DataSet4] - IBM SPSS Statistics Data Editor

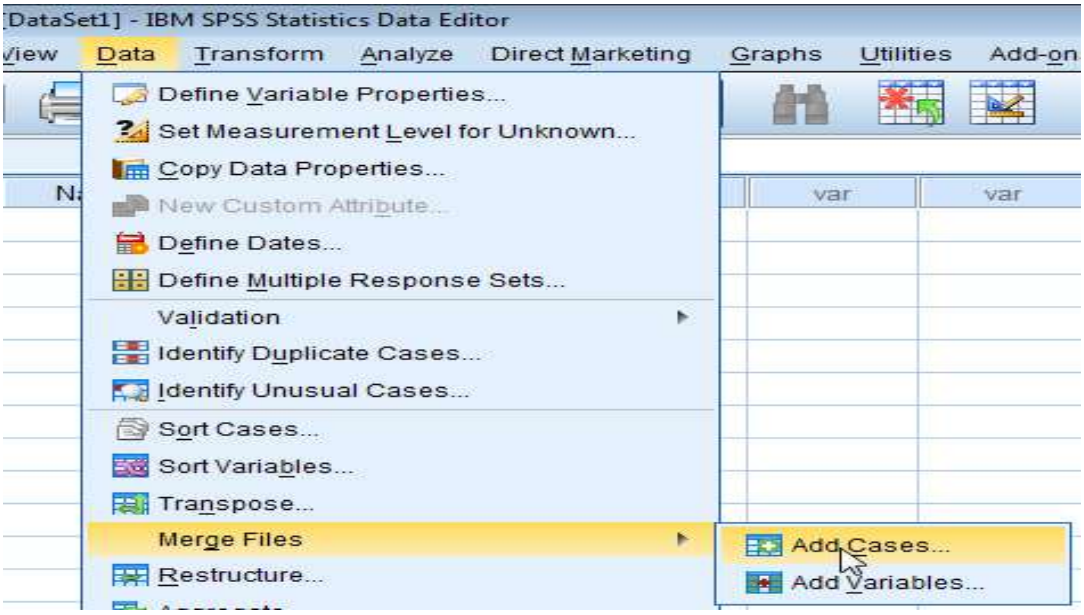
File Edit View Data Transform Analyze Direct Marketing G

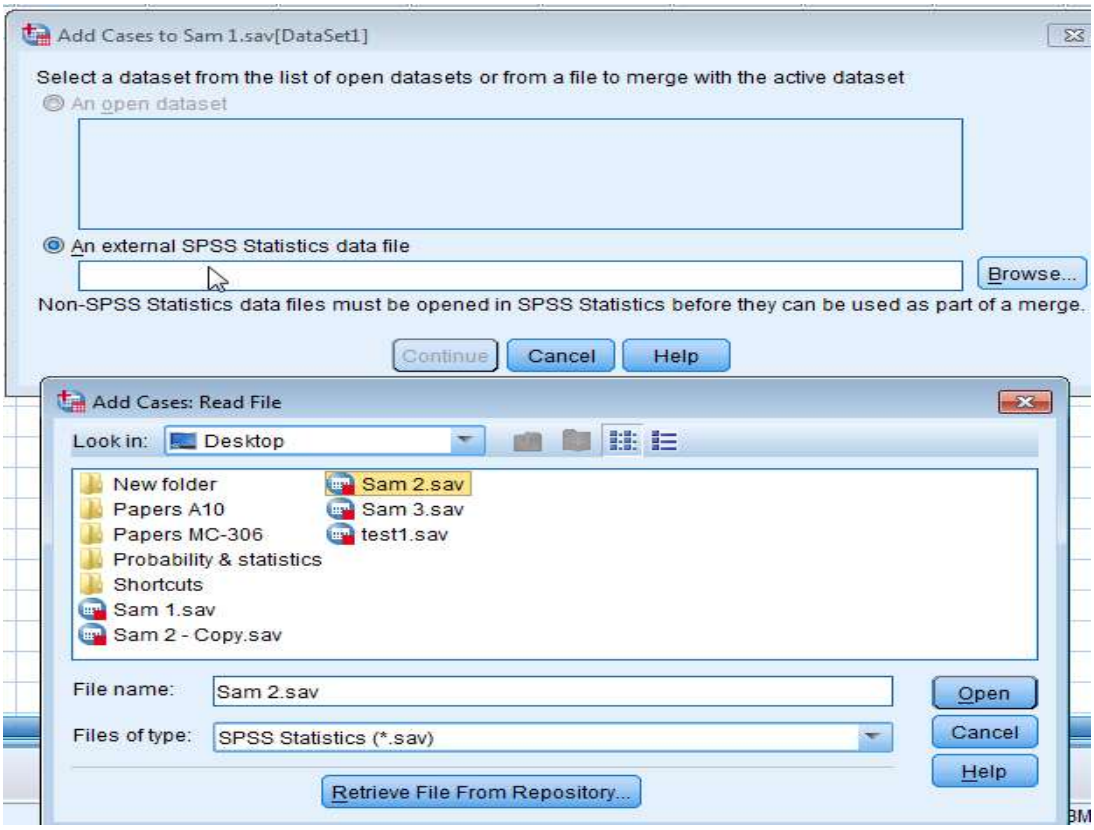
	Name	RollNo	MM	MP
1	E	5	62	66
2	F	6	72	77
3	G	7	82	88
4	H	8	92	99

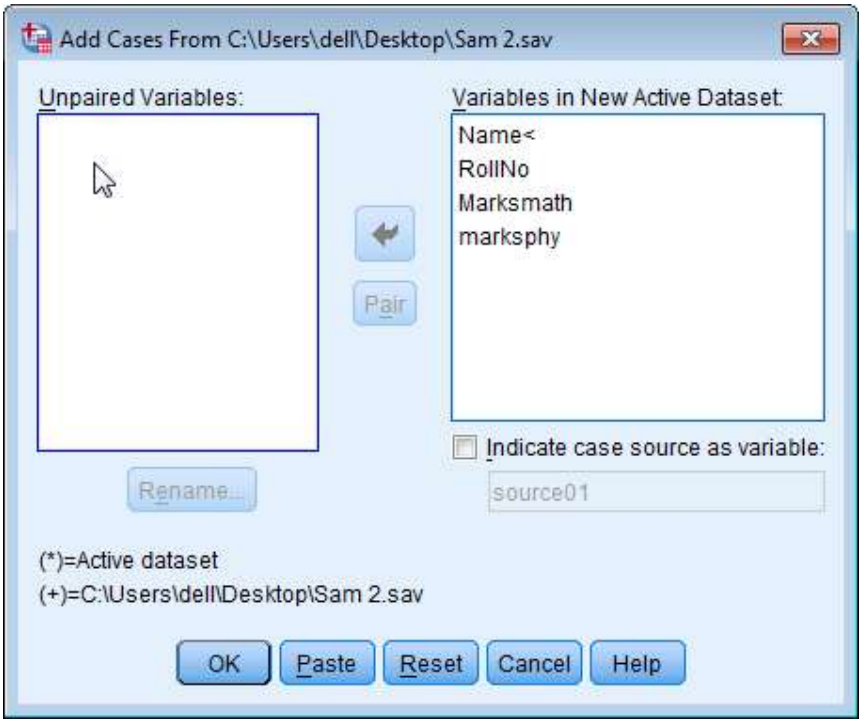
Case 1.1: Merging files with identical variable name

Commands: File > Open > Data > sam1.sav

Merging sam2.sav in sam1.sav Data > merge file > add cases > browse>select file and click open

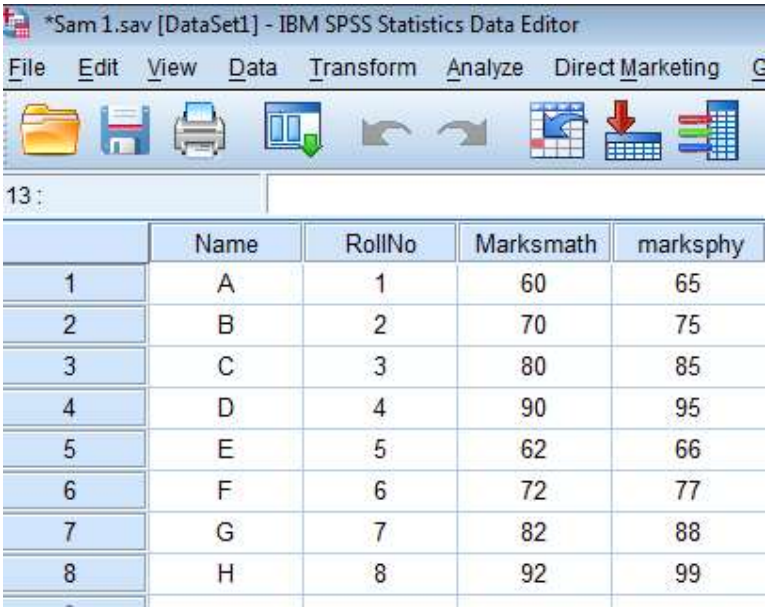






(both files are possessing the identical variable names)

Output window after merger.

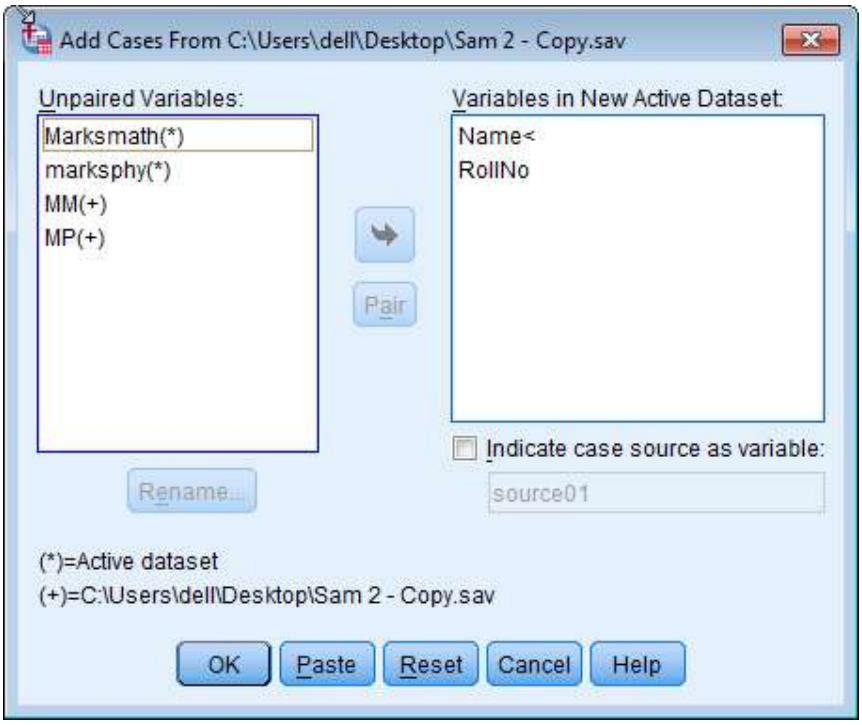


	Name	RollNo	Marksmath	marksphy
1	A	1	60	65
2	B	2	70	75
3	C	3	80	85
4	D	4	90	95
5	E	5	62	66
6	F	6	72	77
7	G	7	82	88
8	H	8	92	99

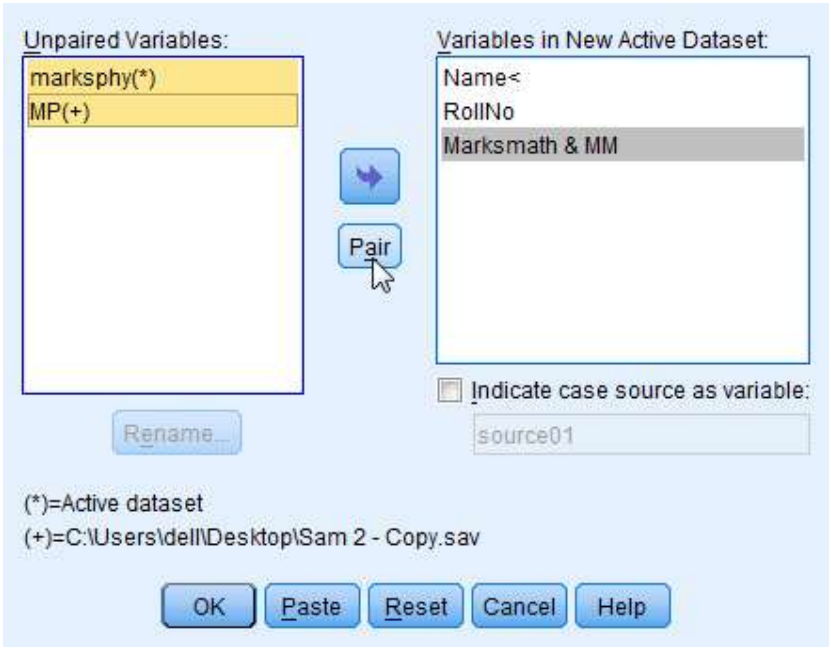
Case 1.2: Merging files with different variable name but same data

Commands: File > Open > Data > sam1.sav

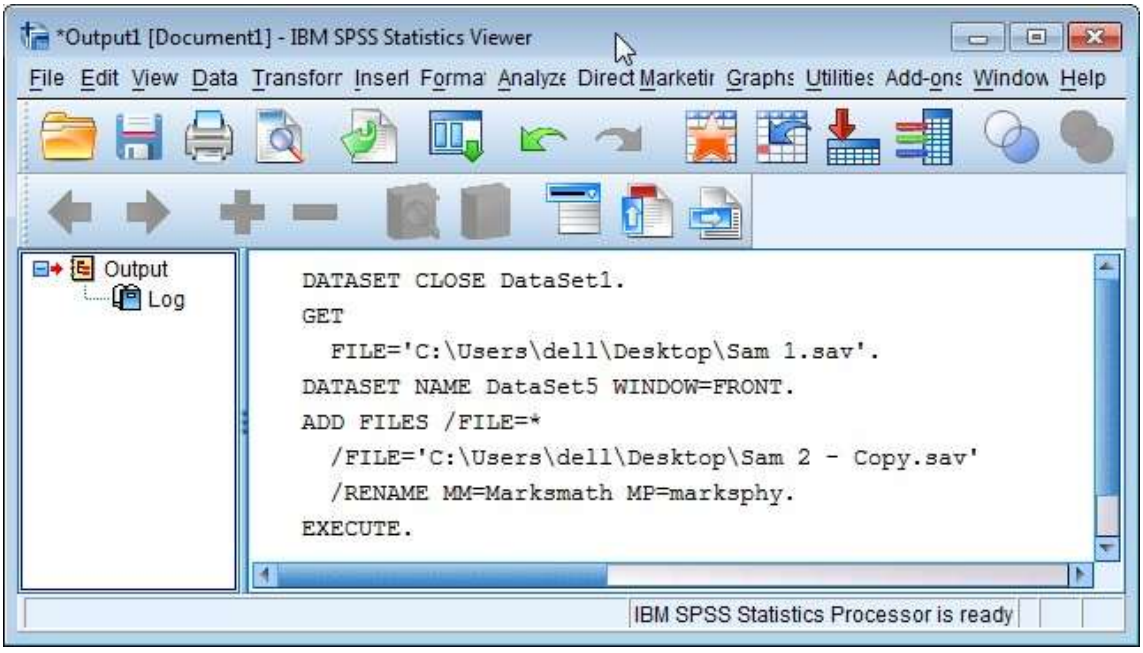
(Merging sam2.sav_copy in sam1.sav) Data > merge file > add cases > browse > select file



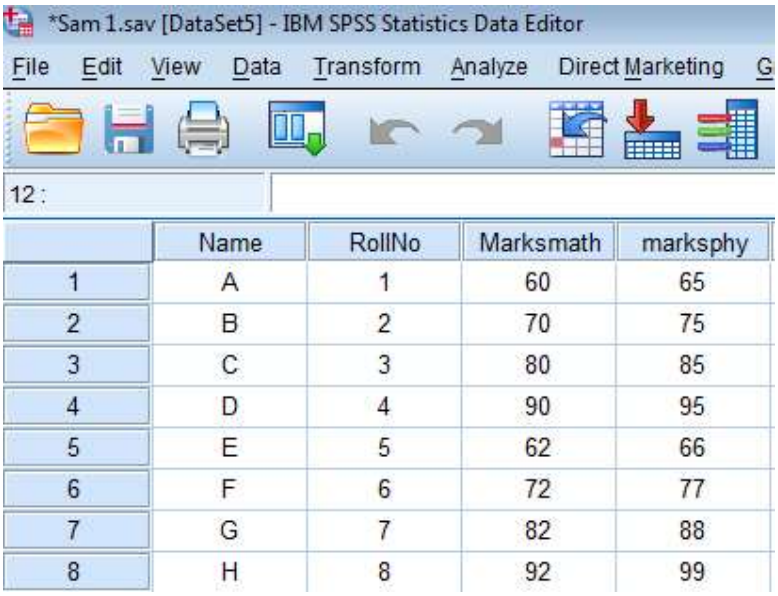
MM & MP contains the marks of math & physics but named different, so they are to be paired. Select both the variables to be paired and click pair.



Doc Output window shows the execution.

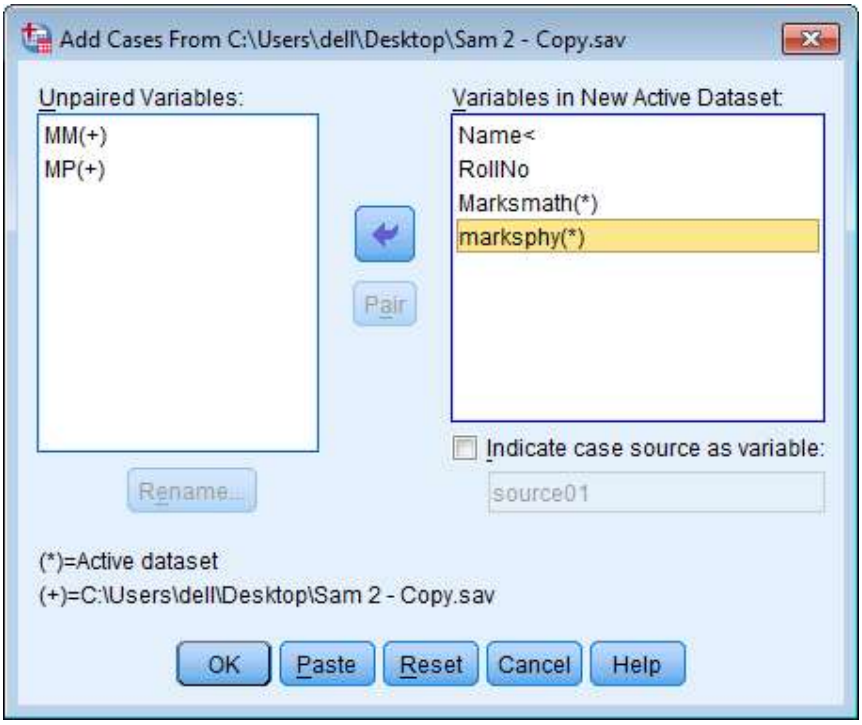


Output window

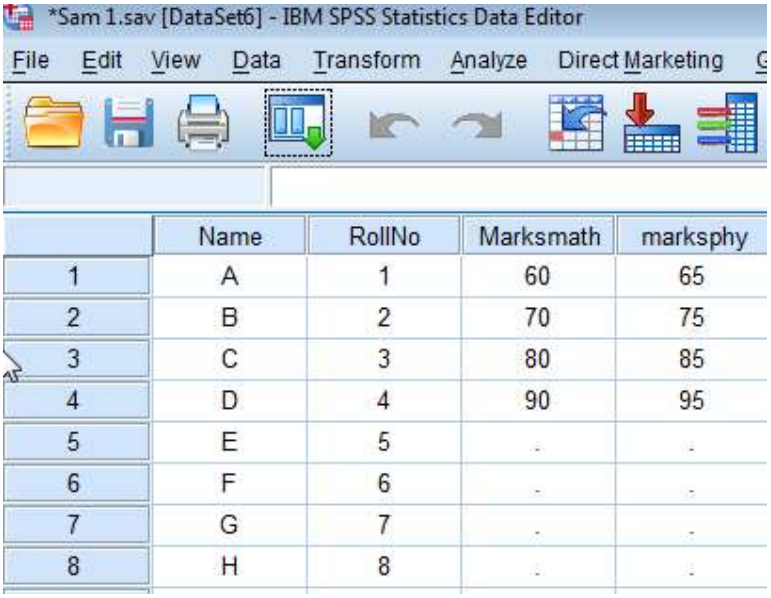


	Name	RollNo	Marksmath	marksphy
1	A	1	60	65
2	B	2	70	75
3	C	3	80	85
4	D	4	90	95
5	E	5	62	66
6	F	6	72	77
7	G	7	82	88
8	H	8	92	99

If the files are merged without pairing,



Output without pairing



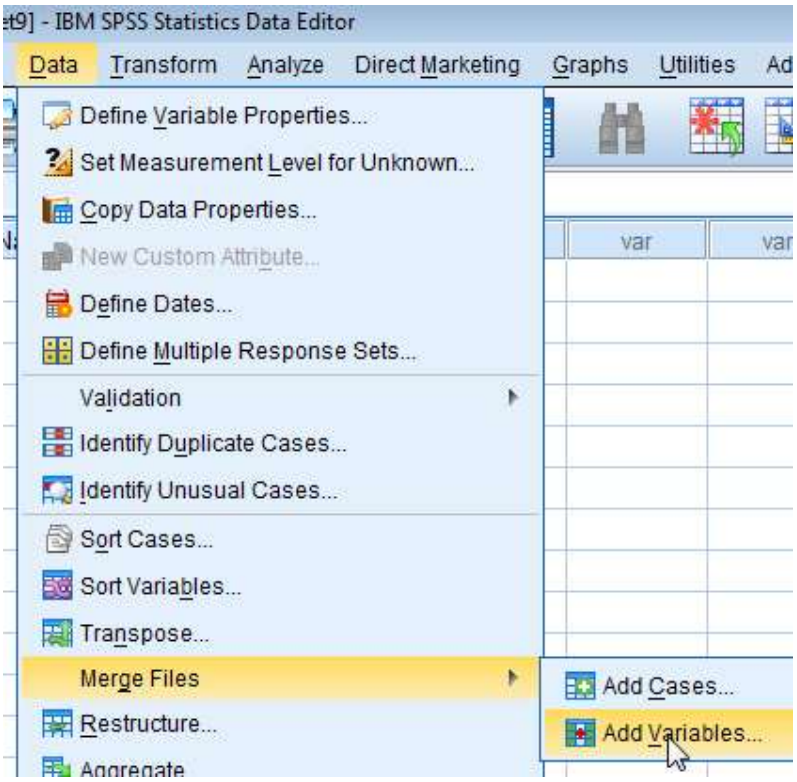
	Name	RollNo	Marksmath	marksphy
1	A	1	60	65
2	B	2	70	75
3	C	3	80	85
4	D	4	90	95
5	E	5	.	.
6	F	6	.	.
7	G	7	.	.
8	H	8	.	.

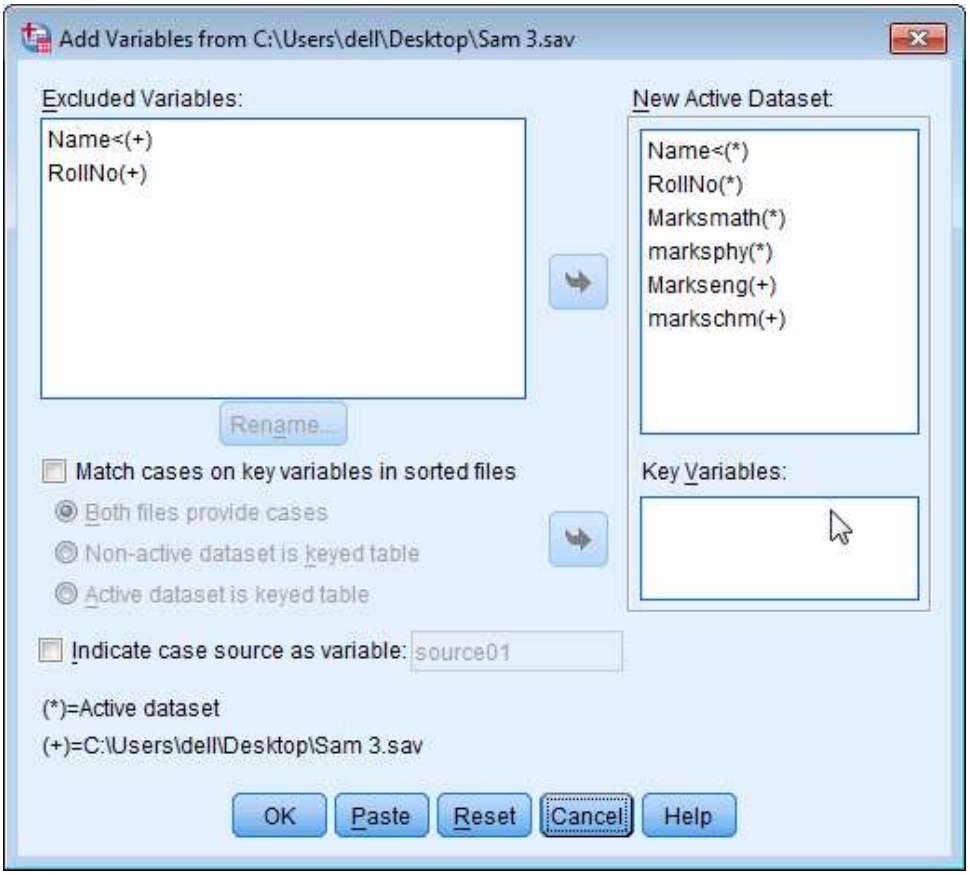
Dotted values are missing values. Names and roll no. are added but marks are not merged.

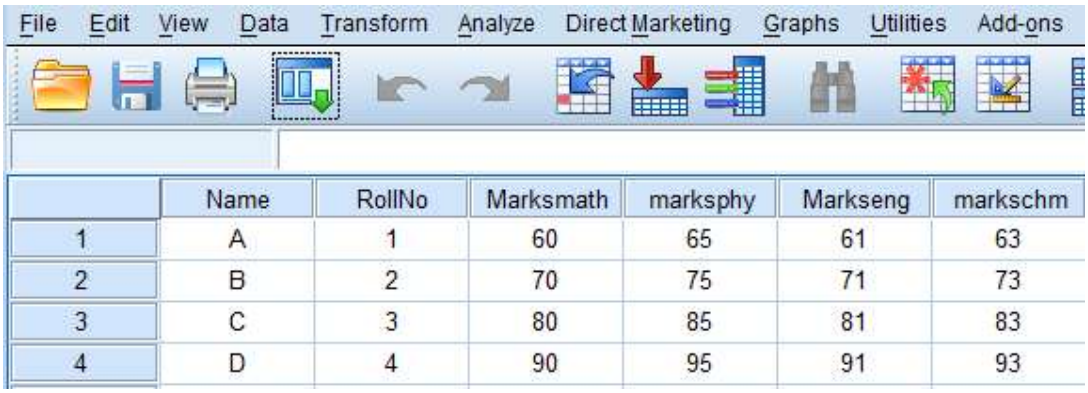
Case 2. Merging data as add variables.

Commands: File > Open > Data> sam1.sav

Merging sam3.sav in sam1.sav Data > merge file > add variables > browse>select file


et9] - IBM SPSS Statistics Data Editor
Data
Transform
Analyze
Direct Marketing
Graphs
Utilities
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Sort Cases...
Sort Variables...
Transpose...
Merge Files
Restructure...
Aggregate
var
Add Cases...
Add Variables...

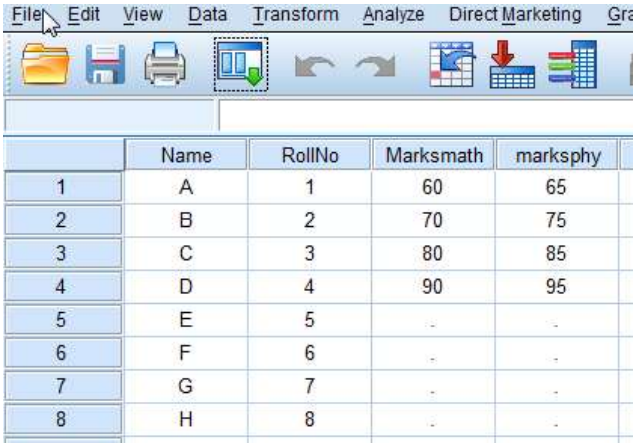

Add Variables from C:\Users\dell\Desktop\Sam 3.sav
Excluded Variables:
Name<(+)
RollNo(+)
Rename...
New Active Dataset:
Name<(*)
RollNo(*)
Marksmath(*)
marksphy(*)
Markseng(+)
markschm(+)
Match cases on key variables in sorted files
Both files provide cases
Non-active dataset is keyed table
Active dataset is keyed table
Key Variables:
Indicate case source as variable:
source01
(*)=Active dataset
(+)=C:\Users\dell\Desktop\Sam 3.sav
OK
Paste
Reset
Cancel
Help



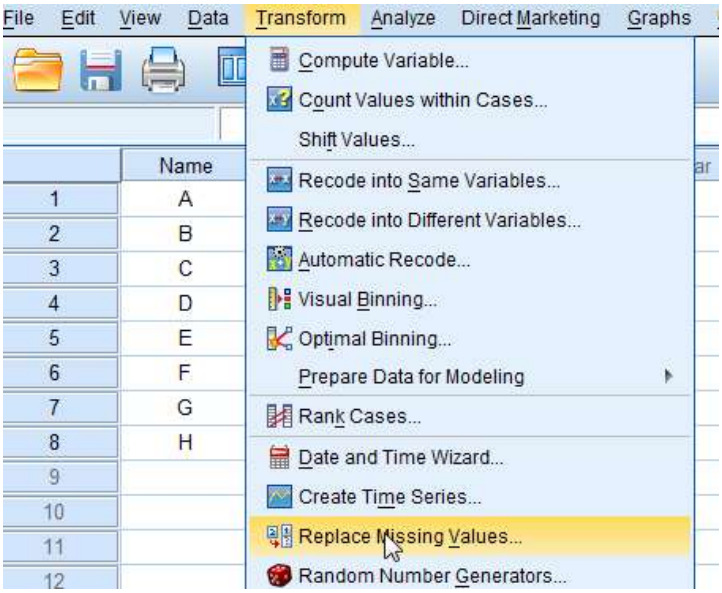
	Name	RollNo	Marksmath	marksphy	Markseng	markschm
1	A	1	60	65	61	63
2	B	2	70	75	71	73
3	C	3	80	85	81	83
4	D	4	90	95	91	93

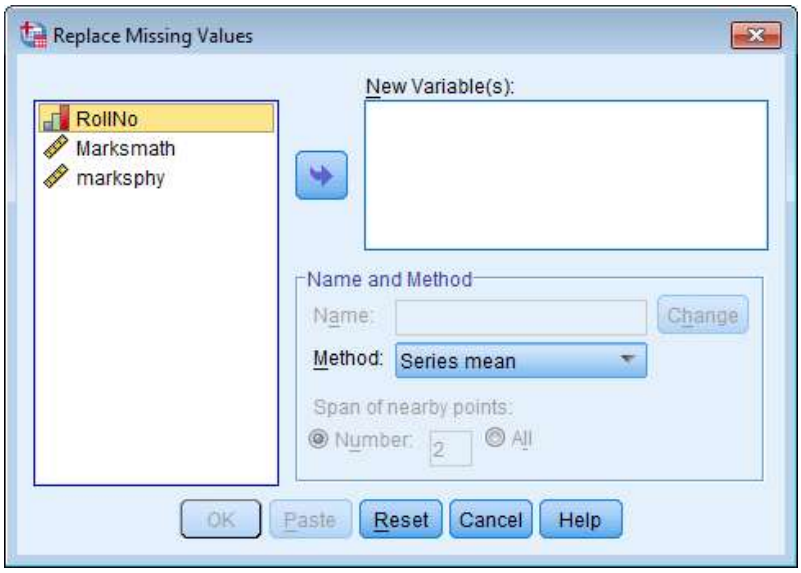
Providing missing values

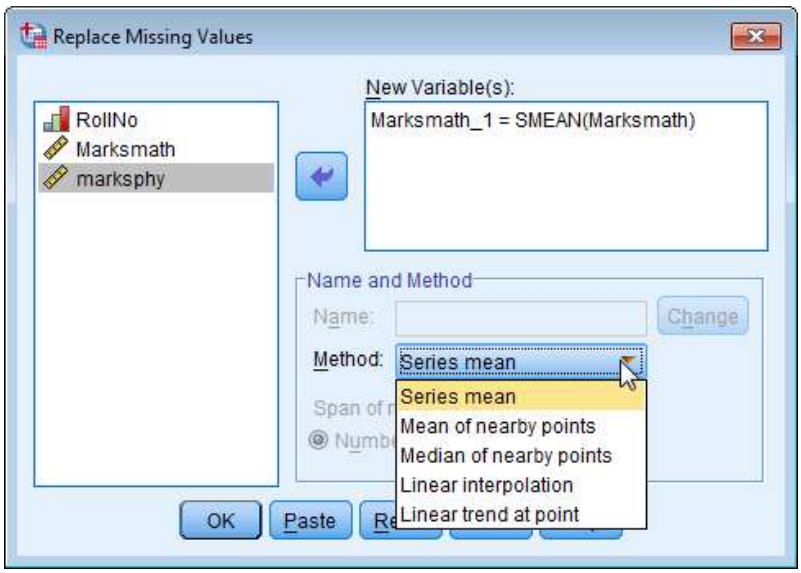
Open the file with sav extension, containing missing value then follow the command transform > replace missing values



	Name	RollNo	Marksmath	marksphy
1	A	1	60	65
2	B	2	70	75
3	C	3	80	85
4	D	4	90	95
5	E	5	.	.
6	F	6	.	.
7	G	7	.	.
8	H	8	.	.







choose the method

Replace Missing Values

```
[DataSet1] C:\Users\dell\Desktop\Files for SPSS\Sam 1.sav
```

Result Variables

	Result Variable	N of Replaced Missing Values	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1	Marksmath_1	4	1	8	8	SMEAN (Marksmath)

Double-click to activate

Final data output window

: Marksmath_1 | 60.0

	Name	RollNo	Marksmath	marksphy	Marksmath_1
1	A	1	60	65	60.0
2	B	2	70	75	70.0
3	C	3	80	85	80.0
4	D	4	90	95	90.0
5	E	5	.	.	75.0
6	F	6	.	.	75.0
7	G	7	.	.	75.0
8	H	8	.	.	75.0

A new variable is created with missing value replaced by the mean of the series.

Conclusion:

Data files are merged by adding cases, and by adding variables. Also the missing values are replaced by (mention the name of method used)… ..... value.

Precautions:

1. Merger is possible only for .sav files, therefore each data should be first imported and saved in .sav file.
2. The pairing of the variables is to be done carefully.

Experiment - 3

Objective: Pictographical representation of data using Bar diagram and Pie chart

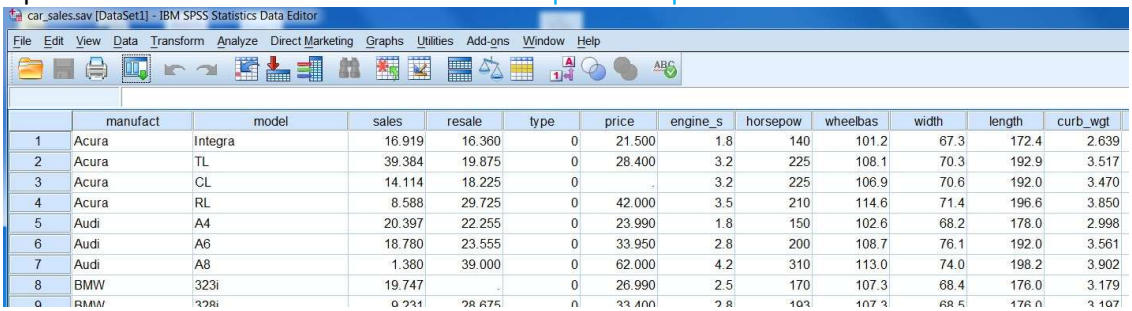
Resources: SPSS editor, Data Source file .

Proceedure And Methodology

A pictographical representation of data helps in quick and better interpretation of the data.

Case1. Bar diagram

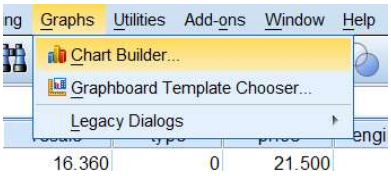
Open the desired file in SPSS editor File > Open > file path

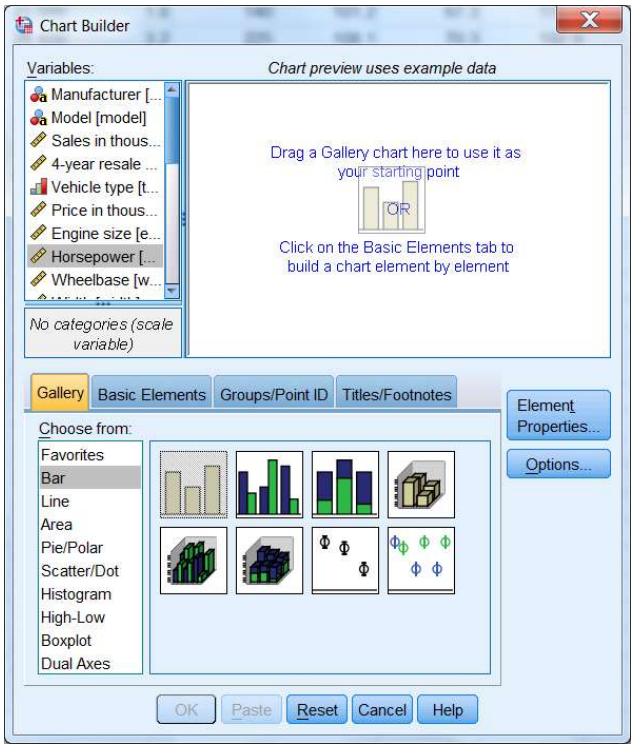


	manufact	model	sales	resale	type	price	engine_s	horsepow	wheelbas	width	length	curb_wgt
1	Acura	Integra	16.919	16.360	0	21.500	1.8	140	101.2	67.3	172.4	2.639
2	Acura	TL	39.384	19.875	0	28.400	3.2	225	108.1	70.3	192.9	3.517
3	Acura	CL	14.114	18.225	0	.	3.2	225	106.9	70.6	192.0	3.470
4	Acura	RL	8.588	29.725	0	42.000	3.5	210	114.6	71.4	196.6	3.850
5	Audi	A4	20.397	22.255	0	23.990	1.8	150	102.6	68.2	178.0	2.998
6	Audi	A6	18.780	23.555	0	33.950	2.8	200	108.7	76.1	192.0	3.561
7	Audi	A8	1.380	39.000	0	62.000	4.2	310	113.0	74.0	198.2	3.902
8	BMW	323i	19.747	.	0	26.990	2.5	170	107.3	68.4	176.0	3.179
9	BMW	328i	9.231	28.675	0	33.400	2.8	193	107.3	68.5	176.0	3.197

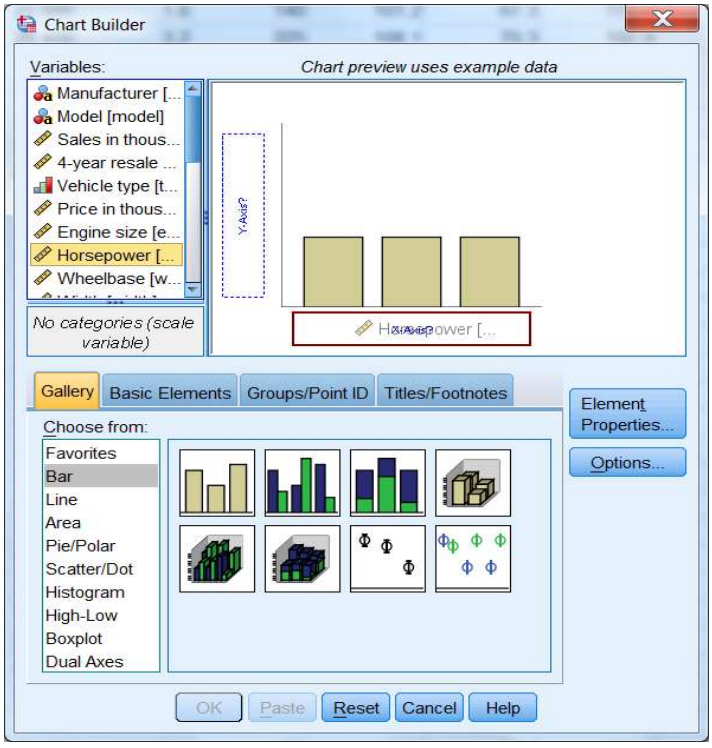
This is car_sales data and we are to plot a bar diagram for the Horse Power

Go to Graph > Chart Builder and click (charts can be drawn using legacy dialogs also)





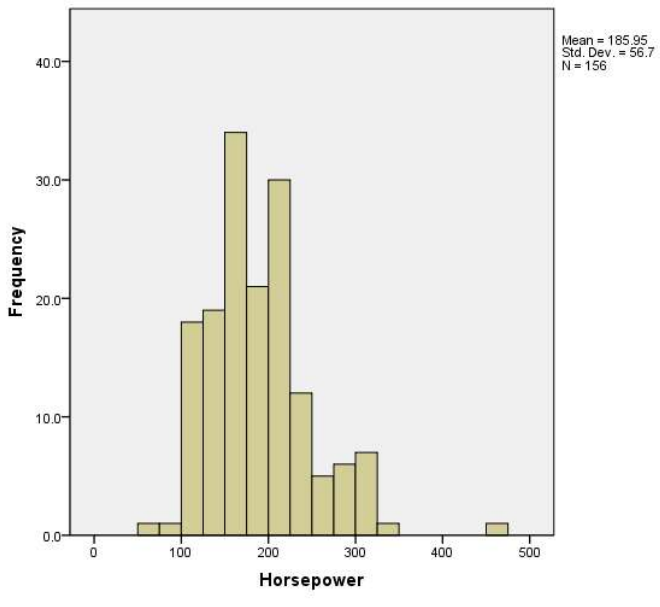
Select the variable Horse Power then Bar and then drag the picture of bar in chart preview area.



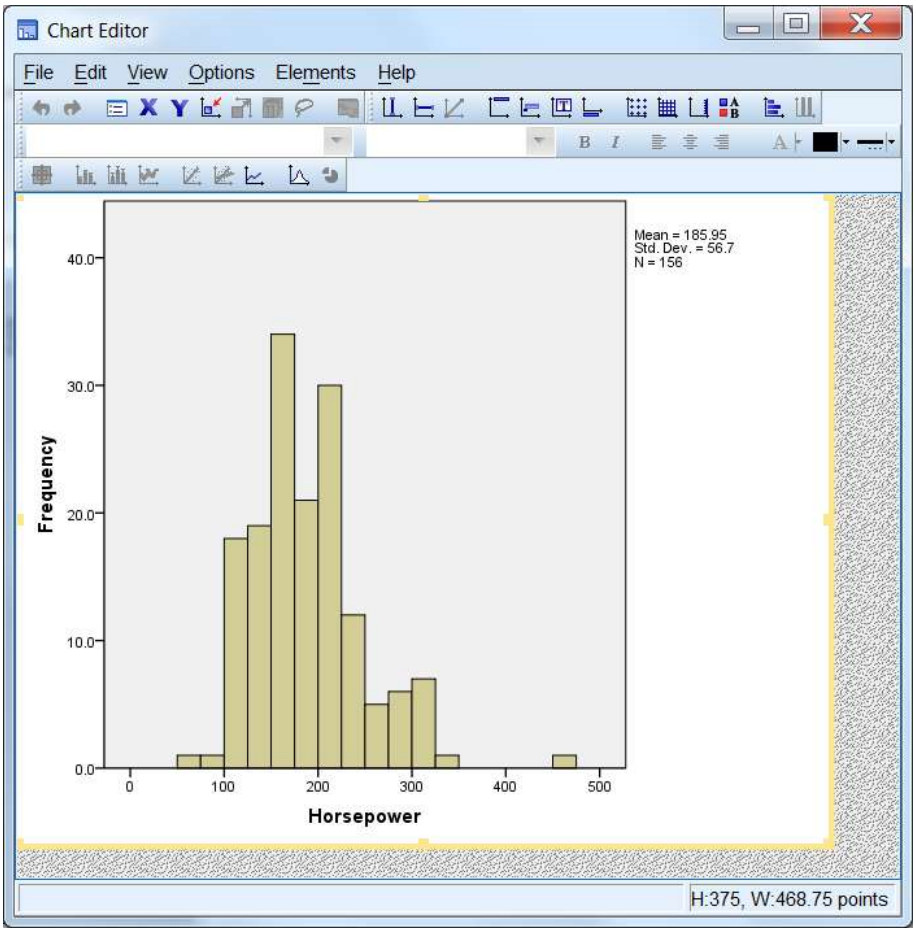
Drag Horsepower to horizontal box and click OK

GGraph

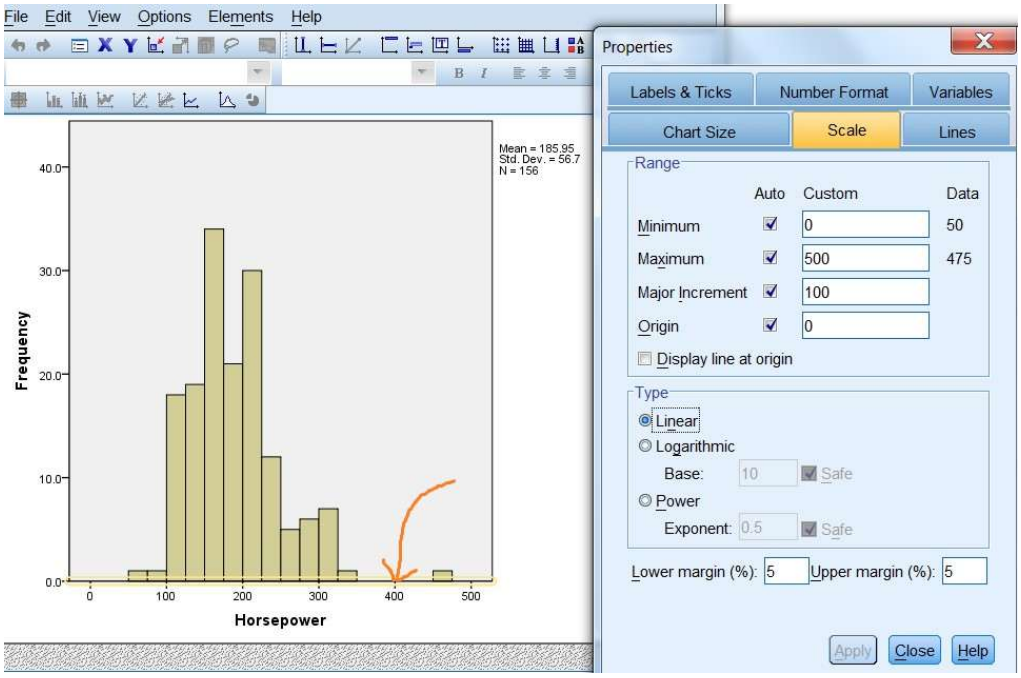
```
[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\car_sales.sav
```



Double click the at graph area and it will go in editable mode



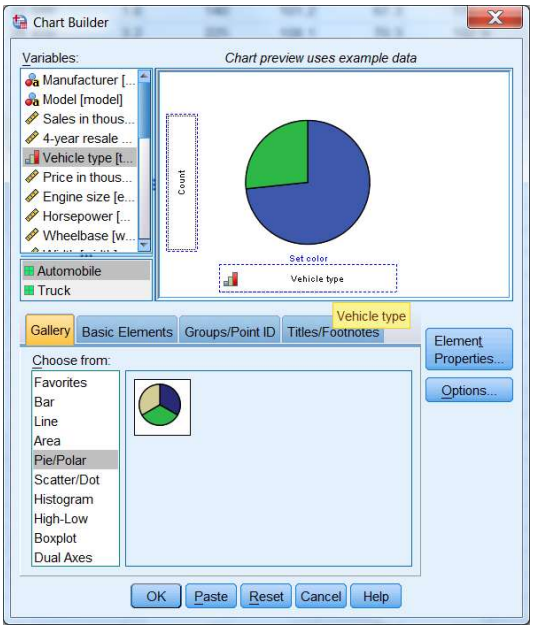
To which part you want to edit double click and a dialog box will open to edit

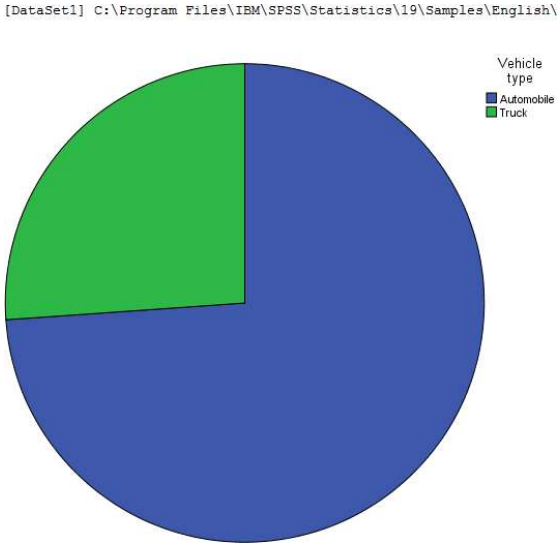


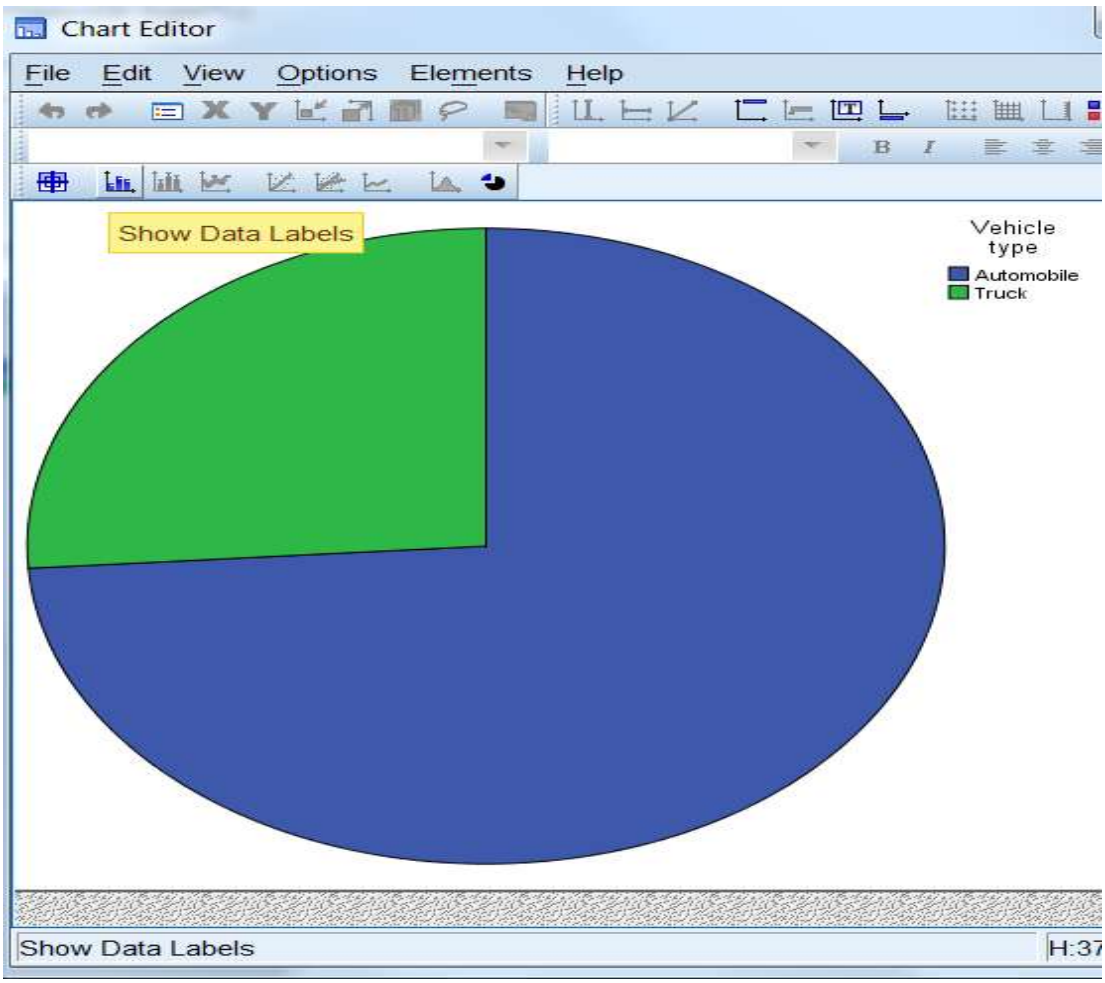
e.g. Horizontal axis is selected for editing.

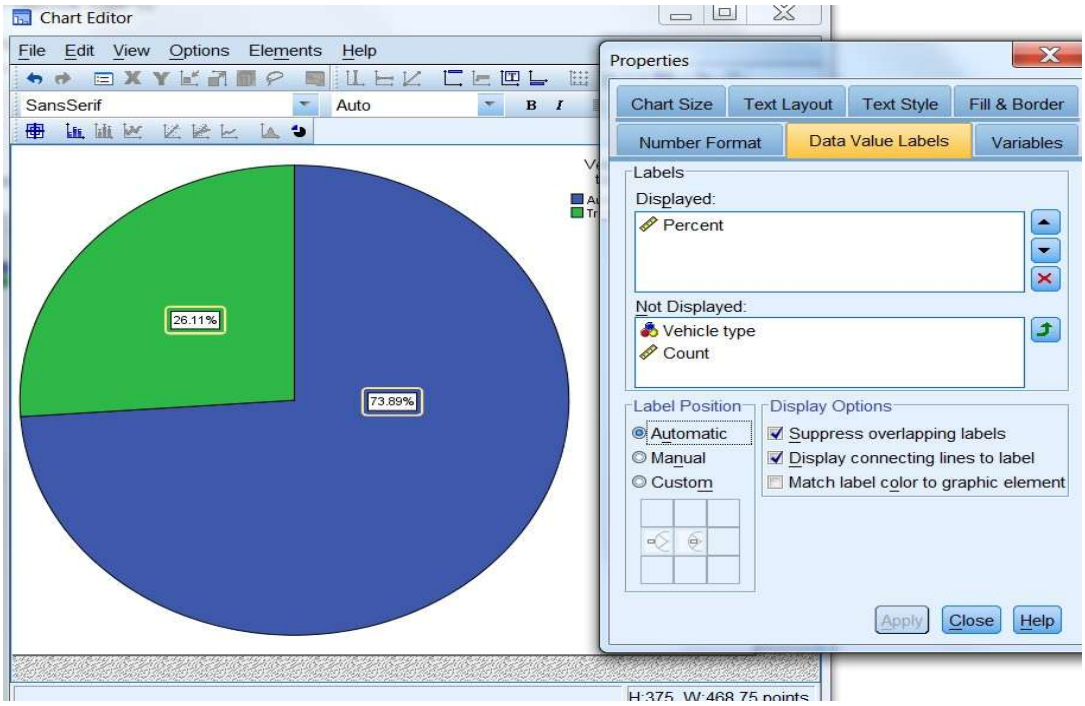
Case2. Pie chart

After opening the file go to Graph > Chart Builder the select the chart type Pie









Conclusion:

The Data set of Car sales from repository is used, and the bar diagram of Horse power of engine variable and Pie chart of Type of vehicle variable is successfully drawn.

Precautions:

1 The variable property should be defined carefully.
2. Missing values if any should be taken care of.

Experiment - 4

Objective: Drawing of Histogram and distribution curve

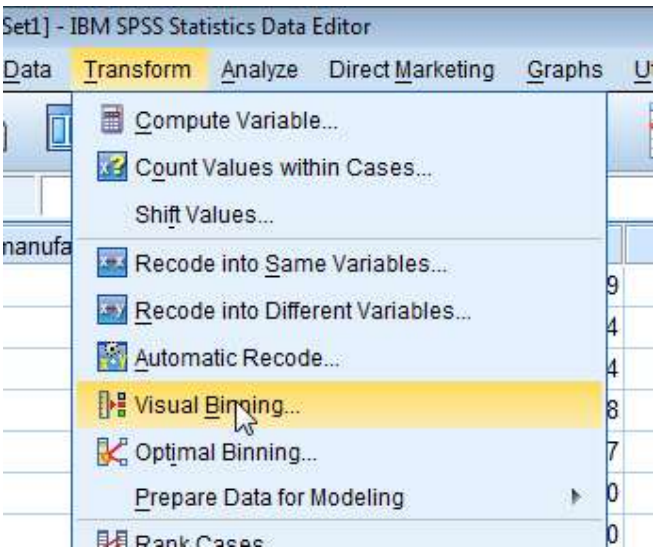
Resources: SPSS editor, Data Source file.

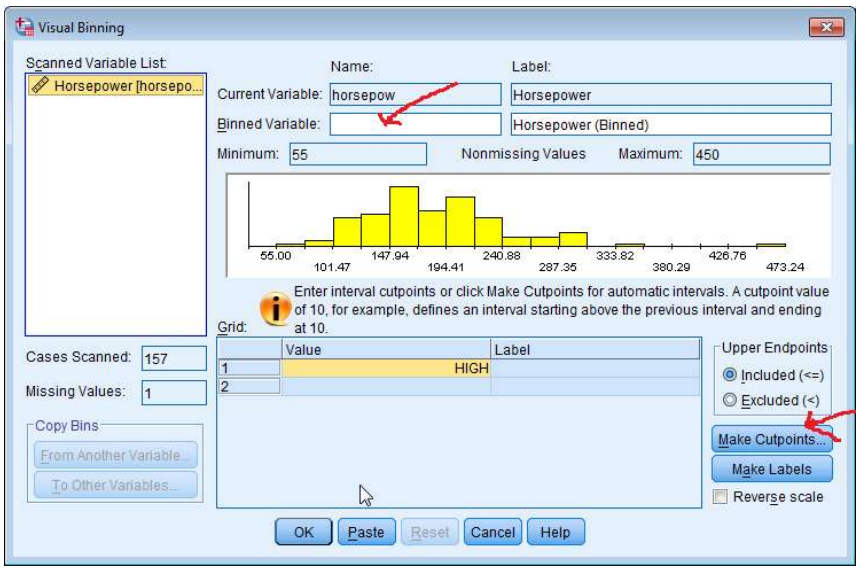
Proceedure And Methodology

Data is arranged in a table with respective frequency. Then Histogram is drawn. The curve obtained by joining the middle point of each bar gives us frequency curve.

Histogram

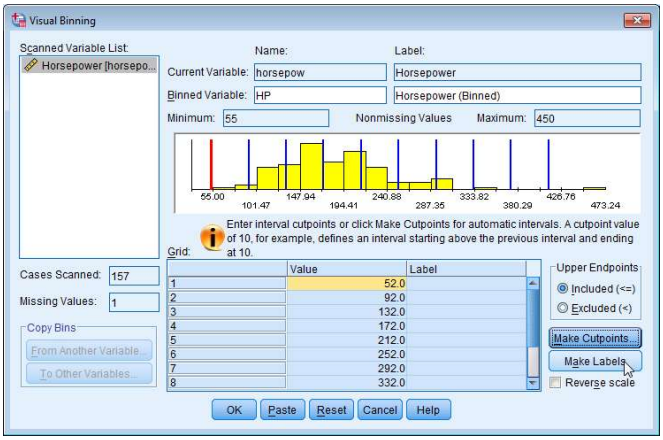
Open the desired file in SPSS editor File > Open > file path
(car_sales.sav from repository is used & the variable Horsepower is considered for frequency)
Transform > visual Binning > enter





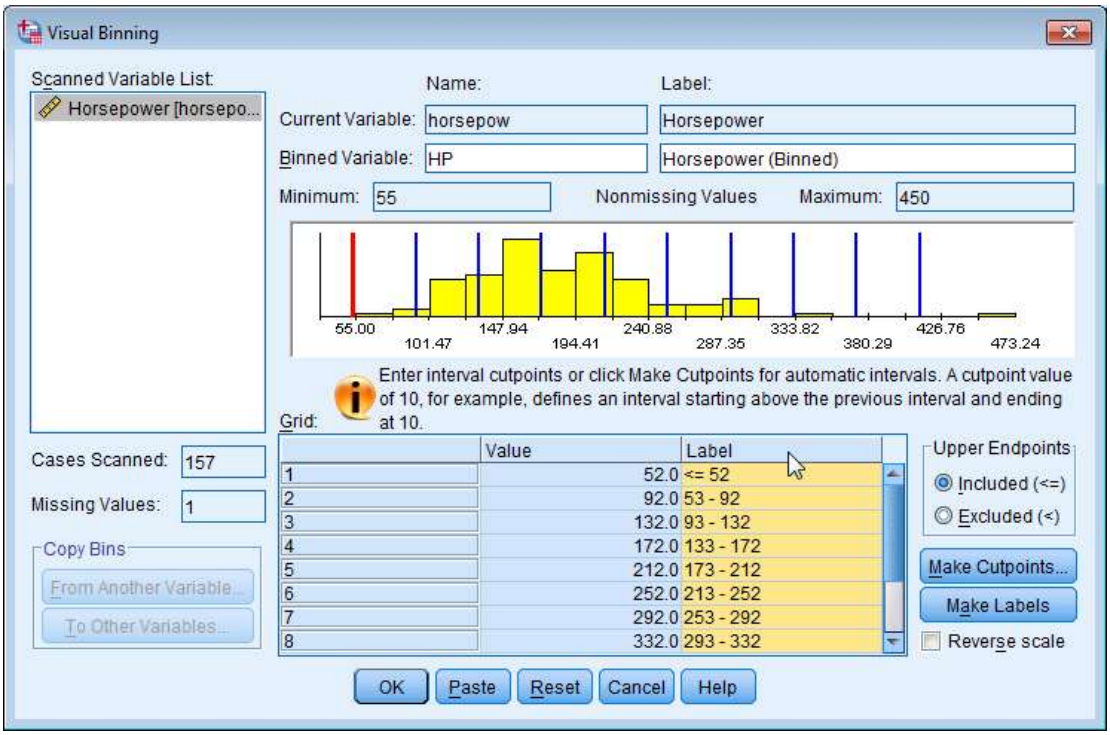
Provide a new name for variable. Then click make cut points.

give desired values then OK



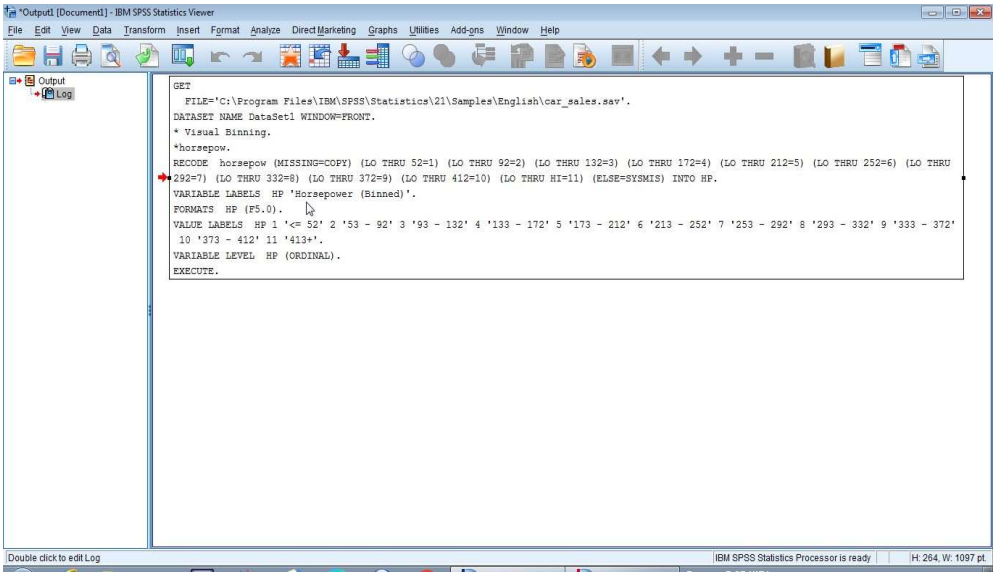
HP is new name given to variable.

Click Make Labels to display the class interval



Then OK.

A document window will open describing the execution.



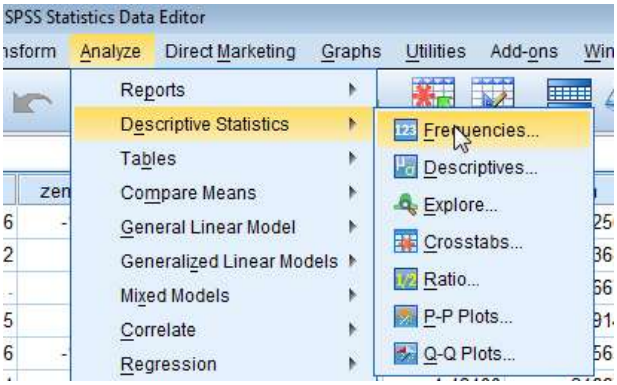
	HP
53	133 - 172
04	213 - 252
53	213 - 252
45	173 - 212
03	133 - 172
45	173 - 212
95	293 - 332
88	133 - 172
54	173 - 212
34	173 - 212
04	173 - 212

Fig.1

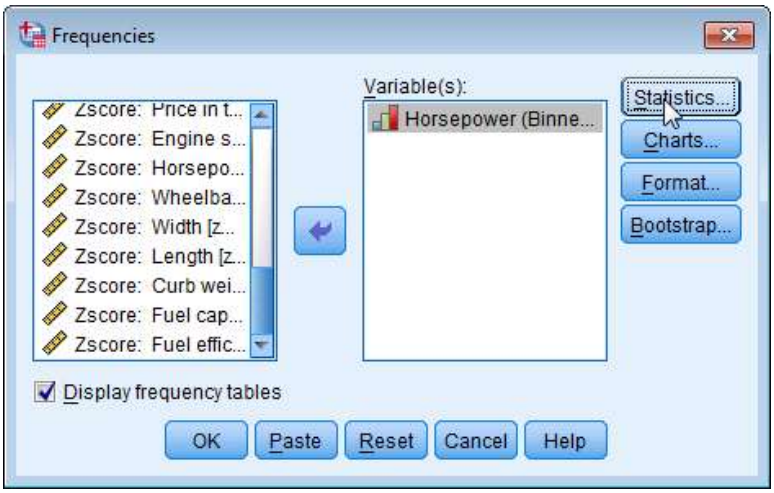
HP
4
6
6
5
4
5
8
4
5
5
5
6
5

Fig.2

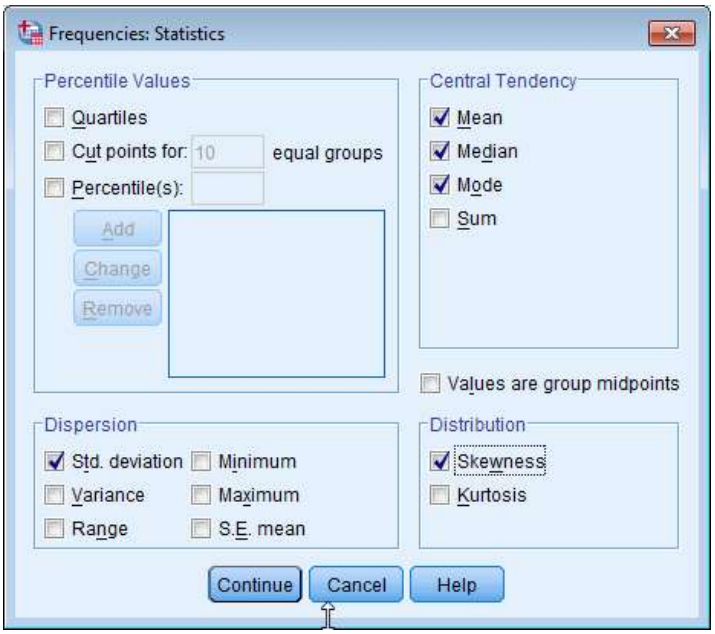
A new variable HP (Horsepower Binned)will be present in editors view. In Fig.1 the class intervals are displayed and in Fig.2 frequencies are displayed. You can toggle between these two by clicking



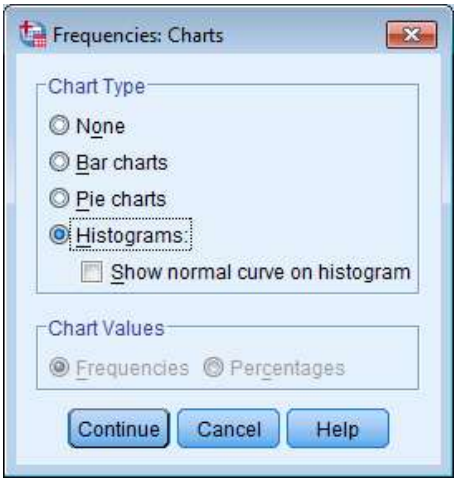
Transfer the variable for which frequencies are to be obtained.



click Statistics



select the desired options and continue. Then click Charts and select Histogram.



Doc window showing Statistical details

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\21\Samples

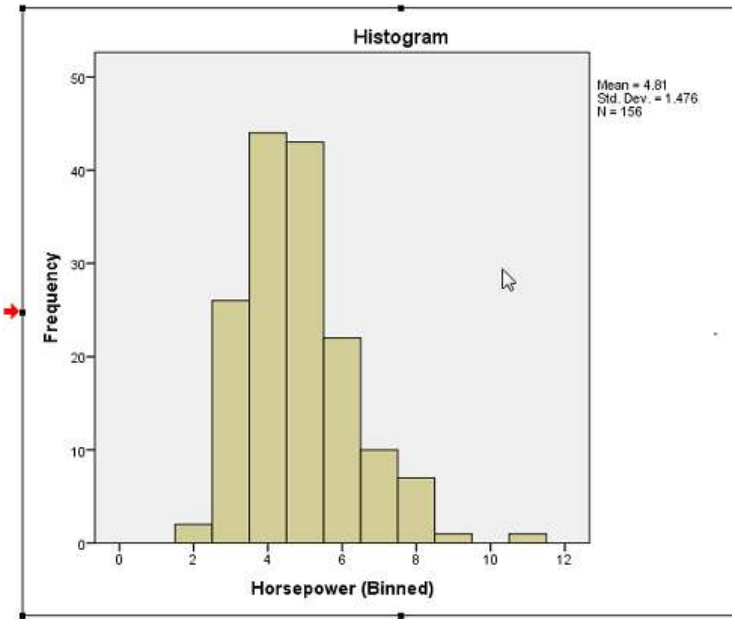
Statistics

Horsepower (Binned)

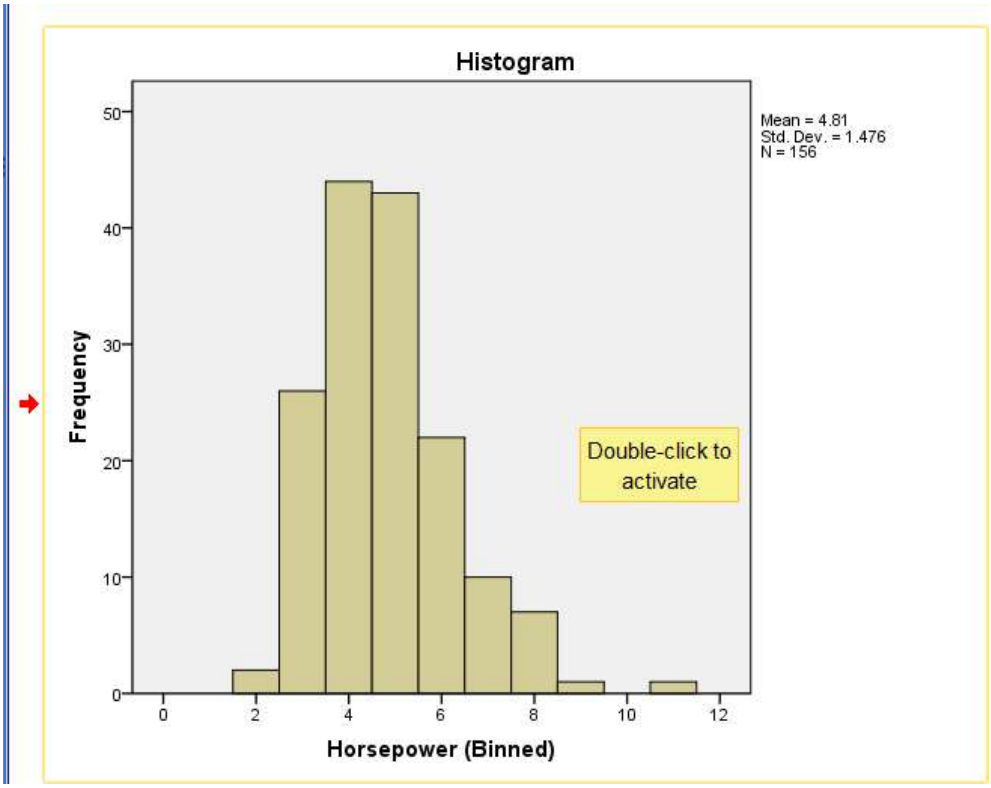
N	Valid	156
	Missing	1
Mean		4.81
Median		5.00
Mode		4
Std. Deviation		1.476
Skewness		.924
Std. Error of Skewness		.194

Horsepower (Binned)

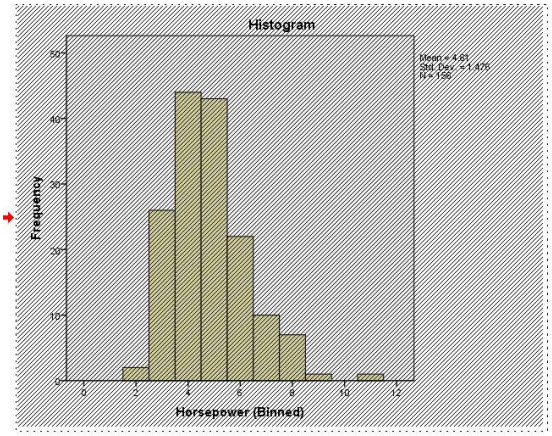
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	53 - 92	2	1.3	1.3	1.3
	93 - 132	26	16.6	16.7	17.9
	133 - 172	44	28.0	28.2	46.2
	173 - 212	43	27.4	27.6	73.7
	213 - 252	22	14.0	14.1	87.8
	253 - 292	10	6.4	6.4	94.2
	293 - 332	7	4.5	4.5	98.7
	333 - 372	1	.6	.6	99.4
	413+	1	.6	.6	100.0
	Total	156	99.4	100.0	

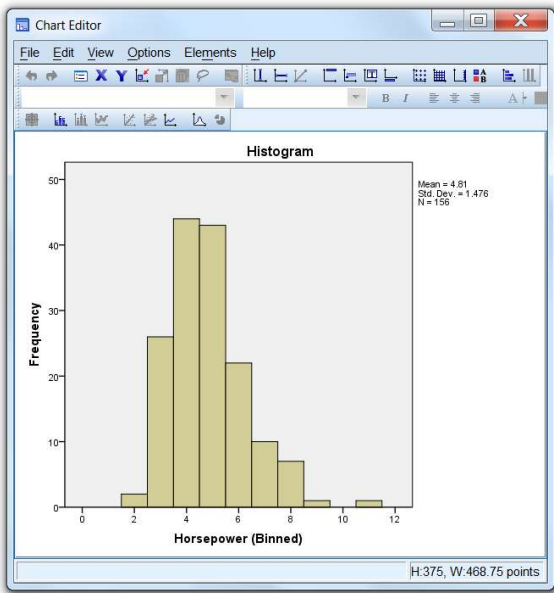


Histogram may be edited by double clicking it.



	253 - 292	10	6.4	6.4	94.2
	293 - 332	7	4.5	4.5	98.7
	333 - 372	1	.6	.6	99.4
	413+	1	.6	.6	100.0
	Total	156	99.4	100.0	
Missing	System	1	.6		
Total		157	100.0		





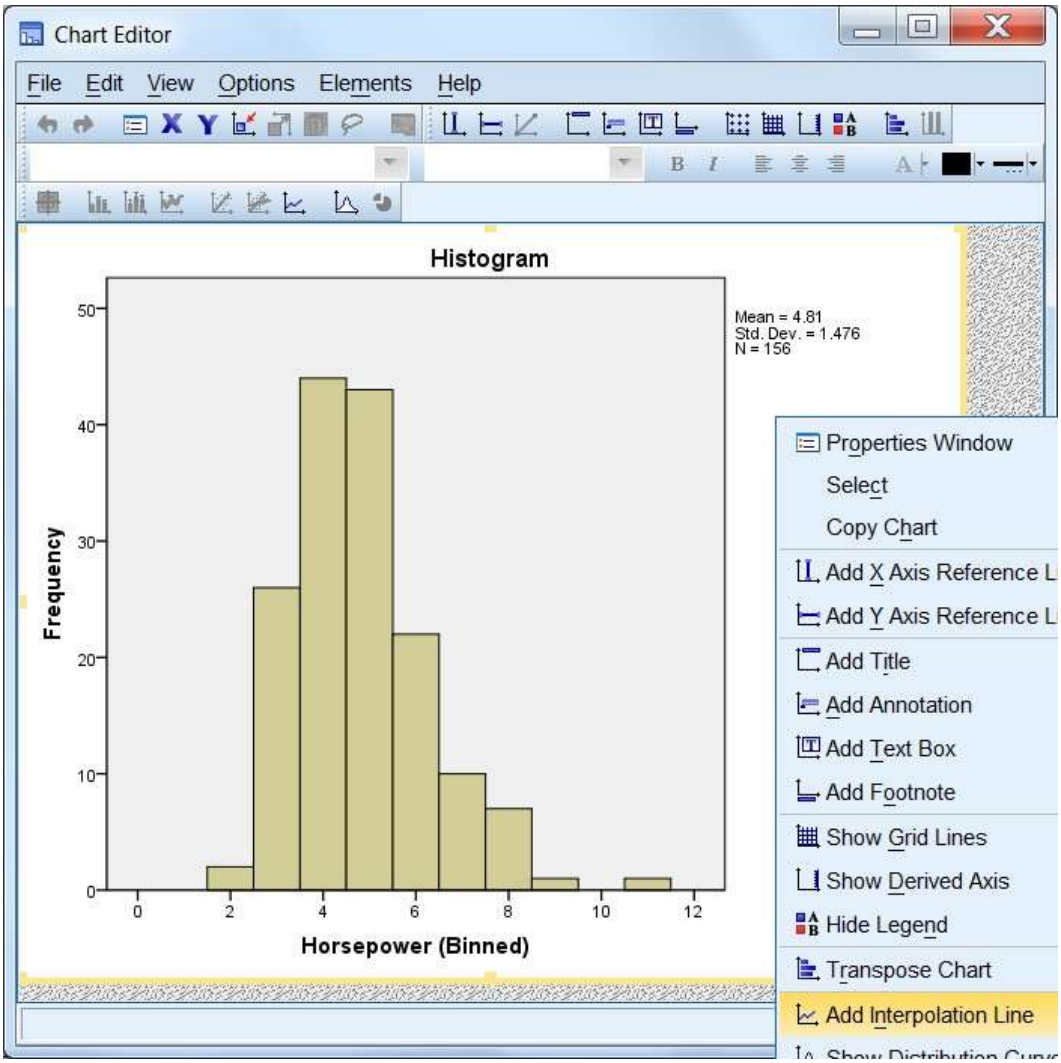
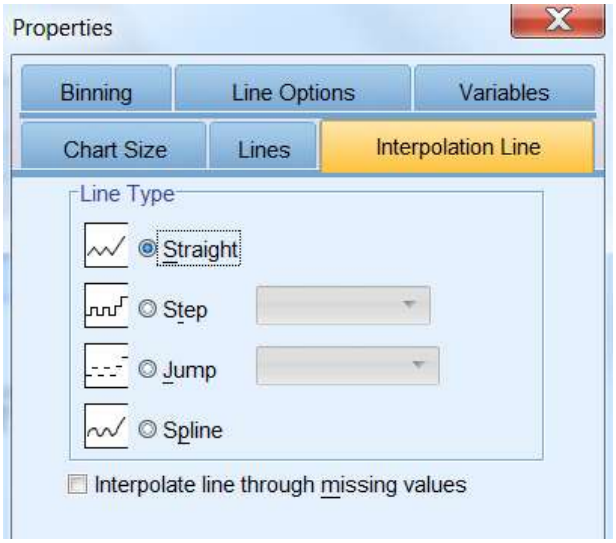
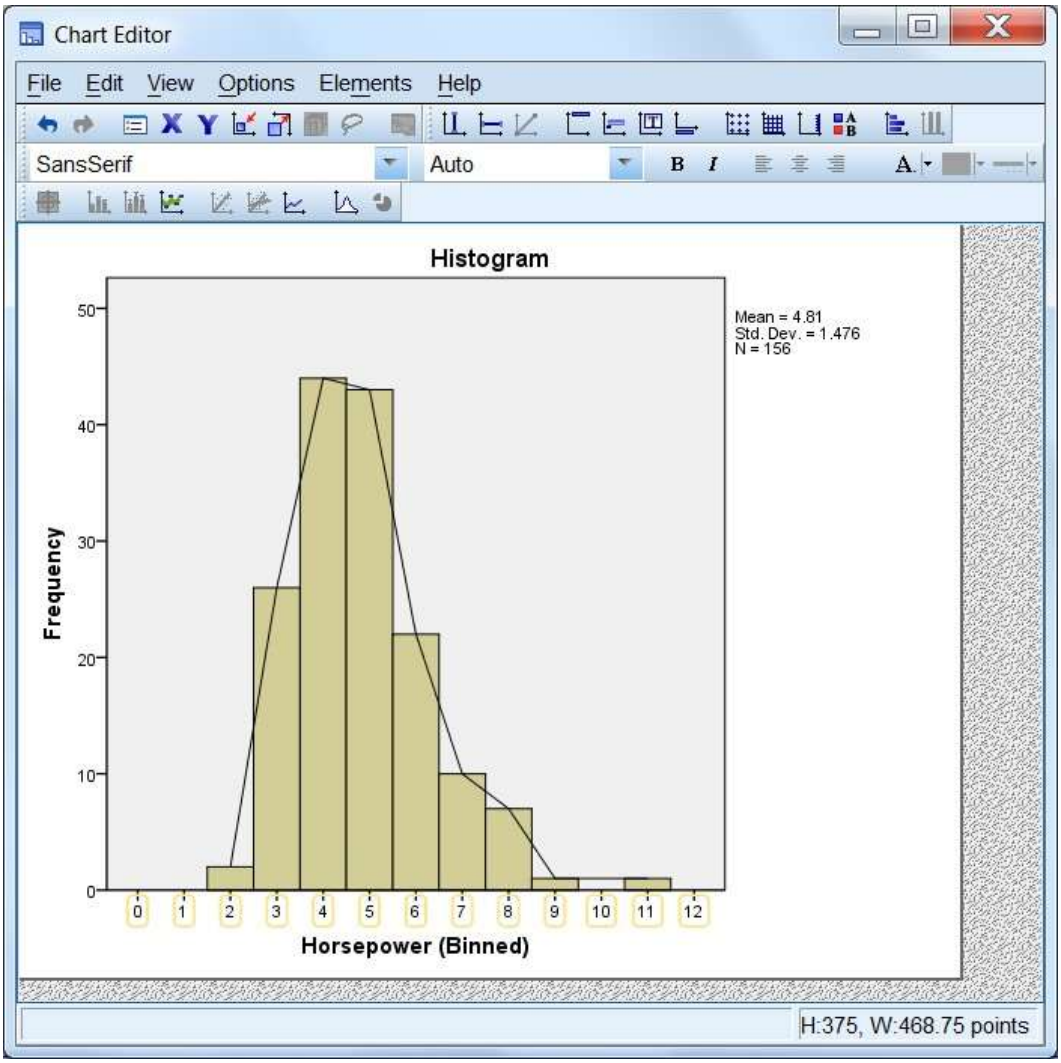


Chart Editor
File Edit View Options Elements Help
Histogram
Mean = 4.81
Std. Dev. = 1.476
N = 156
Frequency
Horsepower (Binned)
Properties Window
Select
Copy Chart
Add X Axis Reference L
Add Y Axis Reference L
Add Title
Add Annotation
Add Text Box
Add Footnote
Show Grid Lines
Show Derived Axis
Hide Legend
Transpose Chart
Add Interpolation Line


Properties
Binning
Line Options
Variables
Chart Size
Lines
Interpolation Line
Line Type
Straight
Step
Jump
Spline
Interpolate line through missing values



Conclusion:

The Data set of Car sales from repository is used, and frequency table is created. Then the Histogram and frequency curve of Horse power of engine variable is successfully drawn. Discuss about class interval and the information revealed by frequency table, histogram and frequency curve

Precautions:

1. The class interval should be decided carefully so that no critical information of variable property is lost.
2. Missing values if any should be taken care of.

Experiment - 5

Objective: Descriptive Statistics

Resources: SPSS editor, Data

Source file. **Theory And**

Methodology

The various statistical constants are to be calculated and interpreted.

$$sample\ mean\ (\bar{x}) = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Median ()

The point in the data set dividing it into two halves is called median.

Mode

It is the value which occurs with the greatest frequency (maybe more than one value). In case of grouped frequency it is given as

Variance and standard deviation

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

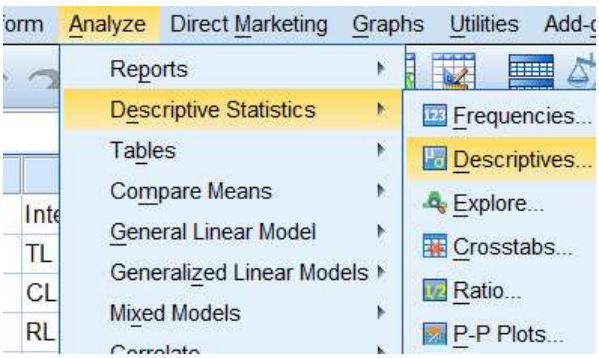
Skewness: Skewness is the measure of deviation from symmetry. The Karl-Pearson's Coefficient of Skewness is given as

$$\frac{mean - mode}{S.D.}$$

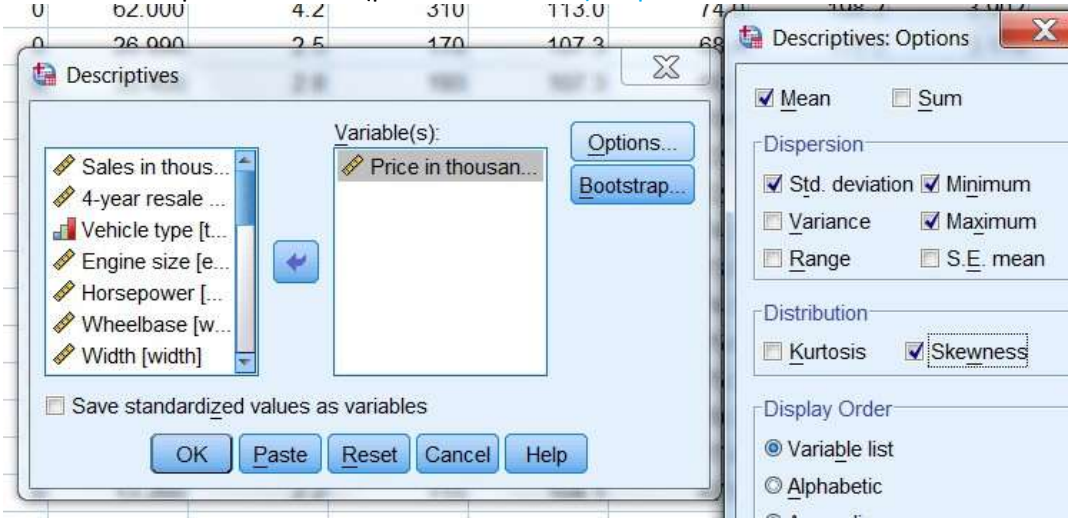
Open the desired file in SPSS editor File > Open > file path
(car_sales.sav from repository is used & the variable Price in thousand is considered for descriptive statistics)

Case1. Statistical values when all type of cases together.

Then Analyze > Descriptive Statistics > Descriptives...



Choose the required variable (price in thousand) > option and check the desired boxes.



Document output window

Descriptives

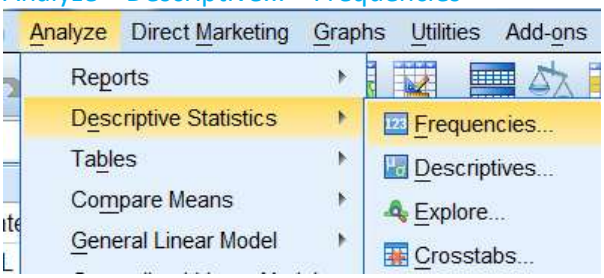
[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\car_sales.sa

Descriptive Statistics

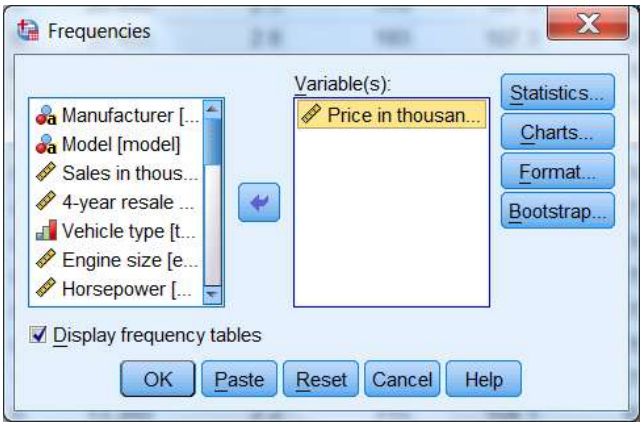
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Price in thousands	155	9.235	85.500	27.39075	14.351653	1.766	.195
Valid N (listwise)	155						

Now follow the path

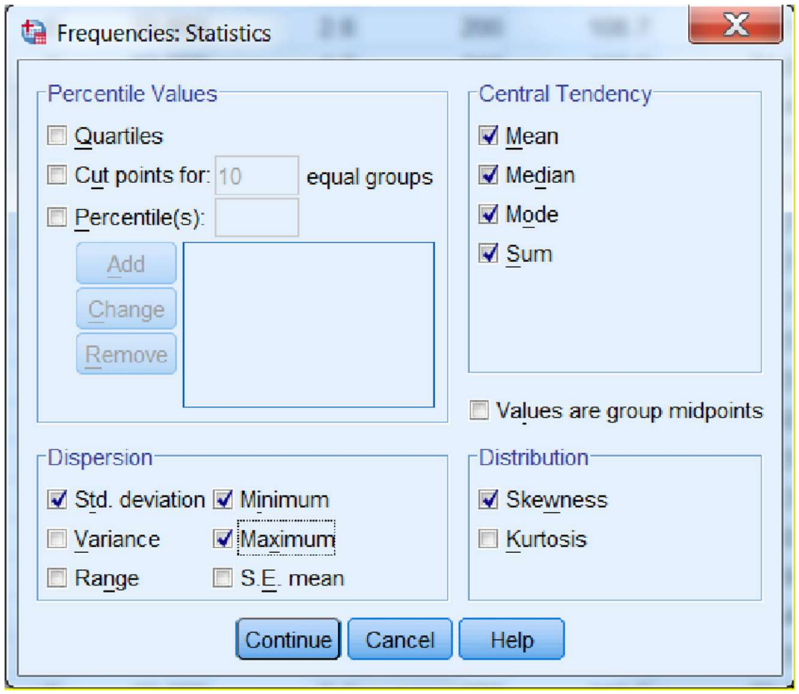
Analyze > Descriptive... > Frequencies



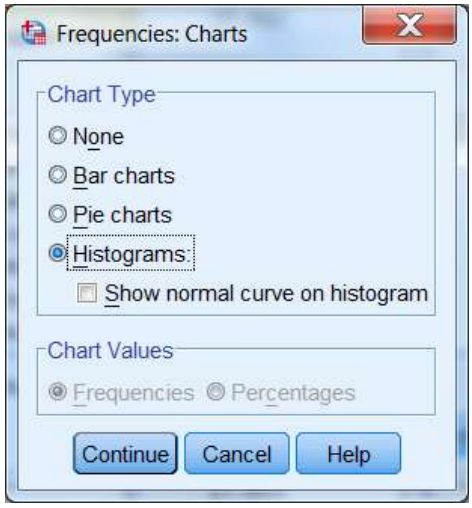
Choose desired variable



Click Statistics



Check the desired boxes, then continue. Click Charts button



select the desired chart type (if chart is not required choose None).

Document output window

Frequencies

[DataSet5] C:\Program Files\IBM\SPSS\Statistics\19\Sam

Statistics

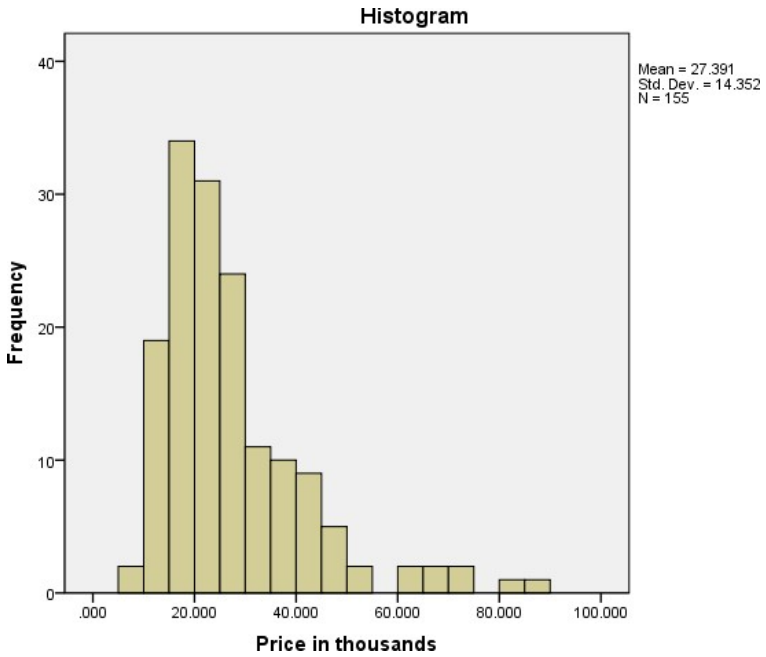
Price in thousands

N	Valid	155
	Missing	2
Mean		27.39075
Median		22.79900
Mode		12.640[a]
Std. Deviation		14.351653
Skewness		1.766
Std. Error of Skewness		.195
Minimum		9.235
Maximum		85.500
Sum		4245.567

a. Multiple modes exist. The smallest value is shown

Price in thousands

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	9.235	1	.6	.6	.6
	9.699	1	.6	.6	1.3
	10.685	1	.6	.6	1.9
	11.528	1	.6	.6	2.6

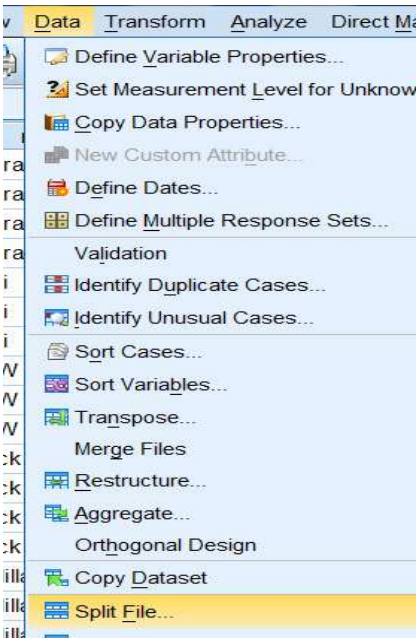


Histogram cab be edited by double clicking it.

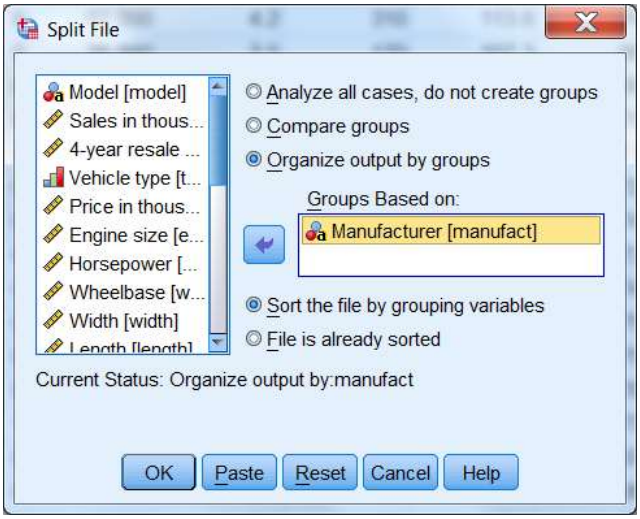
(*Observe the difference of information provided by two mode of retrieving statisticalconstants*.)

Case2. **Statistical values for all type of cases separately.**

Suppose we are interested to have the statistical details with respect to all type of manufacturer separately then follow Data > Split file



Select the desired variable (e.g. manufacturer) for which the splitter information is to be obtained, also select organize output by groups.



Doc output window

Descriptives

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\Eng

Manufacturer = Acura Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Price in thousands	3	21.500	42.000	30.63333	10.430884
Valid N (listwise)	3				

a. Manufacturer = Acura

Manufacturer = Audi Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Price in thousands	3	23.990	62.000	39.98000	19.709406
Valid N (listwise)	3				

a. Manufacturer = Audi

Manufacturer = BMW Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Price in thousands	3	26.990	38.900	33.09667	5.960791
Valid N (listwise)	3				

a. Manufacturer = BMW

Manufacturer = Buick Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Price in thousands	4	21.975	31.965	26.78125	4.218359
Valid N (listwise)	4				

Conclusion:

The Data set of Car sales from repository is used, and Descriptive statistical table is created. Also the (name of the chart) (Histogram) of Price in thousand variable is successfully drawn. Separate statistical summary of different manufacturers are also obtained

Precautions:

1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.

Experiment - 6

Objective: Correlation between two random variables.

Resources: SPSS editor, Data Source file.

Theory And Methodology

As a measure of intensity or degree of linear relationship between two variables Karl Pearson coefficient of correlation denoted as r_{xy} or $r(X, Y)$ between two random variables X & Y s defined as

If (x_i, y_i) , $i = 1,2, \ldots n$ is the bivariate distribution, then

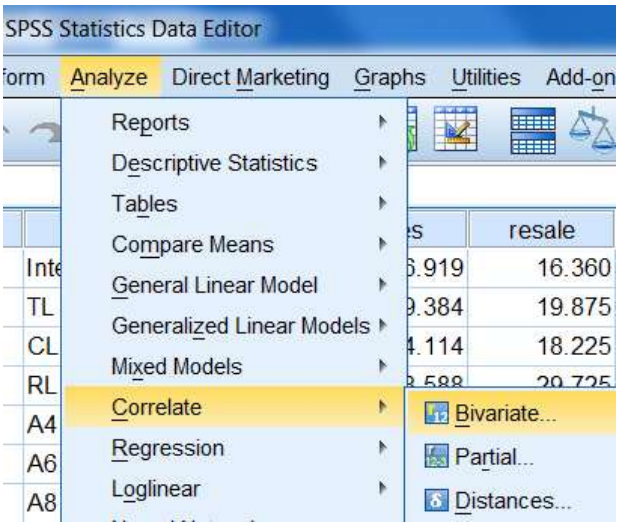
$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

Open the desired file in SPSS editor File > Open > file path

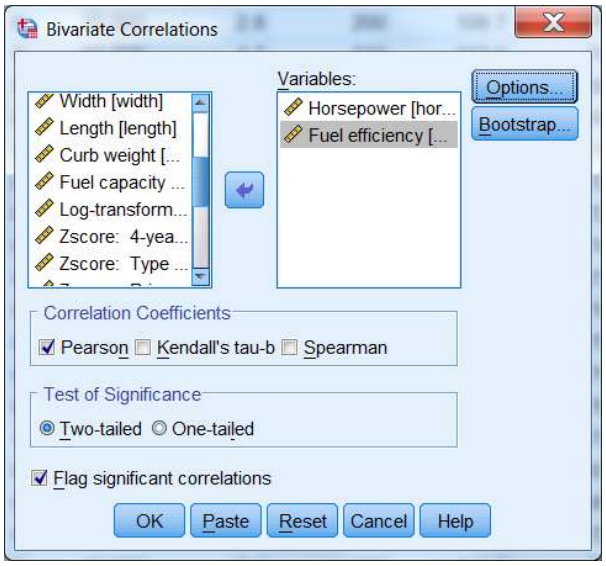
(car_sales.sav from repository is used and the variables Horsepower of the engine & Fuel efficiency is considered)

Case1. Missing values dealt as Exclude cases pairwise

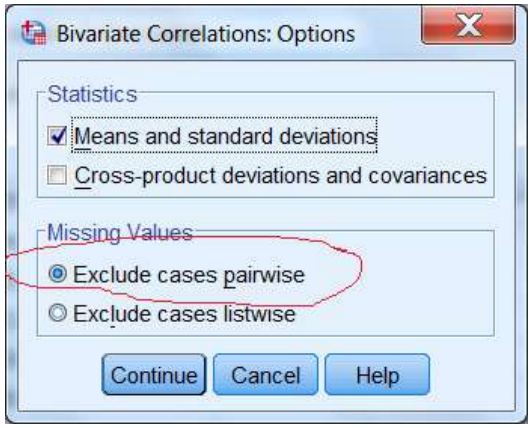
Now follow the path Analyze > Correlate > Bivariate



Choose the desired variable



Click option and check the desired boxes.



(Observe the selection for missing values)

Document output window

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
Horsepower	185.95	56.700	156
Fuel efficiency	23.84	4.283	154

Correlations

		Horsepower	Fuel efficiency
Horsepower	Pearson Correlation	1	-.611**
	Sig. (2-tailed)		.000
	N	156	154
Fuel efficiency	Pearson Correlation	-.611**	1
	Sig. (2-tailed)	.000	
	N	154	154

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

Observe in the descriptive statistics number of entry in Horse power and Fuel efficiency is 156 and 154 respectively. The total data points are extended to 157 points.

Case2. Missing values dealt as Exclude cases listwise

Now follow the path Analyze > Correlate > Bivariate

Correlations

```
[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\S
```

Descriptive Statistics

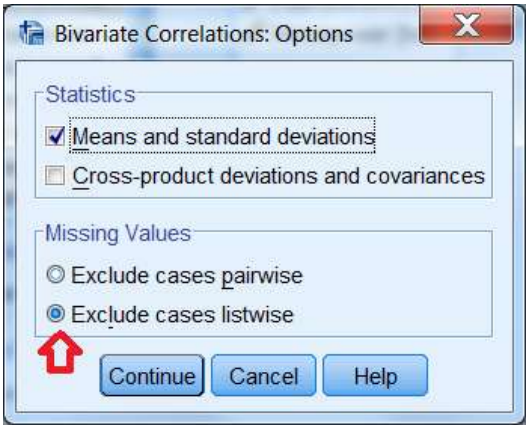
	Mean	Std. Deviation	N
Fuel efficiency	23.84	4.283	154
Horsepower	185.66	57.006	154

Correlations[a]

		Fuel efficiency	Horsepower
Fuel efficiency	Pearson Correlation	1	-.611**
	Sig. (2-tailed)		.000
Horsepower	Pearson Correlation	-.611**	1
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).
a. Listwise N=154

Choose Exclude cases listwise option for missing values



Document output window

Case3. Missing values dealt before calculation of correlation

(Note : if the missing values are dealt before finding the correlation by series mean, then the total number of cases has become 157)

Document output window

Replace Missing Values

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\car_sales.sa

Result Variables

	Result Variable	N of Replaced Missing Values	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1	horsepow_1	1	1	157	157	SMEAN (horsepow)
2	mpg_1	3	1	157	157	SMEAN(mpg)

Correlations

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\car_sales.sa

Descriptive Statistics

	Mean	Std. Deviation	N
SMEAN(horsepow)	185.95	56.518	157
SMEAN(mpg)	23.84	4.241	157

Correlations

		SMEAN (horsepow)	SMEAN(mpg)
SMEAN(horsepow)	Pearson Correlation	1	-.611**
	Sig. (2-tailed)		.000
	N	157	157
SMEAN(mpg)	Pearson Correlation	-.611**	1
	Sig. (2-tailed)	.000	
	N	157	157

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

Conclusion:

The Data set of Car sales from repository is used, and correlation between the variables () Horse power and Fuel efficiency is found to be …….. Explain further the correlation coefficient value according to sign and its value.

Precautions:

1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.

Experiment - 7

Objective: Regression Analysis

Resources: SPSS editor, Data

Source file.**Theory And**

Methodology

Let (x_i, y_i) , $i = 1,2, \dots n$ be a bivariate sample, assuming one of them Independent say(x) and other dependent on first say(y), we predict the value of y by fitting a curve to data. If the curve is line then it is called line of regression and there is said to be a linear regression.

The curve fit for Y on x is called regression curve of y on x and similarly curve fit for X on y is called regression curve of x on y

Let the line of regression of $Y \; on \; x$ be

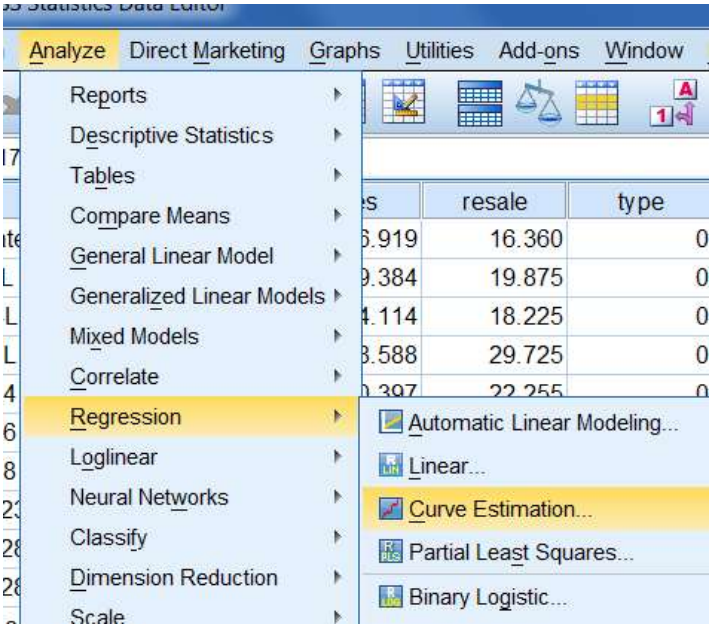
$$Y = a + bx$$

Open the desired file in SPSS editor File > Open > file path

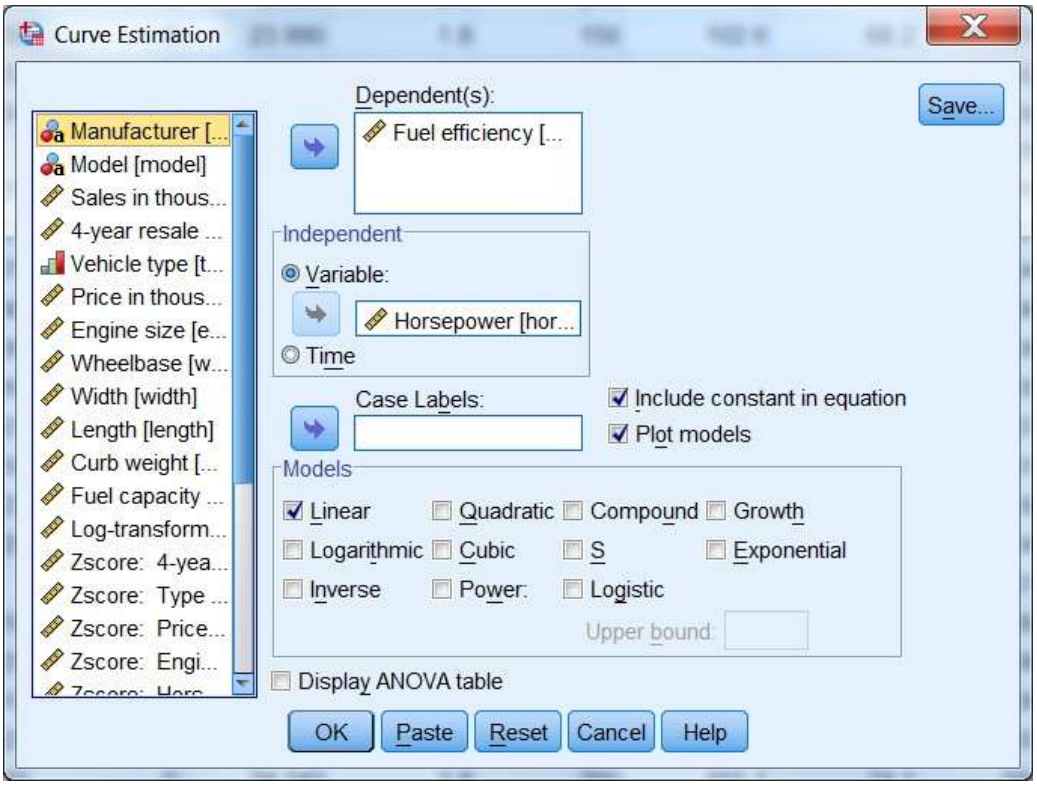
(car_sales.sav from repository is used and the variables Horsepower of the engine & Fuel efficiency is considered)

Case1. Using curve fitting path

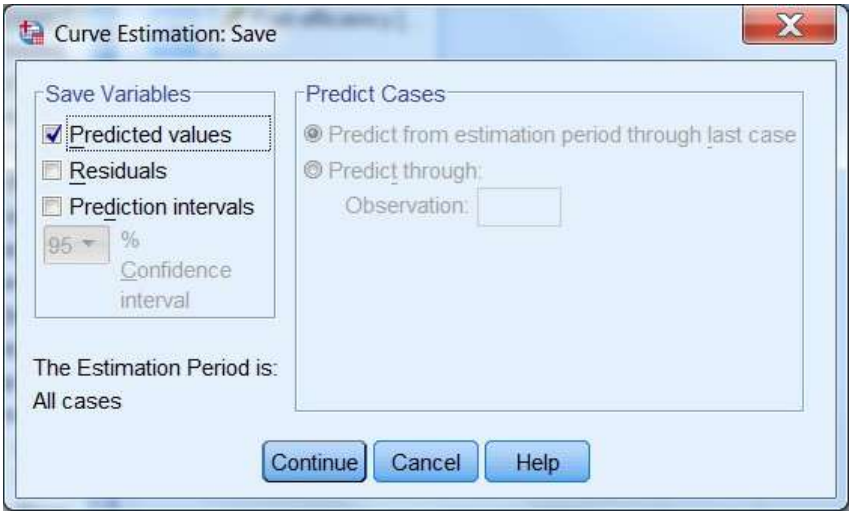
Now follow the path Analyze > Regression > Curve Estimation



Choose dependent and independent variables and check the other desired boxes (Linear Model). Click save



Check the predicted values



Document output window

Curve Fit

Model Description

Model Name		MOD_1
Dependent Variable	1	Fuel efficiency
Equation	1	Linear
Independent Variable		Horsepower
Constant		Included
Variable Whose Values Label Observations in Plots		Unspecified

Case Processing Summary

	N
Total Cases	157
Excluded Cases[a]	3
Forecasted Cases	0
Newly Created Cases	0

a. Cases with a missing value in any variable are excluded from the analysis.

Variable Processing Summary

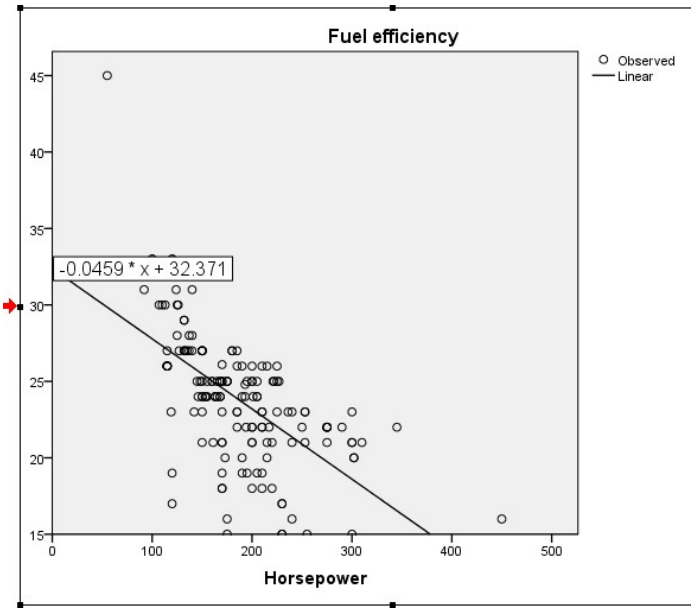
		Variables	
		Dependent	Independent
		Fuel efficiency	Horsepower
Number of Positive Values		154	156
Number of Zeros		0	0
Number of Negative Values		0	0
Number of Missing Values	User-Missing	0	0
	System-Missing	3	1

Model Summary and Parameter Estimates

Dependent Variable:Fuel efficiency

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.374	90.743	1	152	.000	32.371	-.046

The independent variable is Horsepower.

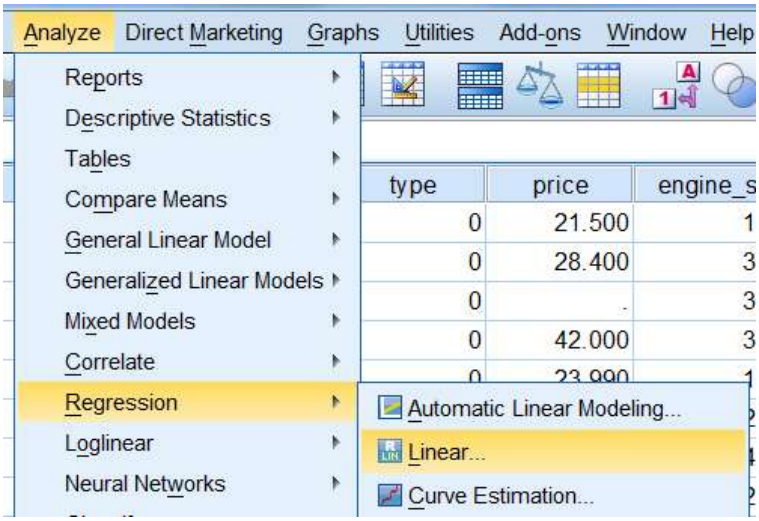


Also in data sheet a new variable for predicted values has been created.

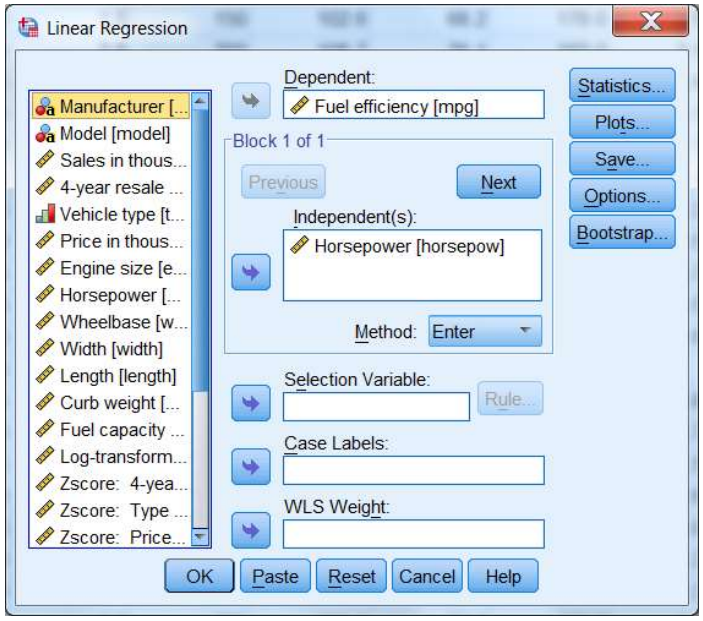
	horsepow	FIT_1
8	140	25.94066
2	225	22.03627
2	225	22.03627
5	210	22.72528
8	150	25.48132
8	200	23.18462
2	310	18.13189
5	170	24.56264
8	193	23.50616

Case2. Using Linear Model approach

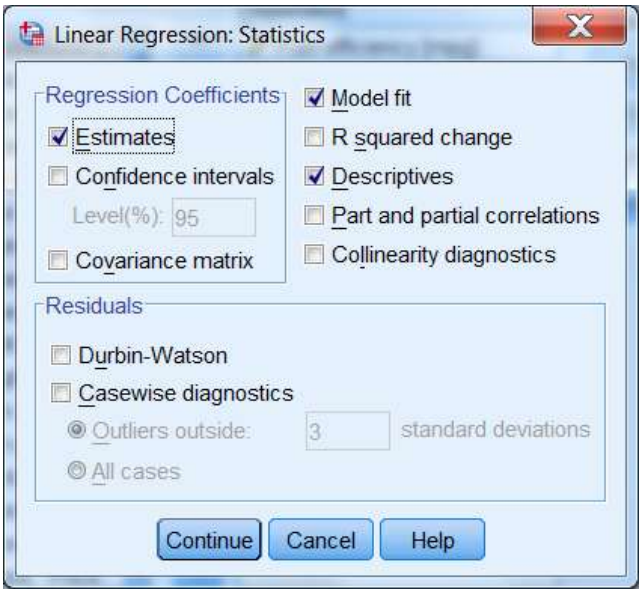
Now follow the path Analyze > Regression > Linear



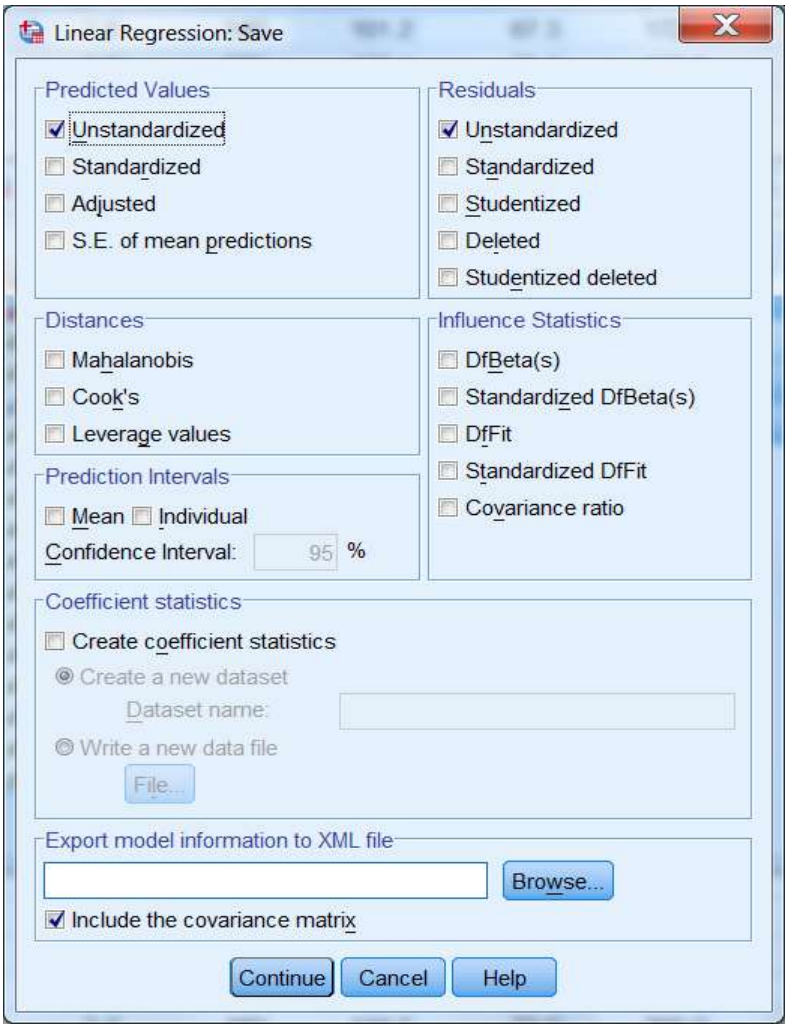
Choose dependent and independent variables and click statistics



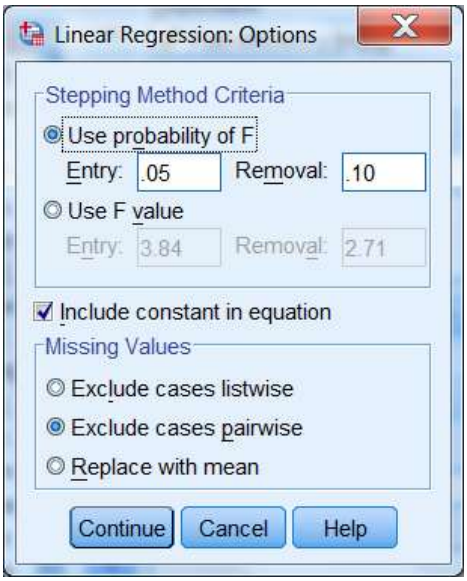
Check the desired boxes and continue



Click Save and check the desired boxes



Click Option



Decide carefully for treating the missing values.

Doc output window

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Fuel efficiency	23.84	4.283	154
Horsepower	185.95	56.700	156

Correlations

		Fuel efficiency	Horsepower
Pearson Correlation	Fuel efficiency	1.000	-.611
	Horsepower	-.611	1.000
Sig. (1-tailed)	Fuel efficiency	.	.000
	Horsepower	.000	.
N	Fuel efficiency	154	154
	Horsepower	154	156

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Horsepower[a]	.	Enter

a. All requested variables entered.
b. Dependent Variable: Fuel efficiency

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.611[a]	.374	.370	3.400

a. Predictors: (Constant), Horsepower
b. Dependent Variable: Fuel efficiency

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1049.053	1	1049.053	90.743	.000[a]
	Residual	1757.226	152	11.561		
	Total	2806.279	153			

a. Predictors: (Constant), Horsepower
b. Dependent Variable: Fuel efficiency

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	32.431	.942		34.421	.000
	Horsepower	-.046	.005	-.611	-9.526	.000

a. Dependent Variable: Fuel efficiency

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	11.65	29.89	23.84	2.619	156
Residual	-9.889	15.109	-.014	3.389	154
Std. Predicted Value	-4.657	2.309	.000	1.000	156
Std. Residual	-2.908	4.444	-.004	.997	154

a. Dependent Variable: Fuel efficiency

Two new variables for predicted values (PRE_1) and difference between predicted and actual (RES_1) have been created.

PRE_1	RES_1
25.96548	2.03452
22.04006	2.95994
22.04006	3.95994
22.73278	-.73278
25.50367	1.49633
23.19460	-1.19460

(There are various plot options also available and can be drawn as per need under plot button)

Conclusion:

The Data set of Car sales from repository is used, and a regression line fitted between the variables (Horse power-independent variable and Fuel efficiency-dependent variable). the value of constant is.... and coefficient is .....Explain further the regression line properties exhibited in the case.

Precautions:

1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.

Experiment - 8

Objective: Hypothesis Testing 't' -

test **Resources:** SPSS editor, Data

Source file. **Theory And**

Methodology

Null Hypothesis **:** Any hypothesis we wish to test and is denoted by H_0 (hypothesis of no difference)

Alternative Hypothesis **:** Any hypothesis which is complementary to null hypothesis is called and alternative hypothesis, denoted by H_1

't' - test for single mean

If a random sample x_i , $i = 1,2, \ldots n$ of size 'n' and is drawn from a normal population with mean μ then we have the statistic

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \quad , \quad or \quad t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$$

't' - test for difference of means

Suppose we want to test if two independent samples x_i , $i = 1,2, \ldots n_1$ & y_j , $j = 1,2, \ldots n_2$ had been drawn from to normal populations with mean μ_X & μ_Y respectively, under the assumption that the population variance are equal,i.e.; $\sigma^2 = \sigma^2 = \sigma^2$ then test statistic

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If samples are drawn from the population with same mean i.e.; $H_0 = \mu_X = \mu_Y$

$t = \frac{(\bar{x} - \bar{y})}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ And degree of freedom (df) $v = n_1 + n_2 - 2$.

Paired 't'- test for difference of means

$\bar{d} = \frac{\sum di}{n}$ $S^2 = \frac{1}{n-1}\sum^{n}_{i=1} (d_i - \bar{d})^2 = \frac{1}{n-1}\left[\sum d_i^2 - \frac{(\sum di)^2}{n}\right]$ with $n - 1$ df

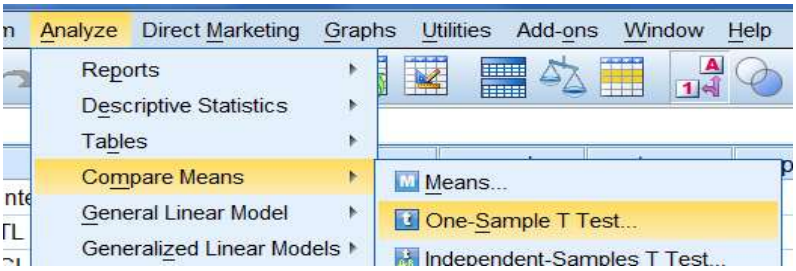
Case1. Test for single mean

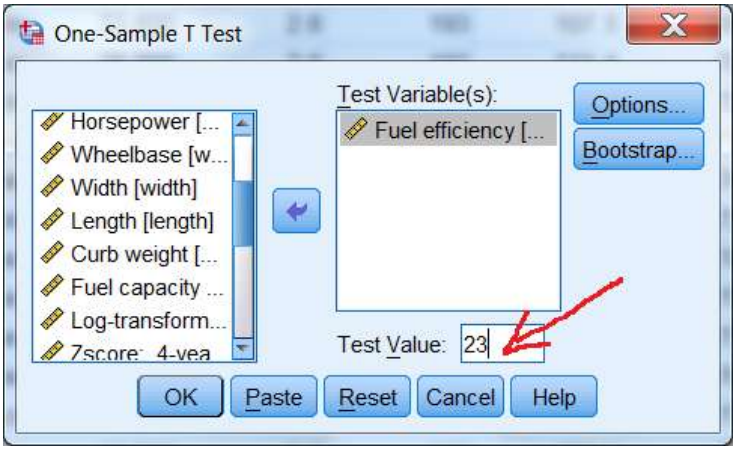
Open the desired file in SPSS editor File > Open > file path

(car_sales.sav from repository is used and the variables Fuel efficiency is considered)

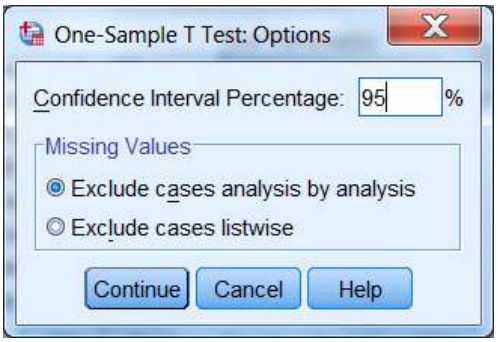
$H_0 : Average\ is\ 23\ km/lt$ or ($\mu = 23\ km\ per\ lt$) $H_1 : \mu \neq 23\ km\ per\ lt$)

Now follow the path Analyze > Compare Means > One Sample t-test





Assign the test value.



fix the confidence interval.

T-Test

```
[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\car_sales.
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Fuel efficiency	154	23.84	4.283	.345

One-Sample Test

	Test Value = 23					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Fuel efficiency	2.444	153	.016	.844	.16	1.53

Note the p value is 0.016 which is less than the assigned value 0.05 hence null hypothesis is rejected, ie., claim $\mu = 23\ km\ per\ lt$ can not be accepted.

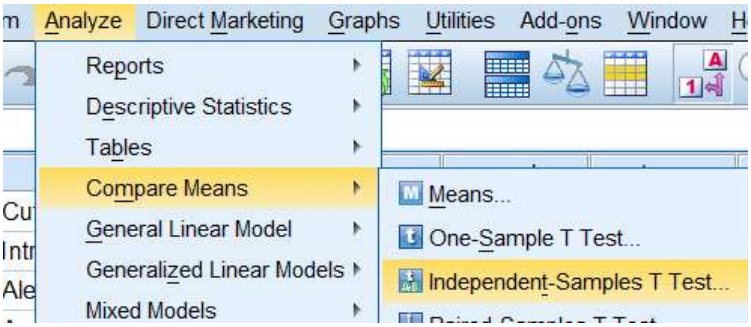
Case2. Test for difference of means

Open the desired file in SPSS editor File > Open > file path

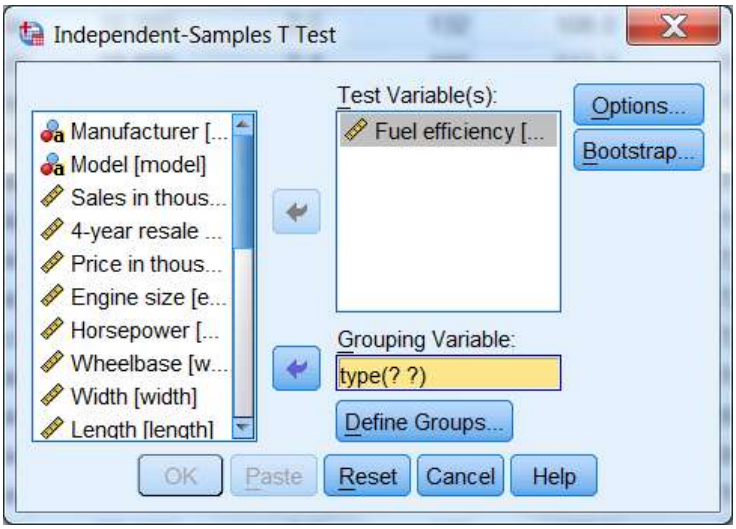
(car_sales.sav from repository is used and the variables Fuel efficiency is considered for testing of means w.r.t. types of vehicle ie. Automobile & Truck)

$H_0 : Fuel\ efficiency\ of\ Automobiles = Fuel\ efficiency\ of\ Trucks\ \ (\mu_A = \mu_T)$

Now follow the path Analyze > Compare Means > Independent Sample test

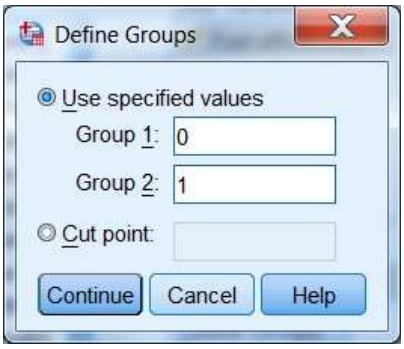


Transfer the the desired variable to test variable then transfer the variable w.r.t. which the means are to be tested (in present case Automobiles & truck are considered)

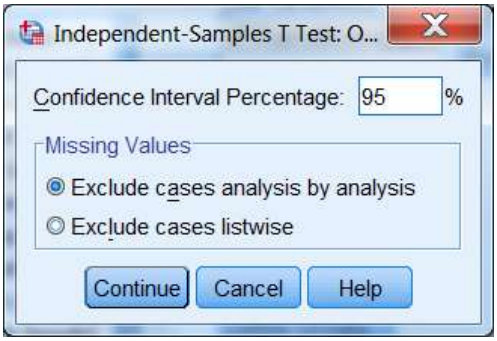


click define group

(Note: in the variable view the value can be assigned)



Enter the desired level of significance



T-Test

```
[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\car_sales.sav
```

Group Statistics

	Vehicle type	N	Mean	Std. Deviation	Std. Error Mean
Fuel efficiency	Automobile	114	25.30	3.646	.341
	Truck	40	19.70	3.107	.491

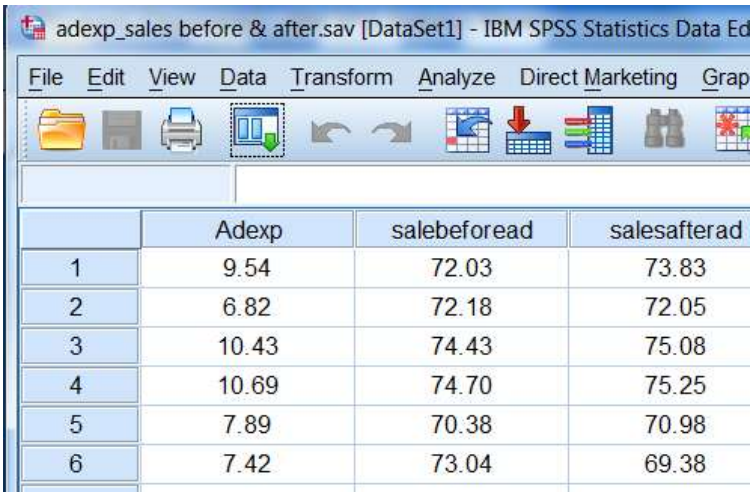
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Fuel efficiency	Equal variances assumed	.004	.948	8.664	152	.000	5.597	.646	4.321	6.874
	Equal variances not assumed			9.356	79.405	.000	5.597	.598	4.407	6.788

(Levene's test is to decide the equivalence of the variances) in the present case in first row reading of p-value is 0.948 > 0.05 hence equivalence of variances can be accepted (Now we will read all information from first row). Also the for t-test p-value is almost zero <0.05 hence null hypothesis is rejected.

Case3. Paired 't'- test for difference of means

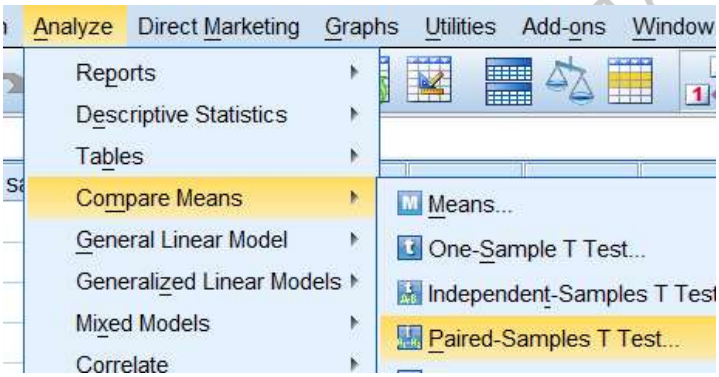
Open the desired file in SPSS editor File > Open > file path
(adexp_salesbefore & after.sav is used) sales before & after the ad is to be tested.

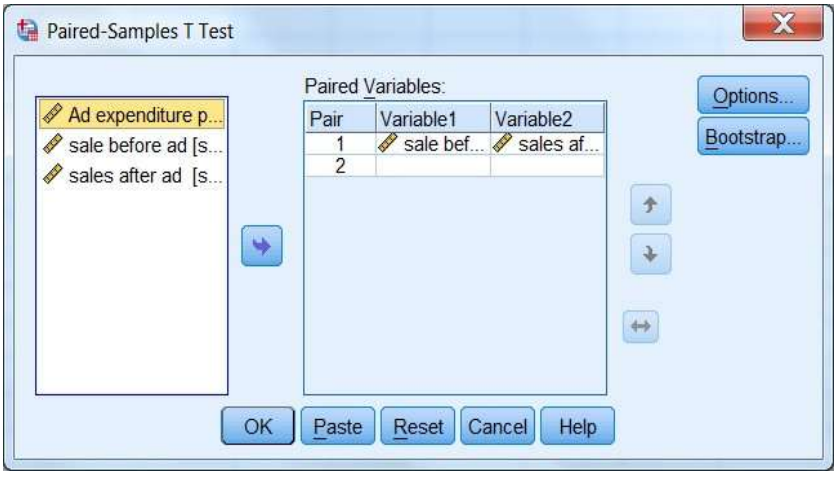


	Adexp	salebeforead	salesafterad
1	9.54	72.03	73.83
2	6.82	72.18	72.05
3	10.43	74.43	75.08
4	10.69	74.70	75.25
5	7.89	70.38	70.98
6	7.42	73.04	69.38

$H_0 : avg\ sales\ before\ ad = avg\ sales\ after\ ad$ ($\mu_b = \mu_a$) or { $\mu_b - \mu_a = 0$ }

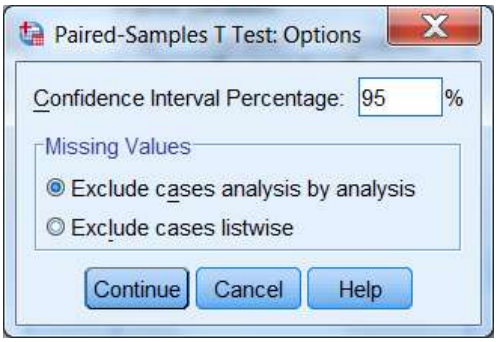
Now follow the path Analyze > Compare Means > Paired-Samples T-test





Transfer the testing variables (Sales before & sales after)

Click options



enter the desired level significance.

Doc output window

T-Test

[DataSet1] C:\Users\Ramesh\Desktop\adexp_sales before & after.sav

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sale before ad	73.5415	20	1.61219	.36050
	sales after ad	74.0845	20	3.62275	.81007

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	sale before ad & sales after ad	20	.714	.000

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	sale before ad - sales after ad	-.54300	2.71706	.60755	-1.81462	.72862	-.894	19	.383

Since the p-value 0.383 which is greater than 0.05 ie, average sales befor and after ad does not differ significantly.Hence null hypothesis can not be rejected.

Conclusion:

Give information of data set used for testing. Also mentioning the p-value write the decision about rejection or non-rejection of null hypothesis.

Precautions:

1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.
3. Null hypothesis is defined carefully and level of significance chosen appropriately.

Experiment - 9

Objective: Chi Square test

Resources: SPSS editor, Data Source file.

Theory And Methodology

Chi Square test in SPSS is test for association (or independence) of two nominal variables.

Case1. Chi-squared test of independence in contingency tables

A cell containing information of two characteristics resulting into a bivariate data. The table displaying data is called contingency table. Horizontal and vertical display of data is referred as $r \times c$ table.

Our interest is to see whether two characteristics are independent, thus

H_0: Two characteristics are independent. H_1: they are dependent.

O_{ij} : observed frequency of i[th] row and j[th] column respectively.

e_{ij} : expected frequency of i[th] row and j[th] column respectively.

Then

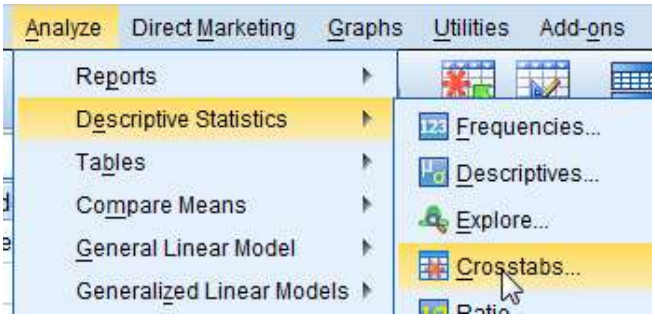
$So, \quad \chi^2 = \sum_i \sum_j \frac{(O_{ij}-e_{ij})^2}{e_{ij}}$ with $(r-1) \times (c-1)$ degree of freedom (df).

Open the desired file in SPSS editor File > Open > file path

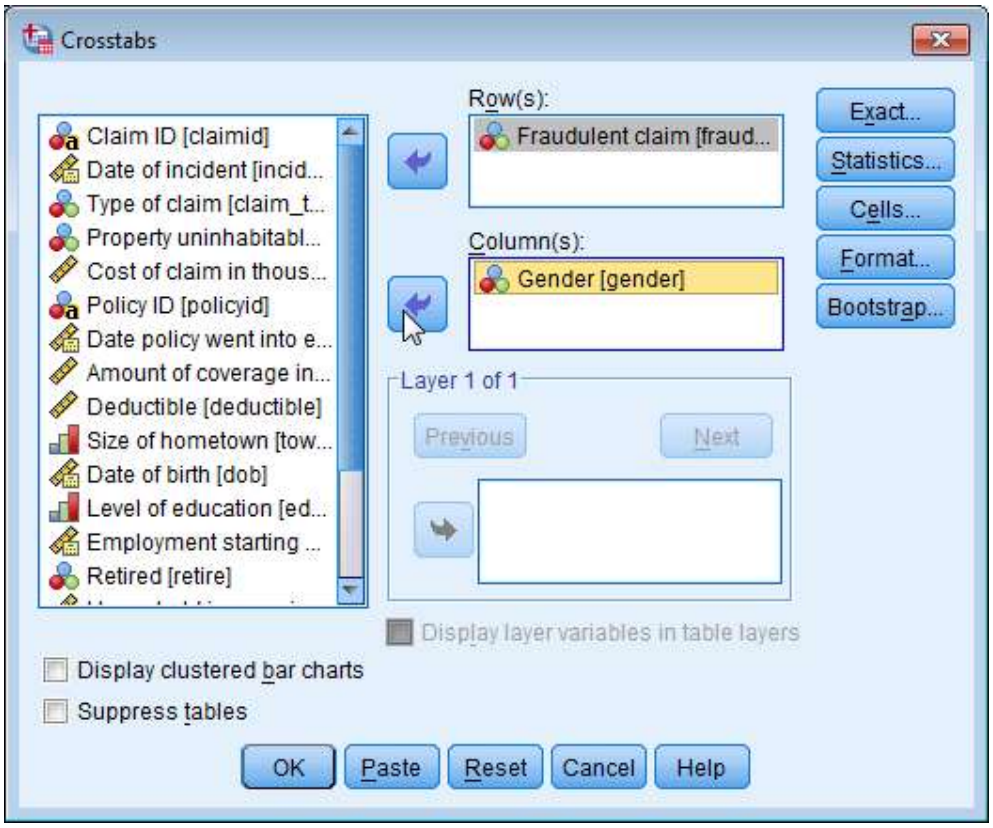
H_0: Fraudulent claims are independent of Gender

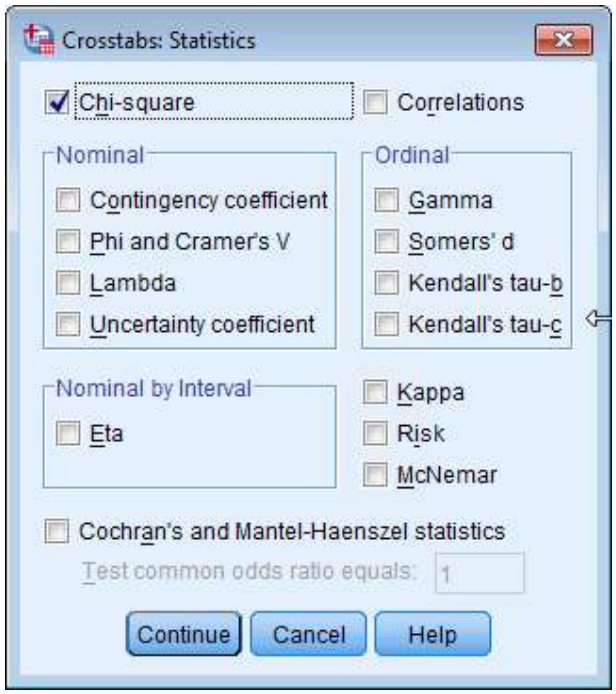
H_1: Fraudulent claim is Gender dependent

Now follow the path Analyze > Descriptive Statistics > Crosstabs

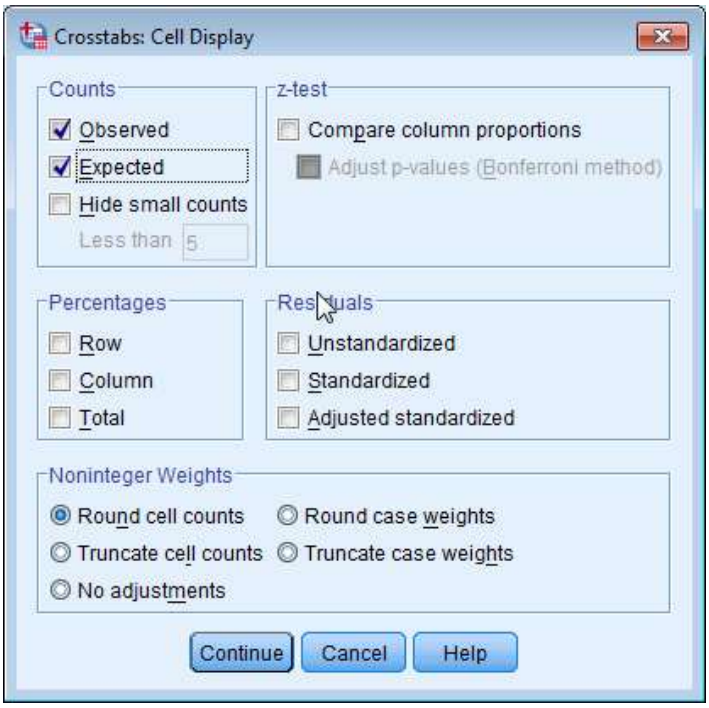


Transfer the desired variables in rows & column box (their order does not matter).





click continue. Then click Cells button



Observed (under count) is checked by default, select Expected.Then OK

Doc output window

Crosstabs

[DataSet2] C:\Program Files\IBM\SPSS\Statistics\21\Samples\English\insu

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Fraudulent claim * Gender	4415	100.0%	0	0.0%	4415	100.0%

Fraudulent claim * Gender Crosstabulation

			Gender		
			Male	Female	Total
Fraudulent claim	No	Count	1964	1988	3952
		Expected Count	1946.9	2005.1	3952.0
	Yes	Count	211	252	463
		Expected Count	228.1	234.9	463.0
Total		Count	2175	2240	4415
		Expected Count	2175.0	2240.0	4415.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.820[a]	1	.093		
Continuity Correction[b]	2.657	1	.103		
Likelihood Ratio	2.824	1	.093		
Fisher's Exact Test				.095	.051
Linear-by-Linear Association	2.819	1	.093		
N of Valid Cases	4415				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 228.09.

b. Computed only for a 2x2 table

The Pearson Chi-Square p-value 0.093 is greater than 0.05 , therefore Null Hypothesis is not rejected. It is concluded thet fraudulent claim is not gender biased.

Case2. Chi-square test for goodness of fit

If a population has a specified theoretical distribution. The test is based onhow good a fit we have between the frequency of occurrence of observationsin an observed sample and the expected frequencies obtained from the hypothesized distribution.

A goodness of fit test between observed and expected frequencies is based on the quantity

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \ ,$$

O_i: observed frequency of the cell e_i: expected frequency of the cell

cell : each possible out comes of the experiment.

Where χ^2 is a value of a random variable whose sampling distribution is approximated very closely by the Chi square distribution with $\nu = k - 1$ degree of freedom (df).

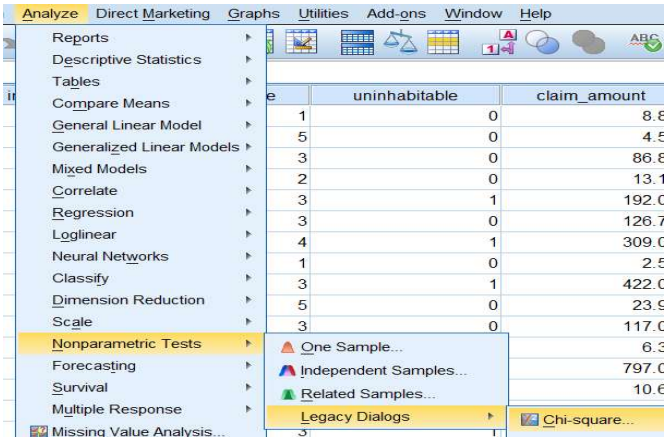
Critical values

If the observed frequencies differ considerately from the expected frequencies, the χ^2 value will be large and fit is poor. A good fit leads to the acceptance of H_0, where as a poor fit leads to rejection.

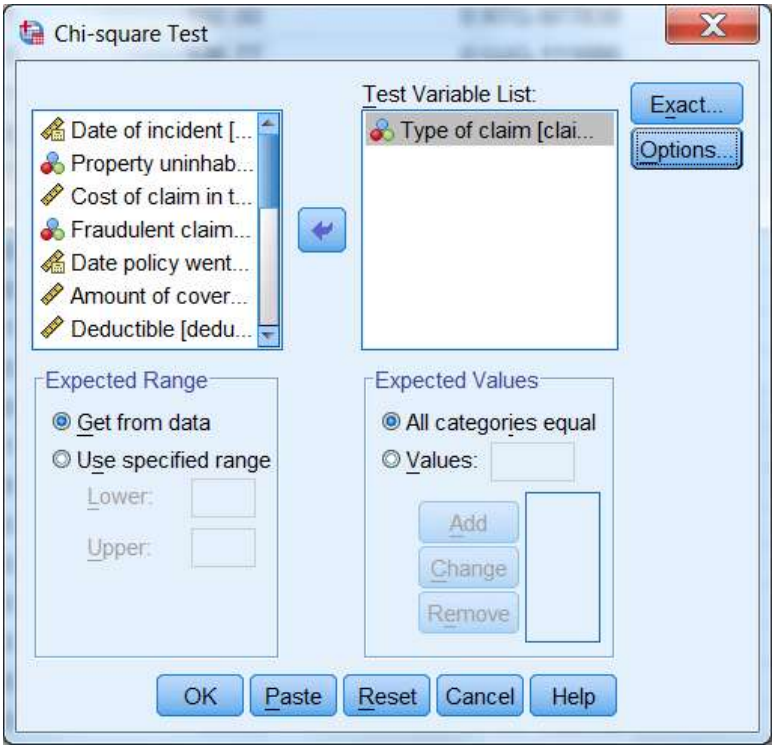
(insurance_claims.sav from repository is used and Claim type is considered)

H_0: Claims are equally distributed H_1: It is not equally distributed

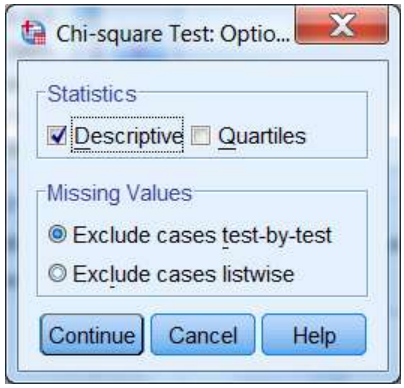
Now follow the path Analyze > Nonparametric Tests > Legacy dialogs > Chi Square



Transfer the desired variable to Test variable list box

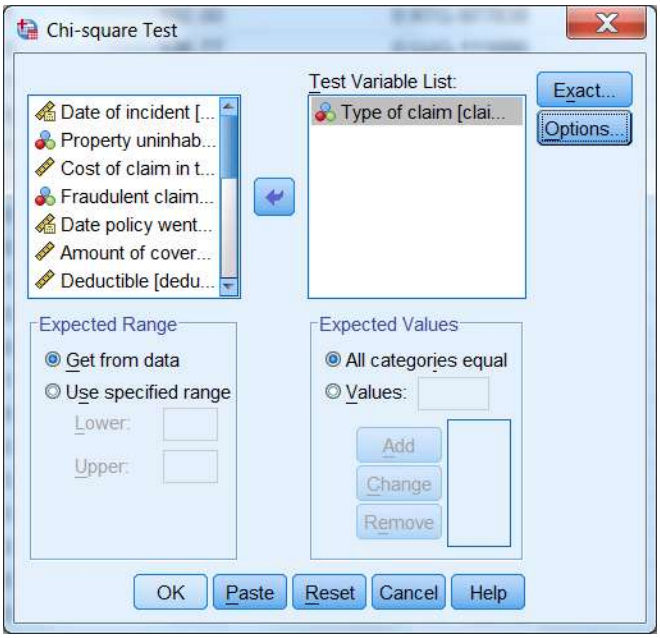


Click option and check the desired box (Descriptive under statistics)



continue

Under Expected Values check the desired option (in present case all categories equal)



click OK

NPar Tests

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\insurance_claims.sav

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Type of claim	4415	3.06	1.535	1	5

Chi-Square Test

Frequencies

Type of claim

	Observed N	Expected N	Residual
Wind/Hail	1054	883.0	171.0
Water damage	627	883.0	-256.0
Fire/Smoke	1039	883.0	156.0
Contamination	404	883.0	-479.0
Theft/Vandalism	1291	883.0	408.0
Total	4415		

Test Statistics

	Type of claim
Chi-Square	583.259[a]
df	4
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 883.0.

The Test Statistics shows the p-value almost zero which is less than 0.05, therefore the null hypothesis is rejected in favour of alternate, ie., type of claim are not equally distributed. This can be visualised through graph also.

Double click the frequency table to bring into editable mode

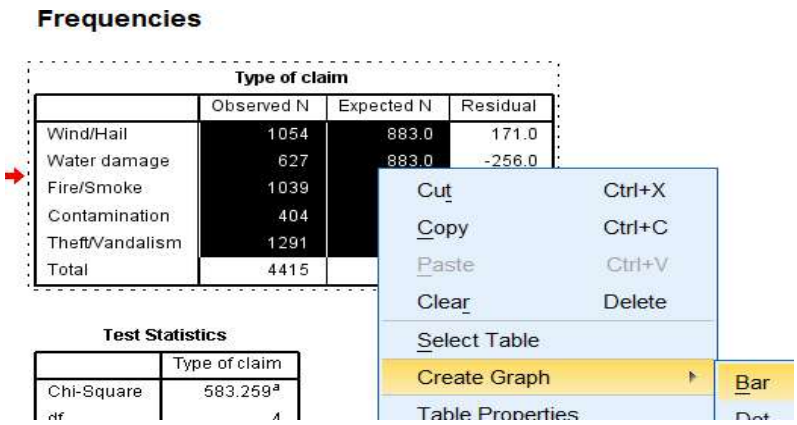
Frequencies

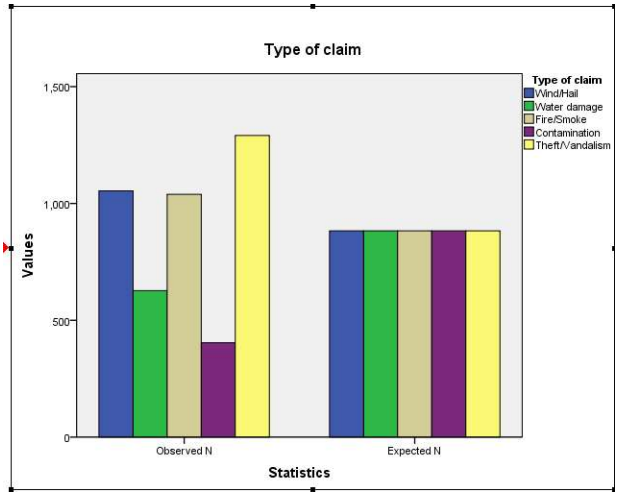
Type of claim

	Observed N	Expected N	Residual
Wind/Hail	1054	883.0	171.0
Water damage	627	883.0	-256.0
Fire/Smoke	1039	883.0	156.0
Contamination	404	883.0	-479.0
Theft/Vandalism	1291	883	
Total	4415		

Double-click to activate

Select both the columns of observed & expected frequency and right click > Create Graph > Bar





Conclusion:

Give information of data set used for testing. Also mentioning the p-value write the decision about rejection or non-rejection of null hypothesis.

Precautions:

1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.
3. Null hypothesis is defined carefully and level of significance chosen appropriately.

Experiment - 10

Objective: ANOVA One Way

Resources: SPSS editor, Data Source file.

Theory And Methodology

Let k random samples of size $n_1, n_2 \ldots n_k$ are selected from normal populations with mean $\mu_1, \mu_2 \ldots \mu_k$, and common variance σ^2 . And let the means are depending on one factor then One Way ANOVA is applied

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ and H_1: at least two means are different.

Procedure: the total variation can be attributed to variation within the class and variation between the class.

Total sum of squares = Sum of square within the class + sum of square between the class

$TSS = SSB + SSW$

The degree of freedom (df) for TSS is $(N-1)$, for SSB is $(k-1)$, & for SSW is $(N-k)$.

Now mean square (MS) for each variation is obtained by dividing each sum of square by its respective degree of freedom (df), i.e.,

$$MSS = \frac{TSS}{N-1}, \quad MSB = \frac{SSB}{k-1}, \quad MSW = \frac{SSW}{N-k}$$

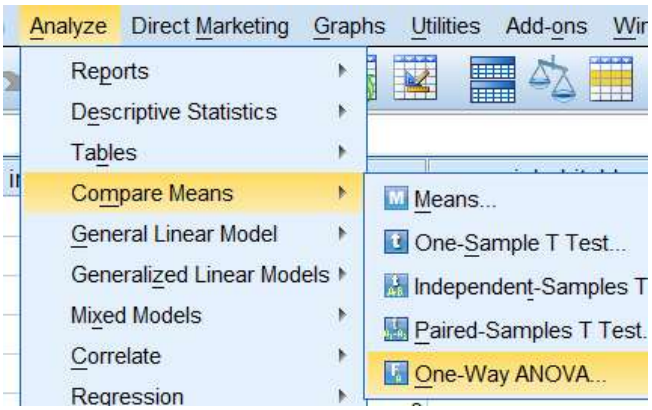
The value of F −statistic will be

$$F = \frac{\sum_i \sum_j n_i (x_{i.} - \bar{x}_{..})^2}{\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2} * \frac{N-k}{k-1}, \quad \nu_1 = k-1, \ \nu_2 = N-k$$

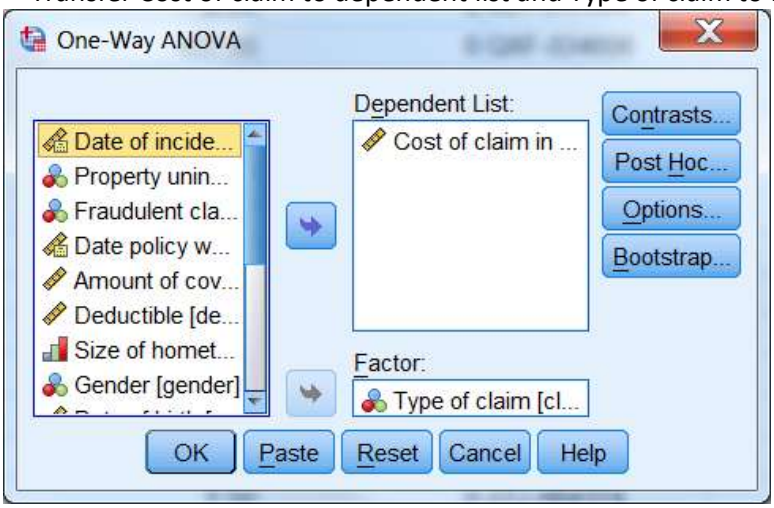
Open the desired file in SPSS editor File > Open > file path

(insurance_claims.sav from repository is used and claim amount & Type of claim variables are considered) (see the data 5-types of claims are there)
H_0: average claim amount in each category is same. Ie., $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
H_1: At least one of these differs

Now follow the path Analyze > Compare Means >

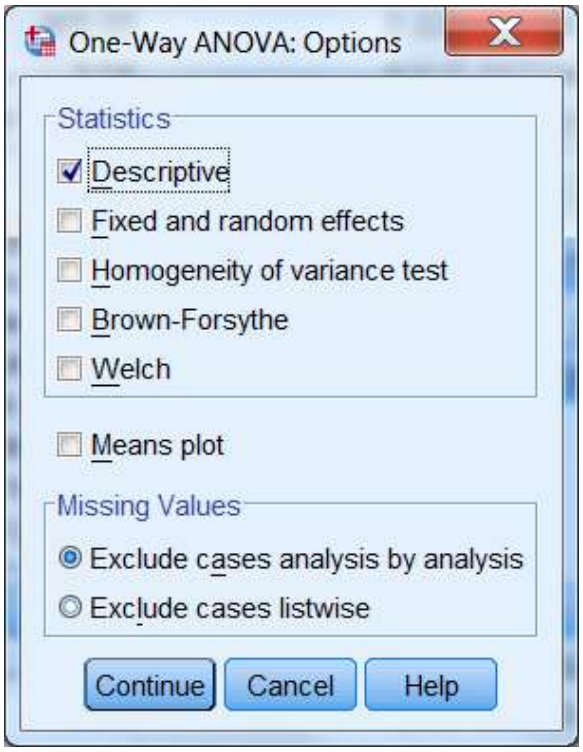


Transfer Cost of claim to dependent list and Type of claim to factor list



Click Option and select the Descriptive to display statistics

Continue & click OK.


One-Way ANOVA: Options
Statistics
Descriptive
Fixed and random effects
Homogeneity of variance test
Brown-Forsythe
Welch
Means plot
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Continue
Cancel
Help

Doc Output Window

Oneway

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\English\insurance_claims.sav

Descriptives

Cost of claim in thousands

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Wind/Hail	1054	16.7819	20.51640	.63195	15.5419	18.0219	1.46	184.38
Water damage	627	35.2896	47.55906	1.89933	31.5598	39.0195	2.31	504.94
Fire/Smoke	1039	171.5793	204.53371	6.34537	159.1281	184.0305	11.59	1635.00
Contamination	404	202.2067	227.66948	11.32698	179.9393	224.4740	18.69	1662.00
Theft/Vandalism	1291	17.4795	25.27650	.70348	16.0994	18.8596	1.57	576.91
Total	4415	73.0109	144.40137	2.17323	68.7503	77.2715	1.46	1662.00

ANOVA

Cost of claim in thousands

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25043718.31	4	6260929.579	412.125	.000
Within Groups	66995929.16	4410	15191.821		
Total	92039647.47	4414			

Since the p-value of F-statistics is almost zero which is less than 0.05, hence the Null Hypothesis is rejected, ie., at least one of the mean is different.

Conclusion:

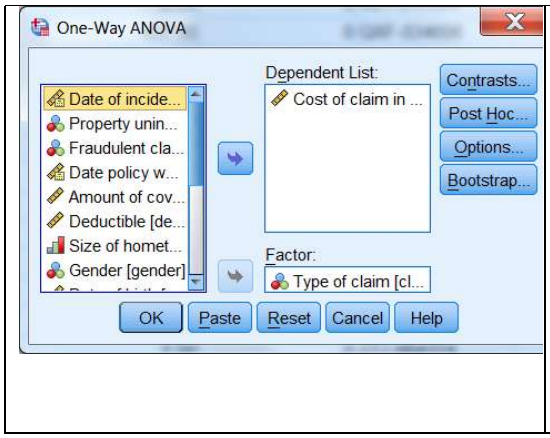
Give information of data set used for testing. Also mentioning the p-value write the decision about rejection or non-rejection of null hypothesis.

Precautions:

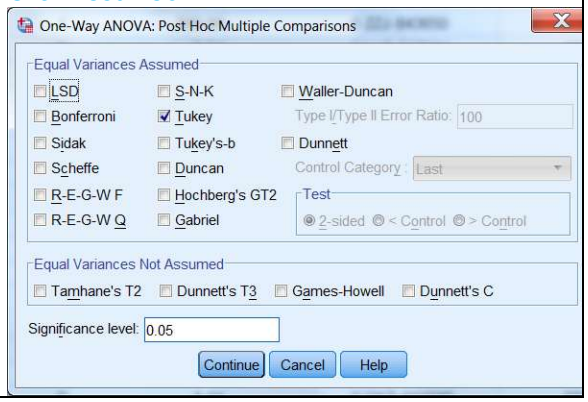
1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.
3. Null hypothesis is defined carefully and level of significance chosen appropriately.

(Optional Procedure)

In case null hypothesis is rejected and you are interested to know which mean is different then



Click Post Hoc..



Select Tukey & continue. Additional windows will open to show the results

Post Hoc Tests

Multiple Comparisons

Cost of claim in thousands
Tukey HSD

(I) Type of claim	(J) Type of claim	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Wind/Hail	Water damage	-18.50775*	6.21634	.024	-35.4715	-1.5440
	Fire/Smoke	-154.79742*	5.38842	.000	-169.5019	-140.0929
	Contamination	-185.42477*	7.21228	.000	-205.1064	-165.7431
	Theft/Vandalism	-.69762	5.11673	1.000	-14.6607	13.2655
Water damage	Wind/Hail	18.50775*	6.21634	.024	1.5440	35.4715
	Fire/Smoke	-136.28967*	6.23305	.000	-153.2991	-119.2803
	Contamination	-166.91702*	7.86339	.000	-188.3755	-145.4586
	Theft/Vandalism	17.81013*	5.99974	.025	1.4374	34.1828
Fire/Smoke	Wind/Hail	154.79742*	5.38842	.000	140.0929	169.5019
	Water damage	136.28967*	6.23305	.000	119.2803	153.2991
	Contamination	-30.62735*	7.22669	.000	-50.3483	-10.9064
	Theft/Vandalism	154.09980*	5.13703	.000	140.0813	168.1183
Contamination	Wind/Hail	185.42477*	7.21228	.000	165.7431	205.1064
	Water damage	166.91702*	7.86339	.000	145.4586	188.3755
	Fire/Smoke	30.62735*	7.22669	.000	10.9064	50.3483
	Theft/Vandalism	184.72715*	7.02645	.000	165.5526	203.9017
Theft/Vandalism	Wind/Hail	.69762	5.11673	1.000	-13.2655	14.6607
	Water damage	-17.81013*	5.99974	.025	-34.1828	-1.4374
	Fire/Smoke	-154.09980*	5.13703	.000	-168.1183	-140.0813
	Contamination	-184.72715*	7.02645	.000	-203.9017	-165.5526

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

Cost of claim in thousands

Tukey HSD[a,b]

Type of claim	N	Subset for alpha = 0.05			
		1	2	3	4
Wind/Hail	1054	16.7819			
Theft/Vandalism	1291	17.4795			
Water damage	627		35.2896		
Fire/Smoke	1039			171.5793	
Contamination	404				202.2067
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 740.086.
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Look to the p-values and where ever it is less than 0.05 , difference is significant.

Experiment - 11

Objective: ANOVA Two Way

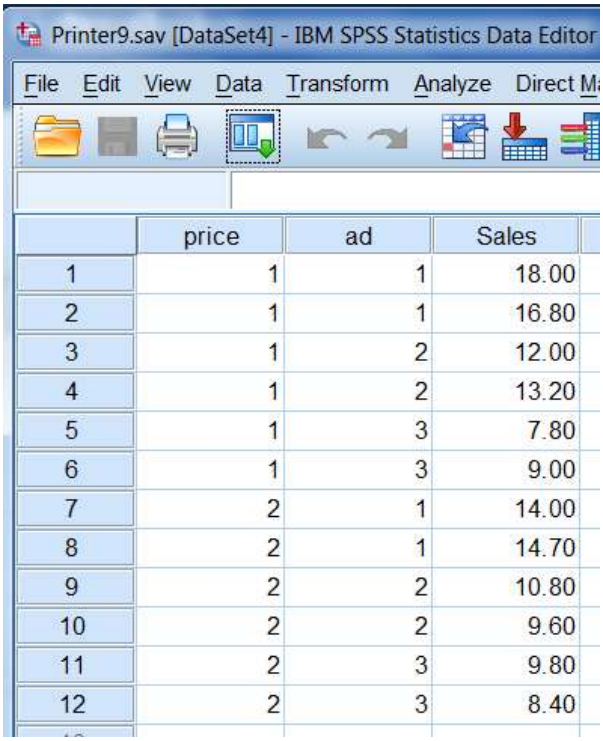
Resources: SPSS editor, Data Source file.

Theory And Methodology

Open the desired file in SPSS editor File > Open > file path

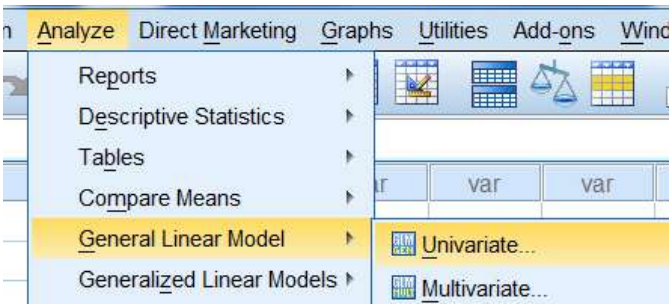
(Printer9.sav is used)
H_0: average claim amount in each category is same.
H_1: At least one of these differs

Sample file

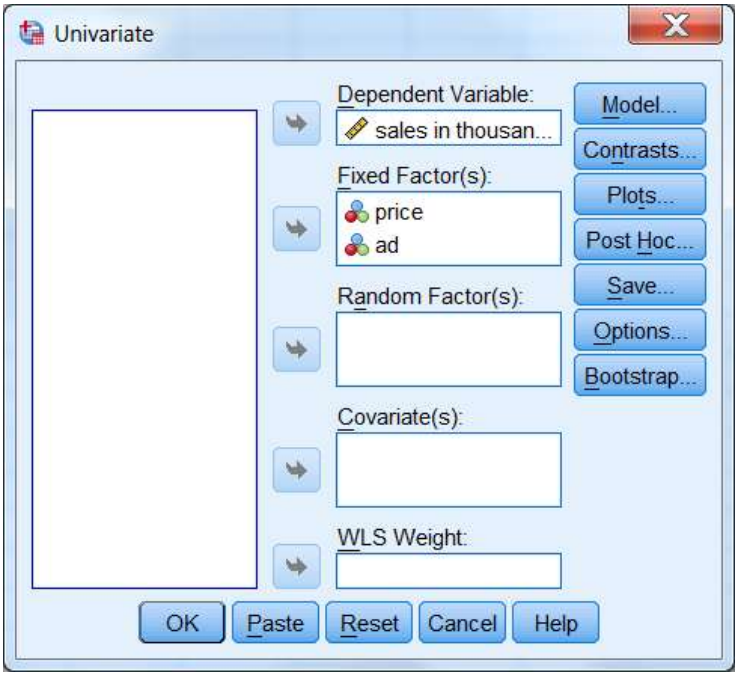


	price	ad	Sales
1	1	1	18.00
2	1	1	16.80
3	1	2	12.00
4	1	2	13.20
5	1	3	7.80
6	1	3	9.00
7	2	1	14.00
8	2	1	14.70
9	2	2	10.80
10	2	2	9.60
11	2	3	9.80
12	2	3	8.40

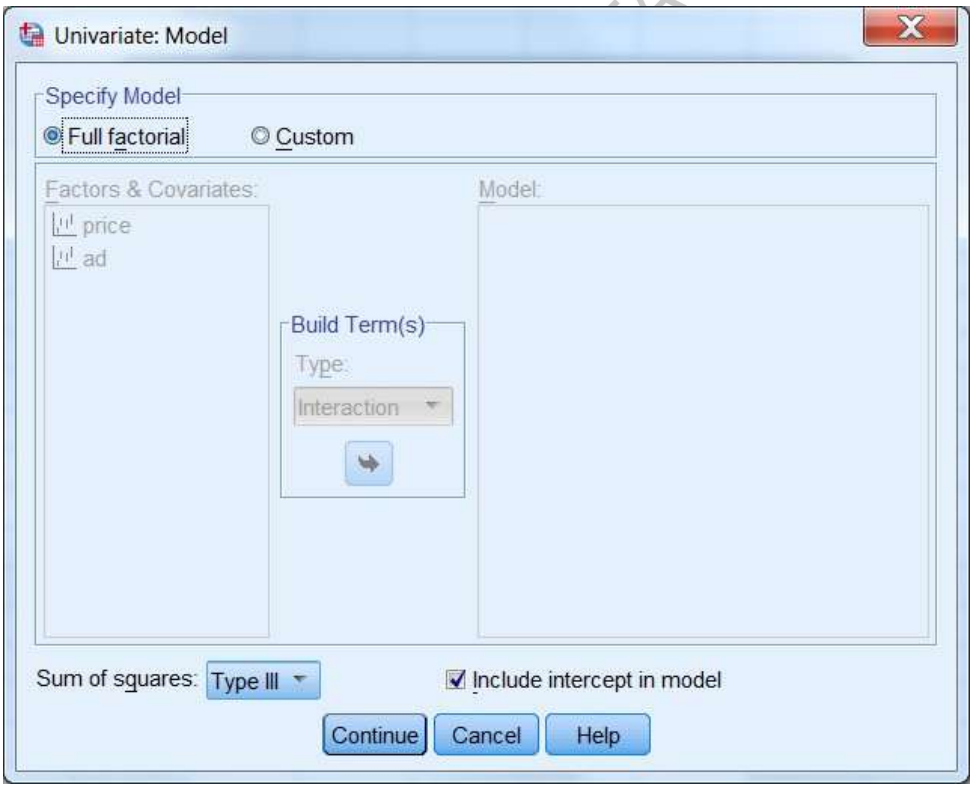
Now follow the path Analyze > General Linear Model > Univariate



Transfer Sales in Thousand in dependent variable box and Price & Ad in Fixed factor box

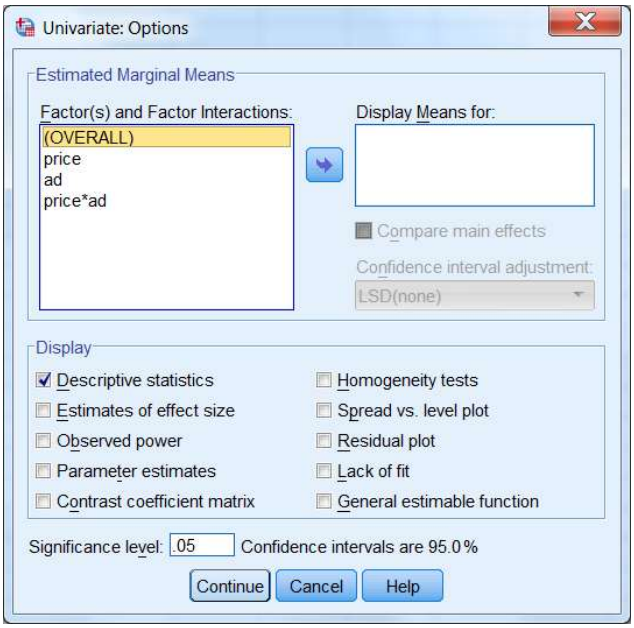


click Model



By default Full factorial is selected, need not to change.

Click Option button and choose the desired option, and continue.



Click OK.

Doc Output Window

Univariate Analysis of Variance

[DataSet4] C:\Users\Ramesh\Desktop\Lab\ANOVA 2-way\Printer9.sav

Between-Subjects Factors

		Value Label	N
price	1	$600	6
	2	$700	6
ad	1	ad-tv	4
	2	ad-radio	4
	3	ad-newspaper	4

Descriptive Statistics

Dependent Variable:sales in thousands

price	ad	Mean	Std. Deviation	N
$600	ad-tv	17.4000	.84853	2
	ad-radio	12.6000	.84853	2
	ad-newspaper	8.4000	.84853	2
	Total	12.8000	4.08118	6
$700	ad-tv	14.3500	.49497	2
	ad-radio	10.2000	.84853	2
	ad-newspaper	9.1000	.98995	2
	Total	11.2167	2.55376	6
Total	ad-tv	15.8750	1.85000	4
	ad-radio	11.4000	1.54919	4
	ad-newspaper	8.7500	.85440	4
	Total	12.0083	3.34948	12

Tests of Between-Subjects Effects

Dependent Variable:sales in thousands

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	119.304[a]	5	23.861	34.876	.000
Intercept	1730.401	1	1730.401	2529.210	.000
price	7.521	1	7.521	10.993	.016
ad	103.752	2	51.876	75.823	.000
price * ad	8.032	2	4.016	5.870	.039
Error	4.105	6	.684		
Total	1853.810	12			
Corrected Total	123.409	11			

a. R Squared = .967 (Adjusted R Squared = .939)

(Interpretation: The p-values along price and ad both are much smaller than 0.05, therefore Null Hypothesis is rejected for both the factors, ie., the average sales are not same along boththe factors separately.)

Conclusion:

Give information of data set used for testing. Also mentioning the p-value write the decisionabout rejection or non-rejection of null hypothesis.

Precautions:

1. Variable Properties are to be defined carefully.
2. Missing values if any should be taken care of.
3. Null hypothesis is defined carefully and level of significance chosen appropriately.