SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
COLLEGE OF SCIENCE AND HUMANITIES
Ramapuram Campus

EX.NO: 01 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

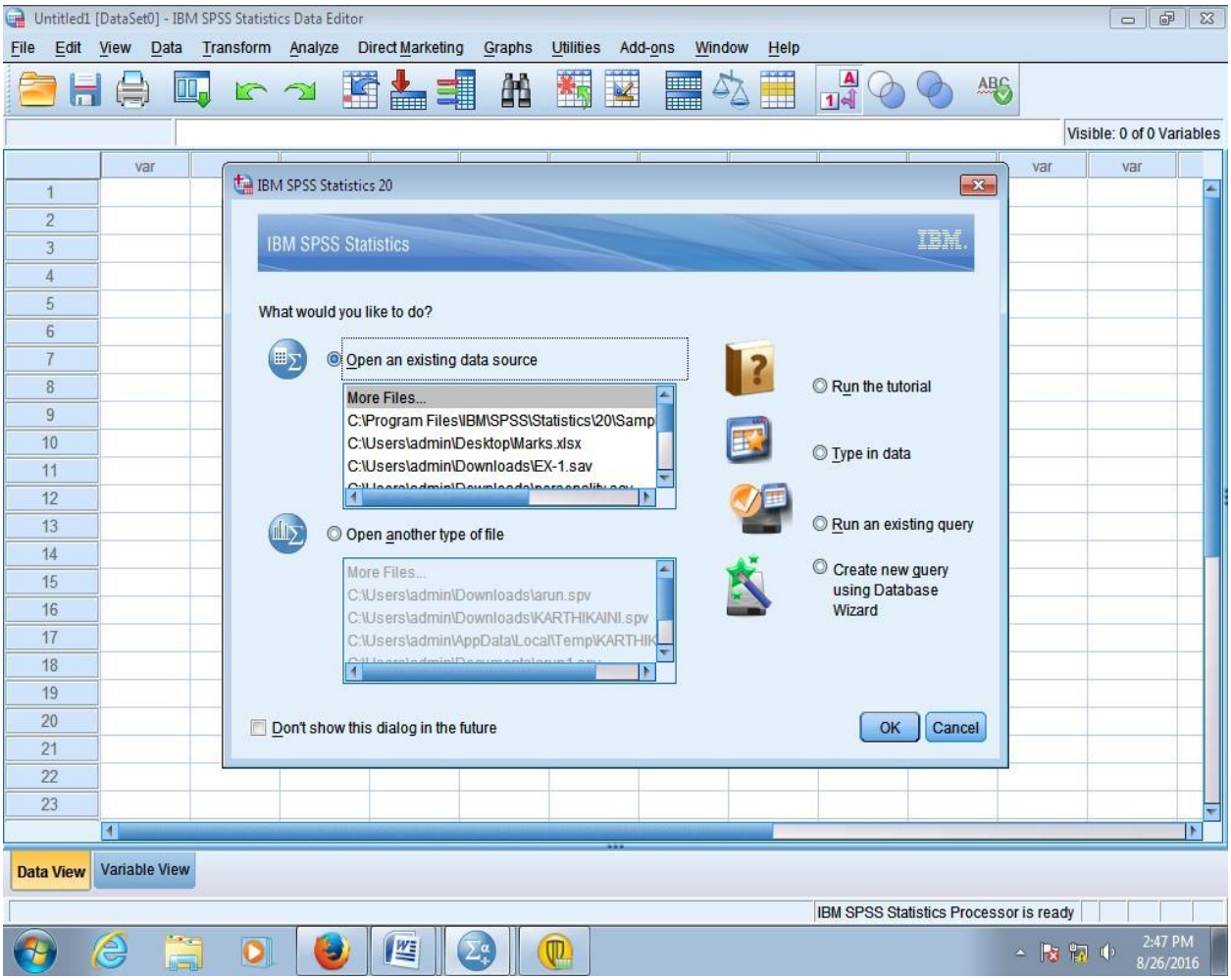
INTRODUCTION TO SPSS (Menus & Tools)

AIM:

- To open SPSS and study about the Menus and Tools in it
- To enter data into SPSS Data Editor Window.

PROCEDURE:

- To start SPSS, choose **IBM SPSS Statistics 20** from the Start menu. The **Data Editor** window will be shown as below:

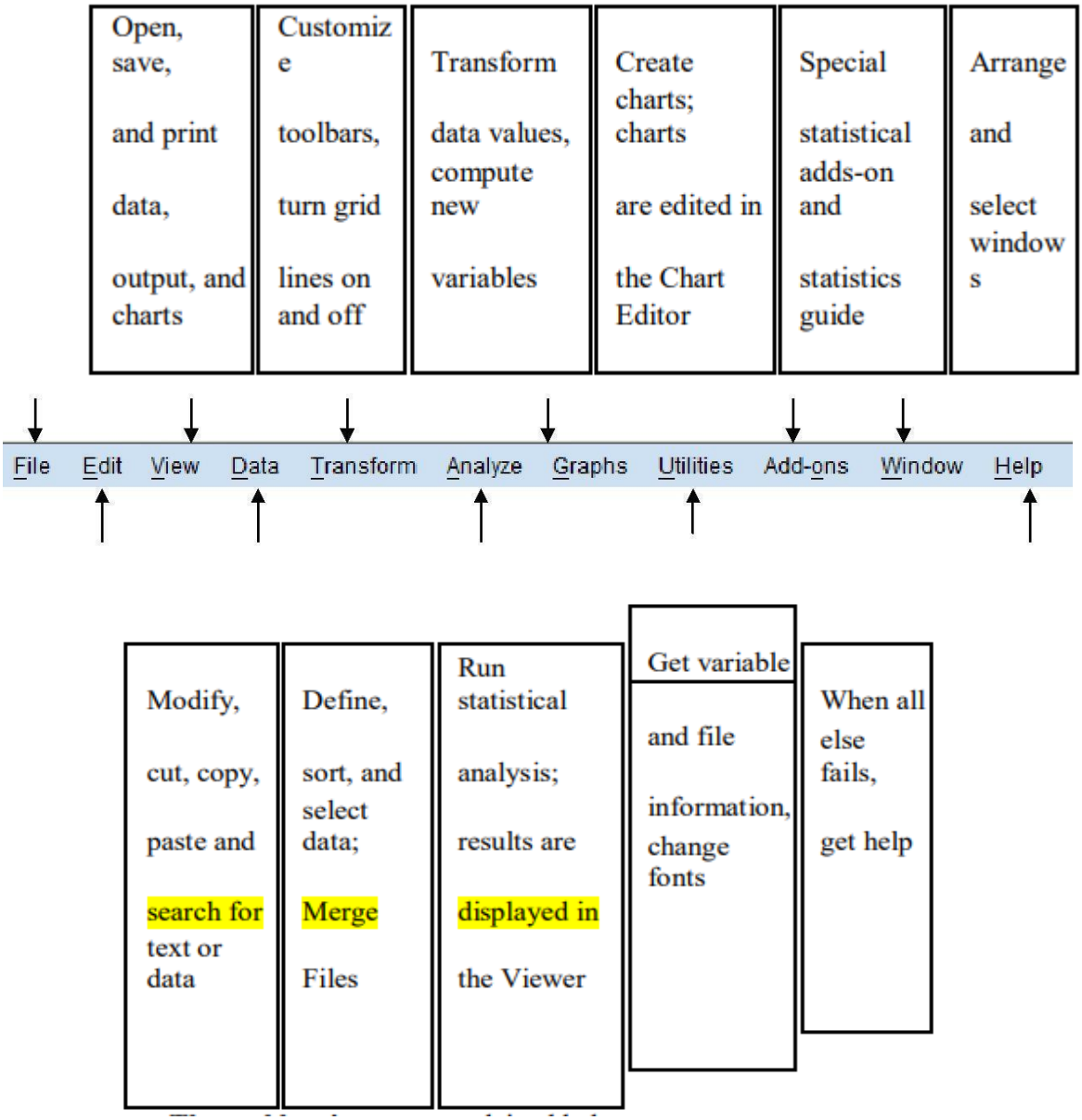


- In order to type any data into SPSS that you already have, you need to select **Type in data** option and then click on **OK.**

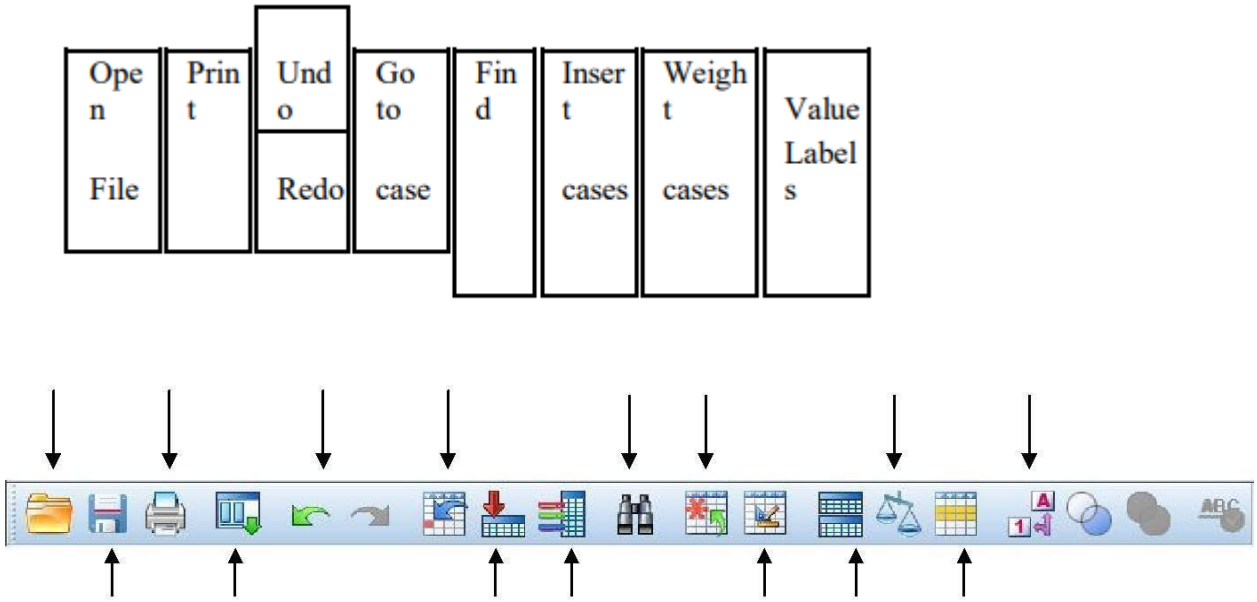
- If you are going to work on the existing data set in SPSS, click OK and select **C drive → Program Files → IBM → SPSS → Statistics → 20 → Samples → English →** and choose the data set you desire.

- In the **data editor window** the Menu bar consists of the usual Windows menus, plus some specific to SPSS, such as **Data, Transform, Analyze and Graph.** Details of the menus are given below:

- The **Menu bar** provides easy access to most SPSS features. It consists of ten drop-down menus:



- **The tool bar icons** are explained below:



Save File	Dialog Recall	Go to Variable	Variables	Insert variable	Split File	Select cases

- **To enter data in SPSSData Editor Window**, follow the steps given for the below example:
- **Example:**The Framinhgham Heart Study followed a cohort of 5209 men and women for over 25years. The study has been important in identifying risk factors associated with cardiovascular disease. The following is a description of the variables we have selected from the study for our purpose:

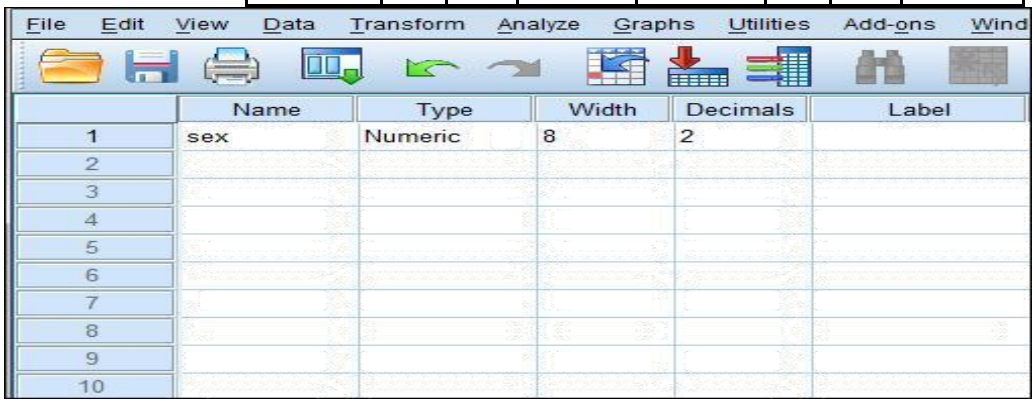
Column	**Description of Variable**
1	Sex (Gender: M-Male, F-Female),
2	Age (30-64 years),
3	Systolic blood pressure (82-300 mm).

- The table below contains data for a random sample of 28 subjects from the study.
- Now you can **enter the data into the Data Editor window**. Do not enter the names of the variables at the top of each column yet. Follow the instructions below:

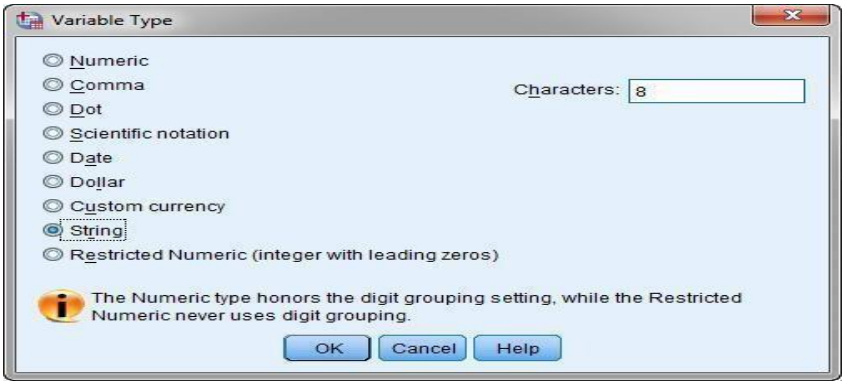
- The variable name must begin with a letter and cannot end with a period. The length of the name cannot exceed 8 characters. To define a variable make the **VariableView** the active window (click the **Variable View**tab at the bottom of the**Data Editor window**). This will obtain the**Variable View window.**

- Enter the new variable name in the column **Name**in any blank row. For example, enter the name **sex**in the first row. After entering the name, the default attributes (**Type, Width**,...) areautomatically assigned.

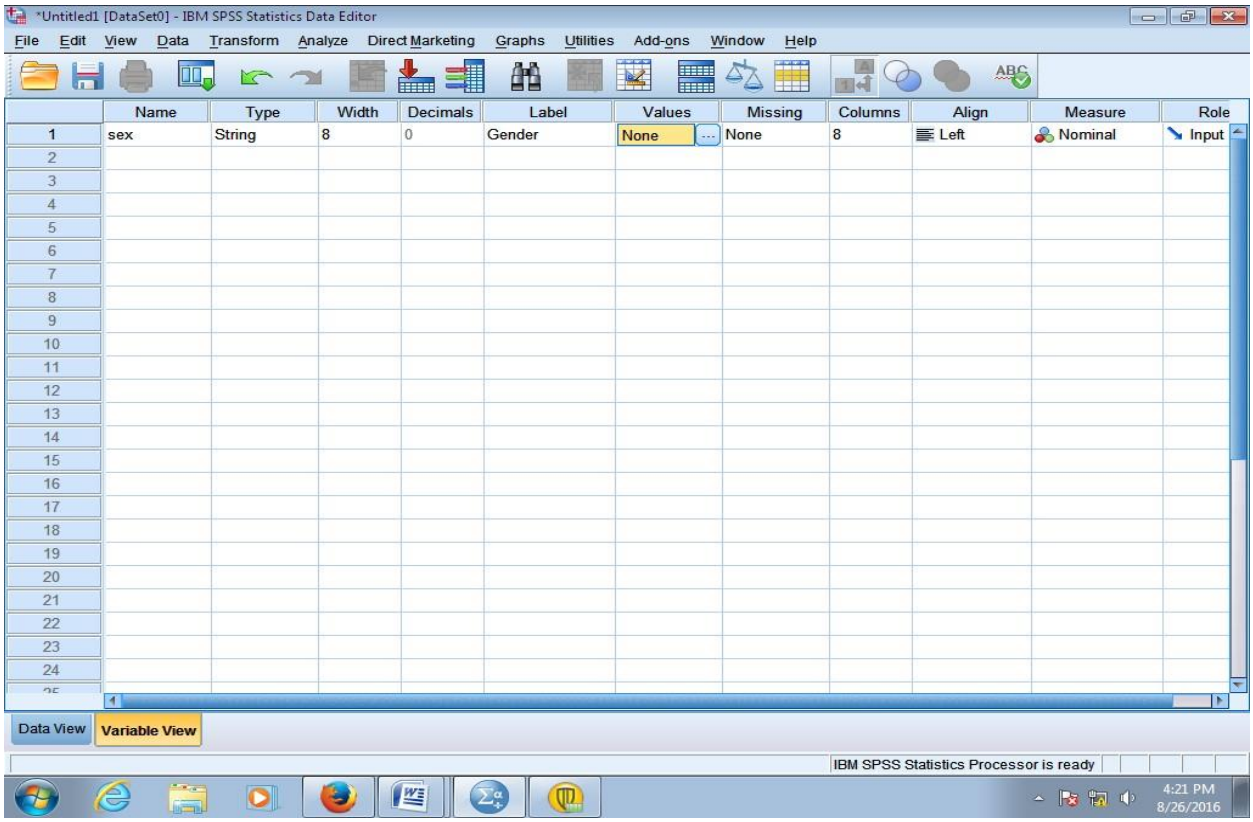
Number	Sex	Age	Systolic	Number	Sex	Age	Systolic
1	F	59	170	15	F	61	156
2	M	35	130	16	F	49	170
3	M	46	136	17	M	32	120
4	F	43	96	18	F	54	162
5	M	53	120	19	F	33	110
6	M	50	110	20	F	41	145
7	M	33	100	21	M	56	134
8	M	57	145	22	F	36	104
9	F	41	132	23	M	56	126
10	F	40	112	24	F	40	100
11	M	54	140	25	F	37	116
12	M	53	148	26	M	38	132
13	F	53	165	27	F	46	132
14	M	49	100	28	M	38	124



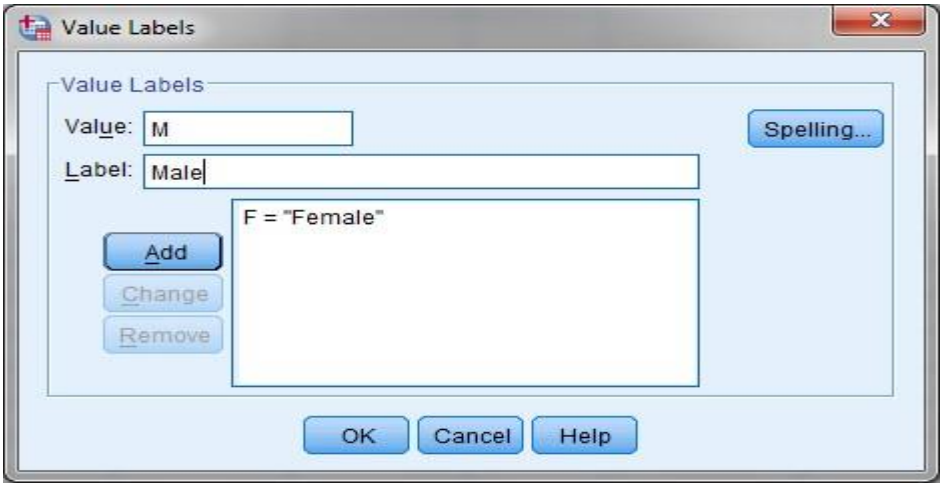
- Clicking the cell **Numeric** and then the button in the cell opens the **Variable Type** dialog box.



- As **sex** is a categorical variable, click the radio button for **String.**
- In our example there are three variables: **sex(categorical), age (numeric), and systolic blood pressure (numeric).** You may provide a description of the variable listed in each row of the Viewer window in the **Label** column. For example, we may assign the label gender to the variable**sex**. Enter **gender** in the **Label** column corresponding to the variable **sex**. • To define possible values of the variable sex (possible values M for male and F for female) click the Values cell in the row for the variable, and then click the button in the cell.



- Define the **Value labels** for the variable **sex**as follows:



- After assigning values for male and female click on **Add** and then click **OK.**
- In the same way enter the remaining variables **Age and Systolic**. **Age**should be defined as a numeric variablewith two digits (to minimize the chances of transcription error) and ***Systolic***as a numeric variable with 3 digits.

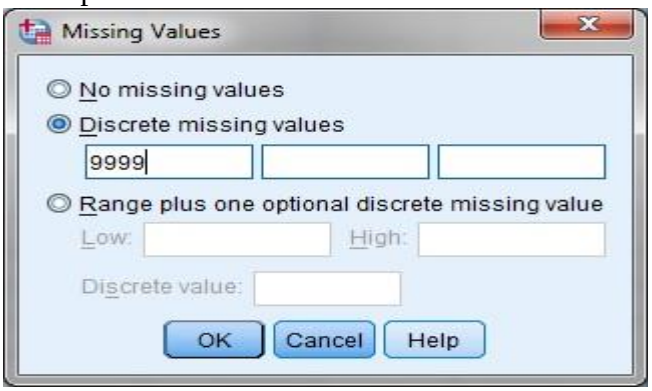
- **To delete a variable (row)**, select the row number that you wish to delete, click **Edit,** and then on **Clear**. The selected variable will be deleted and all variables to the right of the deleted variable will shift to the left. Alternatively, you can select the row and press **Delete** key on your keyboard.

- **To insert a new variable (row)** between existing variables: click on the row that is below the row where you wish to enter a new variable, click **Data**on the menu bar, and then click **Insert Variable** from the pull-down menu.

- **To enter the values for the assigned variables**, switch from the **Variable View** window to the **Data View** window. The three variables **gender, age, and systolic** are represented as columns. Each row represents a case or an observation.
- Enter the values for all cases on one variable (column) and then repeat the procedure for all values in the remaining columns

- The column **Missing** gives the number of data values that is not given in the data view for that particular variable. With SPSS, there are two forms of missing values: **systemmissing and user-defined missing.**
- **System-missing values** are those that SPSS automatically treats as missing. The most common form of this type of value is when there is a "blank" in the data file.
- **User-defined missing values** are those that the user specifically informs SPSS to treat as missing. Rather than leaving a blank in the data file, **numbers** are often entered that are meant to represent data.
- You need to inform SPSS in particular that those **numbers** are to be treated as a missing value; otherwise it will treat it as valid.
- Click the particular cell under the column **Missing**for the variable (Eg. Systolic) we will get the Missing Values dialog box in which we need to select the **Discrete Missing Values** radio button and give the value as **9999**.

- With this definition of the missing values for the variable **systolic**, SPSS will treat **9999** as a missing value of the variable and not include it in any computations involving the systolic blood pressure



- For stringvariables, W**idth** refers to how many characters a value can hold.
- For numeric variables, **Width** refers to how many digits should be displayed. However, SPSS will often override the specified width if it's insufficient. **Columns** refer to the width of the variable's column.
- You can use the **Align** column for setting the alignment for the values of the variables you are entering in the data view.
- The most important assignment for a variable is the **Measures**. In the measures column we need to choose the appropriate option for that variable. There are three types of measures **Nominal, Ordinal and Scale.**
- **Nominal** variables are Qualitative variables without any order. It can be numeric or categorical variables.
- **Ordina**l variables are Qualitative variables with some order. These are categorical variables with intrinsic ranking.
- **Scale** variables are Quantitative variables.
- After giving all the values for the assigned variables in the **Data View Window**, in order to save the data file you have created, go to **File** in the menu bar and choose **Save As**and give the name in the **File name** box and click on**Save** tab.
- **To insert a new case** (row) in between cases that already exists in your data file: click the row below the row where you wish to enter the new case,click **Data**on the menu bar, click **Insert Case** from the pull-down menu.
- **To delete a case**, click the case number that you wish to delete, click **Edit**from the menu, and then on **Clear.** The selected case will be deleted and the rows below will shift upward.

RESULT:

Entered data into SPSS Data Editor Window and learnt various Menu and Tools available with SPSS.

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
COLLEGE OF SCIENCE AND HUMANITIES
Ramapuram Campus

EX.NO: 02 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

ENTERING DATA IN A NEW DATA SET

AIM

To construct frequency table univariant frequency table and cross tabulation.

PROCEDURE:

Step 1: Go to start menu □All program ->IBM SPSS statistics and select IBM SPSS statistics 20

Step 2: The software opens out of the two views->data and variable, choose variable view.

Step 3: In the variable view, enter the field you will have in your frequency table and define their type and size.

Step 4: After entering the field in the, variable view, shift to data view and you can see your fields are ready to take inputs, so enter inputs for the field.

Step 5: Go to analyze in the menu bar->descriptive statistics->choose frequencies. Frequencies dialog box will open. Select the fields you want in your output opens, click OK and output opens.

Step 6: Go to analyze->descriptive statistics ->choose crosstabs, select the fields that you want as rows and the once that you want as column and click OK. output opens.

SAMPLE INPUT:

name	register	gender	mark1	mark2	mark3	mark4	mark5
student1	63	0	65	74	87	91	89
student2	64	0	79	55	86	87	54
student3	65	1	41	41	87	87	73
student4	66	0	70	68	87	90	36
student5	67	1	59	92	72	76	44

Sample Output:

name * mark4 Crosstabulation

Count

		mark4				Total
		76	87	90	91	
name	student1	0	0	0	1	1
	student2	0	1	0	0	1
	student3	0	1	0	0	1
	student4	0	0	1	0	1
	student5	1	0	0	0	1
Total		1	2	1	1	5

gender * mark2 Crosstabulation

Count

		mark2					Total
		41	55	68	74	92	
gender	male	0	1	1	1	0	3
	female	1	0	0	0	1	2
Total		1	1	1	1	1	5

RESULT:

Entered data in SPSS Data Editor Window.

EX.NO: 03 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

CONSTRUCTING BAR DIAGRAM,HISTOGRAM, PIE DIAGRAM

AIM

To represent the data graphically, bar chart pie chart and histogram multi pie chart, subdivided bar diagram

PROCEDURE:

Step 1: Go to start menu ->All program->IBM SPSS statistics and select IBM SPSS statistics 20

Step 2: The software opens out of the two views->data and variable, choose variable view.

Step 3: In the variable view, enter the field you will have in your frequency table and define their type and size.

Step 4: After entering the field in the, variable view, shift to data view and you can see your fields are ready to take inputs, so, enter inputs for the field.

Step 5: Go to analyze in the menu bar->descriptive statistics->choose frequencies. Frequencies dialog box will open. Select the fields you want in your output opens, click OK and output opens.

Step 6: on the right side, you can see option, choose chart from the options. Chart dialog box opens. Select the type of chart->bar, pie or histogram, you want and click ok.

Step 7: Go to graphs->legacy dialogs and choose bar. Bar dialog box opens. Select clustered and click defines or select stacked and click defines.

Step8: Stack or clustered dialog box opens select one variable each for category axis, define stack/cluster by, rows and columns and click OK.

Step 9: Output is displayed.

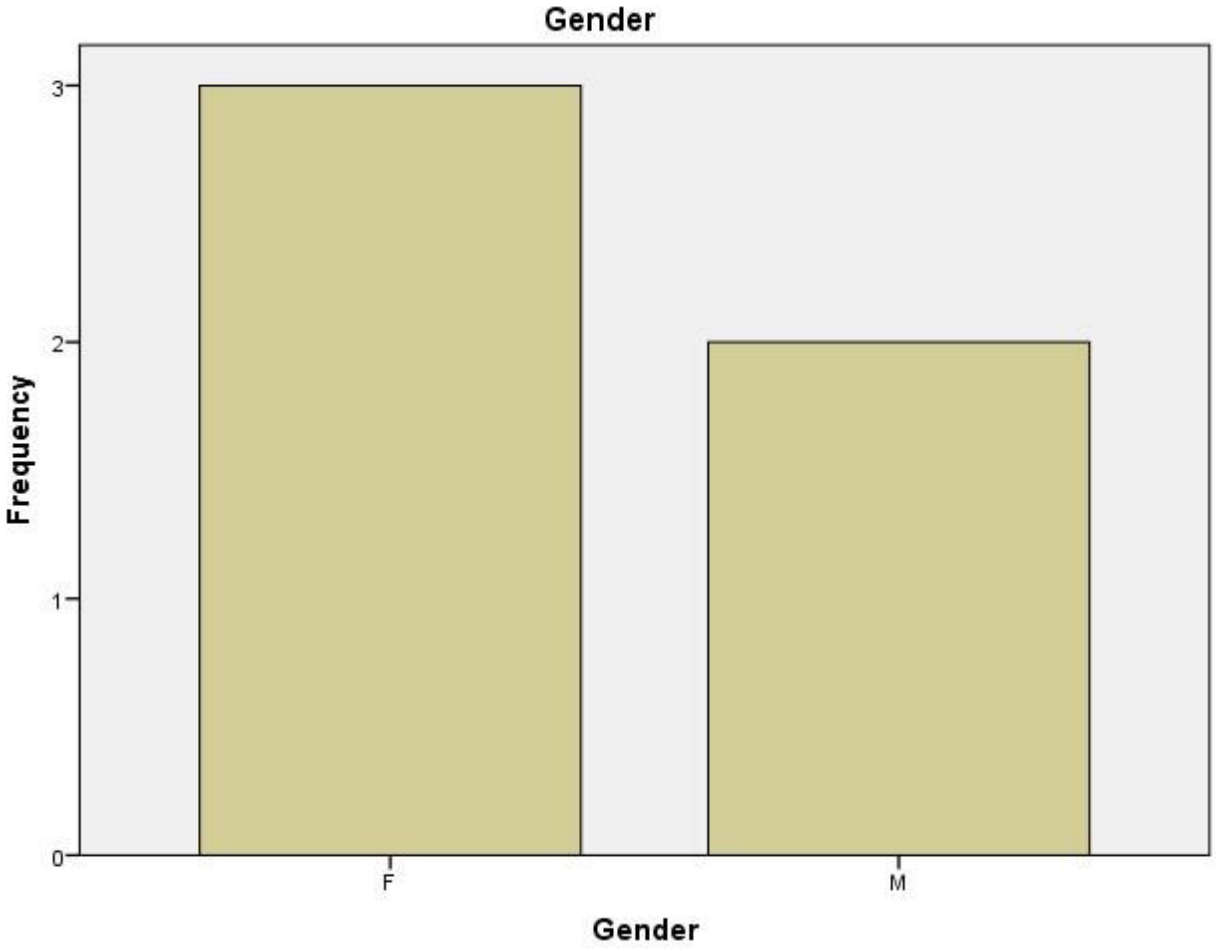
Sample Input:

name	register	gender	mark1	mark2	var
student1	63	0	65	74	
student2	64	0	79	55	
student3	65	1	41	41	
student4	66	0	70	68	
student5	67	1	59	92	

Sample Output:

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	F	3	60.0	60.0	60.0
	M	2	40.0	40.0	100.0
	Total	5	100.0	100.0	



RESULT:

Constructing Piechart, Histogram, Barchart data files using SPSS is done successfully

EX.NO: 04 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

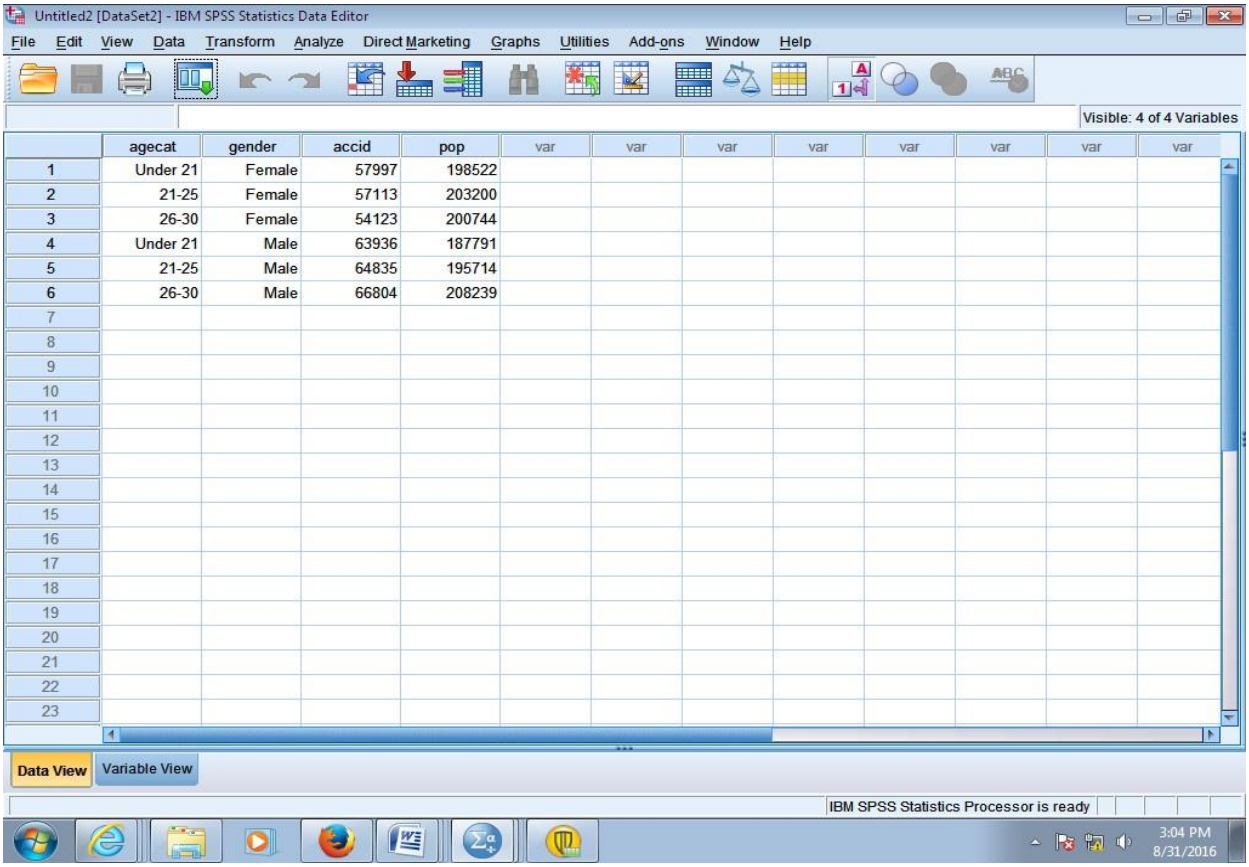
DATA TRANSFORMATIONS, RECODING VARIABLES AND SELECTING CASES

AIM:

To sort, split and merge data files using SPSS.

PROCEDURE:

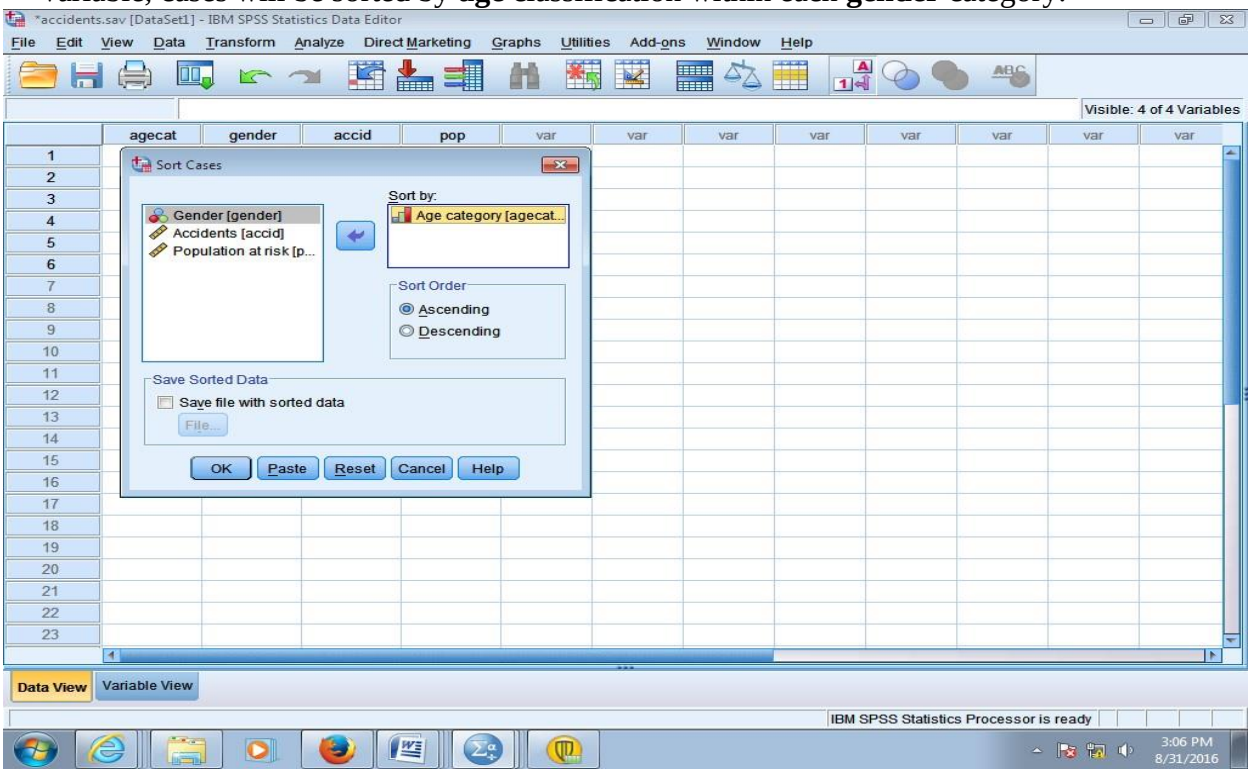
- Suppose that we would like to sort the data in the data file according to the **age** of the subjects enrolled in the study, **select a data set** that is already available in SPSS sample files (Eg:**accidents.sav)** and choose **Data**from the menu bar and then select **SortCases.**



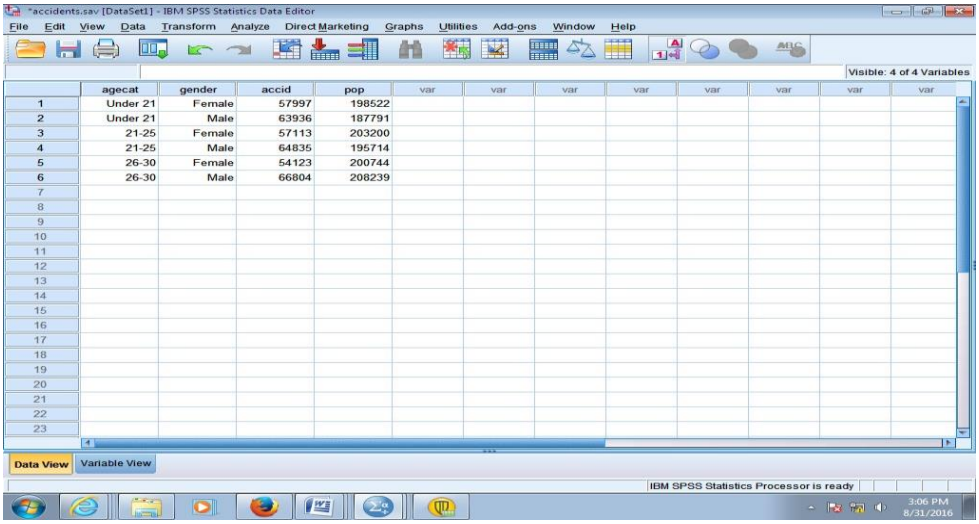
- In order to sort the subjects according to the age, select **agecat** and move it to the **Sort by** box. You can sort cases in ascending or descending order.

If you select multiple sort variables, cases are sorted by each variable within category of the prior variable on the **Sort by:** list.

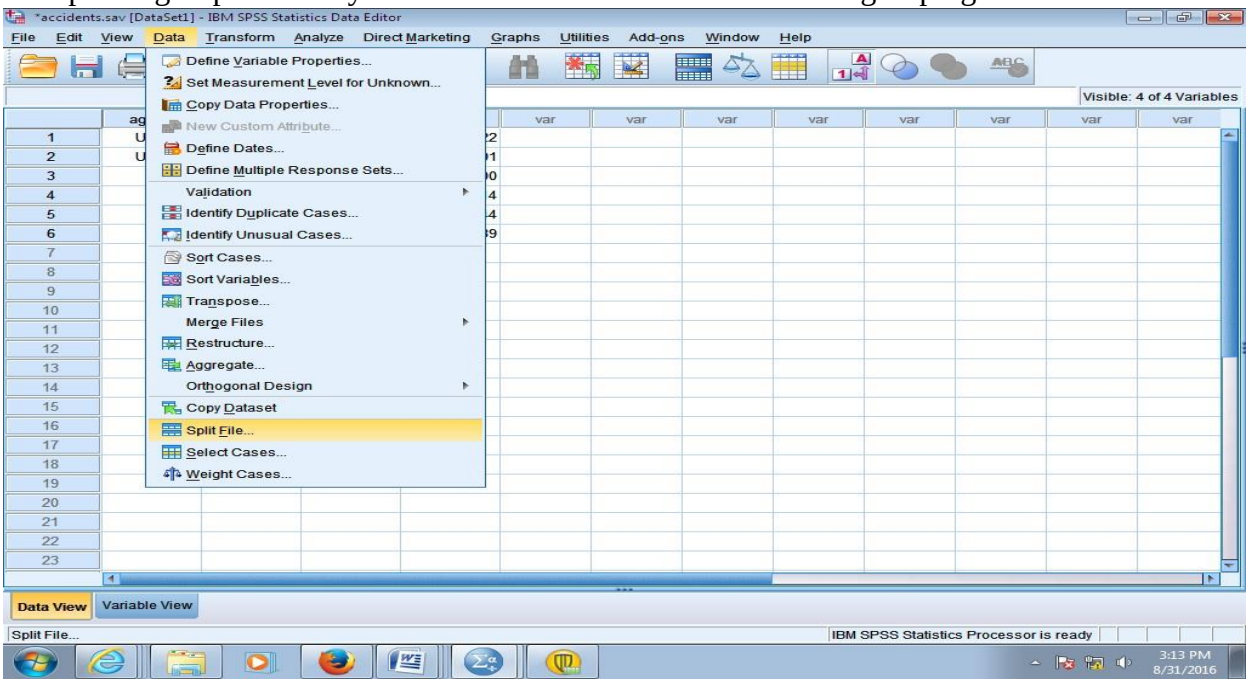
- For example, if you select **gender** as the first sorting variable and***age*** as the second sorting variable, cases will be sorted by **age** classification within each **gender** category.



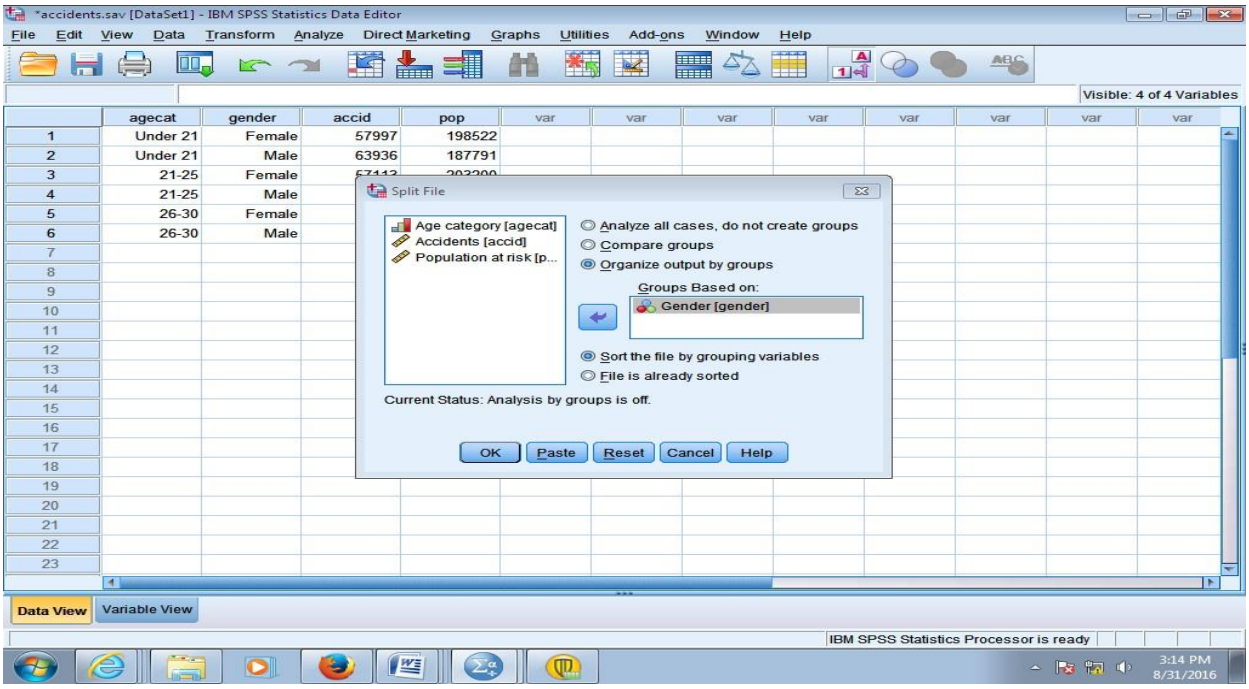
- The below window shows the sorted data for the variable **agecat.**



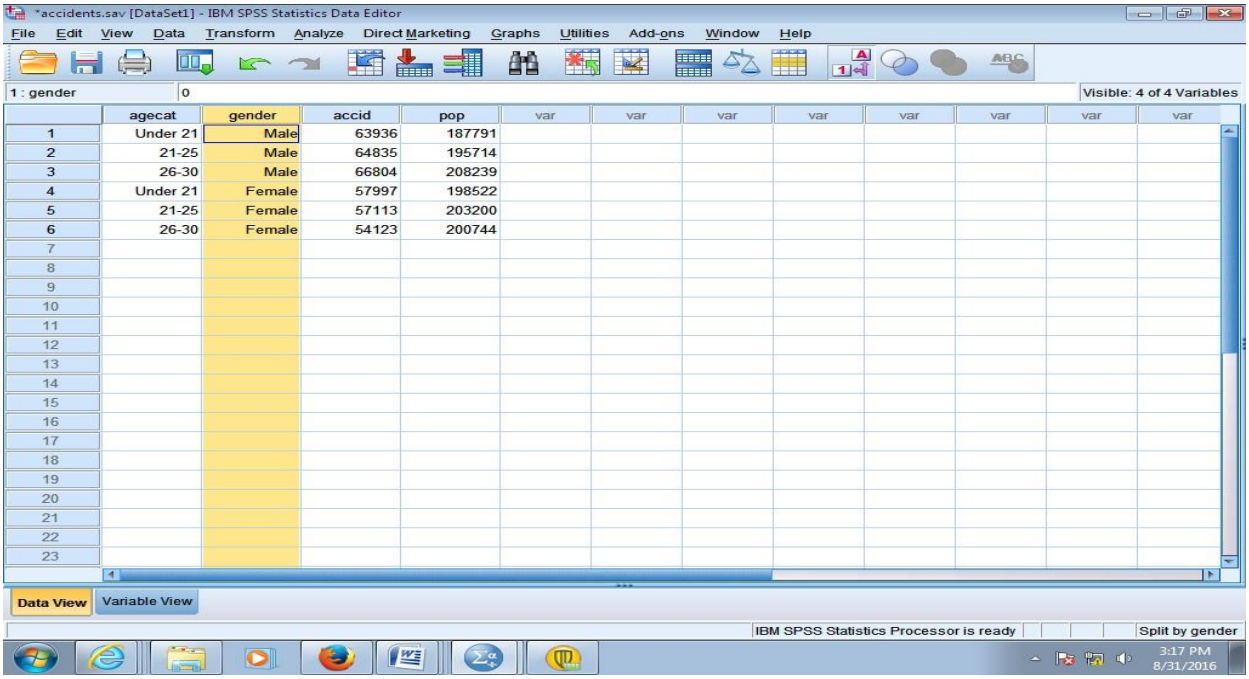
- **Split File** feature can be accessed from the **Data** menu. This feature **split the data file** into separate groups for analysis based on the values of one or more grouping variables.



- If you select multiple grouping variables, cases are grouped by each variable within categories of the prior variable.
- For example, if you select **sex** as the first grouping variable and **agecat** as the second grouping variable, cases will be grouped by age category classification within each gender group.
- If you check **Organize output by groups** radio button, all results from each statistical procedure will be displayed separately for each split-file group (**in this case gender**).

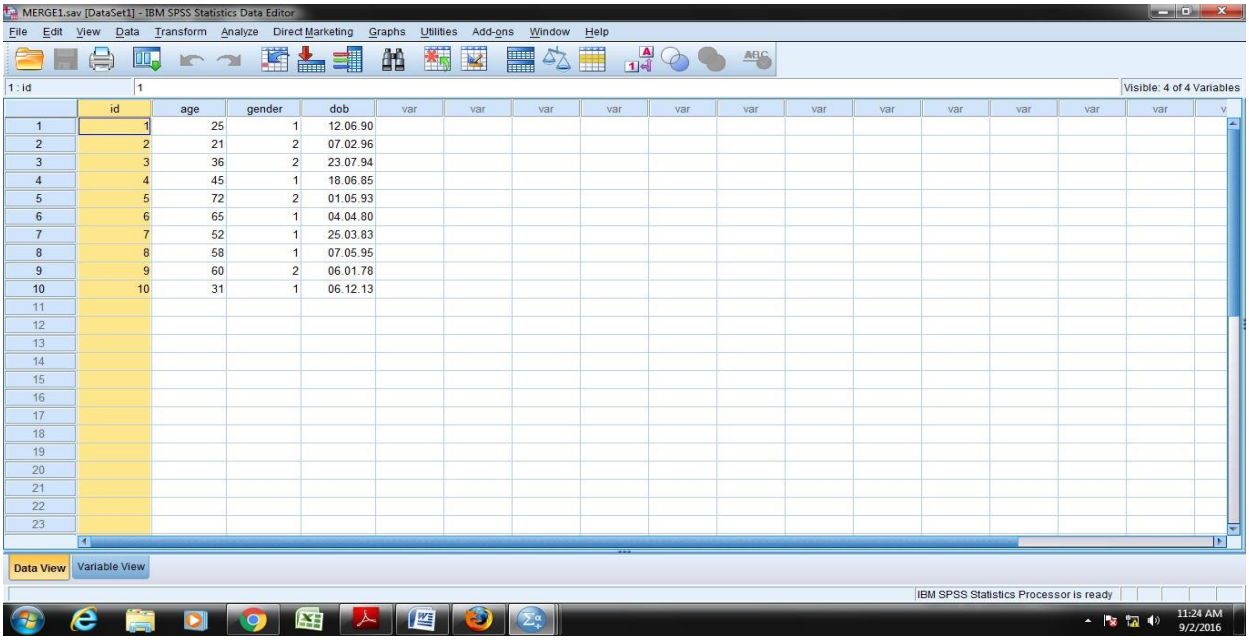


- In the below window we can see that the variable **gender** is split into two different groups say male and female.

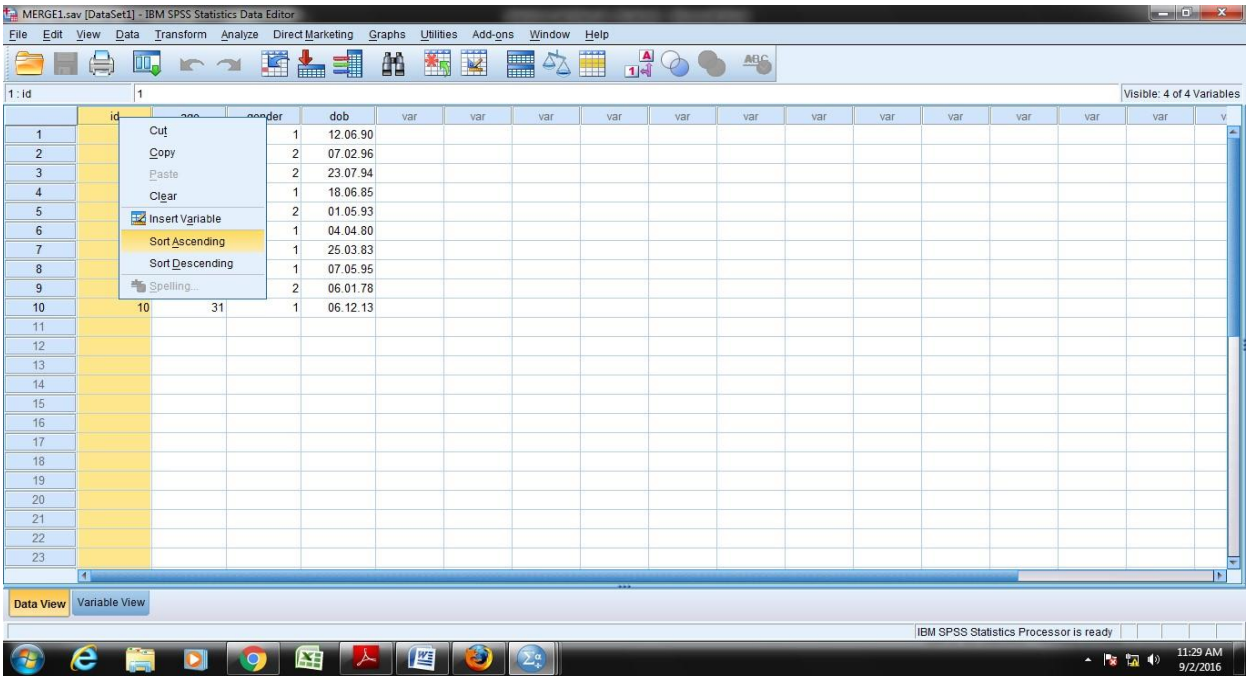


	agecat	gender	accid	pop	var	var	var	var	var	var	var	var
1	Under 21	Male	63936	187791								
2	21-25	Male	64835	195714								
3	26-30	Male	66804	208239								
4	Under 21	Female	57997	198522								
5	21-25	Female	57113	203200								
6	26-30	Female	54123	200744								
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												

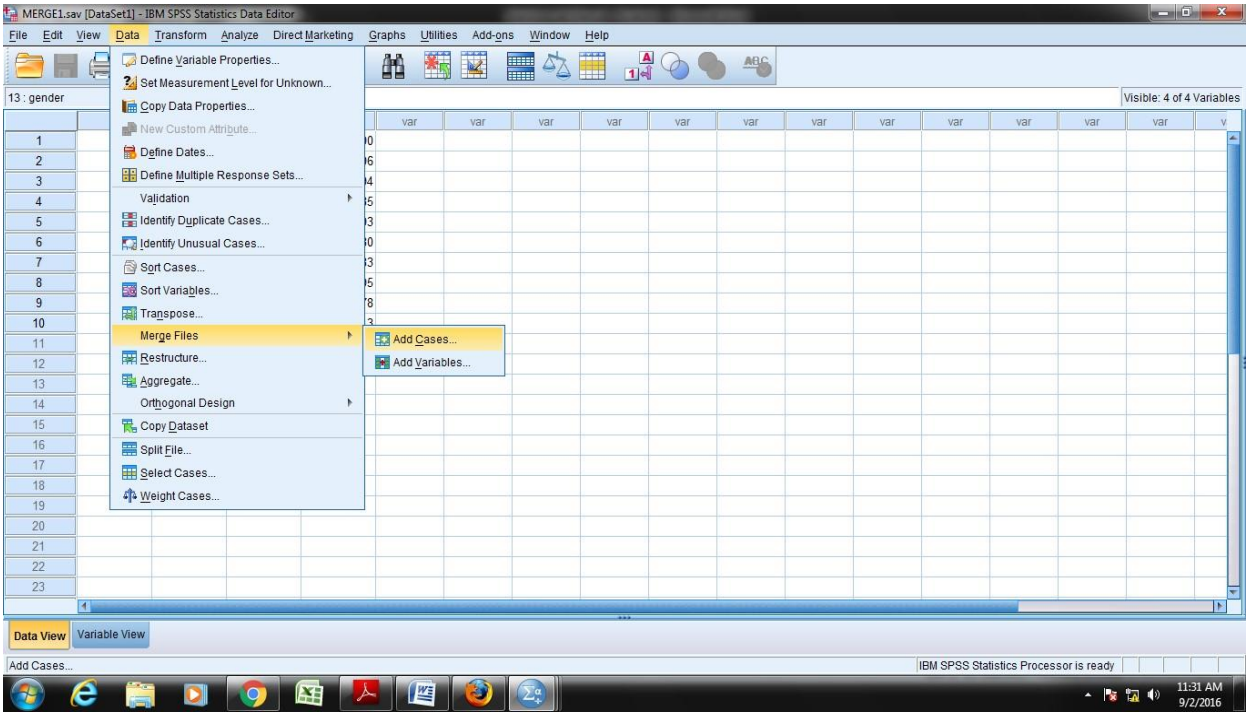
- For **Merging Files** in SPSS, both files need to have a common indexing key (preferably numeric). This would be a **unique identifier** for each variable in your data set.
- The common indexing keys have to be sorted in ascending order for SPSS to be able to merge files, so **make sure both files are sorted in ascending order before trying to merge.**
- Both data files should provide different data for the same set of variables. For example, you might record the same information for customers in two different sales regions and maintain the data for each region in separate files.
- Open one of the data files that you want to merge and that will be your active data set. The cases from this file will appear first in the merged data file. Here we are going to open the file **MERGE1** that is created as shown below:



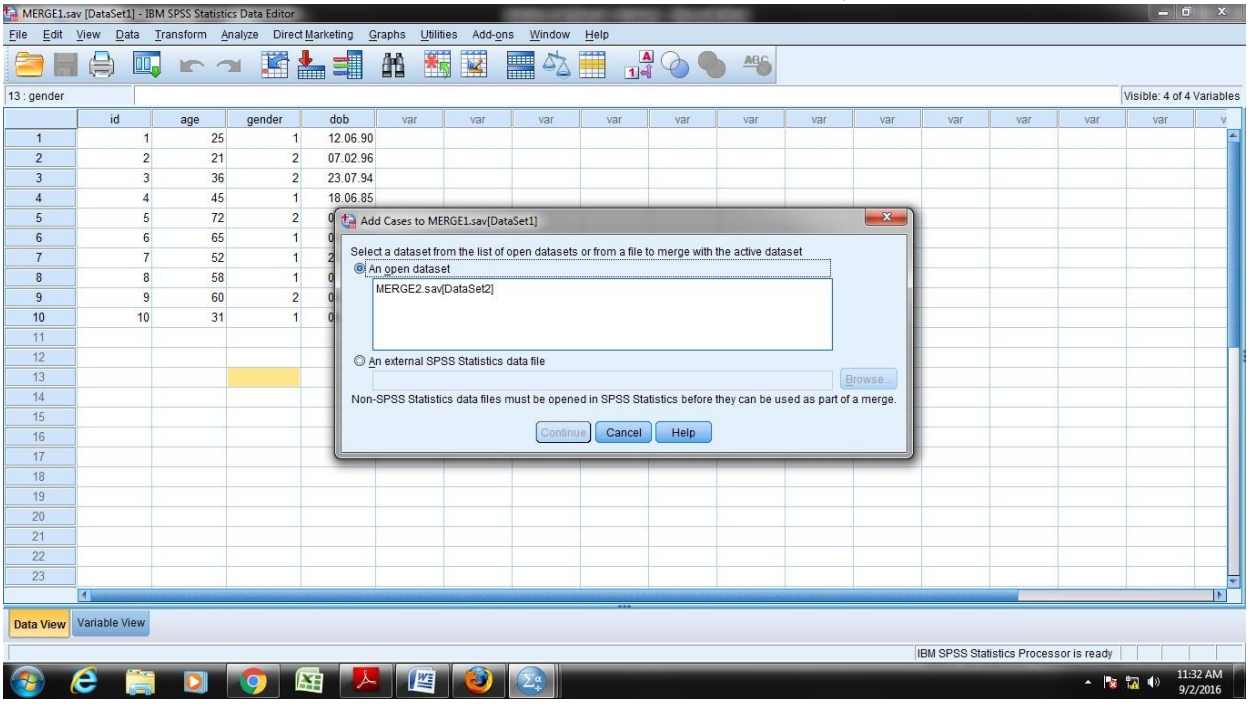
- It is shown above that here the key variable we have chosen is **id**and we have sorted that variable in ascending order either by using **Sort Cases** option or by right-clicking on the variable name **id** and selecting **Sort Ascending** option.

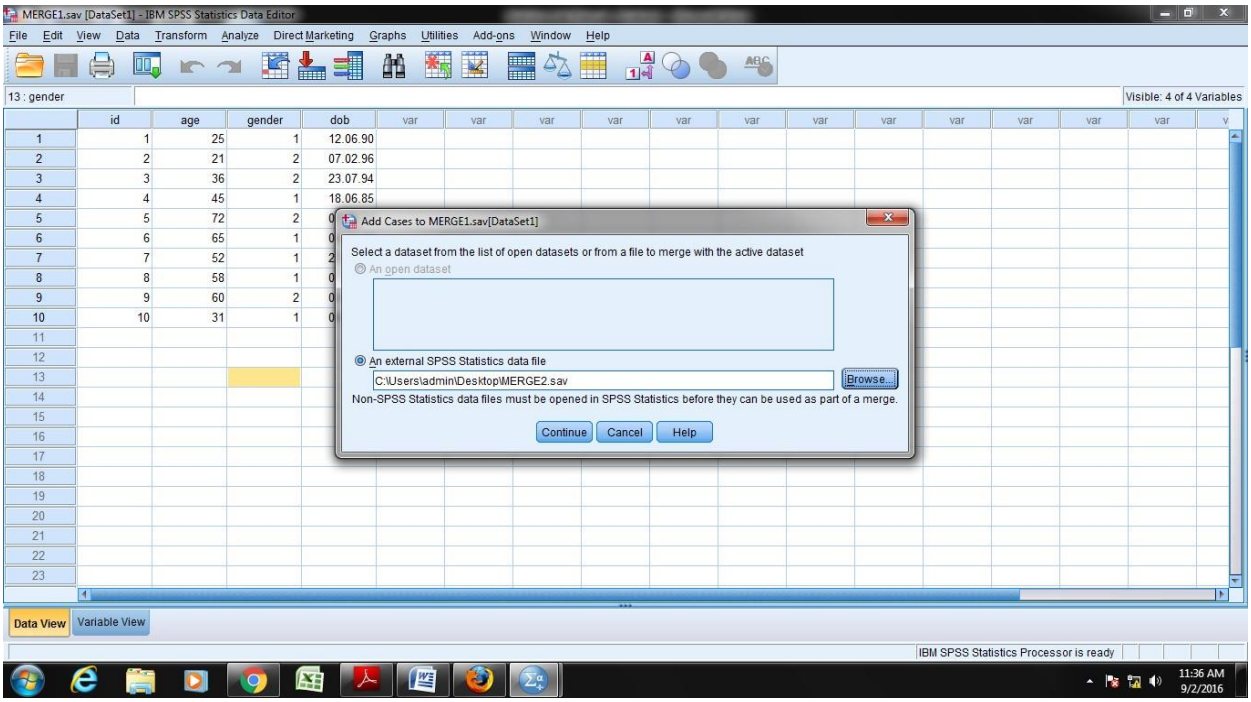


- Choose **Data** from the menu bar and select **Merge Files** and then click on **Add Cases.**

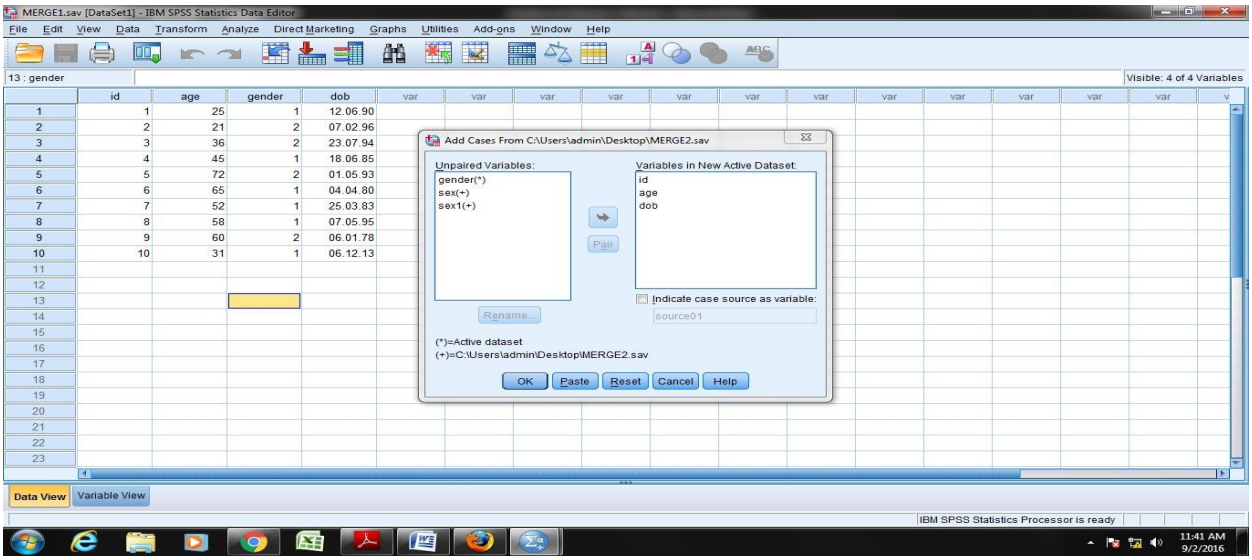


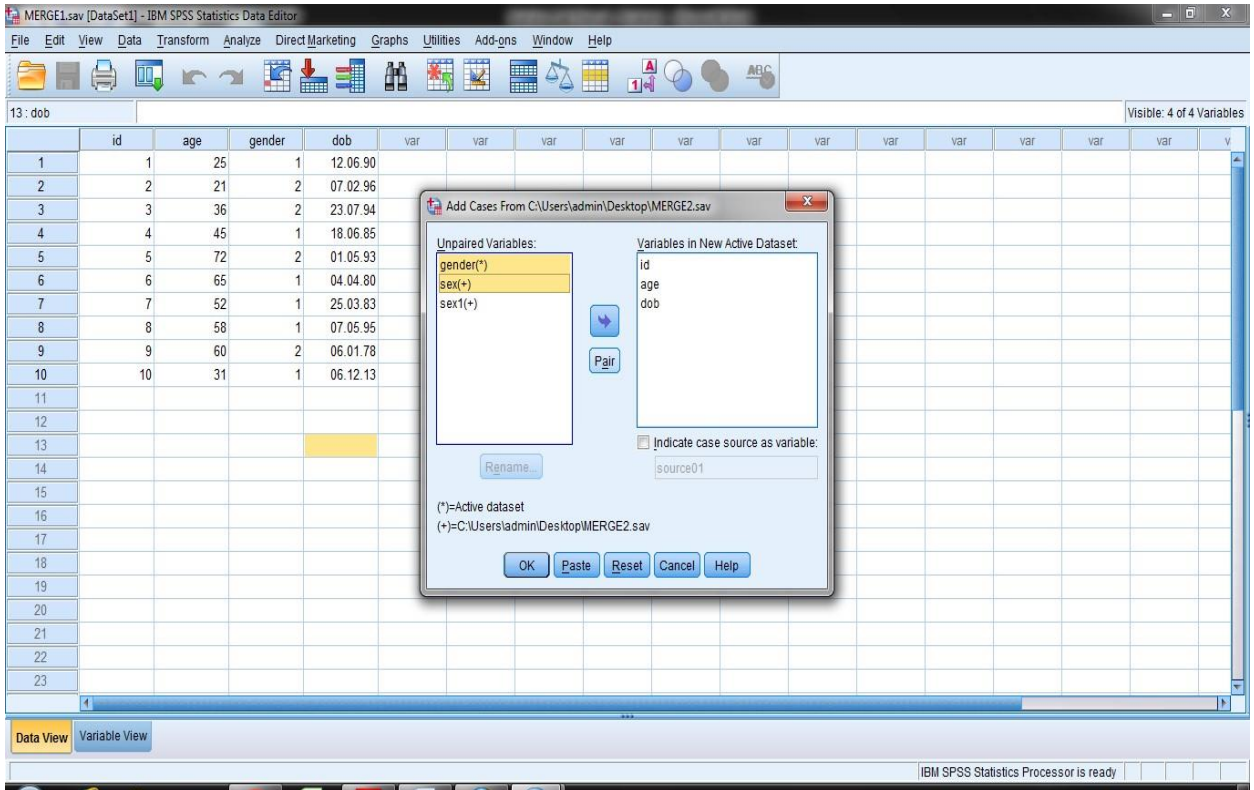
- In the **Add Cases** dialog box shown below we can choose **An open dataset** radio button if the second file we need to merge is opened already or else we can select **An external SPSS Statistics data file** radio button and browse for the second file, i.e. **MERGE2**

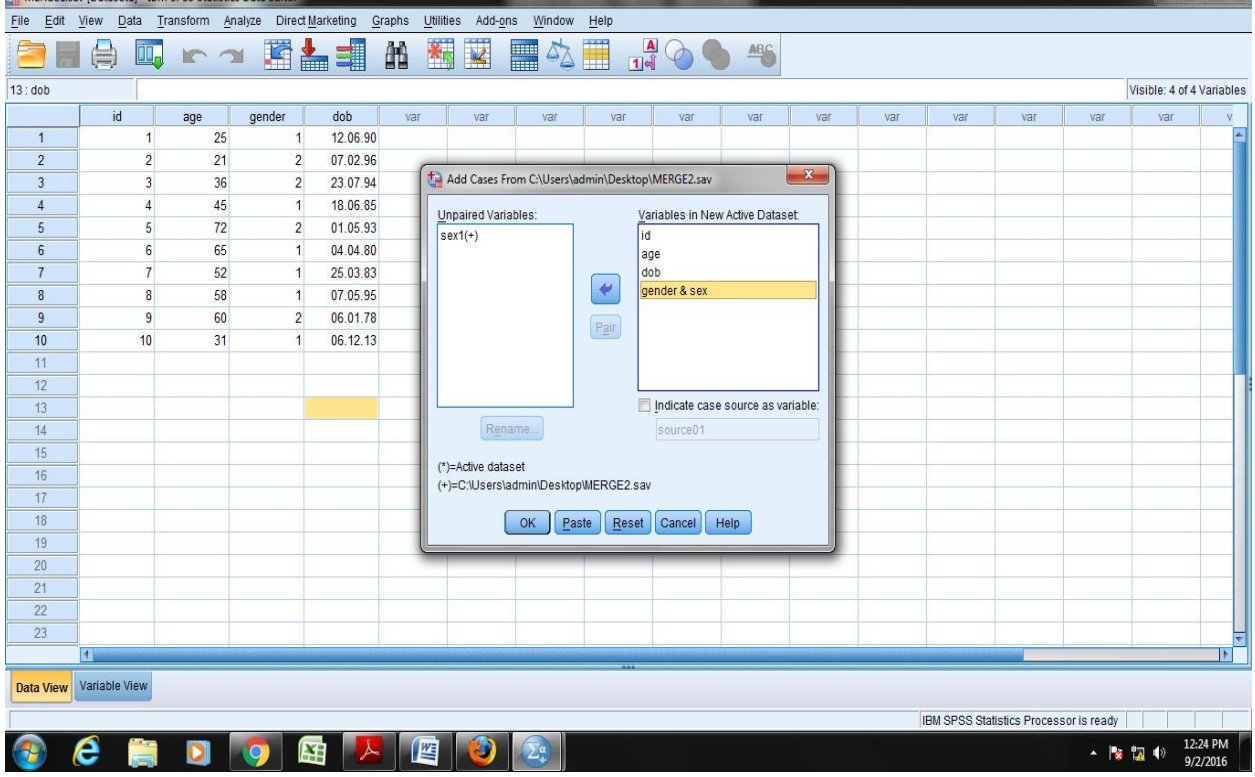




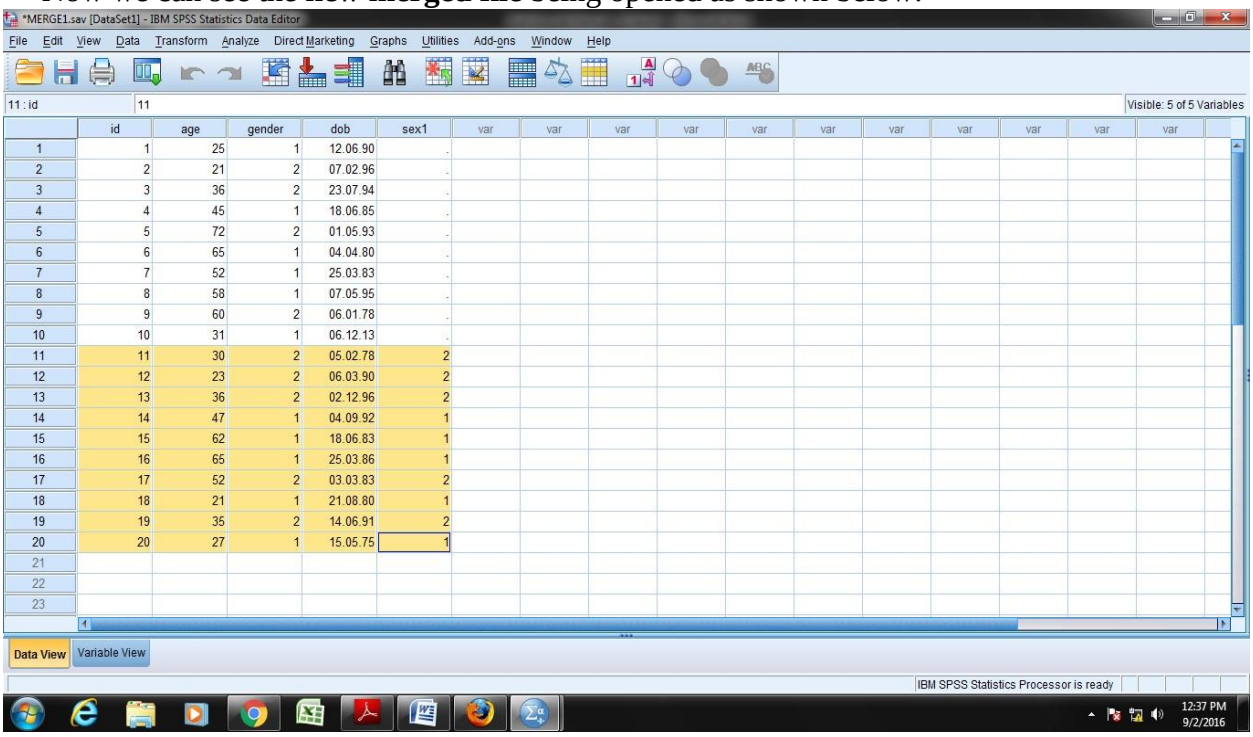
- Then click on **Continue.** We can see that three variables **gender, sex and sex1** are shown under **Unpaired Variables List.**
- Variables from the active dataset are identified with an **asterisk (*)**. Variables from the other dataset are identified with a plus **sign (+).** Identical variables in both the files are grouped under **Variables in New Active Dataset** box namely **id, age &dob.**
- Variables from either data file that do not match a variable name in the other file are given in the **Unpaired Variables List**. You can create pairs from unpaired variables and include them in the newmerged file




MERGE1.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
13 : dob
Visible: 4 of 4 Variables
Add Cases From C:\Users\admin\Desktop\MERGE2.sav
Unpaired Variables:
gender(*)
sex(+)
sex1(+)
Variables in New Active Dataset:
id
age
dob
Pair
Indicate case source as variable:
source01
Rename...
(*)=Active dataset
(+)=C:\Users\admin\Desktop\MERGE2.sav
OK Paste Reset Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
12:24 PM
9/2/2016


MERGE1.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
13 : dob
Visible: 4 of 4 Variables
Add Cases From C:\Users\admin\Desktop\MERGE2.sav
Unpaired Variables:
sex1(+)
Variables in New Active Dataset:
id
age
dob
gender & sex
Pair
Indicate case source as variable:
source01
Rename...
(*)=Active dataset
(+)=C:\Users\admin\Desktop\MERGE2.sav
OK Paste Reset Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
12:24 PM
9/2/2016

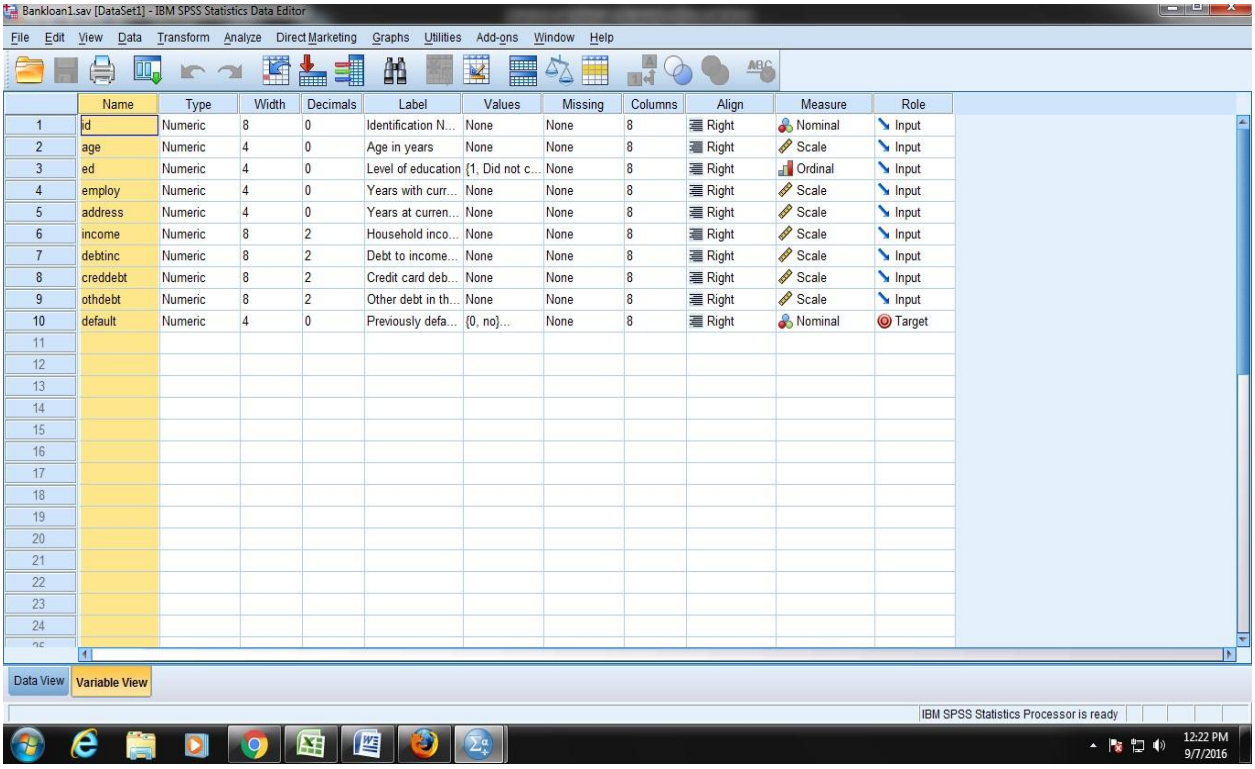
- We can see that there is another variable **sex1** in the **Unpaired Variables** box, we've created this variable as a copy of the variable **sex** in the second data file in order to ensure that the same data has been merged in the appropriate order as it was in the second data file **MERGE2** into the **new merged data file, MERGE1.** So, we need to click on the variable **sex1** and add it to the **Variables in New Active Dataset** box by clicking on the **arrow** tab and then click **OK**
- Now we can see the **new merged file** being opened as shown below:

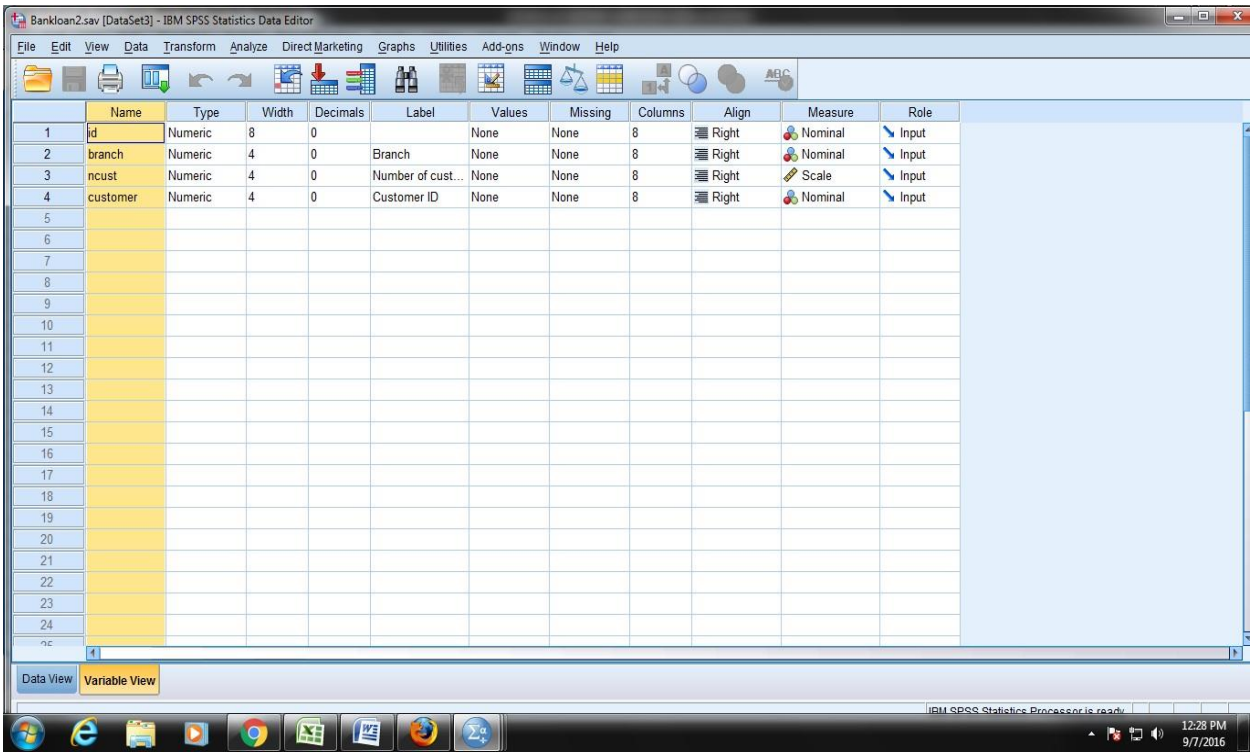


	id	age	gender	dob	sex1	var	var	var	var	var	var	var	var	var	var	var
1	1	25	1	12.06.90	.											
2	2	21	2	07.02.96	.											
3	3	36	2	23.07.94	.											
4	4	45	1	18.06.85	.											
5	5	72	2	01.05.93	.											
6	6	65	1	04.04.80	.											
7	7	52	1	25.03.83	.											
8	8	58	1	07.05.95	.											
9	9	60	2	06.01.78	.											
10	10	31	1	06.12.13	.											
11	11	30	2	05.02.78	2											
12	12	23	2	06.03.90	2											
13	13	36	2	02.12.96	2											
14	14	47	1	04.09.92	1											
15	15	62	1	18.06.83	1											
16	16	65	1	25.03.86	1											
17	17	52	2	03.03.83	2											
18	18	21	1	21.08.80	1											
19	19	35	2	14.06.91	2											
20	20	27	1	15.05.75	1											
21																
22																
23																

- Now in order to **Merge two data files that have Same Cases but Different Variables** we choose **Data → Merge → Add Variables.**

The two data files we are going to merge are **Bankloan1.sav** and **Bankloan2.sav**The active data file is **Bankloan1.sav** and we are going to merge the additional variables in the file **Bankloan2.sav** for the same number of persons into the active data file.





- In our Eg: we take the variable **id as the Key Variable** and Sort it in Ascending order in both the Files given below:

Bankloan1.sav [DataSet4] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

1 : id | 1 | Visible: 10 of 10 Variables

	id	age	ed	employ	address	income	debtinc	creddebt	othdebt	default	var	var	var	var	var	var
1	1	20	3	0	1	17.00	2.30	.04	.35	0						
2	2	20	1	4	0	14.00	9.70	.20	1.16	1						
3	3	21	2	1	2	16.00	18.00	.24	2.64	1						
4	4	21	2	0	1	16.00	6.80	.15	.94	0						
5	5	21	3	0	2	24.00	7.70	.83	1.01	0						
6	6	21	1	5	1	25.00	9.00	.37	1.88	0						
7	7	21	1	4	0	26.00	8.90	1.42	.89	0						
8	8	21	1	1	1	18.00	17.30	.16	2.96	0						
9	9	21	3	0	1	26.00	7.70	1.53	.47	0						
10	10	21	2	0	2	21.00	12.50	.49	2.14	1						
11	11	21	2	2	0	20.00	4.50	.29	.61	1						
12	12	21	1	1	1	16.00	6.30	.14	.87	0						
13	13	21	2	1	0	17.00	10.50	.56	1.23	.						
14	14	21	3	0	1	41.00	19.50	2.37	5.63	.						
15	15	22	1	4	2	24.00	15.60	1.64	2.11	1						
16	16	22	2	0	3	20.00	5.60	.21	.91	0						
17	17	22	1	4	1	16.00	1.20	.08	.11	0						
18	18	22	1	1	3	17.00	18.60	.81	2.36	0						
19	19	22	1	4	3	14.00	17.50	.23	2.22	0						
20	20	22	1	1	0	18.00	4.40	.27	.52	0						
21	21	22	3	0	1	18.00	7.70	.48	.91	0						
22	22	22	1	4	3	19.00	23.10	1.89	2.50	1						
23	23	22	3	1	3	25.00	13.80	1.98	1.47	1						

Data View | Variable View

IBM SPSS Statistics Processor is ready

5:23 PM 9/2/2016

Bankloan2.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

1 : id | 1 | Visible: 4 of 4 Variables

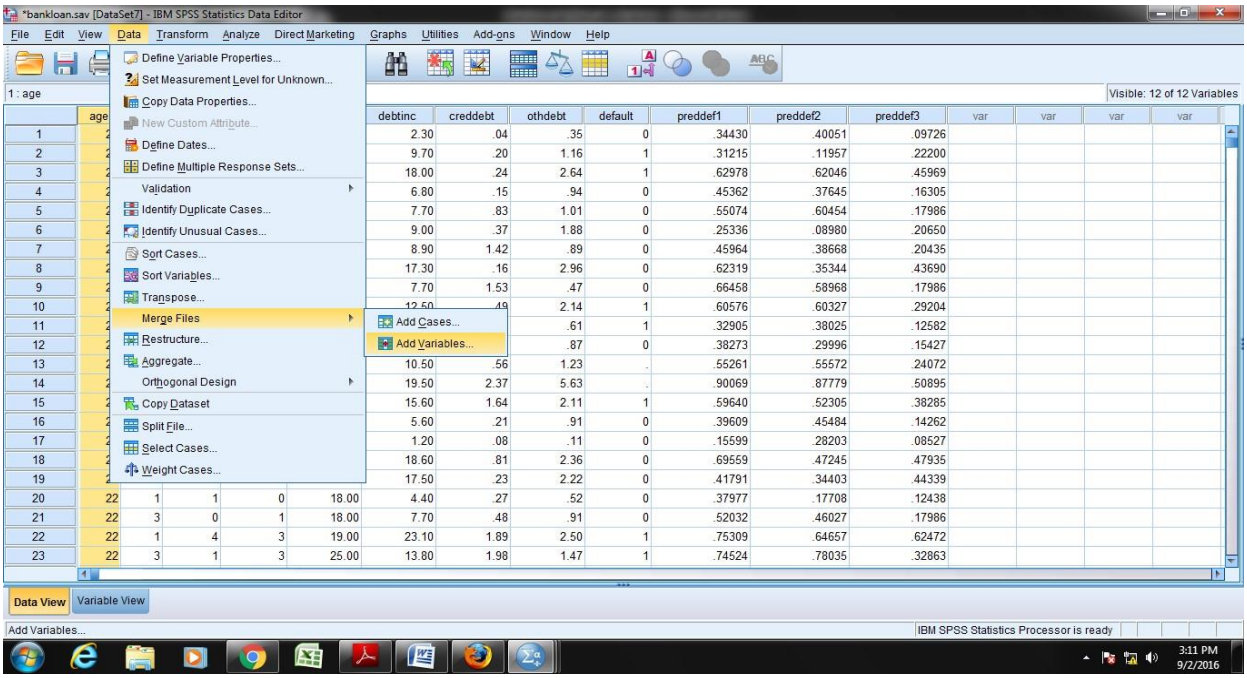
	id	branch	ncust	customer	var	var	var	var	var	var	var	var	var	var	var	var	v
1	1	3	3017	10200													
2	2	3	3017	10234													
3	3	3	3017	10351													
4	4	3	3017	10800													
5	5	3	3017	11859													
6	6	3	3017	12841													
7	7	13	3080	60657													
8	8	13	3080	60949													
9	9	13	3080	61481													
10	10	13	3080	62645													
11	11	13	3080	62744													
12	12	13	3080	62750													
13	13	13	3080	63041													
14	14	15	4809	70567													
15	15	15	4809	70687													
16	16	15	4809	71270													
17	17	15	4809	71699													
18	18	15	4809	71758													
19	19	15	4809	72584													
20	20	15	4809	73864													
21	21	15	4809	74169													
22	22	20	4650	95050													
23	23	20	4650	97629													

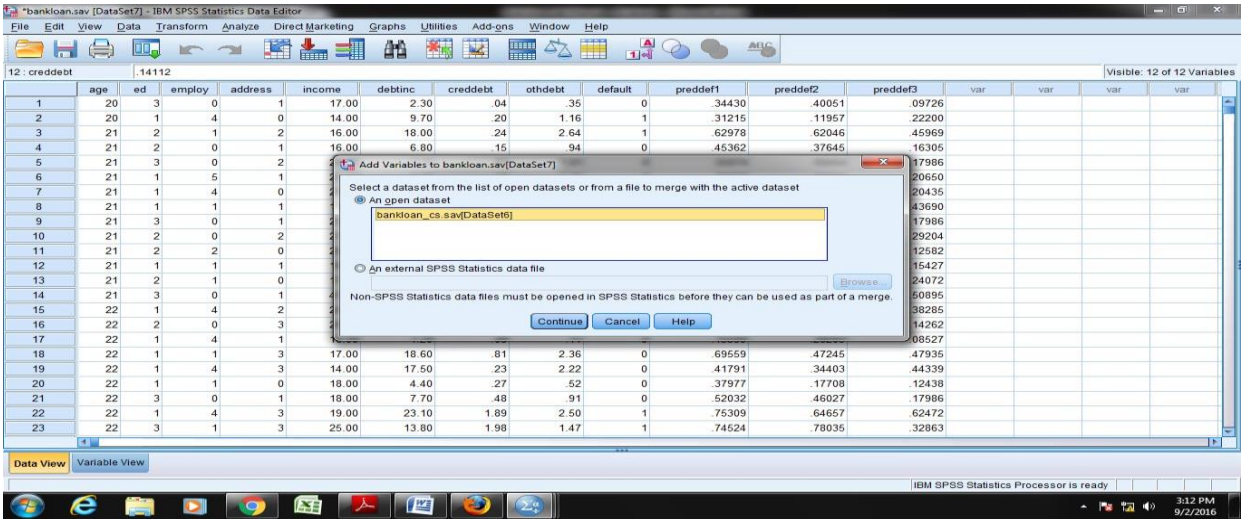
Data View | Variable View

IBM SPSS Statistics Processor is ready

5:23 PM 9/2/2016

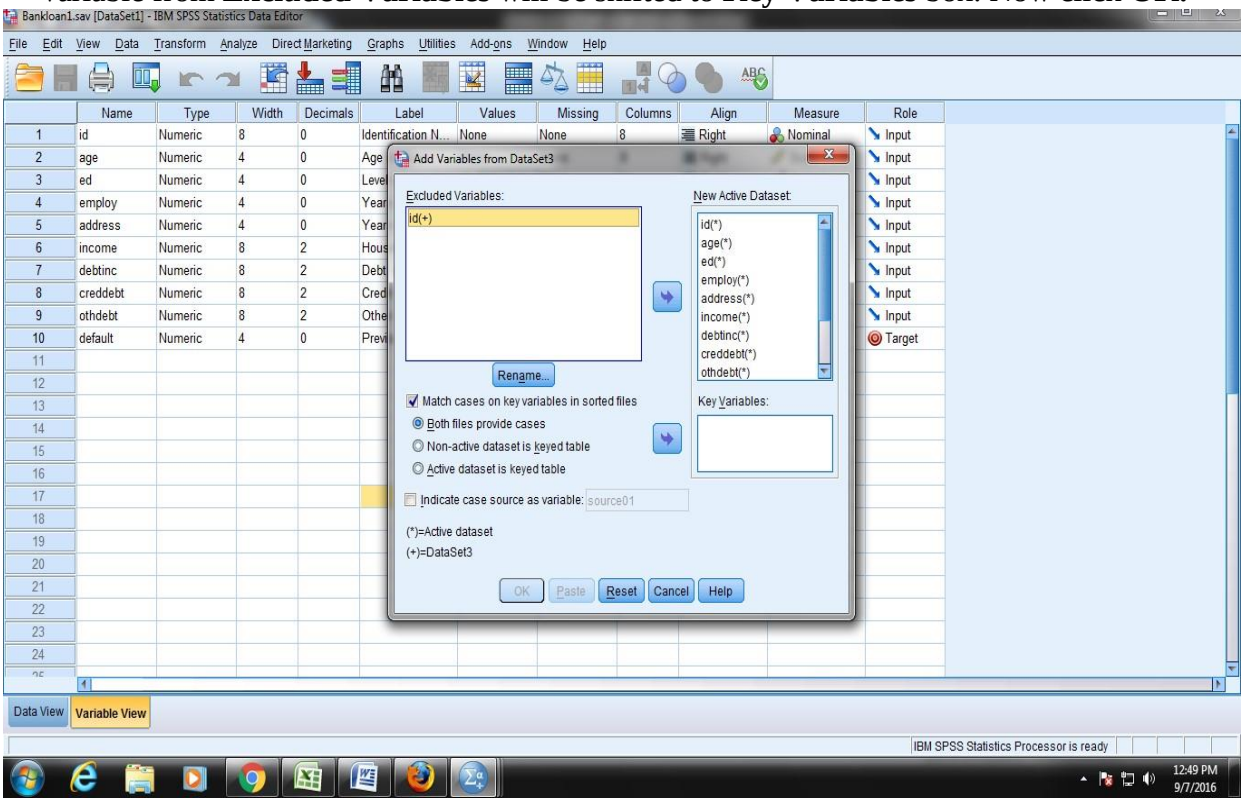
- Now go to the **Data** menu and select the **Merge** option and then select **Add Variables.**

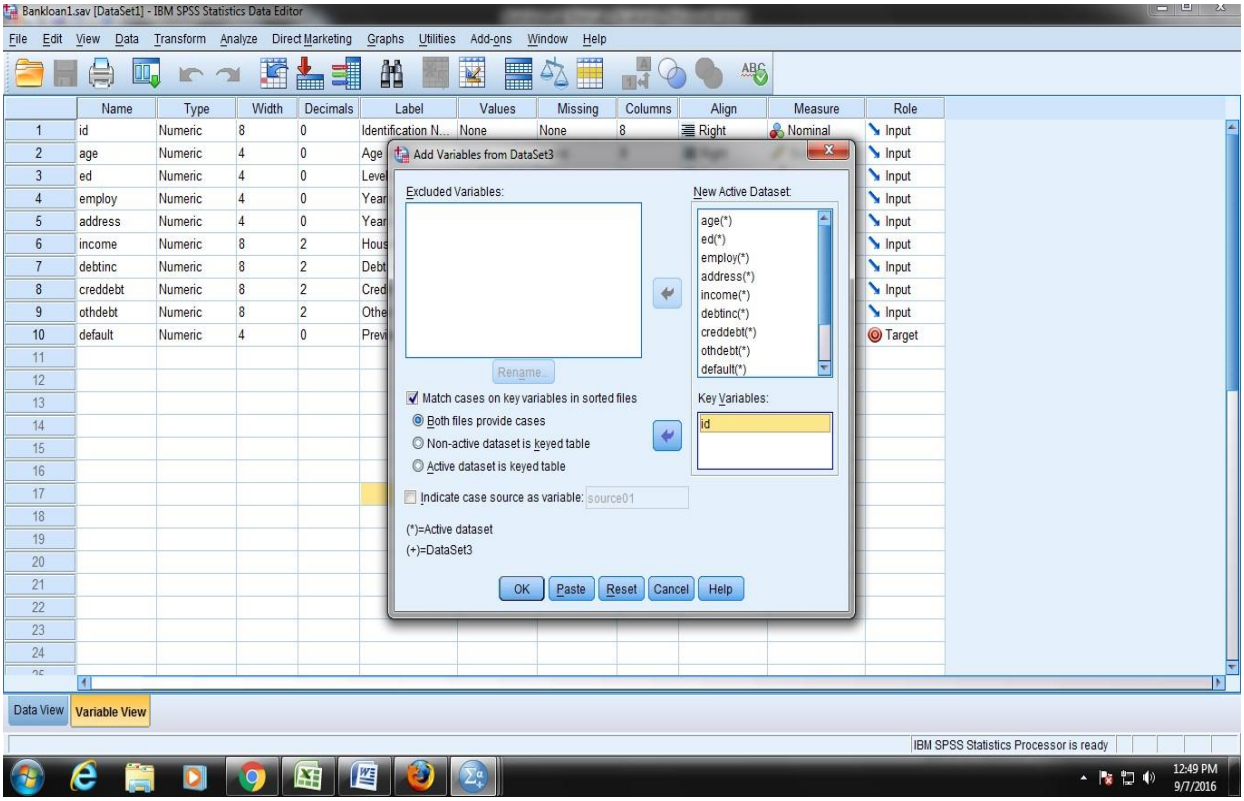




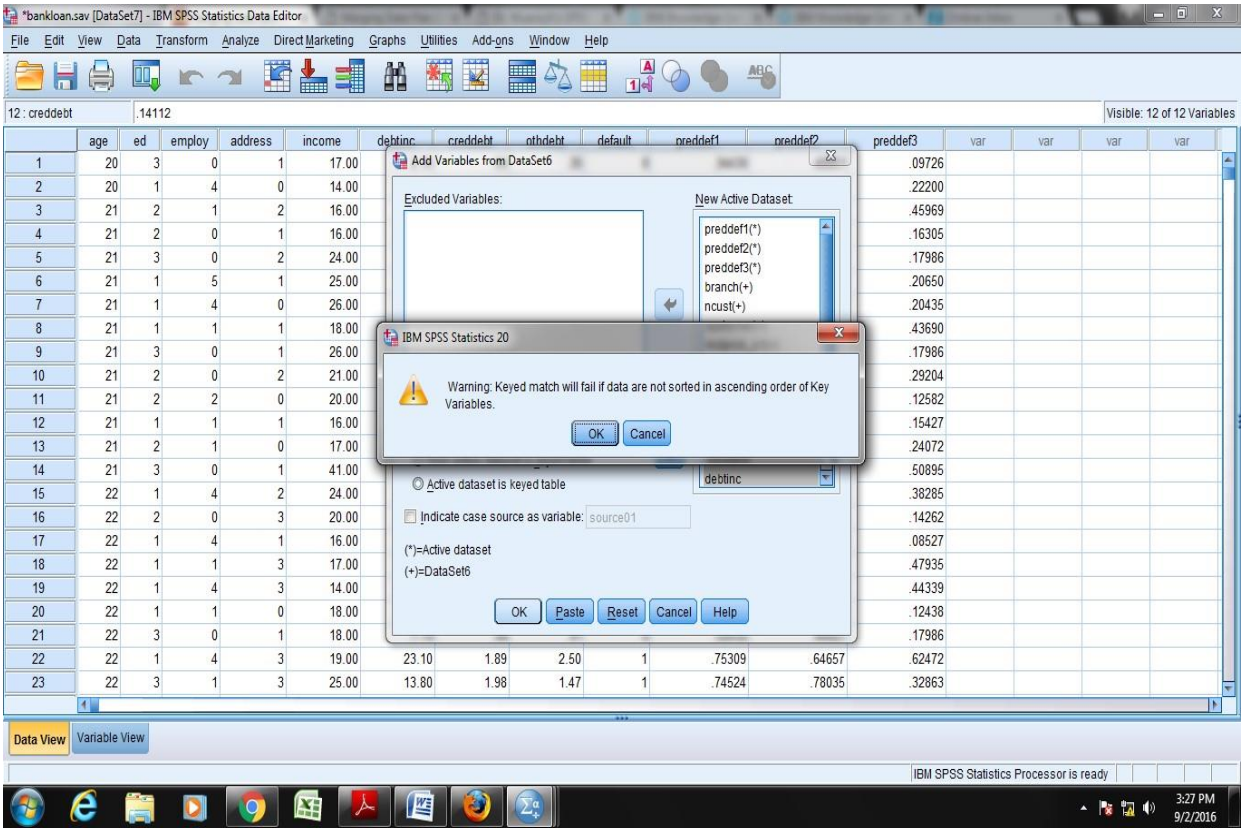
- In the **Add Variables** dialog box shown above we can see that the second file **Bankloan2.sav** is already opened hence it is shown in **An open dataset box** select it and click on **Continue** if it is not already opened **Browse** and open it.
- A dialog box will be opened as shown below in which we can see the **Key Variable** that have same data is shown under **Excluded Variables** box. By default, this list contains any variable names from the other dataset that duplicate variable names in the active dataset. Here the only similar variable is **id.**

- In the **New Active Dataset** box Variables to be included in the new, merged dataset is given. By default, all unique variable names in both datasets are included on this list.
- Select the **Variable, id** and Click on **Match cases on key variables in sorted files** option and also select the **Both files provide cases** radio button and click on the **arrow** tab. The selected variable from **Excluded Variables** will be shifted to **Key Variables** box. Now click **OK.**





- SPSS will give you a warning regarding Sorted Key Variables as shown below. Make sure both files were sorted in ascending order before trying to do a file merge



- Click on **OK** and we can see that the variables in the file **Bankloan2.sav** have been merged in the file **Bankloan1.sav**, for the same number of cases in both the files.

*Bankloan1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	8	0	Identification N...	None	None	8	Right	Nominal	Input
2	age	Numeric	4	0	Age in years	None	None	8	Right	Scale	Input
3	ed	Numeric	4	0	Level of education	{1, Did not c...	None	8	Right	Ordinal	Input
4	employ	Numeric	4	0	Years with curr...	None	None	8	Right	Scale	Input
5	address	Numeric	4	0	Years at curren...	None	None	8	Right	Scale	Input
6	income	Numeric	8	2	Household inco...	None	None	8	Right	Scale	Input
7	debtinc	Numeric	8	2	Debt to income...	None	None	8	Right	Scale	Input
8	creddebt	Numeric	8	2	Credit card deb...	None	None	8	Right	Scale	Input
9	othdebt	Numeric	8	2	Other debt in th...	None	None	8	Right	Scale	Input
10	default	Numeric	4	0	Previously defa...	{0, no}...	None	8	Right	Nominal	Target
11	branch	Numeric	4	0	Branch	None	None	8	Right	Nominal	Input
12	ncust	Numeric	4	0	Number of cust...	None	None	8	Right	Scale	Input
13	customer	Numeric	4	0	Customer ID	None	None	8	Right	Nominal	Input
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											

Data View | Variable View

IBM SPSS Statistics Processor is ready

12:54 PM 9/7/2016

*Bankloan1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

1 : branch | 3 | Visible: 13 of 13 Variables

	id	age	ed	employ	address	income	debtinc	creddebt	othdebt	default	branch	ncust	customer	var	var	var
1	1	20	3	0	1	17.00	2.30	.04	.35	0	3	3017	10200			
2	2	20	1	4	0	14.00	9.70	.20	1.16	1	3	3017	10234			
3	3	21	2	1	2	16.00	18.00	.24	2.64	1	3	3017	10351			
4	4	21	2	0	1	16.00	6.80	.15	.94	0	3	3017	10800			
5	5	21	3	0	2	24.00	7.70	.83	1.01	0	3	3017	11859			
6	6	21	1	5	1	25.00	9.00	.37	1.88	0	3	3017	12841			
7	7	21	1	4	0	26.00	8.90	1.42	.89	0	13	3080	60657			
8	8	21	1	1	1	18.00	17.30	.16	2.96	0	13	3080	60949			
9	9	21	3	0	1	26.00	7.70	1.53	.47	0	13	3080	61481			
10	10	21	2	0	2	21.00	12.50	.49	2.14	1	13	3080	62645			
11	11	21	2	2	0	20.00	4.50	.29	.61	1	13	3080	62744			
12	12	21	1	1	1	16.00	6.30	.14	.87	0	13	3080	62750			
13	13	21	2	1	0	17.00	10.50	.56	1.23	.	13	3080	63041			
14	14	21	3	0	1	41.00	19.50	2.37	5.63	.	15	4809	70567			
15	15	22	1	4	2	24.00	15.60	1.64	2.11	1	15	4809	70687			
16	16	22	2	0	3	20.00	5.60	.21	.91	0	15	4809	71270			
17	17	22	1	4	1	16.00	1.20	.08	.11	0	15	4809	71699			
18	18	22	1	1	3	17.00	18.60	.81	2.36	0	15	4809	71758			
19	19	22	1	4	3	14.00	17.50	.23	2.22	0	15	4809	72584			
20	20	22	1	1	0	18.00	4.40	.27	.52	0	15	4809	73864			
21	21	22	3	0	1	18.00	7.70	.48	.91	0	15	4809	74169			
22	22	22	1	4	3	19.00	23.10	1.89	2.50	1	20	4650	95050			
23	23	22	3	1	3	25.00	13.80	1.98	1.47	1	20	4650	97629			

Data View | Variable View

IBM SPSS Statistics Processor is ready

12:54 PM 9/7/2016

RESULT: Sorting, splitting and merging data files using SPSS is done successfully

EX.NO: 05 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

MEASURE OF CENTRAL TENDENCY MEAN, MEDIAN AND MODE

AIM

To calculate the measures of central tendency mean, median and mode

PROCEDURE:

Step 1: Go to start menu -> All program -> IBM SPSS statistics 20

Step 2: Click variable view and enter the data as Group 1 and Group 2

Step 3: Click data view and enter the appropriate data nearly 6 to 7 values

Step 4: Then Go to Analyze->Descriptive statistics->Frequencies->Statistics->Mark Mean, Median and Mode□Continue->Click Ok->

Step5: The output has been displayed on the output screen

Sample input

group1	group2	var	var
67	34		
56	23		
34	78		
87	92		
45	34		
63	67		
47	45		

Sample output

Statistics

		group1	group2
N	Valid	7	7
	Missing	0	0
Mean		57.00	53.29
Median		56.00	45.00
Mode		34[a]	34

a. Multiple modes exist. The smallest value is shown

RESULT: The measures of central tendency mean, median and mode files using SPSS is done successfully

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
COLLEGE OF SCIENCE AND HUMANITIES
Ramapuram Campus

EX.NO: 06 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

METHODS OF DISPERSION-STANDARD DEVIATION, QUARTILES, SKEWNESS, KURTOSIS

AIM:

To Calculate Methods of Dispersion-Standard Deviation, Quartiles, Skewness, Kurtosis

PROCEDURE:

Step 1: Start home□click All programs □ IBM SPSS Statistics 20

Step 2: Go to variable view and defines the field you want

Step 3: Go to data view and give input for the fields you want to define

Step 4:Go to Analyze->Descriptive Statistics->Frequencies->Statistics->Mark Quartiles, Standard deviation, Skewness, Kurtosis->Continue->Select the Variables and click Ok

Step5: The output has been displayed in output screen.

Sample input :

Name	Mark1	Mark2
A1	54	32
A2	78	68
A3	65	54
A4	45	46
A5	57	36
A6	85	89
A7	98	98

Sample output:

Statistics

		Name	Mark1	Mark2
N	Valid	7	7	7
	Missing	0	0	0
Std. Deviation			18.881	25.625
Skewness			.386	.499
Std. Error of Skewness			.794	.794
Kurtosis			-1.034	-1.377
Std. Error of Kurtosis			1.587	1.587
Percentiles	25		54.00	36.00
	50		65.00	54.00
	75		85.00	89.00

RESULT: Methods of Dispersion-Standard Deviation, Quartiles, Skewness, Kurtosis using SPSS is done successfully

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
COLLEGE OF SCIENCE AND HUMANITIES
Ramapuram Campus

EX.NO: 07 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

REGRESSION TREND

AIM:

Calculation of Regression Trend Trend Line.

PROCEDURE:

Step 1 Go to the variable view and enter the fields you want. Step 2 Go to the data view and give inputs for the fields.

Step 3 Then go to Analyze Regression Linear Move the field names to right Choose plots,move *zpred to Y axis and move *zresid to X axis and then press continue.

Step 4 Select Statistics Descriptions Continue and then press OK.

Sample input

Name	Age	BP
Random1	18	92
Random2	22	78
Random3	56	121
Random4	19	82

Sample output

Correlations

		BP	Age
Pearson Correlation	BP	1.000	.924
	Age	.924	1.000
Sig. (1-tailed)	BP	.	.038
	Age	.038	.
N	BP	4	4
	Age	4	4

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	64.975	9.421	9.24	6.897	.020
	Age	.983	.287		3.425	.076

a. Dependent Variable: BP

RESULT: The Calculation of Regression Trend using SPSS is done successfully

EX.NO: 08 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

CORRELATION ANALYSIS

AIM:

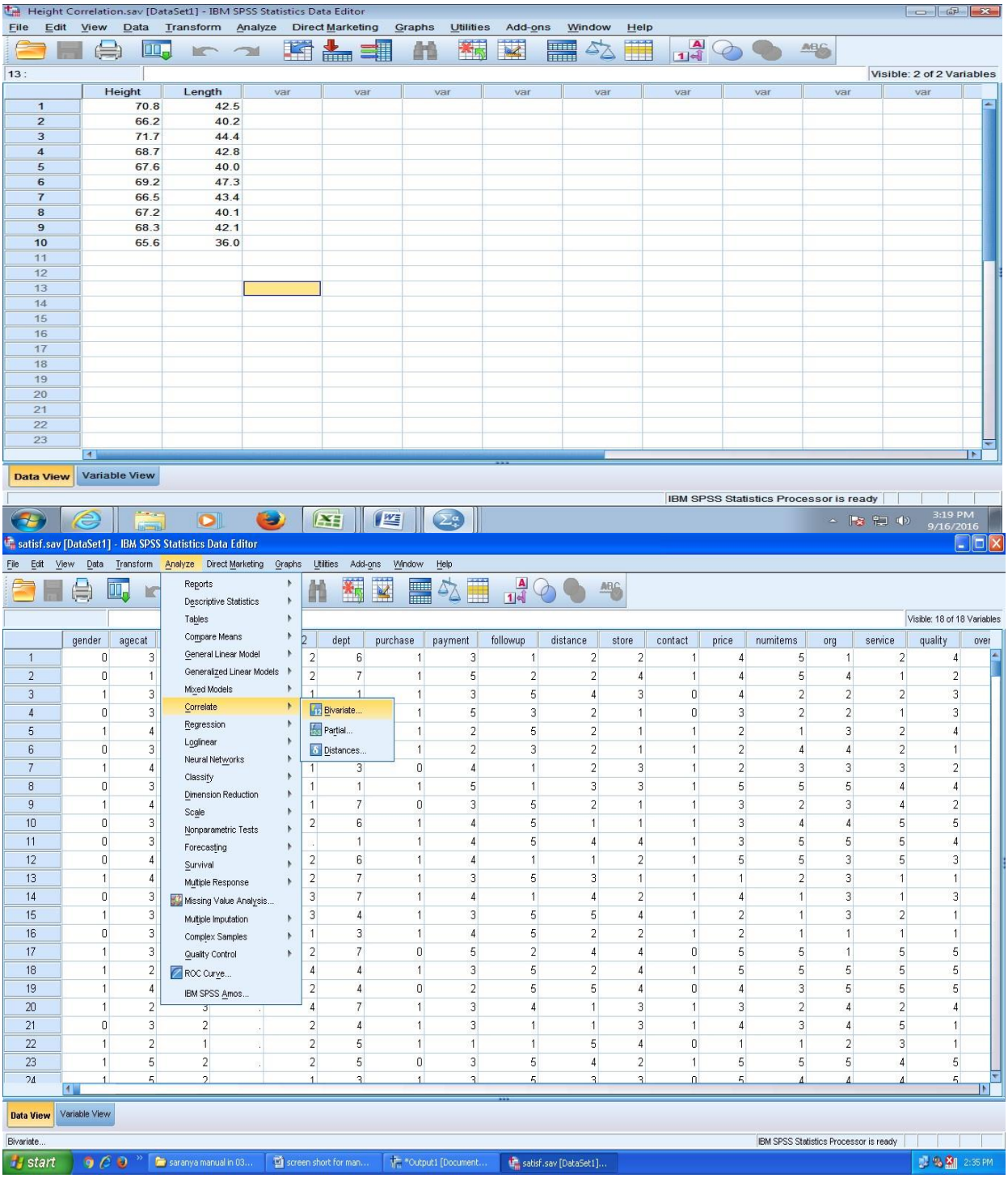
To calculate how strongly the variables are related to each other using SPSS.

INTRODUCTION:

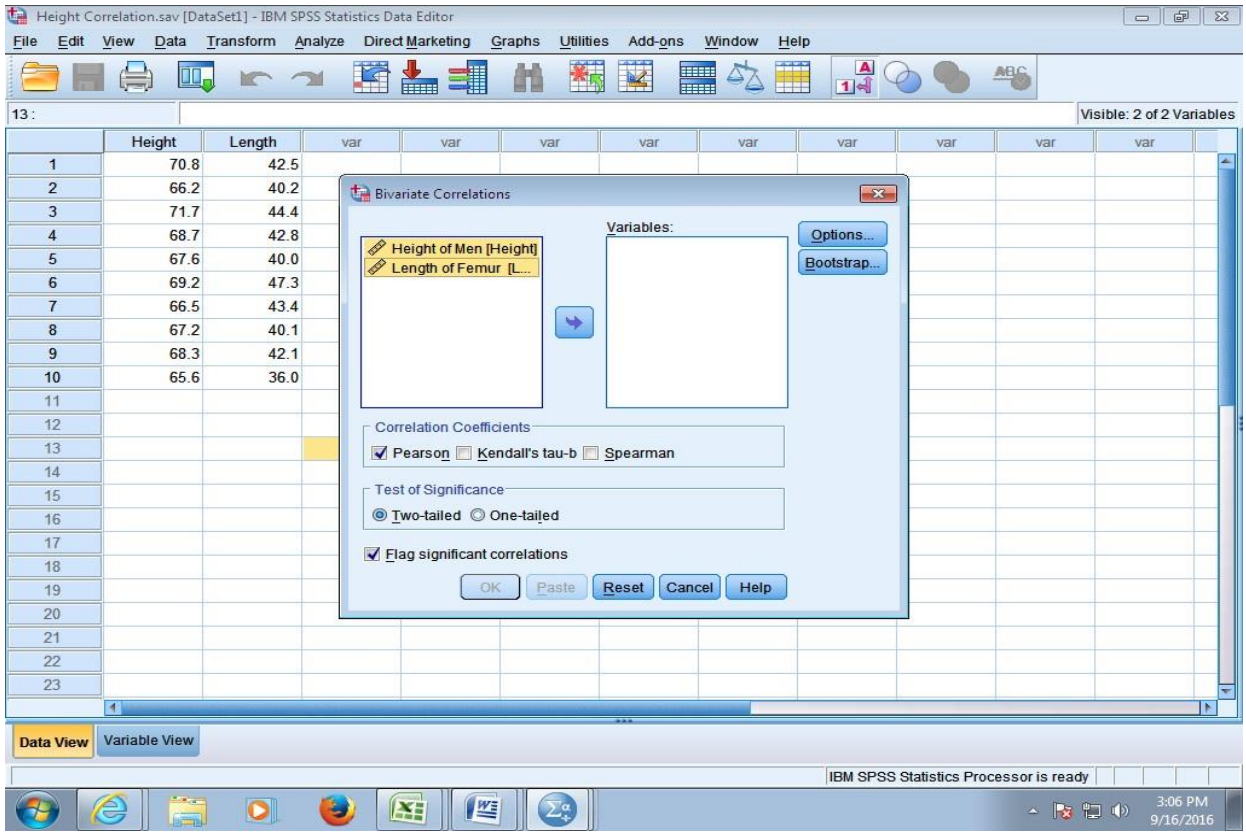
- **Correlation** is a statistical technique that can show whether and how strongly pairs of variables are related. For example, height and weight are related; Correlation can tell you just how much of the variation in peoples' weights is related to their heights.
- Like all statistical techniques, correlation is only appropriate for certain kinds of data. **Correlation works for quantifiable data** in which numbers are meaningful, usually quantities of some sort. It cannot be used for purely categorical data, such as gender, brands purchased, or favorite color.
- The main result of a correlation is called the **correlation coefficient** (or "**r**"). It ranges from -1.0 to +1.0. The closer r is to +1 or -1, the more closely the two variables are related.
- If **'r'**is close to **0**, it means there is no relationship between the variables. If **'r'** is positive, it means that as one variable gets larger the other gets larger. If **'r'**is negative it means that as one gets larger, the other gets smaller (often called an **"inverse"** correlation).
- While correlation coefficients are normally reported as **'r'**, squaring them makes them easier to understand. The square of the coefficient (or **r square**) is equal to the percent of the variation in one variable that is related to the variation in the other.
- A correlation coefficient of +1 indicates a perfect positive correlation. As variable X increases, variable Y increases. As variable X decreases, variable Y decreases.

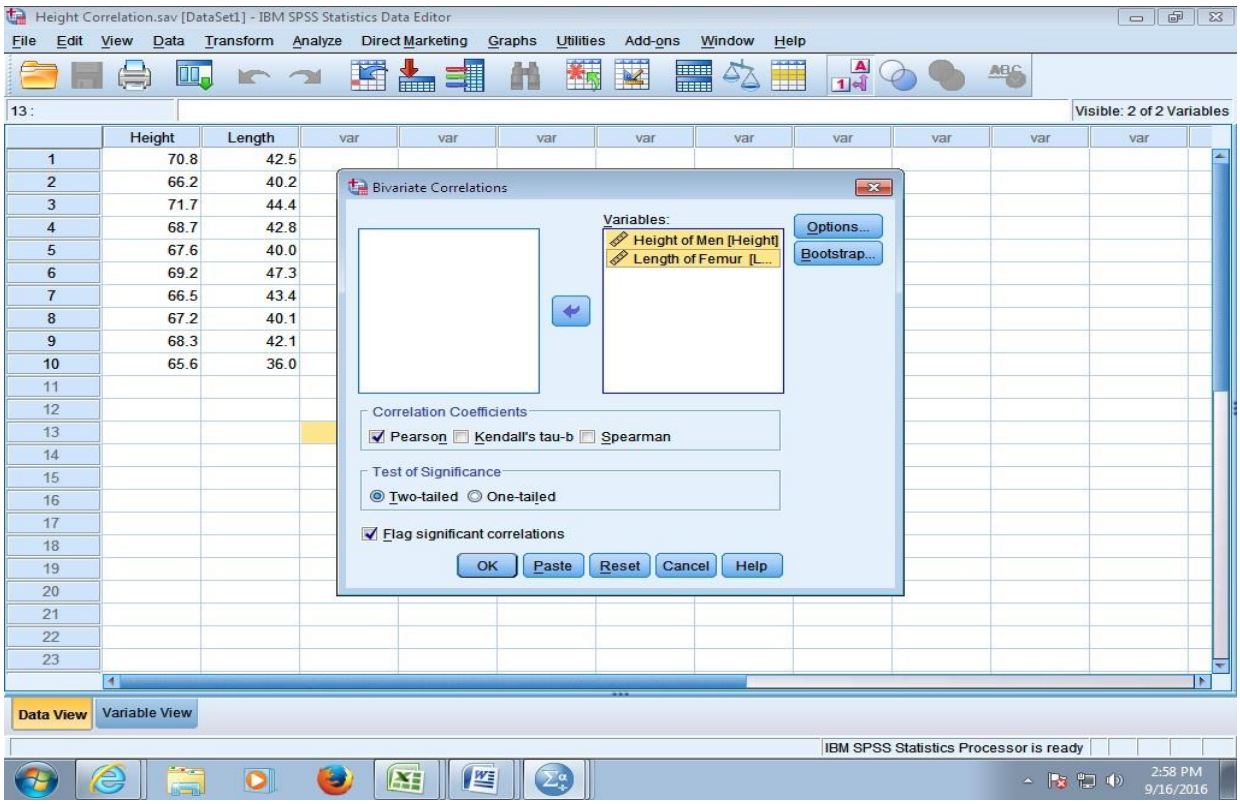
PROCEDURE:

- Select **Height Correlation** data set.
- Go to **Analyze→Correlate→Bivariate…**



- In the **Bivariate Correlations** dialog box, select the two variables **Height** and **Length** and put in the Variable box by clicking on the arrow mark and click on **OK.**


Height Correlation.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
13 :
Visible: 2 of 2 Variables
Height
Length
Bivariate Correlations
Height of Men [Height]
Length of Femur [L...
Variables:
Options...
Bootstrap...
Correlation Coefficients
Pearson
Kendall's tau-b
Spearman
Test of Significance
Two-tailed
One-tailed
Flag significant correlations
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
3:06 PM
9/16/2016


Height Correlation.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
13 :
Visible: 2 of 2 Variables
Height
Length
Bivariate Correlations
Variables:
Height of Men [Height]
Length of Femur [L...
Options...
Bootstrap...
Correlation Coefficients
Pearson
Kendall's tau-b
Spearman
Test of Significance
Two-tailed
One-tailed
Flag significant correlations
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
2:58 PM
9/16/2016

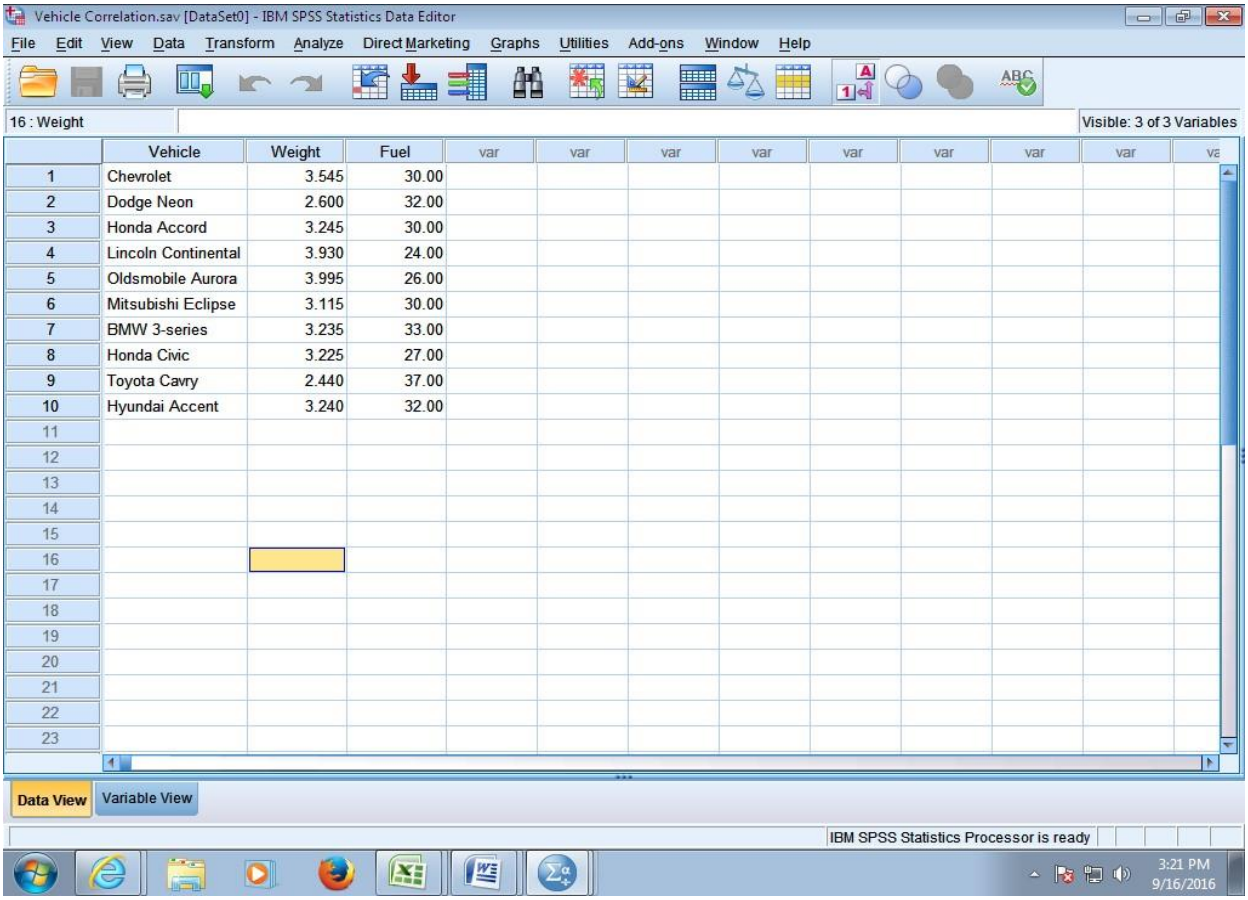
OUTPUT:
Correlations

Correlations

		Height of Men	Length of Femur
Height of Men	Pearson Correlation	1	.651*
	Sig. (2-tailed)		.042
	N	10	10
Length of Femur	Pearson Correlation	.651*	1
	Sig. (2-tailed)	.042	
	N	10	10

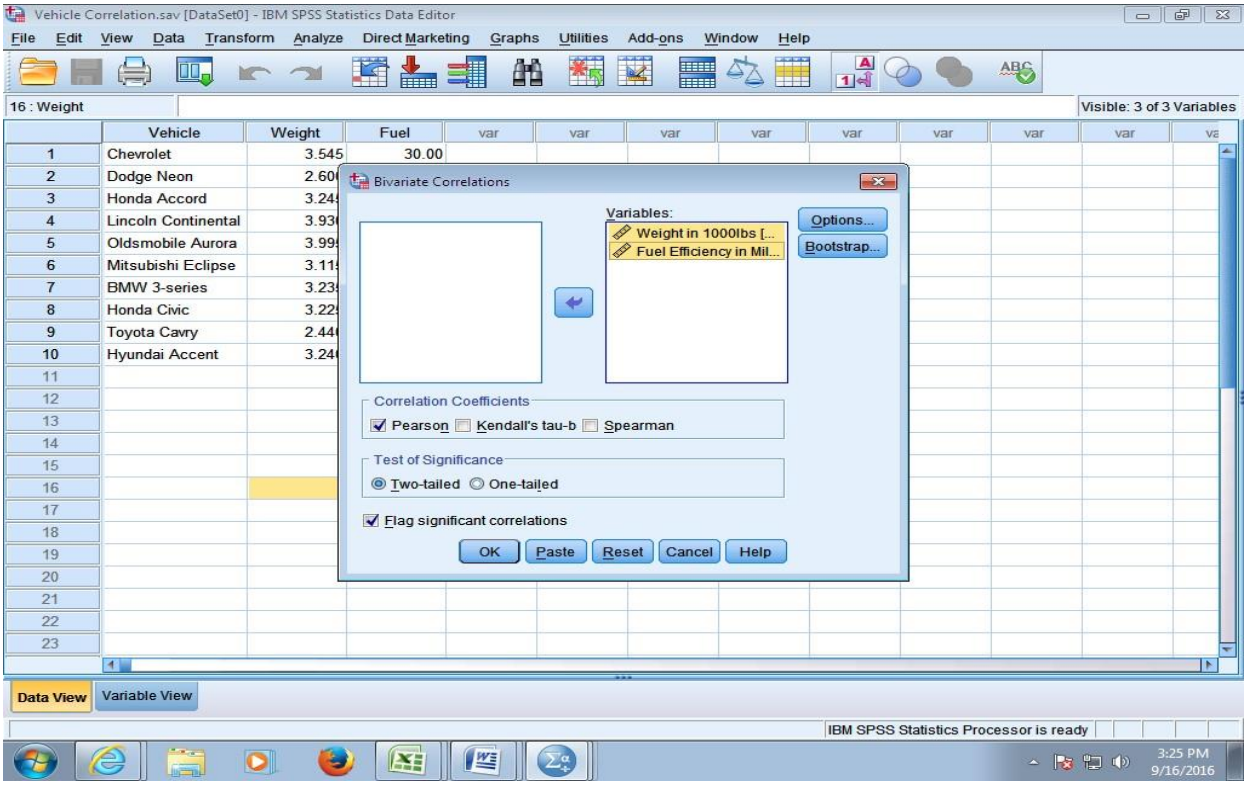
*. Correlation is significant at the 0.05 level (2-tailed).

- The correlation coefficient we obtained is **0.651** which determines that we've obtained a **Positive Correlation** (between 0 to +1). This shows that the value of height increases with that of the Femur Length or vice-versa.
- **Significant value** must be **< 0.05** which means that the "**there is significant difference**" between the variables hence we reject the null hypothesis.
- If the Significance value is> **0.05** then we accept the null hypothesis that **"there is no significant difference".** Here we've got **0.042.**
- Now open the data set, **Vehicle Correlation.sav**



- Again go to **Analyze→Correlate→Bivariate…**

- In the **Bivariate Correlations** dialog box, select the variables and put them in the **Variables** box and click on **OK.**



OUTPUT:

Correlations:

Correlations

		Weight in 1000lbs	Fuel Efficiency in Miles/gallon
Weight in 1000lbs	Pearson Correlation	1	-.839**
	Sig. (2-tailed)		.002
	N	10	10
Fuel Efficiency in Miles/gallon	Pearson Correlation	-.839**	1
	Sig. (2-tailed)	.002	
	N	10	10

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

RESULT:

Calculated how strongly the variables are related to each other for the given data using SPSS.

EX.NO: 09 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

INDEPENDENT SAMPLE T-TEST

AIM:

*Test of significance for single and two samples (1-test for mean and standard deviation;2-test for proposition)

PROCEDURE:

*In variable view enter data as machine. Enter in values and select 1=machine a, 2=machine b. Enter field name as quality and count.

*In data view enter the required values.

TEST FOR MEAN AND STANDARD DEVIATION:

*Analyze->compare means->independent variable- >sample t-test

*Move machine to grouping variables, quality and count to test variable.

*For machine enter into define groups and select group1=1, group2=2.

TO CREATE WEIGHT BY MACHINE.

*Data->weight cases ->enable weight cases by, move machine ->frequency variable.

TEST FOR PROPOSITION:

*Analyze->descriptive statistics->crosstabs.

*Move machine to rows and quality to columns.

*Enter into statistics and enable chi square then select continue, press ok.

Sample input

	machine	quality	count
1	1	2	25
2	2	1	3
3	1	3	40
4	2	2	18

Sample Output

Sample Output:

Test for mean and standard deviation

T-Test

[DataSet0]

Group Statistics

	machine	N	Mean	Std. Deviation	Std. Error Mean
quality	machinea	2	2.50	.707	.500
	machineb	2	1.50	.707	.500
count	machinea	2	32.50	10.607	7.500
	machineb	2	10.50	10.607	7.500

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
quality	Equal variances assumed	.	.	1.414	2	.293	1.000	.707	-2.042	4.042
	Equal variances not			1.414	2.000	.293	1.000	.707	-2.042	4.042

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	29.297[a]	2	.000
Likelihood Ratio	37.141	2	.000
Linear-by-Linear Association	28.642	1	.000
N of Valid Cases	86		

a. 2 cells (33.3%) have expected count less than 5. The minimum expected count is .73.

RESULT: The Test of significance for single and two samples using SPSS is done successfully

EX.NO: 10 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

INDEPENDENT SAMPLE T-TESTS

AIM:

To learn how to do Independent Sample T-test in SPSS

INTRODUCTION:

The independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable.

For example, you could use an independent t-test to understand whether first year graduate salaries differed based on gender (i.e., your dependent variable would be "first year graduate salaries" and your independent variable would be "gender", which has two groups: "male" and "female").

When you choose to analyze your data using an independent t-test, part of the process involves checking to make sure that the data you want to analyze can actually be analyzed using an independent t-test. You need to do this because it is only appropriate to use an independent t-test if your data "passes" six assumptions that are required for an independent t-test to give you a valid result.

Assumption #1: Your dependent variable should be measured on a continuous scale (i.e., it is measured at the interval or ratio level).

Assumption #2: Your independent variable should consist of two categorical, independent groups.

Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves

Assumption #4: There should be no significant outliers. Outliers are simply single data points within your data that do not follow the usual pattern

Assumption #5: Your dependent variable should be approximately normally distributed for each group of the independent variable.

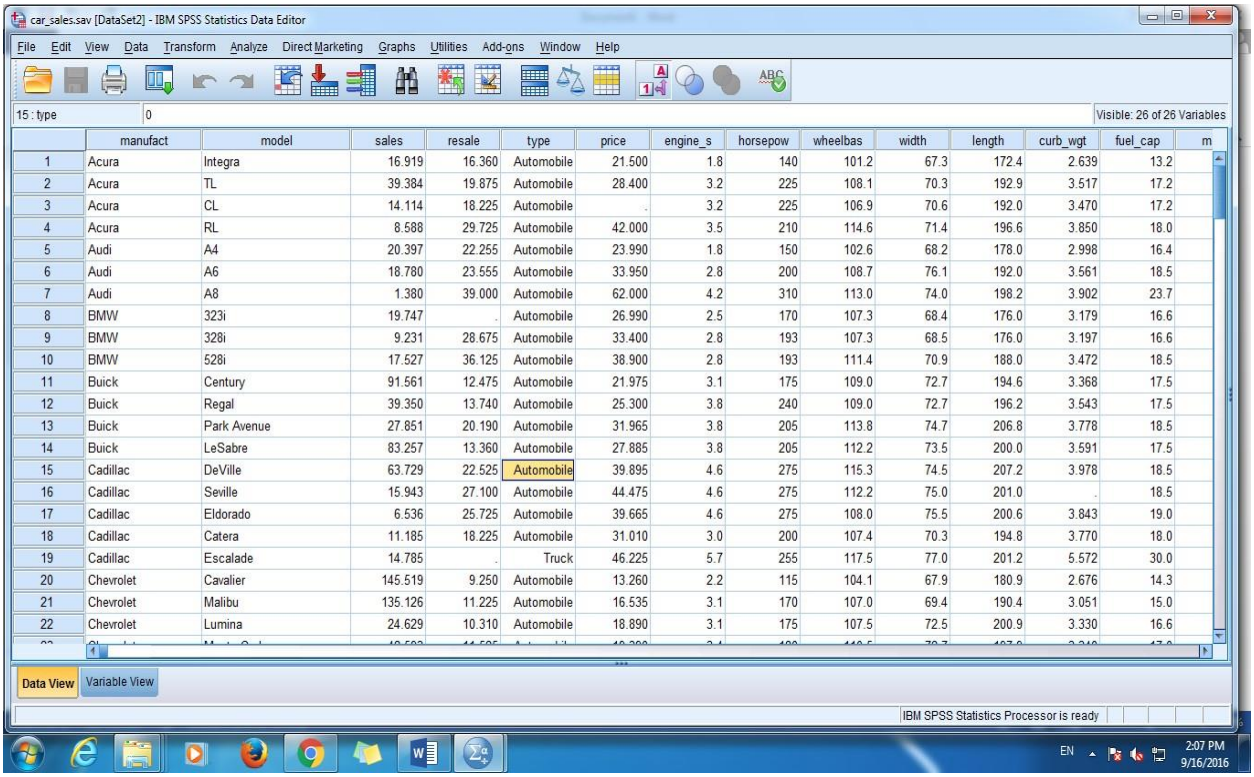
Assumption #6: There needs to be homogeneity of variances. You can test this assumption in SPSS Statistics using Levene's test for homogeneity of variances.

SCENARIO:

Finding whether there is any significant difference between the **AverageSales** based on the **Vehicle type** (Automobile or Truck).

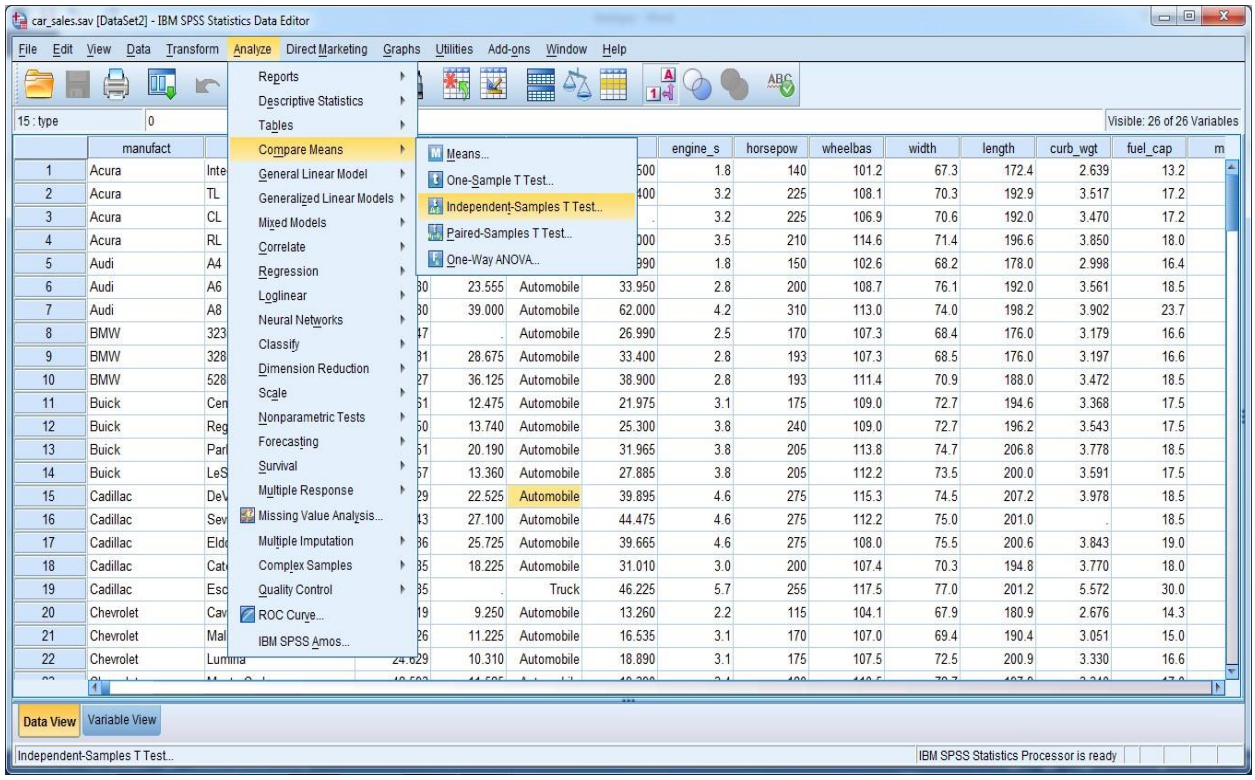
PROCEDURE:

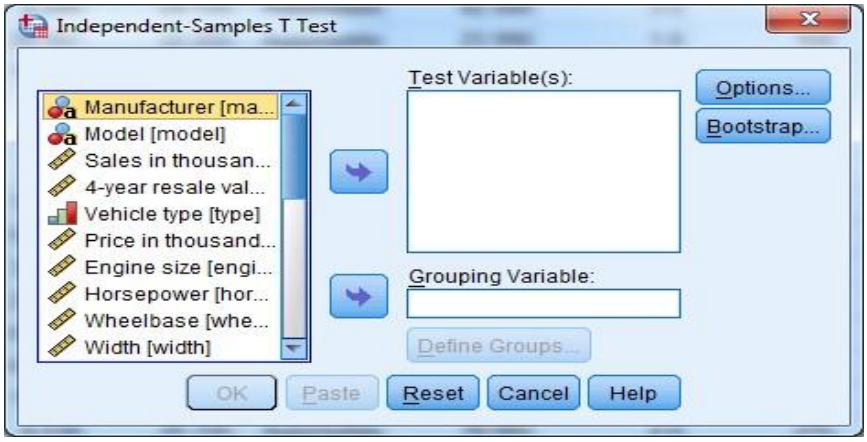
- Open the **Car sales.sav** dataset in **SPSS** using **File-> Open->Data options.**



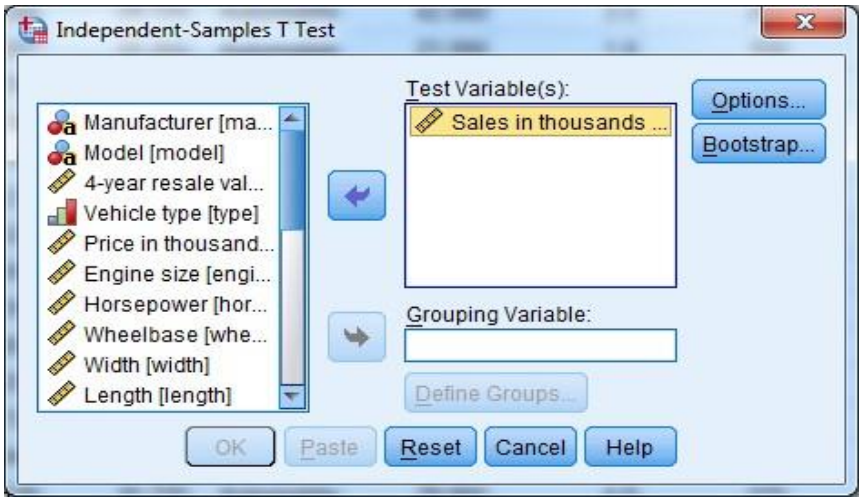
	manufact	model	sales	resale	type	price	engine_s	horsepow	wheelbas	width	length	curb_wgt	fuel_cap	m
1	Acura	Integra	16.919	16.360	Automobile	21.500	1.8	140	101.2	67.3	172.4	2.639	13.2	
2	Acura	TL	39.384	19.875	Automobile	28.400	3.2	225	108.1	70.3	192.9	3.517	17.2	
3	Acura	CL	14.114	18.225	Automobile	.	3.2	225	106.9	70.6	192.0	3.470	17.2	
4	Acura	RL	8.588	29.725	Automobile	42.000	3.5	210	114.6	71.4	196.6	3.850	18.0	
5	Audi	A4	20.397	22.255	Automobile	23.990	1.8	150	102.6	68.2	178.0	2.998	16.4	
6	Audi	A6	18.780	23.555	Automobile	33.950	2.8	200	108.7	76.1	192.0	3.561	18.5	
7	Audi	A8	1.380	39.000	Automobile	62.000	4.2	310	113.0	74.0	198.2	3.902	23.7	
8	BMW	323i	19.747	.	Automobile	26.990	2.5	170	107.3	68.4	176.0	3.179	16.6	
9	BMW	328i	9.231	28.675	Automobile	33.400	2.8	193	107.3	68.5	176.0	3.197	16.6	
10	BMW	528i	17.527	36.125	Automobile	38.900	2.8	193	111.4	70.9	188.0	3.472	18.5	
11	Buick	Century	91.561	12.475	Automobile	21.975	3.1	175	109.0	72.7	194.6	3.368	17.5	
12	Buick	Regal	39.350	13.740	Automobile	25.300	3.8	240	109.0	72.7	196.2	3.543	17.5	
13	Buick	Park Avenue	27.851	20.190	Automobile	31.965	3.8	205	113.8	74.7	206.8	3.778	18.5	
14	Buick	LeSabre	83.257	13.360	Automobile	27.885	3.8	205	112.2	73.5	200.0	3.591	17.5	
15	Cadillac	DeVille	63.729	22.525	Automobile	39.895	4.6	275	115.3	74.5	207.2	3.978	18.5	
16	Cadillac	Seville	15.943	27.100	Automobile	44.475	4.6	275	112.2	75.0	201.0	.	18.5	
17	Cadillac	Eldorado	6.536	25.725	Automobile	39.665	4.6	275	108.0	75.5	200.6	3.843	19.0	
18	Cadillac	Catera	11.185	18.225	Automobile	31.010	3.0	200	107.4	70.3	194.8	3.770	18.0	
19	Cadillac	Escalade	14.785	.	Truck	46.225	5.7	255	117.5	77.0	201.2	5.572	30.0	
20	Chevrolet	Cavalier	145.519	9.250	Automobile	13.260	2.2	115	104.1	67.9	180.9	2.676	14.3	
21	Chevrolet	Malibu	135.126	11.225	Automobile	16.535	3.1	170	107.0	69.4	190.4	3.051	15.0	
22	Chevrolet	Lumina	24.629	10.310	Automobile	18.890	3.1	175	107.5	72.5	200.9	3.330	16.6	

- Go to **Analyze->Compare Means-> Independent Samples T-Test**

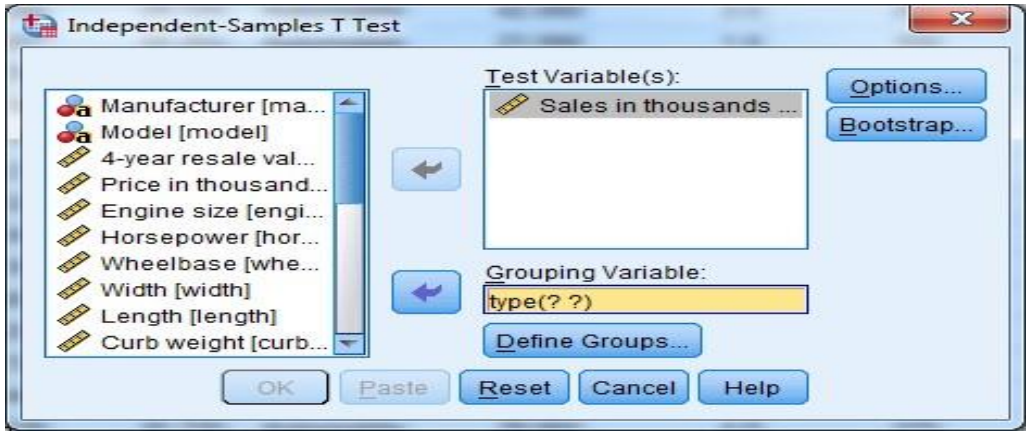


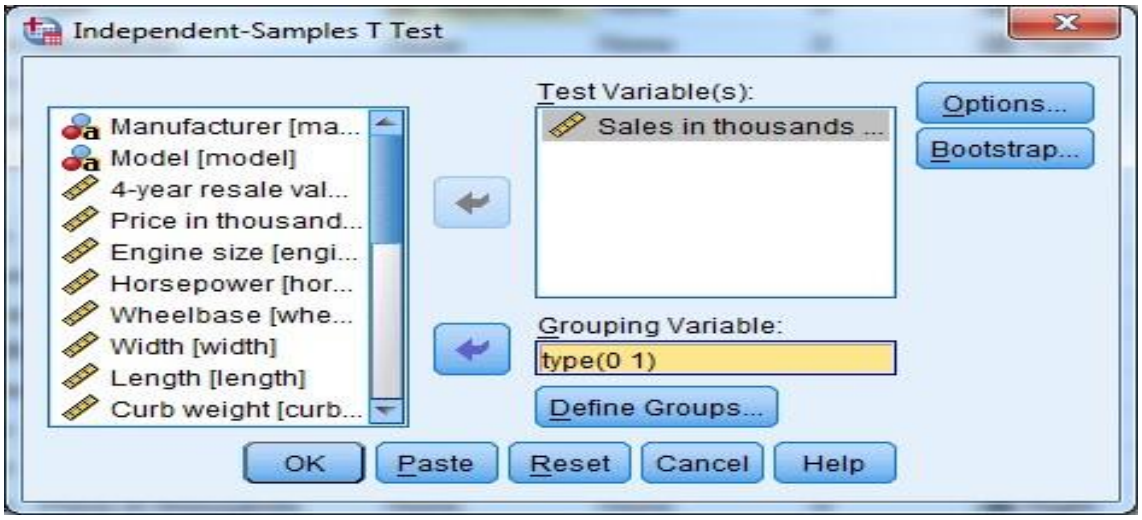


- In the **Independent-Samples T Test** dialog box shown above**,** put the variable **Sales in thousands** in the **Test Variable** box



- In the **Grouping Variable** box, put the variable **Vehicle Type** and then Click on **Define groups** tab and assign **0** in **Group 1** box and assign **1** in **Group 2** box and then click on **Continue** tab.





- Click **OK** in the **Independent-Samples T Test** dialog box.

OUTPUT:

Group Statistics

	Vehicle type	N	Mean	Std. Deviation	Std. Error Mean
Sales in thousands	Automobile	116	43.23434	50.557875	4.694181
	Truck	41	80.62229	98.192626	15.335112

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
Sales in thousands	Equal variances assumed	9.662	.002	-3.108	155	.002	-37.387948	12.030872	-61.153577	-13.622319
	Equal variances not assumed			-2.331	47.702	.024	-37.387948	16.037487	-69.638688	-5.137208

RESULT:

Independent Sample T-test in SPSS is learnt successfully.

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
COLLEGE OF SCIENCE AND HUMANITIES
Ramapuram Campus

EX.NO: 11 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

ONE WAY ANOVA TEST

AIM:

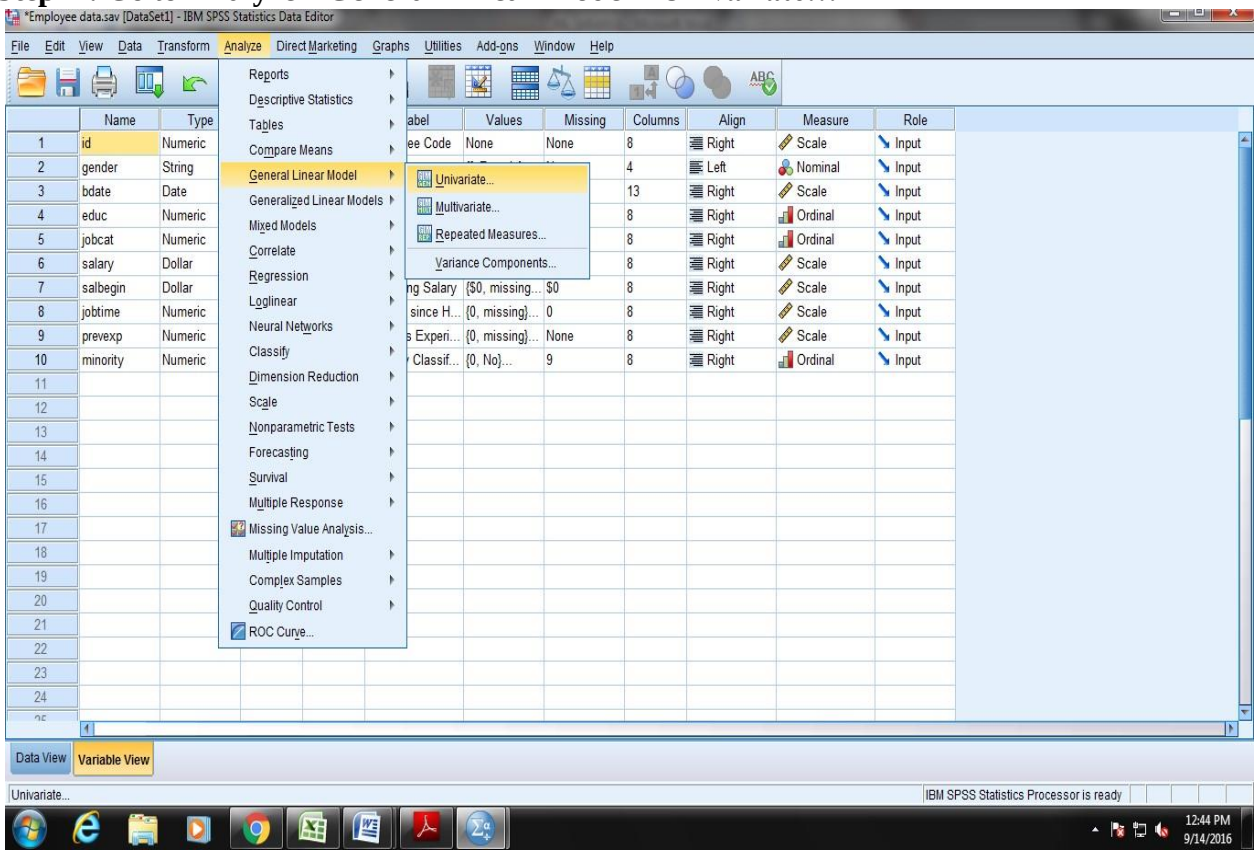
To implement ANOVA test in SPSS using the given dataset.

PROCEDURE:

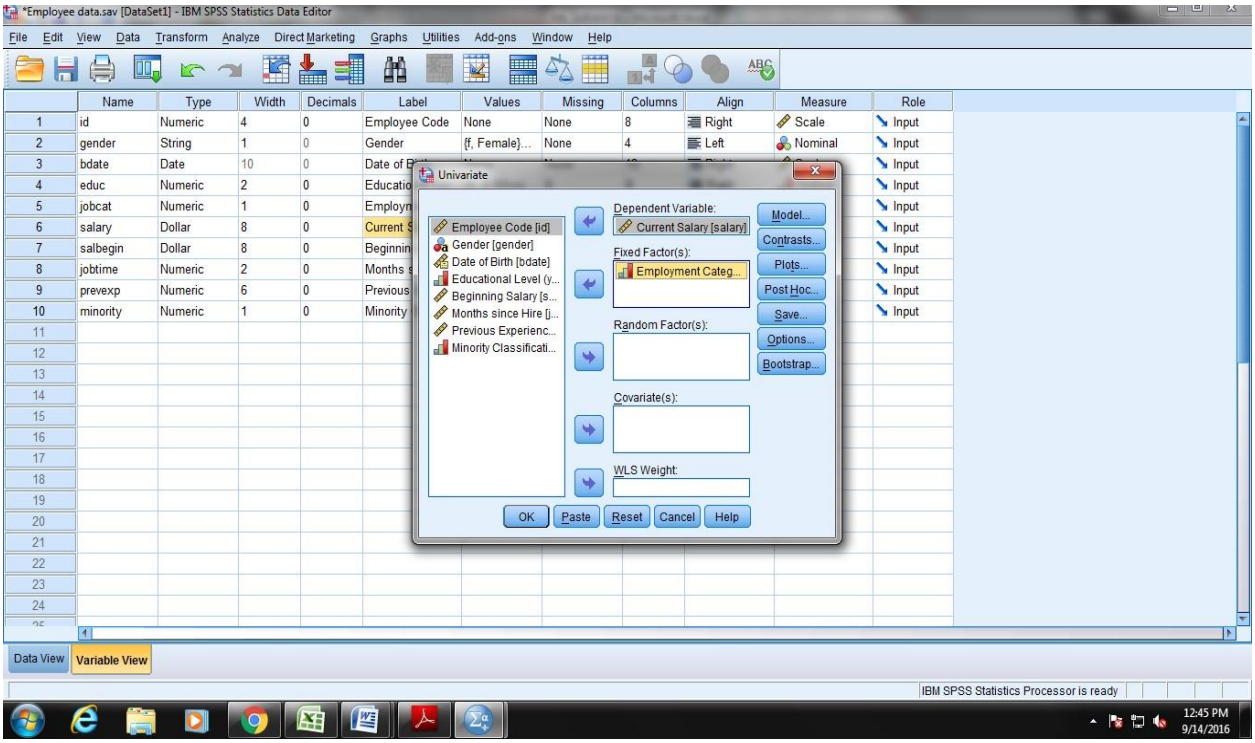
- Analysis of variance (**ANOVA**) is a collection of **statistical** models used to analyze the differences among group means and their associated procedures (such as "variation" among and between groups), developed by statistician and evolutionary biologist Ronald Fisher.
- The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

Create ANOVA TEST using the **Employee Data.sav** dataset.

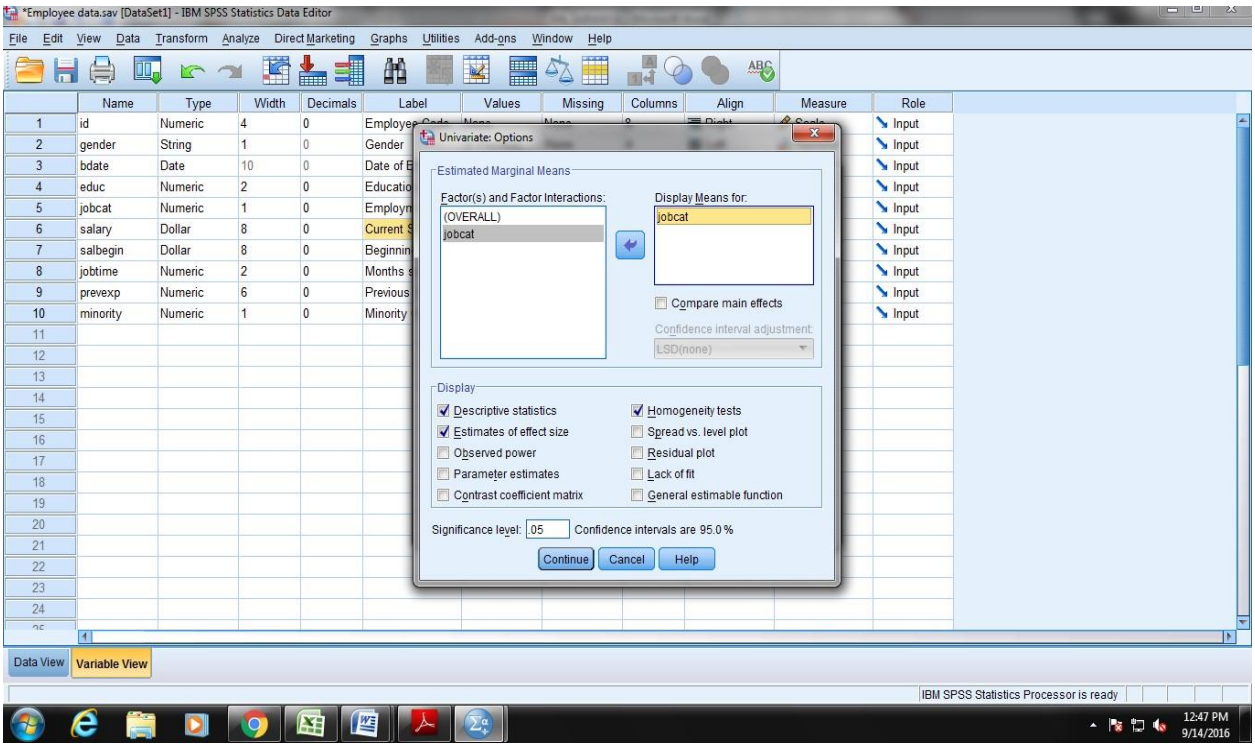
Step 1 : Go to **Analyze->General Linear Model->Univariate…**



Step 2: Put the variable **Salary** (Current salary)in **Dependent variable** box and the variable **jobcat** (Employment Category) in **Fixed Factor** box. Now click **Options** tab to customize the output.



Step 3: In the **Univariate Options** dialog box, move the variable **jobcat** from **Factor(s) and Factor Interactions** box into **Display Means for**: box and click on the check boxes **Descriptive Statistics, Homogeneity tests and Estimate of effect size** and then click on **Continue** and click on **OK** in the **Univariate** dialog box.



OUTPUT:

Between-Subjects Factors

		Value Label	N
Employment Category	1	Clerical	363
	2	Custodial	27
	3	Manager	84

Descriptive Statistics

Dependent Variable: Current Salary

Employment Category	Mean	Std. Deviation	N
Clerical	$27,838.54	$7,567.995	363
Custodial	$30,938.89	$2,114.616	27
Manager	$63,977.80	$18,244.776	84
Total	$34,419.57	$17,075.661	474

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Current Salary

F	df1	df2	Sig.
59.733	2	471	.000

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

a. Design: Intercept + jobcat

Estimated Marginal Means

Employment Category

Dependent Variable: Current Salary

Employment Category	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Clerical	27838.540	532.487	26792.197	28884.883
Custodial	30938.889	1952.451	27102.297	34775.481
Manager	63977.798	1106.935	61802.655	66152.941

RESULT: Implemented ANOVA test in SPSS using the given dataset successfully.

EX.NO: 12 **NAME: S.Yugendran**

DATE: **REG NO: RA2031241020045**

TWO WAY ANOVA TEST

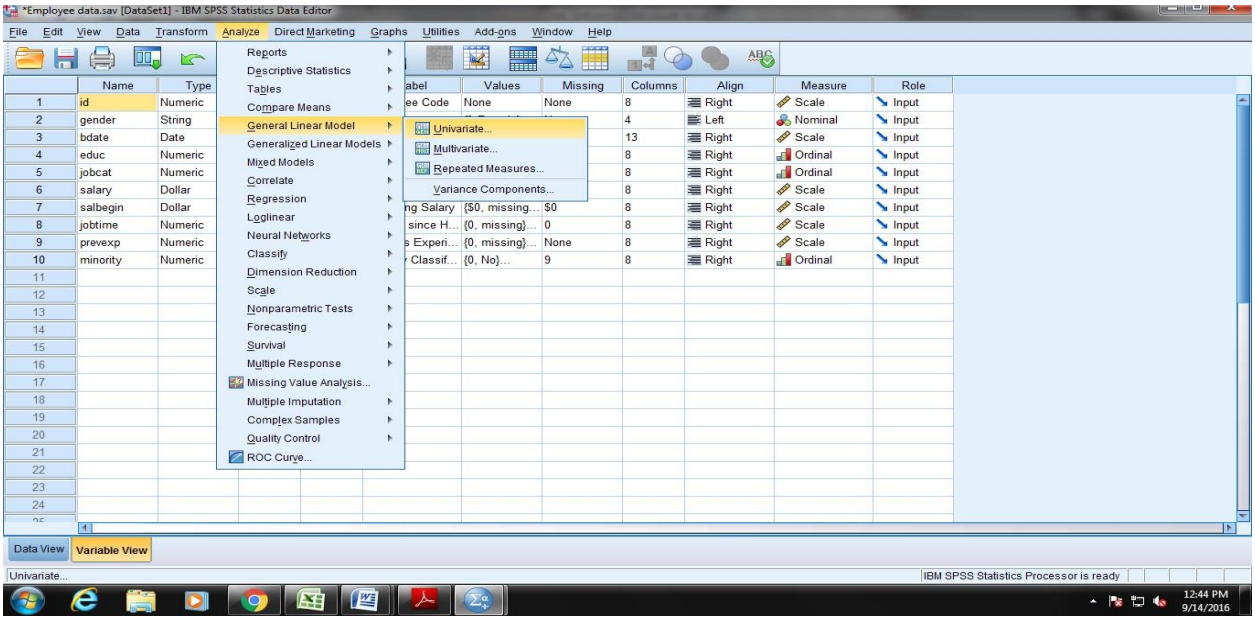
AIM:

To perform Two Way Anova test using SPSS.

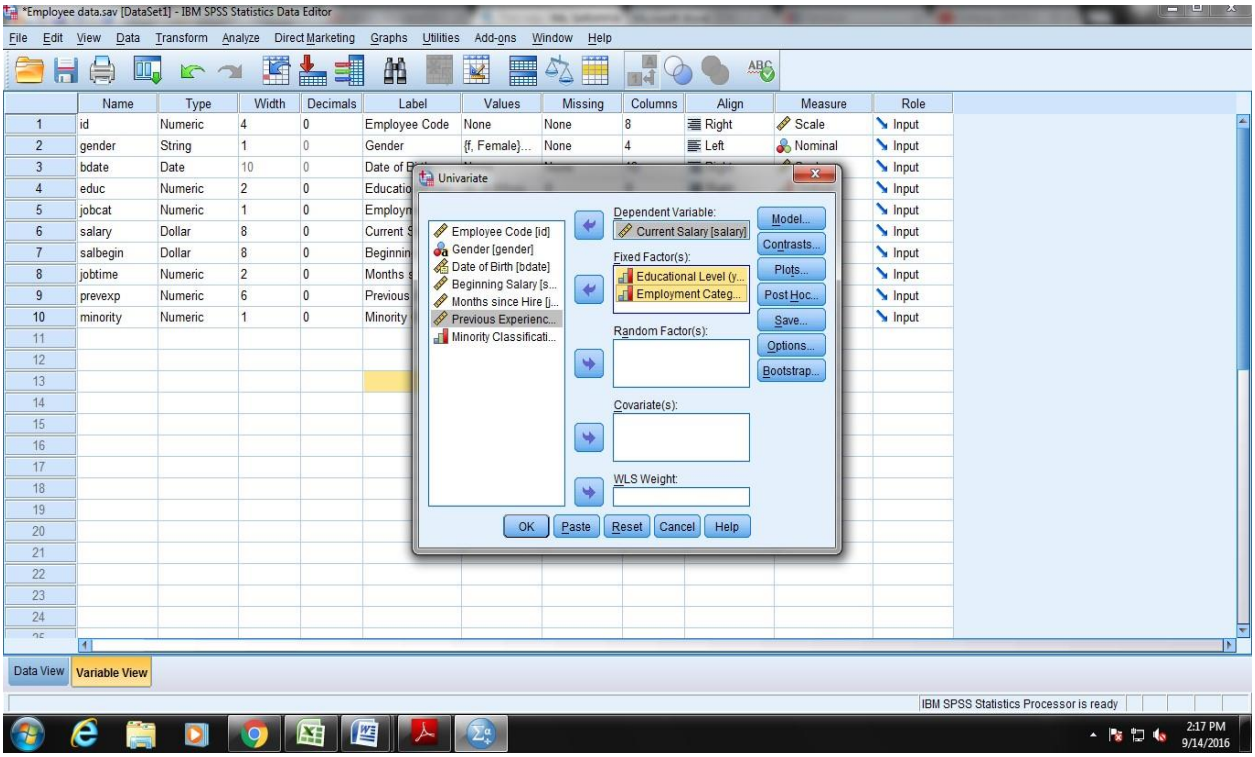
INTRODUCTION:

- The two-way ANOVA compares the mean differences between groups that have been split on two independent variables (called factors).
- The primary purpose of a two-way ANOVA is to understand if there is an interaction between the two independent variables on the dependent variable.
- For example, you could use a two-way ANOVA to understand whether there is an interaction between gender and educational level on test anxiety amongst university students, where gender (males/females) and education level (undergraduate/postgraduate) are your independent variables, and test anxiety is your dependent variable.

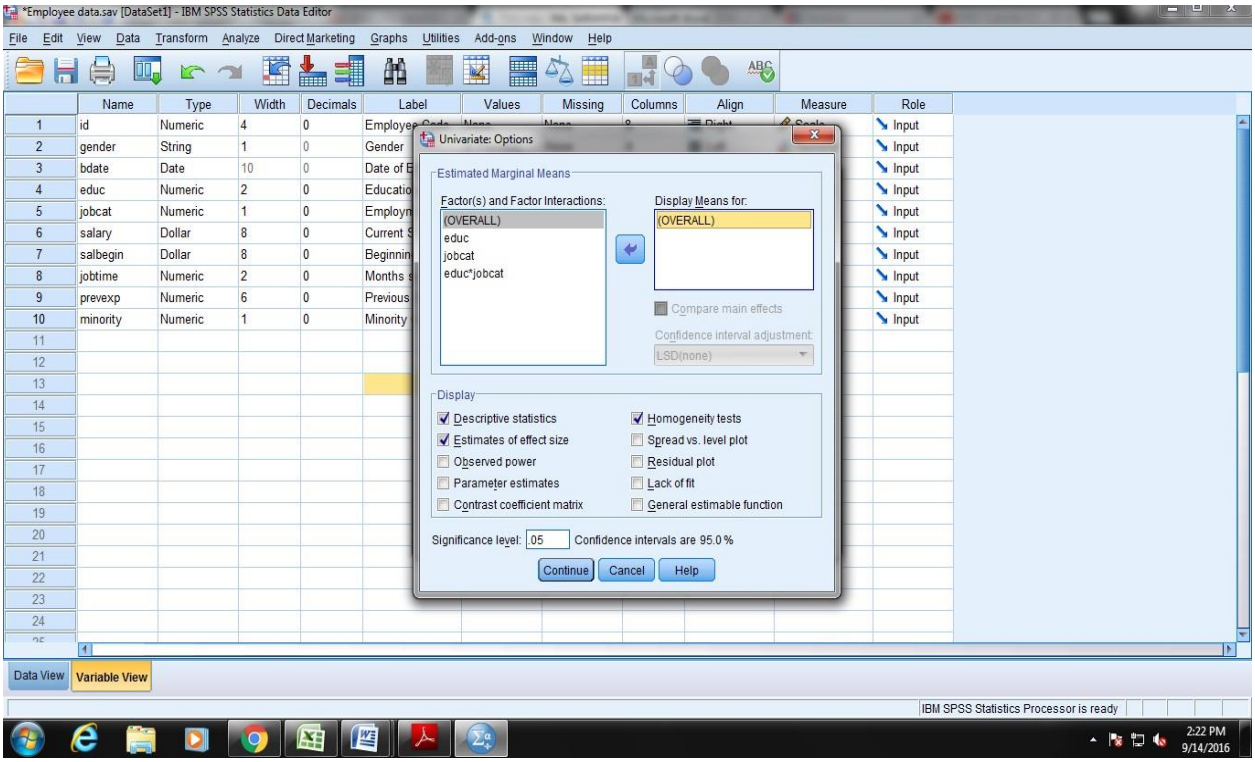
Step 1: Open the **Employee Data.sav** data set. **Goto Analyze->General Linear Model ->Univariate…..**



Step 2: In the **Univariate** dialog box, select one dependent variable for eg: **salary** (Current Salary)in **Dependent Variable** box and two independent variables say **educ and jobcat**(Educational Level and Employment category) in **Fixed Factor** box. Then click **Options** tab to customize the output.



Step 3: In the **Univariate: Options** dialog box, move the (**Overall)** option from **Factor(s) and Factor Interactions** box into **Display Means for**: box and click on the check boxes **Descriptive Statistics, Homogeneity tests and Estimate of effect size** and then click on **Continue.**



Step 4: Click **OK** in the **Univariate** dialog box and the output will be displayed.

OUTPUT:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Current Salary

F	df1	df2	Sig.
7.654	18	455	.000

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

a. Design: Intercept + jobcat + educ + jobcat * educ

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	102578890584.605[a]	18	5698827254.700	73.377	.000	.744
Intercept	78050133390.543	1	78050133390.543	1004.958	.000	.688
jobcat	7643728211.287	2	3821864105.643	49.210	.000	.178
educ	3423321707.953	9	380369078.661	4.898	.000	.088
jobcat * educ	1448112123.442	7	206873160.492	2.664	.010	.039
Error	35337604851.734	455	77665065.608			
Total	699467436925.000	474				
Corrected Total	137916495436.340	473				

a. R Squared = .744 (Adjusted R Squared = .734)

RESULT: **P**erformed Two Way ANOVA test using SPSS successfully.