C-International Research Consultancy

Promoting integrity in research and mentoring researchers across the globe

Research Methodology Workshop

Module 3

Creating a Database, Univariate Analysis & Test of Normality in SPSS

https://doi.org/10.33515/cintarchrescon/rmw/m3

Contact details

- To join our team of "Locum Research Consultancy Specialists and Workshop Facilitators", please click on this link: http://www.cintarch.org/research-consultancy/
- To access our "Research Consultancy Services", please click on this link: http://www.cintarch.org/research-consultancy-request/
- To submit your manuscript for publication in any of our journals, please click on this link: http://www.cintarch.org/submit-manuscripts/

Outline

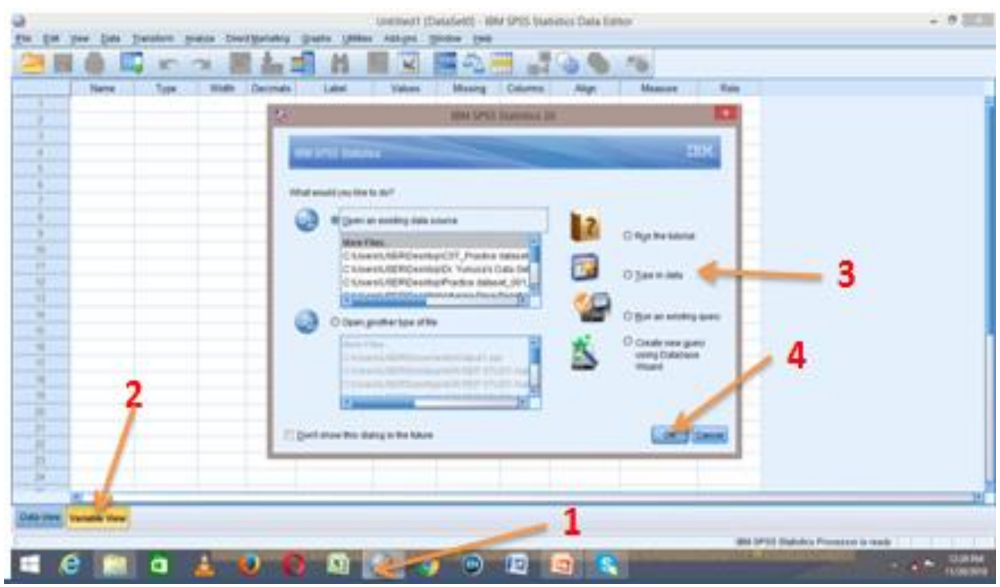
- Creating a Database in SPSS
- Univariate Analysis in SPSS
- Test of Normality

CREATING A DATABASE IN SPSS

Starting SPSS

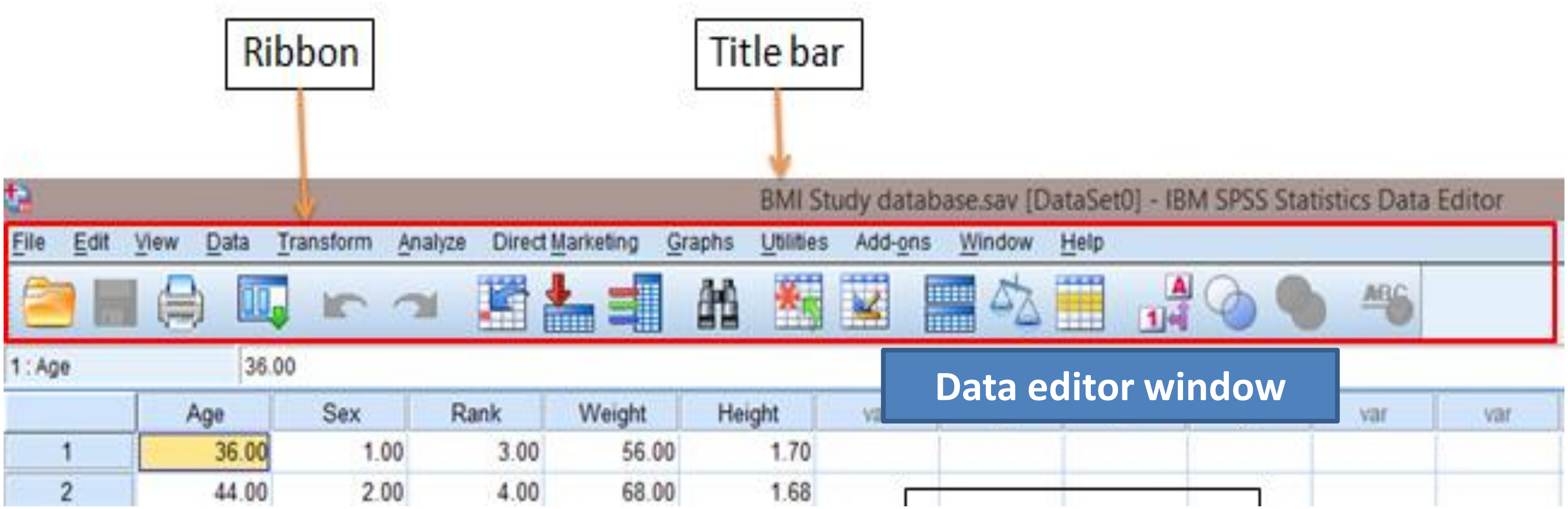
- Double click on the **IBM SPSS** icon [1] (if you have it on your desktop).
- Alternatively, click the start button and then select:

  All programs > IBM SPSS Statistics > IBM SPSS Statistics 20
- This opens the **data editor** window. This is the window where you input data and carry out statistical function (and the default view that opens in this window is the **Variable view** [2]). The other SPSS window is the **viewer**, this is where the results of any analysis appears.
- If you are just creating the database, select **Type in data** [3] and click **OK** [4] in the **IBM SPSS Statistics** box that appears at the center of the window.



Features of IBM SPSS data editor window

- **Title bar**: The title bar appears at the top of the program window and displays the name of the window (i.e., **Data editor**) and program (i.e., **IBM SPSS Statistics**).
- **Ribbon**: The Ribbon is designed to help you quickly find the commands that you need to complete a task. It consists of a set of task-specific **tabs**.
- **Data editor window**: This is where you input data and carry out statistical functions.



Features of IBM SPSS data editor window contd.

The Ribbon

- The **File tab** allows you to save data, graphs or output. It also allows you to open previously saved files and print graphs data or output.
- The **Edit tab** allows you to edit data such as cut and paste blocks of numbers from one part of the data editor to another, insert and clear (i.e., delete) variables.
- The **Transform data tab** allows you to manipulate your variables, such as recoding a continuous quantitative variable (e.g., age) into a qualitative variable (e.g., age groups), and compute a variable from a combination of variables (e.g., BMI from weight and height).
- The **Analyze data tab** allows you to analyze your data and also create charts if you want.
- The **Graphs tab** allows you to create different charts.

Entering variables and giving coding instructions

The first thing to do is to enter the variables and give coding instructions.

- While still in the variable view, type the name of the variable (e.g., age, sex, etc) under the column for name, and press Tab to move to the next column.

- For **age**, **weight** and **height**, since they are **quantitative variables** , in the column for measure select **scale**.

- For **sex** and the other variables, since they are **qualitative variables** (measured on the nominal scale), in the column for measure, select **Nominal**.

In the variable view, each variable occupies a row

Entering variables and giving coding instructions contd.

- In contrast to quantitative variables (either discrete [e.g., age last birthday, number of students attending a lecture, etc] or continuous [e.g., age, weight, height etc], values (also called coding instructions) must be assigned to qualitative variables as all variables in SPSS are entered in Arabic numerals.

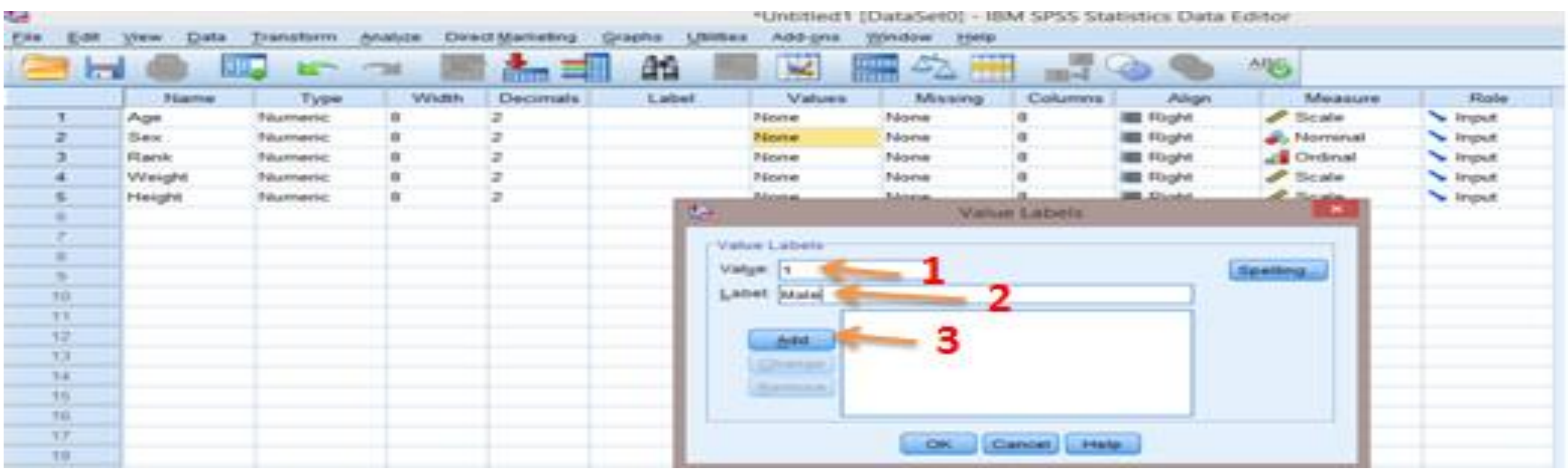
- The value assigned to the available options under the respective variables should be specified in the **coding instructions** (alternatively you may indicate it in the questionnaire used in developing the database).

In the variable view, each variable occupies a row

Entering variables and giving coding instructions contd.

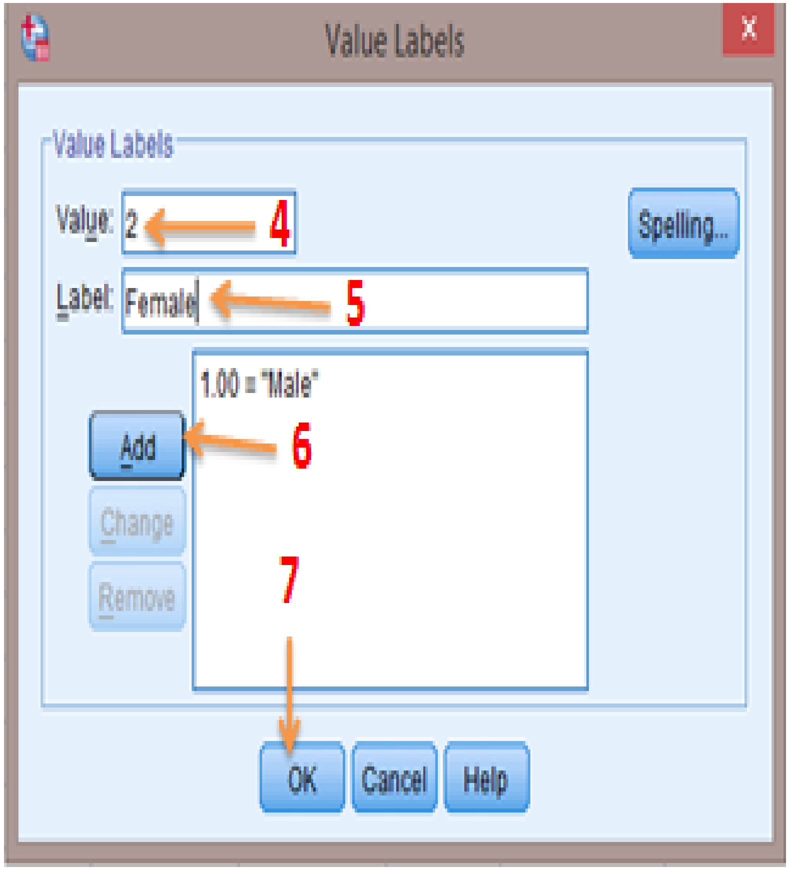
- In the row for the variable names **sex**, click in the cell under the column for value. And then click on the box with dots at the right border of cell.

- In the dialog box that appears type **1** in the box for **Value** [1] , type **Male** [2] in the box for **Label**, and then click **Add** [3].



Entering variables and giving coding instructions contd.

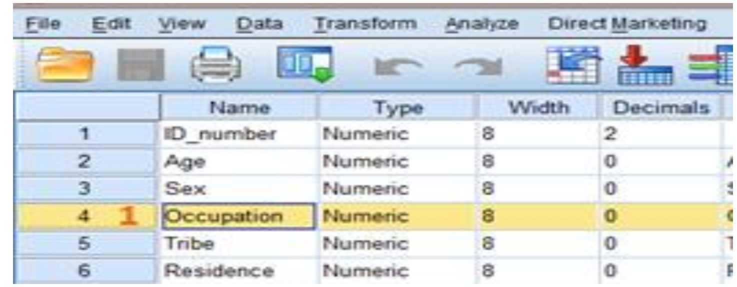
- Type **2** in the box for **Value** [4] , type **Female** [5] in the box for **Label**, and then click **Add** [6].
- Since all the available options have been entered, now click **OK** [7].
- In the cell under the column for measure, select **Nominal**.
- Repeat the procedure for the other qualitative variables.



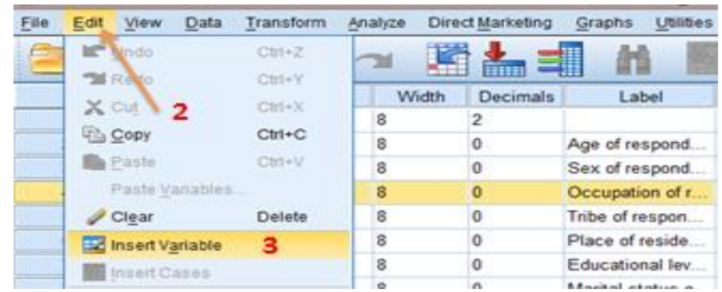
Inserting variables

To insert a variable

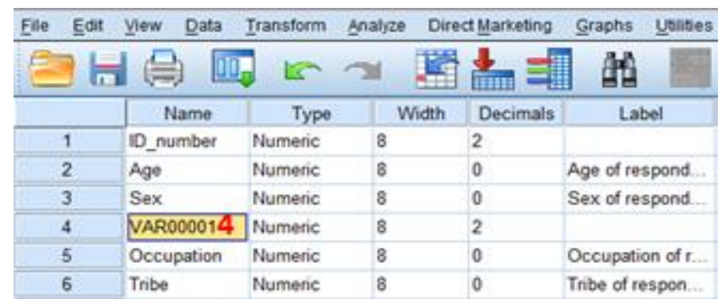
- Click on the heading of the row above which you want to insert a new variable to select the whole row[1].

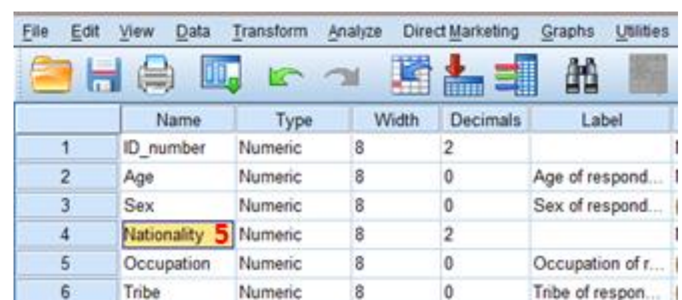


- Click on **Edit tab** [2], and then click **Insert Variable** [3]



- Replace **VAR0001** [4] that appears under the Name column in the new row created with the name of the new variable to be inserted (e.g., **Nationality**) [5].

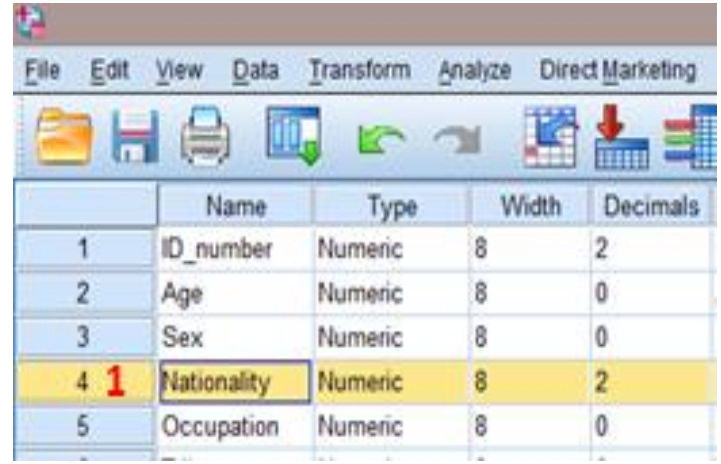




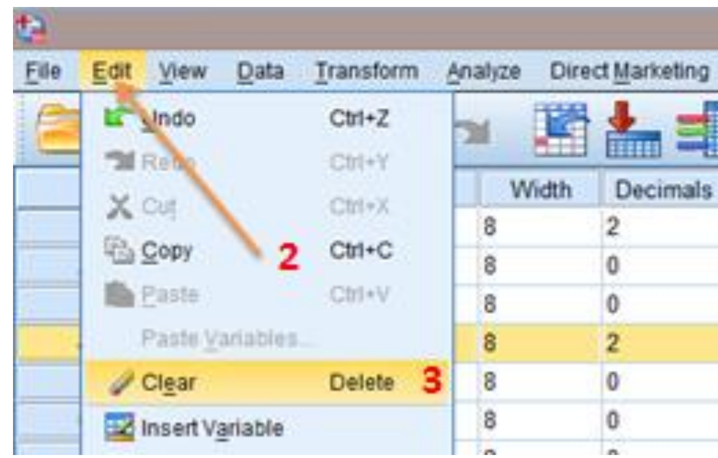
Deleting variables

To delete a variable

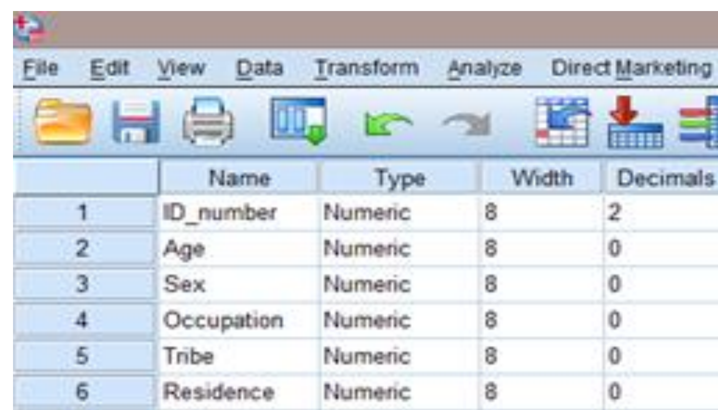
- Click on the heading of the row containing the variable you want to delete (e.g., **Nationality**) [1].



- Click on **Edit tab** [2], and then click **Clear** (i.e., **Delete**) [3]



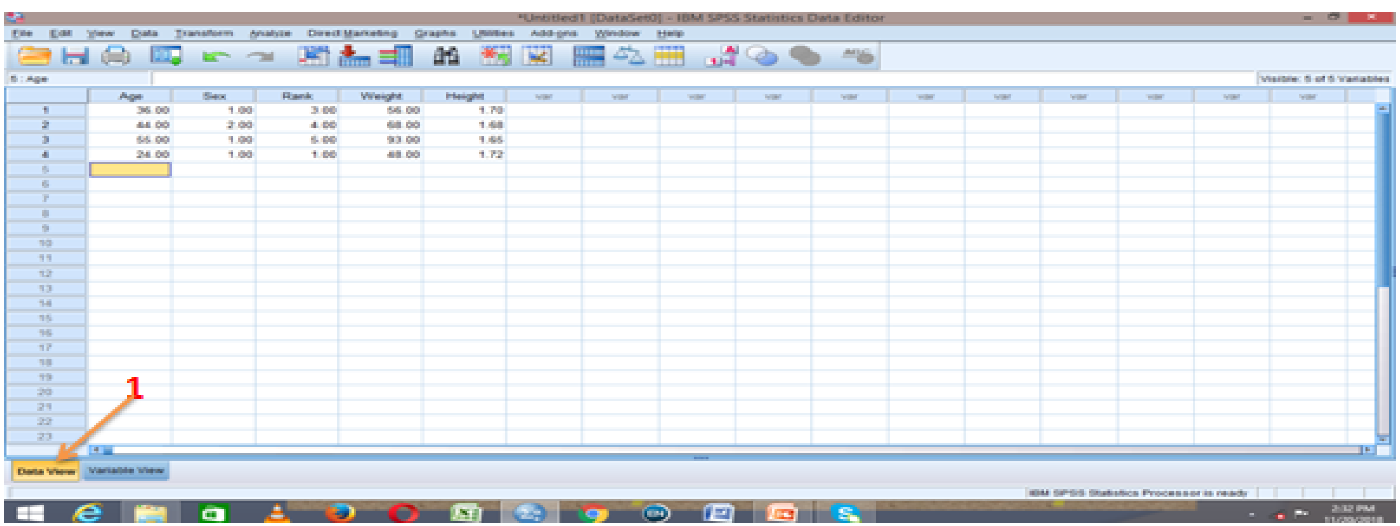
- The selected row (containing the variable to be deleted) becomes deleted.



Data entering

After entering all the **variables** (in the **variable view** of the **data editor** window), the next thing to do is to enter the **data** (in the **data view** of the **data editor** window)

- Click on the **Data view** button to enter your data [1].
- In the Data view, each variable occupies a column, with each row representing a subject.
- It is preferable to assign identification numbers to your questionnaires to enable you trace those concerned in the event of errors in data entering, and to enable several research assistants to enter the data for a study using the same database template.

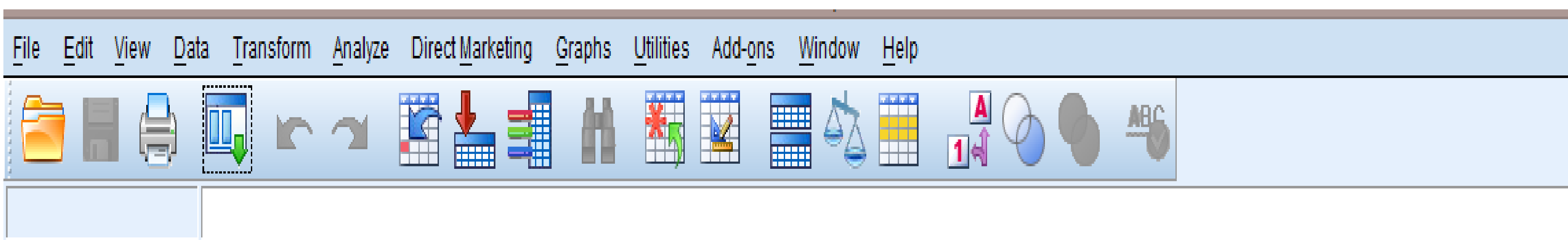


Data entering contd.

Data entering is done as indicated the coding instructions.

- **Subject 1: Age** = **18 years** (entered directly); **Sex** is **Male** (entered as **1**, as indicated in the coding instruction); ……… **Weight** = **40kg** (entered directly); **Height** = **158cm** (entered directly).

- **Subject 2: Age** = **18 years** (entered directly); **Sex** is **Female** (entered as **2**, as indicated in the coding instruction); ……… **Weight** = **62kg** (entered directly); **Height** = **171cm** (entered directly).



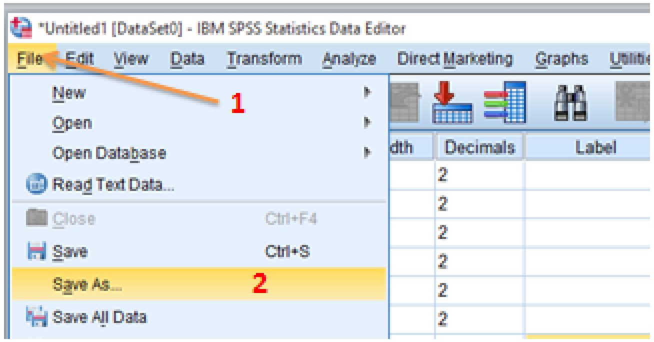
	ID_number	Age	Sex	Occupation	Tribe	Residence	Education	Marital_sta...	Religion	SBP	DBP	Weight	Height
1	1.00	18	1	2	1	1	4	1	1	116.00	80.00	40.00	158.00
2	2.00	18	2	2	1	1	5	1	1	110.00	70.00	62.00	171.00

- **Finally, give your file a name and save it.**

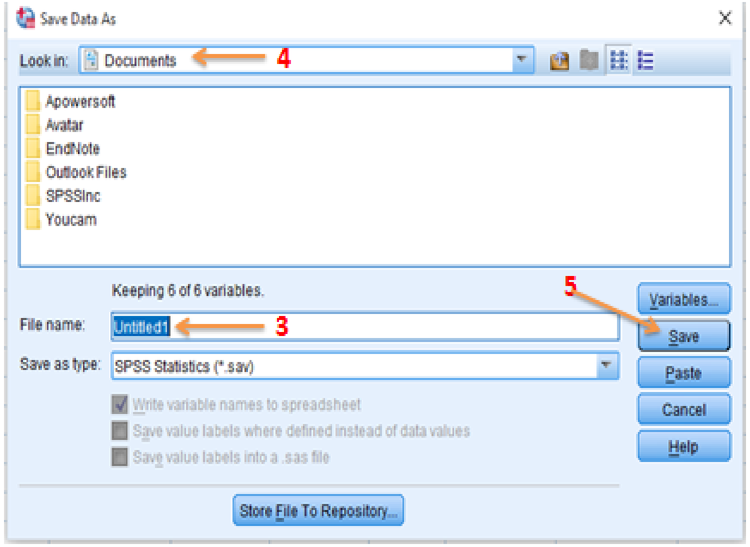
Naming and saving your file

To name and save your file

- Click on the **File tab** [1] and then choose **Save As** [2].

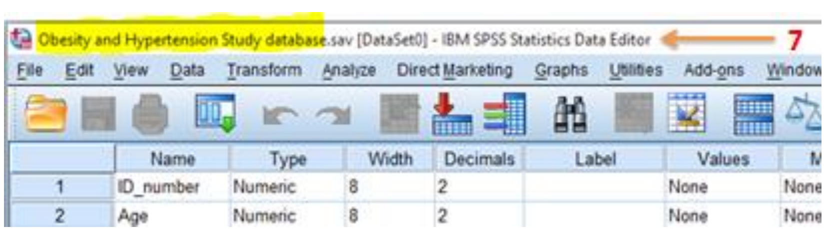


- Type the name (e.g., **Obesity and Hypertension Study database**) you want to give your file in the **File name** box [3], there is no need to clear what is inside the box before typing the name, whatever you type will replace what is already there.

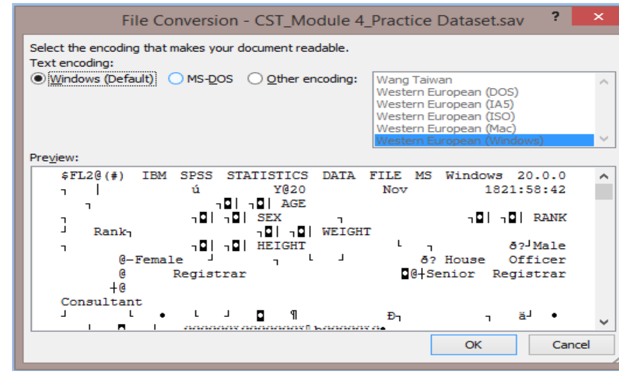


- Select the location where you want to save your file (e.g., **Documents**) [4].
- Finally click **Save** [5].

- Instead of **Untitled1** [6], your file now has a name (**Obesity and Hypertension Study database**) [7].

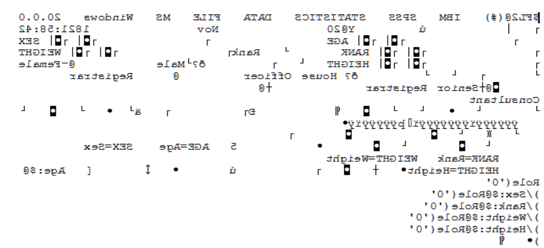


Opening an existing dataset



To open an existing dataset:

- Go to the location (desktop or document) where you saved the file, open the folder if it was saved in a folder, and click on the file.
- In some computers rather than open the dataset a **File Conversion** dialog box will appear.
- Even if you click OK in the box, a Microsoft Word document with some funny signs will appear.

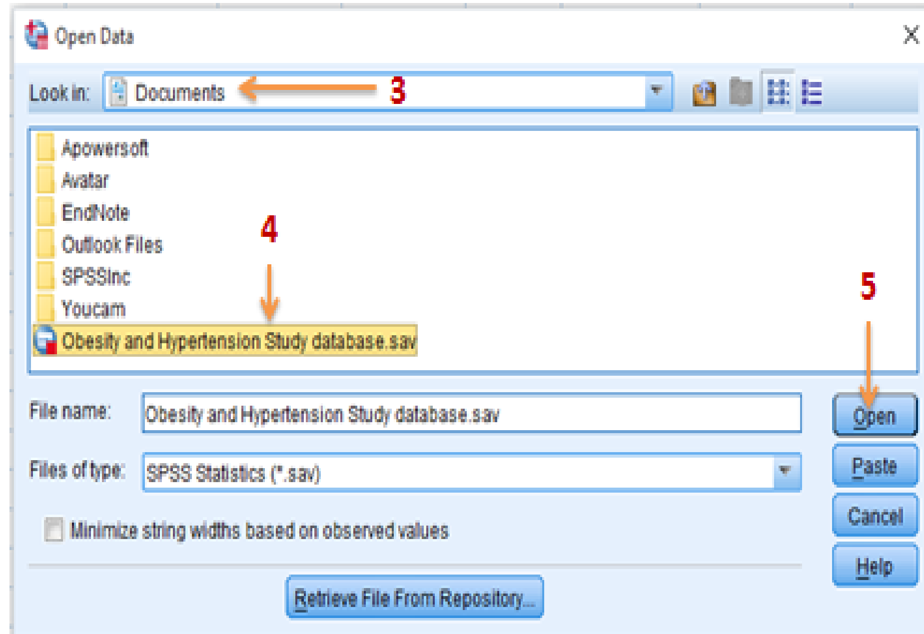


- If this occurs, open the SPSS program as earlier described.
- By default "**Open an existing data source**" is selected [1], so just click **OK** [2].

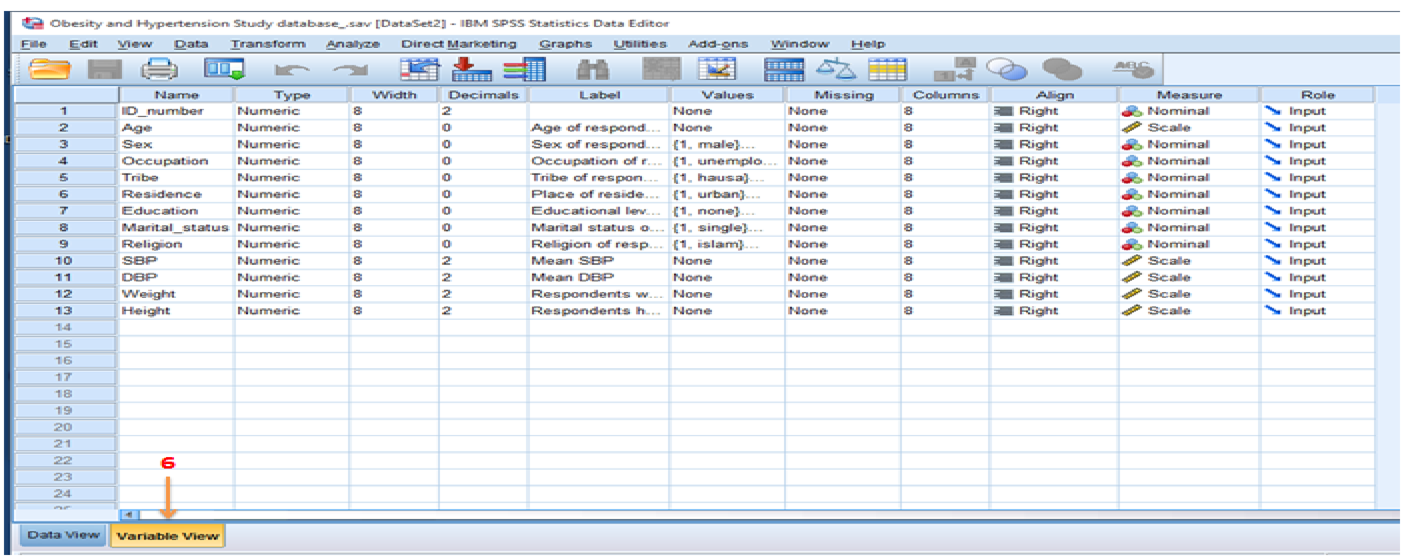
Opening an existing dataset contd.

To open an existing dataset contd.:

- In the new dialog box that appears, use the drop down arrow next to **Look in** box to select the location of the file (e.g., **Documents**) [3], search for the file (e.g **Obesity and Hypertension Study database**) [4], click on it, and then click **Open** [5].



- The **Data Editor** window opens in **Variable view** [6].



Importing data into SPSS from MS Excel

- Data collected with the **Open Data Kit** (**ODK**) software using android phones (which removes the cost of printing questionnaires, creating database in SPSS and the rigorous data entry phase) is downloaded in MS Excel format and then imported into SPSS for analysis.

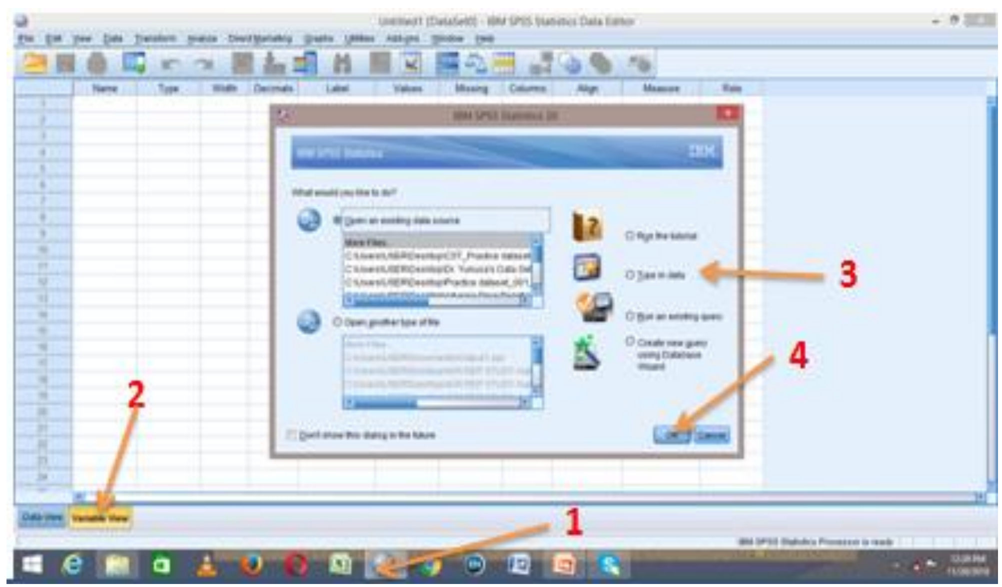
To import data into SPSS from MS Excel:

- Double click on the **IBM SPSS** icon [1] (if you have it on your desktop).
- Alternatively, click the start button and then select:

All programs > IBM SPSS Statistics > IBM SPSS Statistics 20

- This opens the **data editor** window. This is the window where you input data and carry out statistical function (and the default view that opens in this window is the **Variable view** [2]).

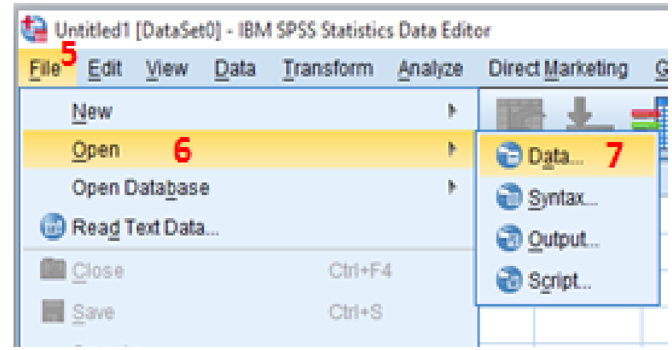
- Since a new database must be created for the database to be imported into SPSS from MS Excel, select **Type in data** [3] and click **OK** [4] in the **IBM SPSS Statistics** box that appears at the center of the window.

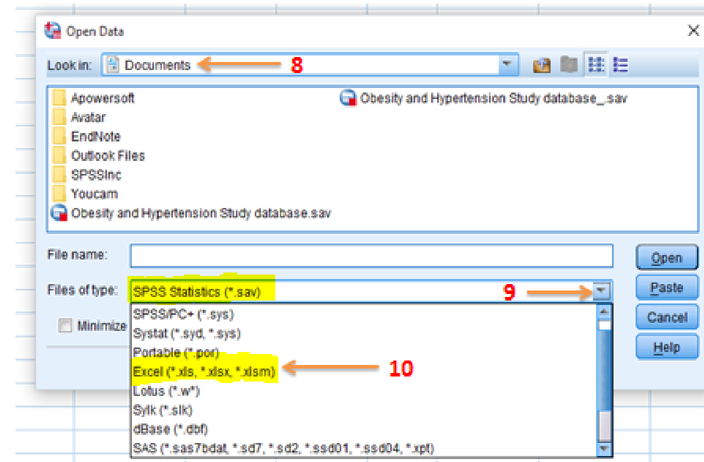


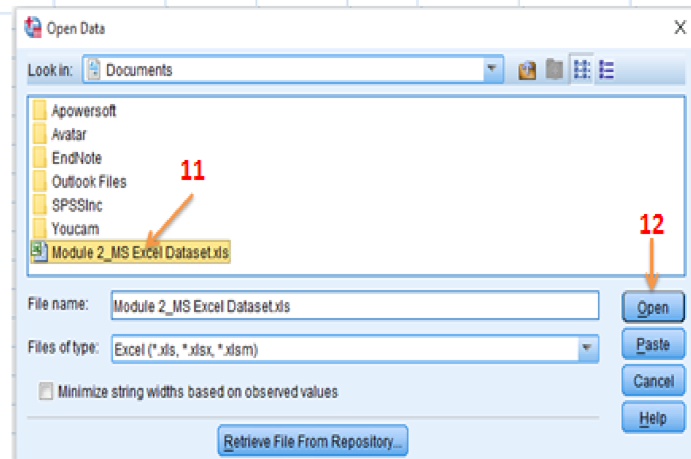
Importing data into SPSS from MS Excel contd.

To import data into SPSS from MS Excel contd.:

- Click on **File** [5], select **Open** [6], and then **Data** [7].
- In the **Open Data** box that appears, locate the drive (e.g., **Documents**)[8]where the MS Excel file was saved.
- To make the file visible, click on the drop-down arrow in the **Files of type** box [9] and select **Excel (“xls,”xlsx,”xlsm)** [10] to replace the **SPSS Statistics (*sav)** displayed.
- The MS Excel file [11] is now visible, and the name (**Module 2_MS Excel Dataset**) appears in the **File name** box once you click on it.
- Click **Open** [12].

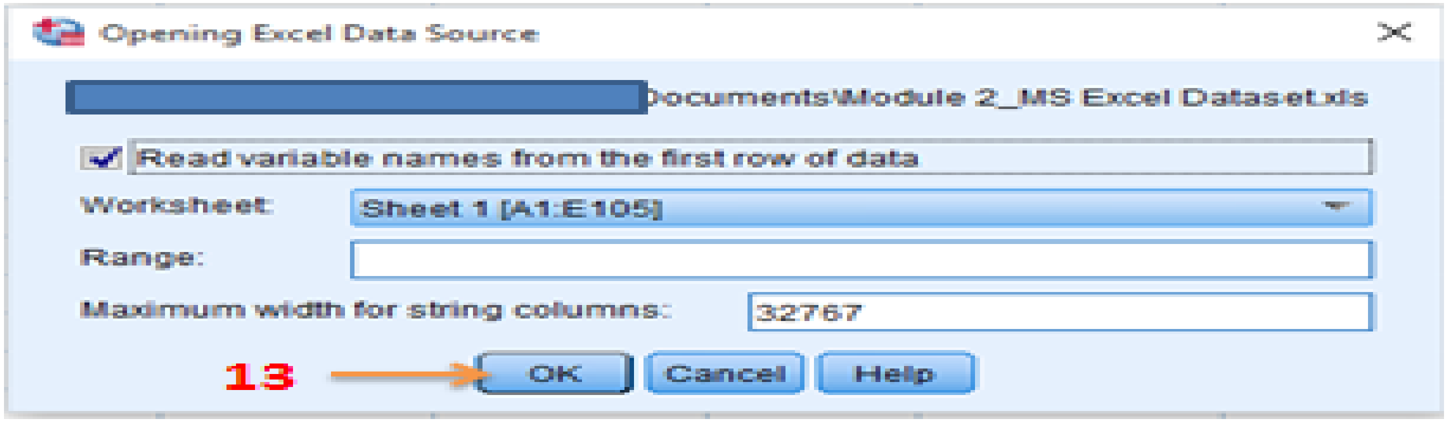




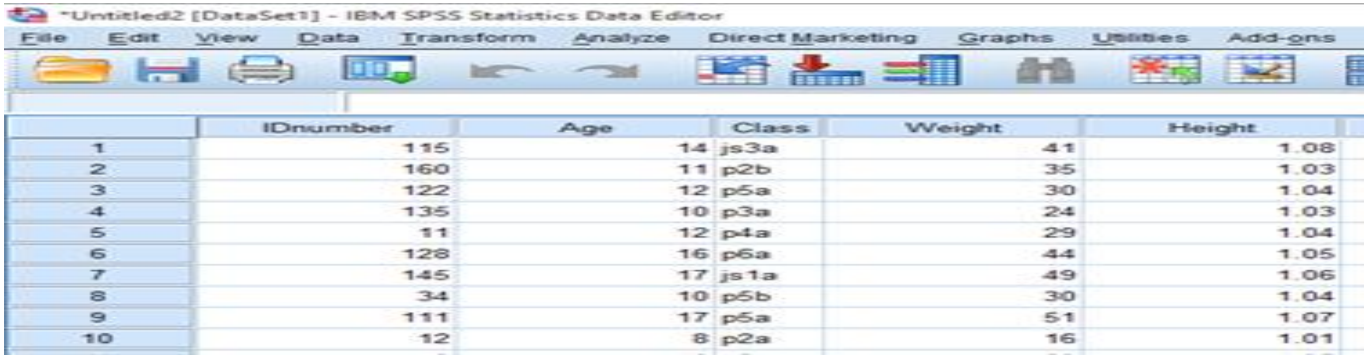


Importing data into SPSS from MS Excel contd.

To import data into SPSS from MS Excel contd.:



- In the **Opening Excel Data Source** box that appears click **OK** [13].
- An untitled SPSS version of the file appears, name and save it as previously illustrated.



The dataset for this module (i.e., Module 3ii) is in MS Excel, please import it into SPSS

UNIVARIATE ANALYSIS IN SPSS

Analysis in SPSS

There are 3 categories of analysis (**univariate**, **bivariate** and **multivariate**). Univariate analysis is covered in this module, while bivariate and multivariate analyses are covered on individual basis in our **Research Consultancy Services** at: **http://www.cintarch.org/research-consultancy-request/**

Univariate analysis

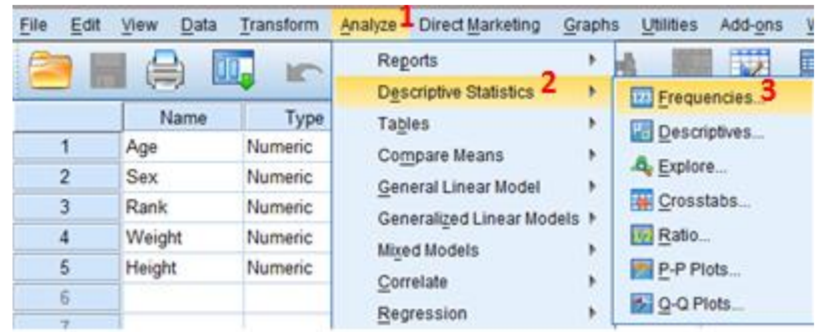
This involves analysis of **one variable** at a time. This form of analysis is referred to as **descriptive statistics** (as it does not compare variables or establish associations).

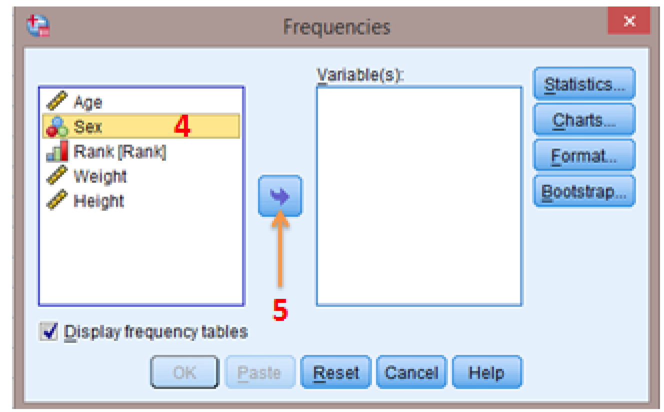
- For **qualitative variables** (e.g., **sex** and **rank** on **nominal** and **ordinal** measurements respectively) **frequencies** procedures are performed.

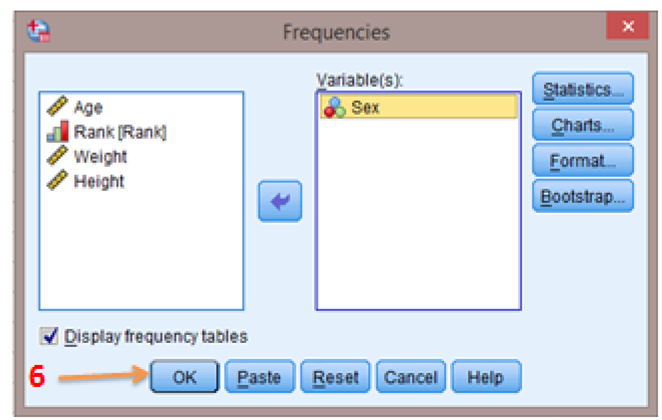
- For **quantitative variables** (e.g., **age**, **weight** and **height** on **scale** measurement) **measures of central tendency** (mean, mode and median) and **measures of dispersion** (range, inter-quartile range, variance and standard deviation) called **descriptives** procedures are performed.

A_To run frequencies on qualitative variables

- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Frequencies** [3].
- Double click on the qualitative variable you want to analyze (e.g., **Sex** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.
- Finally, click **OK** [6].
- Repeat the procedure for the other qualitative variables.
- You can actually move all the qualitative variables in your dataset to the **Variable(s)** box (step 4 or 5) and analyze them together, but the results for each variable will appear in a separate table.

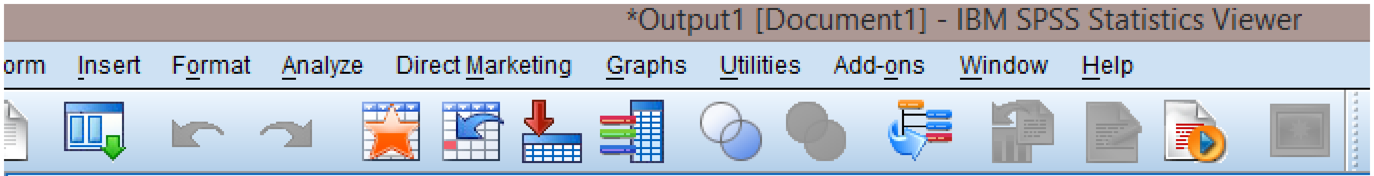






A_To run frequencies on qualitative variables contd.

- The results of the analysis are displayed in a table on the **viewer** window.



Frequencies

Statistics

		Sex of respondents	Occupation of respondents	Tribe of respondents	Place of residence
N	Valid	300	300	300	300
	Missing	0	0	0	0

Frequency Table

Sex of respondents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	161	53.7	53.7	53.7
	female	139	46.3	46.3	100.0
	Total	300	100.0	100.0	

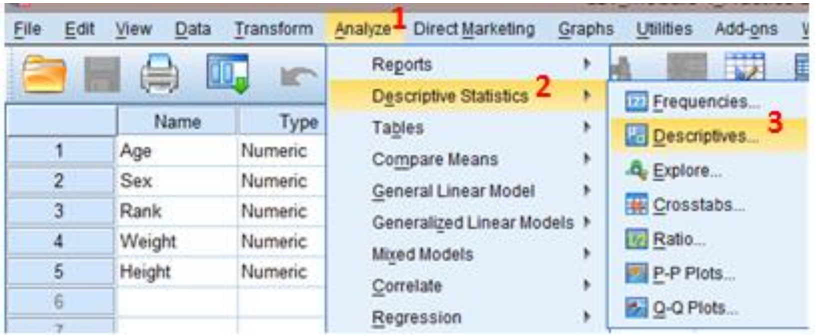
The output shows the following:
All the 300 subjects were entered for all the variables (i.e., no missing values). This is why the Percent (based on the sample size), is the same as the Valid Percent (based on the values entered).

Frequency for male = 161; Valid Percent = 53.7
Frequency for female = 139; Valid Percent = 46.3

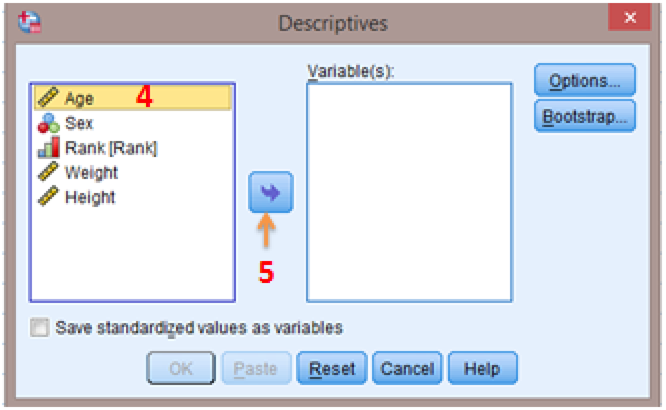
If there are missing values it is the Valid Percent that should be reported).

B1a_To run descriptives on quantitative variables

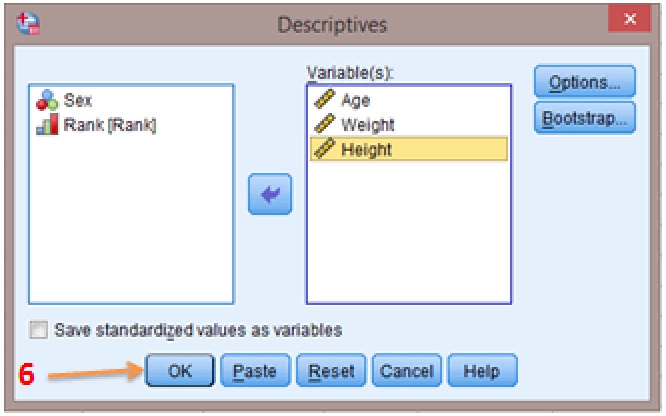
- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Descriptives** [3].



- Double click on the quantitative variable you want to analyze (e.g., **Age** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.



- Repeat the procedure (step 4 or 5) to move the other quantitative variables (i.e., **Weight** and **Height**) to the **Variable(s)** box and click **OK** [6].



B1a_To run descriptives on quantitative variables contd.

- The results of the analysis are displayed in a table on the **viewer** window.

Descriptives

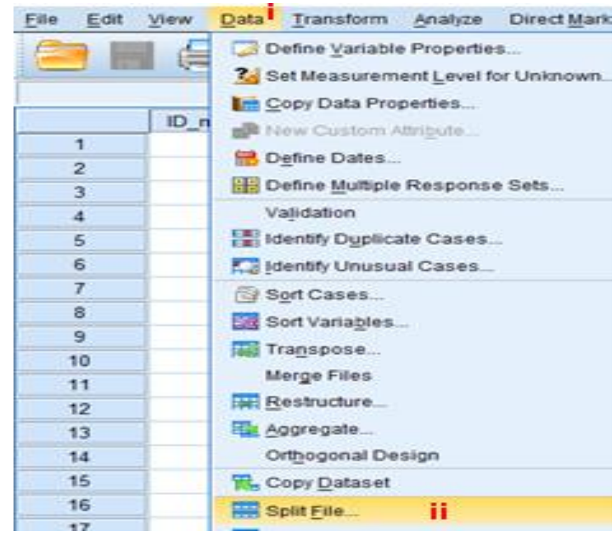
Descriptive Statistics

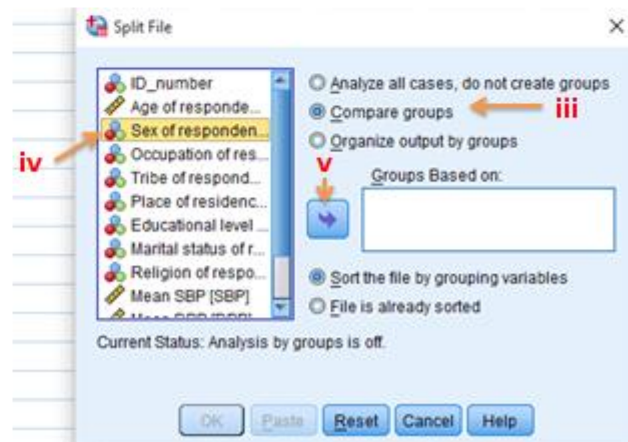
	N	Minimum	Maximum	Mean	Std. Deviation
Age of respondents	300	18	91	40.46	18.060
Mean SBP	300	85.00	229.00	131.5413	23.72248
Mean DBP	300	48.50	130.00	82.9217	13.75592
Respondents weight	300	30.00	129.00	62.4900	13.72411
Respondents height	300	61.00	199.00	164.8533	10.93972
Valid N (listwise)	300				

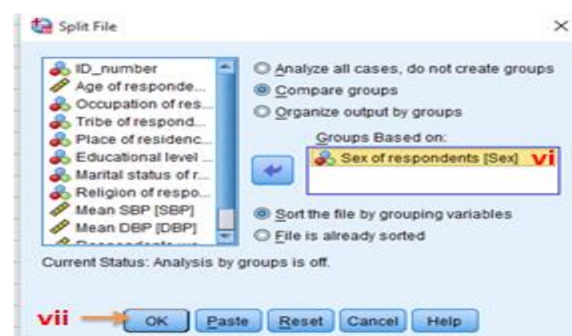
For each of the variables, the number of subjects involved (i.e., 300), Minimum, Maximum, Mean, and Standard Deviation are displayed.

B1b_To run descriptives on quantitative variables for males and females separately

- Running descriptives on quantitative variables for males and females separately requires splitting the file by sex.
- Click on **Data tab** [i], and then click **Split File** [ii].
- In the Split File box that appears, select **Compare groups** [iii], click on the grouping variable (i.e., **Sex of respondents**)[iv], and click on the arrow [v] to move it into the **Groups Based on** box [vi].
- Click **OK** [vii].
- Finally, perform the analysis as described in Steps [1] to [6]







B1b_To run descriptives on quantitative variables for males and females separately contd.

- The results of the analysis are displayed in a table on the **viewer** window.

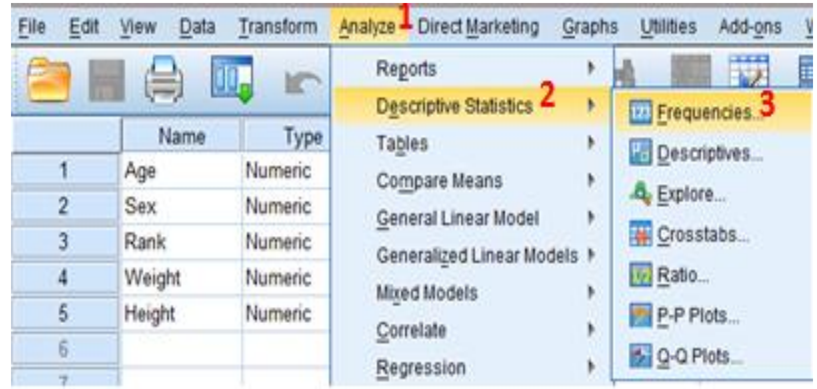
Descriptives

Descriptive Statistics

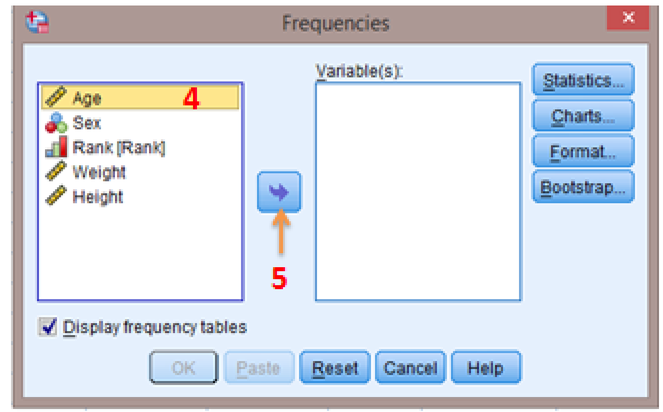
Sex of respondents		N	Minimum	Maximum	Mean	Std. Deviation
male	Age of respondents	161	18	91	41.40	18.244
	Mean SBP	161	96.00	229.00	132.5267	22.19807
	Mean DBP	161	52.50	130.00	82.9814	13.52150
	Respondents weight	161	40.00	129.00	64.9565	13.23128
	Respondents height	161	126.00	199.00	170.0248	8.28247
	Valid N (listwise)	161				
female	Age of respondents	139	18	87	39.37	17.848
	Mean SBP	139	85.00	220.00	130.4000	25.40768
	Mean DBP	139	48.50	124.00	82.8525	14.07131
	Respondents weight	139	30.00	125.00	59.6331	13.77676
	Respondents height	139	61.00	181.00	158.8633	10.60298
	Valid N (listwise)	139				

The results for males and females are presented separately. There were 161 males and 139 females. Also for each of the variables the Minimum, Maximum, Mean, and Standard Deviation are displayed.

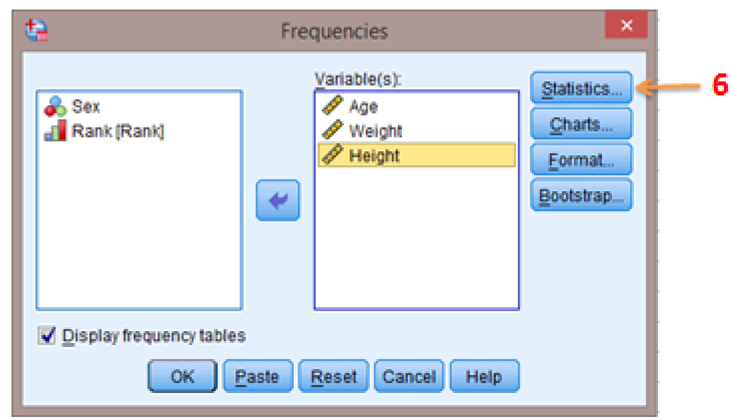
B2a_To run descriptives on quantitative variables (alternative pathway)



- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Frequencies** [3].



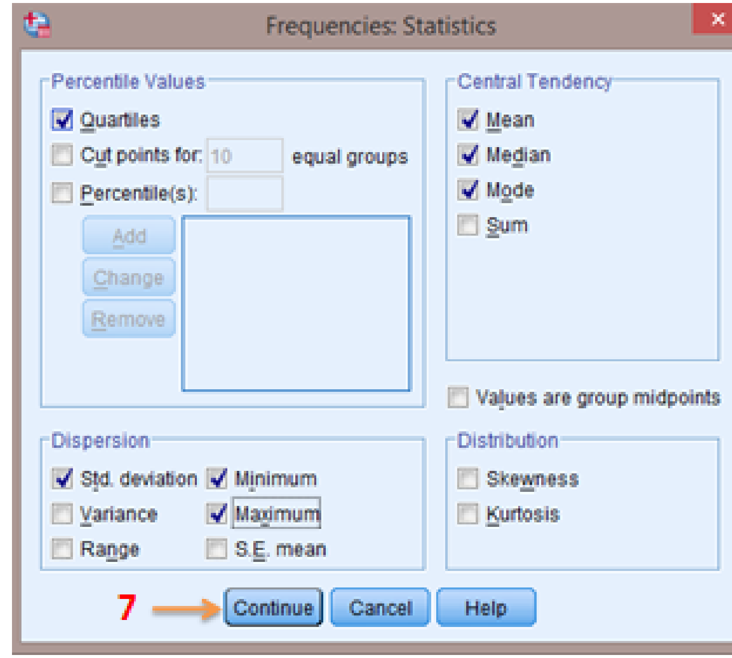
- Double click on the quantitative variable you want to analyze (e.g., **Age** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.



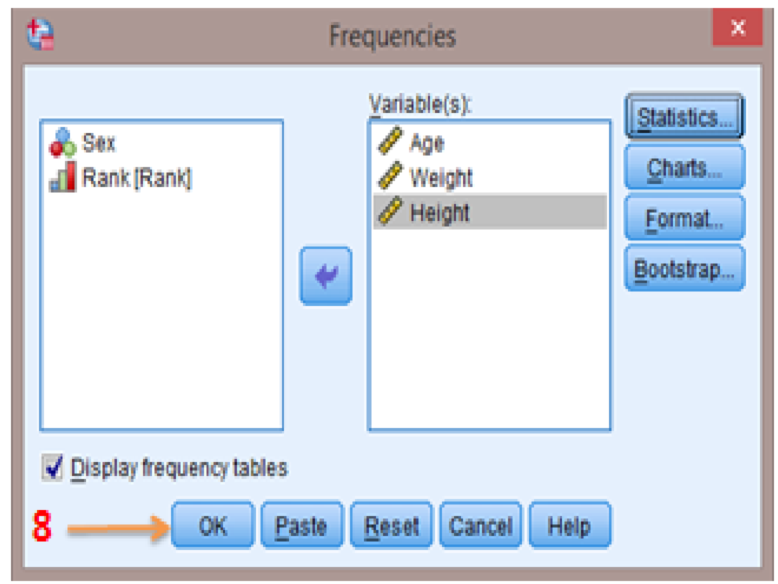
- Repeat the procedure (step 4 or 5) to move the other quantitative variables (i.e., **SBP**, **DBP**, **Weight** and **Height**) to the Variable(s) box and click **Statistics** [6].

B2a_To run descriptives on quantitative variables (alternative pathway) contd.

- In the **Frequencies Statistics** box that appears select the measures of Central Tendency (i.e., **Mean**, **Median**, **Mode**), Dispersion (i.e., **Minimum**, **Maximum**, **Standard deviation**) and Percentile Values (i.e., **Quartiles**) that you want and click **Continue** [7].



- In the **Frequencies** box that re-appears click **OK** [8].



B2a_To run descriptives on quantitative variables (alternative pathway) contd.

- The results of the analysis are displayed in a table on the **viewer** window.

Frequencies

Statistics

		Age of respondents	Mean DBP	Mean SBP	Respondents weight	Respondents height
N	Valid	300	300	300	300	300
	Missing	0	0	0	0	0
Mean		40.46	82.9217	131.5413	62.4900	164.8533
Median		38.50	80.0000	126.5000	60.0000	165.0000
Mode		50	80.00	110.00	55.00	175.00
Std. Deviation		18.060	13.75592	23.72248	13.72411	10.93972
Minimum		18	48.50	85.00	30.00	61.00
Maximum		91	130.00	229.00	129.00	199.00
Percentiles	25	24.00	73.1250	115.0000	53.0000	158.2500
	50	38.50	80.0000	126.5000	60.0000	165.0000
	75	53.00	91.7500	145.0000	70.0000	172.0000

For a normally distributed data, the mean, median and mode are equal. The result obtained showed that none of the variables is normally distributed.

Whereas, the mean is the appropriate measure of Central Tendency for a normally distributed data, the median is the appropriate measure of Central Tendency for a distribution free data.

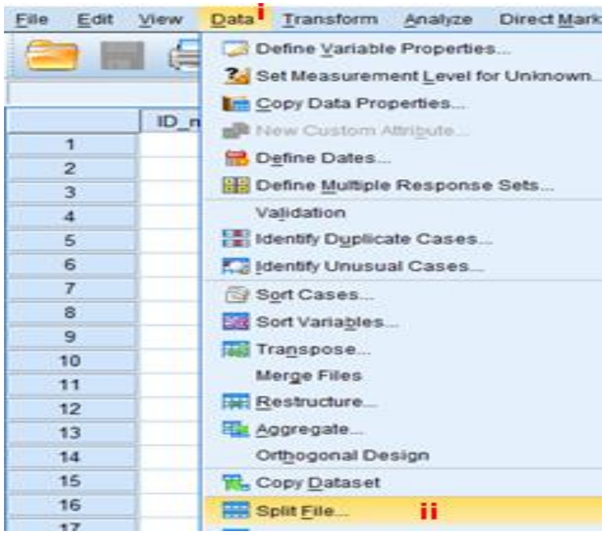
Please note that the 50th percentile is also the median

In the dissertation / project report, the results for age can be presented as:

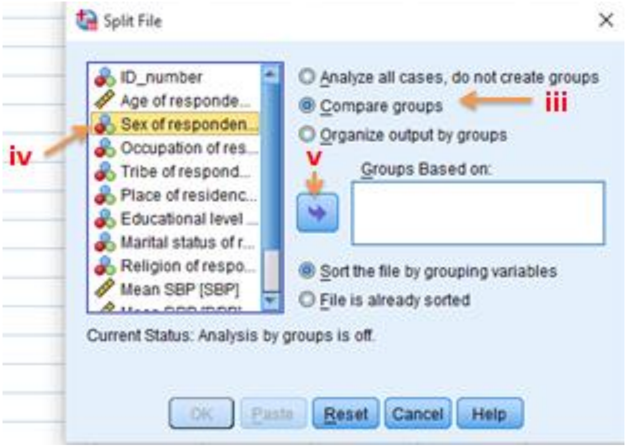
The ages of the study participants ranged from 18 to 91 years with a median age of 38.5 years, and inter-quartile range (IQR) of 24.0 – 53.0 years.

(Please note that the **median** and **IQR** are used as the measures of **central tendency** and **dispersion** respectively here instead of the **mean** and **standard deviation** because the data are not normally distributed).

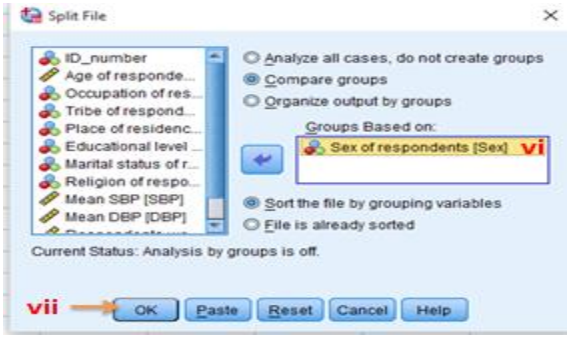
B2b_To run descriptives on quantitative variables for males and females separately (alternative pathway)



- Running descriptives on quantitative variables for males and females separately requires splitting the file by sex.
- Click on **Data tab** [i], and then click **Split File** [ii].



- In the Split File box that appears, select **Compare groups** [iii], click on the grouping variable (i.e., **Sex of respondents**)[iv], and click on the arrow [v] to move it into the **Groups Based on** box [vi].



- Click **OK** [vii].
- Finally, perform the analysis as described in Steps [1] to [8]

B2b_To run descriptives on quantitative variables for males and females separately (alternative pathway) contd.

- The results of the analysis are displayed separately for males and females in a table on the **viewer** window.

Frequencies

Statistics

Sex of respondents			Age of respondents	Mean SBP	Mean DBP	Respondents weight	Respondents height
male	N	Valid	161	161	161	161	161
		Missing	0	0	0	0	0
	Mean		41.40	132.5267	82.9814	64.9565	170.0248
	Median		40.00	130.0000	80.0000	63.0000	171.0000
	Mode		22[a]	110.00	80.00	60.00	174.00
	Std. Deviation		18.244	22.19807	13.52150	13.23128	8.28247
	Minimum		18	96.00	52.50	40.00	126.00
	Maximum		91	229.00	130.00	129.00	199.00
	Percentiles	25	24.00	116.0000	73.2500	55.0000	165.5000
		50	40.00	130.0000	80.0000	63.0000	171.0000
		75	55.00	145.7500	92.0000	72.0000	175.0000
female	N	Valid	139	139	139	139	139
		Missing	0	0	0	0	0
	Mean		39.37	130.4000	82.8525	59.6331	158.8633
	Median		35.00	125.0000	80.0000	55.0000	159.0000
	Mode		50	115.00[a]	80.00	55.00	159.00[a]
	Std. Deviation		17.848	25.40768	14.07131	13.77676	10.60298
	Minimum		18	85.00	48.50	30.00	61.00
	Maximum		87	220.00	124.00	125.00	181.00
	Percentiles	25	24.00	115.0000	73.0000	50.0000	154.0000
		50	35.00	125.0000	80.0000	55.0000	159.0000
		75	50.00	145.0000	90.0000	68.0000	164.0000

a. Multiple modes exist. The smallest value is shown

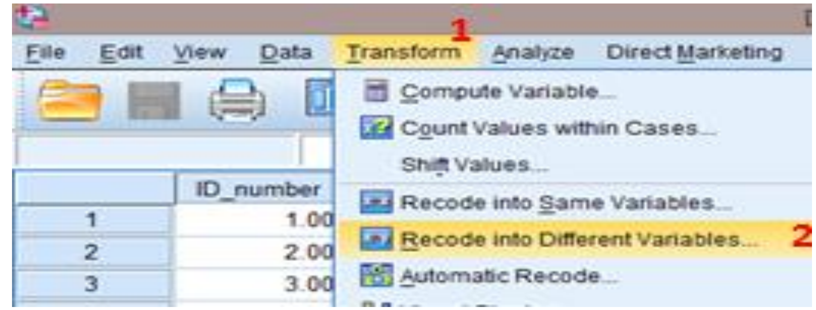
C_To transform quantitative to qualitative variables

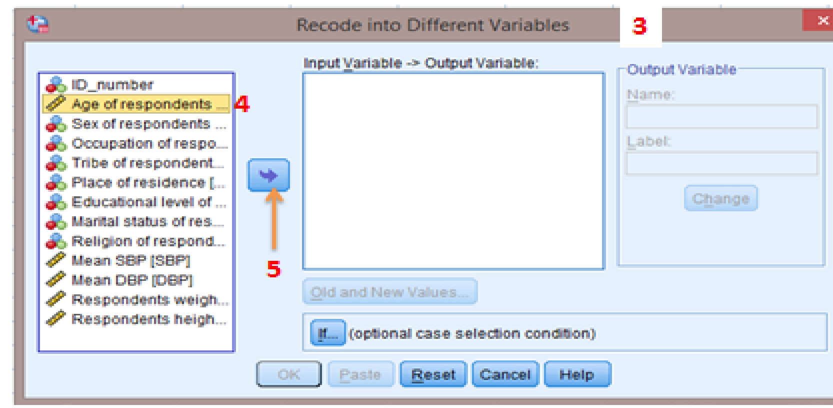
- In addition to reporting the descriptive statistics for age, it is necessary to know the distribution of respondents by age groups (i.e., age categories).
- This requires transforming the data from quantitative to qualitative variable.
- In transforming a variable from quantitative to qualitative, the results of the descriptive statistics are used in forming the categories.
- Since the **minimum age** = **18** and the **maximum age** = **91**, the age distribution can be designed as shown below:

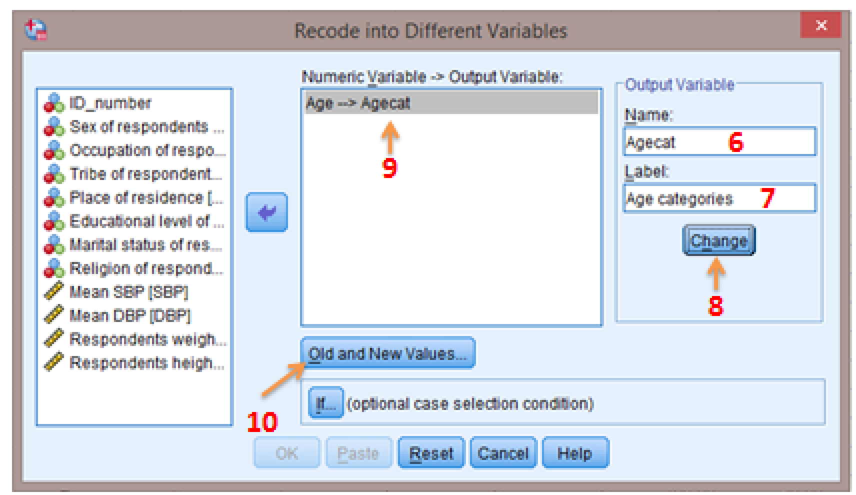
<20 [or below 20 (i.e., 18 – 19) years]
20 – 29
30 – 39
40 -49
50 and above (i.e., 50 – 91 years)

C_To transform quantitative to qualitative variables contd.

- Click **Transform** [1] > **Recode into Different Variables** [2].
- In the **Recode into Different Variables** box [3] that appears, double click on the quantitative variable you want to transform (i.e., **Age of respondents** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.
- Type the name of the new variable you want to create (i.e **Agecat**) [6] in the **Name** box, type the label (i.e., **Age categories**) [7] in the **Label** box, and click **Change** [8].
- The new variable's name appears in front of the old variables name [9] in the box .
- Click **Old and New Values** [10].



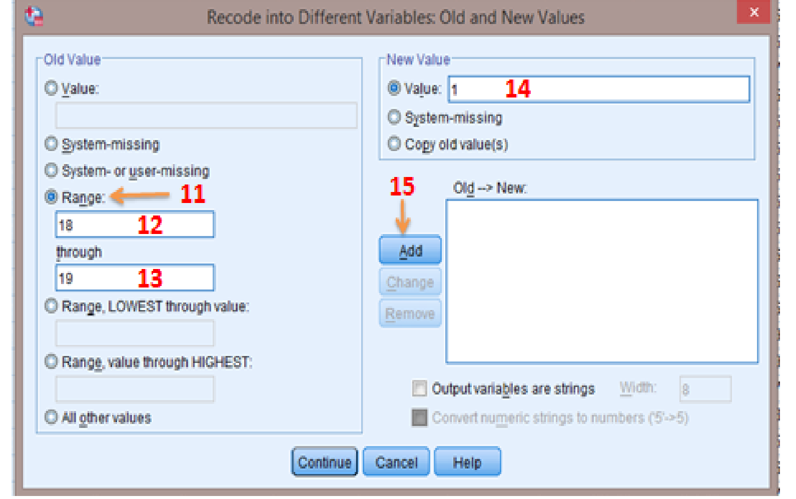


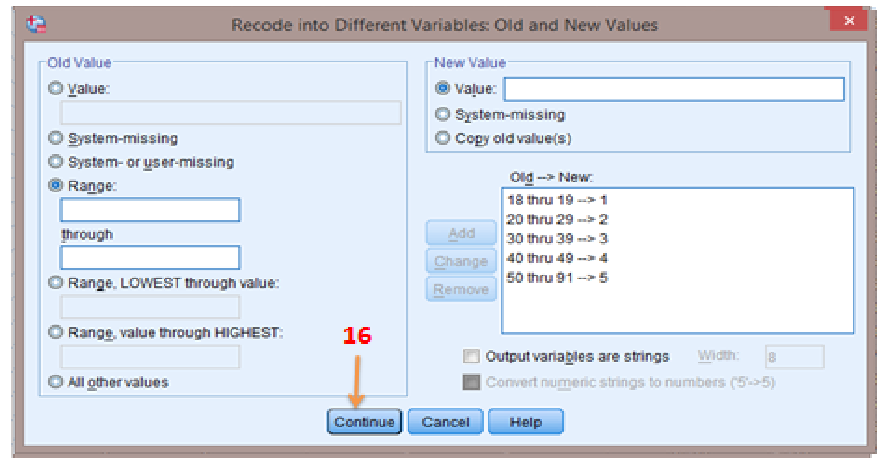


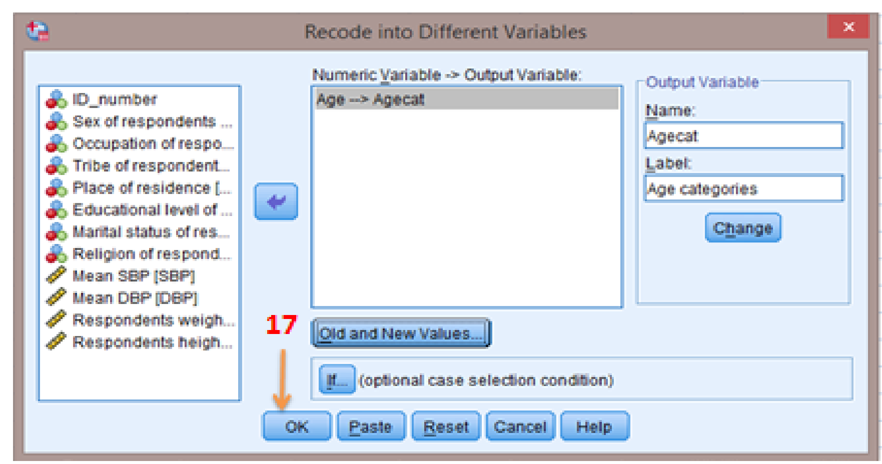
>

C_To transform quantitative to qualitative variables contd.

- In the new box that appears, click on **Range** [11] to activate it.
- Type the lower limit (i.e., **18**)[12] and upper limit (i.e., **19**) [13] of the first class in the boxes shown, type its value (i.e., **1)** [14] in the **New Value** box and click **Add** [15] to move the class into the **Old to New** Box.
- Repeat the procedure for the other classes (assign values of **2**, **3**, **4** and **5** to classes **20-29**, **30-39**, **40-49**, **50 and above** respectively).
- After all the classes have been entered into the box, click **Continue** [16].
- Finally, click **OK** [17] in the **Recode into Different Variables** box that re-appears.

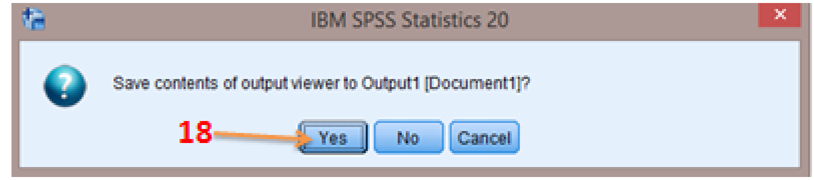


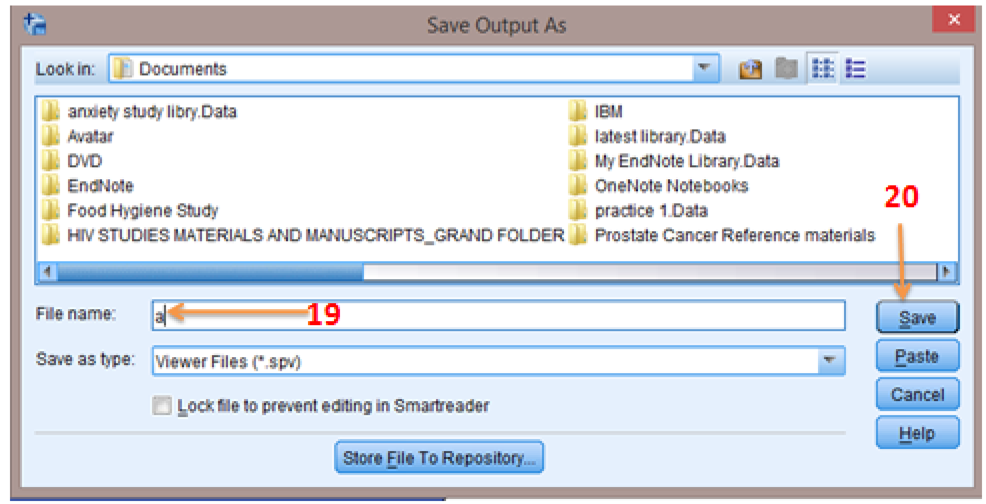


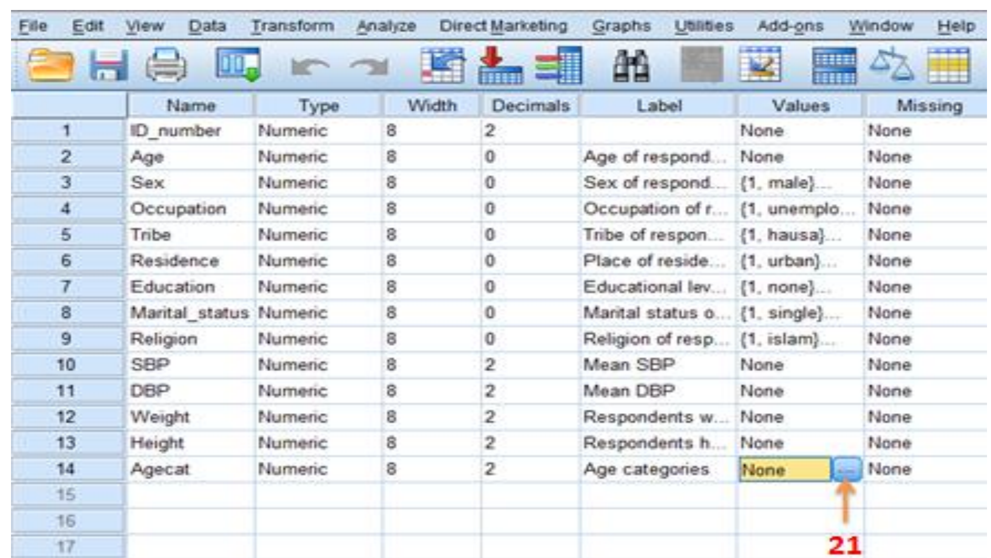


C_To transform quantitative to qualitative variables contd.

- Close the viewer window that appears indicating that the task has been carried out, always select **YES** [18] in all the queries that appear.
- Also, in the **Save Output As** dialog box that appears, type "**a**" [19] in the **File name** box, and click **Save** [20].
- Go to the **variable view** and scroll down to locate the recoded variable (i.e., **Agecat**), which now appears as the last variable.
- Click in the cell under **Values** and click on the small box with 3 dots at the right end of the cell [21].

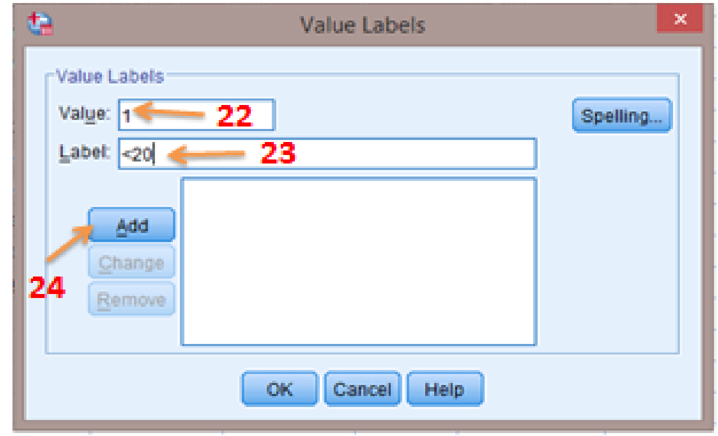


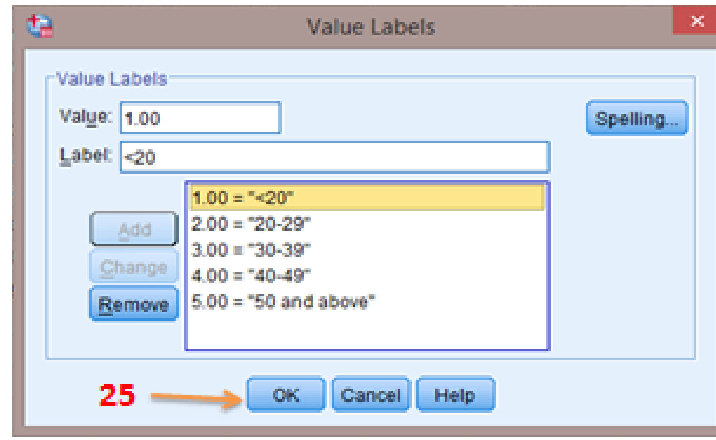


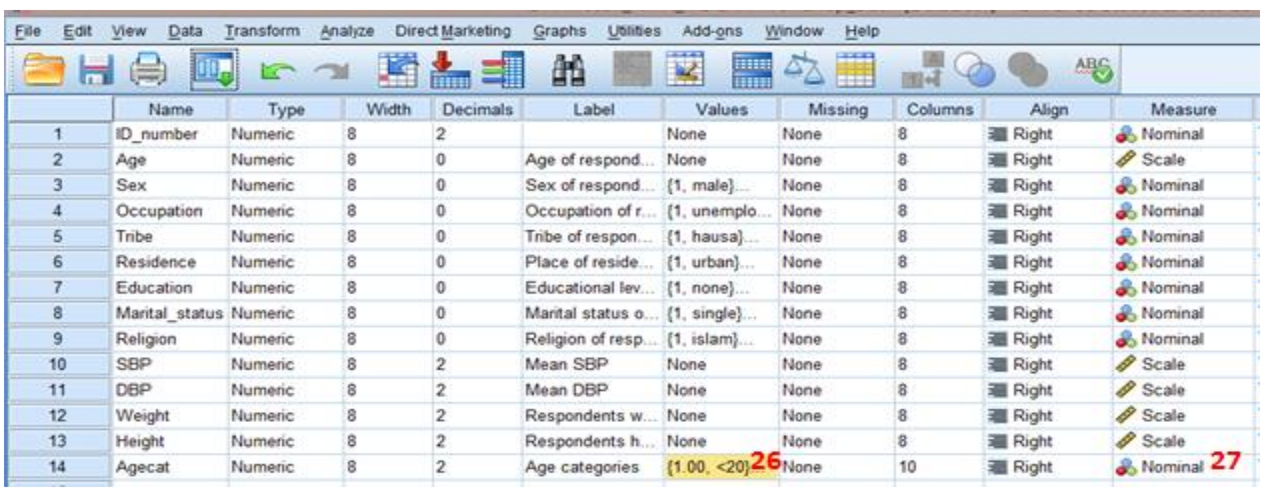


C_To transform quantitative to qualitative variables contd.

- In the **Value labels** dialog box that appears, type **1** in the **Value** box [22], type **<20** in the **Label** box [23], and click **Add** [24] to move them into the box.
- Repeat the procedure for the other values (i.e., **2**,**3**,**4**, and **5**) and their labels (i.e., **30-39**, **40-49**, **50-59** and **50 and above**) respectively, and then click **OK** [25].
- The values are displayed in the cell under the **Values** column [26], and the measure is **Nominal** [27].





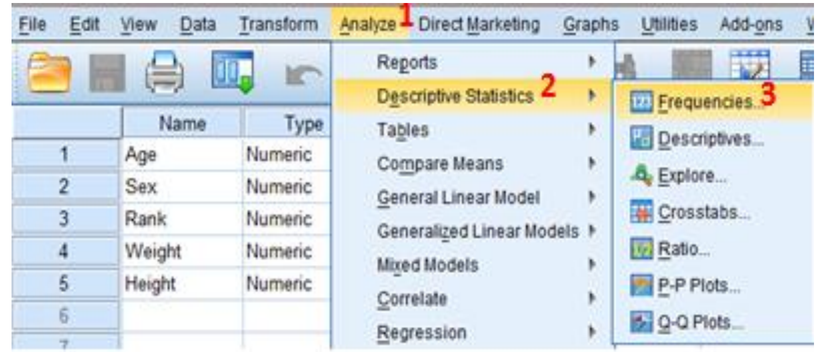


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID_number	Numeric	8	2		None	None	8	Right	Nominal
2	Age	Numeric	8	0	Age of respond...	None	None	8	Right	Scale
3	Sex	Numeric	8	0	Sex of respond...	{1, male}...	None	8	Right	Nominal
4	Occupation	Numeric	8	0	Occupation of r...	{1, unemplo...	None	8	Right	Nominal
5	Tribe	Numeric	8	0	Tribe of respon...	{1, hausa}...	None	8	Right	Nominal
6	Residence	Numeric	8	0	Place of reside...	{1, urban}...	None	8	Right	Nominal
7	Education	Numeric	8	0	Educational lev...	{1, none}...	None	8	Right	Nominal
8	Marital_status	Numeric	8	0	Marital status o...	{1, single}...	None	8	Right	Nominal
9	Religion	Numeric	8	0	Religion of resp...	{1, islam}...	None	8	Right	Nominal
10	SBP	Numeric	8	2	Mean SBP	None	None	8	Right	Scale
11	DBP	Numeric	8	2	Mean DBP	None	None	8	Right	Scale
12	Weight	Numeric	8	2	Respondents w...	None	None	8	Right	Scale
13	Height	Numeric	8	2	Respondents h...	None	None	8	Right	Scale
14	Agecat	Numeric	8	2	Age categories	{1.00, <20} 26	None	10	Right	Nominal 27

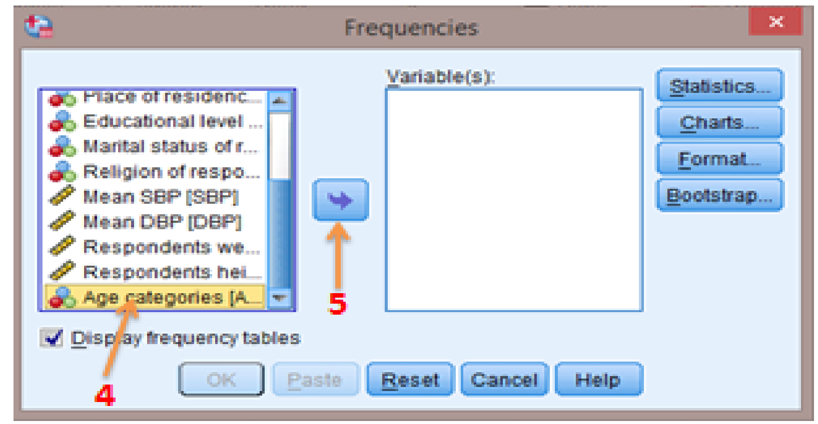
>

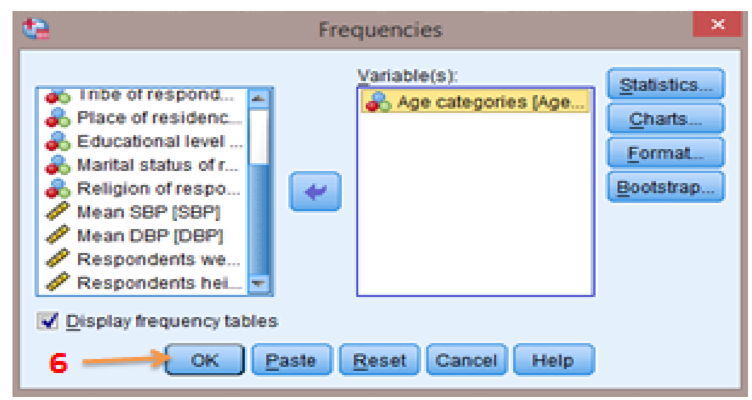
D_To run frequencies on the newly created age categories

- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Frequencies** [3].



- Double click on the qualitative variable you want to analyze (i.e., **Agecat** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.

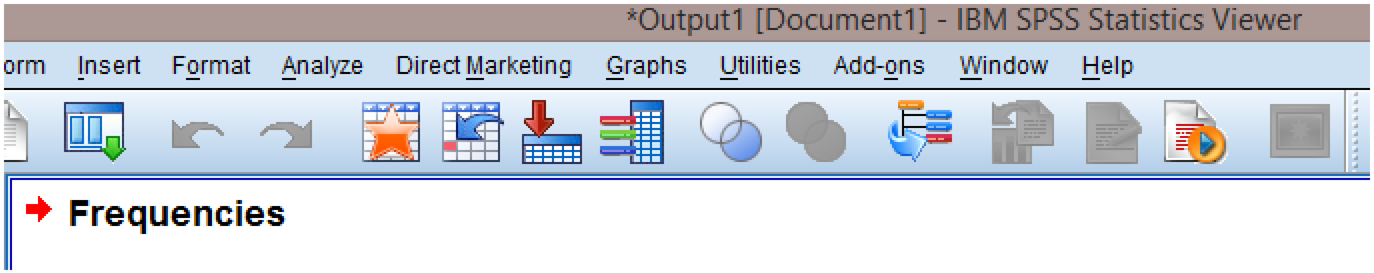




- Finally, click **OK** [6].

D_To run frequencies on the newly created age categories contd.

- The results of the analysis are displayed in a table on the **viewer** window.



Statistics

Age categories

N	Valid	300
	Missing	0

Age categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<20	28	9.3	9.3	9.3
	20-29	81	27.0	27.0	36.3
	30-39	42	14.0	14.0	50.3
	40-49	45	15.0	15.0	65.3
	50 and above	104	34.7	34.7	100.0
	Total	300	100.0	100.0	

The output shows the following:
All the 300 subjects were entered for the variables (i.e., no missing values). This is why the Percent (based on the sample size), is the same as the Valid Percent (based on the values entered)

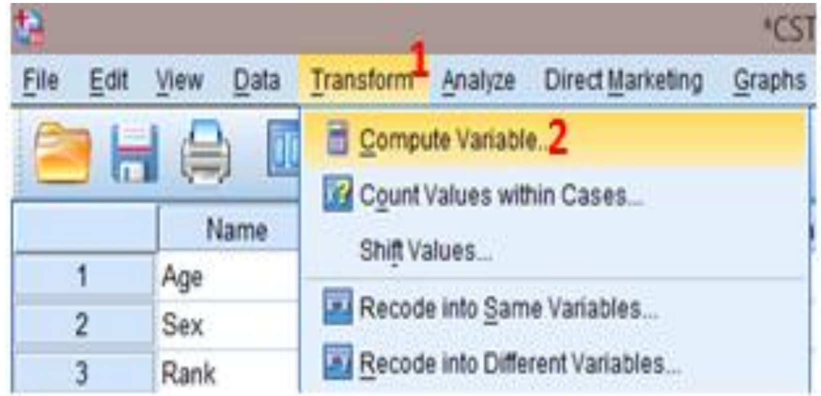
If there are missing values it is the Valid Percent that should be reported).

In the dissertation the results for age categories can be presented as follows:
A larger proportion 104 (34.7%) of the 300 respondents were aged 50 years and above, followed by those that were aged 20-29 years (27.0%), while only a few (9.3%) were aged <20 years.

E_To compute a new variable

To determine the nutritional status of the study participants it is necessary to compute a new variable called **Body Mass Index** (**BMI**) from **Weight** and **Height**.

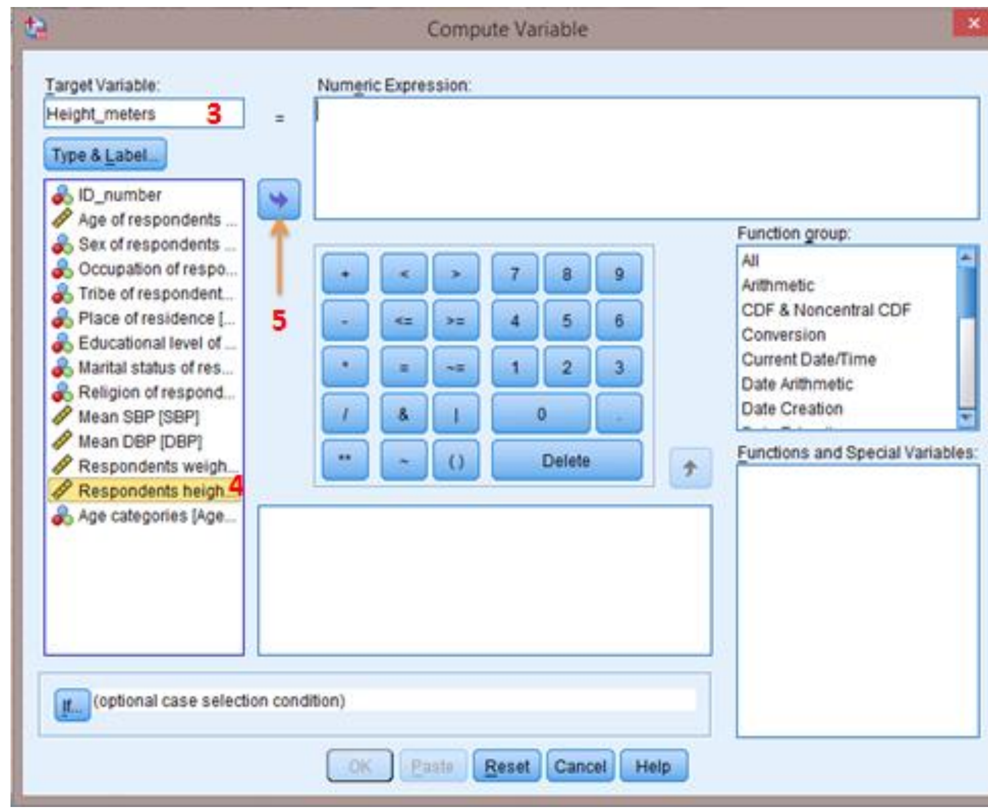
$$BMI = \frac{Weight\ (kg)}{Height^2 (m)^2}$$



Unlike weight which was appropriately entered in kilograms in this dataset, height was entered In centimeters, and as such, it must be converted into meters before computing the BMI.

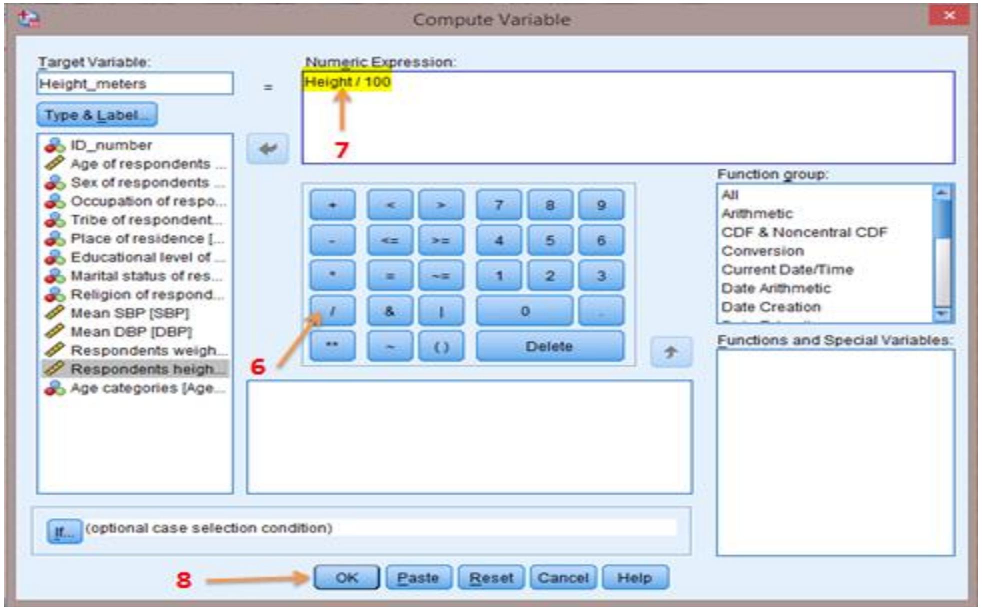
To compute Height (in meters) from Height (in centimeters):

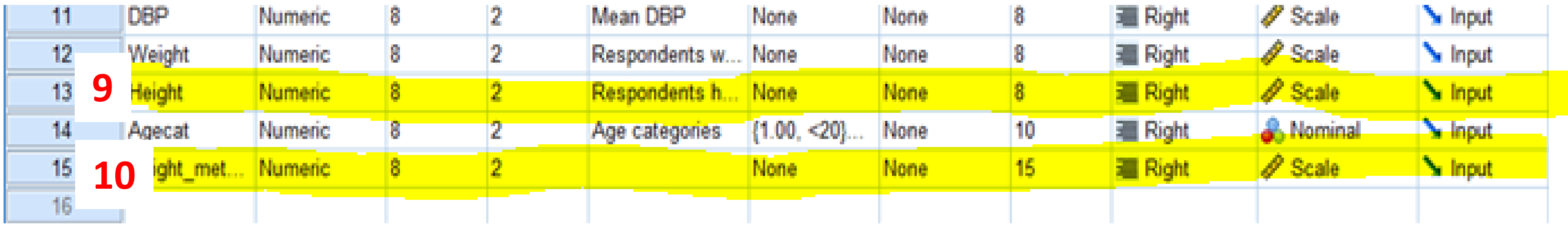
- Click **Transform** [1] > **Compute Variable** [2].
- In the **Compute Variable** box that appears, type **Height_meters** [3] in the **Target Variable** box.
- Click on **Respondents height** [4], and use the arrow [5] to enter it into the **Numeric Expression** box.



E_To compute a new variable contd.

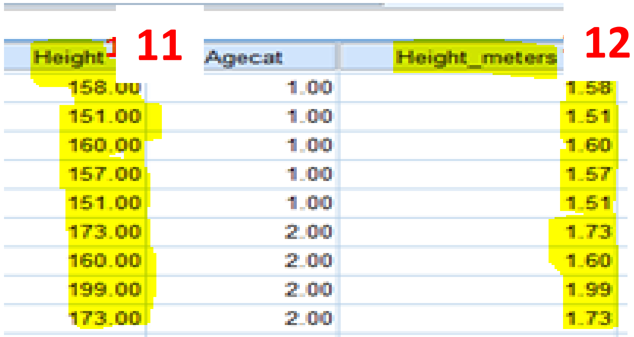
- Click on the **division "/"** button [6] on the calculator to enter it into the box, and then type 100 (i.e., **Height / 100**) [7].
- Click **OK** [8].
- In the Variable view; the originally entered **Height** (in centimeters) [9] is now converted to the newly computed **Height_meters** (in meters) [10] with the measure still retained as scale, and it appears as the last variable.





11	DBP	Numeric	8	2	Mean DBP	None	None	8	Right	Scale	Input
12	Weight	Numeric	8	2	Respondents w...	None	None	8	Right	Scale	Input
13 [9]	Height	Numeric	8	2	Respondents h...	None	None	8	Right	Scale	Input
14	Agecat	Numeric	8	2	Age categories	{1.00, <20}...	None	10	Right	Nominal	Input
15 [10]	ight_met...	Numeric	8	2		None	None	15	Right	Scale	Input
16											

- In the Data view; the originally entered **Height** (in centimeters) [11] now appears as the newly computed **Height_meters** (in meters) [12].

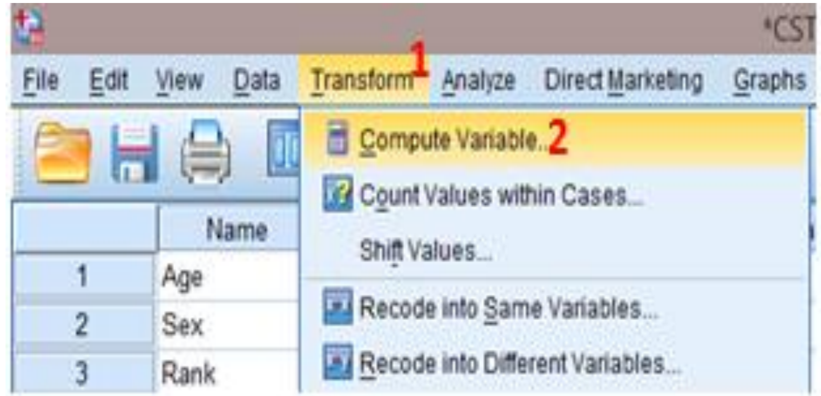


Height [11]	Agecat	Height_meters [12]
158.00	1.00	1.58
151.00	1.00	1.51
160.00	1.00	1.60
157.00	1.00	1.57
151.00	1.00	1.51
173.00	2.00	1.73
160.00	2.00	1.60
199.00	2.00	1.99
173.00	2.00	1.73

E_To compute a new variable contd.

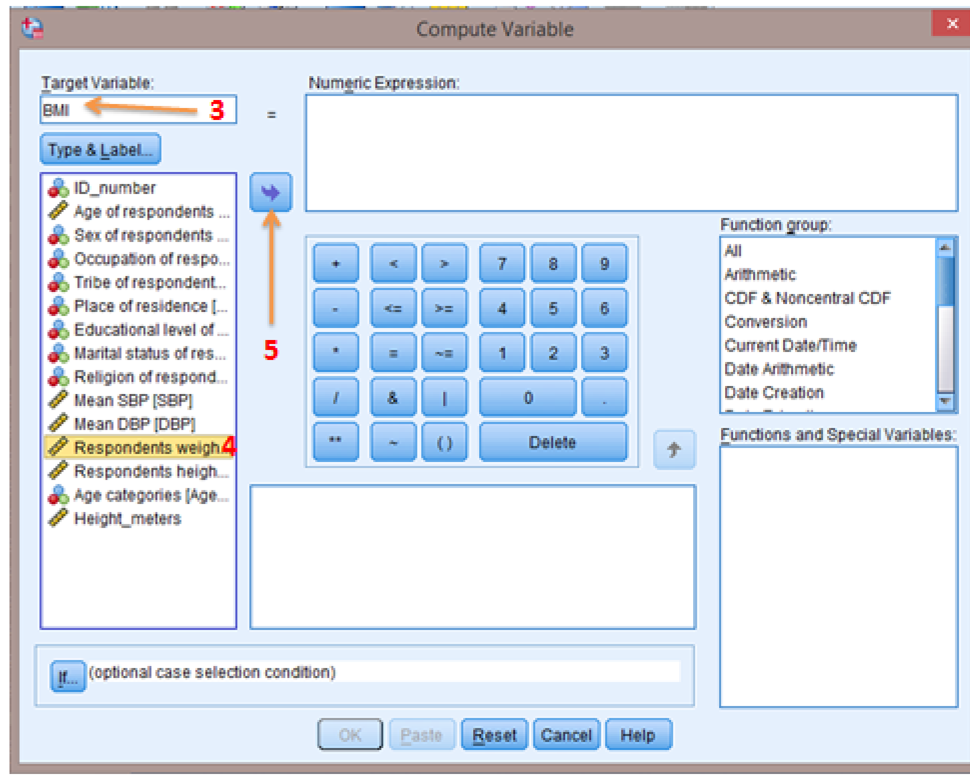
- The BMI can now be computed using the originally entered weight in kg and the newly computed height in meters.

$$BMI = \frac{Weight\ (kg)}{Height^2 (m)^2}$$



To compute the BMI:

- Click **Transform** [1] > **Compute Variable** [2].
- In the **Compute Variable** box that appears, type **BMI** [3] in the **Target Variable** box.
- Click on **Respondents' weight** [4], and use the arrow [5] to enter it into the **Numeric Expression** box.

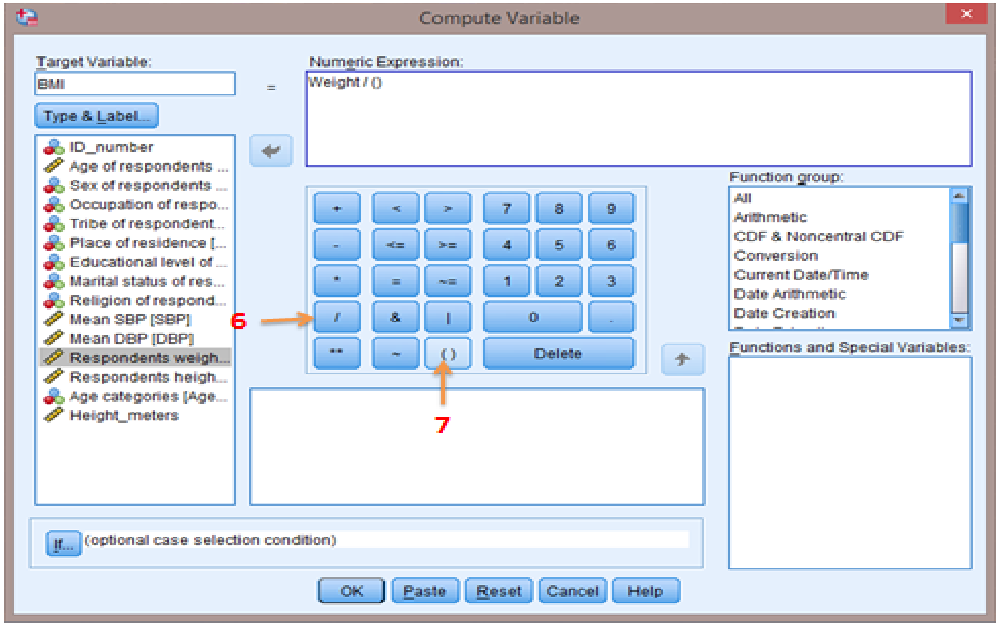


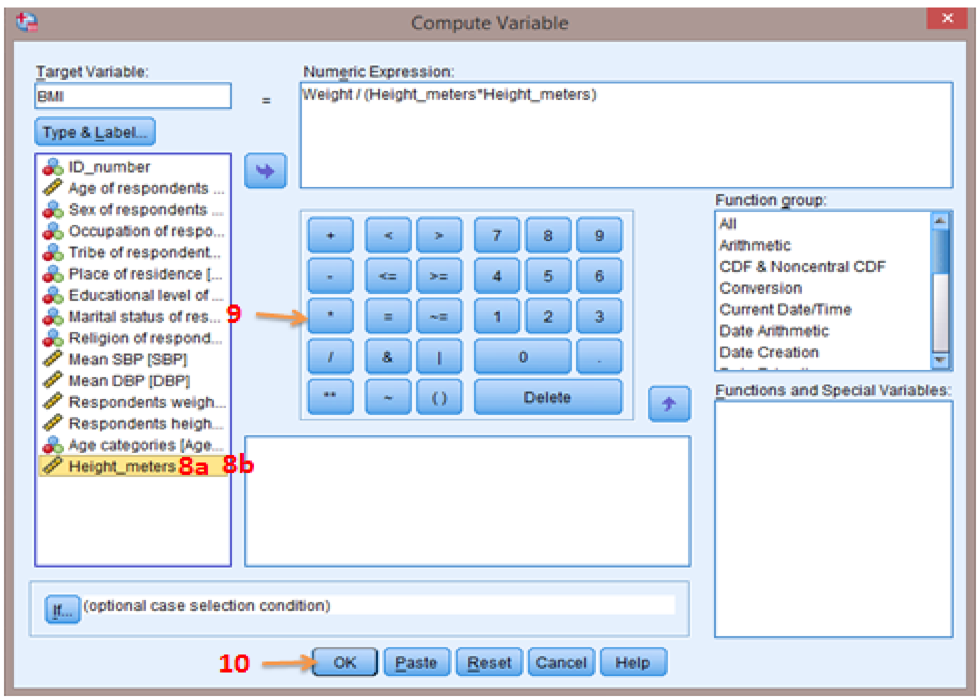
>

E_To compute a new variable contd.

- Click on the **division** "**/**" button [6] on the calculator to enter it into the box, and then click on the **brackets** "**()**" button [7] to enter it into the box.
- With the cursor still inside the brackets, click on **Height_meters** [8a], and use the arrow to enter it inside the brackets.
- Click on the **multiplication** "*****" button [9] to enter it inside the brackets, and click on **Height_meters** again [8b] to enter it inside the brackets.
- Finally, click **OK** [10].
- In the **Variable view**, the computed BMI appears as the last variable, while the values are shown in the **Data view**.
- The next thing to do is to transform the computed **BMI** (quantitative variable) into **BMI categories** (qualitative), through data transformation to be able to determine the nutritional status of the participants based on the World Health Organization's classification:

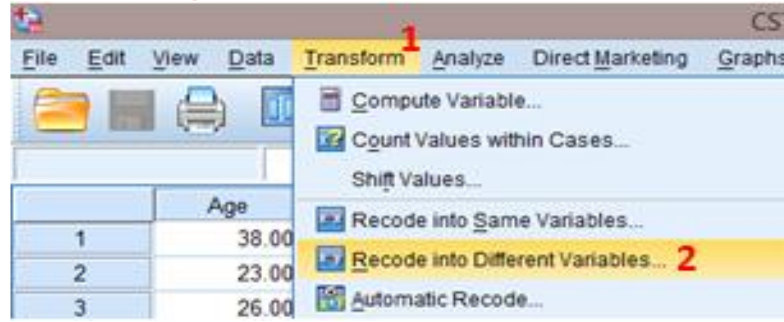
BMI (kg/m^2)	Nutritional status
< 18.5	Underweight
18.5 – 24.9	Normal weight
25.0 – 29.9	Overweight
30.0 and above	Obesity

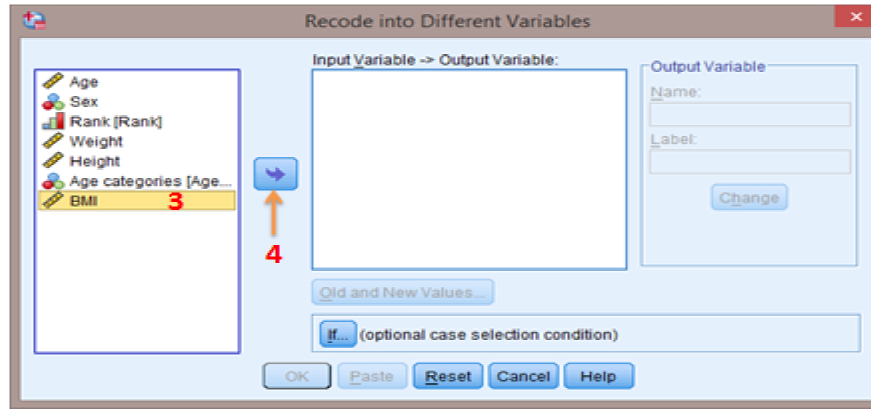


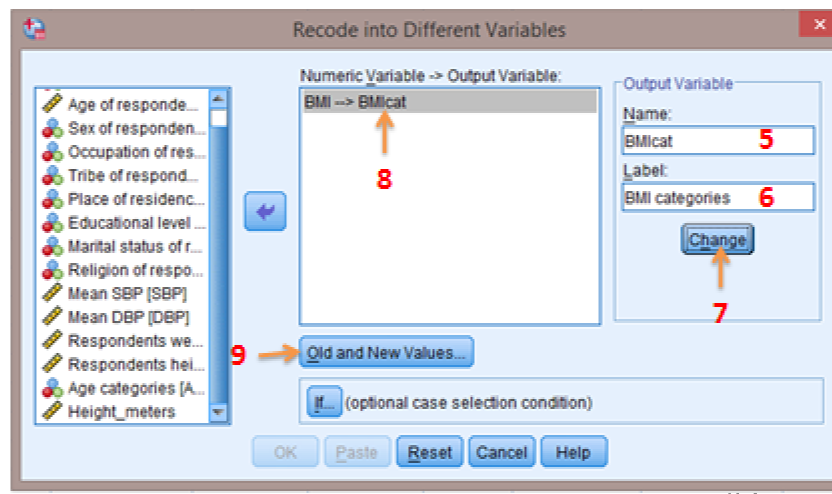


F_To transform the computed BMI into categories

- Click **Transform** [1] > **Recode into Different Variables** [2].
- In the **Recode into Different Variables** box that appears, double click on the quantitative variable you want to transform (i.e., **BMI** [3]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [4] to move it to the **Variable(s)** box.
- Type the name of the new variable you want to create (i.e **BMIcat**) [5] in the **Name** box, type the label (i.e., **BMI categories**) [6] in the **Label** box, and click **Change** [7].
- The new variable's name appears in front of the old variables name [8] in the box .
- Click **Old and New Values** [9].

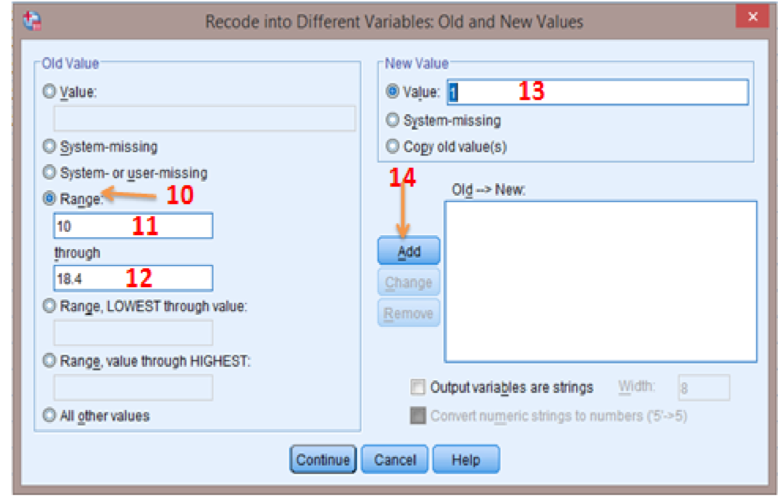


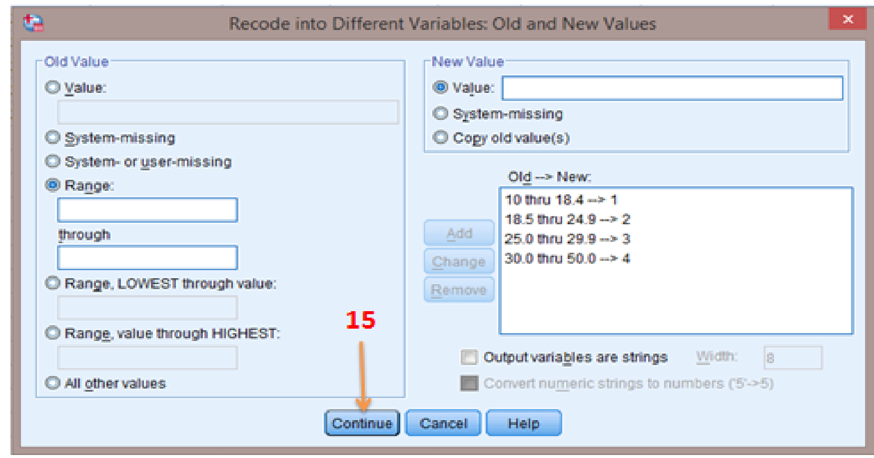


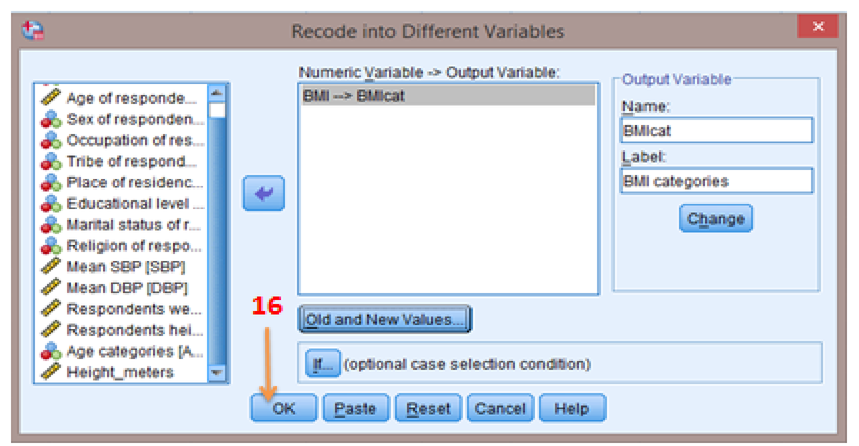


F_To transform the computed BMI into categories contd.

- In the new box that appears, click on **Range** [10] to activate it.
- Type any value substantially lower than 18.5 (e.g., **10**)[11] and the closest value to 18.5 (i.e., **18.4**) [12] in the boxes shown, type its value (i.e., **1)** [13] in the **New Value** box and click **Add** [14] to move the class into the **Old to New** Box.
- Repeat the procedure for the other classes (assign values of **2**, **3**, and **4** to classes **18.5-24.9**, **25.0-29.9** and **30.0 and above** respectively).
- After all the classes have been entered into the box, click **Continue** [15].
- Finally, click **OK** [16] in the **Recode into Different Variables** box that re-appears.

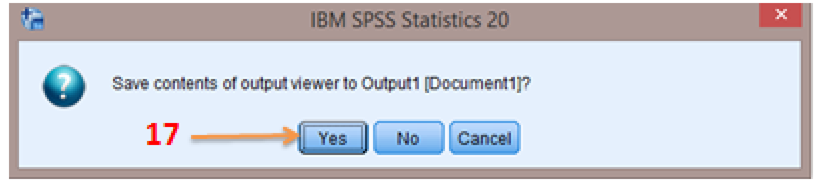


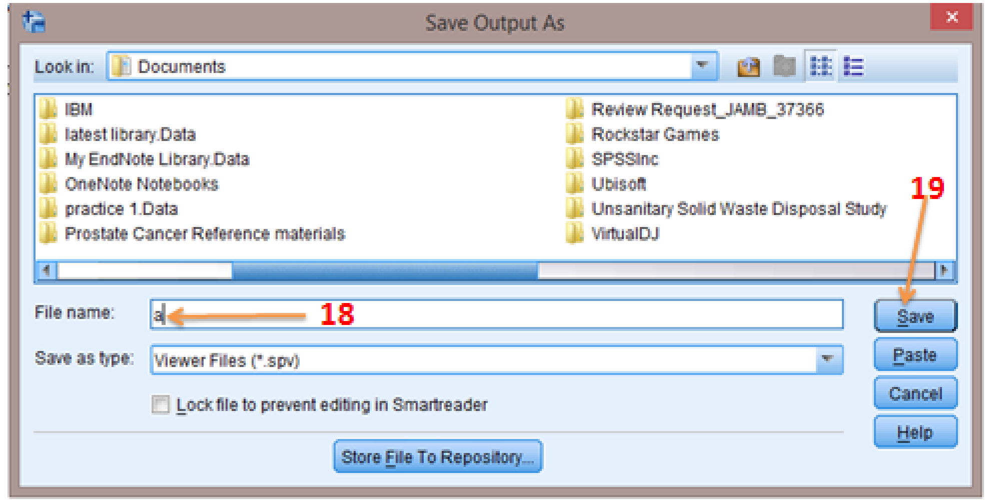


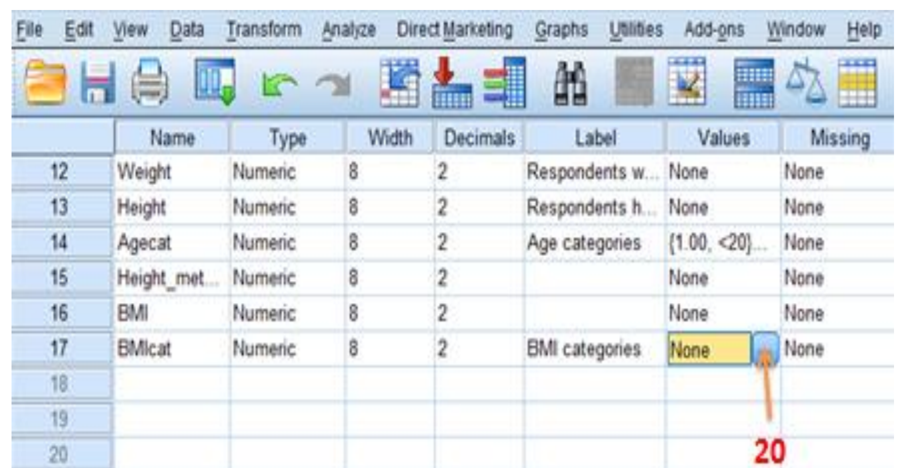


F_To transform the computed BMI into categories contd.

- Close the viewer window that appears indicating that the task has been carried out, always select **YES** [17] in all the queries that appear.
- Also, in the **Save Output As** dialog box that appears, type "**a**" [18] in the **File name** box, and click **Save** [19]. If a dialog box with replace existing "**a**" appears, click **YES**.
- Go to the **variable view** and scroll down to locate the recoded variable (i.e., **BMIcat**), which now appears as the last variable.
- Click in the cell under **Values** and click on the small box with 3 dots at the right end of the cell [20].

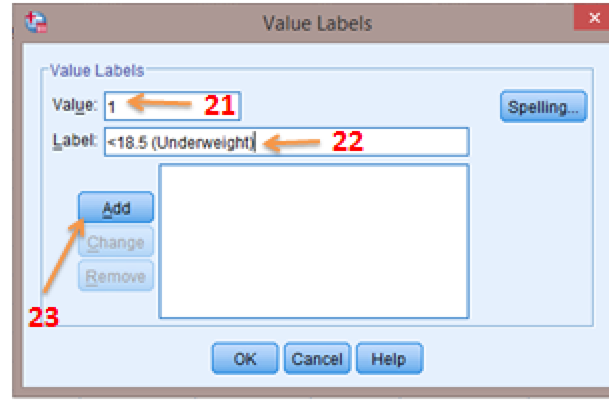


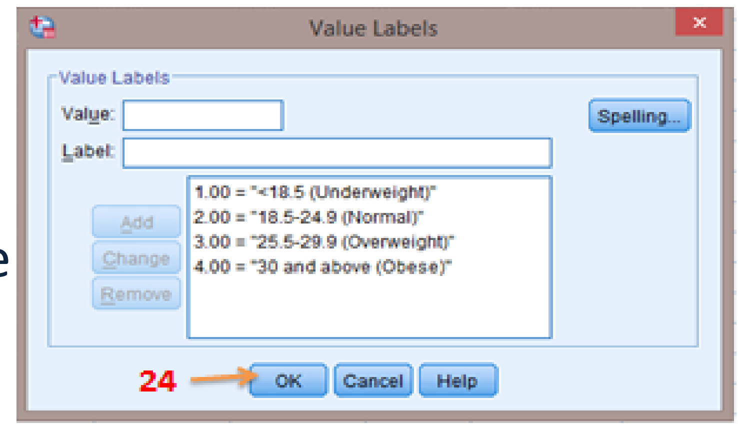


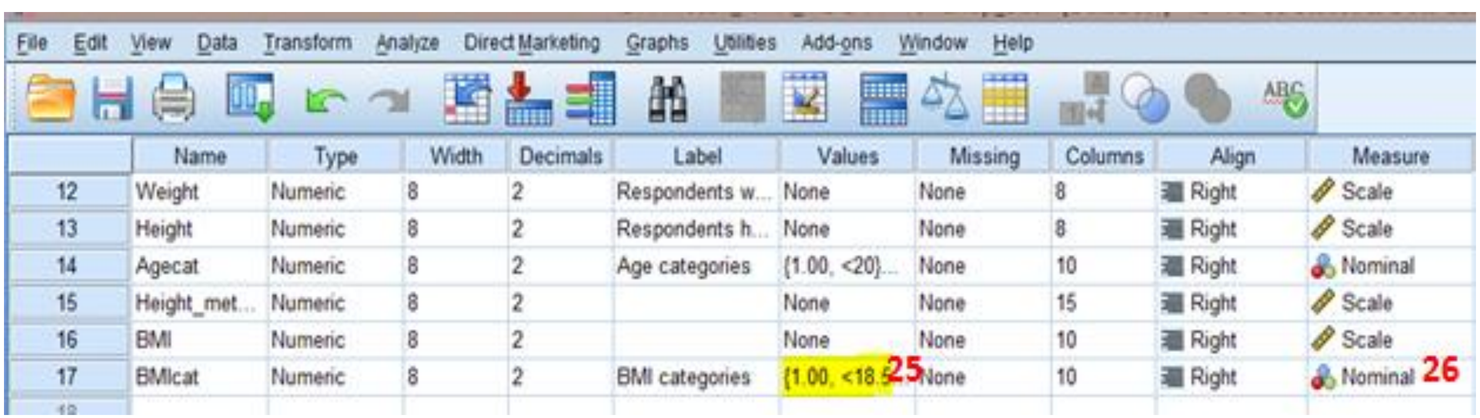


F_To transform the computed BMI into categories contd.

- In the **Value labels** dialog box that appears, type **1** in the **Value** box [21], type **<18.5 (Underweight)** in the **Label** box [22], and click **Add** [23] to move them into the box.
- Repeat the procedure for the other values [i.e., **2**,**3**,and **4**) and their labels (i.e., **18.5-24.9 (Normal)**, **25.0-29.9 (Overweight)**, and **30.0 and above (Obese)**] respectively, and then click **OK** [24].
- The values are displayed in the cell under the **Values** column [25], and the measure is **Nominal** [26].

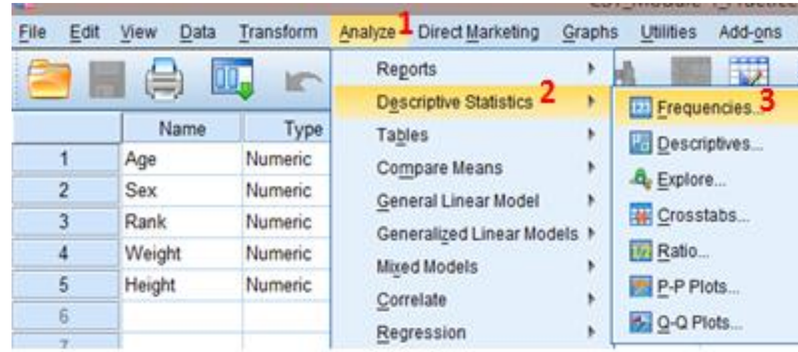




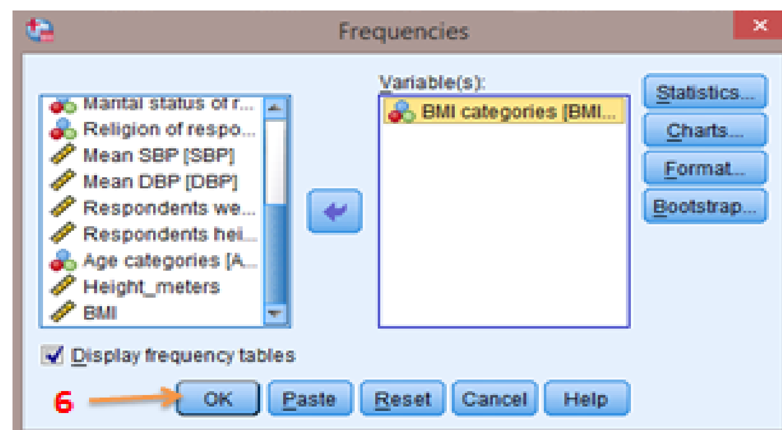


G_To run frequencies on the BMI categories

- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Frequencies** [3].



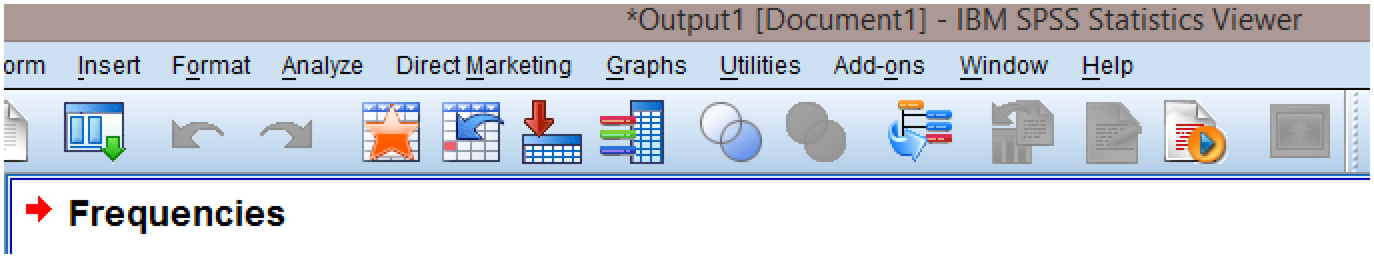
- Double click on the qualitative variable you want to analyze (i.e., **BMI categories** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.



- Finally, click **OK** [6].

G_To run frequencies on the BMI categories contd.

- The results of the analysis are displayed in a table on the **viewer** window.



Statistics

BMI categories

N	Valid	294
	Missing	6

BMI categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<18.5 (Underweight)	37	12.3	12.6	12.6
	18.5-24.9 (Normal)	171	57.0	58.2	70.7
	25.5-29.9 (Overweight)	64	21.3	21.8	92.5
	30 and above (Obese)	22	7.3	7.5	100.0
	Total	294	98.0	100.0	
Missing	System	6	2.0		
Total		300	100.0		

The output shows the following:
There were 6 missing values , this is why the Percent (based on the sample size), is less than the Valid Percent (based on the values entered)

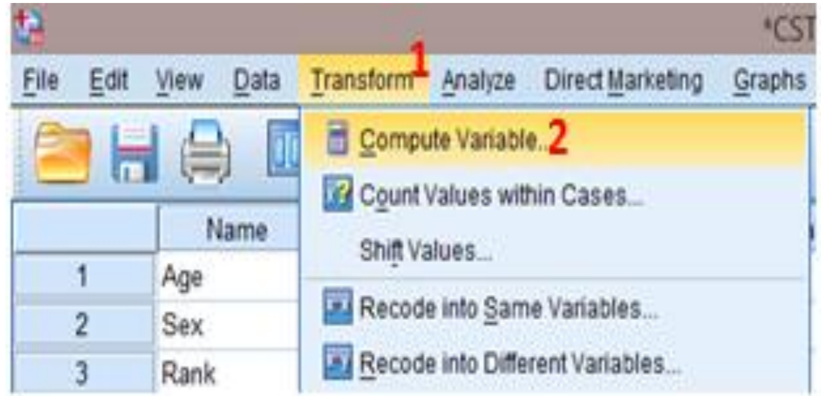
In this case it is the Valid Percent that should be reported).

In the dissertation the results for age categories can be presented as follows:
Majority 171 (58.2%) of the 294 respondents with complete data on BMI had normal weight, about a tenth (12.6%) were underweight, while close to a third (29.2%) were either overweight or obese.

H_To compute the Mean Arterial Pressure

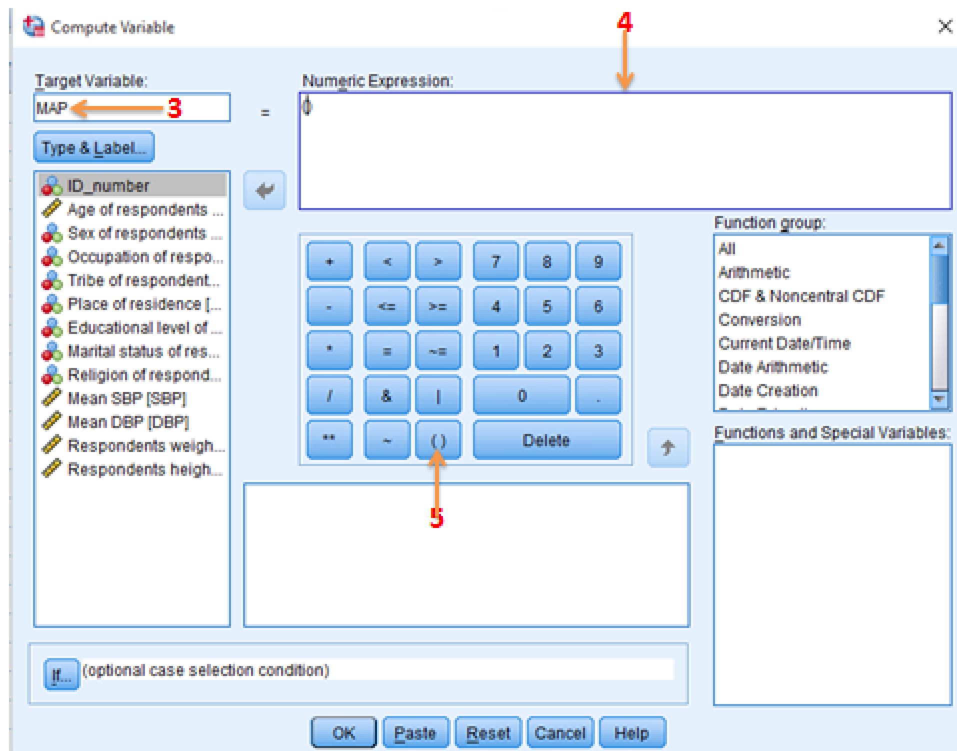
The **Mean Arterial Pressure** (**MAP**) is computed from the **Systolic Blood Pressure** (**SPB**) and **Diastolic Blood Pressure** (**DBP**) using the formula:

$$MAP = \frac{SBP + 2(DBP)}{3}$$



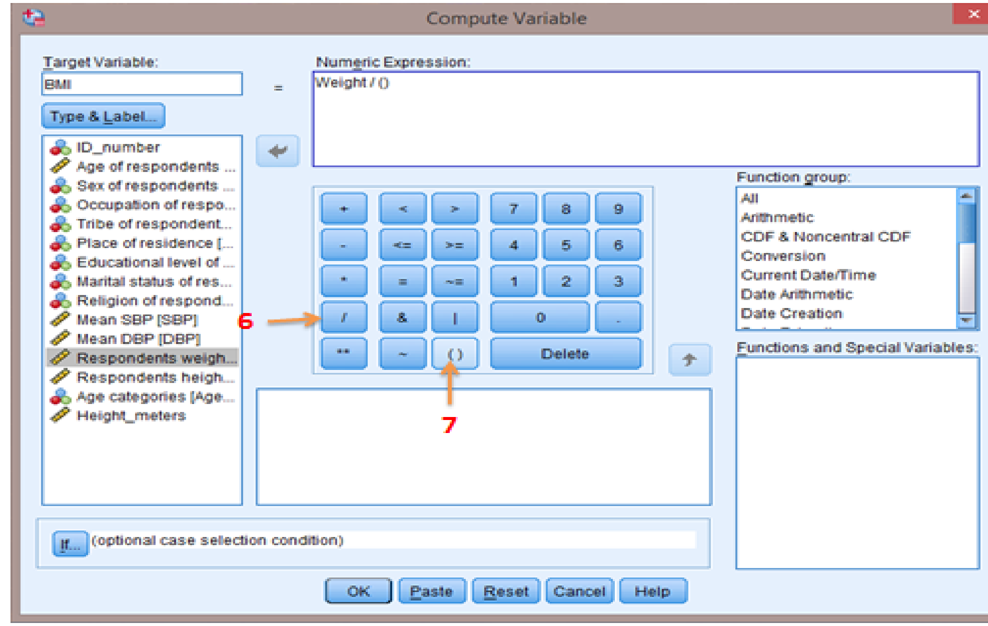
To compute the Mean Arterial Pressure (MAP) from SBP and DBP:

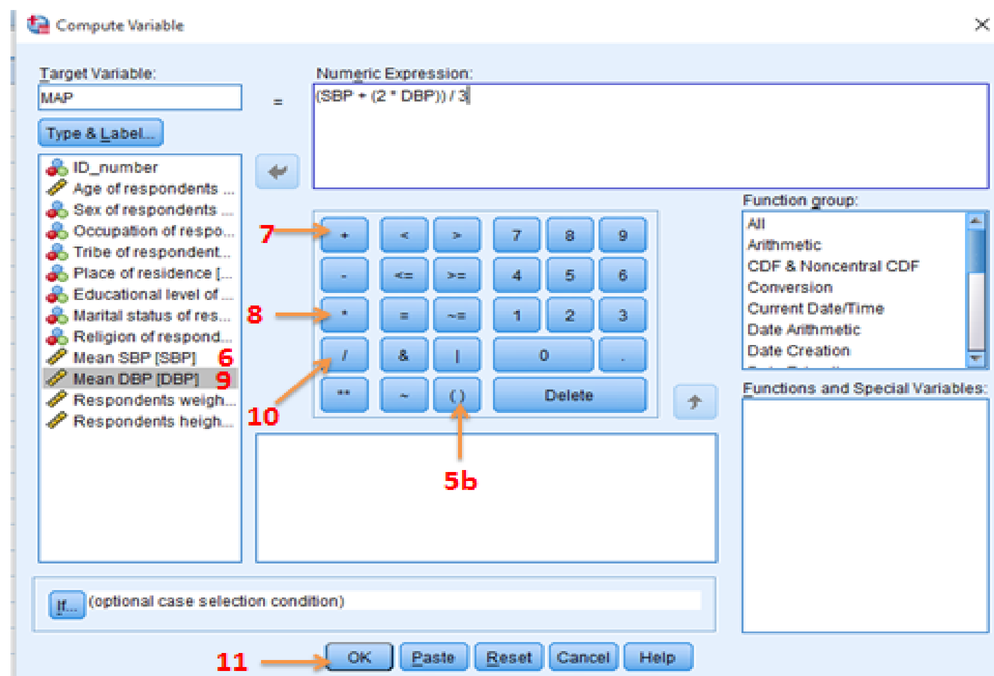
- Click **Transform** [1] > **Compute Variable** [2].
- In the **Compute Variable** box that appears, type **MAP** [3] in the **Target Variable** box.
- Place the cursor in the **Numeric Expression** box [4].
- Click on brackets button “**()**”[5] to enter it into the box.



H_To compute the Mean Arterial Pressure contd.

- With the cursor still inside the brackets, click on the brackets sign “()” button [5b] again to enter another brackets inside the first brackets.
- Click on **Mean SBP** [6], and use the arrow to enter it inside the first brackets (but outside the second brackets).
- Click on the **addition** “**+**” button [7] to enter it inside the first brackets immediately after SBP.
- Place the cursor inside the second brackets and type **2**.
- Click on the **multiplication** “*” button [8] to enter it inside the second brackets immediately after 2.
- Click on **Mean DBP** [9] and use the arrow to enter it inside the second brackets immediately after the multiplication sign.
- Place the cursor after the first brackets again (on the Rt side) and click on the **division** “**/**” button [10] on the calculator to enter it into the box, and then type **3**.
- Finally, click **OK** [11].
- In the **Variable view**, the computed MAP appears as the last variable, while the values are shown in the **Data view**.





I_To determine the prevalence of hypertension

Hypertension shall be defined using the World Health Organization and International Society of Hypertension criteria as systolic blood pressure (SBP) > 140 mmHg and/or diastolic blood pressure (DBP) > 90 mmHg or both or self-reported anti hypertensive medication during the past 1 week.

Determining the prevalence of hypertension among the respondents requires a combination of recoding and computation of variables. These shall be done in stages:

Stage 1: Transform both SBP and DBP into categories; i.e., from quantitative to qualitative (as **hypertensive** and **normal BP** based on the criteria stated) by recoding them into different variables on **Nominal measure**.

Stage 2: Recode use of anti hypertensive medication during the past 1 week into a different variable but still on **Nominal measure**.

Stage 3: Change the measures of the 3 recoded variables from **Nominal** to **Scale measure**.

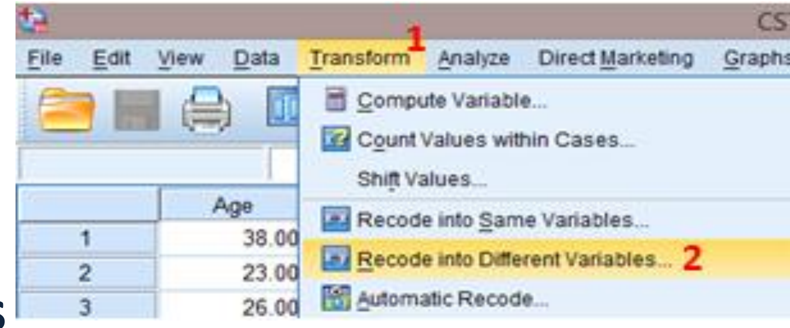
Stage 4: Compute the BP score by adding the scores on the 3 recoded variables (on **Scale measure**).

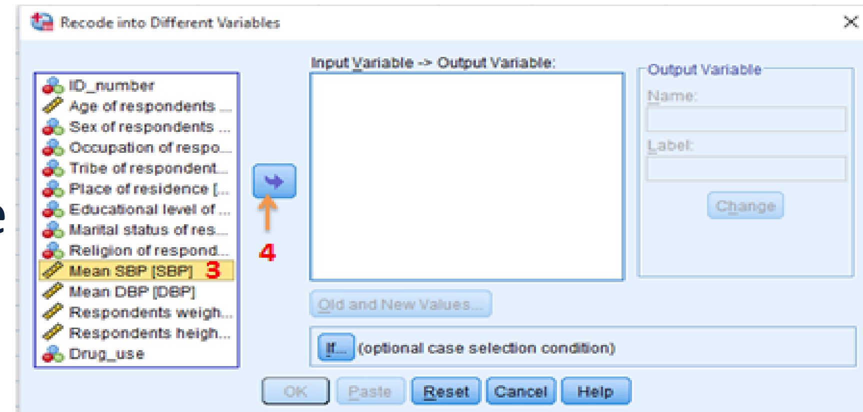
Stage 5: Transform the BP score into categories; i.e., from quantitative to qualitative (as hypertensive and Normal BP) by recoding it into a different variable on **Nominal measure**.

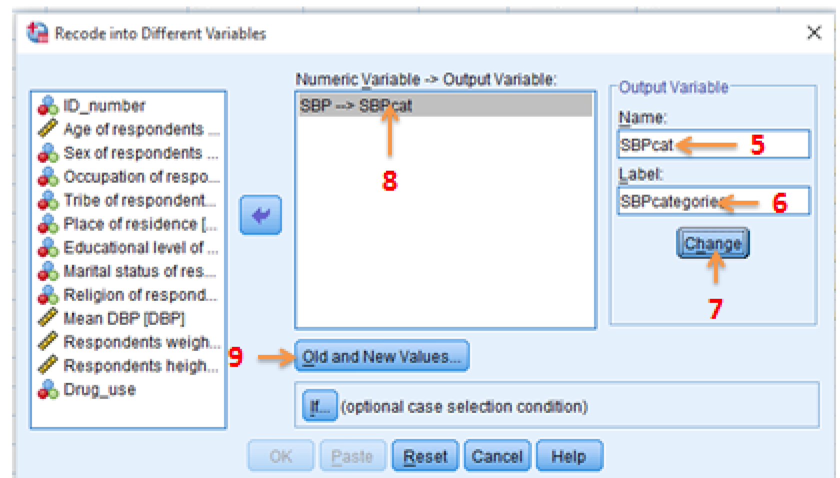
I_To determine the prevalence of hypertension contd.

Stage 1: Transform both SBP and DBP into categories

- Click **Transform** [1] > **Recode into Different Variables** [2].
- In the **Recode into Different Variables** box that appears, double click on the quantitative variable you want to transform (i.e., **SBP** [3]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [4] to move it to the **Variable(s)** box.
- Type the name of the new variable you want to create (i.e **SBPcat**) [5] in the **Name** box, type the label (i.e., **SBP categories**) [6] in the **Label** box, and click **Change** [7].
- The new variable's name appears in front of the old variables name [8] in the box.
- Click **Old and New Values** [9].



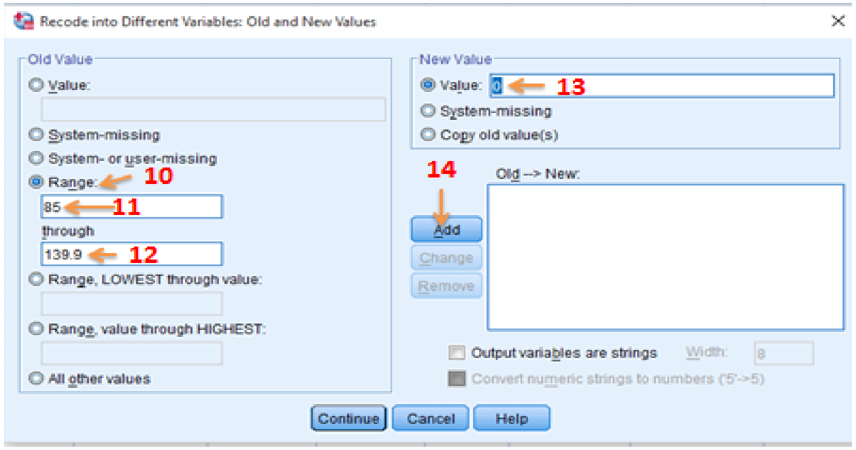


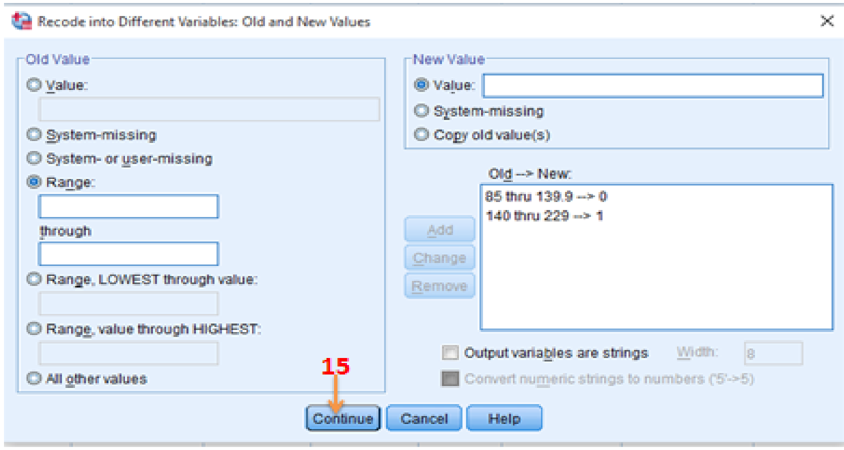


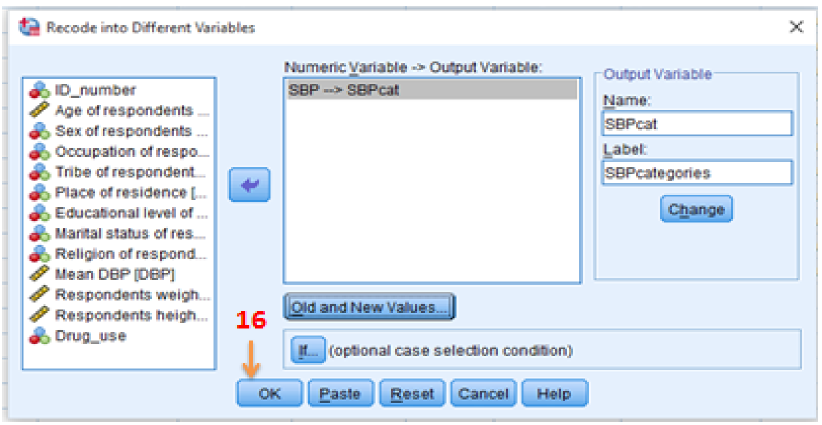
I_To determine the prevalence of hypertension contd.

Stage 1: Transform both SBP and DBP into categories contd.

- In the new box that appears, click on **Range** [10] to activate it.
- The decriptives done (**in B2a**) show that **SBP** ranged from **85-229mmHg**.
- Type **85** [11] and **139.9** [12] in the boxes shown (i.e using 140mmHg as the cut-off for hypertension as specified in the WHO/ISH criteria, type its value (i.e., **0)** [13] in the **New Value** box and click **Add** [14] to move the class into the **Old to New** Box.
- Repeat the procedure for the other class (and assign a value of **1** to class **140-229**).
- Click **Continue** [15].
- Finally, click **OK** [16] in the **Recode into Different Variables** box that re-appears.







I_To determine the prevalence of hypertension contd.

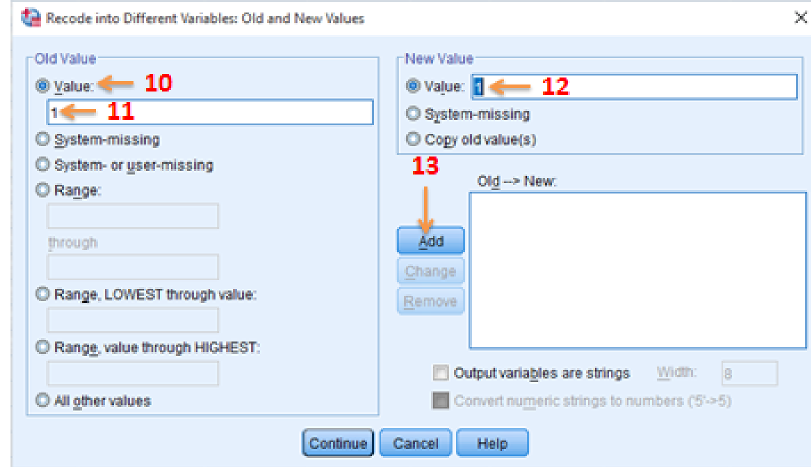
Stage 1: Transform both SBP and DBP into categories contd.

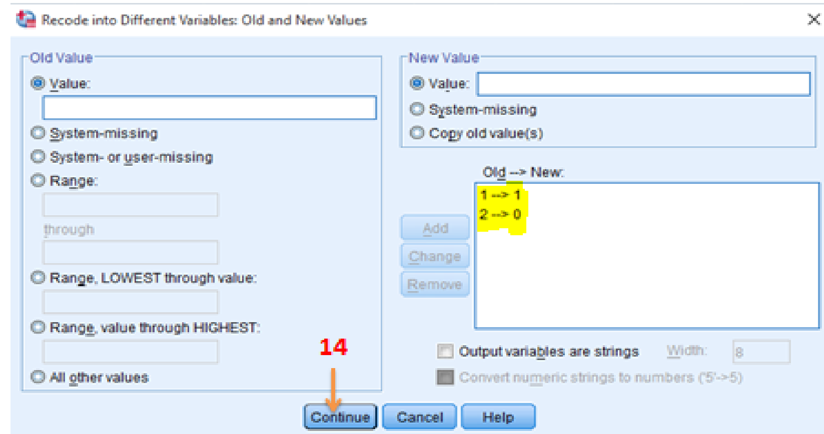
- Repeat the procedure for DBP (assign a value of **0** to **48.5-89.9mmHg**, and a value of **1** to **90-130mmHg**.

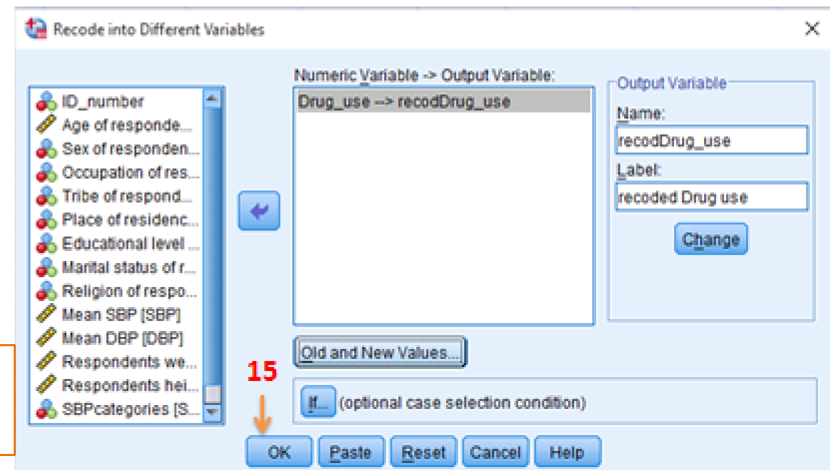
Stage 2: Recode use of anti hypertensive medication during the past 1 week into a different variable but still on Nominal measure.

- Repeat steps [1] to [9] for **Drug_use** (as was done for SBP and DBP) but with the new variable now named **recodDrug_use (**i.e., recoded Drug_use**)**
- In the new box that appears, click on **Value** [10] to activate it.
- Type **1** [11] in the **Old Value** box (being the value originally assigned to use of anti hypertensive drug in the past 1 week in the database) and **1** [12] in the **New Value** box and click **Add** [13] to move the class into the **Old to New** Box.
- Repeat the procedure for the other class (i.e., assign a value of **0** to class **2**).
- Click **Continue** [14].
- Finally, click **OK** [15] in the **Recode into Different Variables** box that re-appears.

Please note that all the 3 recoded variables (i.e., SBPcat, DBPcat and recodDrug_use are in Nominal Measures







I_To determine the prevalence of hypertension contd.

Stage 3: Change the measures of the 3 recoded variables from Nominal to Scale measure.

- Find the column named Measure [1].
- For each of the recoded variables, click on the drop down arrow next to Nominal [2], and select Scale [3] in the options that appear.

15	SBPcat	Numeric	8	2	SBPcategories	None	None	10	Right	Nominal
16	DBPcat	Numeric	8	2	DBP categories	None	None	10	Right	Nominal
17	recodDrug_...	Numeric	8	2	recoded Drug use	None	None	15	Right	Nominal

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure 1	Role
1	ID_number	Numeric	8	2		None	None	8	Right	Nominal	Input
2	Age	Numeric	8	0	Age of respond...	None	None	8	Right	Scale	Input
3	Sex	Numeric	8	0	Sex of respond...	{1, male}...	None	8	Right	Nominal	Input
4	Occupation	Numeric	8	0	Occupation of r...	{1, unemplo...	None	8	Right	Nominal	Input
5	Tribe	Numeric	8	0	Tribe of respon...	{1, hausa}...	None	8	Right	Nominal	Input
6	Residence	Numeric	8	0	Place of reside...	{1, urban}...	None	8	Right	Nominal	Input
7	Education	Numeric	8	0	Educational lev...	{1, none}...	None	8	Right	Nominal	Input
8	Marital_status	Numeric	8	0	Marital status o...	{1, single}...	None	8	Right	Nominal	Input
9	Religion	Numeric	8	0	Religion of resp...	{1, islam}...	None	8	Right	Nominal	Input
10	SBP	Numeric	8	2	Mean SBP	None	None	8	Right	Scale	Input
11	DBP	Numeric	8	2	Mean DBP	None	None	8	Right	Scale	Input
12	Weight	Numeric	8	2	Respondents w...	None	None	8	Right	Scale	Input
13	Height	Numeric	8	2	Respondents h...	None	None	8	Right	Scale	Input
14	Drug_use	Numeric	8	2		{1.00, Yes}...	None	8	Right	Nominal	Input
15	SBPcat	Numeric	8	2	SBPcategories	None	None	10	Right	Nominal	Input
16	DBPcat	Numeric	8	2	DBP categories	None	None	10	Right	Scale	Input
17	recodDrug_...	Numeric	8	2	recoded Drug use	None	None	15	Right	Ordinal	Input
18										Nominal	

- The recoded variables are now in scale measure and have become quantitative in nature.

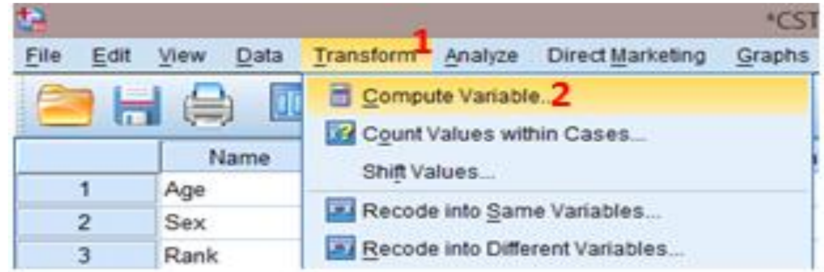
15	SBPcat	Numeric	8	2	SBPcategories	None	None	10	Right	Scale
16	DBPcat	Numeric	8	2	DBP categories	None	None	10	Right	Scale
17	recodDrug_...	Numeric	8	2	recoded Drug use	None	None	15	Right	Scale

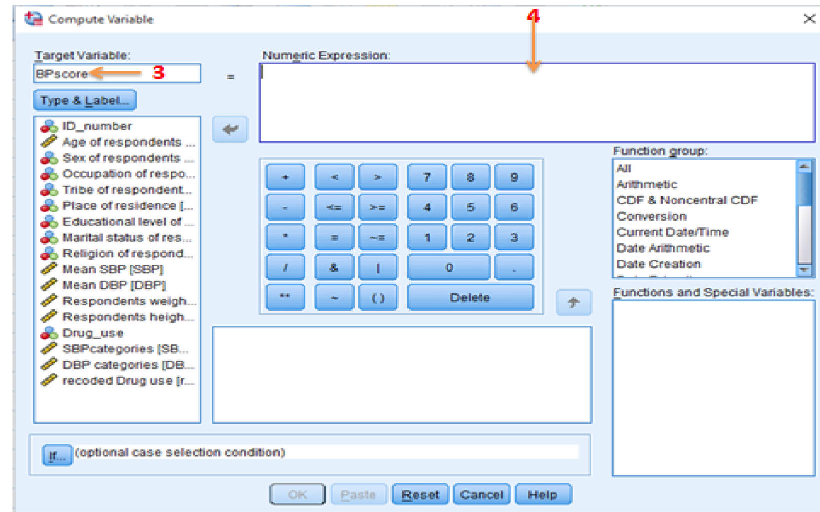
I_To determine the prevalence of hypertension contd.

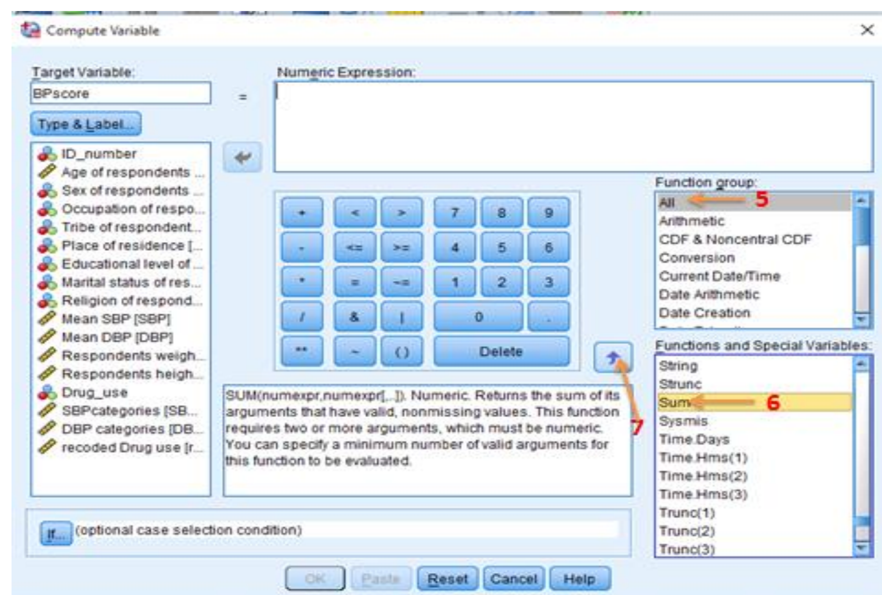
Stage 4: Compute the BP score by adding the scores on the 3 recoded variables (on Scale measure).

To compute the BP score

- Click **Transform** [1] > **Compute Variable** [2].
- In the **Compute Variable** box that appears, type **BPscore** [3] in the **Target Variable** box.
- Place the cursor in the **Numeric Expression** box [4].
- Go to the **Function group** box and select **All** [5].
- Go to the **Functions and Special Variables** box and use the scroll bar to find **Sum** [6], select it and click on the arrow [7] to enter it into the **Numeric Expression** box.





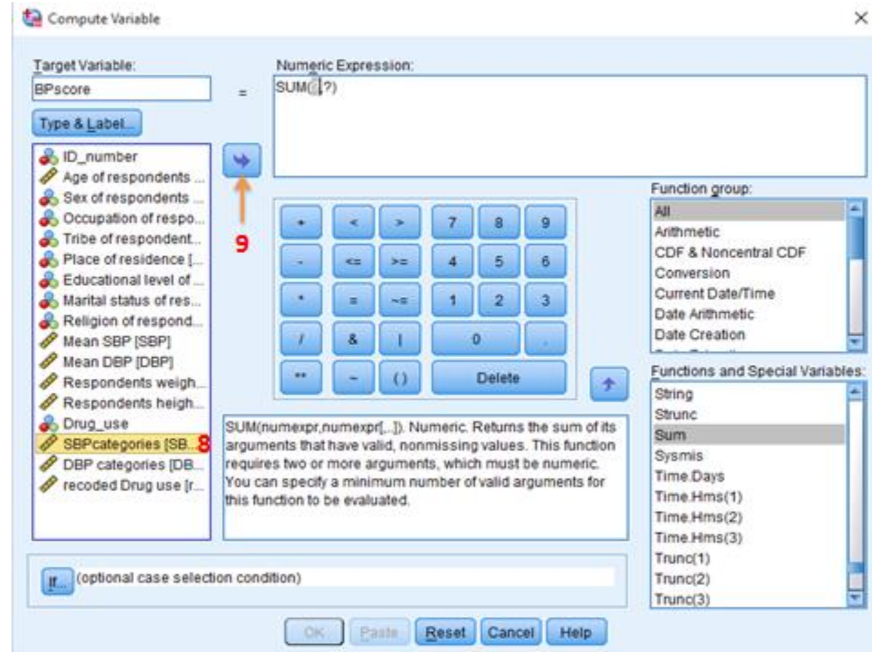


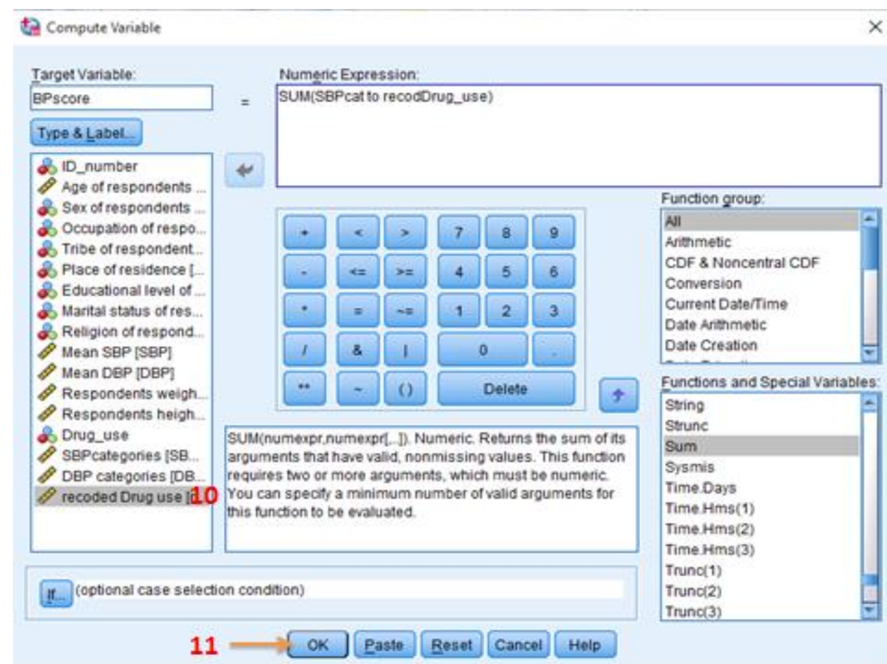
I_To determine the prevalence of hypertension contd.

Stage 4: Compute the BP score contd.

To compute the BP score contd.

- Click on the first of the 3 variables concerned (i.e., **SBP**)[8] and click on the arrow [9] to enter it inside the brackets (i.e., to replace the first question mark "**?**" inside the brackets).
- Press the space bar once, type **to**, and press the space bar once again.
- Click on the last of the 3 variables concerned (i.e., **recodDrug_use**)[10] and click on the arrow [9] to enter it inside the brackets (i.e., to replace the second question mark "**?**" inside the brackets).
- Delete the comma "**,**" and question mark "**?**" signs still remaining in the brackets and press **OK** [11].
- In the **Variable view**, the computed BPscore appears as the last variable, while the values are shown in the **Data view**.
- Go to the Variable view and change the measure of the newly computed **BPscore** from **Nominal** to **Scale**.



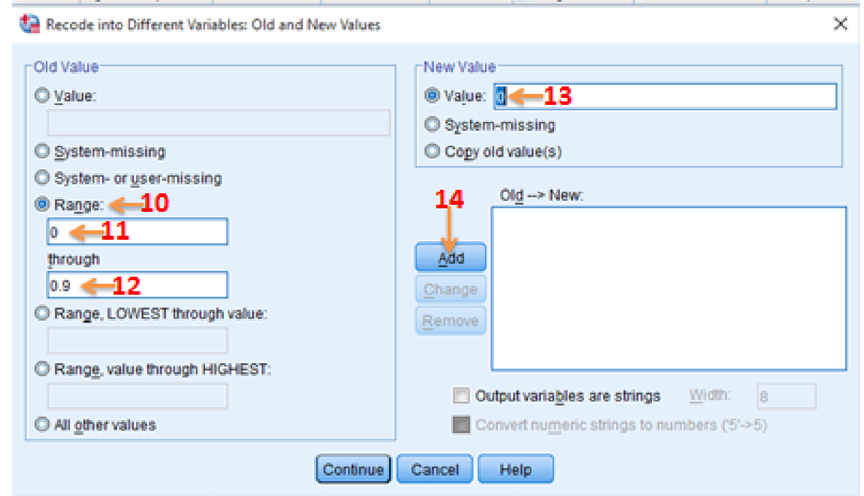


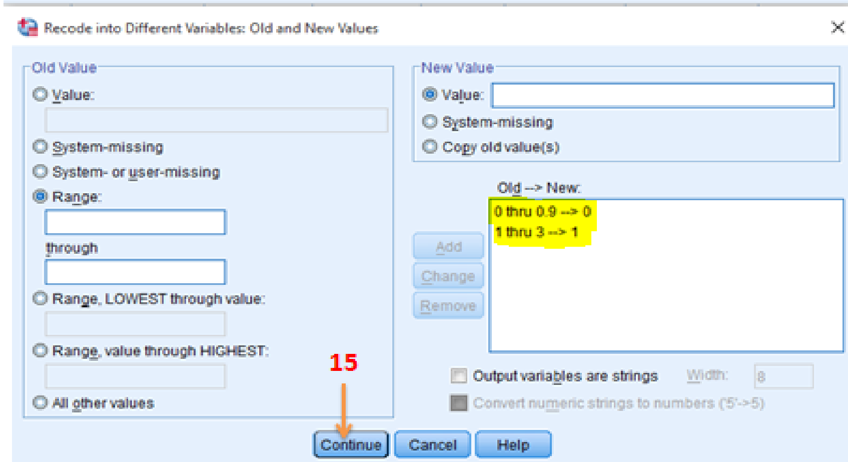
I_To determine the prevalence of hypertension contd.

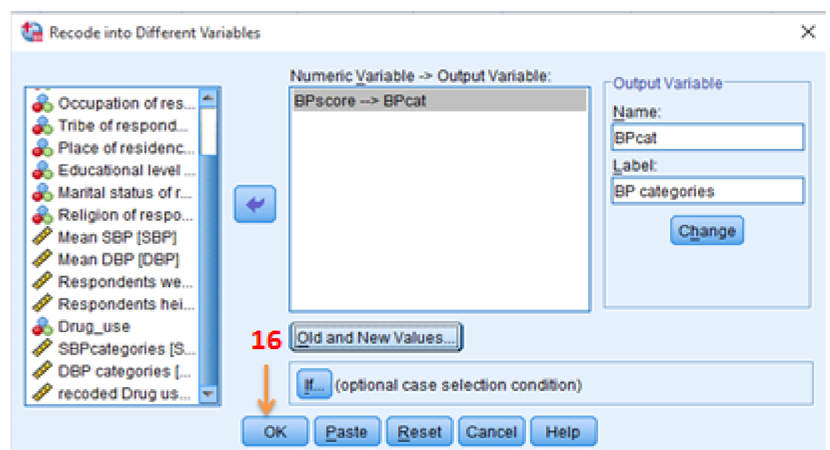
Stage 5: Transform the BP score into categories

To transform the BP score to categories

- Repeat steps [1] to [9] for **BPscore** (as was done for SBP and DBP) but with the new variable now named **BPcat** (i.e., BP categories)
- In the new box that appears, click on **Range** [10] to activate it.
- Type **0** [11] and **0.9** [12] in the boxes shown, type its value (i.e., **0**) [13] in the **New Value** box and click **Add** [14] to move the class into the **Old to New** Box.
- Repeat the procedure for the other class (i.e., assign **1** to class **1-3**), and click **Continue** [15].
- Finally, click **OK** [16] in the **Recode into Different Variables** box that re-appears.

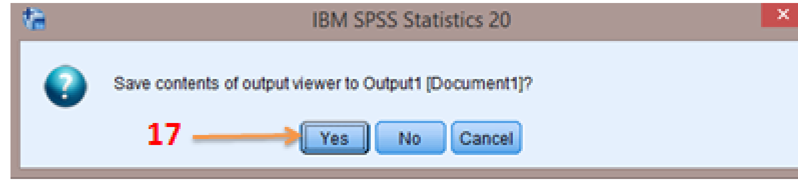


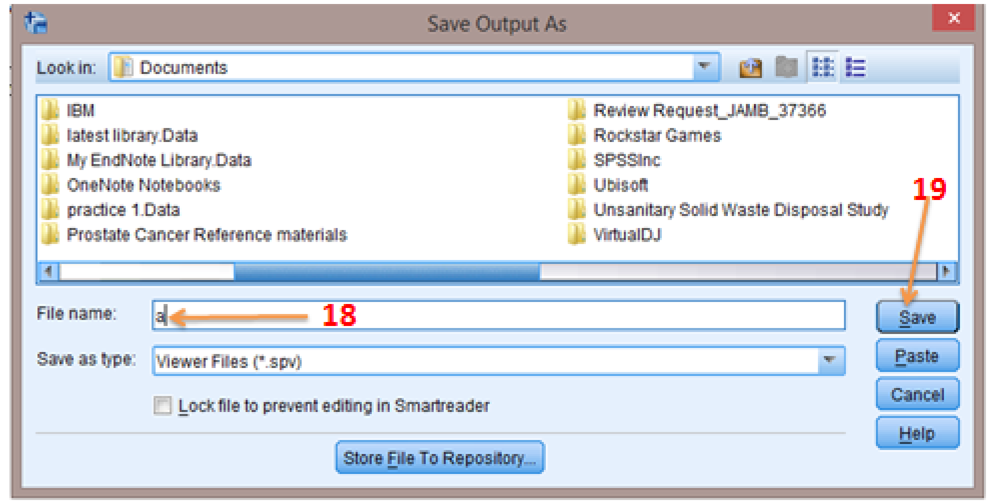




I_To determine the prevalence of hypertension contd.

- Close the viewer window that appears indicating that the task has been carried out, always select **YES** [17] in all the queries that appear.
- Also, in the **Save Output As** dialog box that appears, type "**a**" [18] in the **File name** box, and click **Save** [19]. If a dialog box with replace existing "**a**" appears, click **YES**.
- Go to the **variable view** and scroll down to locate the recoded variable (i.e., **BPcat**), which now appears as the last variable.
- Click in the cell under **Values** and click on the small box with 3 dots at the right end of the cell [20].



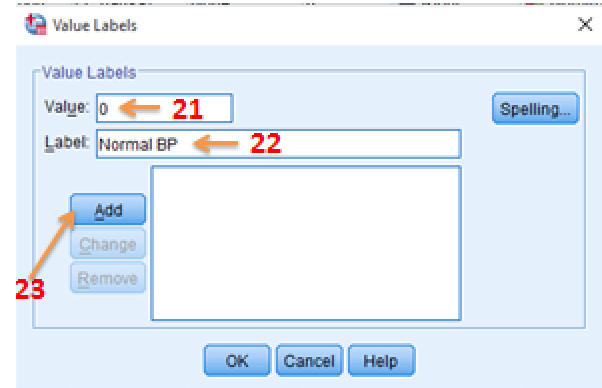


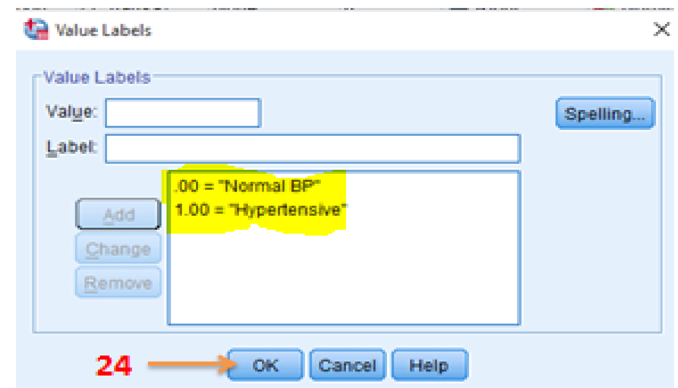
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID_number	Numeric	8	2		None	None	8	Right	Nominal
2	Age	Numeric	8	0	Age of respond...	None	None	8	Right	Scale
3	Sex	Numeric	8	0	Sex of respond...	{1, male}...	None	8	Right	Nominal
4	Occupation	Numeric	8	0	Occupation of r...	{1, unemplo...	None	8	Right	Nominal
5	Tribe	Numeric	8	0	Tribe of respon...	{1, hausa}...	None	8	Right	Nominal
6	Residence	Numeric	8	0	Place of reside...	{1, urban}...	None	8	Right	Nominal
7	Education	Numeric	8	0	Educational lev...	{1, none}...	None	8	Right	Nominal
8	Marital_status	Numeric	8	0	Marital status o...	{1, single}...	None	8	Right	Nominal
9	Religion	Numeric	8	0	Religion of resp...	{1, islam}...	None	8	Right	Nominal
10	SBP	Numeric	8	2	Mean SBP	None	None	8	Right	Scale
11	DBP	Numeric	8	2	Mean DBP	None	None	8	Right	Scale
12	Weight	Numeric	8	2	Respondents w...	None	None	8	Right	Scale
13	Height	Numeric	8	2	Respondents h...	None	None	8	Right	Scale
14	Drug_use	Numeric	8	2		{1.00, Yes}...	None	8	Right	Nominal
15	SBPcat	Numeric	8	2	SBPcategories	None	None	10	Right	Scale
16	DBPcat	Numeric	8	2	DBP categories	None	None	10	Right	Scale
17	recodDrug_...	Numeric	8	2	recoded Drug use	None	None	15	Right	Scale
18	BPscore	Numeric	8	2		None	None	10	Right	Scale
19	BPcat	Numeric	8	2	BP categories	None	None	10	Right	Nominal
20										
21										
22										

20

I_To determine the prevalence of hypertension contd.

- In the **Value label** dialog box that appears, type **0** in the **Value** box [21], type **Normal BP** in the **Label** box [22], and click **Add** [23] to move them into the box.
- Repeat the procedure for the other value [i.e., label **1** as **Hypertensive**, and then click **OK** [24].
- The values are displayed in the cell under the **Values** column [25], and the measure is **Nominal** [26].





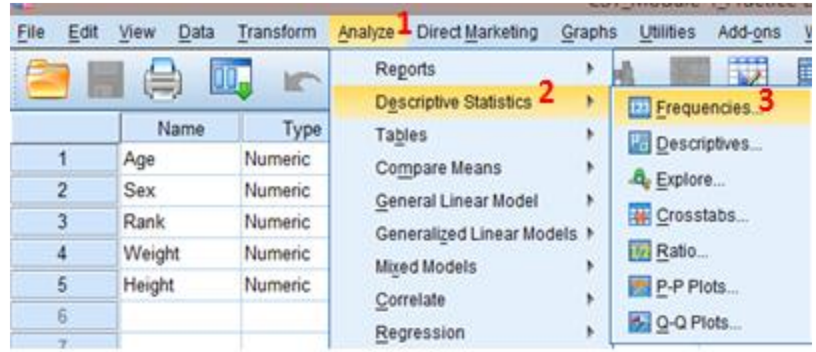
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID_number	Numeric	8	2		None	None	8	Right	Nominal
2	Age	Numeric	8	0	Age of respond...	None	None	8	Right	Scale
3	Sex	Numeric	8	0	Sex of respond...	{1, male}...	None	8	Right	Nominal
4	Occupation	Numeric	8	0	Occupation of r...	{1, unemplo...	None	8	Right	Nominal
5	Tribe	Numeric	8	0	Tribe of respon...	{1, hausa}...	None	8	Right	Nominal
6	Residence	Numeric	8	0	Place of reside...	{1, urban}...	None	8	Right	Nominal
7	Education	Numeric	8	0	Educational lev...	{1, none}...	None	8	Right	Nominal
8	Marital_status	Numeric	8	0	Marital status o...	{1, single}...	None	8	Right	Nominal
9	Religion	Numeric	8	0	Religion of resp...	{1, islam}...	None	8	Right	Nominal
10	SBP	Numeric	8	2	Mean SBP	None	None	8	Right	Scale
11	DBP	Numeric	8	2	Mean DBP	None	None	8	Right	Scale
12	Weight	Numeric	8	2	Respondents w...	None	None	8	Right	Scale
13	Height	Numeric	8	2	Respondents h...	None	None	8	Right	Scale
14	Drug_use	Numeric	8	2		{1.00, Yes}...	None	8	Right	Nominal
15	SBPcat	Numeric	8	2	SBPcategories	None	None	10	Right	Scale
16	DBPcat	Numeric	8	2	DBP categories	None	None	10	Right	Scale
17	recodDrug_...	Numeric	8	2	recoded Drug use	None	None	15	Right	Scale
18	BPscore	Numeric	8	2		None	None	10	Right	Scale
19	BPcat	Numeric	8	2	BP categories	mal BP}... 25	None	10	Right	Nominal 26

I_To determine the prevalence of hypertension contd.

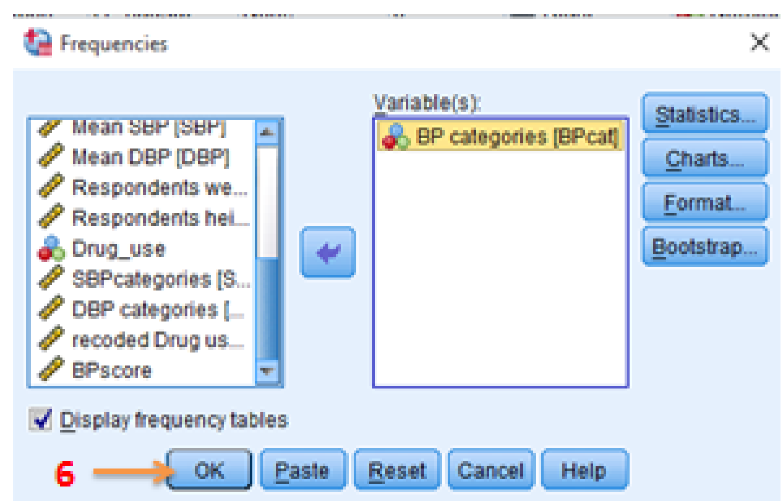
The next thing to do is to run frequencies on the BP categories.

To run frequencies on the BP categories:

- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Frequencies** [3].



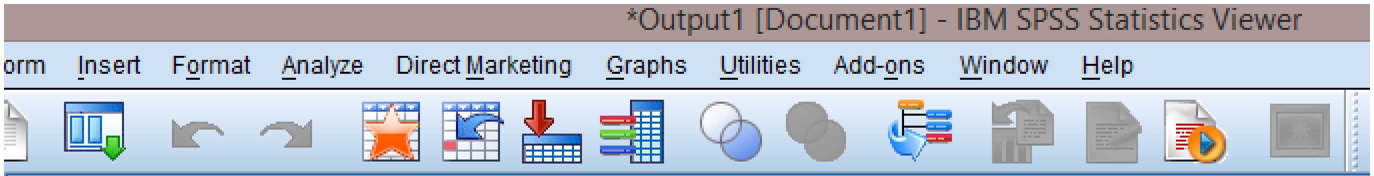
- Double click on the qualitative variable you want to analyze (i.e., **BP categories** [4]) to move it to the **Variable(s)** box, or click on it once and then click on the arrow between the boxes [5] to move it to the **Variable(s)** box.



- Finally, click **OK** [6].

I_To determine the prevalence of hypertension contd.

- The results of the analysis are displayed in a table on the **viewer** window.



Frequencies

Statistics

BP categories

N	Valid	300
	Missing	0

BP categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Normal BP	167	55.7	55.7	55.7
	Hypertensive	133	44.3	44.3	100.0
	Total	300	100.0	100.0	

The output shows that:
There were no missing values, this is why the Percent (based on the sample size), is the same as the Valid Percent (based on the values entered).

In the dissertation the results for BP categories can be presented as:
Close to half 133 (44.3%) of the 300 respondents were hypertensive (i.e., the prevalence of hypertension was 44.3%).

TEST OF NORMALITY

Test of Normality

Normal distribution curve

- If we take a large sample of men or women, measure their heights, and plot them on a frequency distribution, the distribution will almost certainly obtain a symmetrical bell-shaped pattern.

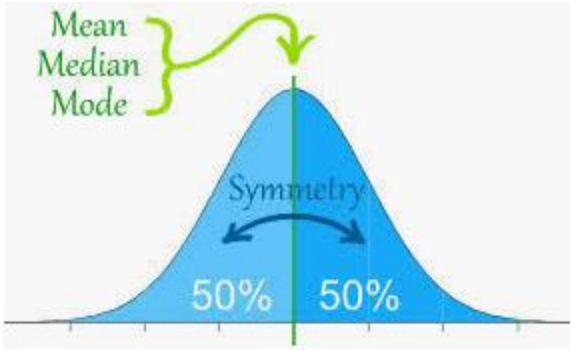
- This is known as the ***normal distribution curve*** *(also called the* ***Gaussian*** **distribution**).

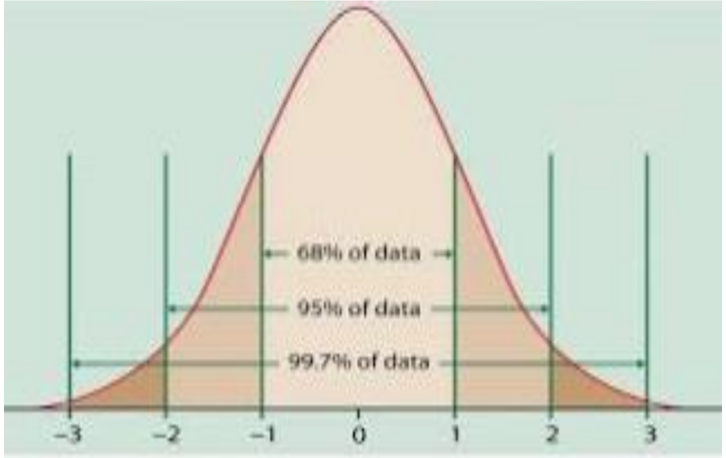
- The least frequently recorded heights lie at the two extremes of the curve.

Test of Normality contd.

Properties of a normal distribution curve

1. The normal distribution curve is bell-shaped.
2. It is symmetrical about the mean.
3. The curve on either side of the mean is a mirror image of the other.
4. The mean, median and mode are equal and located at the center of the distribution.
5. The curve is unimodal (single mode).
6. The curve is continuous.
7. The curve never touches the x-axis.
8. The total area under the normal distribution curve is equal to 1.





9. Area corresponding to 1SD will comprise 68.27% of the total area, 2SD will comprise 95.45% of the total area and 3SD will comprise 99.73% of the total area. (i.e., 68- 95- 99.7 rule).

Test of Normality contd.

- Performing test of normality enables one to know the appropriate measure of central tendency to be used for a dataset. For example, while the mean is appropriate for a normally distributed data, the median is appropriate for a data that is not normally distributed.

- Also, in hypothesis testing, while parametric tests are appropriate for normally distributed data, non-parametric tests are appropriate for data that are not normally distributed.

The methods of testing for normality include:

1. **Graphical assessment of normality**: The graphical representation of a normal distribution as a frequency curve gives a bell shaped pattern.

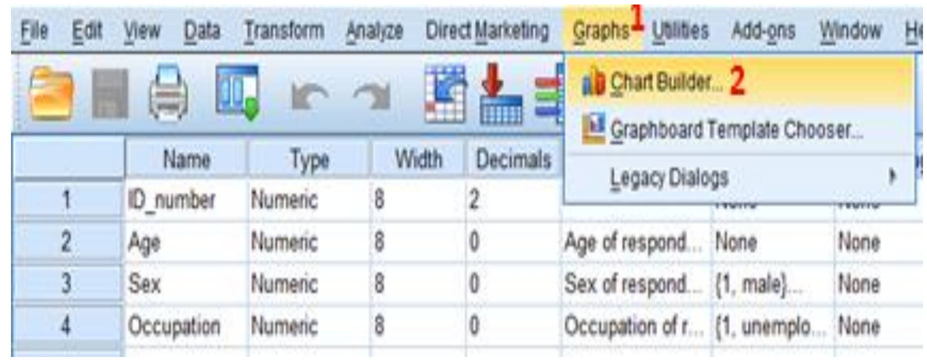
2. **Goodness of fit testing of normality**: **Kolmogorov-Smironov** goodness of fit procedures with appropriate modifications can be used to test the hypothesis of normality in the population distribution. **Shapiro and Wilk method** is another test of normality.

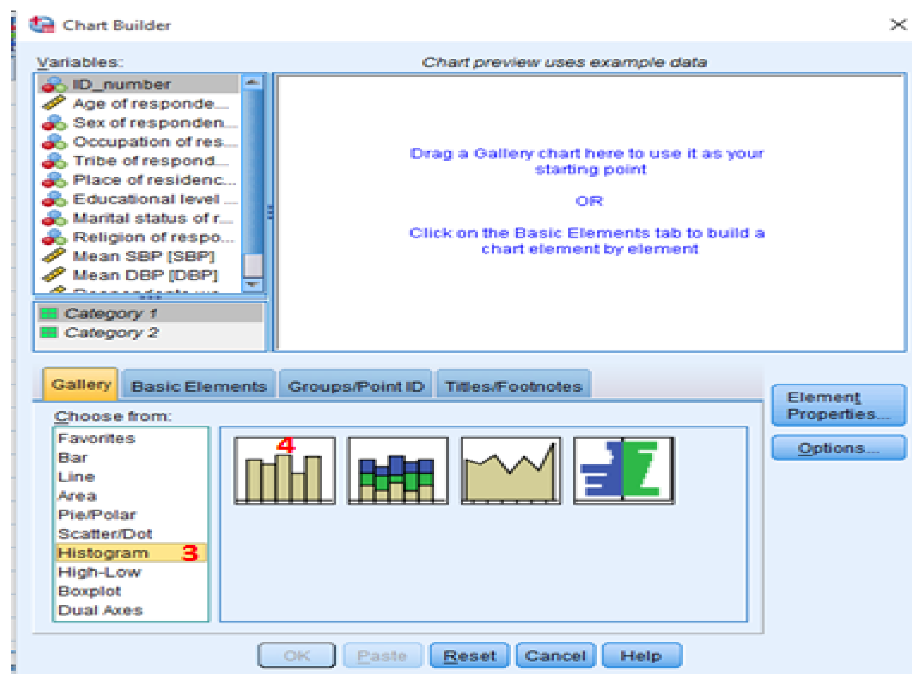
Graphical assessment of normality

- Normality of a dataset for a particular variable can be assessed graphically by the nature of the frequency distribution curve of its histogram.

To create a histogram:

- Open the database.
- Click **Graphs** [1] > **Chart Builder** [2].
- In the **Chart Builder** box that appears, find the **Choose from** box and click **Histogram** [3].
- In the Histogram designs that appear click on the design you want [4].

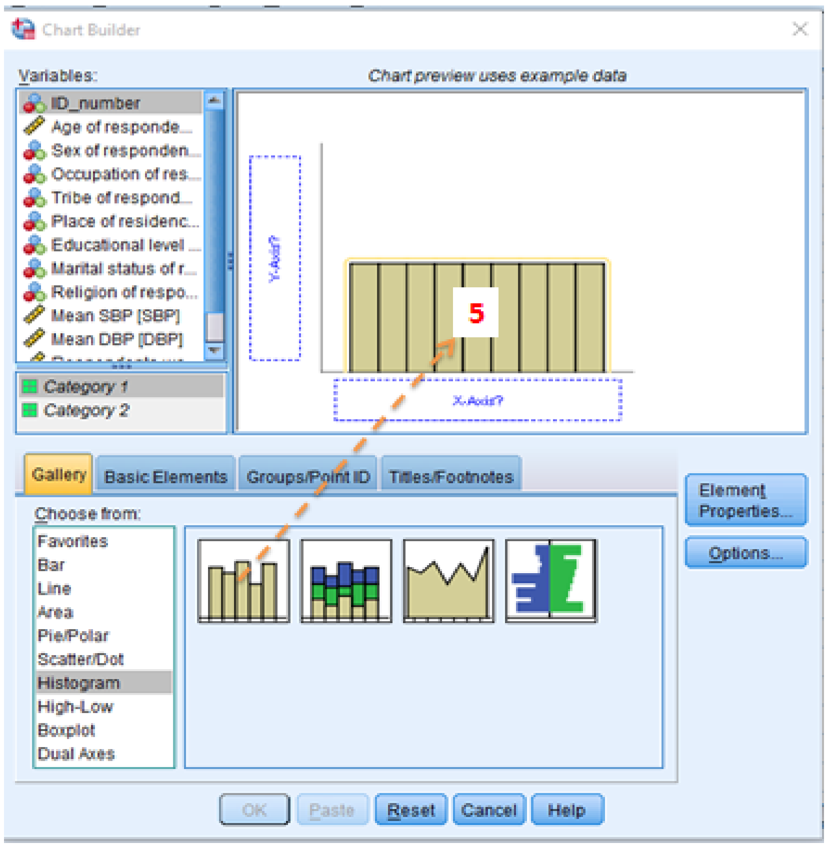


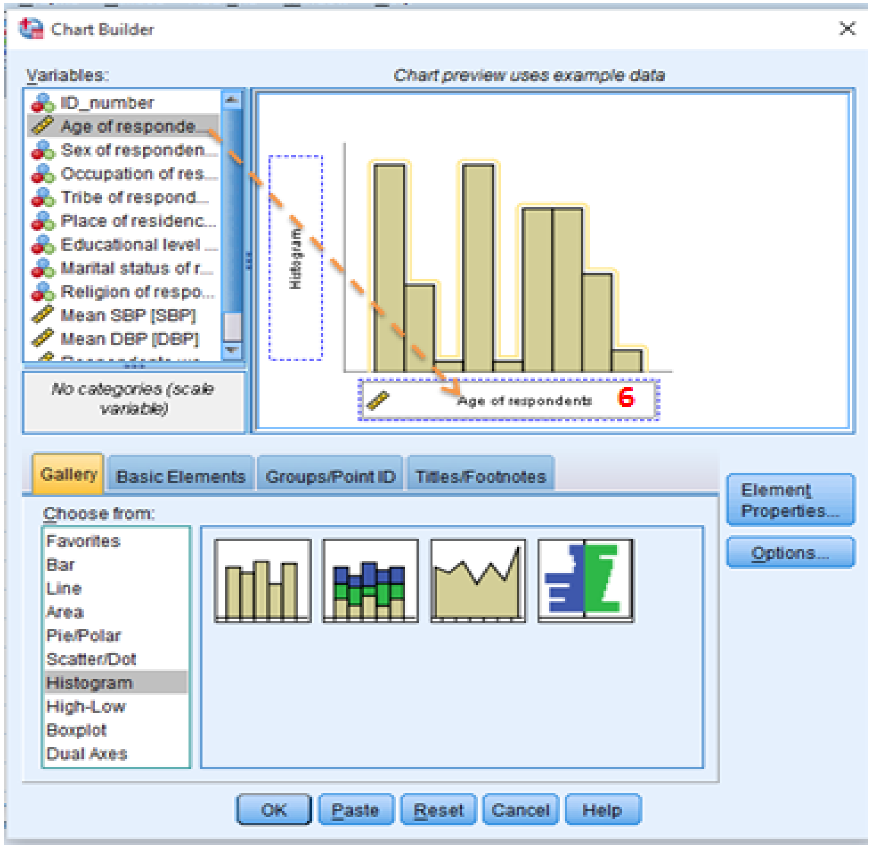


Graphical assessment of normality contd.

To create a histogram contd.:

- Hold down the Left button of the cursor and drag the selected histogram design into the preview area [5].
- From the **Variables** list drag **Age of respondents** into the X axis [6].

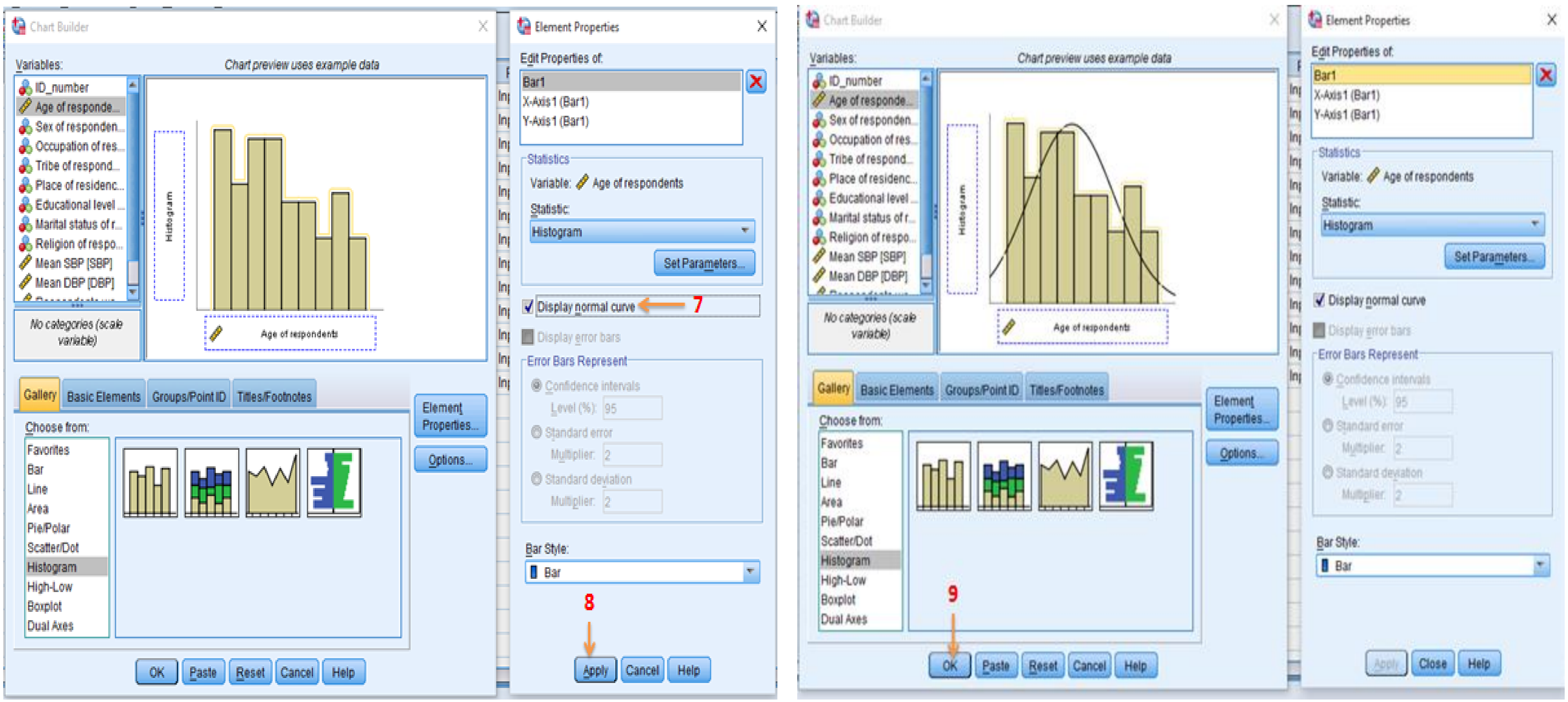




Graphical assessment of normality contd.

To create a histogram contd.:

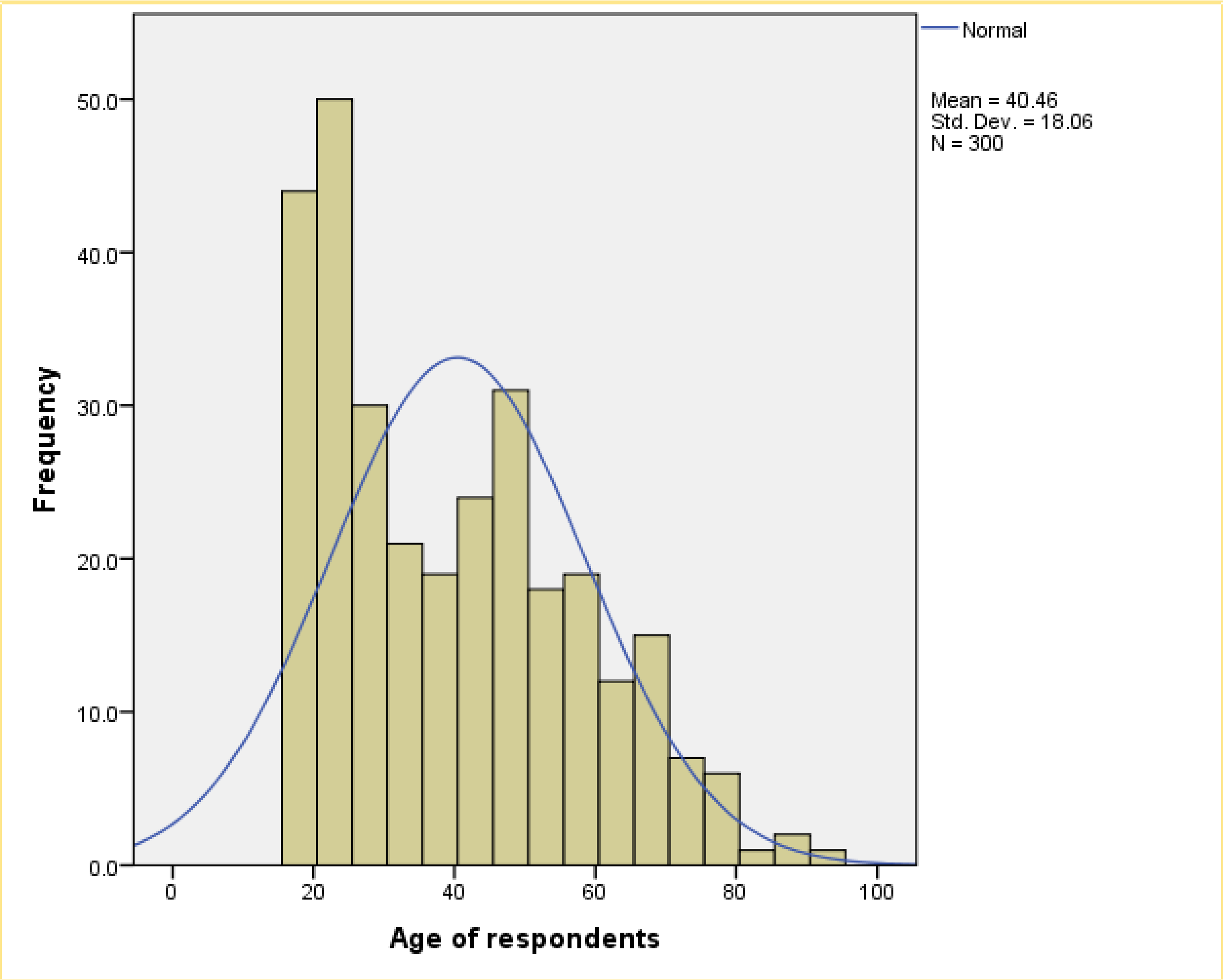
- Click on **Display normal curve** [7], in the **Element properties** box, and then click on **Apply** [8].
- Finally, click **OK** [9].



Graphical assessment of normality contd.

To create a histogram contd.:

- The chart (histogram with the normal curve) is displayed on the **viewer** window.



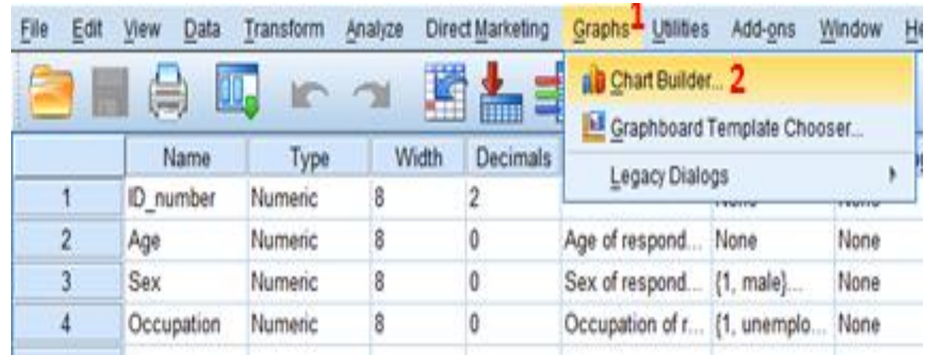
- The distribution curve displayed shows that the dataset for the variable (**Age**) is **not normally distributed** because **the curve on the Right side of the mean is not a mirror image of the one on the Left** (whereas the curve crosses the Y-axis on the Left, it rests on the X-axis on the Right.

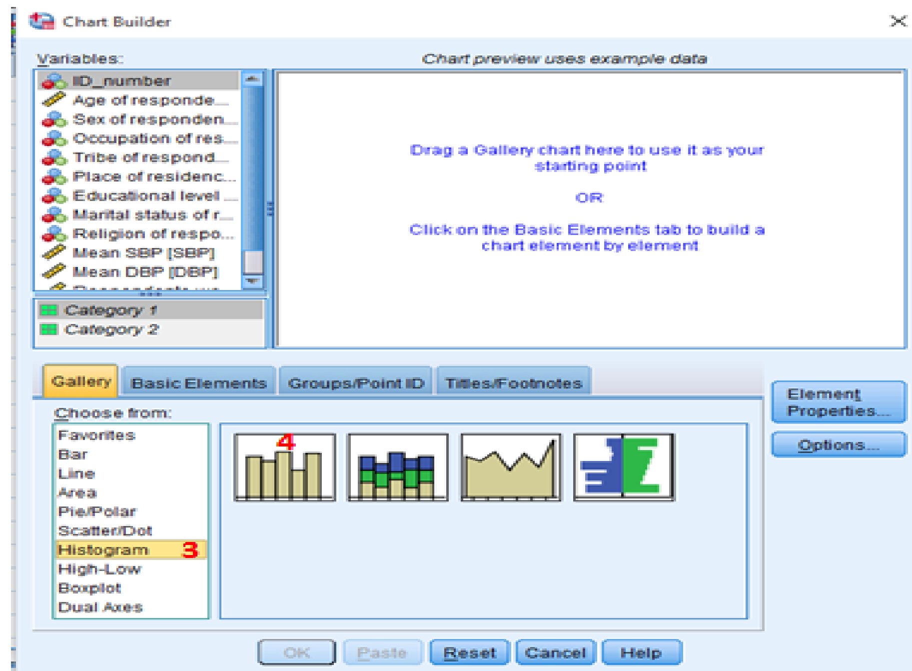
- Since the variable (**Age**) is not normally distributed, the appropriate measures of **central tendency** and **dispersion** for it therefore are the **median** and **inter-quartile range** (**IQR**) respectively.

Graphical assessment of normality contd.

To create separate histograms for males and females:

- Open the database.
- Click **Graphs** [1] > **Chart Builder** [2].
- In the **Chart Builder** box that appears, find the **Choose from** box and click **Histogram** [3].
- In the Histogram designs that appear click on the design you want [4].

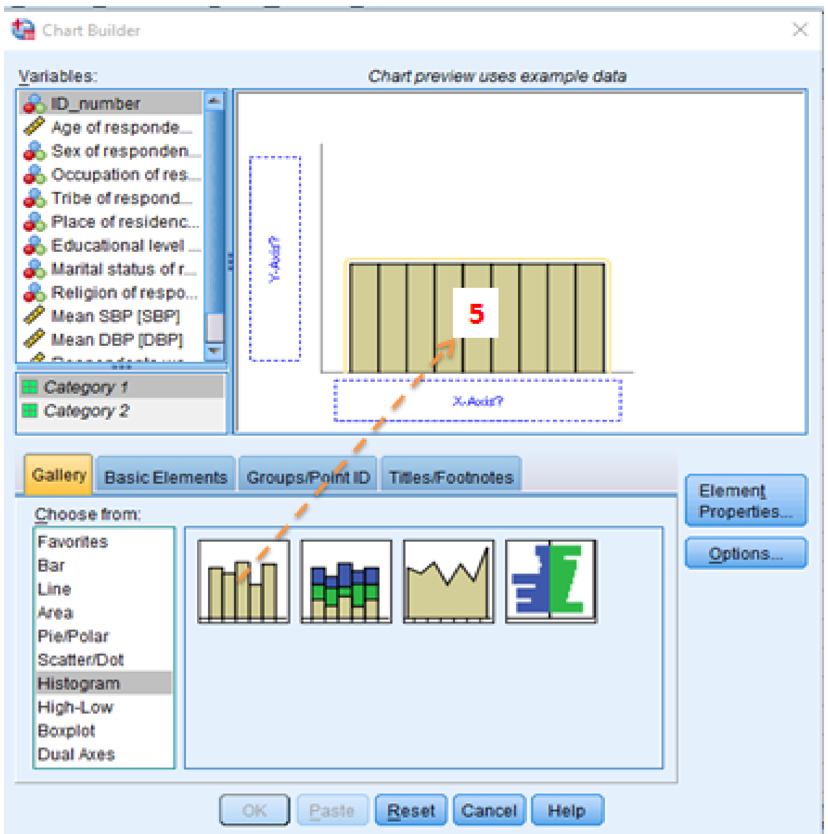


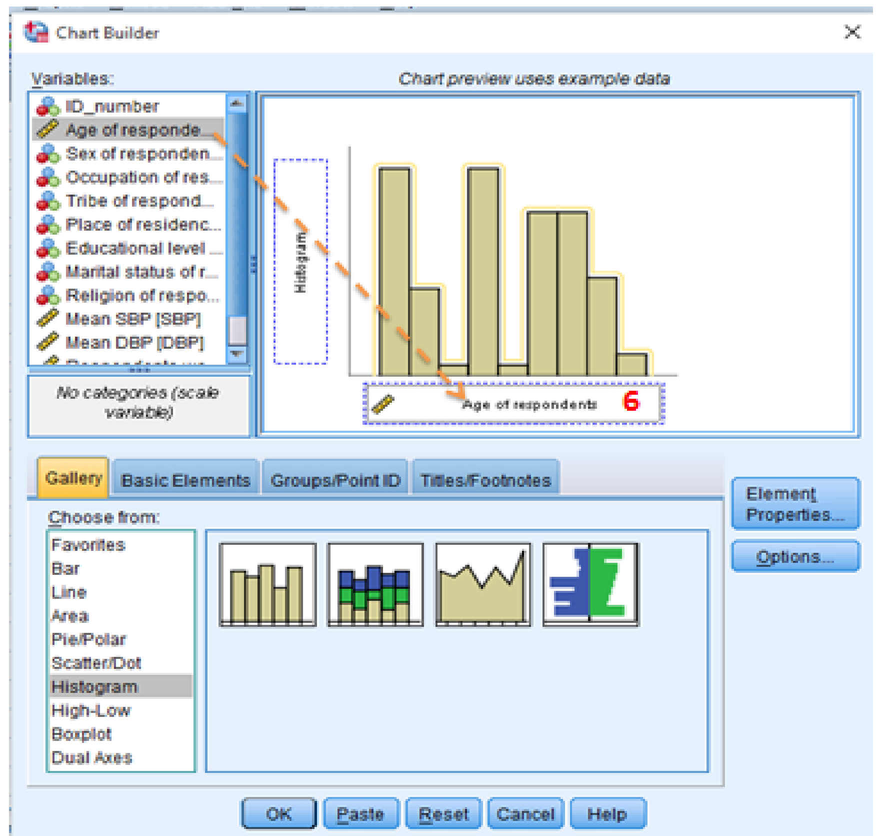


Graphical assessment of normality contd.

To create separate histograms for males and females contd.:

- Hold down the Left button of the cursor and drag the selected histogram design into the preview area [5].
- From the **Variables** list drag **Age** into the **X axis** [6].

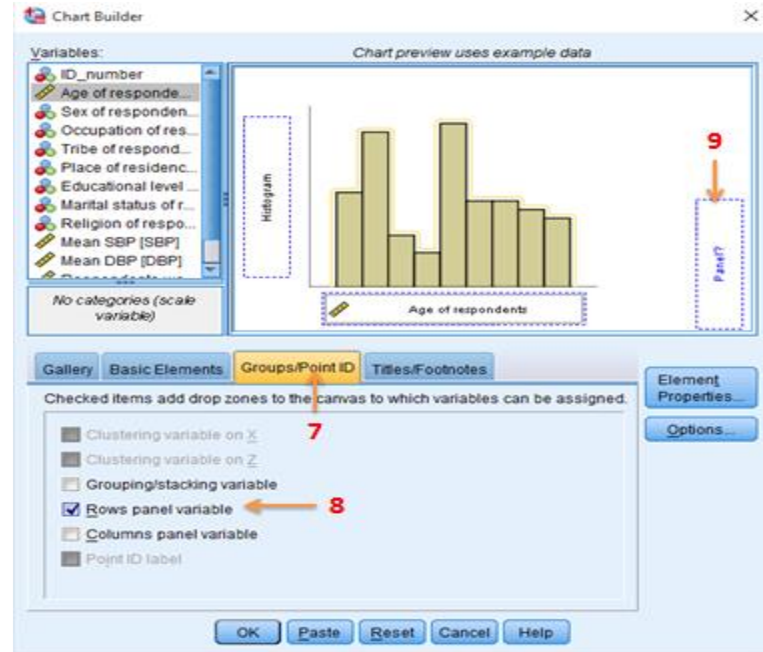


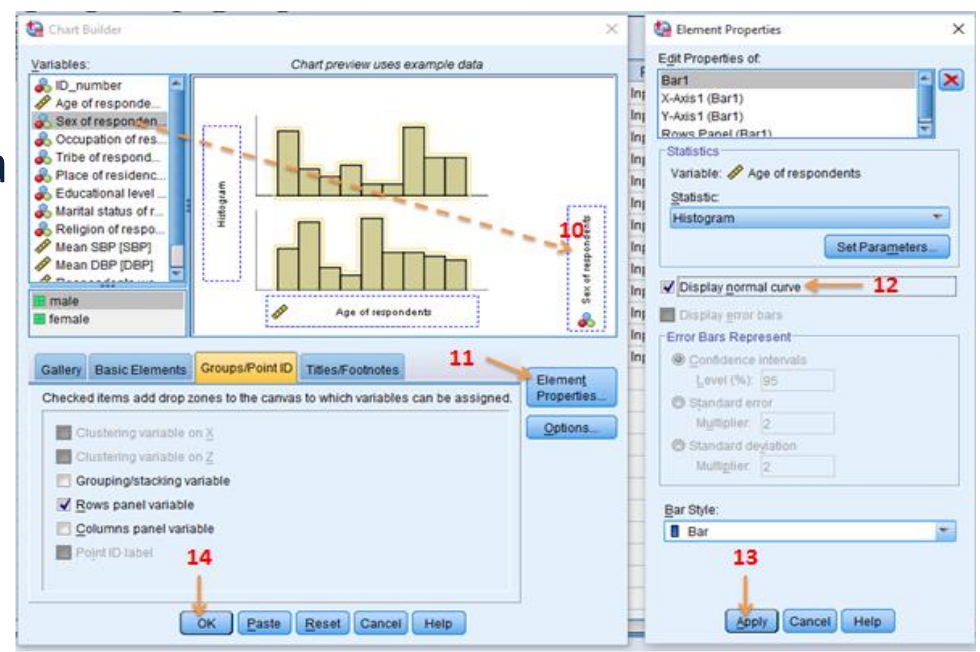


Graphical assessment of normality contd.

To create separate histograms for males and females contd.:

- Click **Groups/Point ID** [7] and put a tick in the **Rows panel variable** checkbox [8] (this will produce a **Panel?** box to the right of the preview area) [9].
- Drag the variable **Sex** into the panel box [10].
- Click on **Element Properties** (to the right of the screen) [11] and place a tick in the box next to **Display normal curve** [12], and then click **Apply** [13].
- Finally, click **OK** [14]

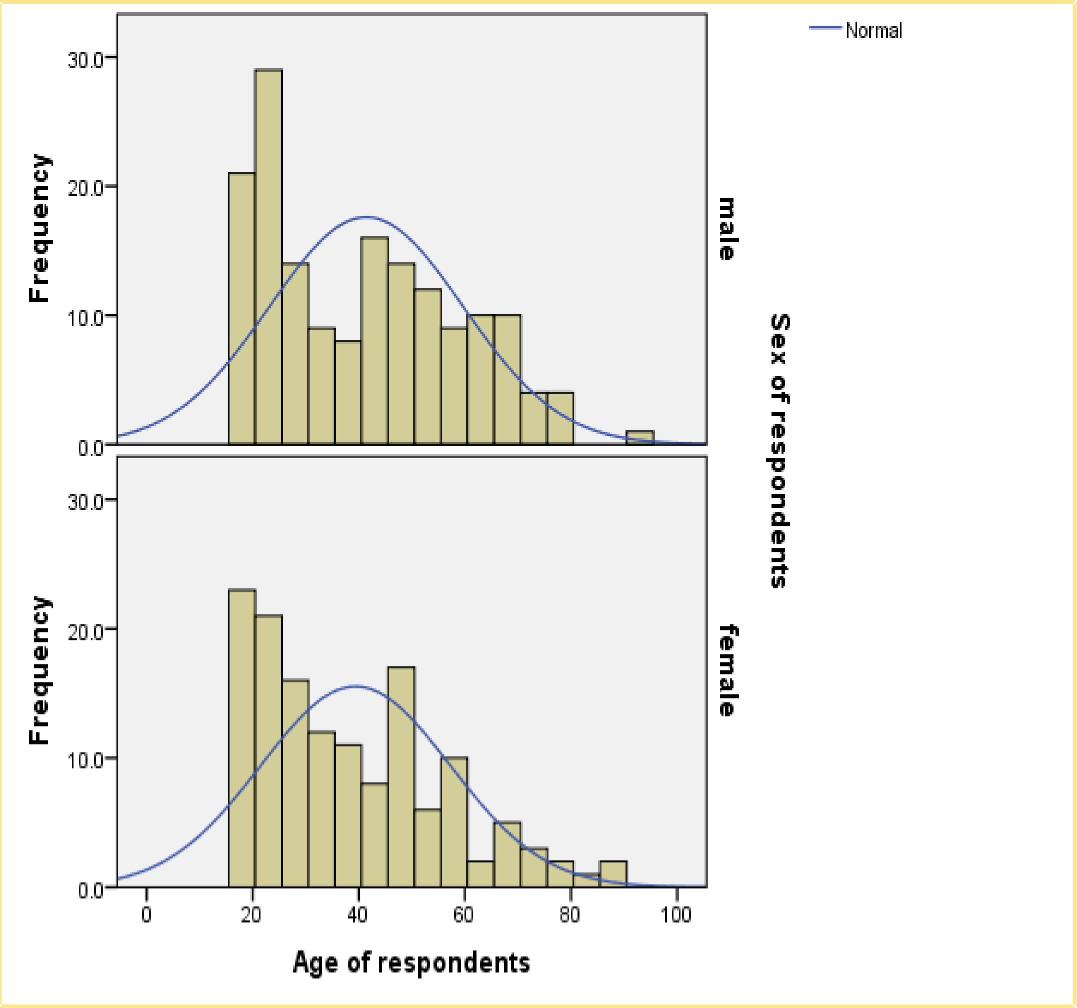




Graphical assessment of normality contd.

To create separate histograms for males and females contd.:

- The separate histograms for males and females are displayed on the **viewer** window.



•The distribution curves displayed for both males and females show that the dataset for the variable (**Age**) is **not normally distributed** in both groups.

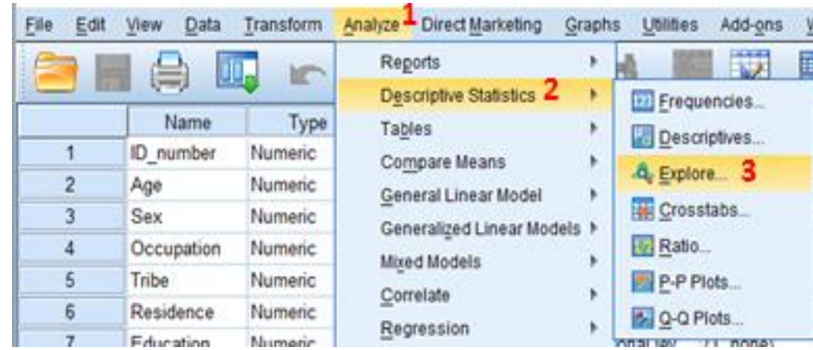
•In situations where the distribution curve appears to be normally distributed, you may go ahead and use the mean and standard deviation as measures of central tendency as dispersion respectively; and also use parametric tests for hypothesis testing.

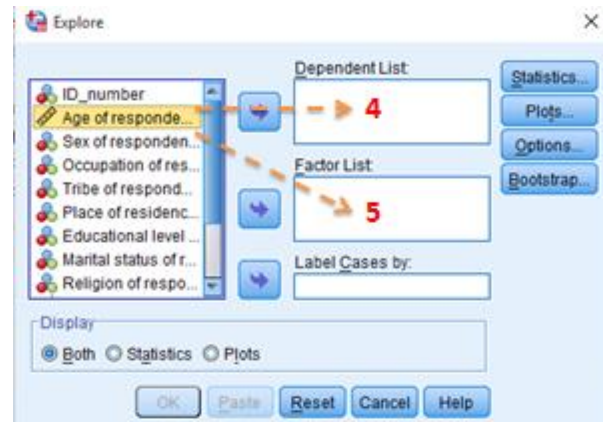
•Alternatively you may confirm if the dataset is actually normally distributed or not by performing Goodness of fit testing of normality (i.e., **Kolmogorov-Smirnov** and **Shapiro- Wilk** tests).

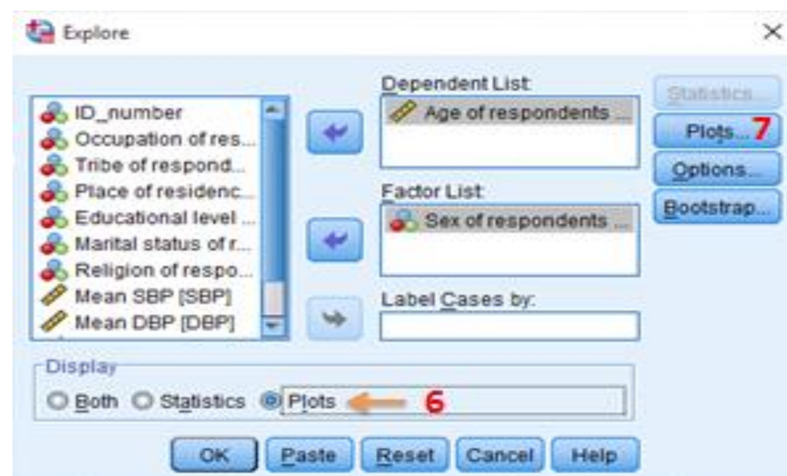
Goodness of fit testing of normality

To perform goodness of fit test:

- Click **Analyze** [1] > **Descriptive Statistics** [2] > **Explore** [3].
- Move **Age** to the **Dependent list** box [4] and **Sex** to the **Factor list** box [5].
- Under **Display**, ensure that there is only a tick next to **Plots** [6].
- Click on the **Plots** button [7] to open the **Plots** dialogue box.



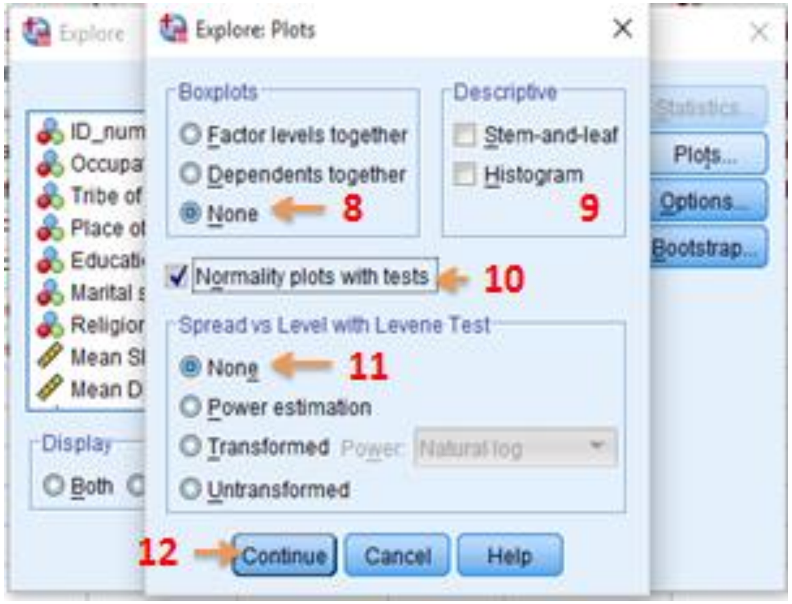


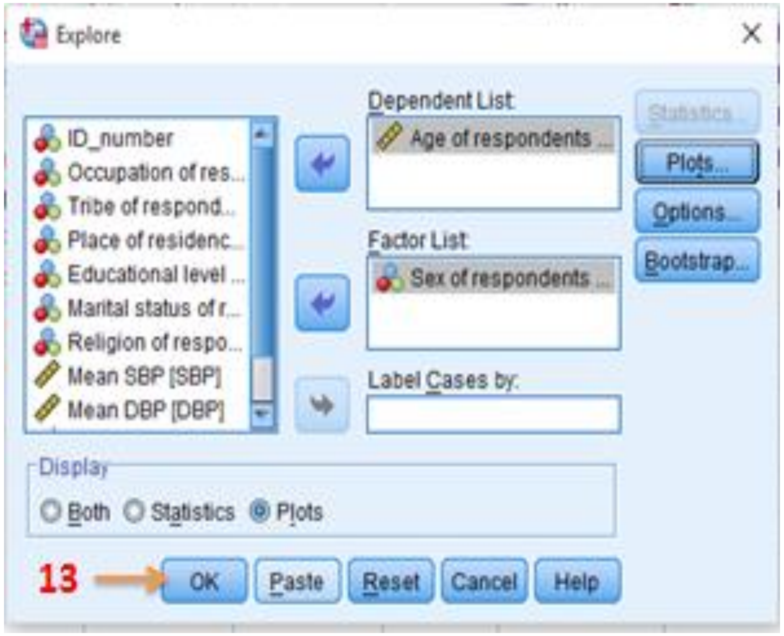


Goodness of fit testing of normality contd.

To perform goodness of fit test contd.:

- Under **Boxplots** click ◦**None** [8] and remove any ticks under **Descriptive** [9].
- Place a tick in **Normality plots with tests** [10].
- Under **Spread vs Level with Levene Test** tick ◦**None** [11].
- Click **Continue** [12].
- Finally, click **OK** [13].





Goodness of fit testing of normality contd.

To perform goodness of fit test contd.:

- The Case Processing Summary and the Tests of Normality results are displayed on the **viewer** window.

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	Sex of respondents	N	Percent	N	Percent	N	Percent
Age of respondents	male	161	100.0%	0	0.0%	161	100.0%
	female	139	100.0%	0	0.0%	139	100.0%

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Sex of respondents	Statistic	df	Sig.	Statistic	df	Sig.
Age of respondents	male	.131	161	.000	.928	161	.000
	female	.132	139	.000	.920	139	.000

a. Lilliefors Significance Correction

- The **Case Processing Summary** table shows that data were analyzed for 161 males and 139 females and there were no missing values.

- The **Tests of Normality** table shows that both the **Kolmogorov-Smirnov** and **Shapiro-Wilk** tests gave statistically significant difference ($p < 0.001$) between the distribution curves for both males and females and a normal distribution; thus confirming that the dataset for age in both males and females were not normally distributed.

C-International Research Consultancy

Our Services

Design or Review of Protocol / Proposal

We provide technical support in the design of study protocol / proposal. We provide guidance on formulation of research topic, specific objectives, research questions, and research hypothesis. We provide technical support regarding choice of appropriate study design, sample size estimation and sampling technique; as well as choice of appropriate method and instrument of data collection, and data management. We also review protocol / proposal and provide guidance on how to improve the quality in compliance with the guidelines of the institution concerned.

Development of Data Collection Instrument

We provide technical support in the development and validation of data collection instrument. We build questionnaires into the Open Data Kit (ODK) software for data collection with android phones. This saves the cost of printing questionnaires, makes data collection easier, eliminates non-response, enables the researcher to monitor the research assistants recruited for data collection (particularly, when and where each questionnaire was administered, taking pictures of relevant locations, etc), and eliminates the stressful data entry stage that usually follows completion of data collection.

Design of Database, Data Analysis and Interpretation of Results

We provide technical support in the design of database, data entry, data analysis and interpretation of results. However, collecting data with the ODK software (instead of printed questionnaires) removes the need for (and the cost of) designing database and data entry.

Design or Review of Dissertation / Thesis / Project Report

We provide technical support in the design of dissertation / thesis / project report (in compliance with the guidelines of the institution concerned). We provide guidance on data presentation (including creation and formatting of tables and charts). We also review dissertation / thesis / project report and provide guidance on how to improve the quality in compliance with the guidelines of the institution concerned

Manuscript Development

We provide technical support in the development of manuscript (for publication of article extracted from completed Dissertation, Thesis and Project report).