Computer Science

Second Stage 2021-2022

1st Semester

Statistical Package for Social Sciences (SPSS)

By:

Assistant Lecturer: Husam K Salih

INDEX

Chapter 1	Introduction to SPSS	
1.1	What is SPSS	1
1.2	Starting SPSS	2
1.3	Data Entry	3
1.4	File Management	10
Chapter 2	Data Analysis	
2.1	Descriptive Statistics	12
2.2	Editing and Modifying The Data	19
Chapter 3	Computing and Graphics	
3.1	Constructing New Variables	21
3.2	Graphics	25
Chapter 4	Exercises	
4.1	Exercise 1	30
4.2	Exercise 2	31
4.3	Exercise 3	31
4.4	Exercise 4	31
4.5	Exercise 5	31

Chapter One
Introduction to SPSS

1.1 What is SPSS?

SPSS Statistics is a software package used for interactive or batched, statistical analysis. Long produced by SPSS Inc., it was acquired by IBM in 2009. SPSS for Windows is a popular and comprehensive data analysis package containing a multitude of features designed to facilitate the execution of a wide range of statistical analyses. It was developed for the analysis of data in the social sciences - SPSS means Statistical Package for Social Science. It is well suited to analyzing data from surveys and database.

SPSS is capable of handling large amounts of data and can perform all of the analyses covered in the text and much more. SPSS is commonly used in the Social Sciences and in the business world, so familiarity with this program should serve you well in the future. SPSS is updated often. This document was written around a last version of SPSS software which is 23.0, but the differences should not cause any problems.

It comprises two parts:

1- Data editor-creating and modifying data sets.

2- Statistical procedures-analysing the given data sets.

Four steps in data analysis using SPSS:

1- Get your data into the Data Editor. You can open a previously saved data file; read a spreadsheet, text file, or database; or enter your data directly in the Data Editor.
2- Select a procedure from the menus to create tables, calculate statistics, or create charts.
3- Select the variables you want to use in the analysis. The variables in the data file are displayed in a dialog box for the procedure.
4- Run the procedure and study the results.

Figure 1

1.2 Starting SPSS

After logging on to Windows 7, the user will be presented with a screen containing a number of different icons. Start SPSS by clicking the **Start** button then selecting:

All Programs ⟶ **IBM SPSS Statistics** ⟶ **IBM SPSS Statistics 23.0**

Then the **SPSS 23.0** for Windows 7 screen will appear called **Untitled – SPSS Data Editor** (Figure 2).

Figure 2

Click the **New Dataset** within **the New Files** option, to get a blank SPSS data screen and the maximise your SPSS window.

1.3 Data Entry

The SPSS Data Editor screen looks like a spreadsheet but there are some important differences. Each row represents the data for a case. A case could be a patient or a laboratory specimen. It could also be a set of results for a patient at a particular time. Each column represents a variable. A variable could be the answer to a question or any other piece of information recorded on each case. Before you enter any data in the spreadsheet you have to create a variable for the information you have collected. You must define a variable for each question in your data set you plan to analyse.

Defining Variables

If you look at the left hand corner at the bottom of the SPSS Data Editor screen, you will see two small tabs labelled: Data View and Variable View. To create a new variable click on Variable View and the following screen will appear (Figure 3).

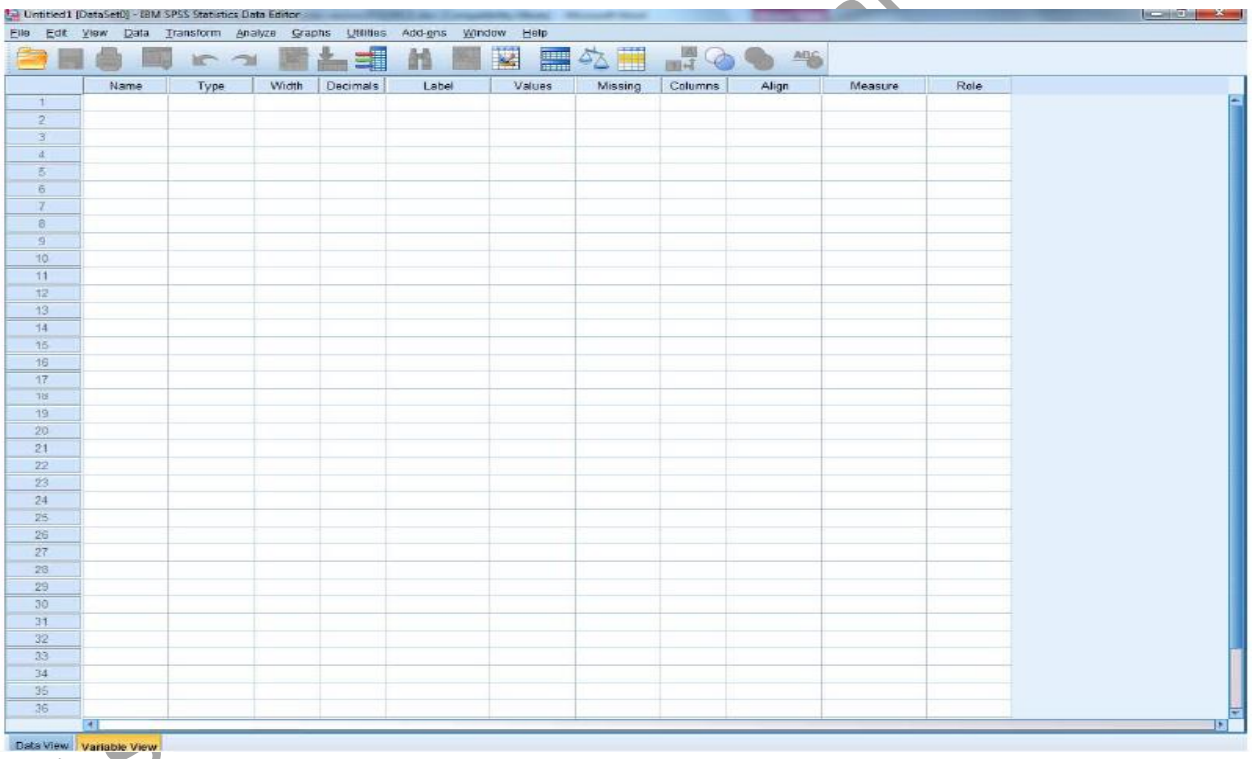


Figure 3

Each row describes the attributes of one variable. Begin by entering a variable name in the **Name** column. A variable name can be up to 64 characters long, must contain no spaces, and should be something meaningful. It is best to stick to alphanumeric characters and start with a letter. Once you have entered a name, SPSS defines the variable type as **Numeric**. You may need to change the variable type, to e.g. **String** if you wanted to use text such as names, or to **Date** if you want to enter dates. To do this, click on the cell within the **Type** column. A little combo button will appear on the right hand side, click the button and the following screen will appear (Figure 4).

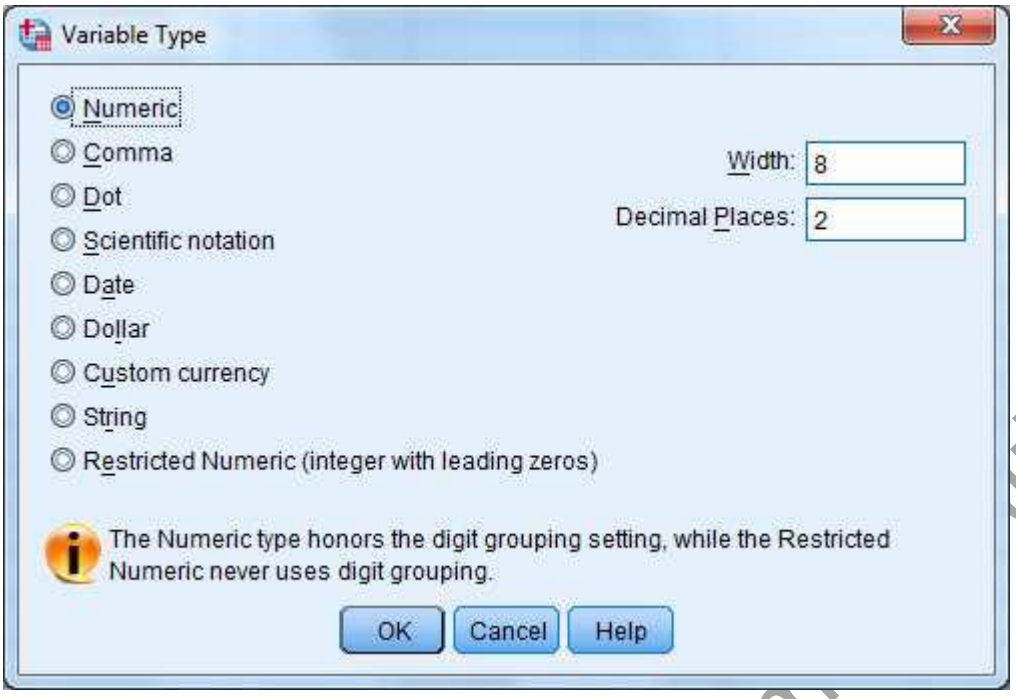


Figure 4

You will usually be working with one of **Numeric**, **Date** or **String** type of data. For Numeric variables you may want to change the decimal places. If the data are integers (whole numbers) such as age in complete years you could alter the decimal places to zero. If the numbers you are planning to enter are very small (0.00072) or you require a high level of precision (21.7865) you may want to increase the number of decimal places. Usually there is no need to change the width from 8, note that width must be larger than the number of decimal places. For a date variable it is best to use a 4 digit year (dd.mm.yyyy).

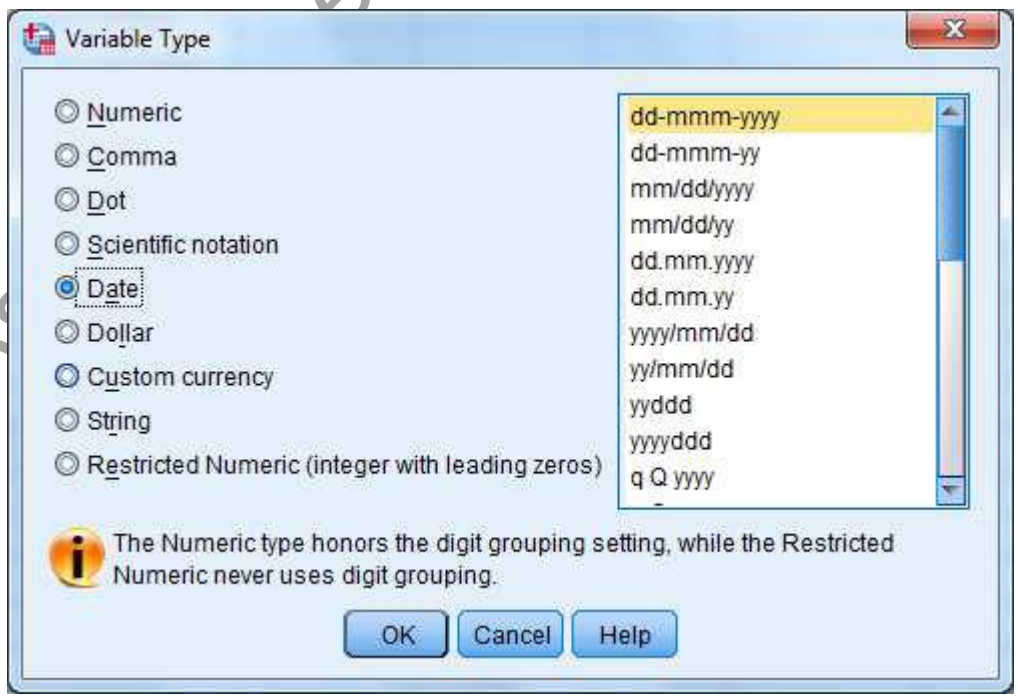


Figure 5

With text strings you are given the option to change the number of characters.

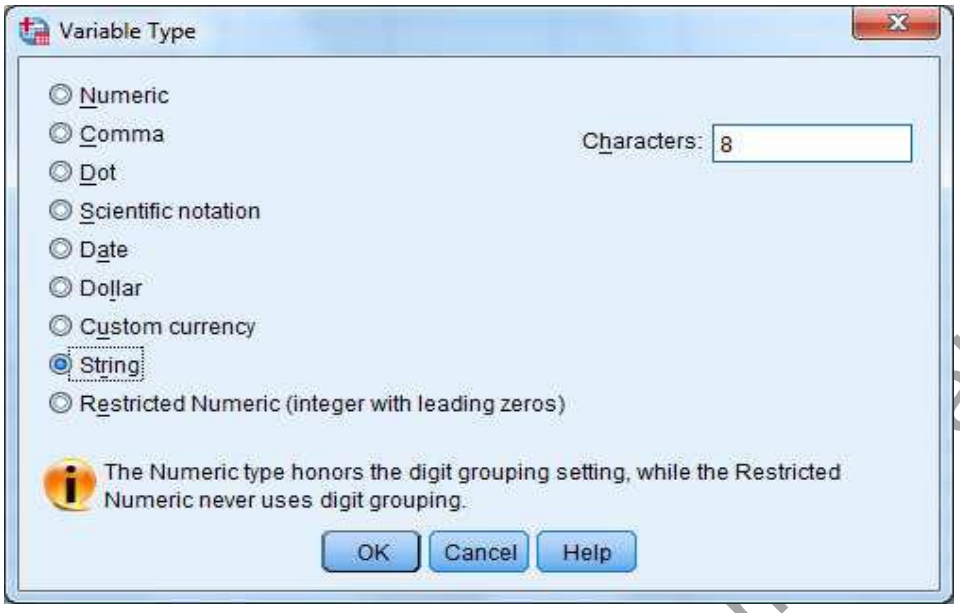


Figure 6

Where possible you are strongly advised to use numerical coding rather than strings as this makes statistical analysis easier. If you are entering string data that is longer than 8 characters, you will need to increase the Width from the default of eight. To be able to fully display the string in the **data view** window you may need to increase the numbers of columns in the **variable view** window. The column missing in the variable view window allows you to define codes that identify a missing value. You can have several values allowing you to distinguish between types of missing data due to the respondent forgetting to answer rather than say not applicable or refused to answer. For example, a code of **-88** could indicate not applicable, and **-99** could indicate the respondent had missed a question out. If a value is defined as a missing value code for a particular variable, subjects with that code will be dropped from the analysis of that variable. To set up missing value codes for a variable, click on a cell followed by the grey square within the **Missing** column as you did with **Type**. Click **Discrete missing values** and enter the values to represent missing in the boxes below (Up to 3 can be entered). To complete the entry press **OK.**

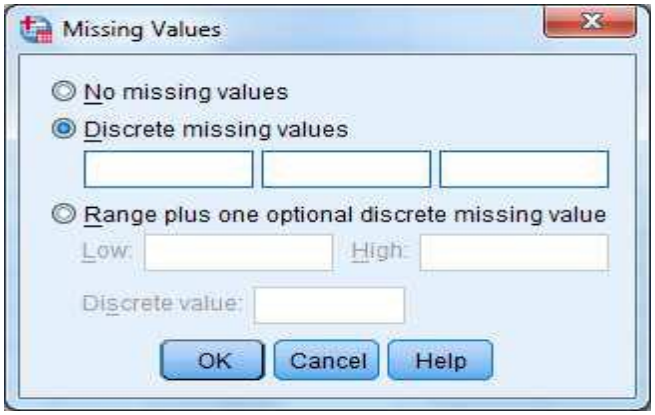


Figure 7

Variable and Value Labels

There are two types of labels in SPSS:

1- **Variable label**, given to a variable gives a clearer description of the variable and will be displayed on the statistical output such as graphs and tables.

2- **Value label** allows you to describe each of the values in a variable. These labels will be displayed on tables improving readability. For example, *Exposure group* in the following practical has two values "Unexposed" and "Exposure to dust" which are coded as "0" and "1". The label option in the variable view window also allows you to define labels for missing values.

To define a variable label click the cell within a Label column screen and enter your description of the variable.

To define Value Labels - click the cell of the value column and then the click on the combo button to the right, then enter the Value and its associated label then press **Add**. The added label will then appear in the (Figure 8).

Figure 8

Once you have entered all the value labels for a variable press **OK**.

Exercise (A): The (table 1) lists the example variables from the construction study. Set-up the following variables:

Variable Name	Description (Variable Label)	Missing Data Code	Value Labels for each code
Id_No	Identification No		
Group	Exposure Group		1 = Exposed to dust 0 = Unexposed
Age	Age at assessment		
Sex			0 = Female 1 = Male
Height	Height in CM		
Asthma	Ever had asthma		0 = No 1 = Yes 2 = Don't Know
Bronchitis	Ever had Bronchitis		0 = No 1 = Yes 2 = Don't Know
Smoke_now	Do you smoke now		0 = No 1 = Yes
Smoke_ever	Have you ever smoked		0 = No 1 = Ex-smoker 2 = Current smoker
Cig_day	No of cigarettes per day		
Cig_year	No of years smoked		

Table 1: Exercise A

Entering Data

When you finish creating all the variables, you enter the **Data View** and the following screen with all the variable names at the top of the spreadsheet.

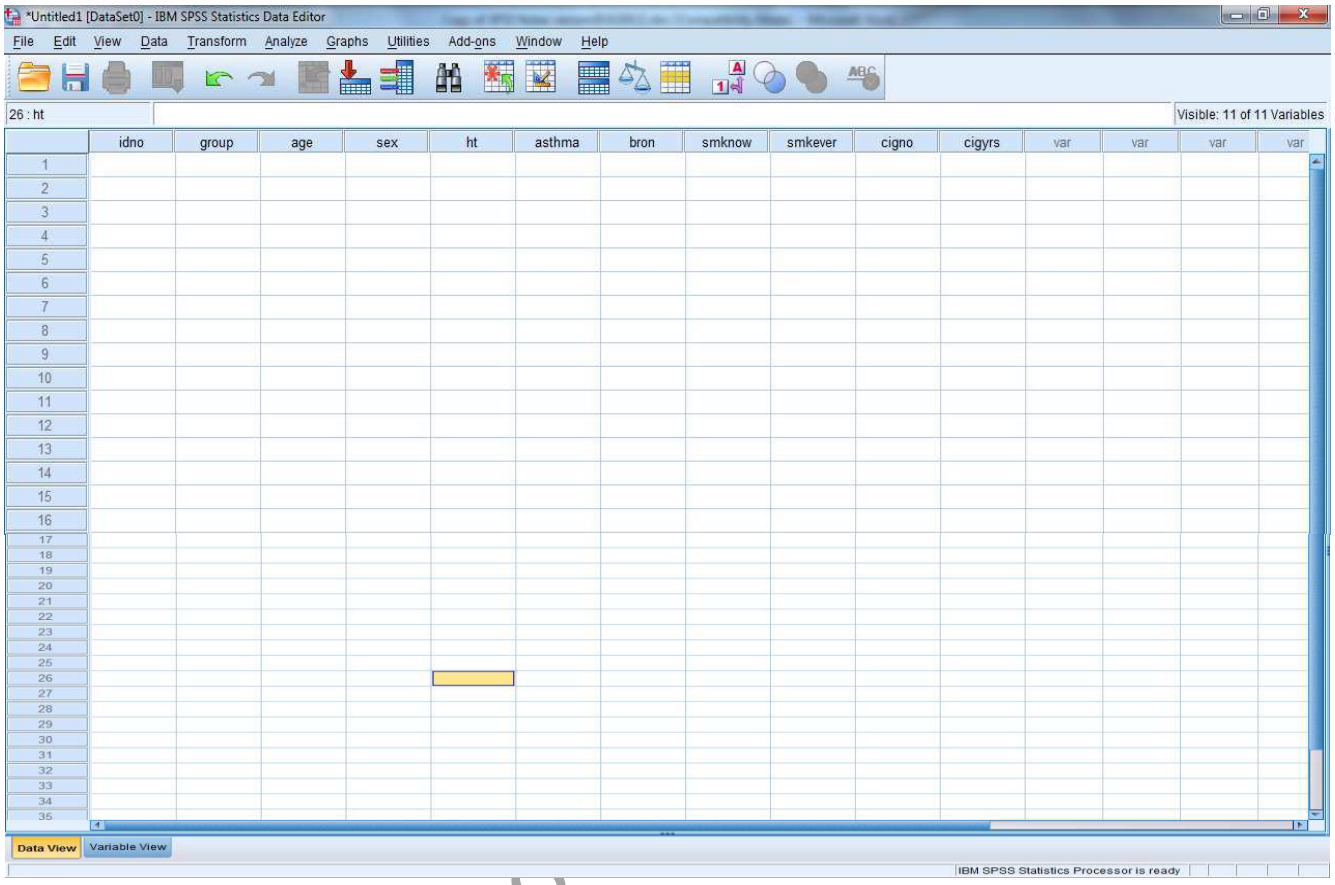


Figure 9

You can now enter the data as you would in an excel spreadsheet. To make an entry in a particular cell on the spreadsheet use the mouse to move the cursor to select that cell and type in the value. The value will appear in the cell. Click on the mouse, press enter or use the cursor keys to enter that value.

If you attempt to enter data of the wrong type into a variable (for example text into a numeric variable) the data will not be accepted. If incorrect data is entered, it can be overtyped or deleted.

فيديو تعليمي لكيفية اعداد وادخال البيانات

Click Here

Exercise (B): The data below are some variables from the foundry study for which you have just entered the variable codes. If you leave a gap in any cell in the worksheet, SPSS will put a dot (.) and treat it as missing data. To enter the cases, either type the number corresponding to the value label or alternatively display the Value Labels of the coded values. These are displayed by using choosing value labels button from the second row of options at the top of either the Data view or Variable View window.

Id_No	Group	Age	Sex	Height	Asthma	Bronchitis	Smoke_now	Smoke_ever	Cig_day	Cig_year
1	**Exp**	**49**	**Female**	**176**	**No**	**No**	**Yes**	**Curr**	**20**	**250**
2	**Exp**	**46**	**Female**	**168**	**Yes**	**No**	**Yes**	**Curr**	**20**	**350**
3	**UnExp**	**34**	**Female**	**180**	**No**	**No**	**No**	**Never**		
4	**UnExp**	**34**	**Male**	**180**	**No**	**No**	**Yes**	**Curr**	**25**	**400**

Table 2: Exercise B

1.4 File Management

Once you have entered some data you should save the file. It is good practice to save data at regular intervals during data entry just in case.

To save the data you have just entered, click the **File** at the top left corner of the screen and then the **Save As...** Sub-option.

Something similar to the following screen will appear:

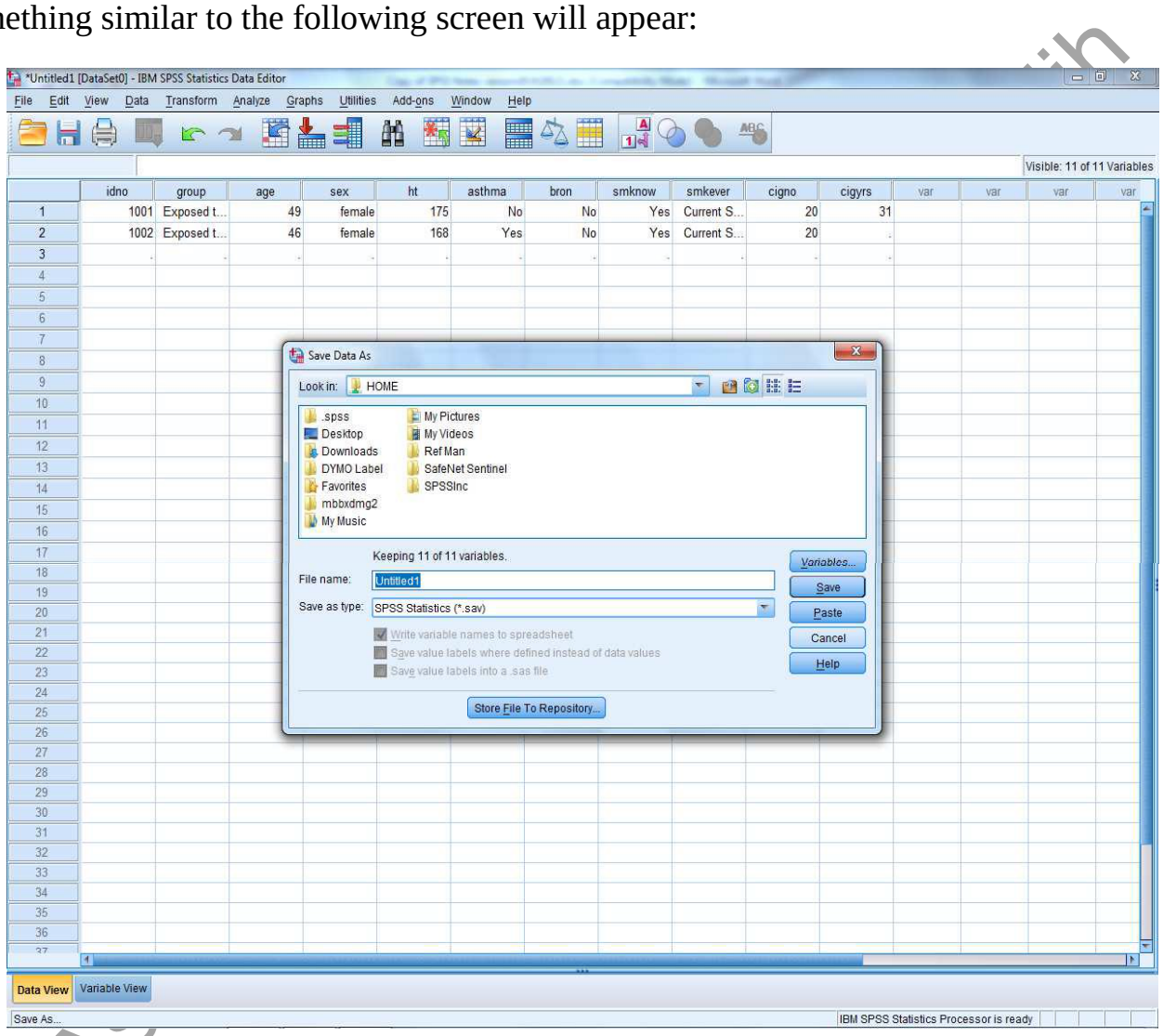


Figure 10

Save a copy of the current **SPSS for Windows 7** file on your P: Drive or your pen drive, under **Drives:** click on ▼in the **Look in** window to generate a list of the drives.

Click on the up/down-arrows to move to the **relevant pen drive** and enter a suitable name in the **File name** window. By default SPSS will add the file extension **.sav** in order to help identify the file as a SPSS data file. Finally, click on the **Save** button.

Backing Up Your Data

It is good practice to save data on different disks and also several names as data entry progresses (e.g. **mydata1 mydata2** etc.). To make a backup copy of your data repeat the **Save Data As** procedure.

Retrieving Data Files

Retrieving an SPSS for Windows 7 File is essentially the reverse of the save process. Click on the **File** option, then the **Open** sub-option followed by the **Data** option. Something similar to the following screen will appear. Then retrieve the required file from the saved location.

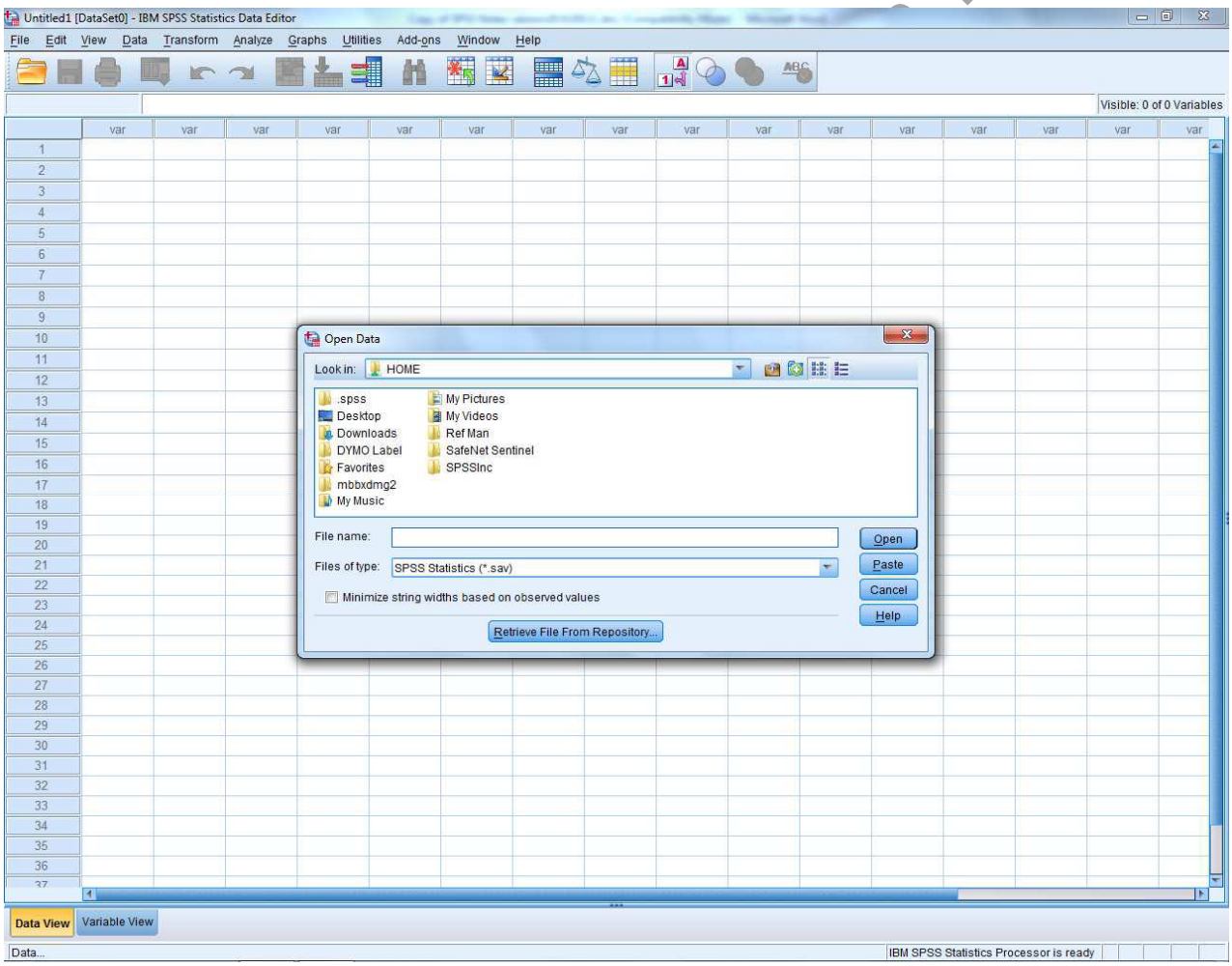


Figure 11

We can also open a data file when we start an SPSS session (see Figure 2).

Chapter two
Data Analysis

2.1 Descriptive Statistics

The first step in data analysis is to generate descriptive statistics. This will give us a feel for the data. It will also help identify any inconsistencies that may be in the data. This is sometimes called data cleaning. Techniques that are commonly used to do this include:

1- Frequency Analyses
2- Descriptive Statistics
3- Cross-tabulations
4- Plots

Frequency Tables

Carrying out a frequencies analysis on variables is the first step when checking for data errors, click on **Analyze** and choose the **Descriptive Statistics** option and then choose **Frequencies.** Move the variables of interest into the **Variables** box on the right-hand side, and then click **Statistics** to select some summary statistics such as range, maximum, minimum, mean and median, which will help you look for errors.

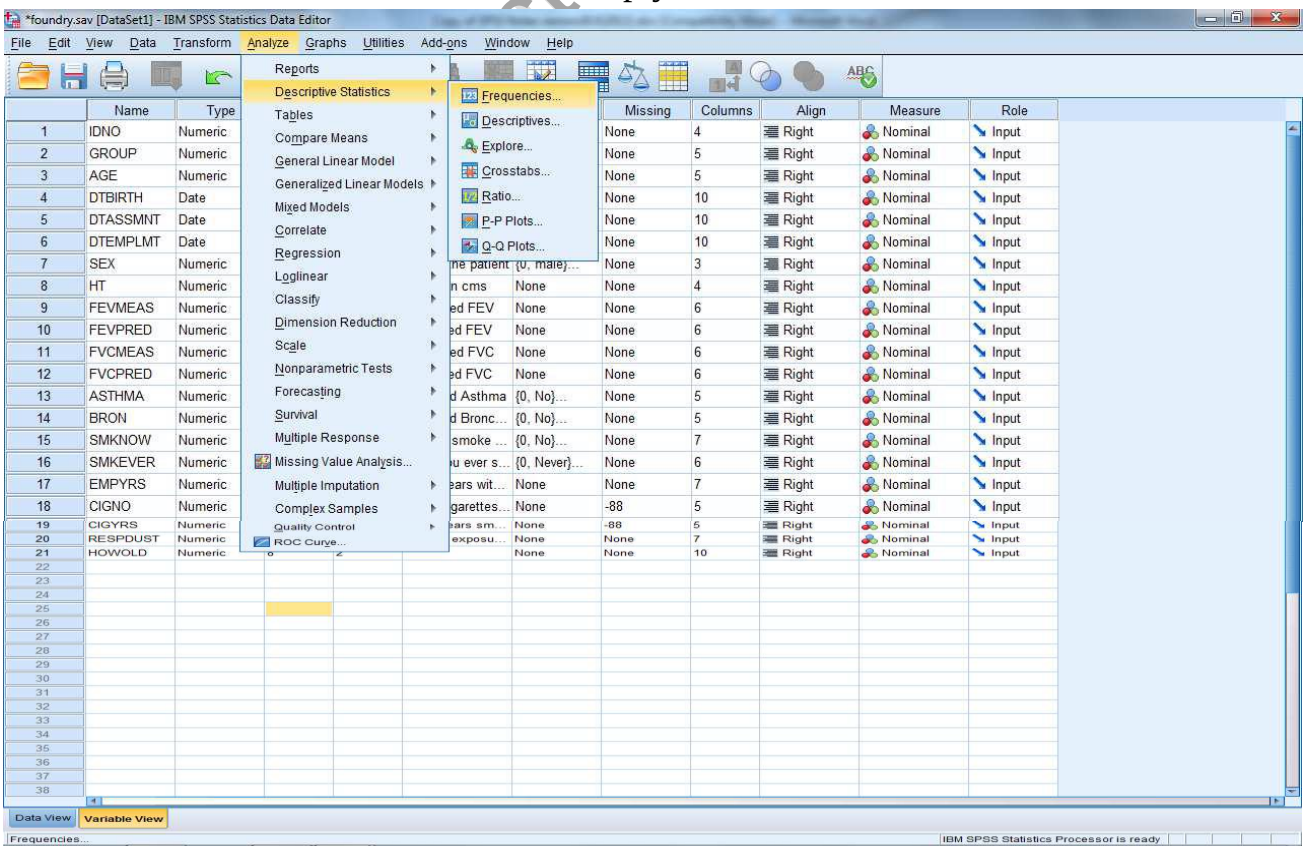


Figure 12

The following screen will appear.

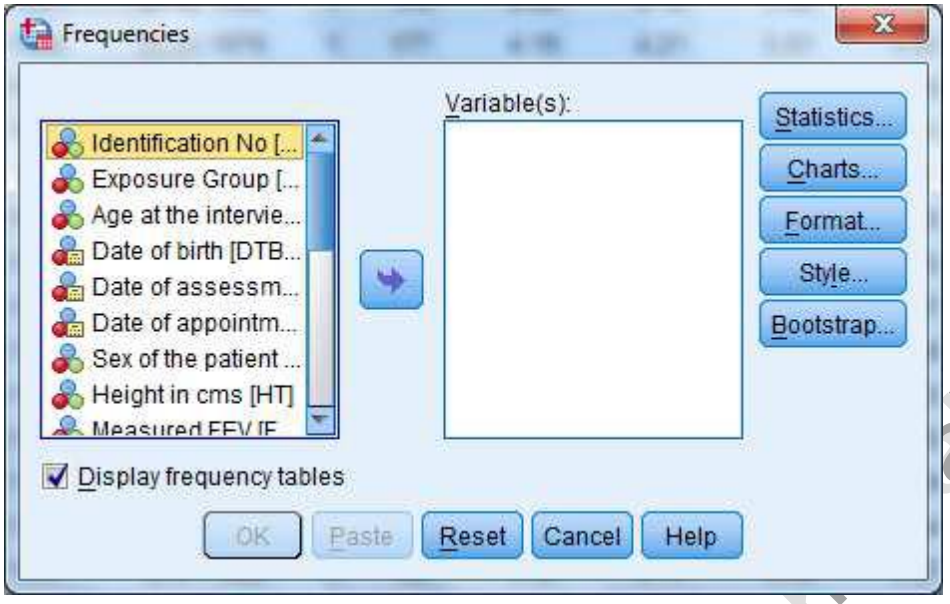


Figure 13

To select the variable to perform a frequency table for example the Exposure group variable, click on its name in the left hand list and then press . Finally click on **OK** and the following output is then generated in the output window.

Exposure Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unexposed	63	46.3	46.3	46.3
	Exposure to Dust	73	53.7	53.7	100.0
	Total	136	100.0	100.0	

Table 3: Frequency Table

To return to the data editor click on **Window** and take the data editor option from the list. With the frequency table you can have a list of summary statistics as well. Click **Analyze, Descriptive Statistics, and Frequencies**. Press reset and then bring the variable (say, **ht**) to the **Variable(s)** window, click on **Statistics** option and select some summary statistics. Click **Continue** and **OK** button.

Once the **OK** button is pressed the results are automatically produced in an **Output window**, if the screen does not appear then the Output window may already exist but is located in the background. All results including the can be copied into word processing documents by clicking on the table and performing a standard copy and paste procedure.

Height in cms

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	158	1	.7	.7	.7
	160	3	2.2	2.2	2.9
	162	1	.7	.7	3.7
	163	6	4.4	4.4	8.1
	165	7	5.1	5.1	13.2
	166	1	.7	.7	14.0
	167	5	3.7	3.7	17.6
	168	14	10.3	10.3	27.9
	170	19	14.0	14.0	41.9
	171	1	.7	.7	42.6
	172	8	5.9	5.9	48.5
	173	7	5.1	5.1	53.7
	174	1	.7	.7	54.4
	175	26	19.1	19.1	73.5
	177	7	5.1	5.1	78.7
	178	5	3.7	3.7	82.4
	180	12	8.8	8.8	91.2
	182	2	1.5	1.5	92.6
	183	2	1.5	1.5	94.1
	185	3	2.2	2.2	96.3
	190	4	2.9	2.9	99.3
	192	1	.7	.7	100.0
	Total	136	100.0	100.0	

Statistics

Height in cms

N	Valid	136
	Missing	0
Mean		172.97
Std. Error of Mean		.567
Median		173.00
Mode		175
Std. Deviation		6.613
Variance		43.732
Skewness		.429
Std. Error of Skewness		.208
Kurtosis		.393
Std. Error of Kurtosis		.413
Range		34
Minimum		158
Maximum		192
Sum		23524

Table 4: Output from Frequencies with some summary statistics

(***Frequency Tables & Descriptive Statistics***) فيديو تعليمي حول

Click Here

Exercise (C): Using the frequencies options find out

1- What proportion of the construction workers were exposed to dust?
2- What proportions had ever suffered from bronchitis?
3- What proportion had ever smoked?
4- What proportion smoked more than 40 cigarettes per day?

Descriptive

The descriptive command in SPSS is useful for summarizing quantitative data. To use this click on the **Analyse** tile choose the **Descriptive Statistics** option and then choose **descriptive.** Move the variables of interest into the **Variables** box on the right-hand side. As with the frequencies command we can obtain descriptive statistics for several variables at once. In the panel below we have chosen some of the quantitative variables in the foundry data set.

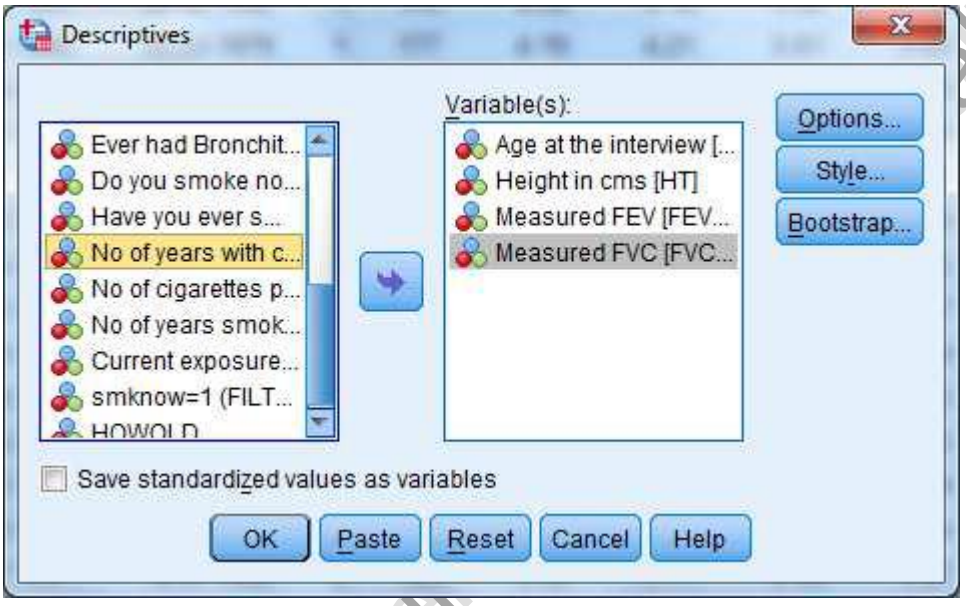


Figure 14

Exercise (D): Use the **descriptive** procedure to determine

1- The current mean exposure to dust per day
2- The mean number of cigarettes smoked per day

For mean number of cigarettes per day you may get a negative answer. Check the missing value codes and redo.

Cross-tabulation

To examine the relationship between two categorical variables, a two way Frequency Table can be used. This is called a cross-tabulation. Click on **Analyze** then **Descriptive Statistics** and then **Crosstabs.** The screen below appears. Suppose we wished to examine how smoking status related to exposure. We could examine this by a cross-tabulation of the variables **group** and **smoke_ever**.

Select the smoking status variable **smoke_ever** labelled **Have you ever smoked** in the source list then click ➔ by the **Row(s)** box to make this the row variable

Select **group** labelled **Exposure Group** in the source list and click ➔ by the **Column's** box to select the column variable. Finally press **OK**

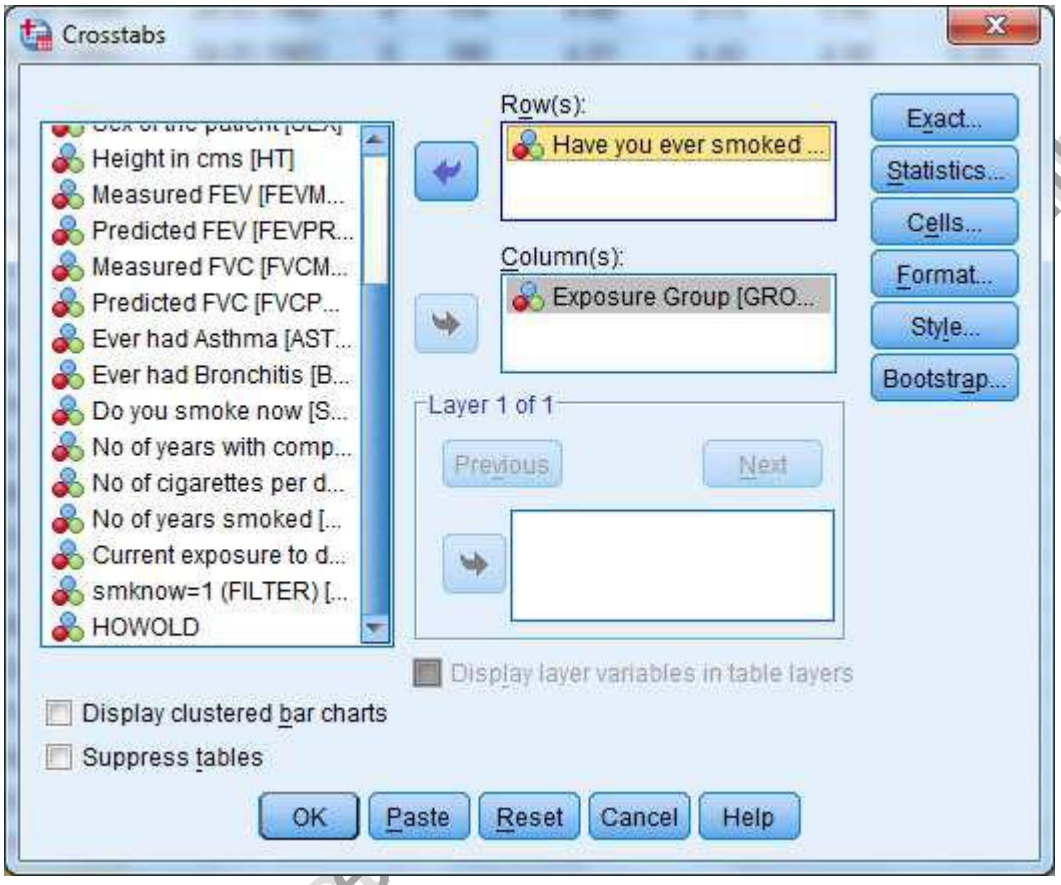


Figure 15

The following result appears when the two frequency table has been completed.

Have you ever smoked * Exposure Group Crosstabulation

Count

		Exposure Group		
		Unexposed	Exposure to Dust	Total
Have you ever smoked	Never	24	20	44
	Ex Smoker	19	19	38
	Curr. Smoker	20	34	54
Total		63	73	136

Table 5: Result of Cross-Tabulation (Simple Cross)

Two way frequency tables are more informative if they include percentages. To add percentages to the table select **Cells** from the Crosstabs screen. On pressing **Cells,** the following screen appears. Column, row, or total percentages can be selected by clicking the appropriate box. Whilst it is tempting to click all three this will make the output confusing. For the table above column percentages are the most useful as they will allow us to compare the smoking status of non-exposed and exposed subjects. By clicking column we get the resulting table.

Have you ever smoked * Exposure Group Crosstabulation

			Exposure Group		
			Unexposed	Exposure to Dust	Total
Have you ever smoked	Never	Count	24	20	44
		% within Exposure Group	38.1%	27.4%	32.4%
	Ex Smoker	Count	19	19	38
		% within Exposure Group	30.2%	26.0%	27.9%
	Curr. Smoker	Count	20	34	54
		% within Exposure Group	31.7%	46.6%	39.7%
Total		Count	63	73	136
		% within Exposure Group	100.0%	100.0%	100.0%

Table 6: Result of Cross-Tabulation (With Percentages)

فيديو تعليمي حول (***Cross-Tabulation***)

Simple Cross ⟶ Click Here

With Percentages ⟶ Click Here

Three-way tables

You may need to do comparisons on three variables. To do this, choose **Analyze** then **Descriptive Statistics** and then **Crosstabs.** Then the following screen appears.

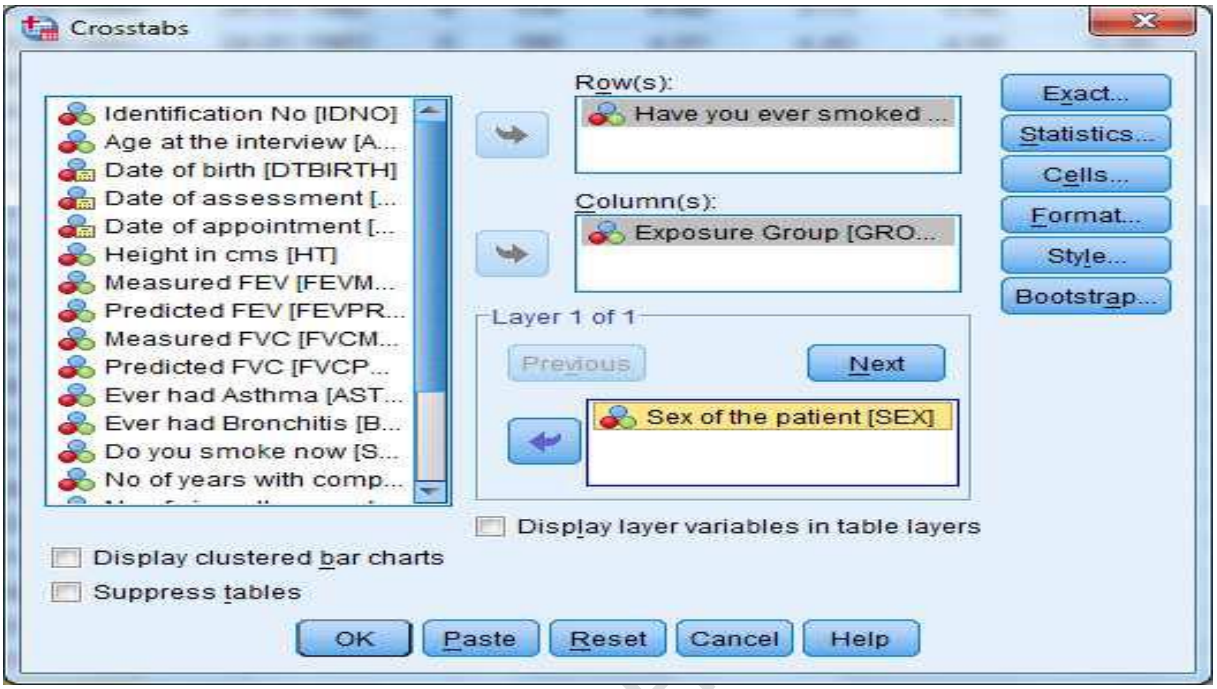


Figure 16

To create a three dimensional table instead of a two dimensional table, click on a variable and move using to layer 1 of 1 box.

If we add the variable sex we will now get separate tables for men and women giving the following output.

Have you ever smoked * Exposure Group * Sex of the patient Crosstabulation

Sex of the patient				Exposure Group		Total
				Unexposed	Exposure to Dust	Unexposed
male	Have you ever smoked	Never	Count	14	6	20
			% within Exposure Group	42.4%	20.0%	31.7%
		Ex Smoker	Count	7	7	14
			% within Exposure Group	21.2%	23.3%	22.2%
		Curr. Smoker	Count	12	17	29
			% within Exposure Group	36.4%	56.7%	46.0%
	Total		Count	33	30	63
			% within Exposure Group	100.0%	100.0%	100.0%
female	Have you ever smoked	Never	Count	10	14	24
			% within Exposure Group	33.3%	32.6%	32.9%
		Ex Smoker	Count	12	12	24
			% within Exposure Group	40.0%	27.9%	32.9%
		Curr. Smoker	Count	8	17	25
			% within Exposure Group	26.7%	39.5%	34.2%
	Total		Count	30	43	73
			% within Exposure Group	100.0%	100.0%	100.0%

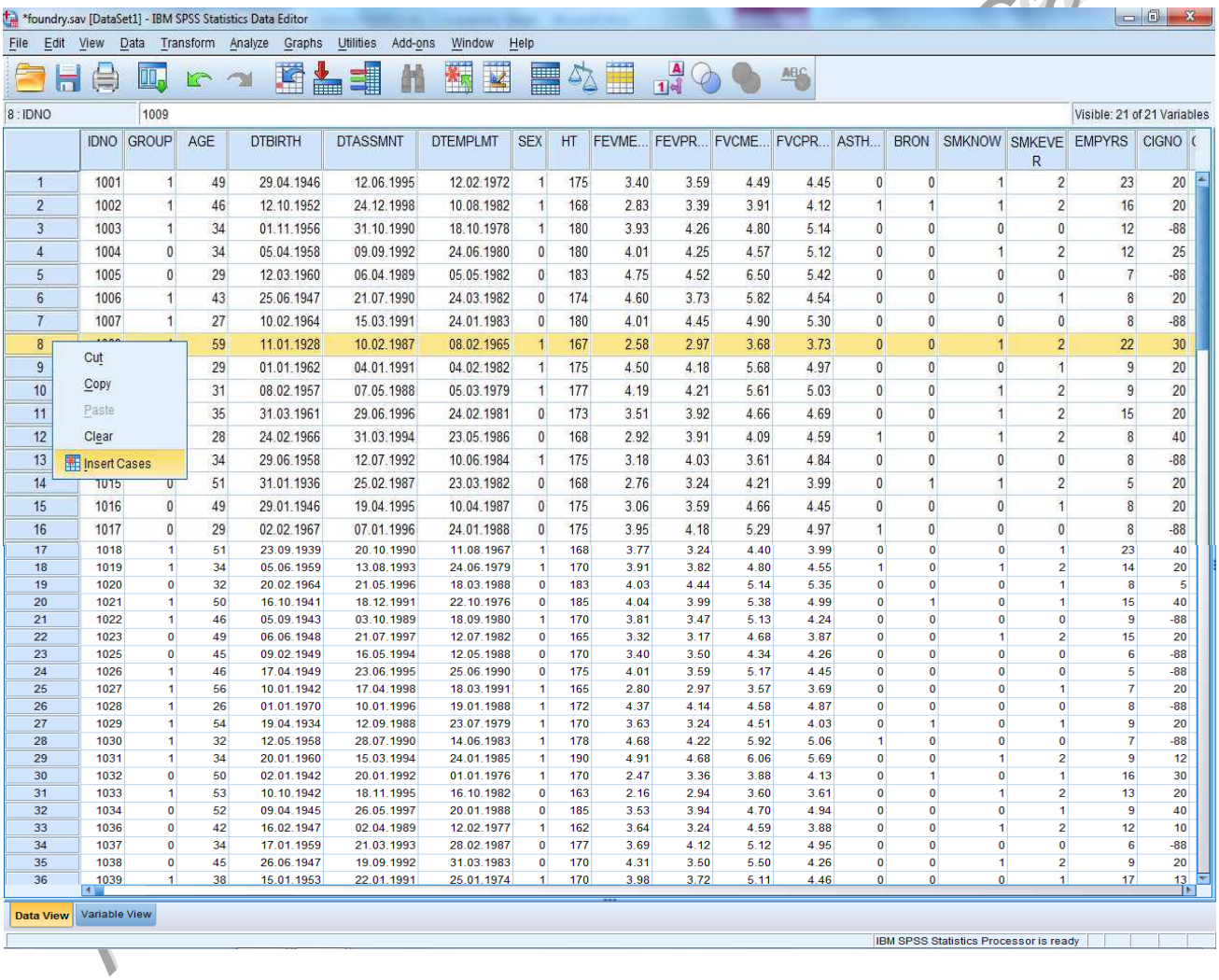
Table 7: Three way tables

2.2 Editing and Modifying The Data

Having done some preliminary analysis we may need to change the data. There are some useful functions for modifying data files.

Inserting Data

To insert data, either click **Edit** then **Insert Case** or right click on the sidebar and click **Insert Case** and a new blank row is added as shown below.



	IDNO	GROUP	AGE	DTBIRTH	DTASSMNT	DTEMPLMT	SEX	HT	FEVME...	FEVPR...	FVCME...	FVCPR...	ASTH...	BRON	SMKNOW	SMKEVER	EMPYRS	CIGNO
1	1001	1	49	29.04.1946	12.06.1995	12.02.1972	1	175	3.40	3.59	4.49	4.45	0	0	1	2	23	20
2	1002	1	46	12.10.1952	24.12.1998	10.08.1982	1	168	2.83	3.39	3.91	4.12	1	1	1	2	16	20
3	1003	1	34	01.11.1956	31.10.1990	18.10.1978	1	180	3.93	4.26	4.80	5.14	0	0	0	0	12	-88
4	1004	0	34	05.04.1958	09.09.1992	24.06.1980	0	180	4.01	4.25	4.57	5.12	0	0	1	2	12	25
5	1005	0	29	12.03.1960	06.04.1989	05.05.1982	0	183	4.75	4.52	6.50	5.42	0	0	0	0	7	-88
6	1006	1	43	25.06.1947	21.07.1990	24.03.1982	0	174	4.60	3.73	5.82	4.54	0	0	0	1	8	20
7	1007	1	27	10.02.1964	15.03.1991	24.01.1983	0	180	4.01	4.45	4.90	5.30	0	0	0	0	8	-88
8			59	11.01.1928	10.02.1987	08.02.1965	1	167	2.58	2.97	3.68	3.73	0	0	1	2	22	30
9			29	01.01.1962	04.01.1991	04.02.1982	1	175	4.50	4.18	5.68	4.97	0	0	0	1	9	20
10			31	08.02.1957	07.05.1988	05.03.1979	1	177	4.19	4.21	5.61	5.03	0	0	1	2	9	20
11			35	31.03.1961	29.06.1996	24.02.1981	0	173	3.51	3.92	4.66	4.69	0	0	1	2	15	20
12			28	24.02.1966	31.03.1994	23.05.1986	0	168	2.92	3.91	4.09	4.59	1	0	1	2	8	40
13			34	29.06.1958	12.07.1992	10.06.1984	1	175	3.18	4.03	3.61	4.84	0	0	0	0	8	-88
14	1015	0	51	31.01.1936	25.02.1987	23.03.1982	0	168	2.76	3.24	4.21	3.99	0	1	1	2	5	20
15	1016	0	49	29.01.1946	19.04.1995	10.04.1987	0	175	3.06	3.59	4.66	4.45	0	0	0	1	8	20
16	1017	0	29	02.02.1967	07.01.1996	24.01.1988	0	175	3.95	4.18	5.29	4.97	1	0	0	0	8	-88
17	1018	1	51	23.09.1939	20.10.1990	11.08.1967	1	168	3.77	3.24	4.40	3.99	0	0	0	1	23	40
18	1019	1	34	05.06.1959	13.08.1993	24.06.1979	1	170	3.91	3.82	4.80	4.55	1	0	1	2	14	20
19	1020	0	32	20.02.1964	21.05.1996	18.03.1988	0	183	4.03	4.44	5.14	5.35	0	0	0	1	8	5
20	1021	1	50	16.10.1941	18.12.1991	22.10.1976	0	185	4.04	3.99	5.38	4.99	0	1	0	1	15	40
21	1022	1	46	05.09.1943	03.10.1989	18.09.1980	1	170	3.81	3.47	5.13	4.24	0	0	0	0	9	-88
22	1023	0	49	06.06.1948	21.07.1997	12.07.1982	0	165	3.32	3.17	4.68	3.87	0	0	1	2	15	20
23	1025	0	45	09.02.1949	16.05.1994	12.05.1988	0	170	3.40	3.50	4.34	4.26	0	0	0	0	6	-88
24	1026	1	46	17.04.1949	23.06.1995	25.06.1990	0	175	4.01	3.59	5.17	4.45	0	0	0	0	5	-88
25	1027	1	56	10.01.1942	17.04.1998	18.03.1991	1	165	2.80	2.97	3.57	3.69	0	0	0	1	7	20
26	1028	1	26	01.01.1970	10.01.1996	19.01.1988	1	172	4.37	4.14	4.58	4.87	0	0	0	0	8	-88
27	1029	1	54	19.04.1934	12.09.1988	23.07.1979	1	170	3.63	3.24	4.51	4.03	0	1	0	1	9	20
28	1030	1	32	12.05.1958	28.07.1990	14.06.1983	1	178	4.68	4.22	5.92	5.06	1	0	0	0	7	-88
29	1031	1	34	20.01.1960	15.03.1994	24.01.1985	1	190	4.91	4.68	6.06	5.69	0	0	1	2	9	12
30	1032	0	50	02.01.1942	20.01.1992	01.01.1976	1	170	2.47	3.36	3.88	4.13	0	1	0	1	16	30
31	1033	1	53	10.10.1942	18.11.1995	16.10.1982	0	163	2.16	2.94	3.60	3.61	0	0	1	2	13	20
32	1034	0	52	09.04.1945	26.05.1997	20.01.1988	0	185	3.53	3.94	4.70	4.94	0	0	0	1	9	40
33	1036	0	42	16.02.1947	02.04.1989	12.02.1977	1	162	3.64	3.24	4.59	3.88	0	0	1	2	12	10
34	1037	0	34	17.01.1959	21.03.1993	28.02.1987	0	177	3.69	4.12	5.12	4.95	0	0	0	0	6	-88
35	1038	0	45	26.06.1947	19.09.1992	31.03.1983	0	170	4.31	3.50	5.50	4.26	0	0	1	2	9	20
36	1039	1	38	15.01.1953	22.01.1991	25.01.1974	1	170	3.98	3.72	5.11	4.46	0	0	0	1	17	13

Figure 17

Deleting a Case

To delete a case, right click on the row number on the far left of the Data Editor to highlight the row containing the case. Press the **Clear** button (alternatively, click on the **Edit** option on the menu bar then click on the **Clear** option) and the case is deleted and the cases below move up to fill the gap.

Inserting a Variable

To insert a variable into the middle of the data, click on the variable after the position at which you wish the variable to appear and then click on **Data** then **Insert Variable**. A blank column is inserted before the selected variable shown here.

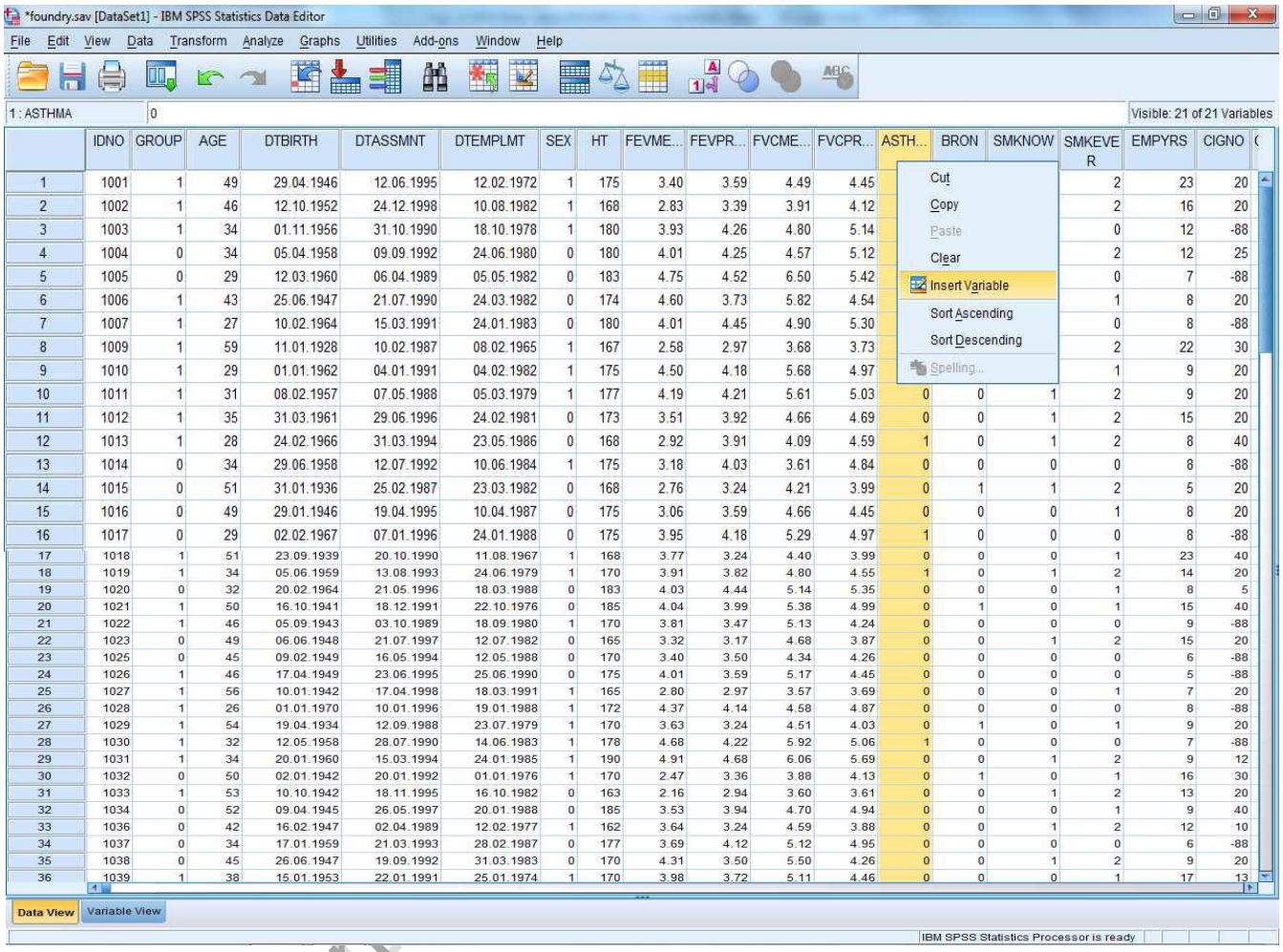


Figure 18

Deleting a Variable

To delete a variable, click on its column name at the top of the Data Editor to highlight the column containing the variable. Then press the **Delete** button. The variable is deleted and the variables to the right move to the left to fill the gap. Now delete the variable you just created.

Moving a Variable

Insert a blank variable as mentioned above in the required position. Click on the name of the variable to be moved (This highlights the column), **Edit** and **Cut.** Click on the name of the blank variable and **Edit** then **Paste**.

Chapter three
Computing and Graphics

3.1 CONSTRUCTING NEW VARIABLES

Sometimes we need to compute new variables from the data entered. For example in the data set we might want to compute the ratio of the measured to predicted fev. Alternatively, we might want to group ages into bands. SPSS has procedures to construct a new variable from existing variables.

Computing a New Variable

For the foundry worker data we shall compute the variable **fevratio** defined as **fevmeas/fevpred.** Click **Transform** then **Compute** and the following screen appears:-

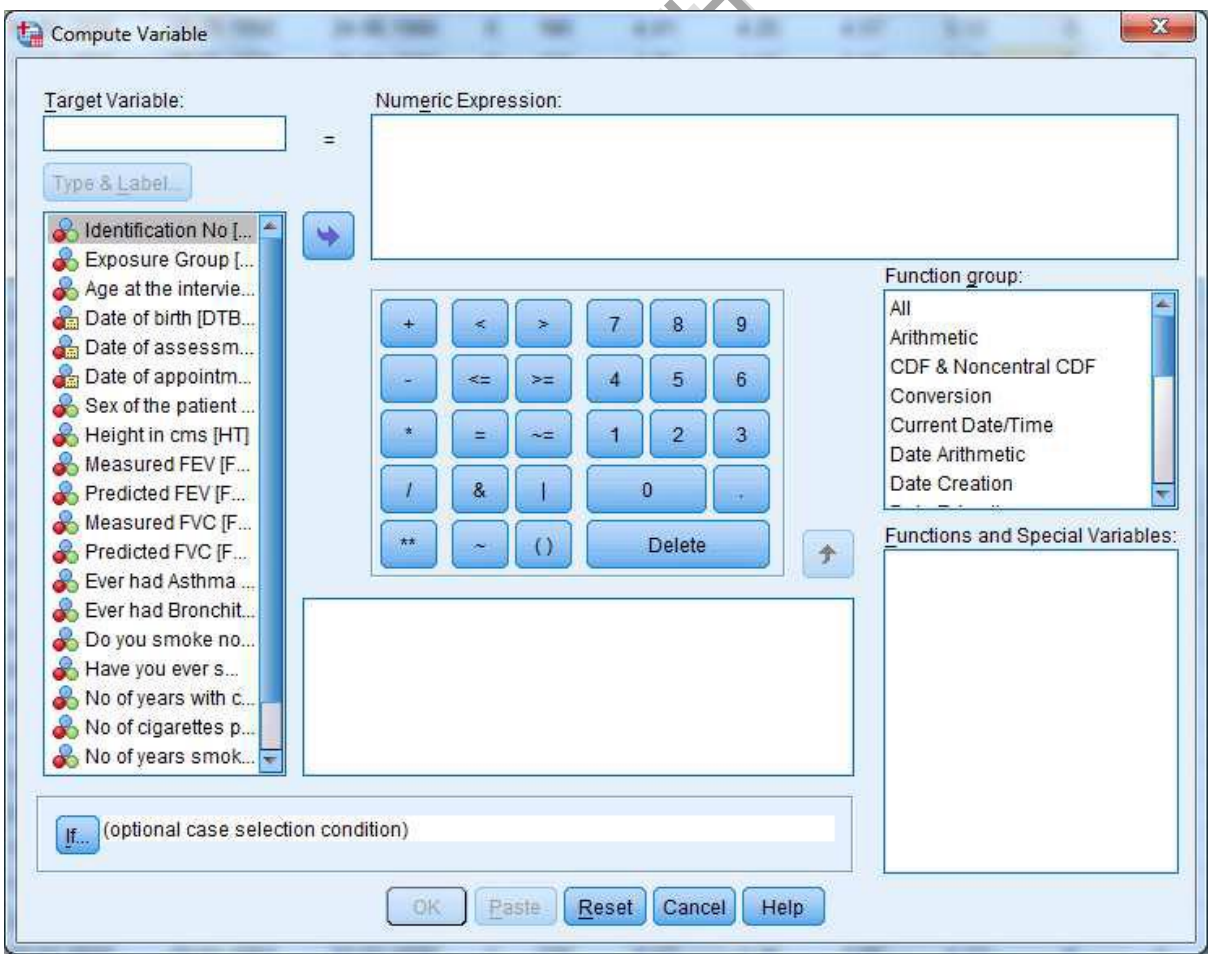


Figure 19

Enter the name **fevratio** in **Target variable** window. If the variable is new, click on **Type & Label** to define the type and variable label. To build up mathematical expression which will create the new variable you can choose variables from the left hand box then click to move them to the **numeric expression** window. You can choose any of the keys on the calculator pad in the centre or any of the functions from the built-in functions box followed by

Select the function using up and down arrow key from the Built in function window and then click on the button .The expression will appear in the **Numeric Expression** window.

These are the functions on the calculator pad are defined as follows.

Operator	Mnemonic form	Description	Operator	Mnemonic form	Description
+		Addition	>=	GE	Greater Than Or Equal To
-		Subtraction	=	EQ	Equals
*		Multiplication	~=	NE	Not Equals
/		Division	&	AND	Logical And
**		Power Of	|	OR	Logical Or
<	LT	Less Than	( )		Parentheses
>	GT	Greater Than	~	NOT	Logical Not
<=	LE	Less Than Or Equal To			

Table 8: Functions Description

To compute **fevratio** we move **fevmeas** and **fevpred** into the **numeric expression** window. You can also type a formula into the numeric expression window. This is illustrated below.

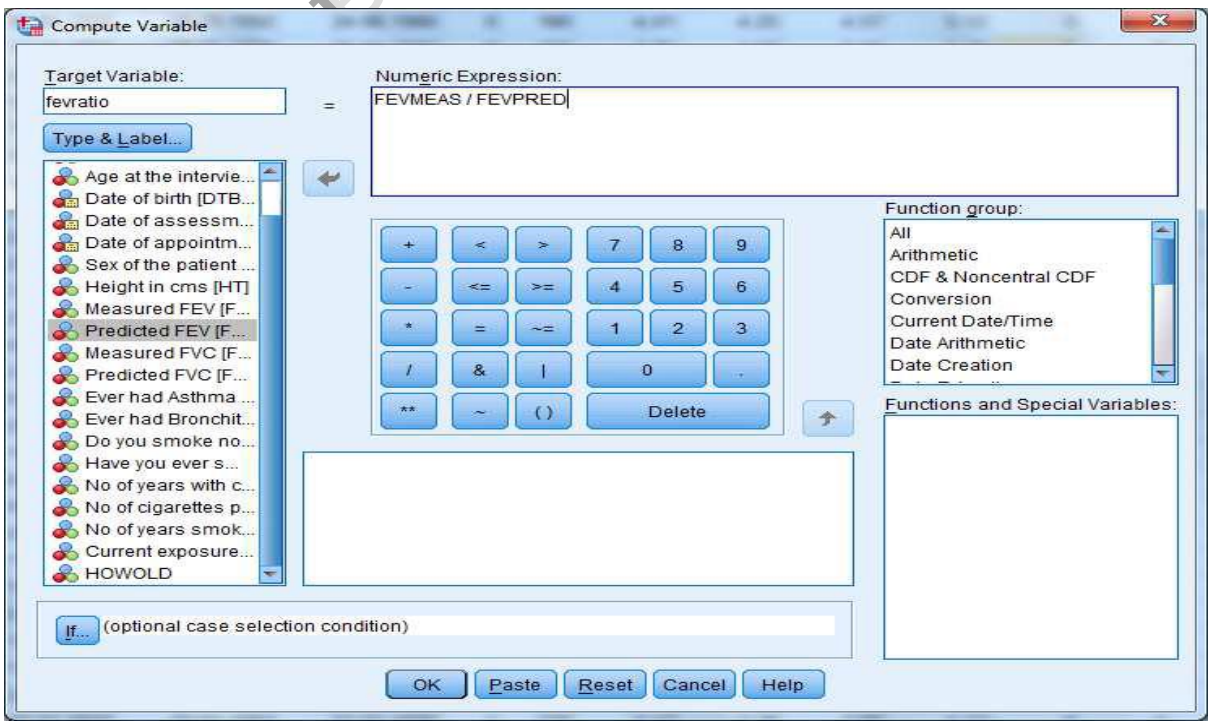


Figure 20

Computing a New Variable by using built-in Functions

In the **Compute** procedure there is a built in functions window which can be used to create a new variable or to transform the values of an existing variable. Transformations such as the square root, or the logarithm, are easily made. Suppose you wish to do a log transformation of the variable called height (**ht**) from the **foundry** data set. First click **Transform** from menu bar and then choose **Compute** from drop down menu, then you get the compute window.

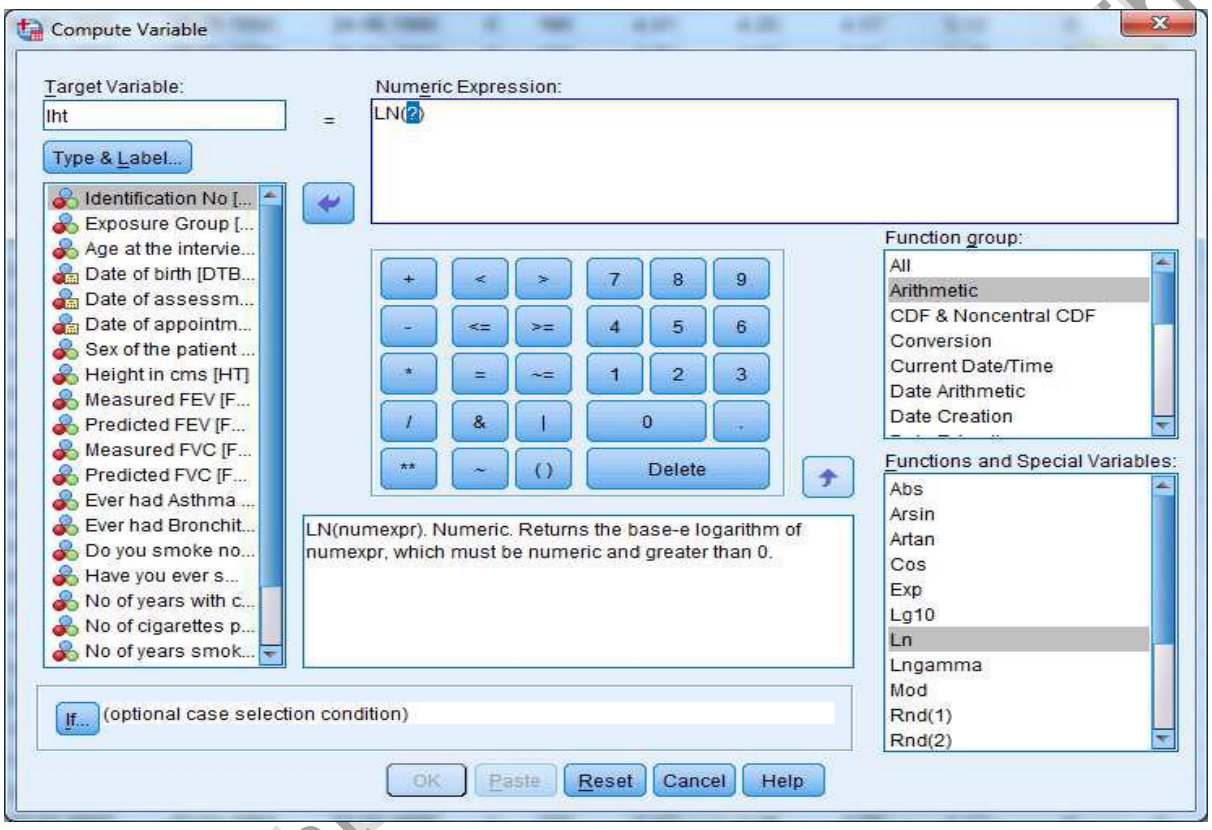


Figure 21

Type a name, say **lht**, in the target variable window. Click on the arrow on the right of the **Functions** box to scroll up and down through the functions. Select **Arithmetic** followed by **Ln** function in the **Functions and Special Variables** box for natural log and click on **Functions** , this will put the function with a **?** In parentheses in the window named **Numeric Expression.** Then select the variable to replace? i.e. **ht** by clicking and then press **OK** button. Then a new variable **lht** will be created (located at the end of the variable list). Having carried out a transformation it is important to check the result. For example, taking a log of a negative value creates a missing value. Other commonly used transformation functions are **LG10, SQRT, ABS, TRUNC** etc.

(***Creating a new variable – log transformation***) فيديو تعليمي حول

Click Here

Computing Duration of Time Difference by built-in Functions

In the same data set there are some variables (date of birth, date of assessment etc) which are stored in date format. One is able to calculate the time difference (in days) by using the functions **Ctime.Days** The age of the patients on the date of assessment can be calculated from the date of birth and assessment date. As before click **Transform** from menu bar and then **Compute** from drop down menu, you then get the compute window. In the target variable window type a name say **howold**, then select the functions group **Time Duration Extraction** followed by **Ctime.Days** in the Functions and Special Variables window using the up and down arrow keys, click on **Functions** , this will put the function with a **?** In parentheses in the box named **Numeric Expression.** Then select the variable to replace? i.e. date of assessment by clicking . Perform the same procedure for date of birth. You can then compute the difference **Time** (in days), then you have to divide the whole thing by 365 (number of days in quarterly leap year) to get **howold** in years. Below is the example.

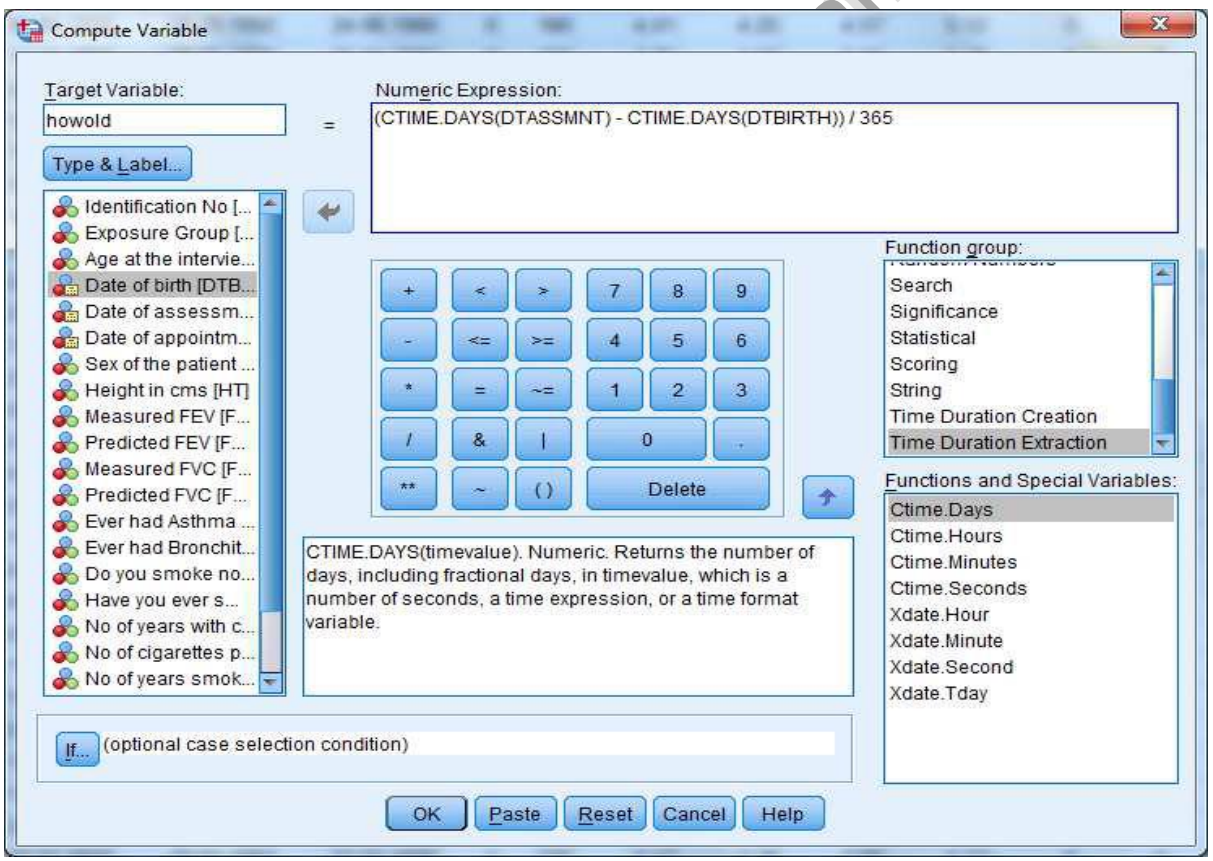


Figure 22

Whenever you compute a new variable from existing data it is important to check that what you have created is sensible. You also need to check that missing values have not been converted into none missing values. Using the **Data view** tab check the value of **howold**.

3.2 GRAPHICS

SPSS will produce good quality high- resolution statistical graphics. We will look at Bar Charts, Histograms, and Scatter Plots with regression lines directly from the data. Please note, that sometimes it is easier in Excel to create bar charts using the frequencies.

Bar Charts

Bar Charts can only be **produced for categorical variables** e.g. ever smoked Asthma etc. To produce a Bar Chart click **Graphs, Legacy Dialogs** then **Bar** and the following screen appears.

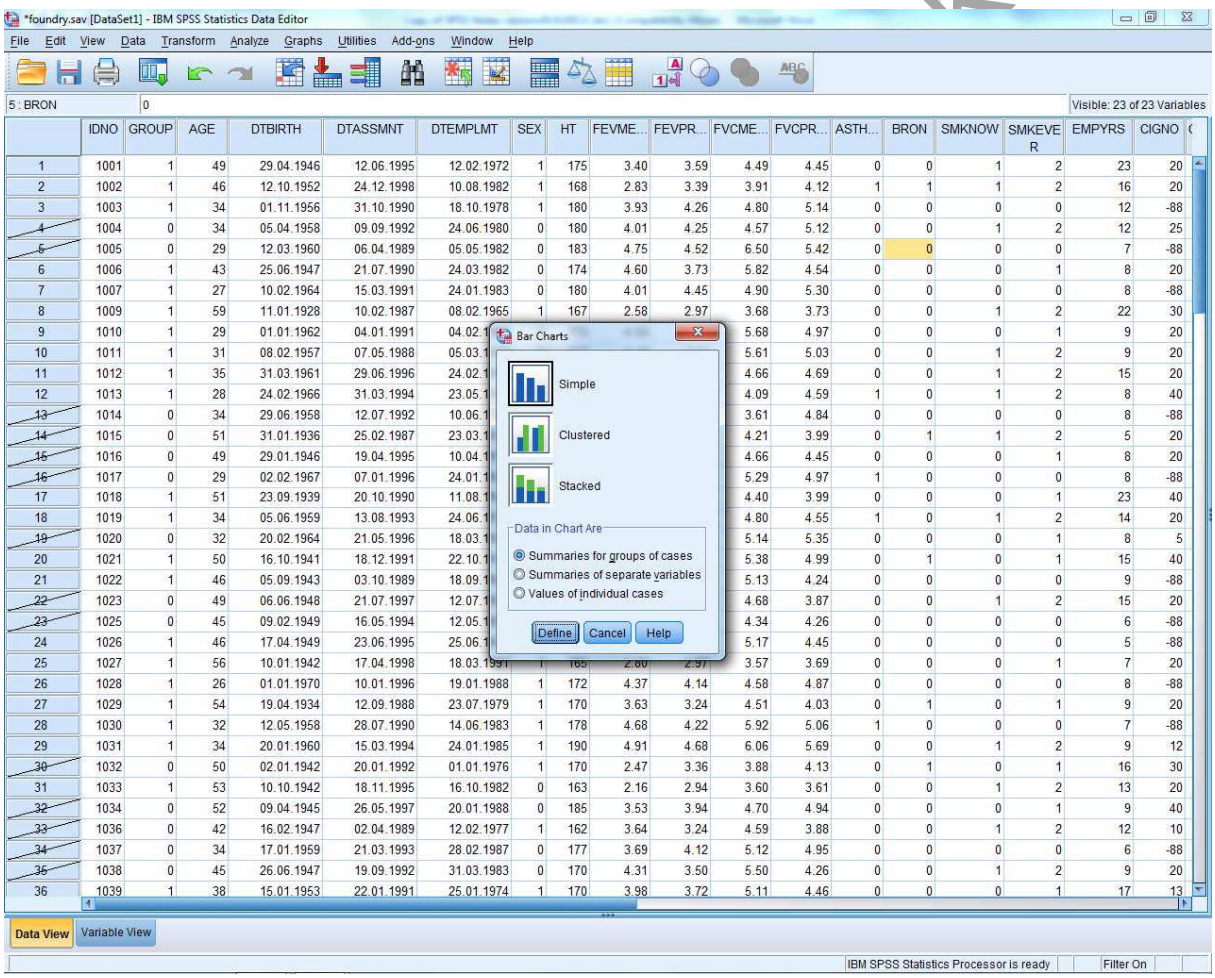


Figure 23

Click on **Simple** and then **Define** and the next screen will appear. Click **No of Cases**, then move your chosen variable from the left hand list to the **Categorical Axis** and press **OK.**

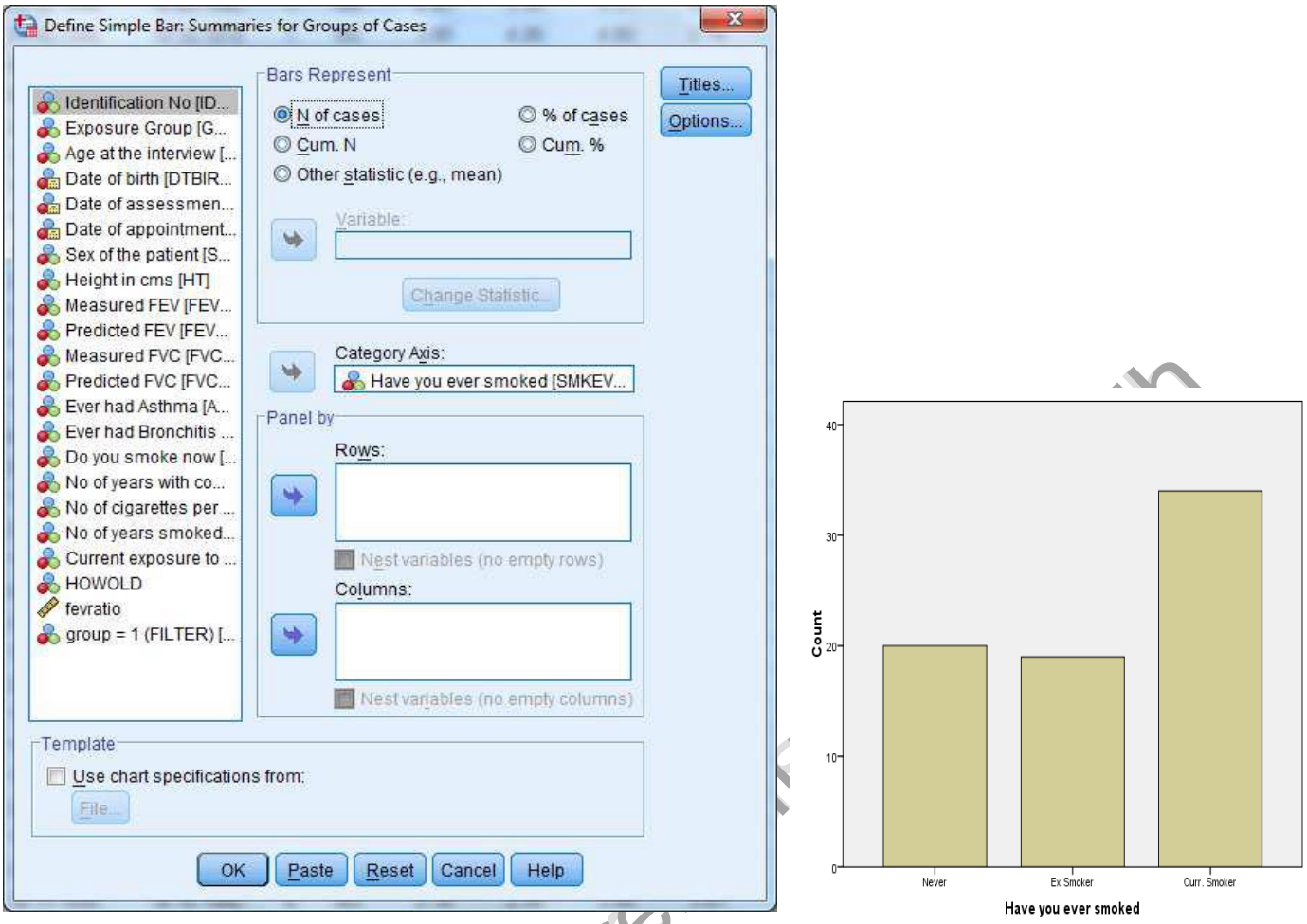


Figure 24

فيديو تعليمي حول (***Bar Chart***)

Click Here

Histograms

At this point it is a good idea to return the select cases back to all data, by **Data, Select Cases**, and then **All Cases** followed by **ok**.

Histograms are **produced for interval variables** e.g. age. To produce a histogram click on **Graphs, Legacy Dialogs** then **Histogram** and the following screen appears.

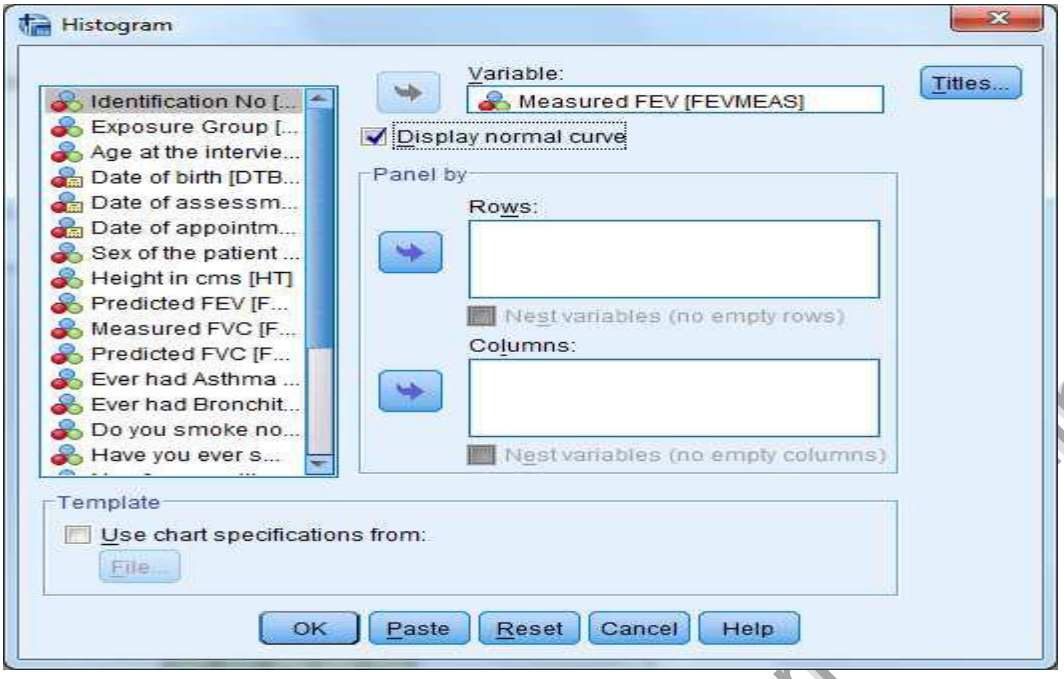


Figure 25

Click on the required variable, in this case FEV, in the left hand side list and press  then press **OK.** If you require a normal curve to be drawn on the graph click on **Display normal curve.**

This is the Histogram produced for measured FEV.

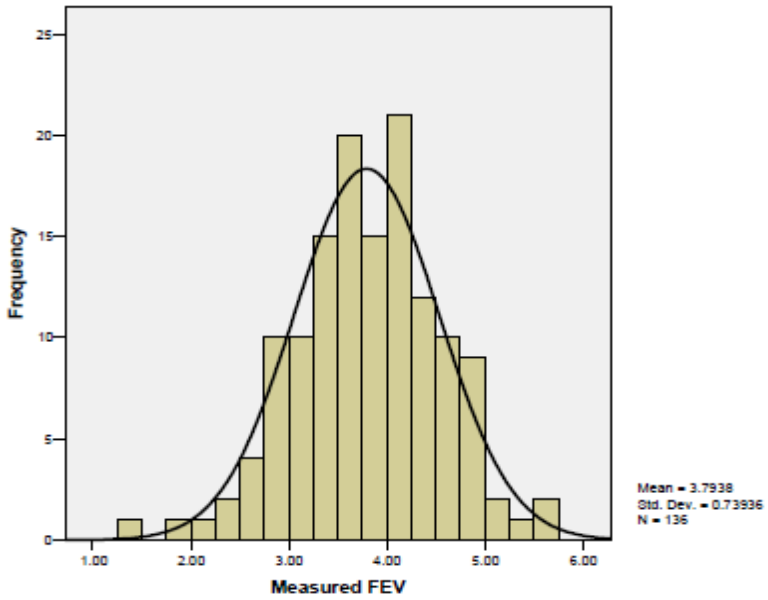


Figure 26

فيديو تعليمي حول (***Histogram***)

Click Here

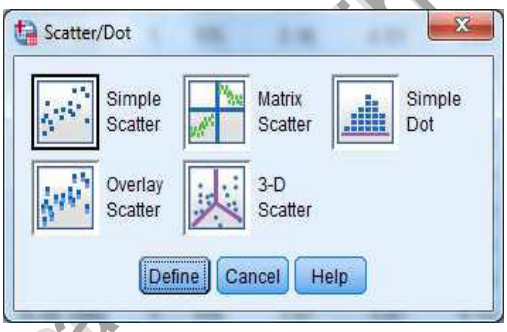
Scatter Plots

Scatter plots show the joint behaviour of two interval variables. If you want to decide whether two interval variables are related in any way you should first draw a scatter plot.

Scatter plots have 2 axes:

1- The value of the dependent or response variable on the y axis.
2- The value of the independent variable on the horizontal axis.

To run a scatter plot click **Graphs – Legacy Dialogs – Scatter/dot** and the following will appear.



Click on **Simple scatter** and then select variables

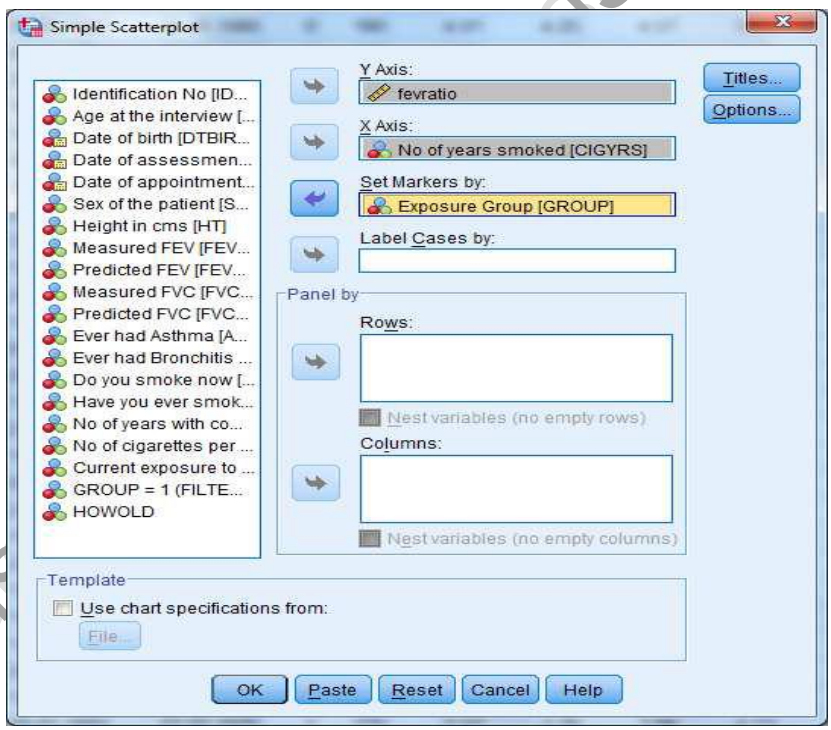
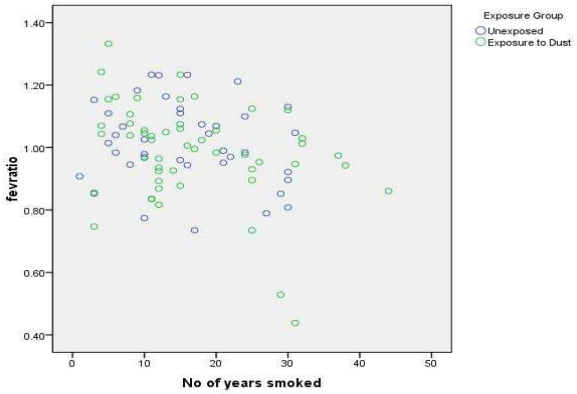


Figure 27

The above selection produces this graph ⟶



Plotting a Regression Line on a Scatter Plot

To fit a line of regression, double left click on the graph. This moves the graph into the Chart Editor. A Regression line can be added by clicking on **Elements** then **Fit Line Total** if you have not defined any markers or **Fit Line Subgroups** if you have defined markers.

This produces the following graph.

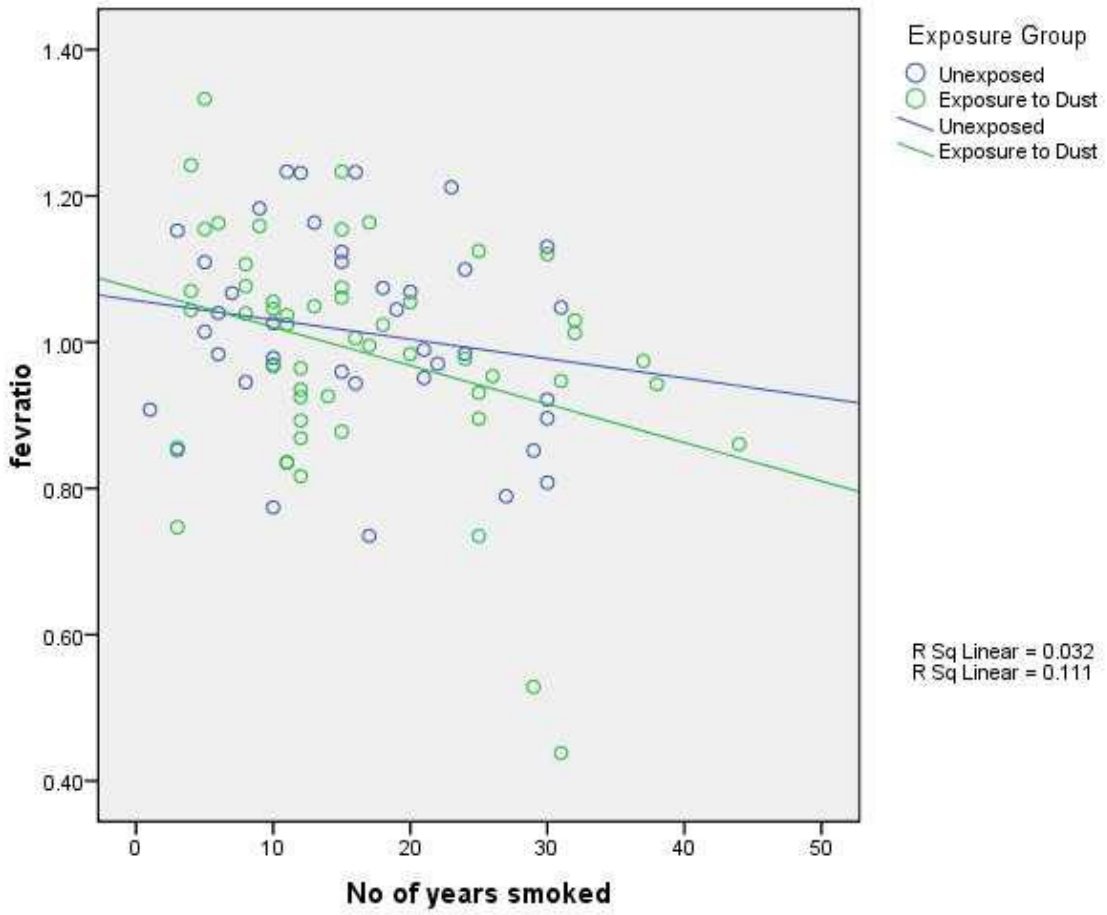


Figure 28

(***Scatter Plot with Regression Line***) فيديو تعليمي حول

Click Here

Chapter Four
Exercises

4.1 Exercise 1:

A- Set-up the following variables:

Variable Name	Description (Variable Label)	Missing Data Code	Value
Name	**Name of the Student**		
Stage	**Year of Study**		**1= first stage** **2= second stage** **3= third stage** **4= fourth stage** **5= fifth stage**
Age	**Age of the Student**		
Gender	**Gender of the Student**		**1= male** **2= female**
Weight	**Weight of the Student**		
Status	**Marital Status**		**1= single** **2= married**
City	**Current City**		
Driving	**Do you know how to Drive?**		**1= yes** **2=no**
Exercise	**How often you do the Sport exercises?**		**1= once a week** **2= twice a week** **3= not often** **4= never**

B- Fill up the form for the 15 students.

4.2 Exercise 2: Using the frequencies options find out:

1- What proportion of the males and females?
2- What proportion of the students those got married?
3- What proportion of the students those doing sport exercises?

4.3 Exercise 3: Use the **descriptive** procedure to determine the relationship between the **stages**, **gender**, and **driving** with **percentages**

4.4 Exercise 4: Use the **descriptive** procedure to determine the median of **(Ages)** & **(weight)**

4.5 Exercise 5: Find out the following graphs:

1- Bar chart of (driving, age, gender).
2- Histogram of (age), display the normal curve.
3- Regression Line on a Scatter Plot of
    a. Ages
    b. Weight
    c. Gender
    d. Sport exercises