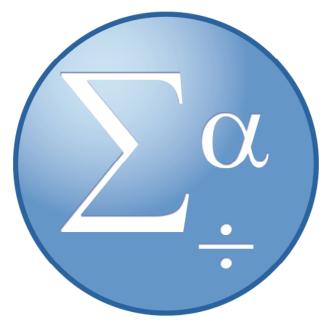


Computer Science

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	Wha	t is SPSS	1			
1.2	Start	ing SPSS	2			
1.3	Dat	ta Entry	3			
1.4	File M	anagement	10			
	Chapter 2	Data Analysis				
2.1	Descrip	tive Statistics	12			
2.2	2.2 Editing and Modifying The Data					
	Chapter 3	Computing and Graphic	S			
3.1	Constructin	g New Variables	21			
3.2	Gi	Graphics				
	Chapter 4	Exercises				
4.1	Exe	ercise 1	30			
4.2	Exe	xercise 2				
4.3	Exe	ercise 3	31			
4.4	Exe	ercise 4	31			
4.5	Exe	ercise 5	31			

Chapter Gne 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 <u>23.0</u>, 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 <u>Start</u> 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

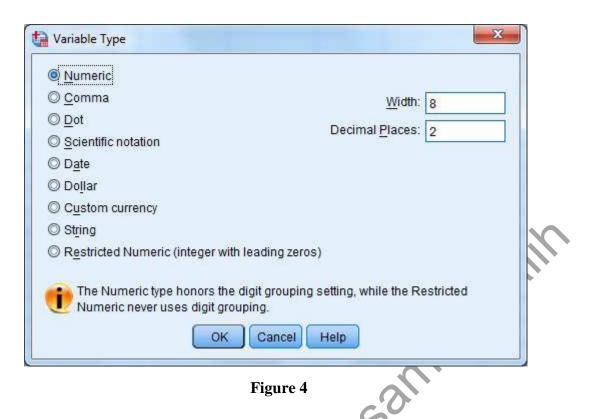
The SPSS Data Editor screen looks like a spreadsheet but there are some important differences. Each row represents the data for a case. A case could be a patient or a laboratory specimen. It could also be a set of results for a patient at a particular time. Each column represents a variable. A variable 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).

Name	Туре	Width	Decimals	Label	Values		Columns	Align	Measure	Role	
Haife	ine	writelli -	Construe's	Love	Falues	in sound	Summas.	cudin	inicapure,	, wee	
6											
-	-										
1											
		-									
	-										
1											
	-	_									
-											
1											
					-						
Variable Viev		_									

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).



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).

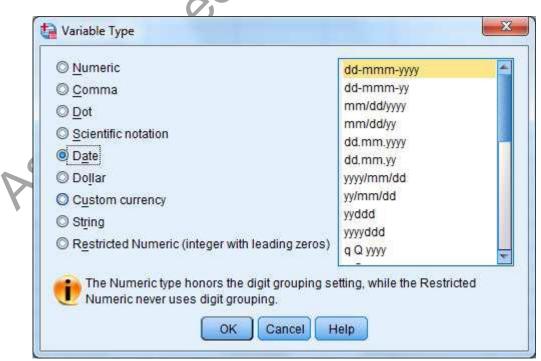
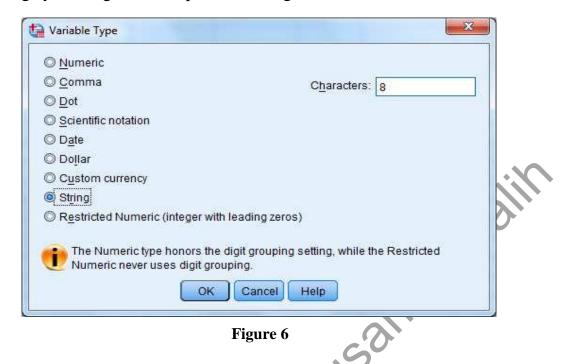


Figure 5



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

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**. Chick **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**.

O No missing v	alues
Oiscrete miss	sing values
© <u>R</u> ange plus o	one optional discrete missing va
Low:	High.
Discrete value	at
CASCING VAILR	H-1

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).

t	Value Labels	×
	Value Labels Value: 1 Label: Exposed to Dust .00 = "Unexposed" Add Change Remove	Spelling
P	OK Cancel Help	

Figure 8

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

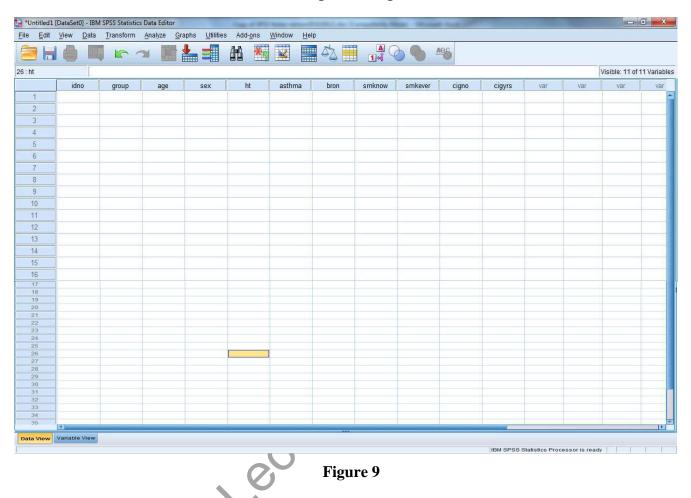
Exercise (A): The (table 1) lists the example variables from the construction study. Set-up the following variables:

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
		S	1 = Yes
			2 = Don't Know
Smoke_now	Do you smoke now		0 = No
		C + 1	1 = Yes
0 1			0 N
Smoke_ever	Have you ever smoked		0 = No
	X		1 = Ex-smoker
Cig_day	No of cigarettes per day		2 = Current smoker
Cig_year	No of years smoked		
AS	Table 1: H	Exercise A	
V			

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.



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.

		-			-				• •	
ld_No	Group	Age	Sex	Height	Asthma	Bronchitis	Smoke_now	Smoke_ever	Cig_day	Cig_yea
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		
1	UnExp	34	Male	180	No	No	Yes	Curr	25	400
					,cill	e 2: Exerc				
			×							

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.

*Untitled1	[DataSet0] - IBM	SPSS Statistics	Data Editor		Case of Pro-	-	-		-	Total State			100		0 X
ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform		_		Vindow <u>H</u> elp	7/240 7/2442	-		(i)					
2 H		100	M 🖉		# 🔠		A3 ■								
										-				Visible: 11 of 1	11 Variab
	idno	group	age	sex	ht	asthma	bron	smknow	smkever	cigno	cigyrs	var	var	var	var
1	1001	Exposed t	49	fema	e 175	No	No	Yes	Current S	20	31				
2	1002	Exposed t	46	i fema	e 168	Yes	No	Yes	Current S	20	6				
3			1		s - 6	-	14			-	18				
4															
5					_										
6															
7			(t	Save Data A	5						X				
9				-						Por Halls Have					
10				Lookin: 📱 HOME 🗾 👩 🔢 🗉											
11				.spss	📕 My Pic										
12	1			Desktop	ds 🔒 Ref Ma										
13	1			DYMO La		et Sentinel									
14	1			Favorites	🚺 SPSSI	nc									
15				mbbxdm											
16				My Music											
17					Keeping 11 of 11	variables.				Var	iables				
18	1			File name:	Untitled1						Save				
19				Save as type:	SPSS Statistics	* sav)									
20					Write variable		Contractor and				Paste				
22					Save value la			data values			ancel				
23	1				Save value la						Help				
24															
25						Store File T	o Repository								
26	1		Ľ	-	1		-	_		-					
27															
28 29															
30	1														
31	-														
32								-							
33															
34	1														
35															
36															
	4			à contratore	di na seraiti			1120100							1
ata View	Variable View														
ave As	X A										IBM SPSS S	tatistics Proc	essor is rea	dv	
1	X						Figure	. 10							

Something similar to the following screen will appear:

Save a copy of the current **SPSS for Windows 7** file on your P: Drive or your pen drive, under **Drives:** click on **V**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.

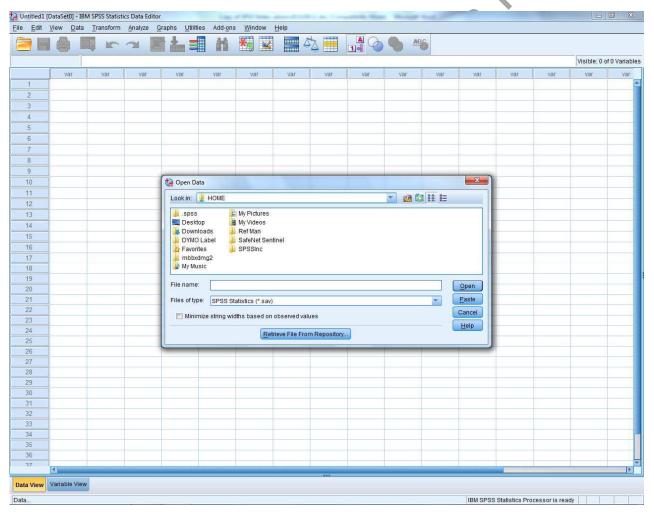


Figure 11

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.

Sam

			Reports		*		in the second second	- 42 III			ABG				
	Name	Туре	Tables	e Statistics			uencies	Missing	Columns	Align	Measure	Role			
1	IDNO	Numeric	Compare	Heene		Desi	criptives	None	4	端 Right	\delta Nominal	S Input			
2	GROUP	Numeric				A Explo	ore	None	5	端 Right	🚴 Nominal	N Input			
3	AGE	Numeric		inear Model		Cros	stabs	None	5	遍 Right	& Nominal	S Input			
4	DTBIRTH	Date	and an and a second second second	ed Linear Models	1	Ratio)	None	10	Right	🙈 Nominal	> Input			
5	DTASSMNT	Date	Mixed Mod	Jeis	1	P-P I	Plots	None	10	Right	\delta Nominal	S Input			
6	DTEMPLMT	Date	Correlate		1		Plots	None	10	Right	\delta Nominal	S Input			
7	SEX	Numeric	Regressio	on			{u, maie}	None	3	Right	\delta Nominal	🔪 Input			
8	HT	Numeric	L <u>o</u> glinear			n cms	None	None	4	Right	🙈 Nominal	> Input			
9	FEVMEAS	Numeric	Classify		*	ed FEV	None	None	6	I Right	🚴 Nominal	S Input			
10	FEVPRED	Numeric		n Reduction	P.	ed FEV	None	None	6	📲 Right	🚴 Nominal	S Input			
11	FVCMEAS	Numeric	Scale		۴	ed FVC	None	None	6	疆 Right	🙈 Nominal	N Input			
12	FVCPRED	Numeric	Nonparan	Nonparametric Tests Image: Constraint of the second seco	۴	ed FVC	None	None	6	疆 Right	🙈 Nominal	> Input			
13	ASTHMA	Numeric	Forecastir		۶	d Asthma	{0, No}	None	5	🚟 Right	🙈 Nominal	S Input			
14	BRON	Numeric	Survival		Response 🕨		d Bronc	{0, No}	None	5	🗏 Right	\delta Nominal	💊 Input		
15	SMKNOW	Numeric	Multiple R				smoke	{0, No}	None	7	遍 Right	\delta Nominal	🔪 Input		
16	SMKEVER	Numeric	Missing V		8		u ever s	{0, Never}	None	6	🗃 Right	🙈 Nominal	💊 Input		
17	EMPYRS	Numeric	Multiple In	1000 T		ears wit	None	None	7	🚟 Right	🗞 Nominal	💊 Input			
18	CIGNO	Numeric	Complex	Samples) }			garettes	None	-88	5	🗃 Right	🚴 Nominal	💊 Input	
19 20	CIGYRS	Numeric Numeric	Quality Co				ears sm		-88 None	5	Right	Sominal Nominal	Input		
20	HOWOLD	Numeric	ROC Cury	e		exposu	None	None	10	Right Right	Nominal	Input			
22 23															
24															
25															
26 27															
28															
29 30	-														
31			-												
32 33	_														
33 34															
35															
36 37	-		-												
37															

The following screen will appear.

Date of appointm Sex of the patient Height in cms [HT] Measured FEV/IF Display frequency tables OK Paste Reset Cancel Help	Bootstrap	
--	-----------	--

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 \square . 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.

			leight in chis		
		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	

Height in cms

	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 Skewr	ness	.208
Kurtosis		.393
Std. Error of Kurtos	iis	.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.

	Variable(s):	Options
b Ever had Bronchit 📤 Do you smoke no	Age at the interview [& Height in cms [HT]	Style
Alave you ever s	💑 Measured FEV [FEV	Bootstrap
No of years with c	Measured FVC [FVC	
No of cigarettes p		
Current exposure		
smknow=1 (FILT		
Save standardized values a	s variables	
OK Pas	te <u>R</u> eset Cancel Help	
×	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 **source** by the **Row(s)** box to make this the row variable

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

	Row(s):	Exact
Height in cms [HT]	Have you ever smok	ed
& Measured FEV [FEVM		Statistics
Redicted FEV [FEVPR		C <u>e</u> lls
🗞 Measured FVC (FVCM	Column(s):	Eormat
Predicted FVC [FVCP	Exposure Group [GR	Style
Ever had Asthma [AST Ever had Bronchitis [B		
Do you smoke now [S	Layer 1 of 1	Bootstrap
No of years with comp	(manual)	
💫 No of cigarettes per d	Previous	
💦 No of years smoked [
Current exposure to d	*	
🗞 smknow=1 (FILTER) [🗞 HOWOLD 🔤		
	Display layer variables in table	e layers
Display clustered <u>b</u> ar charts		
Suppress tables		
ОКП	aste Reset Cancel Help	
UN	aste Reset Cancel Heip	

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

1

Have you ever smoked * Exposure Group Crosstabulation

	Count				
6			Exposure	e Group	
				Exposure	
X			Unexposed	to Dust	Total
v	Have you	Never	24	20	44
	ever smoked	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 ev	ver smoked * Exposure Gro	oup Crosstabi	ulation	
			Exposure	e Group	
			Unexposed	Exposure to Dust	Total
Have you	Never	Count	24	20	44
ever smoked		% 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 \rightarrow Click Here Assistan

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.

	Row(s):	Exact
🚴 Identification No [IDNO]	💦 💦 Have you ever smoked	
Age at the interview [A		Statistics
Date of birth [DTBIRTH]		C <u>e</u> lls
Date of assessment [<u>C</u> olumn(s):	Format
Height in cms [HT]	Exposure Group [GRO	Style
Measured FEV [FEVM		
Predicted FEV [FEVPR	Layer 1 of 1	Bootstrap
Measured FVC [FVCM	Layer I of T	
💫 Predicted FVC [FVCP	Previous	S.
💫 Ever had Asthma [AST		1
Ever had Bronchitis [B	Sex of the patient [SEX]	3
Do you smoke now [S		
No of years with comp		
	📃 Display layer variables in table lay	ers
Display clustered <u>bar chart</u>		
Suppress tables		
-		

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.

Sex of the patient			-	Exposur	e Group	Total
					Exposure	
			-	Unexposed	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%

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

 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 GI 1 1001 2 1002 3 1003 4 1004 5 1005 6 1006 7 1007 8 Cut 9 Copy 11 Paste 12 Clear 13 Insert Cas 14 1015 15 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1026 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		9	12.10.1952 01.11.1956 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	DTASSMNT 12.06.1995 24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988 29.06.1996	DTEMPLMT 12.02.1972 10.08.1982 18.10.1978 24.06.1980 24.03.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982 05.03.1979	SEX 1 1 1 0 0 0 0 1	HT 175 168 180 180 183 174 180		FEVPR 3.59 3.39 4.26 4.25 4.52 3.73 4.45	FVCME 4.49 3.91 4.80 4.57 6.50 5.82	FVCPR 4.45 4.12 5.14 5.12 5.42 4.54	ASTH 0 1 0 0 0	BRON 0 1 0 0 0	1 1 0 1	R 2 2	Visible: 21 o EMPYRS 23 16 12 12	810-21-20-21-24
IDNO Gi 1 1001 2 1002 3 1003 4 1004 5 1005 6 1006 7 1007 8 Cut 9 Cut 11 Paste 12 Clear 13 Insert Cas 14 1015 15 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031	GROUF	P AGE 1 45 1 45 1 45 1 45 1 45 1 22 1 45 25 37 35 28 28	DTBIRTH 29.04.1946 12.10.1952 01.11.1952 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	DTASSMNT 12.06.1995 24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	DTEMPLMT 12.02.1972 10.08.1982 18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	SEX 1 1 1 0 0 0 0 1	HT 175 168 180 180 183 174 180	FEVME 3.40 2.83 3.93 4.01 4.75 4.60	FEVPR 3.59 3.39 4.26 4.25 4.52 3.73	4.49 3.91 4.80 4.57 6.50	FVCPR 4.45 4.12 5.14 5.12 5.42	0 1 0 0	0 1 0 0	1 1 0 1	R 2 2 0	EMPYRS 23 16 12	CIGNO 20 20 -88
IDNO Gi 1 1001 2 2 1002 3 3 1003 4 5 1005 6 7 1007 8 9 Cut 2 10 Qopy Paste 12 Clgar 1016 15 1016 1017 17 1018 1019 18 1019 1020 20 1021 22 21 1022 22 22 1023 23 23 1025 24 24 1026 25 25 1027 26 28 1030 29	GROUF	P AGE 1 49 1 49 1 49 1 34 0 34 0 29 1 43 1 20 2 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	29.04.1946 12.10.1952 01.11.1952 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	12.06.1995 24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	12.02.1972 10.08.1982 18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	1 1 0 0 0 0 1	175 168 180 180 183 174 180	3.40 2.83 3.93 4.01 4.75 4.60	3.59 3.39 4.26 4.25 4.52 3.73	4.49 3.91 4.80 4.57 6.50	4.45 4.12 5.14 5.12 5.42	0 1 0 0	0 1 0 0	1 1 0 1	R 2 2 0	EMPYRS 23 16 12	CIGNO 20 20 -88
1 1001 2 1002 3 1003 4 1004 5 1005 6 1006 7 1007 8 Cut 9 Copy 101 Copy 101 Copy 11 Copy 12 Clear 13 Insert Cas 14 1015 15 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		1 45 1 46 1 32 0 32 0 25 1 43 1 27 55 22 33 35 28 28	29.04.1946 12.10.1952 01.11.1952 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	12.06.1995 24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	12.02.1972 10.08.1982 18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	1 1 0 0 0 0 1	175 168 180 180 183 174 180	3.40 2.83 3.93 4.01 4.75 4.60	3.59 3.39 4.26 4.25 4.52 3.73	4.49 3.91 4.80 4.57 6.50	4.45 4.12 5.14 5.12 5.42	0 1 0 0	0 1 0 0	1 1 0 1	R 2 2 0	23 16 12	20 20 -88
1 1001 2 1002 3 1003 4 1004 5 1006 6 1006 7 1007 8 Cut 9 Cut 11 Paste 12 Clear 13 Insert Cas 14 1016 15 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1028 28 1030		1 45 1 46 1 32 0 32 0 25 1 43 1 27 55 22 33 35 28 28	29.04.1946 12.10.1952 01.11.1952 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	12.06.1995 24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	12.02.1972 10.08.1982 18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	1 1 0 0 0 0 1	175 168 180 180 183 174 180	3.40 2.83 3.93 4.01 4.75 4.60	3.59 3.39 4.26 4.25 4.52 3.73	4.49 3.91 4.80 4.57 6.50	4.45 4.12 5.14 5.12 5.42	0 1 0 0	0 1 0 0	1 1 0 1	R 2 2 0	23 16 12	20 20 -88
2 1002 3 1003 4 1004 5 1005 6 1006 7 1007 8 Cut 9 Copy 11 Clear 13 Insert Cas 14 1015 15 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		1 440 1 34 0 34 0 25 1 43 1 27 55 25 33 34 28 28	12.10.1952 01.11.1956 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	10.08.1982 18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	1 1 0 0 0 0 1	168 180 180 183 174 180	2.83 3.93 4.01 4.75 4.60	3.39 4.26 4.25 4.52 3.73	3.91 4.80 4.57 6.50	4.12 5.14 5.12 5.42	1 0 0	1 0 0	1 0 1	2 2 0	16 12	20 -88
2 1002 3 1003 4 1004 5 1005 6 1006 7 1007 8 Cut 9 Copy Paste Clear 13 Inset cas 14 1015 15 1016 16 1017 17 1018 18 1019 1020 20 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1033		1 440 1 34 0 34 0 25 1 43 1 27 55 25 33 34 28 28	12.10.1952 01.11.1956 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	24.12.1998 31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	10.08.1982 18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	1 1 0 0 0 0 1	168 180 180 183 174 180	2.83 3.93 4.01 4.75 4.60	3.39 4.26 4.25 4.52 3.73	3.91 4.80 4.57 6.50	4.12 5.14 5.12 5.42	1 0 0	1 0 0	1 0 1	2 0	16 12	-8
3 1003 4 1004 5 1006 6 1006 7 1007 8 Cut 9 Copy Paste Cigar 13 Egar 13 Elast Cas 14 1015 15 1016 16 1017 17 1018 18 1019 1020 20 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1032 29 1031		1 34 0 34 0 29 1 43 1 27 59 29 39 39 20	01.11.1956 05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	31.10.1990 09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	18.10.1978 24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	1 0 0 0 0	180 180 183 174 180	3.93 4.01 4.75 4.60	4.26 4.25 4.52 3.73	4.80 4.57 6.50	5.14 5.12 5.42	0	0	0	0	12	-8
4 1004 5 1005 6 1006 7 1007 8 Cut 9 Cut 101 Paste 12 Clear 13 Inset Cas 14 1015 15 1016 16 1017 17 1018 18 1019 1020 1021 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1039 29 1031		0 34 0 29 1 43 1 27 29 39 39 39 28	05.04.1958 12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	09.09.1992 06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	24.06.1980 05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	0 0 0 0	180 183 174 180	4.01 4.75 4.60	4.25 4.52 3.73	4.57 6.50	5.12 5.42	0	0	1			
5 1005 6 1006 7 1007 8 Cut 9 Copy 101 Paste 102 Clear 11 Paste 12 Clear 131 Inset Cas 1016 1016 16 1017 17 1018 18 1019 1020 1021 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030		0 29 1 43 1 27 55 29 34 36 28	12.03.1960 25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	06.04.1989 21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	05.05.1982 24.03.1982 24.01.1983 08.02.1965 04.02.1982	0 0 0 1	183 174 180	4.75 4.60	4.52 3.73	6.50	5.42				2	12	
6 1006 7 1007 8 Cut 9 Cut 100 Paste 11 Paste 12 Clear 13 ■ Inset Cas 14 1016 15 1016 16 1017 17 1018 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1031		1 43 1 27 29 39 39 28	25.06.1947 10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	21.07.1990 15.03.1991 10.02.1987 04.01.1991 07.05.1988	24.03.1982 24.01.1983 08.02.1965 04.02.1982	0 0 1	174 180	4.60	3.73	1000		0	0			-	
7 1007 8 9 Cut 9 Cut Copy 10 Copy Clear 13 Imsert Cas Imsert Cas 14 1016 16 16 1017 1018 18 1019 1020 20 1021 21 21 1022 22 22 1023 23 23 1025 24 25 1027 26 26 1028 27 26 1029 28 29 1031 29		1 27 29 3 3 28	10.02.1964 11.01.1928 01.01.1962 08.02.1957 31.03.1961	15.03.1991 10.02.1987 04.01.1991 07.05.1988	24.01.1983 08.02.1965 04.02.1982	0	180			5.82				0	5	7	-81
8 Cut 9 Cut 10 Copy 11 Paste 12 Clear 13 Imsert Cas 14 T015 15 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		29 29 31 36 28	11.01.1928 01.01.1962 08.02.1957 31.03.1961	10.02.1987 04.01.1991 07.05.1988	08.02.1965 04.02.1982	1		4.01	1 15			0	0	0		8	2
9 Cut 9 Copy 11 Paste 12 Clear 13 ■ Insert Cas 14 1015 15 1016 16 1017 17 1018 1019 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031	ases	29 31 38 28	01.01.1962 08.02.1957 31.03.1961	04.01.1991 07.05.1988	04.02.1982		167		4.40	4.90	5.30	0	0	0	0	8	-8
9 Copy 0 Paste 12 Clgar 13 Insert cas 14 1015 15 1016 16 1017 17 1018 18 1019 1020 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031	ases	31 38 28	08.02.1957 31.03.1961	07.05.1988		1		2.58	2.97	3.68	3.73	0	0	1	2	22	3
Image: book of the state of the s	ases	35 28	31.03.1961		05.03.1979		175	4.50	4.18	5.68	4.97	0	0	0	1	9	2
Image: Constraint of the sector of	ases	28		29.06.1996		1	177	4.19	4.21	5.61	5.03	0	0	1	2	9	2
Image: Clear product Image: Clear product Image: Clear product	ases	28		2010011000	24.02.1981	0	173	3.51	3.92	4.66	4.69	0	0	1	2	15	2
Image: Section 2016 Image: Section 2016 101 1015 101 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 255 1027 26 1028 27 1029 28 1030 29 1031	ases		24.02.1500	31.03.1994	23.05.1986	0		2.92	3.91	4.09	4.59	1	0	1		8	4
Interference 4 1015 5 1016 6 1017 17 1018 18 1019 19 1020 102 1021 11 1022 1023 1025 122 1025 1025 1027 105 1027 106 1028 17 1030 19 1030	ases		29.06.1958	12.07.1992	10.06.1984	1		3.18	4.03	3.61	4.84	0	0	0		8	-8
5 1016 16 1017 17 1018 18 1019 19 1020 20 1021 21 1023 22 1023 23 1025 24 1026 255 1027 266 1028 27 1029 288 1030 29 1031			1.														
1017 1018 1019 1020 200 1021 21 1022 2023 1023 23 1025 24 1027 26 1027 26 1027 26 1028 27 1029 1031		0 51		25.02.1987	23.03.1982	0		2.76	3.24	4.21	3.99	0	1	1		5	2
17 1018 18 1019 19 1020 19 1021 21 1022 22 1023 23 1026 244 1026 25 1027 26 1028 27 1028 28 1030 29 1031		0 49		19.04.1995	10.04.1987	0	175	3.06	3.59	4.66	4.45	0	0	0		8	2
18 1019 19 1020 20 1021 21 1022 22 1023 23 1026 24 1026 27 1028 27 1028 27 1029 28 1030 29 1031		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	-8
19 1020 20 1021 21 1022 22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		1 5		20.10.1990	11.08.1967	1		3.77	3.24	4.40	3.99	0	0				4
20 1021 21 1022 22 1023 33 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		1 34 0 30		13.08.1993	24.06.1979				3.82	4.80	4.55	1	0		-		2
21 1022 22 1023 23 1025 24 1026 25 1027 266 1028 277 1029 28 1030 29 1031		0 32		21.05.1996 18.12.1991	18.03.1988 22.10.1976			4.03	4.44 3.99	5.14 5.38	5.35 4.99	0	1	0			4
22 1023 23 1025 24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		1 40		03.10.1989	18.09.1980		170		3.33	5.13	4.33	0	0				-8
24 1026 25 1027 26 1028 27 1029 28 1030 29 1031		0 49		21.07.1997	12.07.1982			3.32	3.17	4.68	3.87	0	0				2
25 1027 26 1028 27 1029 28 1030 29 1031		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	-8
26 1028 27 1029 28 1030 29 1031		1 40	17.04.1949	23.06.1995	25.06.1990	0	175	4.01	3.59	5.17	4.45	0	0	0	0		-8
27 1029 28 1030 29 1031		1 50		17.04.1998	18.03.1991					3.57	3.69	0	0				2
28 1030 29 1031		1 20		10.01.1996	19.01.1988				4.14	4.58	4.87	0	0			8	-8
29 1031		1 54		12.09.1988	23.07.1979					4.51 5.92	4.03	0	1	0			2
		1 32		28.07.1990 15.03.1994	14.06.1983 24.01.1985			4.68 4.91	4.22	5.92 6.06	5.06 5.69	1	0				-8 1
		0 50		20.01.1992	01.01.1976				3.36	3.88	4.13	0	1	0	-		3
31 1032		1 53		18.11.1995	16.10.1982					3.60	3.61	0	0				2
1034		0 52		26.05.1997	20.01.1988					4.70	4.94	0	0				4
33 1036		0 42		02.04.1989	12.02.1977		162	3.64	3.24	4.59	3.88	0	0	1	2	12	1
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	-8
35 1038		0 45		19.09.1992	31.03.1983			4.31	3.50	5.50	4.26	0	0				2
36 1039		1 30	15.01.1953	22.01.1991	25.01.1974	1	170	3.98	3.72	5.11	4.46	0	0	0	1	17	1
View Variable Vie																	
														tatistics Proc			

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.

e <u>E</u> dit !	View [ata <u>T</u> ra	nsform	Analyze Graphs	Utilities Add-o	ns <u>W</u> indow <u>H</u>	lelp											
a H			5	<u>→</u> 🖺		*		5		A Q		ARG						
ASTHMA		0															Visible: 21 of	f 21 Vari:
	IDNO	GROUP	AGE	DTBIRTH	DTASSMNT	DTEMPLMT	SEX	HT	FEVME	FEVPR	FVCME	FVCPR	ASTH	BRON SMK	NOW SI	MENTE	EMPYRS	CIGNO
	IDINO	UNCON	AUL	DIDIKITI	DIAGONINI	DIEW EWI	ULA			1 E VI IS	I VONL	i vor i	Aonta.	DICON ONIC	1011 31	R		CICINO
1	1001	1	49	29.04.1946	12.06.1995	12.02.1972	1	175	3.40	3.59	4.49	4.45		Cut	-	2	23	2
2	1002	1	46	12.10.1952	24.12.1998	10.08.1982	1	168		3.39	3.91	4.12		Сору	-	2	16	2
															-			
3	1003	1	34	01.11.1956	31.10.1990	18.10.1978	1	180	3.93	4.26	4.80	5.14		Paste		0	12	-8
4	1004	0	34	05.04.1958	09.09.1992	24.06.1980	0	180	4.01	4.25	4.57	5.12		Clear		2	12	2
5	1005	0	29	12.03.1960	06.04.1989	05.05.1982	0	183	4.75	4.52	6.50	5.42		Insert Variable		0	7	-8
6	1006	1	43	25.06.1947	21.07.1990	24.03.1982	0	174	4.60	3.73	5.82	4.54		insen v <u>a</u> nable		1	8	2
7	1007	1	27	10.02.1964	15.03.1991	24.01.1983	0	180	1000	4.45	4.90	5.30		Sort Ascending		0	8	-8
							-							Sort Descending	-	~		
8	1009	1	59	11.01.1928	10.02.1987	08.02.1965	1	167	2.58	2.97	3.68	3.73	der			2	22	3
9	1010	1	29	01.01.1962	04.01.1991	04.02.1982	1	175	4.50	4.18	5.68	4.97	-0	Spelling		1	9	2
10	1011	1	31	08.02.1957	07.05.1988	05.03.1979	1	177	4.19	4.21	5.61	5.03	C	0	1	2	9	2
11	1012	1	35	31.03.1961	29.06.1996	24.02.1981	0	173	3.51	3.92	4.66	4.69	C	0	1	2	15	2
12	1013	1	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	4
13	1014	0	34	29.06.1958	12.07.1992	10.06.1984	1	175		4.03	3.61	4.84	0		0	0	8	-6
																č		
14	1015	0	51	31.01.1936	25.02.1987	23.03.1982	0	168		3.24	4.21	3.99	C		1	2	5	2
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	2
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	-8
17	1018	1	51	23.09.1939	20.10.1990	11.08.1967	1	168	3.77	3.24	4_40	3.99	C	0	0	1	23	4
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	2
19	1020	0		20.02.1964	21.05.1996	18.03.1988	0	183		4.44	5.14	5.35	C		0	1	8	
20	1021	1		16.10.1941	18.12.1991	22.10.1976	0	185		3.99	5.38	4.99	C		0	1	15	4
21	1022	1		05.09.1943	03.10.1989	18.09.1980	1	170		3.47	5.13	4.24	0		0	0	9	-8
22 23	1023 1025	0		06.06.1948	21.07.1997 16.05.1994	12.07.1982	0	165 170		3.17	4.68	3.87			1	2	15 6	2 -8
23	1025	1		09.02.1949	23.06.1995	12.05.1988 25.06.1990	0	175		3.59	5.17	4.45			0	0	5	-6
25	1020			10.01.1942	17.04.1998	18.03.1991	1	165		2.97	3.57	3.69			0	1	7	2
26	1028	1		01.01.1970	10.01.1996	19.01.1988	1	172		4.14	4.58	4.87			0	0	8	-8
27	1029	1		19.04.1934	12.09.1988	23.07.1979	1	170		3.24	4.51	4.03	c		0	1	9	2
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	-8
29	1031	1	34	20.01.1960	15.03.1994	24.01.1985	1	190	4.91	4.68	6.06	5.69	c	0	1	2	9	81
30	1032	0		02.01.1942	20.01.1992	01.01.1976	1	170	2.47	3.36	3.88	4.13	C	1	0	1	16	3
31	1033	1		10.10.1942	18.11.1995	16.10.1982	0	163		2.94	3.60	3.61	C		1	2	13	2
32	1034	0		09.04.1945	26.05.1997	20.01.1988	0	185		3.94	4.70	4.94		-	0	1	9	4
33	1036	0		16.02.1947	02.04.1989	12.02.1977	1	162		3.24	4.59	3.88	C		1	2	12	1
34	1037	0		17.01.1959	21.03.1993	28.02.1987	0	177		4.12	5.12	4.95	0		0	0	6	-8
35 36	1038	0		26.06.1947	19.09.1992	31.03.1983	0	170		3.50	5.50	4.26	0	0	1	2	9	2
36	1039	1	38	15.01.1953	22.01.1991	25.01.1974	1	170	3.98	3.72	5.11	4.46	C	U	0	1	1/	1

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:-

Compute Variable Target Variable: Type & Label Compute Variable: Type & Label Composed of the latent Composed of the latent of the latent Composed of the latent Compo	$=$ Num <u>eric Expression:</u> $+ < 2789$ $- < 2789$ $- < 2456$ $* = - = 123$ $/ & 10$ $* \sim () Delete$	Function group: All Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arithmetic Date Creation
💫 No of cigarettes p	n condition) OK Paste Reset Cancel Help	

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 \bigcirc 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 \bigcirc

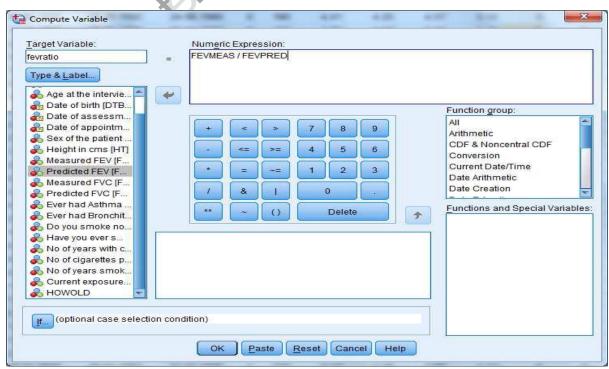
Select the function using up and down arrow key from the Built in function window and then click on the button *solution*. The expression will appear in the **Numeric Expression** window.

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			-
		То			

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

Table 8: Functions Description

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



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.

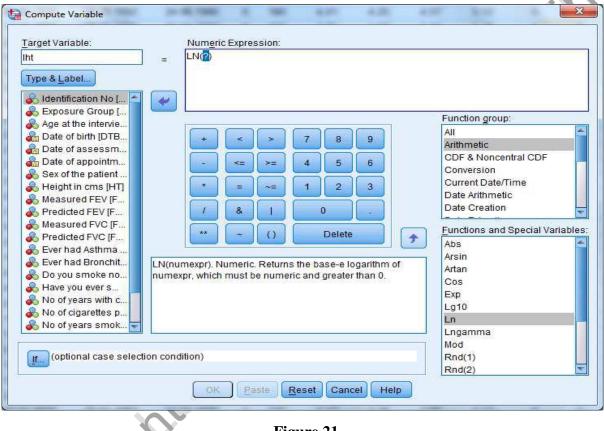


Figure 21

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.

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

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.

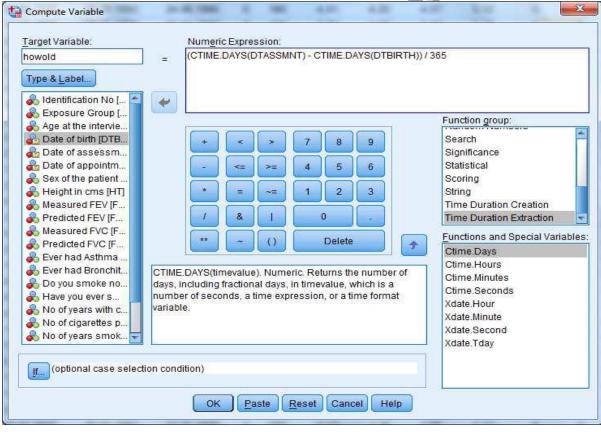


Figure 22

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.

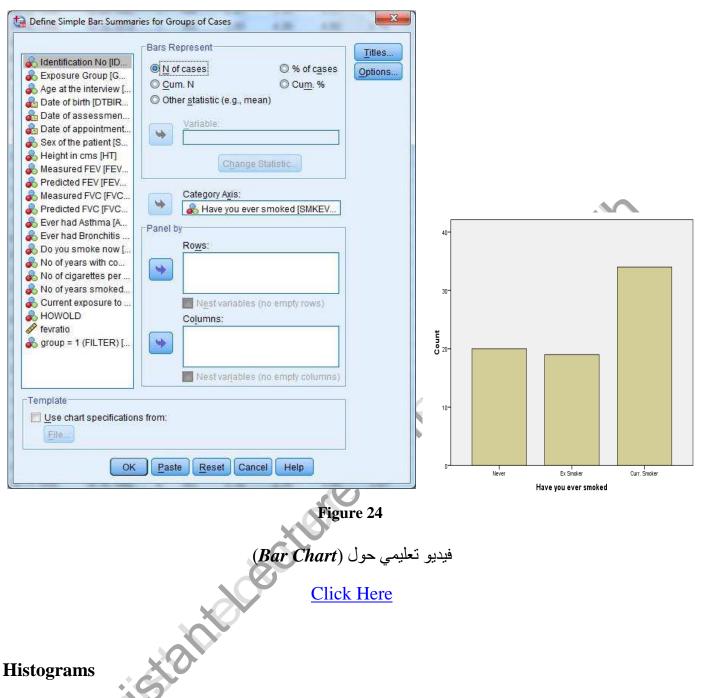
Bar Charts

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.

						ABG	•			52	Help		Utilities Add-o	Analyze Graphs		Data Tra		le <u>E</u> dit
23 Variabl	Visible: 23 of													500 S (0, 500		0		BRON
CIGNO	EMPYRS	SMKEVE R	SMKNOW	BRON	ASTH	FVCPR	VCME	FEVPR F	FEVME	HT	SEX	DTEMPLMT	DTASSMNT	DTBIRTH	AGE	GROUP	IDNO	
20	23	2	1	0	0	4.45	4.49	3.59	3.40	175	1	12.02.1972	12.06.1995	29.04.1946	49	1	1001	1
20	16	2	1	1	1	4.12	3.91	3.39	2.83	168	1	10.08.1982	24.12.1998	12.10.1952	46	1	1002	2
-88	12		0	0	0	5.14	4.80	4.26	3.93	180	1	18.10.1978	31.10.1990	01.11.1956	34	1	1003	3
25	12		1	0		5.12	4.57	4.25	4.01	180		24.06.1980	09.09.1992	05.04.1958	34	0	1004	4
-88	7		0	0		5.42	6.50	4.52	4.75	183		05.05.1982	06.04.1989	12.03.1960	29	0	1005	-6
20	8	2.	0	0	0	4.54	5.82	3.73	4.60	174		24.03.1982	21.07.1990	25.06.1947	43	1	1006	6
-88	8		0	0		5.30	4.90	4.45	4.01	180		24.01.1983	15.03.1991	10.02.1964	27	1	1007	7
30	22	~	1	0		3.73	3.68	2.97	2.58	167	1	08.02.1965	10.02.1987	11.01.1928	59		1009	8
20	9	22	0	0	0	4.97	5.68	×		arts	Bar Ch		04.01.1991	01.01.1962	29	1	1010	9
20	9		1	0		5.03	5.61			1		05.03.1	07.05.1988	08.02.1957	31	1	1011	10
20	15	~	1	0		4.69	4.66		e	Simpl	I .	24.02.1	29.06.1996	31.03.1961	35		1012	11
40	8		1	0		4.59	4.09					23.05.1	31.03.1994	24.02.1966	28	1	1013	12
-88	8		0		0	4.84	3.61			-		10.06.1	12.07.1992	29.06.1958	34	0	1014	13
20	5		1	1		3.99	4.21		ered	Cluste		23.03.1	25.02.1987	31.01.1936	51	0	1015	14
20	8		0	0		4.45	4.66			-		10.04.1	19.04.1995	29.01.1946	49		1016	15
-88	8		0	0	1	4.97	5.29		ed	Stack		24.01.1	07.01.1996	02.02.1967	29	0	1017	16
40	23		0	0		3.99	4.40					11.08.1	20.10.1990	23.09.1939	51	1	1018	17
20	14		1	0		4.55	4.80		vre	Chart A	-Data ir	24.06.1	13.08.1993	05.06.1959	34	1	1019	18
5	8		0	0		5.35	5.14	fenene	for groups (amariae	O Sun	18.03.1	21.05.1996	20.02.1964	32	0	1020	19
40	15		0	1		4.99	5.38	A DESCRIPTION OF	of separate			22.10.1	18.12.1991	16.10.1941	50	1	1021	20
-88	9		0	0	0	4.24	5.13	-	dividual cas			10.03.1	03.10.1989	05.09.1943	46	1	1022	21
20	15		1	0	0	3.87	4.68			10000480		12.07.1	21.07.1997	06.06.1948	49	0	1023	22
-88	6	~	0	0		4.26	4.34	ielp	Cancel	afine	D	12.05.1	16.05.1994	09.02.1949	45	0	1025	23
-88	5		0	0		4.45	5.17				(com	25.06.1	23.06.1995	17.04.1949	46		1026	24
20	7		0	0		3.69	3.57	2.97	2.80	165		18.03.1991	17.04.1998	10.01.1942	56	1	1027	25
-88	8		0	0	0	4.87	4.58	4.14	4.37	172		19.01.1988	10.01.1996	01.01.1970	26	1	1028	26
20	9		0	1		4.03	4.51	3.24	3.63	170		23.07.1979	12.09.1988	19.04.1934	54	1	1029	27
-88	7	100	0	0		5.06	5.92	4.22	4.68	178		14.06.1983	28.07.1990	12.05.1958	32	1	1030	28
12	9	61	1	0		5.69	6.06	4.68	4.91	190		24.01.1985	15.03.1994	20.01.1960	34	1	1031	29
30	16	2.	0	1		4.13	3.88	3.36	2.47	170		01.01.1976	20.01.1992	02.01.1942	50	0	1032	30
20	13		1	0	0	3.61	3.60	2.94	2.16	163		16.10.1982	18.11.1995	10.10.1942	53		1033	31
40	9		0	0	0	4.94	4.70	3.94	3.53	185		20.01.1988	26.05.1997	09.04.1945	52	0	1034	32
10	12	10	1	0		3.88	4.59	3.24	3.64	162		12.02.1977	02.04.1989	16.02.1947	42		1036	33
-88	6		0	0		4.95	5.12	4.12	3.69	177	0	28.02.1987	21.03.1993	17.01.1959	34	0	1037	34
20	9	-	1	0		4.26	5.50	3.50	4.31	170		31.03.1983	19.09.1992	26.06.1947	45	0	1038	.35
13	17	1	0	0	0	4.46	5.11	3.72	3.98	170	1	25.01.1974	22.01.1991	15.01.1953	38	1	1039	36

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**.



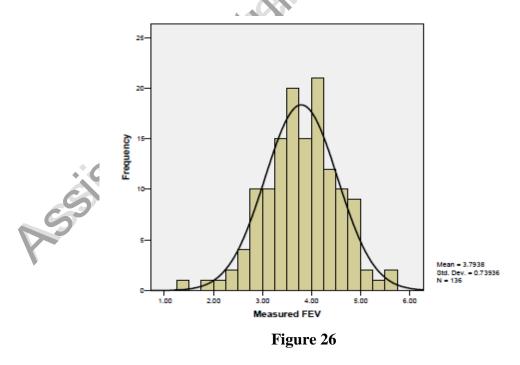
At this point it is a good idea to return the select cases back to all data, by **Data, Select Cases**, and then **All Cases** followed by **ok**.

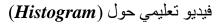
Histograms are **produced for interval variables** e.g. age. To produce a histogram click on **Graphs, Legacy Dialogs** then **Histogram** and the following screen appears.

💰 Identification No [🗲	Variable:	Titles
Exposure Group [Age at the intervie Date of birth [DTB	Display normal curve Panel by	
Date of bittingDTB Date of assessm Date of appointm	Rows:	
Sex of the patient		
Redicted FEV [F	Nestvariables (no empty rows)	
Predicted FVC [F Ever had Asthma	Columns:	
& Ever had Bronchit Do you smoke no		
Have you ever s	Nest variables (no empty columns)	
Template	ns from:	
OK	Paste Reset Cancel Help	

Click on the required variable, in this case FEV, in the left hand side list and press **N** 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.





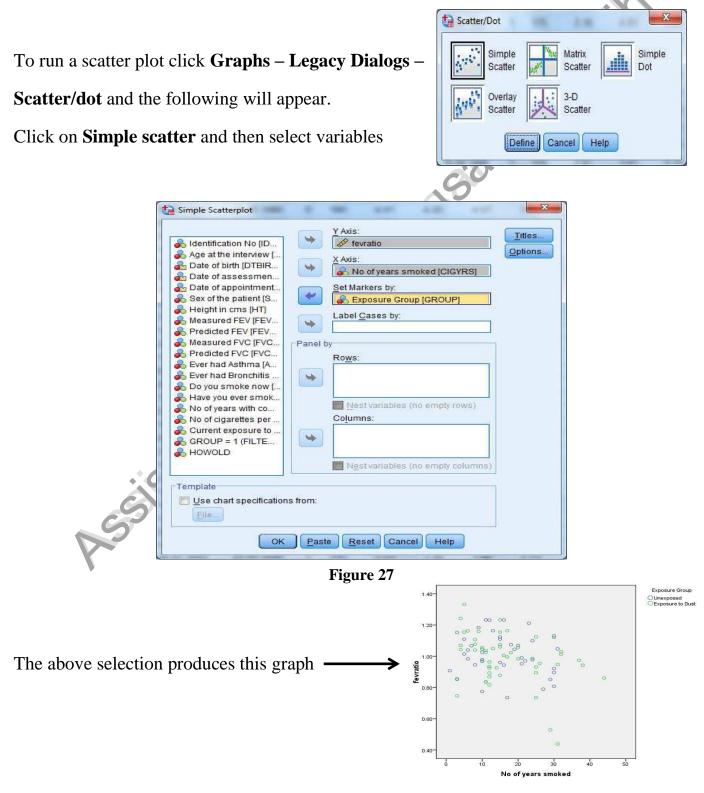
Click Here

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:

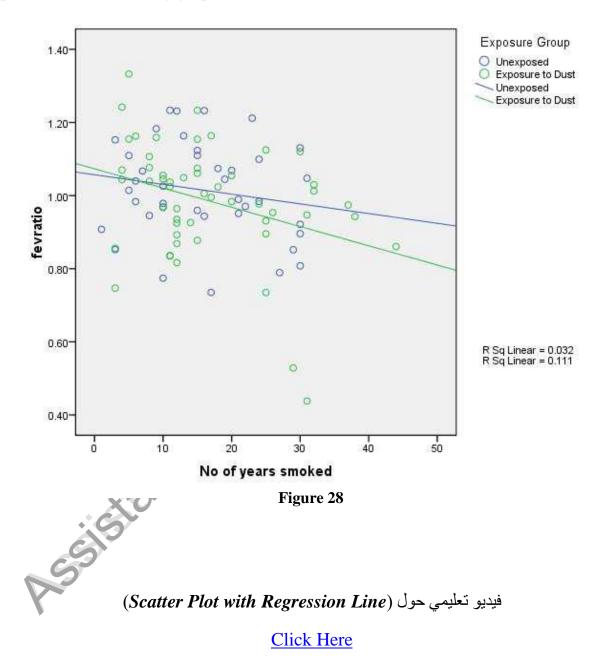
- 1- The value of the dependent or response variable on the y axis.
- 2- The value of the independent variable on the horizontal axis.



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.



Chapter Four Exercises

4.1 <u>Exercise 1</u>:

A- Set-up the following variables:

1	tono wing 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	19	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		1= yes
C	Drive?		2=no
Exercise	How often you do the		1= once a week
	Sport exercises?		2= twice a week
			3= not often
			4= never

B- Fill up the form for the 15 students.

<u>4.2 Exercise 2</u>: 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?

<u>4.3 Exercise 3</u>: 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)

<u>4.5 Exercise 5</u>: 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