INTRODUCTION TO STATSDIRECT

PART 1	2
INTRODUCTION	2
Why Use StatsDirect	2
ACCESSING STATSDIRECT FOR WINDOWS XP	4
DATA ENTRY	5
Missing Data	6
Opening an Excel Workbook	6
Moving around Data Editor	8
INTRODUCING ANALYSIS COMMANDS	8
Descriptive Statistics	8
Analysis By Column Or By Identifier	8
Frequency Tables	9
Descriptives	11
Cross-tabulation, Chi-squared Test and Fisher's Exact Test	13
FILE HANDLING	17
Saving The Work	17
Backup The Work	18
Retrieving a StatsDirect File	18
MODIFYING THE DATA	19
Modifying a new variable by Search and Replace	19
PART II	22
CONTINUOUS OUTCOME MEASURES	22
Histogram	22
Comparison of Means Using a t-test	25
REARRANGING DATA	26
Sorting Data	26
Group Split	27
EXAMINING THE RELATIONSHIP BETWEEN TWO CONTINUOUS VARIABLES	29
Scatter Plots	29
Linear Regressions	30
Model Checking	32
NON-PARAMETRIC METHODS	34
COMPARISONS OF RELATED OR PAIRED VARIABLES	36
Continuous Outcome Measures	36
Analysis of Binary Outcomes that are Related	37
	38
SUMMARY STATISTIC METHODS	39
t-test Using summary Data	39
	41
CHOOSING THE APPROPRIATE STATISTICAL PROCEDURE	42

PART 1

INTRODUCTION

This handbook designed to introduce **StatsDirect**. It assumes familiarity with Microsoft windows and standard windows-based office productivity software such as word processing and spreadsheets.

Why Use StatsDirect

StatsDirect is an easy-to-use package designed for medical researchers. It uses an interface similar to Micro-soft Excel and is able to read **EXCEL** Workbooks. It has some procedures specifically designed for medical researchers. **StatsDirect** has an extensive **Help** Facility that can be accessed by clicking on the local Help Tile. The help provided depends on what part of the system is being used.

StatsDirect is not suitable for analysis of large data sets with many variables and missing data. If your dataset is large and has missing data and many variables, as you might have with a survey or data base, you may find it easier to use SPSS or STATA. For example in SPSS it is easier to select subsets of data for a particular analysis. In SPSS variables can be labelled and text labels can be added to numerical codes making statistical output much more readable. This may be important when analysing large data sets.

This tutorial will assume some familiarity with concepts in statistical inference including hypothesis testing and confidence intervals. This was covered in years 1 and 2 of the medical undergraduate curriculum as part of the Information stream of the course. If you are unfamiliar with these concepts, it is suggested that you read an introductory text in medical statistics such as Campbell and Machin "Medical Statistics A Common Sense Approach". Some examples are given at the Medical Statistics support web site at

http://research.bmh.manchester.ac.uk/biostatistics/teaching/statisticalsupport

StatsDirect has different type of analysis command

 Commands are based on raw data of individual subjects. Data needs to be entered into a spread-sheet with each row representing each subject. This might be into an Excel spreadsheet that is opened by StatsDirect or it may be entered directly into the StatsDirect spreadsheet. We advise the former as you can enter your data, even when you have not got access to StatsDirect.

- Commands based on summary statistics. For this summary statistics data such as frequencies, mean and standard deviation are entered into a separate panel associated with that analysis. For example suppose one is comparing mean outcome for two groups of patients using a t-test. If one already has the mean standard deviation and number of patients for each group of patients, just these six pieces of information can be entered. You can also use StatsDirect to calculate confidence intervals using proportions/ percentages using frequency data. This is very useful where you do not have access to the original data for example where we are extracting data from published research.
- StatsDirect also has commands for meta-analysis, which is the statistical method used to combine summary data from several studies to obtain an overall result.

This tutorial will focus on analysis based on the first type of data. Nevertheless you may wish to use the second two forms of analysis particularly if you want to use data from published research in your study. A brief introduction is given for summary statistic methods at the end of the tutorial.

No statistical package is completely comprehensive. For example you may decide to use SPSS for the analysis of your raw data, but you might use the summary statistics methods for comparison of your results with other.

The tutorial uses a set of data from a cross-sectional survey of respiratory function and dust levels amongst foundry workers. The object of the survey was to determine whether the dust levels found in the foundries have any effect on the respiratory health of the work force.

The purpose of this study was to examine whether dust increased respiratory morbidity. In this study the measure of respiratory morbidity are "Ever had asthma", "Ever had bronchitis", "Measured FEV" and "Measured FVC". The variable "Predicted FEV" and "Predicted FVC" are the values that are expected for a person's demographic characteristics including Age, Height and Sex. Exposure to dust is measured by two variables "Exposed/Un-exposed" and dust level recorded only for exposed workers. Because smoking is a confounding factor in this study, smoking behaviour has been recorded in terms of current smoking status, smoking history, and consumption and duration of smoking.

To make the most of this tutorial, work though generating the screens shown at each stage. There are several short exercises in the text to allow you to reinforce your skills.

ACCESSING STATSDIRECT FOR WINDOWS XP

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

2009
$$\rightarrow$$
 MHS \rightarrow Cluster \rightarrow StatsDirect.



StatsDirect will now start and the screen below will appear.

There are two sections:

Create which has two radio buttons:

- New StatsDirect Workbook
- New StatsDirect Report

Open which displays data files recently accessed by **StatsDirect** from which you may select a file. Once you have made your choice clicking the **OK** will complete start-up process. The default will open the file highlighted in the Open window.

The file **test.sdw** which should be visible in the window is a work book that contains example data prepared by the developers of **StatsDirect** to illustrate the analysis commands of the program.

For the purpose of this tutorial you should select **New StatsDirect Workbook** button and click **OK** button. A spreadsheet (**Book 1** highlighted) will be displayed for data entry as shown below

latsDirect -	[Book 1]	ata Analusis	Graphics	Tools use	dow Helo										
He LUK MIS	art morniat b	ata Analysis	Graphics	TOOIS WIT	uuw neip										
ook 1	De U ente		0					 							
- 6	43 % C	E J 2	¥												
A1	-				-	2							-	-	
A	В	С	D	E	F	G	н	J	K	L	M	N	0	Р	Q
	•														
-															
-															
-															

This has the following features at the top:

Title bar Displays StatsDirect and the name of the active document

Menu bar Contains a list of menus

Document Tabs Contains a list of worksheets and reports available with current one highlighted. If you have more than one, then you can move between them by clicking on the tab.

Tool bar Provides quick access with the mouse to frequently used tools and commands for formatting text and numbers.

Formula bar Contains a cell reference and value

Below the spreadsheet grid at the bottom there is a tab to select the work sheet within the work book.

DATA ENTRY

Data can be entered directly into **StatsDirect**. Alternatively you can enter the data into an EXCEL spreadsheet and then read it into StatsDirect. Enter the data for each subject as a row of the spreadsheet and enter the data for each variable (question or measurement). Add a label for each variable the first row of each column. In case you should need to transfer the data to another statistical package such as SPSS, it is suggested that the label naming each variable should be alpha-numeric without punctuation as illustrated below. The data should be anonymous, but there should be a unique identifying number so that you can check your data against paper records.

After inputting the first 10 subjects from the foundry workers study the screen might look like that given below.



The first row of the spreadsheet contains the entire variable names. Please note that the shaded column on the left-hand side of **StatsDirect** spreadsheet with a number from 1 onwards is the line number, it can not be copied or changed. Do not depend on the grey row numbers on the left as an identifier, as the relationship between data and row number will change if the data is sorted by row.

Missing Data

It is strongly recommended that you use the character * to represent a missing value. This is different to other statistical packages such as SPSS were it is advisable to specify a numerical or STATA where . (period) is used.

Opening an Excel Workbook

You can retrieve an EXCEL workbook containing the data from the study of foundry workers and dust. This is stored on the Shared Data Area located on the desk top. Double click on the Shared Data icon which can be found on the left hand side of the desktop, follow Health Methodology Course Data and save the Excel file to your Work Space. Alternatively the File can be found by going to the following internet address

http://research.bmh.manchester.ac.uk/biostatistics/teaching/statisticalsupport

click the link foundry.sav for the dataset followed by **save** and place it in your workspace. Under the File menu select Open File and then find the file. Once open the screen will look like this

Stats	Direct - [f	oundry.xls] Format Data Ar	nalysis Granhics To	ols Window Help										
Rook	1 6	under	anapires re											
) 🗳		Bak Ba ∽	Σ 💡											
F1	DT	EMPLMT	Al Constant											
	Α	B C	D	E	F	G	Н	I	J	К	L	M	N	0
1	IDNO	GROUP AGE	DTBIRTH	DTASSMNT	DTEMPLMT	SEX	HT	FEVMEAS	FEVPRED	FVCMEAS	FVCPRED	ASTHMA	BRON	SMKNO
2	1001	1 49	29.04.1946	12.06.1995	12.02.1972	1	175	3.40	3.59	4.49	4.45	0	0	
3	1002	1 90	01 11 1956	31 10 1990	18 10 1978	1	180	2.03	3.37	4 80	4:12	1	1	
5	1004	0 34	05.04 1958	09.09.1992	24.06.1980	ô	180	4.01	4.25	4.57	5.12	ů.	ŭ	
6	1005	0 29	12.03.1960	06.04.1989	05.05.1982	0	183	4.75	4.52	6.50	5.42	0	0	
7	1006	1 43	25.06.1947	21.07.1990	24.03.1982	0	174	4.60	3.73	5.82	4.54	0	0	
8	1007	1 27	10.02.1964	15.03.1991	24.01.1983	0	180	4.01	4.45	4.90	5.30	0	0	
9	1010	1 59	01 01 1928	10.02.1987	08.02.1965	1	120	2.58	2.97	3.68	3.73	0	0	
11	1010	1 29	01.01.1962	07.05.1991	09.02.1982	1	175	4.50	4.18	5.68	4.97	0	0	
12	1012	1 35	31 03 1961	29 06 1996	24 02 1981	ń	173	3.51	3 92	4.66	4 69	ň	ů	
13	1013	1 28	24.02.1966	31.03.1994	23.05.1986	0	168	2.92	3.91	4.09	4.59	1	0	
14	1014	0 34	29.06.1958	12.07.1992	10.06.1984	1	175	3.18	4.03	3.61	4.84	0	0	
15	1015	0 51	31.01.1936	25.02.1987	23.03.1982	0	168	2.76	3.24	4.21	3.99	0	1	
16	1016	0 49	29.01.1946	19.04.1995	10.04.1987	0	175	3.06	3.59	4.66	4.45	0	0	
1/	1017	0 29	02.02.1967	07.01.1996	24.01.1988	0	1/5	3.95	4.18	5.29	4.97	1	U	
10	1010	1 24	00 06 1909	12 00 1992	24 06 1979	1	170	2.91	3,24	4.40	3.77 4 EE	1	0	
20	1020	0 32	20.02.1964	21 05 1996	18 03 1988	ń	183	4 03	4 44	5 14	5 35	â	ő	
21	1021	1 50	16.10.1941	18.12.1991	22.10.1976	Ő	185	4.04	3.99	5.38	4.99	Ő	1	
22	1022	1 46	05.09.1943	03.10.1989	18.09.1980	1	170	3.81	3.47	5.13	4.24	0	0	
23	1023	0 49	06.06.1948	21.07.1997	12.07.1982	0	165	3.32	3.17	4.68	3.87	0	0	
24	1025	0 45	09.02.1949	16.05.1994	12.05.1988	0	170	3.40	3.50	4.34	4.26	0	0	
25	1026	1 46	17.04.1949	23.06.1995	25.06.1990	0	175	4.01	3.59	5.17	4.45	0	0	
20	1027	1 55	10.01.1942	17.04.1998	18.03.1991	1	165	2.80	2.97	3.57	3.69	0	0	
20	1028	1 26	10 04 1024	10.01.1996	19.01.1988	1	172	4.37	4.14	4.50	4:07	0	1	
29	1029	1 32	12 05 1958	28 07 1990	14 06 1983	1	178	4 68	4 22	5 92	5.06	1	, i	
30	1031	1 34	20 01 1960	15 03 1994	24 01 1985	1	190	4 91	4 68	6.06	5 69	Ô	ň	
31	1032	0 50	02.01.1942	20.01.1992	01.01.1976	1	170	2.47	3.36	3.88	4.13	0	1	
32	1033	1 53	10.10.1942	18.11.1995	16.10.1982	0	163	2.16	2.94	3.60	3.61	0	0	
33	1034	0 52	09.04.1945	26.05.1997	20.01.1988	0	185	3.53	3.94	4.70	4.94	0	0	
34	1036	0 42	16.02.1947	02.04.1989	12.02.1977	1	162	3.64	3.24	4.59	3.88	0	0	
30	1037	0 34	17.01.1959	21.03.1993	28.02.1987	U	1/7	3.69	4.12	5.12	4.95	0	0	
37	1038	1 20	15 01 1947	22 01 1992	25 01 1974	1	170	4.31	3.50	5.50	4.26	0	0	
38	1040	0 47	09 01 1946	20 03 1993	13 01 1971	0	180	3.97	3 87	5.11	4.40	0	0	
39	1042	0 24	23.07.1966	30.09.1990	24.07.1982	ĭ	180	4,80	4,51	6.27	5.36	0	ŏ	
40	1043	0 35	11.07.1954	11.08.1989	12.06.1974	1	175	4.16	4.00	4.84	4.81	1	0	
41	1044	1 51	02.02.1944	13.01.1995	19.04.1970	0	170	2.72	3.33	3.90	4.11	0	0	
42	1045	1 48	19.06.1949	27.09.1997	20.04.1980	1	175	3.39	3.63	4.17	4.47	0	0	
43	1046	1 38	01.01.1950	12.03.1988	31.01.1981	1	190	4.81	4.56	5.90	5.59	0	1	
44	1047	1 47	21.08.1945	10.06.1992	24.03.1986	1	175	3.59	3.66	4.88	4.49	0	0	
40	1048	1 39	16 02 1020	11 05 1001	12 02 1982	1	120	5.25	4.53	b./4 2.72	5.57	0	0	
47	1050	0 62	17 07 1961	23 09 1997	02 08 1991	1	173	3.68	2.74	4.69	4.67	0	0	
48	1051	0 35	20 03 1955	17 04 1990	31 03 1982	ñ	170	3.97	3 80	3.97	3.80	0	0	
49	1052	1 40	06.09.1948	20.11.1988	28.10.1979	1	190	5.05	4.50	6.26	5.54	Ő	ŏ	
50	1053	0 34	31.12.1959	11.12.1993	29.10.1976	0	182	4.54	4.33	5.65	5.24	0	0	

Using the cursor keys you can quickly examine the data set. Notice the IDNO in column 1. There are 136 subjects in this study (so there are 137 rows in the work sheet).

Variable Name	Variable Description	Value Labels for each code
IDNO	Identification No	
GROUP	Exposure Group	1 = Exposed to dust, $0 =$ Unexposed
SEX		1 = male, 0 = female
HT	Height in cms	
FEVMEAS	Measured FEV	
FEVPRED	Predicted FEV	
FVCMEAS	Measured FVC	
FVCPRED	Predicted FVC	
ASTHMA	Ever had asthma	0 = No, 1 = Yes, 2 = Don't Know
BRON	Ever had Bronchitis	0 = No, 1 = Yes, 2 = Don't Know
SMKNOW	Do you smoke now	1 = Yes $0 = $ No
SMKEVER	Have you ever smoked	0 = No, 1 = Ex smoker, 2 = Current smoker
CIGNO	No of cigarettes per day	-88if never smoked
CIGYRS	No of years smoked	-88 if never smoked
EMPYRS	No of Years with company	
RESPDUST	Current exposure to dust	

Moving around Data Editor

You may wish to try the following commands that will assist you in moving around the Data Editor.

Key	Description
Home	First Variable of the same case
End	Last Variable of the same case
Ctrl and (Up Arrow)	First Case of the same variable
Ctrl and (Down Arrow)	Last Case of the same variable
Ctrl and Home	First Value
Ctrl and End	Last Value

INTRODUCING ANALYSIS COMMANDS

Descriptive Statistics

The first step in data analysis is to generate descriptive statistics. This will help you check the data. This will also give you a feel for the data and help you identify any inconsistencies that there may be in the data. For example it is useful to look at the maximum and minimum values to check that they are in the range of acceptable values. This is sometimes called data cleaning. Techniques that are commonly used to do this include:

- Frequency Tables
- Descriptive Statistics
- Cross-tabulations
- Plots such as bar charts, histograms and scattergrams

The first three are in the analysis menu. Plots are obtained from the graphics menu.

Analysis By Column Or By Identifier

When you select a command that uses raw data it will then ask you to select columns of data in your spreadsheet. Instructions as to which columns to select appear at the top of the screen. Several commands, including some of the descriptive statistics commands, give you the option to **Group by column** or **Group by identifier**. If you select **Group by identifier** you will be asked to select a column that contains a variable that identifies groups. The data in columns that are subsequently selected will be divided according to the values in the identifier. How this option works will be illustrated by the **Frequencies** and **Descriptive** commands.

Frequency Tables

For variables that are categorical or take integer values it is useful to construct a frequency table. To do this in **StatsDirect**, first select the variables for which you want to generate frequencies. If the variables are consecutive you can block them by holding down the left mouse button and drag the mouse across the cells required. If they are not consecutive then select one by one by pressing down the **ctrl** key. Now choose **Analysis** from the Menu bar. From the drop down menu you select **Frequencies**.

Alternatively you can choose **Analysis** (without choosing the variables on the first) from the Menu bar. From the drop down menu you select **Frequencies.** Since the variables are not chosen previously, at this point a message will appear near the top of the screen (just below the menu bar) saying **Select data for FREQUENCY ANALYSIS [Min1: Max 200]**. Shown below.

-la	Felt Insert	Format Data	mahnin Granbina Tra	de Window Helo										الكارك
	Loc men	- I and the second	unaryste draptera, no	cre minuon meg										- 1.7
<u>UK</u>	Land	Select data	for FREQUENCY AN	ALYSIS (Min 1: I	Max 200)	S. F. N.								
sation	mode (*	Groups by column	 Groups by identif 	er To select non	-adjoined data, hold dow	in Util and use	e the mouse.							
F1	DT	EMPLMT												
	A	B C	D	E	F	G	н	1	J	ĸ	L	м	N	0
	1001	GROUP AGE	29 DA 1946	12 06 1995	12 02 1972	SEX	175	FEVREAS	PEVPRED	FUCHEAS	FVCPRED	ASTHNA	BRON	SHEN
	1002	1 46	12 10 1952	24 12 1998	10.08.1982	1	168	2.83	3.39	3.91	4.12	1	1	
	1003	1 34	01.11.1956	31.10.1990	18.10.1978	1	180	3.93	4.26	4.80	5.14	0	Ö	
	1004	0 34	05.04.1958	09.09.1992	24.06.1980	0	180	4.01	4.25	4.57	5.12	0	0	
	1005	0 29	12.03.1960	06.04.1989	05.05.1982	0	183	4.75	4.52	6.50	5.42	0	0	
	1006	1 43	25.06.1947	21.07.1990	24.03.1982	0	174	4.60	3.73	5.82	4.54	0	0	
	1007	1 59	10.02.1964	10 02 1997	24.01.1983	1	167	4.01	4.45	4.90	2 72	0	0	
	1010	1 29	01 01 1962	04 01 1991	04 02 1982	1	126	4 50	4 10	5 69	4 92	0	ň	
	1011	1 31	08.02.1957	07.05.1988	05.03.1979	1	177	4.19	4.21	5.61	5.03	ŏ	ŭ	
	1012	1 35	31.03.1961	29.06.1996	24.02.1981	0	173	3.51	3.92	4.66	4.69	0	0	
	1013	1 28	24.02.1966	31.03.1994	23.05.1986	0	168	2.92	3.91	4.09	4.59	1	0	
	1014	0 34	29.06.1958	12.07.1992	10.06.1984	1	175	3.18	4.03	3.61	4.84	0	0	
2	1015	0 51	31.01.1936	25.02.1987	23.03.1982	0	168	2.76	3.24	4.21	3.99	0	1	
-	1015	0 49	29.01.1946	19.04.1995	10.04.1987	0	175	3.06	3.59	4.66	4.45	0	0	
	1019	1 51	23 09 1939	20 10 1990	11 08 1962	1	168	3.70	3 24	4 40	3.99	0	0	
	1019	1 34	05 06 1959	13 08 1993	24 06 1979	1	170	3 91	3 82	4 80	4.55	1	ň	
	1020	0 32	20.02.1964	21.05.1996	18.03.1988	0	183	4.03	4.44	5.14	5.35	õ	õ	
	1021	1 50	16.10.1941	18.12.1991	22.10.1976	0	185	4.04	3.99	5.38	4.99	0	1	
	1022	1 46	05.09.1943	03.10.1989	18.09.1980	1	170	3.81	3.47	5.13	4.24	0	0	
	1023	0 49	06.06.1948	21.07.1997	12.07.1982	0	165	3.32	3.17	4.68	3.87	0	0	
	1025	0 45	09.02.1949	16.05.1994	12.05.1988	0	170	3.40	3.50	4.34	4.26	0	0	
	1026	1 45	17.04.1949	23.06.1996	25.06.1990	0	1/5	4.01	3.59	5.17	9.95	0	0	
	1027	1 26	01 01 1942	10 01 1996	19 01 1990	1	172	4.27	4 14	3.57	3.67	0	0	
	1020	1 54	19 04 1934	12 09 1988	23 07 1979	1	120	3.63	3 24	4.50	4.03	0	1	
	1030	1 32	12 05 1958	28 07 1990	14 06 1983	1	178	4.68	4 22	5 92	5.06	ĩ	â	
	1031	1 34	20.01.1960	15.03.1994	24.01.1985	1	190	4.91	4.68	6.06	5.69	0	0	
	1032	0 50	02.01.1942	20.01.1992	01.01.1976	1	170	2.47	3.36	3.88	4.13	0	1	
	1033	1 53	10.10.1942	18.11.1995	16.10.1982	0	163	2.16	2.94	3.60	3.61	0	0	
	1034	0 52	09.04.1945	26.05.1997	20.01.1988	0	185	3.53	3.94	4.20	4.94	0	0	
_	1036	0 42	16.02.1947	02.04.1989	12.02.1977	1	162	3.64	3.24	4.59	3.88	0	0	
	1037	0 34	26 06 1947	10 00 1002	20.02.1987	0	170	3.69	4.12	5:12	4.95	0	0	
	1039	1 38	15 01 1953	22 01 1991	25 01 1974	1	170	3 98	3 72	5.11	4.46	0	0	
	1040	0 47	09.01.1946	20.03.1993	13.01.1971	Ô	180	3.97	3.87	5.18	4.78	ŏ	ŏ	
	1042	0 24	23.07.1966	30.09.1990	24.07.1982	1	180	4.80	4.51	6.27	5.36	0	Û	
	1043	0 35	11.07.1954	11.08.1989	12.06.1974	1	175	4.16	4.00	4.84	4.81	1	0	
	1044	1 51	02.02.1944	13.01.1995	19.04.1970	0	170	2.72	3.33	3.90	4.11	0	0	
	1045	1 48	19.06.1949	27.09.1997	20.04.1980	1	175	3.39	3.63	4.17	4.47	0	0	
	1046	1 38	01.01.1950	12 03 1988	31 01 1981	1	190	4.81	4.56	5 90	5.59	0	1	
	1047	1 47	02 02 1945	04 09 1992	24.03.1986	1	175	5.59	3.66	4.88	4.49	0	0	
	1049	0 62	16 03 1929	11 05 1994	13 03 1982	1	175	3.04	2 74	3.73	3.49	0	0	
	1050	0 36	17 07 1961	23 09 1997	02 08 1991	1	173	3.68	3 89	4 69	4 67	0	0	
	1051	0 35	20.03.1955	17.04.1990	31.03.1982	ô	170	3.97	3.80	3.97	3.80	ŏ	ŏ	
	1052	1 40	06.09.1948	20.11.1988	28.10.1979	1	190	5.05	4.50	6.26	5.54	0	Û	
	1053	0 34	31.12.1959	11.12.1993	29.10.1976	0	182	4.54	4.33	5.65	5.24	0	0	

If you select the Group by column option using the radio button, descriptive statistics will be produced for all data for each variable selected. You select the data the same way as explained above and then click **OK** button. If instead you select the radio button **Group by identifiers** the screen below appears, you could have selected a categorical variable. Separate frequency tables would then be produced for each category. For example if you chose GROUP as the identifier variable separate frequency table would be produced for exposed and non-exposed workers. Output for this will be shown later.

d Stats	Direct - [fo	undry. xls	1									
₩ Ele	Edit Insert	Format	lata <u>A</u> r	alysis <u>G</u> raphics <u>T</u> o	ols <u>W</u> indow Help							- @ ×
<u>©</u> K I⊤ Batch F1	mode C	Selec Groups by MPLMT	t GROU column	IP IDENTIFIERS for Groups by identif	r FREQUENCY AN	IALYSIS (Min 1: Ma adjoined data, hold dow	x 20) In Ctrl and use t	he mouse.				
	A	В	C	D	E	F	G	н	I I	J	ĸ	L 🔺
1	IDNO	GROUP	AGE	DTBIRTH	DTASSMNT	DTEMPLMT	SEX	HT	FEVMEAS	FEVPRED	FVCMEAS	FVCPRED
2	1001	1	49	29.04.1946	12.06.1995	12.02.1972	1	175	3.40	3.59	4.49	4.45
3	1002	1	46	12.10.1952	24.12.1998	10.08.1982	1	168	2.83	3.39	3.91	4.12
4	1003	1	34	01.11.1956	31.10.1990	18.10.1978	1	180	3.93	4.26	4.80	5.14
5	1004	0	34	05.04.1958	09.09.1992	24.06.1980	0	180	4.01	4.25	4.57	5.12
6	1005	0	29	12.03.1960	06.04.1989	05.05.1982	0	183	4.75	4.52	6.50	5.42
7	1006	1	43	25.06.1947	21.07.1990	24.03.1982	0	174	4.60	3.73	5.82	4.54
8	1007	1	27	10.02.1964	15.03.1991	24.01.1983	0	180	4.01	4.45	4.90	5.30
9	1009	1	59	11.01.1928	10.02.1987	08.02.1965	1	167	2.58	2.97	3.68	3.73
10	1010	1	29	01.01.1962	04.01.1991	04.02.1982	1	175	4.50	4.18	5.68	4.97
11	1011	1	31	08.02.1957	07.05.1988	05.03.1979	1	177	4.19	4.21	5.61	5.03
12	1012	1	35	31.03.1961	29.06.1996	24.02.1981	0	173	3.51	3.92	4.66	4.69
13	1013	1	28	24.02.1966	31.03.1994	23.05.1986	0	168	2.92	3.91	4.09	4.59
14	1014	0	34	29.06.1958	12.07.1992	10.06.1984	1	175	3.18	4.03	3.61	4.84
15	1015	0	51	31.01.1936	25.02.1987	23.03.1982	0	168	2.76	3.24	4.21	3.99
16	1016	0	49	29.01.1946	19.04.1995	10.04.1987	0	175	3.06	3.59	4.66	4.45
1/	1017	U	29	02.02.1967	07.01.1996	24.01.1988	U	175	3.95	4.18	5.29	4.97
18	1018	1	51	23.09.1939	20.10.1990	11.08.1967	1	168	3.77	3.24	4.40	3.99
19	1019	1	34	05.06.1959	13.08.1993	24.06.1979	1	170	3.91	3.82	4.80	4.55
20	1020	U	32	20.02.1964	21.05.1996	18.03.1988	U	183	4.03	4.44	5.14	5.35
21	1021	1	50	16.10.1941	18.12.1991	22.10.1976	U	185	4.04	3.99	5.38	4.99
22	1022	1	46	05.09.1943	03.10.1989	18.09.1980	1	170	3.81	3.47	5.13	4.24
23	1023	U	49	06.06.1948	21.07.1997	12.07.1982	U	165	3.32	3.17	4.68	3.87
24	1025	0	45	09.02.1949	16.05.1994	12.05.1988	U	170	3.40	3.50	4.34	4.26

You will then be asked where to put the result. You can either select existing output **report1** or **New report**. If you choose **Report1** your output will be appended with the existing output. **New Report** will open a separate page for this output. For example, suppose the three variables **group**, **sex** and **asthma** are chosen to run the frequencies, the output window will look like below

d StatsDirect - [Re	eport 11								
<u></u> ∃ Eile Edit Insert	Format Analy	∕sis <u>T</u> ools <u>W</u> indow	Help	- 7					
Book 1 fou	Indry Re	port 1							
D 📽 🖬 🎒 🛙	<u>ا</u> ه ک 🛍 🛛	ν Σ 🦓							
Arial	• 10	🔹 😝 B I	U × ₂ x ²	e = = = A					
12 1 1 1 1 2 1 1 3 1 1 4 1 1 5 1 1 6 1 1 7 1 1 8 1 1 9 1 1 10 1 11 1 12 1 13 1 14 1 15 1 16 1 17 1 18 1 2									
<u>Frequencies</u>									
Frequency analys	is for GROUF	5							
Total = 136									
Value	Frequency	Relative %	Cumulative	Cumulative Relative %					
1	63 73	46.323529 53.676471	63 136	46.323629 100					
Frequency analys	is for SEX:								
Total = 136									
Value	<u>Frequency</u>	Relative %	<u>Cumulative</u>	Cumulative Relative %					
0 1	63 73	46.323529 53.676471	63 136	46.323529 100					
Frequency analys	is for ASTHM	<u>A</u> :							
Total = 136									
Value	Frequency	Relative %	Cumulative	Cumulative Relative %					
1	125 11	91.911765 8.088235	125 136	91.911/65 100					
09/03/2006 15:12				INS CAPS NU					

If instead you had chosen the Group by Identifier option, can only select one variable for the frequency table. Suppose you wanted separate frequencies for smoking status by gender, one would select **sex** as the group identifier variable and SMKEVER as the data getting the results below in the output window.

Frequencies

Frequency ana	Frequency analysis for SMKEVER_SEX_1:										
Total = 73											
<u>Value</u> 0 1 2	<u>Frequency</u> 24 24 25	Relative % 32.8767 32.8767 34.2466	<u>Cumulative</u> 24 48 73	<u>Cumulative Relative %</u> 32.8767 65.7534 100							
Frequency ana	Ilysis for SMKEV	ER_SEX_0:									
Total = 63											
<u>Value</u> 0 1 2	<u>Frequency</u> 20 14 29	<u>Relative %</u> 31.746 22.2222 46.0317	<u>Cumulative</u> 20 34 63	Cumulative Relative % 31.746 53.9683 100							

Exercise 1 Using the frequencies options find out

- what proportion of the foundry workers were exposed to dust?
- what proportions had ever suffered from bronchitis?
- what proportion had ever smoked?
- what proportion smoked more than 20 cigarettes per day?

Descriptives

The **descriptives** menu in StatsDirect is useful for summarizing quantitative data. It gives the most important descriptive statistics. The output also tells you the number of cases in the analysis and the numbers of missing values. To use this click on the **Analyse** tile choose the **Descriptives** option.

d Stat	Direct	l - [foundry	.xls]													E	
I File	Edit	Insert Forma	t Data	Analysis Graphics	Tools W	indow He	lp										- 8
Bool	k1 [foundry	I Rep	Chi-Square Tests													
		🍜 🖻 🔉		Proportions													
EJ1	7	125		Rates													
	4	1		Distributions													
	C	D		Sample Size			G	н		J	ĸ	L	M	N	0	Р	Q
1	AGE	29.04	1946	Randomization		1972	SER	125	FEVREAS	FEVPRED	FVCMEAS	FUCPRED	ASTHMA	BRON	SHKNOW	SHREVER	EMP
3	45	12 10	1952	Miscellaneous		1982	1	168	2.83	3.39	4.42	4.45	1	1	1	2	
4	34	01.11	1956			1 and	:	100	3.93	4.26	4.80	5.14	Ô	Ô	ô	Ő	
5	34	05.04	.1958	Descriptive		Univar	ate Summary		4.01	4.25	4.57	5.12	0	0	1	2	
6	29	12.03	.1960	Parametric		weight	ed Univariate :	summary	4.75	4.52	6.50	5.42	0	0	0	0	
1	43	25.06	1947	Non-parametric		Quick L	Inivariate Sumr	nary	4.60	3.73	5.82	4.54	0	0	0	1	
ğ	59	11 01	1928	Analysis of Variance	:e •	1965	1	167	2 58	2 97	3,68	3 73	0	0	1	2	
10	29	01.01	1962	Regression & Corr	elation +	1982	1	175	4.50	4.18	5.68	4.97	Ő	Ŭ	ô	ĩ	
11	31	08.02	.1957	Agreement		1979	1	177	4.19	4.21	5.61	5.03	0	0	1	2	
12	35	31.03	.1961	Survival Analysis	•	1981	0	173	3.51	3.92	4.66	4.69	0	0	1	2	
13	28	24.02	1966	Meta-Analysis		1986	0	168	2.92	3.91	4.09	4.59	1	0	1	2	
15	51	29.06	1936	Crosstabs		1982	<u>.</u>	168	2 76	3 24	4 21	3.99	0	1	1	2	
16	49	29.01	1946	Frequencies		1987	ŏ	175	3.06	3.59	4.66	4.45	ŏ	Ô	ô	ĩ	
17	29	02.02	.1967	Ontions		1988	0	175	3.95	4.18	5.29	4.97	1	0	0	0	
18	51	23.09	.1939			1967	1	168	3.77	3.24	4.40	3.99	0	0	0	1	
19	34	05.06	.1959	13.08.1993	24.06	.1979	1	170	3.91	3.82	4.80	4.55	1	0	1	2	
20	32	20.02	1964	21.05.1995	22 10	1988	0	183	4.03	4.44	5.14	5.35	0	1	0	1	
22	46	05.09	1943	03 10 1989	18 09	1980	1	170	3 81	3 47	5.13	4 24	ő	ô	0	ñ	
23	49	06.06	1948	21.07.1997	12.07	.1982	ō	165	3.32	3.17	4.68	3.87	ő	Ő	ĩ	2	
24	45	09.02	. 1949	16.05.1994	12.05	.1988	0	170	3.40	3.50	4.34	4.26	0	0	0	0	
25	46	17.04	.1949	23.06.1995	25.06	.1990	0	175	4.01	3.59	5.17	4.45	0	0	0	0	
26	56	10.01	1020	17.04.1998	18.03	1000	1	155	2.80	2.97	3.57	3.69	0	U	0	1	
28	54	19 04	1934	12 09 1988	23.07	1979	1	170	3 63	3 24	4.50	4.07	0	1	0	1	
29	32	12.05	1958	28.07.1990	14.06	.1983	1	178	4.68	4.22	5.92	5.06	1	ō	0	- O	
30	34	20.01	.1960	15.03.1994	24.01	.1985	1	190	4.91	4.68	6.06	5.69	0	0	1	2	
31	50	02.01	.1942	20.01.1992	01.01	.1976	1	170	2.47	3.36	3.88	4.13	0	1	0	1	
32	53	10.10	. 1942	18.11.1995	16.10	.1982	0	163	2.16	2.94	3.60	3.61	0	0	1	2	
34	42	16 02	1947	02 04 1999	12 02	1977	1	162	2.53	3.74	4.70	4.74	0	0	1	2	
35	34	17.01	1959	21.03.1993	28.02	1987	ô	177	3.69	4.12	5.12	4,95	ŏ	0	ô	Õ	
36	45	26.06	.1947	19.09.1992	31.03	.1983	0	170	4.31	3.50	5.50	4.26	Ő	Ū.	ĩ	2	
37	38	15.01	1953	22.01.1991	25.01	.1974	1	170	3.98	3.72	5.11	4.46	0	0	0	1	
38	47	09.01	. 1946	20.03.1993	13.01	. 1971	0	180	3.97	3.87	5.18	4.78	0	0	0	1	
39	24	23.07	1054	30.09.1990	24.07	1982	1	180	4.80	4.51	6.27	5.35	0	U	0	1	
41	51	02.02	1944	13 01 1995	19 04	1970	n i	170	2 72	3 33	3.90	4.01	ń	0	0	1	
42	48	19.06	1949	27.09.1997	20.04	.1980	1	175	3.39	3.63	4.17	4.47	Ŭ	0	Ŭ	Ô	
43	38	01.01	1950	12.03.1988	31.01	.1981	ĩ	190	4.81	4.56	5.90	5.59	Ő	î	î	2	
44	47	21.08	.1945	10.06.1992	24.03	.1986	1	175	3.59	3.66	4.88	4.49	0	0	0	0	
45	39	07.07	. 1955	04.09.1994	25.06	. 1982	1	190	5.25	4.53	6.74	5.57	0	0	1	2	
40	62	16.03	1929	11.05.1991	13.03	1983	0	175	3.04	2.74	3.73	3.49	0	U	0	1	
48	35	20.03	1955	12 04 1990	31 03	1982	ń	120	3 97	3 80	3 97	3 80	0	0	1	2	
49	40	06.09	1948	20.11.1988	28.10	.1979	ĭ	190	5.05	4.50	6.26	5.54	ŏ	ő	ô	õ	
50	34	31.12	.1959	11.12.1993	29.10	.1976	0	182	4.54	4.33	5.65	5.24	0	0	0	0	
A The	founde	. [

If you select **Descriptives Report** the screen below will appear

d Stats	🗗 StatsDirect - [foundry.xls]													
₩ Ele	<u>E</u> dit Ir	nsert F <u>o</u> rmat <u>D</u> ata	<u>A</u> nalysis <u>G</u> raphics	<u>T</u> ools <u>W</u> indow <u>H</u> el	P				-	Ξ×				
<u>0</u> K		Cancel Select GR	OUP IDENTIFIER	S (Min 1: Max 20)										
F Batch	n mode	C Groups by colum	n 🧿 Groups by id	entifier To select no	n-adjoined dat	a, hold down	Ctrl and use th	ne mouse.						
1		[105	an an Anna an Anna	ACCEPTERS. Derected and accepted and a		anta de a serve e		and the second						
HI HI	7	175	75											
	C	D	E	F	G	Н	T	J	K	•				
1	AGE	DTBIRTH	DTASSMNT	DTEMPLMT	SEX	HT	FEVMEAS	FEVPRED	FVCMEAS	FV(
2	49	29.04.1946	12.06.1995	12.02.1972	1	175	3.40	3.59	4.49					
3	46	12.10.1952	24.12.1998	10.08.1982	1	168	2.83	3.39	3.91					
4	34	01.11.1956	31.10.1990	18.10.1978	1	180	3.93	4.26	4.80					
5	34	05.04.1958	09.09.1992	24.06.1980	0	180	4.01	4.25	4.57					
6	29	12.03.1960	06.04.1989	05.05.1982	0	183	4.75	4.52	6.50					
7	43	25.06.1947	21.07.1990	24.03.1982	0	174	4.60	3.73	5.82					
8	27	10.02.1964	15.03.1991	24.01.1983	0	180	4.01	4.45	4.90					
9	59	11.01.1928	10.02.1987	08.02.1965	1	167	2.58	2.97	3.68					
10	29	01.01.1962	04.01.1991	04.02.1982	1	175	4.50	4.18	5.68					
11	31	08.02.1957	07.05.1988	05.03.1979	1	177	4.19	4.21	5.61					
12	35	31.03.1961	29.06.1996	24.02.1981	0	173	3.51	3.92	4.66					
13	28	24.02.1966	31.03.1994	23.05.1986	0	168	2.92	3.91	4.09					
14	34	29.06.1958	12.07.1992	10.06.1984	1	175	3.18	4.03	3.61					
15	51	31.01.1936	25.02.1987	23.03.1982	0	168	2.76	3.24	4.21					
16	49	29.01.1946	19.04.1995	10.04.1987	0	175	3.06	3.59	4.66					
17	29	02.02.1967	07.01.1996	24.01.1988	0	175	3.95	4.18	5.29					
18	51	23.09.1939	20.10.1990	11.08.1967	1	168	3.77	3.24	4.40					
19	34	05.06.1959	13.08.1993	24.06.1979	1	170	3.91	3.82	4.80					
20	32	20.02.1964	21.05.1996	18.03.1988	0	183	4.03	4.44	5.14	-				
1.	foundry	1			•									

If you select the Group by column option using the radio button, descriptive statistics will be produced for all data for each variable selected. For example if you choose the variables FEVMEAS and FVCMEAS, you get the following in the report window.

Descriptive statistics

Variables	FEVMEAS	FVCMEAS
Valid Data Missing Data Mean Variance SD SEM Lower 95% CL Upper 95% CL Geometric Mean Skewness Kurtosis Maximum Upper Quartile Median Lower Quartile Minimum Range Variance coeff. Sum	136 0 3.793824 0.546659 0.739364 0.0634 3.668438 3.919209 3.714979 -0.241417 3.228908 5.61 4.285 3.79 3.3525 1.45 4.16 0.194886 515.96	136 0 4.813456 0.705058 0.839678 0.072002 4.671059 4.955853 4.738105 0.041158 3.037622 7.21 5.3775 4.805 4.24 2.68 4.53 0.174444 654.63
Centile 5	2.5885	3.4935

If instead you had selected group by identifier, you could have selected a categorical variable SEX as the identifier. Separate descriptive statistics would then be produced for men and women as shown below. It is only possible to obtain descriptive statistics for one variable at a time. Having selected sex as the identifier and then FEVMEAS as the data you get the result below. The first column give descriptive statistics where SEX = 1, that's men, and the second for SEX=0 which is women.

Descriptive statistics

Variables	FEVMEAS_SEX_1	FEVMEAS_SEX_0
Valid Data	73	63
Missing Data	0	0
Mean	3.8033	3.7829
Variance	0.6703	0.4117
SD	0.8187	0.6416
SEM	0.0958	0.0808
Lower 95% CL	3.6123	3.6213
Upper 95% CL	3.9943	3.9444
Geometric Mean	3.7043	3.7274
Skewness	-0.3677	0.0477
Kurtosis	2.9996	3.3013
Maximum	5.56	5.61
Upper Quartile	4.435	4.21
Median	3.81	3.78
Lower Quartile	3.24	3.4
Minimum	1.45	2.16
Range	4.11	3.45
Variance coeff.	0.2153	0.1696
Sum	277.64	238.32
Centile 5	2.428	2.728

If you select **Quick Univariate Summary**, you cannot divide the data by an identifier. **StatsDirect** will then ask you to select a variable and prepare a summary in a separate window as shown.

Summary Statistics	
	<u> </u>
Title: CIGNO	<u>^</u>
Valid data Missing Sum Mean Variance Standard deviation Variation coefficient Standard error of mean 95% Upper CL of mean 95% Lower CL of mean Geometric mean Skewness Kurtosis Maximum 95th percentile Upper quartile Sth percentile Sth percentile Sth percentile Median Lower quartile	136 0 -1987 -14.610294 2652.091449 51.498461 -3.524807 4.415957 -5.876891 -23.343697 * -0.682363 1.576957 60 30 20 16.5 -88 -88
Range	148

Note that the mean numbers of cigarettes given per day is -14.6102. This is because data for nonsmokers is recorded as -88 to represent not applicable. Dealing with missing values will be considered below.

Exercise: Use the **descriptive** procedure to determine the mean median and range of the ages of the workers. Using the group by identifier option obtain the mean median and range of ages for exposed and non-exposed workers.

Cross-tabulation, Chi-squared Test and Fisher's Exact Test

To examine the relationship between two categorical variables, a two way Frequency Table can be used called a cross-tabulation. Rows of the data are categories for one variable and columns are categories for the second. Click on **Analyze** then **Crosstabs.** The screen below will appear.

d Stats	Direct - [foun	dry.xls]									
I Ele	<u>E</u> dit <u>I</u> nsert Fo	ormat <u>D</u> ata	<u>A</u> nalysis	<u>G</u> raphics <u>T</u> ools	<u>W</u> indow	Help				-	- ×
ОК	Cancel	select da	ta for FIBS			(1 column)					
E Datab	and a second sec			. on oon in	. ((1 0010111)					
Daten	mode										
B1	CIGNO)									
í -	Q	R	s	т	U	v	w	X	Y	Z	-
1	EMPYRS	CIGNO	CIGYRS	RESPDUST							
2	23	20	31	. 66							
3	16	20	11	1.71							
4	12	-88	-88	1.69							_
5	12	25	16	. 00							
6	7	-88	-88	. 00							
7	8	20	15	1.96							
8	8	-88	-88	1.23							_
9	22	30	12	. 6 9							
10	9	20	8	1.18							
11		20	17	1.20							
12	15	20	25	2.04							
14	8	40		1.00							
14	2	-00	-00								
16		20	23								
17	0	20		. 00							
18	22	-00	-00	2.04							
19	2.3	20	10	2.04							
20	14	20	10	2.46							
	oundru /		1	.00		-				1	

Select the column SMKNOW then click OK. In the next screen select the column GROUP.

		ormat Data		iraphics Tools	Window He	do				3.8	>
LIK	Cancel	select dat	ta for SECO			IMNS1 (Min	1: Max 100	e.			manual designed there
Batch	mode	1			To select r	ion-adjoined d	lata, hold down	Ctrl and use th	e mouse.		
01	SMKNO	W									
	M	N	0	Р	Q	R	S	т	U	v	1
1	ASTHMA	BRON	SMKNOW	SMKEVER	EMPYRS	CIGNO	CIGYRS F	RESPDUST			
2	0	0	1	2	23	20	31	. 66			
3	1	1	1	2	16	20	11	1.71			
4	0	0	0	0	12	-88	-88	1.69			
5	0	0	1	2	12	25	16	.00			
6	0	0	0	0	7	-88	-88	.00			
7	0	0	0	1	8	20	15	1.96			
8	0	0	0	0	8	-88	-88	1.23			
9	0	0	1	2	22	30	12	. 69			
10	0	0	0	1	9	20	8	1.18			
11	0	0	1	2	9	20	17	1.20			
12	0	0	1	2	15	20	25	2.04			
13	1	0	1	2	8	40	3	1.88			
14	0	0	0	0	8	-88	-88	. 00			
15	0	1	1	2	5	20	29	. 00			
16	0	0	0	1	8	20	3	. 00			
17	1	0	0	0	8	-88	-88	. 00			
18	0	0	0	1	23	40	17	2.04			
19	1	0	1	2	14	20	18	2.46			
20	0	0	0	1	8	5	1	. 00			

A third screen appears, which allows you to select a third variable for a three way table. Since we do not wish to do this, click Cancel. The follow panel appears.

Select options for contract	ontingency t	able analysis	×
	<u>0</u> K	<u>H</u> elp	<u>C</u> ancel
Try exact test Show expected counts Show trend scores		Show percentag Show cell chi-squ Specify trend sco	es Jare Dres

Click OK to get the results part of which are shown below

Crosstabs

Row variable (firs Column variable (t classifier): (second classifie	r): G	MKNOW ROUP						
0 1	0 43 20	1 39 34							
Contingency tab	le analysis								
Observed % of row % of col	43 52.44% 68.25%	39 47.56% 53.42%	82 60.29%						
Observed % of row % of col	20 37.04% 31.75%	34 62.96% 46.58%	54 39.71%						
Total % of n	63 46.32%	73 53.68%	136						
TOTAL number of cells = 4									
NOMINAL INDEF Chi-square = 3.10 G-square = 3.130 Fisher-Freeman-H	P <u>ENDENCE</u> 06252 1666 Halton exact	DF = 1 DF = 1	P = 0.078 P = 0.0768 not calculated						

From the table above it can be seen that the percentage of workers who currently smoke is higher for those exposed to dust than those who are not, 47% (34/73) as compared to 32% (20/63). This is not statistically significant with a 5% level test. It is recommended that you report the p-value rather than just say significant or not significant. Since the p-value is only slightly larger than 5%, you might write "There was a slight suggestion that workers exposed to dust were more likely to smoke (p=0.078)".

If the cross-tabulation has more than 2 rows or 2 columns, a different set of statistics are produced. For example suppose one compared the two dust exposed and not dust exposed in terms of the smoking status variable SMKEVER, using SMKEVER as the row variable and GROUP as the column variable one gets the following table and output.

StatsDirect	- [Report 2]	1							
🗐 Eile Edit I	(nsert Format	<u>Analysis T</u> oo	ols <u>Window</u> Help	_ 8 ×					
foundry	Report 1	Report 2							
🗅 🚔 🖬 🤅	5 Pa 🐰	🛍 🗠 Σ							
Arial	-	10 💽 🛞) B <i>I</i> <u>U</u> × ₂ x ² <u>≡</u> ≡ ≡ <u>⊟</u> <u>#</u>						
当…1…2	113114	1151116	(+++7+++8+++9+++10+++11+++12+++13+++14+++15+++16+++17+++18+++2						
Contingenc	y table anal	<u>ysis</u>		~					
Observed	24	20	44						
% of row % of col	54.55% 38.1%	45.45% 27.4%	32.35%						
Observed	19	19	38						
% of row % of col	50% 30.16%	50% 26.03%	27.94%						
Observed	20	34	54						
% of row % of col	37.04% 31.75%	62.96% 46.58%	39.71%						
Total % of n	63 46.32%	73 53.68%	136						
TOTAL numb	per of cells =	6							
<u>NOMINAL IN</u> Chi-square =	DEPENDEN 3.275682	<u>CE</u> DF = 2	P = 0.1944						
G-square = 3 Fisher-Freem	3.299606 nan-Halton ex	DF = 2 act	P = 0.1921 not calculated	E					
ANOVA Chi-square for equality of mean column scores = 3.058527 DF = 1 P = 0.0803									
LINEAR_TREND Sample correlation (r) = 0.150518 Chi-square for linear trend (MP) = 3.058527 DF = 1 P = 0.0803									
<u>NOMINAL</u> Phi = 0.1551 Pearson's co Cramér's V =	SSOCIATION 96 omingency = 1 = 0.155196	<u>4</u> D. 1 6 336							
ORDINAL Goodman-Kr Approximate Approximate	uskal gamma test of gamn test of indep	a = 0.246264 na = 0 SE endence: SE	E = 0.133721 P = 0.0655 95% CI = -0.015824 to 0.508351 E = 0.137166 P = 0.0726 95% CI = -0.022577 to 0.515105						
09/03/2006 16:	39			INS CAPS NUM					

If one considers the categories, of smoking to have an order None, Ex, Current, the appropriate statistic is Chi^2 with Linear Trend. If instead one did not consider them to be ordered, the appropriate test is Total Chi^2 might be appropriate. From this table there is a slight suggestion (p=0.0803) that those exposed to dust (the column Failures) are more likely to smoke and less likely to be non-smokers or ex-smokers.

Exercise: Using the cross-tabs procedure examine whether there is a relationship between current smoking status and bronchitis symptoms.

Are the expected numbers greater than 5 for all cells?

Fill in the spaces and delete as appropriate in the following statement:

"Amongst those that currently smoked ____% had experienced symptoms of bronchitis whereas

____% of non-smokers experience such symptoms. This was statistically <u>significant/non significant</u> at a 5% level using a two-tailed continuity corrected chi-squared test with p=_____"

Exercise: Now use the cross-tabs procedure to examine the relationship between Exposure to dust and symptoms of bronchitis and asthma. Record your conclusions below using either the continuity corrected chi-squared or Fisher's exact test as appropriate.

You should find no statistically significant relationship between exposure to dust and either asthma or bronchitis symptoms. Whilst 15% (11/73) of the exposed worker had symptoms of bronchitis and only 6% (4/63) of non-exposed, this difference was not statistically significant at the 5% level (p=0.169). There are several explanations for this. There may be no relationship between the exposure to dust and respiratory disease. Alternatively, the study may have lacked statistical power to detect small differences. It should be noted also that only 11% (15/136) of the sample reported such symptoms.

FILE HANDLING

Saving The Work

Worksheets and Reports are all saved separately. To save the report that is created here:

Either:

a) Select Save Report As from the file menu,

Or: b) Click on the **Report 1** tab with the right mouse button and select **Save As** from the menu of commonly used options which appears below

🗗 StatsDirect - [Report 2]		ΞX
🗩 File Edit Insert Format Analysis Tools	Window Help	× B ×
foundry Report 1 Report 2		
	8	
Arial 💌 🚺 🕶 🍪 1	B ℤ ⊻ ×₂ ײ 批 書 ≡ ∷ AA	
141111211311411511611	17 + 1 + 8 + 1 + 9 + 1 + 10 + 1 + 11 + 1 + 12 + 1 + 13 + 1 + 15 + 146 + 147 + 148 + 172	
Configuration to blo and built		^
Conungency table analysis		
Observed 24 20 4		
% of col 38.1% 27.4% 33	2 Save Report As 2	
	Save in: 💼 🚥 🗾 🔶 👉 📰 -	
Ubserved 19 19 3 % of row 50% 50%		1
% of col 30.16% 26.03% 2		
Observed 20 34 5	Documents	
% of row 37.04% 62.96%		
% of col 31.75% 46.58% 3	a Dation	
Total 63 73 1		
% of n 46.32% 53.68%		
TOTAL number of cells = 6	My Documents	
NOMINAL INDEPENDENCE		
Chi-square = 3.275682 DF = 2		
Fisher-Freeman-Halton exact	My Lomputer	
1101/1	File name: Report 2 Save	
ANUVA Chi-square for equality of mean column sco	MuNetwork Save as type: RTF File (".ntf)	
DF = 1 P = 0.0803	Places	
LINEAR TREND		
Sample correlation (r) = 0.150518		
Chi-square for linear trend (MF) = 3.058527 DF = 1 P = 0.0803		
Phi = 0.155196		
Pearson's contingency = 0.15336		
Cramer's V = 0.155196		
ORDINAL		
Goodman-Kruskal gamma = 0.246264 Approximate test of gamma = 0 SE = 0	1 133721 P = 0.0655 95% CI = -0.015824 to 0.608351	
Approximate test of independence: SE = 0).137166 P = 0.0726 95% CI = -0.022577 to 0.515105	~
09/03/2006 16:57	INS CA	PS NUM

The standard **Save As** dialogue box will be displayed as above. You have to choose a drive and a suitable name of the file. For Report the file extension is **.rtf** (which stands for rich text format. Files in that format can be opened by **Word** to edit). Then click the **Save** button. Say, you saved the Report with the name called **survey1** in the pen drive.

Similarly, to save the Workbook (**Book 1**):

Either:

a) Select Save Workbook As from the file menu,

Or:

b) Click on the **Book 1** tab with the right mouse button and select **Save As** from the menu of commonly used options which appears below

d' Stats	Direct -	[foundry.x	ls]															X
I File	Edit Inse	ert Format	Data A	nalysis (Graphics T	ools Wir	ndow Help											- 8 ×
found	dry Y	Report 1	Report	2														
D 🗳	8 6	Ba % 6	2 🔊	Σ	8													
B1	1	ROUP																
	A	В	С		D		E		F	G	Н	1	J	К	L	М	N	0 🔺
1	IDNO	GROUF	P AGE		DTBIRTH	D	FASSMNT		DTEMPLMT	SEX	HT	FEVMEAS	FEVPRED	FVCMEAS	FVCPRED	ASTHMA	BRON	SMKNO
2	1001	1	L 49	29.	04.1946	12.1	06.1995	12	.02.1972	1	175	3.40	3.59	4.49	4.45	0	0	
3	1002	1	1 3/	01	11 1952	24	10 1998	19	10 1978	1	168	2.83	3.39	3.91	9:12	1	1	
5	1004	-	34	05.	04 4050	00		10	04 1000		100	1 04	4.20	4.57	5.12	Ő	Ő	
6	1005	C	29	12.	Save W	orkbool	k As						?	6.50	5.42	0	0	
7	1006	1	L 43	25.	9	18 10					1	1.000 C 100 C 10		5.82	4.54	0	0	
8	1007	1	L 27	10.	2	Save in:	🛛 🌋 Vol1 o	n 'Uk-a	ac-man-ser-fs2' (L:)	-	🗈 💣 🎹	•	4.90	5.30	0	0	
10	1019	1	1 29	01			Con (1)							5 68	3.73	0	0	
11	1011	1	31	08.	1	2	Coehj							5.61	5.03	0	Ő	
12	1012	1	L 35	31.	d 14		person							4.66	4.69	0	0	
13	1013	1	L 28	24.	Docu	ments	QUrive							4.09	4.59	1	0	
14	1014	L L	J 34	29.		3	SETUP							3.61	4.84	0	0	
15	1015	6	1 49	29										4.21	3.99 A.45	0	1	
17	1017	č	29	02	Dee	kton								5.29	4.97	ĭ	ŏ	
18	1018	1	L 51	23.	000	NUP								4.40	3.99	0	Ő	
19	1019	1	L 34	05.	ر ا									4.80	4.55	1	0	
20	1020	9	32	20.	9 💦	7								5.14	5.35	0	0	
21	1021	1	L 50	16.	MuDoc	umante								5.38	4.99		1	
23	1022	ċ	1 49	05.	ing bloc	Jumping								4 68	3.87	0	0	
24	1025	č	45	09.	1 -	6								4.34	4.26	ŏ	ŏ	
25	1026	1	L 46	17.	d 🛌									5.17	4.45	0	0	
26	1027	1	L 56	10.	MuCo	mnuter								3.57	3.69	0	0	
2/	1028	1	L 26	01.	9									4.58	4.87	0	0	
20	1029	1	L 54	12		1	El		Constructor			121	C.m.n	4.51	4.03 E 06	1	1	
30	1031	1	34	20		2	File rialite.		proundly.xis			<u> </u>	5400	6.06	5.69	Ô	ŏ	
31	1032	Č	50	02.	My Ne	etwork	Save as typ	e:	Excel 8/97 (*	.xls)		-	Cancel	3.88	4.13	0	1	
32	1033	1	L 53	10.	1 Pla	ces							1	3.60	3.61	0	0	
33	1034	9	52	09.	9								Help	4.70	4.94	0	0	
34	1036	2	1 42	16.	4 1969	21	17 1992	.,	112 1992		177	7.69	1 5 1	4.59	3.88	0	0	
36	1037	6	1 45	26	06 1947	19	19 1992	31	03 1983	0	170	4 31	3.50	5.12	4.26	0	0	
37	1039	1	1 38	15.	01.1953	22.1	01.1991	25	.01.1974	1	170	3.98	3.72	5.11	4.46	Ő	Ő	
38	1040	C) 47	09.	01.1946	20.1	03.1993	13	.01.1971	Ō	180	3.97	3.87	5.18	4.78	Ő	Ő	
39	1042	9	24	23.	07.1966	30.1	09.1990	24	.07.1982	1	180	4.80	4.51	6.27	5.36	0	0	
40	1043	9	J 35	11.	07.1954	11.1	J8.1989	12	. U6.1974	1	175	4.16	4.00	4.84	4.81	1	0	
41	1044	1	1 40	19	02.1944	27	10 1007	20	.04.1970	1	175	2.72	3.33	3.90	4.11	0	0	
43	1045	1	1 38	01	01.1950	12	33.1988	31	.01.1981	1	190	4,81	4.56	5,90	5.59	n n	1	
44	1047	i	47	21	08.1945	10.1	06.1992	24	.03.1986	î	175	3.59	3.66	4.88	4.49	Ő	Ô	
45	1048	1	L 39	07.	07.1955	04.1	09.1994	25	.06.1982	1	190	5.25	4.53	6.74	5.57	0	0	
46	1049	G	62	16.	03.1929	11.0	05.1991	13	.03.1983	0	175	3.04	2.74	3.73	3.49	0	0	
4/	1050	9	J 36	17.	07.1961	23.1	J9.1997	02	.08.1991	1	173	3.68	3.89	4.69	4.67	0	0	
40	1051	1	35	20.	03.1955	20	11 1982	20	10 1979	1	190	3.97 5.05	3.80	5.97	3.80	0	0	
50	1052		34	31	12.1959	11.	12.1993	29	10.1976	0	182	4.54	4.33	5.65	5.24	0	0	
ITAN.	foundry	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~						/			2.75							, Č

The standard **Save As** dialogue box will be displayed as above. You have to choose a drive and a suitable name. For Workbook file the file extension is **.sdw**. Then click the **Save** button. Say, you saved the workbook with the name called **survey1** in the pen drive.

Backup The Work

When saving Workbook or Report, make sure that you save onto at least two USB pen drives just in case one of your USB pen drives fail. To make a backup copy of your work repeat the **Save As** procedure.

Retrieving a StatsDirect File

Having returned to the computer, access the system as before and click on the **File** option from the menu bar, then on the **Open File** sub-option.

To open a copy of the **StatsDirect** file on your pen drive, under **Drives:** click on \checkmark in the **Look in** window to generate a list of the drives. Select the appropriate **Pen Drive** where you saved your file, a list of file names will be displayed. Then choose **survey1.sdw** and click **Open** button (shown below)



MODIFYING THE DATA

StatsDirect contains the same functions for inserting and deleting rows and columns as a standard spreadsheet such as EXCEL. These are found under the Edit and Insert menus. It is possible to move columns and rows using cut and paste. StatsDirect also contains some additional functions to calculate new variables from old variables.

Modifying a new variable by Search and Replace

The Search and Replace option in the data menu allows more complex editing than that available in the standard editing menu. It should be noted that this is not as powerful as the recode procedures in SPSS or STATA but it still allows you to do some simple recoding.

In the exercise above we noted that number of cigarettes per day is recorded in the data set with -88 for missing. You might want to change this to the missing value code *, if you want to calculate the mean numbers of cigarettes smoked by smokers. Alternatively you might change this to 0 if you wanted to calculate the mean number of cigarettes smoked by each worker regardless of whether they were a smoker. For this part of the tutorial we will change the "-88" in the variables **cigno** to *. Select the column and search and replace.

	Toundy Image: Close CCION A B 1070 GROUI 1071 1091 1034 1086 1018 1018 1048 1048 1052 1052 1052 1 1026 1 1027 1 1030 1 1026 1 1030 1 1027 1 1028 1 1095 1 1030 1 1027 1 1039 1 1039 1 1039 1 1039 1 1039 1	Sort Random Number Pairwise Grouping Apply Function Clear Missing Do Date 8, Time Inte Durmy Variable: Fill Series Normal Scores: Rank Rotate Data Bios Standardized Data Bios Standard	sta sta ervals s ck ck ck ck 1 0 0 0 1 1 1 1 1	HT 177 160 165 165 175 170 170 170 170 170 170 170 170 170 170	F FEVMEAS 5.61 4.51 4.62 4.60 5.56 3.77 4.79 5.25 4.26 4.27 3.97 5.05 3.63 3.43 4.01 4.46 4.68 3.84 3.84	G FEVPRED 4 21 3 45 3 72 3 73 4 67 3 24 4 51 3 69 3 53 4 50 3 53 4 50 3 24 4 51 3 07 3 24 4 51 3 47 3 59 4 01 4 22 3 47 3 47	H FVCHEAS 7 .21 5 .08 5 .31 5 .82 6 .22 6 .22 6 .22 6 .22 5 .53 6 .74 5 .45 5 .16 5 .01 6 .26 5 .01 6 .26 5 .01 6 .26 5 .01 6 .26 5 .01 5 .53 6 .74 5 .22 5 .11 6 .26 5 .01 5 .53 6 .74 5 .02 5 .01 5 .53 6 .74 5 .02 5 .01 5 .02 5 .01 6 .02 5 .01 5 .02 5 .01 6 .02 5 .01 6 .02 5 .01 6 .02 5 .01 5 .02 5 .01 6 .02 5 .01 6 .02 5 .01 6 .02 5 .01 6 .02 5 .01 6 .02 6 .02 5 .01 6 .02 5	I FVCFRED 5 03 4 02 4 02 4 36 5 54 5 54 4 54 5 54 4 42 4 29 5 54 4 42 4 03 3 82 4 45 4 45 4 45 4 45 4 45 4 45 5 54 6 57 5 50 6 10 5 50 5 50 5 70 5 70 5 70 5 70 5 70 5 7	J ASTHMA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	K BRON S 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	L SHKNOV SHI 1 0 1 1 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0	M 22 0 1 1 2 2 2 2 2 2 2 2 0 1 1 0 0 0 0	N - (G)(Q) - 88 - 88 - 88 - 88 - 88 - 88 - 88	0 CIGYRS 5 -88 15 5 5 5 5 5 5 88 -88 -88 -88 -88 -88 -88	P EMPYRS R 12 5 7 7 8 9 9 12 8 8 12 12 19 9 9 9 10 5 6
	CIGNO A B 1070 GROU 1071 1071 1071 1070 1071 1071 1070 1071 10700 1070 1070 1070 1070 1070 1070 1070 1070 1070 1070 1070 1070	Random Number Pairwise Grouping Apply Function Clear Missing Da Date & Time Inte Dummy Variables Fill Series Normal Scores Rank Rotate Data Bio Search & Replac Standardization Text to Numbers 5 5 1 5 4 1 5 6 1 3 2 1 3 9 1 4 6 1 3 2 1 3 9 1 4 6 1 3 2 1 3 9 1 4 6 1 3 2 1 3 9 1 4 6 1 3 2 1 3 9 1 4 6 1 3 2 1 3 9 1 4 5 6 1 3 2 1 3 9 1 3 9 1 4 5 6 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1	rs + + + + + + + + + + + + + + + + + + +	HT 177 160 165 174 185 176 175 170 170 170 170 170 170 170 170 170 165 173 173 175 173 175	F FEVHEAS 5.61 4.62 5.56 3.77 4.79 5.25 4.26 4.27 3.97 5.05 3.63 3.43 4.01 4.46 4.68 3.84 3.84	G FEVPRED 4 21 3 75 3 73 4 67 3 24 4 53 3 69 3 70 3 53 3 69 3 70 3 53 3 24 3 07 3 24 4 01 4 22 3 47 3 59 4 01 4 22 3 47 3 47	H FVCMEAS 7 .21 5 .08 5 .31 5 .82 6 .22 4 .40 5 .53 6 .24 4 .51 6 .26 4 .51 6 .26 4 .51 6 .26 4 .51 1 4 .22 5 .17 5 .92 5 .27 5 .92 4 .96	I FVCPRED 5 03 4 5 3 9 4 5 5 5 5 5 5 4 29 5 5 4 29 5 4 4 29 5 4 4 29 5 4 4 29 5 4 4 29 5 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7	J ASTHMA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	K BRON S 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	L SMKNOW SMI 1 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0	M KEVER 0 2 0 1 1 1 2 2 2 2 2 2 0 0 1 1 0 0 0 0	N 	0 CIGYRS 5 -88 4 15 -88 17 6 9 5 5 5 5 5 5 5 5 88 -88 -88 -88 -88 -88	P EMPYRS R 5 7 8 9 9 23 3 9 9 12 8 12 8 12 9 9 9 9 10 5 5 6
	CIGNO A B DNO GROU 1070 1091 1134 1134 1133 1066 1086 1018 1133 1048 1061 1048 1052 1135 1029 1135 1029 1135 1029 1135 1029 1140 1052 1140 1052 1140 1052 1140 1095 11007 1007 1	Pairwise Grouping Apply Function Clear Missing Do Date & Time Inte Dummy Variables Fill Series Normal Scores Rank Rotate Data Bloc Standardzaton Text to Numbers Standardzaton Text to Numbers Standardzaton Text to Numbers Standardzaton Text to Numbers Standardzaton Stal Stal Stal Stal Stal Stal Stal Stal	+ + + + + + + + + + + + + + + + + + +	HT 177 160 165 174 185 168 175 170 170 170 190 170 190 170 170 170 170 175 173 178 168 175 173 173	F FEVMEAS 5.61 4.51 4.60 5.56 3.77 4.79 5.25 4.26 4.27 3.97 5.05 5.3 63 3.43 4.01 4.46 4.68 3.84 3.84 3.84	G FEVPRED 3 45 3 72 3 73 4 67 3 24 4 12 4 53 3 69 3 70 3 53 4 50 3 70 3 53 4 50 3 24 4 01 4 22 3 59 4 01 4 22 3 47 3 47 3 47 3 47 3 47 3 47 3 47 3 47	H FVCMEAS 7.21 5.08 5.82 6.22 4.40 5.53 6.74 5.42 5.42 5.42 5.16 6.26 4.51 4.22 5.17 5.27 5.92 5.27 5.92 4.96	I FVCPRED 5 03 4 5 3 9 4 5 3 9 4 5 5 7 5 5 4 4 2 5 5 5 5 4 2 5 5 5 5 5 5 5 6 7 5 6 4 5 6 6 6	J ASTHMA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	K SRON S 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	L SMKNOW SMI 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0	M 2 2 1 1 1 2 2 2 2 2 2 0 1 1 0 0 0 0 0 0	N - (G)(0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 -	0 CIGYRS 5 -88 4 15 -88 15 15 25 -88 30 -88 -88 -88 -88	P EMPYRS R 5 7 8 9 23 9 23 9 23 9 23 8 12 8 12 12 9 9 9 10 5 6
	A B IDNO GROUI 1070 1091 1091 1134 1006 1086 1018 1133 1048 1016 1052 1061 1052 11052 1029 1 1026 1 1095 1 1087 1 10807 1 1097 1 1097 1 1097 1 1097 1 1097 1 1097 1 1097 1 1097 1	Grouping Apply Function Clear Missing Da Date & Time Inte Fill Series Normal Scores Rank Rotate Data Bilos Standardization Text to Numbers 56 1 56 1 56 1 32 1 39 1 46 1 32 1 39 1 46	+ ata ervals s ck cc 1 0 0 1 1 1 1 1 1	HT 177 160 165 174 185 168 175 170 170 170 190 170 190 170 190 170 190 170 168 175 173 178 165 173	F FEVMEAS 5.61 4.51 4.62 4.60 5.56 4.60 5.25 4.60 4.27 3.97 5.05 3.63 3.43 4.01 4.46 4.68 4.28 3.43 4.01 4.46 4.68 4.69 5.38 4.27 5.05 5.3,61 4.45 5.55 5.55 5.55 5.55 5.55 5.55 5.5	G FEVPRED 4 21 3 72 3 73 4 67 3 24 4 53 3 70 3 24 4 53 3 70 3 59 3 70 3 53 4 50 3 24 4 50 3 59 4 01 4 22 3 59 4 01 4 22 3 47 3 59 4 01 4 22 3 47 3 73 3 73 3 74 3 75 3 75 3 75 3 75 3 75 3 75 3 75 3 75	H FVCHEAS 7 . 21 7 . 20 5 . 82 6 . 22 4 . 40 5 . 53 6 . 74 5 . 42 5 . 16 6 . 26 6 . 4 . 51 4 . 22 5 . 17 5 . 92 6 . 27 5 . 92 6 . 4 . 60 7 . 92 7 . 9	I FVCPRED 5.03 4.02 4.36 5.59 3.99 4.91 5.57 4.29 5.54 4.42 5.54 4.03 3.82 4.45 4.77 5.06 4.15	J ASTHMA 0 0 0 0 0 0 0 0 0 0 0 0 0	K BRON S 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	L SMKNOW SHI 1 0 1 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0	M KEVER 2 0 2 1 1 1 1 2 2 2 2 2 2 2 0 0 1 1 0 0 0 0	N 	0 CIGYRS -88 4 15 -88 17 6 9 5 15 25 -88 -88 30 -88 -88 -88	P EMPVRS R 12 5 7 8 9 23 9 23 9 12 8 12 12 8 12 19 9 9 10 5 6
	A B IDNO GROUI 1070 GROUI 1071 1091 1134 1006 1086 1018 1018 1133 1048 1061 1052 1052 1042 1 1040 1052 1052 1 1052 1 1049 1 1052 1 1030 1 1030 1 1027 1 1039 1 1039 1 1037 1	Apply Function Clear Missing Date & Time Inte Durmmy Variables Fill Series Normal Scores Rank. Rotate Data Bloc Search & Replat. Standardization Text to Numbers 1 56 1 32 1 39 1 46 1 29 1 56	ata ervals s ck c t 0 0 0 1 1 1 1 1 1	HT 177 160 165 174 185 176 175 170 170 170 170 170 170 170 170 170 170	F FEVMEAS 5.61 4.62 4.60 5.56 3.77 4.79 4.26 4.27 3.97 5.05 3.63 3.43 4.01 4.68 4.68 3.84 3.84	G FEVPRED 4 21 3 45 3 72 3 73 4 67 3 24 4 12 4 53 3 69 3 59 3 59 4 01 4 22 3 47 3 47 3 47 3 47 3 47 3 47 3 47 3 47	H FVCHEAS 7 21 5 08 5 31 5 82 6 22 6 22 6 22 6 22 6 22 5 53 6 74 5 53 6 74 5 51 6 26 5 51 6 22 5 51 6 22 5 51 6 22 5 51 7 52 5 51 6 22 5 51 7 21 5 51 5 51 5 51 5 51 5 51 5 51 5 51 5	FVCPRED 5.03 4.02 4.36 4.54 5.59 3.99 4.91 5.57 4.52 4.44 4.29 5.54 4.44 4.29 5.54 4.42 4.29 5.54 4.42 4.29 5.54 4.42 4.77 5.06 4.77 5.06 4.15	J ASTHMA 0 0 0 0 0 0 0 0 0 0 0 0 0	K BRON S 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	L SHKNOV SHI 5MKNOV SHI 0 1 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0	M KEVER 2 0 2 1 1 1 2 2 2 2 2 2 2 2 0 0 1 1 0 0 0 1 0 0 0 0	N C1GN0 100 -88 15 200 -88 40 10 3 20 3 20 30 -88 20 -88 20 -88 -88 -88 -88	0 CIGYRS 5 -88 4 15 -88 17 6 9 5 15 25 -88 30 -88 30 -88 -88 -88	PP EMPYRS R 12 5 7 8 9 23 9 23 9 12 8 12 12 19 9 9 10 5 6
	IDNO GROUI 1070 1091 1091 1091 1086 1018 1016 1018 1016 1018 1052 1 1052 1 1052 1 1052 1 1026 1 1030 1 1030 1 1022 1 1010 1 1027 1 1039 1 1097 1 1097 1 1097 1 1097 1 1097 1 1097 1	Clear Missing Da Date & Time Inte Jummy Variables Fill Servies Normal Scores Rank Rotate Data Blot Search & Replac Standardization Text to Numbers Text to Numbers 1 54 1 56 1 32 1 39 1 46 1 29 1 56	ata ervals s ck te 1 0 0 0 1 1 1 1 1	HT 177 160 165 174 185 176 175 190 170 170 170 170 170 170 170 17	FEVMEAS 5.61 4.51 4.62 5.56 3.56 3.77 4.79 5.25 4.27 3.97 5.05 3.63 3.4.01 4.46 4.68 3.84 3.84	FEVPRED 4 21 3 45 3 72 3 73 4 67 3 24 4 12 4 53 3 69 3 700 3 53 4 50 3 53 4 50 3 24 3 02 4 3 02 3 53 4 50 3 24 3 02 4 3 02 3 4 7 3 7 4 7 3 7 3 7 3 7 3 7 4 7 3 7 3 7 3 7 3 7 3 7 3 7 3 7 4 7 3 7 4 7 3 7 3 7 3 7 3 7 4 7 3 7 4 7 3 7 3 7 3 7 3 7 3 7 4 7 3 7 4 7 3 7 4 7 3 7 4 7 3 7 4 7 3 7 3 7 7 3 7 7 3 7 7 3 7 7 3 7 7 3 7 7 7 7	FVCHEAS 7.21 5.08 5.31 5.82 6.22 4.40 5.53 6.74 5.53 6.74 5.51 6.74 5.16 5.01 6.26 4.51 6.26 4.51 7.52 5.27 5.27 5.27 5.92 5.72	FVCPRED 5 03 4 02 4 36 5 59 3 99 4 91 5 57 4 52 4 491 5 57 4 52 4 491 5 57 4 52 4 491 5 57 4 52 4 42 4 03 3 82 4 45 4 77 5 06 4 15	ASTHMA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	BRON 5 1 0 0 0 0 0 0 0 0 0 0 0 0 0	SMKNOW SHI 0 1 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0	KEVER 2 2 2 1 1 1 1 2 2 2 2 2 2 2 0 1 0 0 0 0	C+CSN0 -88 -15 20 -88 40 10 -88 -88 -88 -88 -88 -88 -88	CIGYRS 5 -88 4 15 -88 17 6 9 9 5 15 25 25 25 25 88 30 -88 30 -88 -88 -88	EMPYRS R 12 5 7 8 9 23 8 9 12 8 9 12 19 9 10 5 6
	1070 1091 1134 1086 1086 1088 1088 1133 1048 1048 1048 1048 1048 1052 1135 1029 1135 1020 1140 1052 1055 1075 1075 1087 1087 1087 1087 1087 1087 1095 1087 1095 1005 1095 1005 1095 1005	Date & Time Inte Dummy Variables Fill Series Normal Scores Rank Rotate Data Bloc Standardization Text to Numbers 1 56 46 32 1 46 29 56 456	ervals s ck te s 1 0 0 1 1 1 1 1 1	177 160 165 174 185 175 170 175 170 170 170 170 170 170 170 170	5, 61 4, 62 4, 60 5, 56 3, 77 4, 79 5, 25 4, 26 4, 27 3, 97 5, 05 3, 63 3, 43 4, 01 4, 46 4, 68 3, 84 3, 84	4.21 3.45 3.72 3.73 4.67 3.24 4.12 4.53 3.69 3.70 3.53 4.50 3.24 4.53 3.70 3.53 4.50 3.24 4.53 3.70 3.53 4.50 3.24 4.50 3.24 4.53 3.24 4.53 3.24 4.53 3.72 3.24 4.53 3.72 3.24 4.53 3.72 3.24 4.53 3.72 3.24 4.53 3.72 3.72 3.72 3.72 3.72 3.72 3.72 3.7	7.21 5.08 5.31 5.82 6.22 4.40 5.53 6.74 5.42 5.16 6.26 4.51 6.26 4.51 4.22 5.17 5.27 5.92 4.96 6.21	$\begin{array}{c} 5 & 0.3 \\ 4 & 0.2 \\ 4 & 36 \\ 5 & 59 \\ 3 & 99 \\ 4 & 91 \\ 5 & 57 \\ 4 & 52 \\ 4 & 44 \\ 4 & 29 \\ 5 & 54 \\ 4 & 03 \\ 3 & 82 \\ 4 & 45 \\ 4 & 77 \\ 5 & 06 \\ 4 & 15 \\ \end{array}$	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		1 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0	2 2 1 0 1 1 2 2 2 2 2 0 1 0 0 0 1 0 0 0 0	-88 15 20 -88 40 10 30 20 15 30 -88 20 -88 -88 -88 -88 -88	5 -88 4 15 -88 17 6 9 5 5 5 5 5 5 5 25 25 -88 30 -88 -88 -88 -88	12 5 7 8 9 23 9 12 8 12 19 9 9 9 9 10 5 6
	1134 1006 1006 1018 1133 1048 1018 1133 1048 1051 1052 1029 1140 1029 1140 1029 11026 1039 1002 1002 1002 1002 1002 1002 1002 1005	Dummy Variables Fill Series Normal Scores Rank Rotate Data Bloc Search & Raplac Standardization Text to Numbers 1 54 1 56 1 32 1 32 1 39 1 46 1 29 1 56	s ck s 1 0 0 1 1 1 1 1 1	165 174 185 168 175 190 175 170 170 170 170 170 170 168 175 173 178 165 170	4.62 4.60 5.56 3.77 4.79 5.25 4.26 4.27 3.97 5.363 3.63 3.43 4.01 4.46 4.68 3.84 3.84	3.72 3.73 4.67 3.24 4.12 4.53 3.69 3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47 3.47	5.31 5.82 6.22 4.40 5.53 6.74 5.42 5.16 5.01 6.26 4.51 4.22 5.17 5.27 5.27 5.29 4.96	$\begin{array}{c} 4 & 36 \\ 4 & 54 \\ 5 & 59 \\ 3 & 99 \\ 4 & 91 \\ 5 & 57 \\ 4 & 52 \\ 4 & 44 \\ 4 & 29 \\ 5 & 54 \\ 4 & 03 \\ 3 & 82 \\ 4 & 45 \\ 4 & 77 \\ 5 & 06 \\ 4 & 15 \end{array}$	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			2 1 0 1 2 2 2 2 0 1 0 0 0 0	15 20 -88 40 10 30 20 15 30 -88 20 -88 -88 -88 -88 -88	4 15 -88 17 6 9 5 15 25 -88 30 -88 -88 -88 -88	7 8 9 12 8 12 8 12 9 9 9 9 9 9 10 5 6
	1006 1006 1018 1133 1148 1133 1046 1061 1059 1135 1052 1140 1029 1140 1030 1 1032 1 1040 1 1040 1 1040 1 1040 1 1040 1 1040 1 1047 1 1047 1 1049 1 1040 1 1047 1 1049 1 1043 1	Fill Series Normal Scores Rank Rotate Data Bloc Standardizer Standardizer 1 54 1 56 1 46 1 32 1 39 1 46 1 32 1 39 1 46 1 29 1 56	ck 5 1 0 0 1 1 1 1 1 1	174 185 168 175 190 175 170 170 190 170 170 170 170 170 170 175 173 178 165 170	4.60 5.56 3.77 4.79 5.25 4.26 4.26 4.27 3.97 5.05 3.63 3.43 4.01 4.46 4.68 3.84 3.84	3.73 4.67 3.24 4.12 4.53 3.69 3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	5.82 6.22 4.40 5.53 6.74 5.42 5.16 5.01 6.26 4.51 4.22 5.17 5.27 5.27 5.92 4.96	$\begin{array}{c} 4.54\\ 5.59\\ 3.99\\ 4.91\\ 5.57\\ 4.52\\ 4.44\\ 4.29\\ 5.54\\ 4.03\\ 3.82\\ 4.45\\ 4.77\\ 5.06\\ 4.15\end{array}$	0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 1 1 1 1 0 0 0 0 0 0 0 0	1 0 1 2 2 2 2 0 1 0 0 0 0	20 -88 40 10 3 20 15 30 -88 20 -88 -88 -88 -88	15 -88 17 6 9 5 15 25 -88 30 -88 -88 -88 -88 -88	8 9 23 9 12 8 12 19 9 9 9 10 5 6
	1086 1 1018 1 1133 1 1661 1 1059 1 1055 1 1052 1 1040 1 1059 1 1040 1 1059 1 1095 1 1087 1 1087 1 1087 1 1000 1 1001 1 1097 1 1039 1 1000 1 1097 1 1039 1 1039 1	Normal Scores Rank Rotate Data Blor Standardization Text to Numbers 54 54 54 56 1 32 1 32 1 32 1 39 1 46 1 29 1 56	ck 5 1 0 0 1 1 1 1 1 1	185 168 179 170 170 170 170 170 170 170 170 170 170	5.56 3.77 4.79 5.25 4.26 4.27 3.97 5.05 3.63 3.43 4.01 4.46 4.68 3.84 3.84	4.67 3.24 4.12 4.53 3.69 3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47 3.47	6.22 4.40 5.53 6.74 5.42 5.16 5.01 6.26 4.51 4.22 5.17 5.27 5.92 4.96	5.59 3.99 4.91 5.57 4.52 4.44 4.29 5.54 4.03 3.82 4.45 4.77 5.06 4.15	0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		0 0 1 1 1 1 0 0 0 0 0 0 0 0 0	0 1 2 2 2 2 0 1 0 0 0 0 0	-88 40 10 3 20 15 30 -88 20 -88 -88 -88 -88	-88 17 6 9 5 15 25 -88 30 -88 -88 -88 -88	9 23 9 12 8 12 19 9 9 9 9 10 5 6
	1018 1133 1140 1061 1099 1135 1052 1029 1140 1095 1030 1030 1022 10102 1010 1010 1097 1010 1030 1010 1022 1010 1039	Rank Rotate Data Bloo Search & Replac Standardization Text to Numbers 1 54 1 56 1 46 1 32 1 32 1 32 1 39 1 46 1 29 1 56	ck s 1 0 0 1 1 1 1 1 1	168 175 190 175 170 170 170 170 170 168 175 173 178 165 170 120	3 . 7 / 4 . 7 / 5 . 25 4 . 26 4 . 27 3 . 97 5 . 05 3 . 63 3 . 43 3 . 43 3 . 43 4 . 01 4 . 46 4 . 68 3 . 84	3.24 4.12 4.53 3.69 3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	4.40 5.53 6.74 5.42 5.16 5.01 6.26 4.51 4.22 5.17 5.27 5.92 4.96	3.99 4.91 5.57 4.52 4.44 4.29 5.54 4.03 3.82 4.45 4.77 5.06 4.15	0 0 1 0 0 0 0 0 0 0 0 0 0 1 0		0 1 1 1 0 0 0 0 0 0 0 0	1 1 2 2 2 2 2 0 1 0 0 0 0 0 0	40 10 3 20 15 30 -88 20 -88 -88 -88 -88	17 6 9 5 15 25 -88 30 -88 -88 -88 -88	23 9 12 8 12 19 9 9 9 10 5 6
	1045 1041 1061 1099 1135 1052 1029 1140 1126 1030 1030 1087 1022 10100 10100 1097 1097 1097 1099	Starth 201 Starth	ck ce 5 1 0 0 1 1 1 1 1 1	170 175 170 170 170 170 170 168 175 173 178 165 170	5,25 4,26 4,27 3,97 5,05 3,63 3,43 4,01 4,46 4,68 3,84 3,81	4.53 3.69 3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	6.74 5.42 5.16 5.01 6.26 4.51 4.22 5.17 5.27 5.92 4.96	5,57 4,52 4,44 4,29 5,54 4,03 3,82 4,45 4,77 5,06 4,15	0 1 0 0 0 0 0 0 0 0 0 0 1		1 1 1 0 0 0 0 0 0 0	2 2 2 2 0 1 0 0 0 0	20 15 30 -88 20 -88 -88 -88	9 5 15 25 -88 30 -88 -88 -88 -88	12 8 12 19 9 9 10 5 6
	1061 1099 1135 1029 1140 1026 1095 10030 1087 10102 10102 10107 1010 1097 1010 1093 1039	Search & Replac Standardization Text to Numbers 5 4 1 5 6 1 4 6 1 3 2 1 3 2 1 3 2 1 3 9 1 4 6 1 2 9 1 5 6	s 1 0 0 1 1 1 1 1 1	175 170 170 190 170 168 175 173 178 165 170	4.26 4.27 3.97 5.05 3.63 3.43 4.01 4.46 4.68 3.84 3.84	3.69 3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	5.42 5.16 5.01 6.26 4.51 4.51 5.17 5.27 5.92 4.96	4.52 4.44 4.29 5.54 4.03 3.82 4.45 4.77 5.06 4.15	1 0 0 0 0 0 0 0 1	0 0 0 1 0 0 0 0 0	1 1 0 0 0 0 0 0	2 2 0 1 0 0 0 0	20 15 30 -88 20 -88 -88 -88 -88	5 15 25 -88 30 -88 -88 -88 -88	8 12 19 9 9 10 5 6
	1099 1135 1052 1029 1140 1095 1095 1006 1095 1007 1010 1010 1010 1010 1039 1039	Standardization Text to Numbers 1 54 1 56 1 46 1 32 1 39 1 46 1 39 1 46 1 29 1 56	s 1 0 1 1 1 1 1 1	170 170 190 170 168 175 173 173 178 165 170	4.27 3.97 5.05 3.63 4.01 4.46 4.68 3.84 3.81	3.70 3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	5.16 5.01 6.26 4.51 4.22 5.17 5.27 5.92 4.96 4.96	4.44 4.29 5.54 4.03 3.82 4.45 4.77 5.06 4.15	0 0 0 0 0 0 1	0 0 1 0 0 0 0	1 0 0 0 0 0 0	2 2 0 1 0 0 0 0	15 30 -88 20 -88 -88 -88 -88	15 25 -88 30 -88 -88 -88	12 19 9 10 5 6
	1135 1052 1029 1140 1026 1095 1005 1007 1007 1010 1010 1099 1010 1010 1099 1097 1099	Text to Numbers 1 54 1 56 1 46 1 32 1 32 1 39 1 46 1 29 1 56	s 1 0 1 1 1 1 1 1	170 190 170 168 175 173 178 165 170 170	3.97 5.05 3.63 4.01 4.46 4.68 3.84 3.81	3.53 4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	5.01 6.26 4.51 4.22 5.17 5.27 5.92 4.96 4.96	4.29 5.54 4.03 3.82 4.45 4.77 5.06 4.15	0 0 0 0 0 0 1 0	0 0 1 0 0 0	1 0 0 0 0 0	2 0 1 0 0 0	30 -88 20 -88 -88 -88	25 -88 30 -88 -88 -88	19 9 10 5 6
	1029 1 1029 1 1140 1 1026 1 1030 1 1087 1 1010 1 1010 1 1097 1 1039 1	1 54 1 56 1 46 1 32 1 32 1 39 1 46 1 29 1 56	1 0 1 1 1 1 1 1	170 168 175 173 178 165 170	3.63 3.43 4.01 4.46 4.68 3.84 3.81	4.50 3.24 3.07 3.59 4.01 4.22 3.47 3.47	4.51 4.22 5.17 5.27 5.92 4.96	3.54 4.03 3.82 4.45 4.77 5.06 4.15	000000000000000000000000000000000000000	0 1 0 0	0 0 0 0	0	-88 -88 -88	-88 -88 -88	9 10 5
	1140 1 1026 1 1095 1 1030 1 1087 1 1022 1 1010 1 1097 1 1097 1 1039 1	1 56 1 46 1 32 1 32 1 39 1 46 1 29 1 56	0 0 1 1 1 1 1 1	168 175 173 178 165 170	3.43 4.01 4.46 4.68 3.84 3.81	3.07 3.59 4.01 4.22 3.47 3.47	4.22 5.17 5.27 5.92 4.96	3.82 4.45 4.77 5.06 4.15	0 0 1	0 0 0	0	0	-88 -88 -88	-88 -88 -88	10
	1026 1 1095 1 1030 1 1087 1 1022 1 1010 1 1097 1 1039 1	1 46 1 32 1 32 1 39 1 46 1 29 1 56	0 1 1 1 1 1	175 173 178 165 170	4.01 4.46 4.68 3.84 3.81	3.59 4.01 4.22 3.47 3.47	5.17 5.27 5.92 4.96	4.45 4.77 5.06 4.15	0 0 1	0 0 0	0	0	-88 -88	-88 -88	5
	1095 1 1030 1 1087 1 1022 1 1010 1 1097 1 1039 1	1 32 1 32 1 39 1 46 1 29 1 56	1 1 1 1	173 178 165 170	4.46 4.68 3.84 3.81	4.01 4.22 3.47 3.47	5.27 5.92 4.96	4.77 5.06 4.15	0	0	0	0	-88	-88	6
	1030 1 1087 1 1022 1 1010 1 1097 1 1039 1	1 32 1 39 1 46 1 29 1 56	1 1 1 1	1/8 165 170	4.68	4.22	4.96	4.15	1	0				0.01	
	1022 1 1010 1 1097 1 1039 1	1 46 1 29 1 56	1	170	3.81	3 47	4.50 E.10	4.13		0	0	1	-88	-88	17
	1010 1 1097 1 1039 1	1 29 1 56	1	120			5 1 3	4 24	0	0	0	ń	-88	-88	
	1097 1 1039 1	1 56		175	4.50	4.18	5.68	4.97	Ő	Ô	0	1	20	8	9
	1039 1		0	177	3.74	3.48	4.58	4.38	0	0	0	1	60	15	19
	2 B K K B K K K K K K K K K K K K K K K	1 38	1	170	3.98	3.72	5.11	4.46	0	0	0	1	13	4	17
	1029 1	1 37	1	172	4.19	3.95	5.05	4.75	0	0	1	2	_ 0 0		/
	1075 1	1 31	1	180	4.58	4.34	5.39	5.20	0	0	1	2	30	10	8
	1046 1	1 38	1	190	4.81	4.56	5.90	5.59	0	1	1	2	25	20	7
	1031 1	1 34	1	190	4.91	4.68	6.06	5.69	0	0	1	2	12	13	9
	1085 1	1 32	1	182	4.59	4.39	5.96	5.29	0	0	0	1	20	10	- 8
	1126 1	1 29	0	178	4.05	3.88	4.69	4.56	0	0	1	2	15	4	9
	1148 1	1 31	ŏ	168	3.96	3.82	4.87	4.51	Ő	ŏ	Ô	1	20	11	Ś
	1078 1	1 59	1	173	3.30	3.19	4.24	4.09	0	0	0	0	-88	-88	7
	1059 1	1 50	1	170	3.46	3.36	4.40	4.13	0	0	1	2	25	32	22
	1019 1	1 24	1	175	3.80	3.71	4.75	4.55	1	0	1	2	20	10	21
	1021 1	1 50	Ó	185	4.04	3.99	5.38	4.99	0	1	ò	1	40	32	19
	1094 1	1 50	1	163	3.08	3.05	3.68	3.73	Ő	ō	Ō	0	-88	-88	10
	1054 1	1 49	0	170	3.40	3.38	4.11	4.16	0	0	0	1	20	16	16
	1129 1	1 28	0	175	4.22	4.21	4.96	4.99	0	0	0	0	-88	-88	9
	1011 1	1 37	1	177	4.01	4.02	e.81 5.61	4.68	0	0	1	2	-00	17	9
	1153 1	1 33	Ô	180	4.21	4.28	5.89	5.15	Ő	1	1	2	25	20	17
	1047 1	1 47	1	175	3.59	3.66	4.88	4.49	Ő	0	0	0	-88	-88	6
	1131 1	1 49	1	172	3.39	3.47	4.05	4.27	0	0	0	1	8	24	24
	1122 1	1 53	0	175	3.39	3.48	5.30	4.34	0	0	1	2	30	37	- 7
1	1099 1	1 32	1	171	3.78	4.29	4 55	5.07	0	1	1	2	20	12	16
1	A 4 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	A		170	3.70	3,88	5,40	4.59	ő	1	1	2	40	26	10

When the panel below appears insert the values as shown. It is worth clicking on **Count**, to check how many values satisfy the condition. Then click Replace to complete the change.

Count Replace Delete Help Exit Search for entries Image: Search for equal to Search for equal to Search for equal to	🗗 Search		X
C less than -88 C greater than or equal to Replace with: C not equal to 1	<u>C</u> ount <u>Replace</u> Search for entries	Delete Help Search for: -88 Replace with: 1	Exit Search mode Numerical Text

You can also code a numeric value to a string value and vice versa by this **Search and Apply** method. Suppose you want to code the variable **sex**, 0 as male and 1 as female. The screen below is an example.

d Search			×
<u>C</u> ount <u>R</u> eplace	Delete	<u>H</u> elp	Exit
equal to greater than greater than greater than greater than or equal to less than or equal to not equal to match expression	Search for: 0 Replace with: male	Sea O	arch mode Numerical Text

This recoding is very useful for some analyses and graphs, such as crosstabs boxplots etc.

Creating a new variable by calculation

It is recommended that you store your data to disk just in case you make a mistake. Suppose you want to convert the variable HT in meters. To do that you highlight the variable HT, then click **Data** then **Apply Function.** The following screen will be displayed.

Apply function to data											
elp <u>C</u> ancel											
V1: HT Enter an expression using V1 etc.											

You can see a little window called "Enter an expression using V1 etc." You have to write the mathematical expression there. In our case it should be V1/100 (V1 represents HT). Then press OK button.

Then you have options where to put the variable (shown below).

Select worksheet for output		
	Select	<u>C</u> ancel
Output destination:		
Book 1 Sheet1 New Sheet in Book 1		
foundry.xls foundry New Sheet in foundry.xls		
New Book		

It is always advisable to put the new variable in the existing data sheet, so choose **foundry.xls**, then press **Select** button. The new variable name will be **HT/100** and it will be placed right at the end of the variable list. You can replace the variable called **HT/100** by any suitable name (say, **HTmeter**).

Exercise 1: A variable FEV ratio is need for a later part of this tutorial. This is defined as FEV Measured/ FEV Predicted. Block these two columns on the spreadsheet and then use **Data** then **Apply Function** to construct this new variable.

Exercise 2: Use the integer divide operator \setminus to create a new variable **agegp** grouping ages by decade

[Hint The integer divide operator is \. Select **AGE** then use Apply function in the Data menu with the formula v1\10. Compare the **AGE** and **AGEGP** to check your result]

PART II

CONTINUOUS OUTCOME MEASURES

Perhaps the first step in analysing a continuous variable is to examine its distribution. This can be done graphically using a histogram. Histogram can also be used to detect outlier, that is values very different from the other data points.

Histogram

Histograms are produced for interval variables e.g. age, height etc. Here we will produce a histogram of measured FEV. To do that, click **Graphs** on the menu bar and select **Histogram** from drop down menu. Then select the data and click **OK**. You will get a little window asking for the title for the histogram.

d Histogram			×
	<u>0</u> K	<u>H</u> elp	<u>C</u> ancel
Histogram title			
Histogram for F	EVMEAS		

Choose the title and click **OK** button. Then you will get another window like that below.

d Histogran	n Bin Setup	
Data ma	aximum: 5.61	
Data mir	nimum: 1.45	
Number	of bins: 12	<u>A</u> uto Bins
Minimum bin mi	d-point: 1.4	Auto Mid-pts
Bin mid-point i	interval: 0.4	<u>M</u> anual Reset
Bin values:		
Low 1.2 1.6 2.4 2.8 3.2 3.6 4 4.4 4.8 5.2	Mid-point 1.4 1.8 2.2 2.6 3 3.4 3.8 4.2 4.6 5 5.4	High 1.6 2 2.4 2.8 3.2 3.6 4 4.4 4.8 5.2 5.6
<u> </u>	<u>H</u> elp	Cancel

Click **OK** then you will be asked whether you want to overlay the normal curve. If you say "yes" it will overlay the normal curve, otherwise it will produce only the histogram. Then select the title of the axis and the output destination and you will get the histogram (shown below)

Introduction to StatsDirect, 15/03/2017



Exercise: Draw a histogram of AGE or of HT

We will now consider the lung function measurements. Given that lung function is age and size dependent it is usual to divide measured lung function by the expected lung function.

We now want to examine whether workers exposed to dust have reduced lung function. First you might examine this graphically with a box plot also called a Box-and-Whisker plot. To examine whether dust exposure affects lung function you want a separate Box-and-Whisker for exposed and not exposed separately. Make sure no variables are highlighted. Go to the graph menu, select **Box & Whisker**. Then the following screen will appear. Whether a subject is exposed is given by the identifying variable **Group**. Use **Groups by Identifier** rather than **Groups by Column**. Select the radio button for **Groups by Identifier** if necessary. You will then be asked for the column for the group identifier which in this case is the variable **Group**. Then select the variable for FEV Ratio.

C Stats	Direct -	foundry.xl]													EF	×
II File	Edit Inse	rt Format I	Data An	alysis Graphi	cs Tools N	Mindow Help	0									- 8	×
<u>0</u> K	Ca	ncel Selec	t GROU	IP IDENTIFIE	ERS (Min 1	: Max 20)											
F Batch	mode	Groups by	column	 Groups by 	y identifier	To select non	-adjoined data	, hold down D	rl and use the m	iouse.							
B1	G	ROUP															1
	A	В	С	D	E	F	G	Н	I	J	К	L	M	N	0	Р	
1	IDNO	GROUP	AGE	SEX	HT	FEVMEAS	FEVPRED	FVCMEAS	FVCPRED	ASTHMA	BRON	SHKNOV	SHKEVER	CIGNO	CIGYRS	EMPYRS RES	
2	1070	1	31	0	1//	5.61	4.21	7.21	5.03	0	1	1	2	10	5	12	
4	1134	1	30	, i	165	4.62	3.72	5.00	4.36	0	0	1	2	15	-00	7	
5	1006	1	43	0	174	4.60	3.73	5.82	4.54	Û	0	0	1	20	15	8	
6	1086	1	27	1	185	5.56	4.67	6.22	5.59	0	0	0	0	*	-88	9	
7	1018	1	51	1	168	3.77	3.24	4.40	3.99	0	0	0	1	40	17	23	
9	1048	1	31	1	1/5	4.79	4.12	5.53	4.91	0	0	1	2	10	9	12	
10	1061	i	46	1	175	4.26	3.69	5.42	4.52	1	0	1	2	20	5	8	
11	1099	1	38	1	170	4.27	3.70	5.16	4.44	ō	Ō	1	2	15	15	12	
12	1135	1	44	0	170	3.97	3.53	5.01	4.29	0	0	1	2	30	25	19	
13	1052	1	40	1	190	5.05	4.50	6.26	5.54	0	0	0	0	*	-88	9	
14	1140	1	54	1	168	3.63	3.24	4.51	4.03	0	1	0	1	20	-88	10	
16	1026	î	46	ŏ	175	4.01	3.59	5.17	4.45	Ő	ŏ	Ŭ	Ŭ	*	-88	5	
17	1095	1	32	1	173	4.46	4.01	5.27	4.77	0	0	0	0	*	-88	6	
18	1030	1	32	1	178	4.68	4.22	5.92	5.06	1	0	0	0	*	-88	7	
19	1087	1	39	1	165	3.84	3.47	4.96	4.15	0	0	0	1	20	8	17	
20	1022	1	29	1	170	4.50	3.47	5.13	4.24	0	0	0	1	20	-88	9	
22	1097	î	56	ô	177	3.74	3.48	4.58	4.38	Ő	ŏ	Ŭ	1	60	15	19	
23	1039	1	38	1	170	3.98	3.72	5.11	4.46	0	0	0	1	13	4	17	
24	1065	1	37	0	175	4.19	3.95	5.05	4.76	0	0	1	2	20	15	7	
25	1028	1	26	1	172	4.37	4.14	4.58	4.87	0	0	0	0	*	-88	8	
20	10/5	1	31	1	190	4.58	4.34	5.39	5.20	0	1	1	2	25	20	7	
28	1031	1	34	1	190	4.91	4.68	6.06	5.69	0	ô	1	2	12	13	9	
29	1085	1	32	1	182	4.59	4.39	5.96	5.29	0	0	0	1	20	10	8	
30	1090	1	29	0	168	4.05	3.88	4.69	4.56	0	0	1	2	15	4	9	
31	1126	1	28	0	178	4.51	4.34	5.87	5.16	0	0	1	2	3	8	9	
32	1079		51	1	173	3.96	3.82	4.87	4.51	0	0	0	1	20	_99	2	
34	1059	1	50	1	170	3.46	3.36	4.40	4.13	0	0	1	2	25	32	22	
35	1124	1	45	0	175	3.80	3.71	4.75	4.55	0	0	0	1	10	11	21	
36	1019	1	34	1	170	3.91	3.82	4.80	4.55	1	0	1	2	20	18	14	
37	1021	1	50	0	185	4.04	3.99	5.38	4.99	0	1	0	1	40	32	15	
30	1054	1	50	1	153	3.08	3.05	3.68	3.73	0	0	0	1	20	-88	10	
40	1129	1	28	0	175	4 22	4 21	4.96	4 99	0	0	ů.	1		-88		
41	1114	1	39	0	178	4.01	4.02	4.81	4.88	Ô	Ō	0	0	*	-88	12	
42	1011	1	31	1	177	4.19	4.21	5.61	5.03	0	0	1	2	20	17	9	
43	1153	1	33	0	180	4.21	4.28	5.89	5.15	0	1	1	2	25	20	17	
44	1121	1	47	1	172	3.59	3.66	4.88	4.49	0	0	0	1	*	-88	24	
46	1074	1	53	0	175	3,39	3,48	5,30	4.34	0	0	1	2	30	37	7	
47	1132	1	25	i	175	4.15	4.29	5.37	5.07	Ő	î	1	2	20	10	8	
48	1098	1	32	0	171	3.78	3.92	4.55	4.66	0	0	1	2	20	12	16	
49	1123	1	37	1	170	3.70	3.88	5.40	4.59	0	1	1	2	40	26	8	
	undru (-	45	1	163	3.04	3.19	4.18	3.85		1	U			-88	22	-

Next you will be asked for the statistics used to construct the boxplot. The usual choice is **median and quartile.** Select this then provide titles as requested for the graph and the groups as requested.

Box & whisker plotting options											
	<u>0</u> K	<u>H</u> elp	<u>C</u> ancel								
Title: Box & whisker plot from fou	ndry.xls										
Series labels: Non-exposed Exposed Non-exposed			=								
Outlier gate values (limits for plotting	g whiskers):										
Lower: Upper:	Leave these b whiskers to rep 0.438066 to 1	poxes blank if you present the range .332542.	ı want the e, which is								
Use fence values to define outliers for each data set individually.											
Type of plot: Median, quartiles and range Mean, standard deviation and r Mean, confidence interval and	range range										

The boxplot is displayed below.

FEV Ratio Exposed and Non-Exposed Workers



From this one can see that the groups have similar median and quartiles although the range is much larger for the exposed subjects. Descriptive statistics for each exposure group can be obtained using the **Descriptive** option in the **Analysis** menu. Before you do this make sure no variables are highlighted in order that you can select groups by an indicator variable.

Exercise: Use the **Descriptives** options to compare lung function of exposed with non-exposed workers using fvcratio and fevratio. Record the results below.

	Mean	Standard Deviation	Median	Max	Min	Ν
Exposed						
Non Exposed						

Comparison of Means Using a t-test

The t-test procedure can be used for statistical comparison of the mean FEV ratio of the exposed compared to non-exposed workers. It will also give the confidence interval for the difference of the two means. Under **Analysis** select **Parametric** then **Unpaired t**

d Stats	Direct - [fou	ndry.xls]													
TR File	Edit Insert F	format Data	Analysis Graphics	Tools V	/indow Help										- 6 >
found	try Repo	t1	Exact Tests on Co	ounts •											
D 🖨		× 🙉 🖌	Chi-Square Tests												-
			Proportions												
T1	FEVr	atio	Distributions												
	A	B C	Sample Size		F	G	Н	1	J	K	L	M	N	0	P _
1	IDNO	GROUP AG	Bandomization		EVHEAS	FEVPRED	FVCMEAS	FVCPRED	ASTHMA	BRON	SMKNOV	SHKEVER	CIGNO	CIGYRS	EMPYRS RES
2	1070	1 3	Miccellaneous		5.61	4.21	7.21	5.03	0	1	1	2	10	5	12
3	1124	1 3	Historichoods	22	4.51	3.45	5.08	4:02	0	0	1	2	16	-88	2
5	1006	1 4	Descriptive		4.02	2 72	F 82	4.54	ŏ	Ŭ	Ô	1	20	15	8
6	1086	1 2	Parametric		Paired t		22	5.59	0	0	0	0	*	-88	9
7	1018	1 5	Non-parametric	,	Single Sar	nple t	40	3.99	0	0	0	1	40	17	23
8	1133	1 3	Analysis of Varian	ce 🔹 🕨	Unpaired	ŧ	53	4.91	0	0	0	1	10	6	
10	1040	1 4	Regression & Con	relation 🕨	Summary	Data t	► 42	4 52	1	0	1	2	20	5	- 12
11	1099	1 3	Agreement	•	F (Variand	e Ratio)	16	4.44	Ô	ŏ	1	2	15	15	12
12	1135	1 4	Survival Analysis		 Single Sar 	nple z	01	4.29	0	0	1	2	30	25	19
13	1052	1 4	Meta-Analysis	•	Unpaired	z	26	5.54	0	0	0	0	*	-88	9
14	1029	1 5	Crosstabs		Reference	e Range	51	4.03	0	1	0	1	20	30	
16	1026	1 4	Frequencies		Poisson C	onfidence In	terval 17	4 45	0	0	0	0	*	-88	5
17	1095	1 3	Ontions		Shapiro-V	/lk	27	4.77	Ő	õ	0	0	*	-88	6
18	1030	1 3.	opdoris		4.68	4.22	5.92	5.06	1	0	0	0	*	-88	7
19	1087	1 3	9 1	165	3.84	3.47	4.96	4.15	0	0	0	1	20	8	17
20	1022	1 4		170	3.81	3.4/	5.13	4.24	0	U	0	1	20	-88	
22	1097	1 5	1	177	3 74	3 48	4 58	4.37	0	0	0	1	60	15	19
23	1039	1 3	1	170	3.98	3.72	5.11	4.46	Ō	Ō	Ū.	1	13	4	17
24	1065	1 3	7 0	175	4.19	3.95	5.05	4.76	0	0	1	2	20	15	7
25	1028	1 21	5 1	172	4.37	4.14	4.58	4.87	0	0	0	0	*	-88	8
26	1075	1 3.	1	190	4.58	4.34	5.39	5.20	0	1	1	2	30	20	
28	1031	1 3	1 1	190	4.91	4.68	6.06	5.69	Ő	Ô	1	2	12	13	9
29	1085	1 33	2 1	182	4.59	4.39	5.96	5.29	0	0	0	1	20	10	8
30	1090	1 2	9 0	168	4.05	3.88	4.69	4.56	0	0	1	2	15	4	9
31	1126	1 21	3 0	178	4.51	4.34	5.87	5.16	0	0	1	2	3	8	9
33	1078	1 5	1 1	173	3.30	3.02	4.07	4.51	0	0	0	1	20	-88	7
34	1059	1 51	1	170	3.46	3.36	4.40	4.13	Ő	ŏ	1	2	25	32	22
35	1124	1 4	5 0	175	3.80	3.71	4.75	4.55	0	0	0	1	10	11	21
36	1019	1 3	1 1	170	3.91	3.82	4.80	4.55	1	0	1	2	20	18	14
3/	1021	1 5	J U	165	4.04	3.99	5.38	4.99	0	1	0	1	40	32	15
39	1054	1 4	3 0	170	3.40	3 38	4 11	4.16	0	0	0	1	20	-00	16
40	1129	1 21	a Ö	175	4.22	4.21	4.96	4.99	Ő	ő	Ő	Ô	*	-88	9
41	1114	1 3	9 0	178	4.01	4.02	4.81	4.88	0	0	0	0	*	-88	12
42	1011	1 3	1	177	4.19	4.21	5.61	5.03	0	0	1	2	20	17	
43	1153	1 3	5 U 2 1	180	4.21	4.28	5.89	5.15	0	1	1	2	25	20	- 1/
45	1131	1 4	1	172	3.39	3.47	4.05	4.27	0	0	0	1	8	24	24
46	1074	1 5	3 0	175	3.39	3.48	5.30	4.34	0	ŏ	1	2	30	37	7
47	1132	1 2	5 1	175	4.15	4.29	5.37	5.07	0	1	1	2	20	10	8
48	1098	1 33	2 0	171	3.78	3.92	4.55	4.66	0	0	1	2	20	12	16
49	1123	1 4	5 1	163	3.70	3.88	4 18	4.59	0	1	1	2	40	-88	22 .
	foundry (200	5.04	0.10	4.10	0.00	1.1	1					

To compare mean FEV ratio for exposed and unexposed select GROUP as the identifier variable then FEVRATIO. This gives the following in the report window.

Unpaired t test

Mean of FEVRAT_GROUP_1 = 1.0003 (n = 73) Mean of FEVRAT_GROUP_0 = 1.0158 (n = 63)

$$\label{eq:asymptotic standard error} \begin{split} \underline{\text{Assuming equal variances}} \\ \hline \text{Combined standard error} = 0.0239 \\ \text{df} = 134 \\ t = 0.6467 \\ \hline \text{One sided P} = 0.2595 \\ \hline \text{Two sided P} = 0.519 \end{split}$$

95% confidence interval for difference between means = -0.0627 to 0.0318

Assuming unequal variances Combined standard error = 0.0236 df = 133.999 t(d) = 0.6536 One sided P = 0.2572 Two sided P = 0.5145

95% confidence interval for difference between means = -0.0622 to 0.0313

Comparison of variances Two sided F test is not significant No need to assume unequal variances

Two analyses are given. The first assumes variances are equal and the second are unequal. It also gives the results of statistical test comparing the variance. Unfortunately it just summarizes this as significant or non-significant without specifying the values of the magnitude of the variances, standard deviation or the significance level. This is an omission of the software. Above we saw that the standard deviation was 0.1479 for exposed and 0.1278 for non-exposed. As the F-test suggests that they are not significantly different, we take the unequal variance analysis as the t-test results although in this case it makes little difference. The difference of the two means is -0.0155. The result can be summarised as "there was no evidence of increased FEV ratio for workers exposed to dust (mean diff=-0.0155, 95% c.i - 0.0627 to 0.0318 p=0.519)"

Exercise: Compare mean FVC ratio for the exposed and non-exposed subjects using a t-test

REARRANGING DATA

From the analyses there appears to be no evidence that exposure to dust affects respiratory function. It may be argued nevertheless that being categorised as "exposed" or "not exposed" is a crude assessment for exposure. Dust exposure has been recorded for subjects in the exposed group. You will now carry out some analysis on just the exposed subjects, but to do this you need to select just those subjects.

Sorting Data

There are several ways you can do this. You could sort the data by group then delete the nonexposed subjects. Before you do this on your own data, make sure it is saved to disk in case anything goes wrong. Data in the spreadsheet can be sorted using the **Sort** option under **Data**. Check the help to find details of this procedure as this is a little complicated.

Introduction to StatsDirect, 15/03/2017

Highlight all the data involved excluding the column labels. Then select **Data** then **Sort** then **Manipulate Worksheet** the panel (right) will appear. Change the cell reference under Key Reference to select the column by which the data is to be sorted. To sort the data by exposure group insert put B1. If a nested sort is required, i.e. sort by exposure group then by sex within each exposure group, alter Key 1 to Key 2 and enter the appropriate Key reference. In this case enter G1.



Exercise:

- (1) Sort the rows of data by group.
- (2) By using two Keys sort the data first by group and then by age within group.

Group Split

Alternatively a new spreadsheet can be created in which there are separate columns for exposed and non-exposed workers for each variable. Select **Group split** on the **Data** menu. Then select the column that identifies the groups into which you want to split the data, in this case the variable **group**. It then asks us to choose the variables to which we want to apply **Group split**. For this exercise select all the columns RESPDUST and FEVRAT. If you need to select columns that are not adjacent you need to use the **Ctrl** key.

Cuic	insert r	ormac Data	a Analysis (Window He	lp .									- 1
	Cancel	Select d	ata to SPLIT	BY GROUP	(Min 1: Max 2	200)									
mode	e				To select n	on-adjoined (data, hold dov	vn Ctrl and use th	e mouse.						
	RESP	DUST													
	F	G	Н	1	J	ĸ	L	M	N	0	Р	Q	R	S	т
FE	VMEAS	FEVPRED	FVCHEAS	FVCPRED	ASTHMA	BRON	SMKNOV	SMKEVER	CIGNO	CIGYRS	EMPYRS	RESPONST	FEVRAT	HTmeter	FEVratio
	4 51	4.21	7.21 E 0.9	5.03	0	1	1	2	10	00	12	1.15	1.33	1.77	1.332541568
	4.51	3 72	5 31	4.02	0	0	1	2	15	-00	7	2 34	1 24	1.65	1 241935484
	4.60	3.73	5.82	4.54	ŏ	ŏ	ô	ĩ	20	15		1.96	1.23	1.74	1.233243968
	5.56	4.67	6.22	5.59	Ő	0	0	õ	*	-88	9	1.68	1.19	1.85	1.190578158
	3.77	3.24	4.40	3.99	0	0	0	1	40	17	23	2.04	1.16	1.68	1.163580247
	4.79	4.12	5.53	4.91	0	0	0	1	10	6	9	1.83	1.16	1.75	1.162621359
	5.25	4.53	6.74	5.57	0	0	1	2	3	9	12	1.45	1.16	1.9	1.158940397
	4.26	3.69	5.42	4.52	1	0	1	2	20	5	8	2.43	1.15	1.75	1.154471545
	4.27	3.70	5.16	4.44	0	0	1	2	15	15	12	. 5 5	1.15	1.7	1.154054054
	3.97	3.53	5.01	4.29	0	0	1	2	30	25	19	1.03	1.12	1.7	1.124645892
	5.05	4.50	6.26	5.54	U	U	0	0	*	-88	9	. 76	1.12	1.9	1.122222222
	3.63	3.24	4.51	4.03	U	1		1	20	30	10	1.50	1.12	1.7	1.12037037
	3.43	3.07	9.22	3.82	0	0	0	0		-88	10	1.19	1 12	1.00	1.11/203044
	4.01	4 01	5.17	4.45	0	0	0	0		-00	2	.72	1.12	1.75	1 110001040
	4.40	4.01	E 92	5.06	3	0	0	0	-	-00	7	1 72	1 11	1 70	1 100004720
	2.94	2 47	4 96	4 15	â	0	ň	1	20	00	17	2 72	1 11	1 65	1 106629243
	3,81	3 47	5 13	4.15	ő	0	ň	ñ		-88		29	1 10	1 7	1 097982709
	4.50	4.18	5.68	4.97	Ő	õ	Ő	1	20	8	9	1 18	1.08	1.75	1.076555024
	3 74	3 48	4 58	4 38	ň	ň	ň	1	60	15	19	86	1 07	1 77	1 074712644
	3.98	3.72	5.11	4.46	0	0	0	1	13	4	17	1.12	1.07	1.7	1.069892473
	4.19	3.95	5.05	4.76	0	0	1	2	20	15	7	. 40	1.06	1.75	1.060759494
	4.37	4.14	4.58	4.87	0	0	0	0	*	-88	8	1.88	1.06	1.72	1.055555556
	4.58	4.34	5.39	5.20	0	0	1	2	30	10	8	2.29	1.06	1.8	1.055299539
	4.81	4.56	5.90	5.59	0	1	1	2	25	20	7	1.66	1.05	1.9	1.054824561
	4.91	4.68	6.06	5.69	0	0	1	2	12	13	9	1.12	1.05	1.9	1.049145299
	4.59	4.39	5.96	5.29	0	0	0	1	20	10	8	1.51	1.05	1.82	1.045558087
	4.05	3.88	4.69	4.56	0	0	1	2	15	4	9	1.45	1.04	1.68	1.043814433
	4.51	4.34	5.87	5.15	0	U	1	2	3	8	9	1.27	1.04	1.78	1.039170507
	3.20	3.02	4.07	4.51	0	0	0	1	20	11	2	1 20	1.04	1.00	1.030047213
	2 44	2 26	4.40	4.05	0	0	1	2	26	-00		1 24	1 03	1 7	1 020761000
	3.80	3.71	4.40	4.55	0	0	ń	1	10	11	21	83	1 02	1 75	1 02425876
	3 91	3 82	4 80	4.55	1	0	1	2	20	18	14	46	1 02	1 7	1 023560209
	4.04	3.99	5,38	4.99	ô	ĩ	ñ	1	40	32	15	1.22	1.01	1.85	1.012531328
	3.08	3.05	3.68	3.73	õ	ô	Ő	Ō	*	-88	10	.97	1.01	1.63	1.009836066
	3.40	3.38	4.11	4.16	0	Ő	0	î	20	16	16	2.22	1.01	1.7	1.00591716
	4.22	4.21	4.96	4.99	0	0	0	0		-88	9	1.77	1.00	1.75	1.002375297
	4.01	4.02	4.81	4.88	0	0	0	0	*	-88	12	3.31	1.00	1.78	0.9975124378
	4.19	4.21	5.61	5.03	0	0	1	2	20	17	9	1.20	1.00	1.77	0.9952494062
	4.21	4.28	5.89	5.15	0	1	1	2	25	20	17	. 60	. 98	1.8	0.9836448598
	3.59	3.66	4.88	4.49	0	0	0	0	*	-88	6	1.95	. 98	1.75	0.9808743169
	3.39	3.47	4.05	4.27	0	0	0	1	8	24	24	1.90	. 98	1.72	0.976945245
	3.39	3.48	5.30	4.34	0	0	1	2	30	37	7	1.65	. 97	1.75	0.974137931
	4.15	4.29	5.37	5.07	0	1	1	2	20	10	8	1.39	. 97	1.75	0.9673659674
	3.78	3.92	4.55	4.66	0	0	1	2	20	12	16	. 90	. 96	1.71	0.9642857143
	3.70	3.88	5.40	4.59	U	1	1	2	40	26	8	. 29	. 95	1.7	0.9536082474
	3.04	3.19	4.18	3.86	U	U	U	U	.	-88	22	1.29	. 95	1.63	0.7529/80564

Group split then confirms the number of groups in

the panel below.



Click **OK** to give the panel to the right. To save confusion it is suggested that you put the split data into a new worksheet in the same work book. A new worksheet is now created as shown below

Select worksheet for output		
	Select	<u>C</u> ancel
Output destination:		
foundry.xls foundry		
New Sheet in foundry.xls New Book		

d Stats	Direct - [foundry.xls]										
I Eile	Edit Insert Format Dat	ta <u>A</u> nalysis <u>G</u> raphics	Iools Window Help								- 8 ×
found	dry Report 1										
0 🗳	8888	ν Σ 💡									
E2	3										
	Α	В	C D	E	F	G	н	1	J	К	L _
1	RESPDUST~GROUP=1	FEVRAT~GROUP=1	RESPDUST~GROUP=0 FEVRAT~GROUP=0								
2	1.15	1.332541568	0 1.23325062								
4	2.34	1.241935484	0 1.231428571								
5	1.96	1.233243968	0 1.211538462							-	
6	1.68	1.190578158	0 1.182897862								
1	2.04	1.16358024/	0 1.1809/44/8								
9	1.05	1 158940397	0 1.168341709								+
10	2.43	1.154471545	0 1.164983165								
11	0.55	1.154054054	0 1.163507109								
12	1.03	1.124645892	0 1.158501441								
13	U.76 1 C	1.1222222222	0 1.152631579								
15	1 19	1 117263844	0 1 130573248								
16	0.72	1.116991643	0 1.12345679						3		
17	0.91	1.112219451	0 1.111398964								
18	1.73	1.109004739	0 1.109725686								
20	2.72	1.105528242	0 1.109489051								
21	1.18	1.076555024	0 1.073979592								
22	0.86	1.074712644	0 1.06870229								
23	1.12	1.069892473	0 1.067092652								
24	0.4	1.060759494	0 1.064301552								
20	1.88	1.055555555	0 1.050884956								
27	1.66	1 054824561	0 1.030293030								
28	1.12	1.049145299	0 1.047318612						3		
29	1.51	1.045558087	0 1.044736842								
30	1.45	1.043814433	0 1.04								
31	1.27	1.0391/050/	0 1.028277635								
33	1.39	1.034482759	0 1.021226415								
34	1.74	1.029761905	0 1.017191977								
35	0.83	1.02425876	0 1.014336918								
36	0.46	1.023560209	0 0.9950617284						1		
3/	1.22	1.012531328	0 0.9896193772								
39	2 22	1 00591716	0 0.9836065574					-			
40	1.77	1.002375297	0 0.9785522788								
41	3.31	0.9975124378	0 0.9714285714								
42	1.2	0.9952494062	0 0.9702702703								
43	U.6 1 QC	0.9835448598	0 0.9690/21649							-	
45	1.95	0.976945245	0 0.9625360231								
46	1.65	0.974137931	0 0.9595687332								
47	1.39	0.9673659674	0 0.9510703364								
48	0.9	0.9642857143	0 0.9460154242					_			
49	0.29	0.9536082474	0 0 9451754386						-		-
41.0	Phoot1 & founday	0.0320/00304	0 0.7449760768			1		1	1		
	Sileer Anoundry /										100

Either the split work-sheet or the original can be selected by the tab at the bottom of the screen.

EXAMINING THE RELATIONSHIP BETWEEN TWO CONTINUOUS VARIABLES

Scatter Plots

The first step in examining the relationship between two continuous variables such as respiratory dust and lung function is to construct a scatter plot. This can be obtained from the graphics menu and is shown below.

Scatter plots have two axes:

the value of the dependent or response variable on the y axis

the value of the independent variable on the horizontal axis.

Since we expect respiratory dust to affect FEV rather than the other way round you should choose. FEVRATIO as the y-axis and RESPDUST as the x-axis. If you have not already created the variable FEVRATIO, return to on page 721 (Creating new variables) to create it. You will also need to sort the data so that you can select just the expose subjects (Group 1) as described on page 28.

To run the scatter plot click **Graphs** from the menu bar and **Scatter** from the drop down menu. Then a little window will appear like below

Catter Plot			×
	<u>0</u> K	<u>H</u> elp	Cancel
How many XY se	eries to plot	on graph? (N	lax=10)
0			

Type the number "how many xy series to plot on graph" and click **OK**. (In our case it will be 1) followed by selecting the Y and X axis data (FEVratio and RESPDUST respectively). Then you see another window with chart setting options. When you amend your options then click **OK**.

Chart settings	
	<u>OK</u> <u>H</u> elp <u>C</u> ancel
Title: Scatter plot from foundry.xls	
Y axis Label: FEVRAT Label position: C Top (horizontal text)	 Left (vertical text)
X axis Label: RESPDUST	
Axis data type: 💿 Numerical	C Dates
Use automatic scaling of axes	
🔲 Box in axes (lines top and right)	
🦳 Join markers with lines	

The scatter plot of FEV ratio compared to dust for subjects for the exposed group is displayed below.



There is some suggestion from this that respiratory function may be reduced for those with higher exposure.

Linear Regressions

To test this we will use linear regression to fit a straight line of the form Y=A + BX. Where Y is the dependent variable **fevratio** and X is independent variable **respdust**. If the gradient B is negative, this would indicate reduced respiratory function with increased dust. To do this in StatsDirect go to Regression & Correlation in the Analysis menu then select simple linear regression. And the following screen should appear.

StatsDire	ect - [foundry.xis]											
File Edit	Insert Format Data	Analysis Graphics Tools V	/indow Help									- 6
foundry	Report 1	Chi Emure Tests										
i 🖉 🖬	A B X B +	Dispositions 1										
		Proporcions										
E11	170	Rates										
	B C D	Distributions	G H I			К	1	M	N	0	P	0
	GROUP AGE	Sample Size	FEVPRED FUCHEAS FUCP	RED	ASTHMA	BRON	SMKNOW	SMKEVER	CIGNO	CIGYRS	EMPYRS	RESPDUST
2	1 31	Randomization	4.21 7.21 5	.03	0	1	1	2	10	5	12	1.15
3	1 32	Miscellaneous /	3.45 5.08 4	.02	0	0	0	0	*	-88	5	. 29
1	1 30	Descriptive	3.72 5.31 4	. 36	0	0	1	2	15	4	7	2.34
2	1 43	Parametric I	4 67 6 22 6	54	0	0		1	20	15	8	1.96
7	1 51	Non-parametric	3 24 4 40 3	99	0	0	0	1	40	17	23	2 04
8	1 31	Apabyris of Variance	4.12 5.53 4	.91	ŏ	Ő	ŏ	î	10	6	9	1.83
)	1 39	Democran & Conclusion	Simple Linear & Consulation	7	0	0	1	2	3	9	12	1.45
0	1 46	Agreement .	Multiple Linear	2	1	0	1	2	20	5	8	2.43
1	1 38	Agreement	Multiple Linear	4	0	0	1	2	15	15	12	. 55
2	1 44	Jurviva Andrysis	Bula de al Casa as asta		0	0	1	2	30	25	19	1.03
í	1 54	meta-witarysis	Principal Components	3	0	1	0	1	20	-00	9	1 50
5	1 56	Crosstabs	Linearized Estimates	2	ñ	ñ	ň	Â	*	-88	10	1 19
	1 46	Frequencies	Polynomial	5	0	0	Ő	0	*	-88	5	.72
7	1 32	Ontions	Logistic	7	0	0	0	0	*	-88	6	. 91
3	1 32		. Conditional Logistic	6	1	0	0	0	*	-88	7	1.73
	1 39	1 165 3.8	Poisson	5	0	0	0	1	20	8	17	2.72
	1 20	1 170 3.8	Probit Analysis	• 2	0	0		1	20	-88	9	1 10
	1 56	0 177 3 7	Kendal Rank Correlation	8	0	0	0	1	60	15	19	1.10
3	1 38	1 170 3.9	Spearman Rank Correlation	6	ŏ	Ő	ŏ	î	13	4	17	1.12
	1 37	0 175 4.1	Nonparametric Linear Regression	6	0	0	1	2	20	15	7	. 40
	1 26	1 172 4.3	Cox Regression	7	0	0	0	0	*	-88	8	1.8
	1 31	1 180 4.5	6		0	0	1	2	30	10	8	2.2
	1 38	1 190 4.8	1 4.56 5.90 5	.59	0	1	1	2	25	20	/	1.6
	1 34	1 192 4.2	4 4 30 D 00 5	29	0	0	1	2	20	10	2	1.1.
	1 29	0 168 4.0	5 3.88 4.69 4	56	0	0	1	2	15	4	9	1 4
	1 28	0 178 4.5	1 4.34 5.87 5	.16	Ő	0	ĩ	2	3	8	9	1.2
	1 31	0 168 3.9	6 3.82 4.87 4	.51	0	0	0	1	20	11	5	.7
	1 59	1 173 3.3	0 3.19 4.24 4	.09	0	0	0	0	*	-88	7	1.3
_	1 50	1 170 3.4	6 3.36 4.40 4	.13	0	0	1	2	25	32	22	1.7
	1 45	1 170 3.8	0 3.71 4.75 4	.55	1	0	1	1	20	11	21	.8
	1 54	0 100 4.0	1 3.02 4.00 4	. 55	1	1	1		20	22	14	1.2
	1 50	1 163 3.0	8 3.05 3.68 3	73	0	0	ň	Â	*	-88	10	9
	1 49	0 170 3.4	0 3,38 4,11 4	16	0	0	0	1	20	16	16	2.2
	1 28	0 175 4.2	2 4.21 4.96 4	. 99	0	0	0	0	*	-88	9	1.7
	1 39	0 178 4.0	1 4.02 4.81 4	. 88	0	0	0	0	*	-88	12	3.3
_	1 31	1 177 4.1	9 4.21 5.61 5	.03	0	0	1	2	20	17		1.2
	1 33	U 180 4.2	1 4.28 5.89 5	.15	0	1	1	2	25	20	17	.6
	1 49	1 172 2.2	2 3.55 4.88 4 9 3.47 4.05 4	27	0	0	0	1		-88	04	1.9
	1 53	0 175 3.3	9 3 48 5 30 4	34	0	0	1	2	30	37	24	1.5
	1 25	1 175 4 1	5 4.29 5.37 5	.07	Ŭ	1	1	2	20	10	8	1 3
	1 32	0 171 3.7	8 3.92 4.55 4	.66	õ	Ô	ĩ	2	20	12	16	. 90
9	1 37	1 170 3.7	0 3.88 5.40 4	. 59	0	1	1	2	40	26	8	. 29
	1 45	1 163 3.0	4 3.19 4.18 3	.86	0	0	. 0	0	*	-88	22	1.29

When requested select FEVRATIO as the Outcome or Y variable. Select RESPDUST as the Predictor or X variable. The following output should appear. This gives an equation for the line.

Regression and Correlation		
Simple linear regression Interpolate X to Y Interpolate Y to X Analysis Confidence intervals Plot regression Plot residuals	Interpolate X to Y Type value for X and press Enter to calculate	<u>Calculate</u> <u>H</u> elp <u>Cjose</u>
Simple linear regression		
Equation: FEVRAT~GROUP=1 = -	-0.036595 RESPDUST~GROUP=1 + 1.054884	
Standard Error of slope = 0.022004 95% Cl for population value of slope	e = -0.08047 to 0.007279	
Correlation coefficient (r) = -0.19364	t (r² = 0.037496)	
95% CI for r (Fisher's z transformed	I) = -0.405638 to 0.038126	
t with 71 DF = -1.663119 Two sided P = 0.1007 Power (for 5% significance) = 37.09	3%	
Correlation coefficient is not signific	antly different from zero	

It also gives the correlation coefficient r and statistical test of whether the gradient of the line is significantly different from zero. It gives a p-value of 0.1007 for the test that the gradient differs from zero. There is some suggestion of a negative gradient, but this is not significant at the conventional 5% significance level. Note that the output states "Correlation coefficient is not significantly different from zero". This assumes that one is using a two-tailed test with a 5% significance level.

The result for r^2 is also useful. This is an estimate of the proportion of the variance explained by the model. A line that fits the data perfectly will have an r^2 equal to 1. Where as a line that does not explain anything in the data will have an r^2 of zero. The value of r^2 equal to 0.037 is therefore not at all good – only 3.7% of the variation in the data is being explained by the regression line

Model Checking

The linear regression model described by the coefficients allows one to estimate a predicted value. The difference between the observer value and the predicted value is called a residual. Where a model fits badly the regression line will have large residuals. If you consider the scatter plot above for FEV ratio compared to respiratory dust the residuals will be large. It is possible to get a plot of the regression line with the data by selecting **Plot regression** shown below.



From the plot above two points appear to have large negative residuals suggesting that they are outliers. One of the assumptions of a regression model is that the residuals will have a normal distribution. In StatsDirect this may be carried out by plotting the residuals. Just select plot residual options.



This gives three plots, the first two are of the residuals against the fitted values and residuals against predicted values. From this it can be seen that there are two observations with large negative residuals. The second plot is a normal probability plot shown below. If the residuals are normally distributed the plotted points are in a straight line. Unlike SPSS StatsDirect does not show a line through the normal probability plot so that it is difficult to assess this plot. If one covers up the two points with large negative residuals in the plot below, it suggests that the data are approximately normally distributed. If the data were skewed the points would bulge away from the line.





Large negative residuals.

To check the robustness of the analysis it could be rerun excluding the two points with large residuals. The multiple regression procedure (rather than the simple regression) is required to calculate residuals and carry-out more complex statistical analyses.

Exercise: Examine the relationship between FVC ratio and dust levels using the methods above.

NON-PARAMETRIC METHODS

Where data is not normally distributed, statistical analyses that assume a normal distribution may be inappropriate. This is especially a concern where the sample size is small (<50 in total). Variables that are discrete (take only integer values) or have an upper or lower limit are by definition non-normal. Sometimes the distribution of the data is approximately normal so this is not a problem, particularly where the sample size is large, but for some variables it may be unreasonable to treat the data as normally distributed. To illustrate this you can compare the number of cigarettes smoked by "exposed" and "non-exposed" workers who currently smoke.

Either sort the data in the main work sheet so that only current smokers can be selected or use **Group split** to create a new work sheet. Return to the Foundry work sheet. In the Group split option under Data select SMKNOW as the variable to split the data then select the variables GROUP and CIGNO to create separate columns for current smokers and current non-smokers as shown below.

d Stat	sDirect - [foundry	r.xls]											5 🔀
Tile File	Edit Insert Forma	at Data Analysis Graph	iics Tools Window H	elp								-	a x
four	ndry Report1	m s Σ ?											
F	22												_
	A	В	С	D	E	F	G	н	1	J	к	L	-
2	GROUP SHKNOW	1 LIGNO"SHKNOW-1	GROUP"SMENOW=0	LIGNO"SMKNOV=U									-
3		1 19	3	L 20									
5		1 20	2	L 40									
7		1 30		10									
9		1 20)	L 20									
10		1 25	2										
12		1 19	5	20									
14		1 25											
16		1 20)	L 20									
1/		1 25	5	L 13						-			-
19 20		1 20) : 1	L 20 20									
21		1 4											
23		1 20))	L 10 L 40									
24		1 30)	L 20									-
26		1 20) :										
28		1 30)										
30		1 10	3	L 8									-
31		1 20) : 1	20									
33		1 20											
34		1 20)	L 30									
36		0 20	3	30									-
38		0 15	5	10									
40		0 10	5	10									
41 42		0 30	5 ()) ()	20									-
43		0 21) (18									
45		0 25											
46		0 15)									
48 49		0 15	5 (20									-
50		0 25	5 (Test 1						-
• • /	Sheet1 A Sheet2	A toundry /					•						•

The two left-hand columns are related to the current smokers. You may wish to replace the column headings to prevent confusion.

Exercise: Use **Descriptive** and **Frequencies** to generate summary statistics for the numbers of cigarettes smoked by current smokers for exposed and un-exposed workers separately.

From the **Descriptive** command you should obtain the following f tables

Variables	CIGNO (Group=1)	CIGNO (Group=0)					
Valid Data	34	20					
Mean	20.7059	18.75					
Variance	76.9412	33.8816					
SD	8.7716	5.8208					
SEM	1.5043	1.3016					
Lower 95% CL	17.6453	16.0258					
Upper 95% CL	23.7664	21.4742					
Geometric Mean	18.2073	17.5964					
Skewness	0.0916	-0.5262					
Kurtosis	3.1335	3.3146					
Maximum	40	30					
Upper Quartile	26.25	20					
Median	20	20					
Lower Quartile	15	15					
Minimum	3	5					
Range	37	25					
Variance coeff.	0.4236	0.3104					
Sum	704	375					
Centile 5	3	5.25					

Descriptive statistics

Examining this data one might wish to test if smokers exposed to dust were more likely to smoke more as the mean is larger and the upper quartile also (although the medians are the same). To test these use the Mann-Whitney U-test. Select Non-parametric from the Analysis menu then select Mann-Whitney. Choose GROUP as the group identifier and then select CIGNO. This should generate the following output.

Mann-Whitney U test

```
Observations (x) in CIGNO~SMKNOW=1_GROUP~SMKNOW=1_1 = 34 median = 20 rank sum = 976
Observations (y) in CIGNO~SMKNOW=1_GROUP~SMKNOW=1_0 = 20 median = 20
U = 381 U' = 299
```

Exact probability (adjusted for ties): Lower side P = 0.225 (H₁: x tends to be less than y) Upper side P = 0.775 (H₁: x tends to be greater than y) Two sided P = 0.4499 (H₁: x tends to be distributed differently to y)

```
95.1% confidence interval for difference between medians or means: K = 231 median difference = 0 CI = 0 to 5
```

In the tables above note the median for each group and the significance level. Hence, one can conclude that there is no difference between the median number of cigarettes smoked by "exposed" and "non-exposed" workers. The output also gives difference in the medians (0) and the confidence interval for the difference of the median (0 to 5). Note also that the coverage is given as 95.1% rather than 95%.

COMPARISONS OF RELATED OR PAIRED VARIABLES

For most of the analysis above you have compared the "exposed" and "non-exposed" groups of workers. In some circumstances one wants to compare measures within the same subject. Such comparisons are sometimes referred to as **paired** or **pair-matched** comparisons.

Continuous Outcome Measures

One might want to compare the mean of a continuous measure at one time point with the mean of the same measure at a different time point. Whilst this may not be a sensible analysis for this data, we can illustrate this for a continuous variable by comparing FEV measured with FVC measured.

To compare the mean measured FEV with mean measured FVC select a **Paired T** in the **Parametric** submenu. Then chose Group by Column and select the columns FEVMEAS and FEVPRED as shown below.

	meer [p	ouniary.xaj													ElE
e I	Edit Insert	Format Data	Analysis Grapt	tics Tools	Window Help										- 6
<u>IK</u>	<u>C</u> anc	el Select 2 m	atched colum	ns (or 1 of d	ifferences) (Min 1: Max 2	?) 								
aucrii	TTUE TTUE	CMEAC		by identifier	TO SERECT NOT	raujuneu uala	, noid down c	ui anu use me m	ouse.						
HI	A	B C	D	F	F	G	н		1	к	1	М	N	0	D
- 1	IDNO	GROUP AGE	SEX	HT	FEVNEAS	FEVERED	EVONEAS	EVCPRED	ASTHMA	BRON	SHKNOW	SHKEVER	EMPYRS	CIGNO	CLOVES R
	1001	1 49	1	175	3.40	3.59	4.49	4.45	0	0	1	2	23	20	31
	1002	1 46	5 1	168	2.83	3.39	3.91	4.12	1	1	1	2	16	20	11
	1003	1 34	1 1	180	3.93	4.26	4.80	5.14	0	0	0	0	12	*	*
-	1004	0 34	0	180	4.01	4.25	4.57	5.12	0	0	1	2	12	25	16
-	1005	0 23		183	4.75	4.52	6.50	5.42	0	U	0	1	6	*	1.5
	1000	1 9.	2 0	1/4	4.60	3.73 4 AC	3.02	4.54	0	0	0	1		20	15
	1009	1 50	1	167	2 58	2 97	3 68	3.73	ő	0	1	2	22	30	12
	1010	1 29	1	175	4.50	4.18	5.68	4.97	ŏ	ŏ	Ô	ĩ	9	20	8
	1011	1 31	l î	177	4.19	4.21	5.61	5.03	Ő	Ű	1	2	9	20	17
	1012	1 39	5 0	173	3.51	3.92	4.66	4.69	0	0	1	2	15	20	25
	1013	1 28	3 0	168	2.92	3.91	4.09	4.59	1	0	1	2	8	40	3
	1014	0 34	1 1	175	3.18	4,03	3.61	4.84	0	0	0	0	8	*	*
	1015	0 51	. 0	168	2.76	3.24	4.21	3.99	0	1	1	2	5	20	29
-	1015	0 45		175	3.06	3.59	4.65	4.45	0	U	0	1	8	20	3
-	1017	1 51	2 0	1/5	3.75	9.10	5.27	4.97	1	0	0	1	22	40	17
-	1019	1 3/	1	170	3 91	3 82	4 80	4.55	1	0	1	2	14	20	18
	1020	0 33	, î	183	4 03	4 44	5 14	5 35	ô	ň	Ô	ĩ		5	1
	1021	1 50	i ŏ	185	4.04	3.99	5.38	4.99	ŏ	1	Ő	ĩ	15	40	32
	1022	1 46	5 1	170	3.81	3.47	5.13	4.24	0	0	0	0	9	*	*
	1023	0 49	9 0	165	3.32	3.17	4.68	3.87	0	0	1	2	15	20	31
	1025	0 45	5 0	170	3.40	3.50	4.34	4.26	0	0	0	0	6	*	*
	1026	1 46	5 0	175	4.01	3.59	5.17	4.45	0	0	0	0	5	*	*
_	1027	1 56	1	165	2.80	2.97	3.57	3.69	0	0	0	1	7	20	38
_	1028	1 20	1	1/2	4.37	4.14	4.58	4.87	0	0	0		8	*	*
	1029	1 54	1	170	3.63	3.24	4.51	4.03	0	-	0	1	9	20	30
-	1021	1 32	1	190	4.00	9.22	6 04	5.00	1	0	1	2	á	12	12
	1032	0 50	1 1	170	2 47	3 36	3.88	4 13	ő	ĩ	Ô	ĩ	16	30	17
	1033	1 53	i Ö	163	2.16	2.94	3.60	3.61	ů.	ō	1	2	13	20	25
	1034	0 52	2 0	185	3.53	3.94	4.70	4.94	0	0	Ō	1	9	40	30
	1036	0 42	2 1	162	3.64	3.24	4.59	3.88	0	0	1	2	12	10	15
	1037	0 34	1 0	177	3.69	4.12	5.12	4.95	0	0	0	0	6		*
_	1038	0 45	5 0	170	4.31	3.50	5.50	4.26	0	0	1	2	9	20	12
_	1039	1 38	1	170	3.98	3.72	5.11	4.46	0	0	0	1	17	13	4
	1040	0 43		180	3.97	3.87	5.18	4.78	0	0	0	1	22	20	10
-	1042	0 24		120	4.80	4.51	6.2/	5.35	1	0	0	1	10	20	* 2
-	1043	1 51	1	175	2 72	3 33	4.04	4.01	1	0	0	1	25	20	12
	1045	1 48	1	175	3 39	3 63	4 17	4 47	0	ñ	0	, î	17	*	*
	1046	1 38	1	190	4 81	4.56	5.90	5.59	0	1	1	2	7	25	20
	1047	1 47	1	175	3.59	3.66	4.88	4.49	Ő	ô	Ô	õ	6	*	*
	1048	1 39) î	190	5.25	4.53	6.74	5.57	Ő	Ű	1	2	12	3	9
	1049	0 62	2 0	175	3.04	2.74	3.73	3.49	0	0	0	1	8	10	5
	1050	0 36	5 1	173	3.68	3.89	4:69	4.67	0	0	0	0	6	*	*
	1051	0 35	5 0	170	3.97	3.80	3.97	3.80	0	0	1	2	8	25	19
	1052	1 40	1	190	5.05	4.50	6.26	5.54	0	0	0	0	9	*	*
	1053	0 34	1 0	182	4.54	4.33	5.65	5.24	0	0	0	0	17		*

Results are given below

Paired t test

For differences between FEVMEAS and FVCMEAS: Mean of differences = -1.0196 (n = 136) Standard deviation = 0.373Standard error = 0.032

95% CI = -1.0829 to -0.9564

df = 135 t = -31.8789

One sided P < 0.0001 Two sided P < 0.0001 It is readily apparent that mean *measured FVC* is greater than mean *measured FEV*. You could report this as "Measured FVC was significantly higher than measured FEV (diff=1.02, 95% c.i. 0.96 to 1.08, p<0.0001)"

Exercise: Compare the mean FEV ratio with the mean FVC ratio.

The above method of analysis compares the mean value for the two variables. It does not tell one how close individual values are for the same subject. A visual way in which one can do this is with a scatter plot of the two variables as shown below. One gets a visual impression that FEV and FVC are quite strongly correlated. By choosing the same numerical range for both axes we can see also that the values for FVC are systematically larger than for FEV.



Analysis of Binary Outcomes that are Related

Suppose we wish to compare the proportion of workers who had bronchitis symptoms with the proportion who had asthma symptoms. Data are paired and the appropriate method statistical inference is McNemar's test. First one should construct the **Crosstabs** procedure in **Analysis menu**. Select ASTHMA as the **First Classifier Row** and BRON as the **Second Classifier row** to get a 2x2 cross-classification. We obtain the following screen from which we can see that only 3 workers had symptoms of both and 8 had symptoms of just Asthma or 12 just Bronchitis. From this table we can calculate that 11% (15/136) of workers reported bronchitis whilst only 8% (11/136) had asthma. These two proportions can be compared using McNemar's test, which takes account of the pairing of the variable, by selecting **Yes**.

🗗 Statsî)irect - [Report	1]			_ # X
🗩 File 🛛	Edit Insert Forma	at Analysis	Taols Window Help		L di X
foundry	y Report 1	٦			
0 🛩 I		n 🔁 🛍	Σ 💡		
Arial	<u>.</u>	10 💌	🍎 B I 🗓 ×2	* E = = :: A	
<u>X</u>	1 2 1 3 1 1	4 + 1 + 5 + 1	-6-1-7-1-8-1-9) + +10+++11+++12+++13+++14+++15+++16+++17+++18+++2	
Crossta	abs				
Row var Column	riable (first classi ı variable (second	ifier): d classifier):	ASTHMA BRON		
0	0 113 8	1 12 3			
	5	5			
				Select options for contingency table analysis	
				QK Help Cancel	
1				Investment test Investment test Show percentages Show cell chi-square	
				Show trend scores Specify trend scores	
1					
I					
I					
10/03/200	16:34				VS ICAPS NUM

By clicking **Analysis**, **Exact Test on Counts** then **Matched Pairs** the panel below will appear. In this select Matched pair's radio button then click **Calculate** to obtain the results for McNemar's test comparing the proportion of workers with Asthma symptoms as compared to Bronchitis symptoms.

d McNemar and exact	matched pairs test					
Exact Tests C Sign C Eicher's C Egpanded Ficher's C Dy Grieher's Matched pairs C Ogde ratio Cl C Ecisson rate Cl Digar Results	Control/reference cate Present 113 8 95 v Confidence	gory (press F1 for an ex Absent 12 3 e (%)	iample); Case/index category; Present Absent	<u>C</u> alculate	<u>H</u> elp	Cl <u>o</u> se
McNemar and exact	(Liddell) test for a	paired fourfold ta	ible			
Input table: 113 12 8 3						
Uncorrected Chi² = 0.8 Yates' continuity corre	8 (1 DF) P = 0.3711 ected Chi ² = 0.45 (1 [DF) P = 0.5023				
After Liddell (1983): Point estimate of relati	ive risk (R) = 1.5					
Exact 95% confidence	e interval = 0.563826	to 4.230397				
F = 1.333333 Two sided P = 0.5034						

The p-value for the McNemar test with Yates continuity correction is 0.5023. This is not significant at a 5% level so we conclude that symptoms of bronchitis are no more common in this population than symptoms of asthma.

Related Ordinal Data

For ordered categorical or quantitative variables that are not plausibly normal the suggested procedure is to use the **Wilcoxon** procedure.

Introduction to StatsDirect, 15/03/2017

SUMMARY STATISTIC METHODS

StatsDirect has a set of procedures that are based on summary statistics. These methods are based on statistics obtained from previous analysis. For example 0.11 or 11% (15/136) workers had symptoms of Bronchitis. A confidence interval for that proportion using StatsDirect follows. In the **Analysis** menu, choose **Proportions** then **Single** to obtain the following screen in to which the numerator and denominator of the proportion have been entered.

d Tests on proportion:	S				(_ 🗆 🛛
Proportion Tests Single Paired Iwo independent	15 136 .5 95 v Confider	Total number of observ Number responding Probability of success of cce (%)	ations In each trial	Calculate	Help	Cl <u>o</u> se
Clear Results						
Single proportion						
Total = 136, response Proportion = 0.110294	e = 15 1					
Exact (Clopper-Pears	on) 95% confidence	interval = 0.06306 t	0.175383			
Using null hypothesis Binomial one sided P Binomial two sided P	that the population P < 0.0001 P < 0.0001	proportion equals O.	5			
Approximate (Wilson)	95% mid-P confide	nce interval = 0.0679	988 to 0.174011			
Binomial one sided m Binomial two sided m	nid-P < 0.0001 hid-P < 0.0001					

The 95% confidence interval is 0.0631 to 0.1754. One might report this as "the proportion of workers who have experienced symptoms of Bronchitis was 11% (15/136, 95% c.i. 6% to 17%)"

t-test Using summary Data

Summary methods are particularly useful where we want to compare data from different studies where the raw data is not available. Suppose we wanted to compare the data in this study with that reported in say that from workers exposed to paper dust. Let us assume that the summary statistics reported for the FEV Ratio for workers in a study of the effects of paper dust are: mean=0.97, standard deviation=0.14, sample size n=115. From the **Descriptives** option the corresponding values for workers exposed to dust are: mean=1.0007, standard deviation=0.1479, sample size n=73). To carry out a t-test comparing mean FEV Ratio for the foundry workers with the paper workers on selects the option **Parametric** then **Summary Data t** then **Un-paired**.

E Stats	Direct - [found	ry.xls]												×
I File	Edit Insert For	nat Data	Analysis Graphics Tools	Window Help									- 6	i ×
found	try Report 1		Exact Tests on Counts	•										
D 🚅	H 4 h	X 🙉 🖌	Chi-Square Tests	1										_
		2020224	Proportions											_
R2	0.7572	3830734	Distributions	i l										
	F	G	Sample Size	, К	L	M	N	0	P	Q	R	S	Т	
1	FEVMEAS FI	EVPRED	Randomization	FHMA BRON	SMKNON	SMKEVER	EMPYRS	CIGNO	CIGYRS R	ESPDUST	FEVratio			
2	3.40	3.59	Miscellaneous	▶ <u>1</u> 1	1	2	23	20	31	1 71	0.7237851662			
4	3,93	4.26	120000000	0 0	ô	õ	12	*	*	1.69	0.81875			-
5	4.01	4.25	Descriptive		1	2	12	25	16	. 00	0.8774617068			
6	4.75	4.52	Perametric	Pareo C		0	7	*	*	.00	0.7307692308			-
8	4.60	3.73 A 45	Non-parametric	 Single sample c 		1	8	20	15	1.96	0.7903780069			-
9	2.58	2.97	Analysis or variance	Unpared t		Single Sample	22	30	12	. 69	0.7010869565			
10	4.50	4.18	Agreement	E (Variance Datio)		Unpaged	9	20	8	1.18	0.7922535211			
11	4.19	4.21	Survival Analysis	Single Sample 7	-	onpored	9	20	17	1.20	0.7468805704			+
12	3.51	3.92	Mata Analysis	 billige banple z billiopaired a 		2	15	20	25	2.04	0.7532188841			+
14	3.18	4.03	Crosstabe	Reference Range		ō	8	*	*	.00	0.8808864266			T
15	2.76	3.24	Frequencies	Roisson Confidence In	erval	2	5	20	29	. 00	0.6555819477			
16	3.06	3.59		Shaniro-Wilk	or ru	1	8	20	3	.00	0.6566523605			+
18	3.95	3 24	Options			1	23	*	17	2.04	0.7466918715			-
19	3.91	3.82	4.80 4.55	1 0	1	2	14	20	18	2.46	0.8145833333			1
20	4.03	4.44	5.14 5.35	0 0	0	1	8	5	1	. 00	0.7840466926			
21	4.04	3.99	5.38 4.99	0 1	0	1	15	40	32	1.22	0.750929368			-
22	3.81	3.47	4.68 3.87	0 0	1	2	15	20	31	. 29	0.7426900585			
24	3.40	3.50	4.34 4.26	0 0	Ô	Õ	6		*	.00	0.7834101382			
25	4.01	3.59	5.17 4.45	0 0	0	0	5	*	*	. 72	0.7756286267			
26	2.80	2.97	3.57 3.69	0 0	0	1	7	20	38	.74	0.7843137255			-
21	4.37	3 24	4.58 4.87	0 0	0	1	8	20	30	1.50	0.9541484/16			-
29	4.68	4.22	5.92 5.06	1 0	Ő	ô	2		*	1.73	0.7905405405			+
30	4.91	4.68	6.06 5.69	0 0	1	2	9	12	13	1.12	0.8102310231			
31	2.47	3.36	3.88 4.13	0 1	0	1	16	30	17	.00	0.6365979381			+
32	2.16	3 94	4 70 4 94	0 0	1	2	13	20	25	1.07	0.7510638298			-
34	3.64	3.24	4.59 3.88	0 0	1	2	12	10	15	.00	0.7930283224			+
35	3.69	4.12	5.12 4.95	0 0	0	0	6	*	*	. 00	0.720703125			
36	4.31	3.50	5.50 4.26	0 0	1	2	9	20	12	.00	0.7836363636			4
38	3.98	3.72	5.11 4.45	0 0	0	1	22	20	10	1.12	0.7788649706			-
39	4.80	4.51	6.27 5.36	0 0	ŏ	ô	8	*	*	.00	0.7655502392			+
40	4.16	4.00	4.84 4.81	1 0	0	1	15	20	6	. 00	0.8595041322			
41	2.72	3.33	3.90 4.11	0 0	0	1	25	10	12	2.27	0.6974358974			-
42	3.39	3.63	5.90 5.59	0 0	1	2	17	25	20	1 66	0.8129496403			
44	3.59	3.66	4.88 4.49	0 0	Ô	Ū.	6	*	*	.95	0.7356557377			+
45	5.25	4.53	6.74 5.57	0 0	1	2	12	3	9	1.45	0.7789317507			
46	3.04	2.74	3.73 3.49	0 0	0	1	8	10	5	.00	0.8150134048			+
47	3.68	3.89	4.57 4.57	0 0	1	0	6	25	19	.00	0.7845481875			+
49	5.05	4.50	6.26 5.54	0 0	Ó	0	9	*	*	. 76	0.8067092652			+
50	4.54	4.33	5.65 5.24	0 0	0	0	17	*	*	. 00	0.803539823			-
4 + 1	Sheet1 λ found	ry /						4					1	+

This gives the following panel into which the summary

statistics can be entered as shown.

Clicking on OK gives the output

Unpaired t test

Mean of * sample 1 from summary = 1.007 (n = 115) Mean of * sample 2 from summary = 0.97 (n = 73)

Assuming equal variances Combined standard error = 0.0214df = 186t = 1.7277One sided P = 0.0429Two sided P = 0.0857

95% confidence interval for difference between means = -0.0792 to 0.0052

 $\frac{Assuming unequal variances}{Combined standard error = 0.0217} df = 147.1343 t(d) = 1.7065 One sided P = 0.045 Two sided P = 0.09$

95% confidence interval for difference between means = -0.0798 to 0.0058

<u>Comparison of variances</u> Two sided F test is not significant No need to assume unequal variances

🗗 Student's	t Tests from Summary Data 🔀
	Sample 1 size: 115
	Sample 1 mean: 0.97
Sample 1	standard deviation: 0.14
	Sample 2 size: 73
	Sample 2 mean: 1.007
Sample 2	standard deviation: 0.1479
	,
<u>0</u> K	<u>C</u> ancel

As the standard deviations are very similar (0.14 and 0.1479) and this is confirmed by the F-test, one considers the results under Assuming Equal Variance. Whilst it is tempting to present the One-sided P because it is statistically significant at a 5% level rather than the Two-sided P, usual practice in Medical research is to report the latter unless strong arguments can be made for a One-sided test a priori. Based on these results one might report "the difference in mean FEV ratio for paper workers exposed to dust as compared to foundry workers exposed to dust was -0.04 (95% c.i. - 0.079 to 0.005, p=0.086).

Comparison of Proportions

Suppose now that for the paper workers it had been reported symptoms of Bronchitis was 22% (25/115). For dust exposed workers in the foundry the percentage is 15% (11/73). A z-test comparing these two proportions/percentages can be carried out using the summary data. Going to **Proportions** then **Two independent** on the **Analysis** menu one gets the following panel into which the data is entered.

Tests on proportions	S						
Proportion Tests				<u>C</u> alculate	<u>H</u> elp		Cl <u>o</u> se
C <u>P</u> aired	25	Total observations in SAMPLE 1					
	115	Number responding in SAMPLE 1					
	11	Total observations in SAMPLE 2					
	73	Number responding in SAMPLE 2					
95 💌 Confidence (%)							
Clear Results							
Two independent proportions Total 1 = 115, response 1 = 25 Proportion 1 = 0.217391 Total 2 = 73, response 2 = 11 Proportion 2 = 0.150685							
Proportion difference = 0.066706							
Approximate (Miettinen) 95% confidence interval = -0.052548 to 0.175028							
Exact two sided (mid) P = 0.2622							
Standard error of proportion difference = 0.058882 Standard normal deviate (z) = 1.132879 Approximate two sided P = 0.2573 Approximate one sided P = 0.1286							
I							

The difference in proportions is 0.0667 with confidence interval -0.0525 to 0.175. This might be reported as "The percentage of foundry workers with symptoms of Bronchitis was 15% (11/73) compared to paper workers in study X of 22% (25/115). This difference was not statistically significant at the 5% level (Diff.=7%, Exact two-sided p=0.26, 95% c.i. -5% to 18%). "In making comparison with published data one should of course considered whether the method of data collection were comparable. For example were symptoms of Bronchitis ascertained in the same way.

CHOOSING THE APPROPRIATE STATISTICAL PROCEDURE

In this tutorial we have illustrated some of the basic statistical procedures available in **StatsDirect**. These are summarised in the table below.

	Plausibly Continuous	Ordinal or Ordered	Binary and
	and Normal	Categorical	Unordered
			Categories
Comparison of	Box-plot	Box-plot or Cross-	Cross-tabulation
Independent Two	Independent groups t-	tabulation of ordered	Chi-squared test+
Groups	test	categories	Fisher's exact test+
		Mann-Whitney U-test	
Comparison of more	Analysis of variance*	Kruskal Wallis	Cross-tabulation
than Two groups	(ANOVA)	analysis of Variance [*]	Chi-squared test+
Comparison of two	Paired samples t-test	Wilcoxon Matched	McNemar's Test+
related outcomes		Pairs*	
Relationship between	Scatter plot	Spearman correlation	Phi coefficient
a dependent variable	Regression	or Kendall's	
and one or more	Pearson's correlation	correlation coefficient	
independent	coefficient		
variables			

* Not illustrated

+ Method that can be used with summary statistics

For a more comprehensive chart for selecting methods see

http://www.graphpad.com/www/book/choose.htm