MODULE 3

Gathering Evidence:
How to Investigate Crime Statistics

Module 3
Gathering Evidence:
How to Investigate Crime Statistics

In this module:

- You learn about devising measures for concepts
- You learn how to describe large sets of numbers using one or two numbers
- You learn how to make and interpret straightforward tables and graphs
- You learn how to use an undemanding data analysis program

Overview

The previous module introduced two approaches to crime trends, a 'moral panic' approach, which suggested public ignorance of a falling crime rate, and a more radical approach which suggested a class-based split to concern over crime.

   This module and the next will demonstrate how you can begin to investigate these questions yourself using computers.
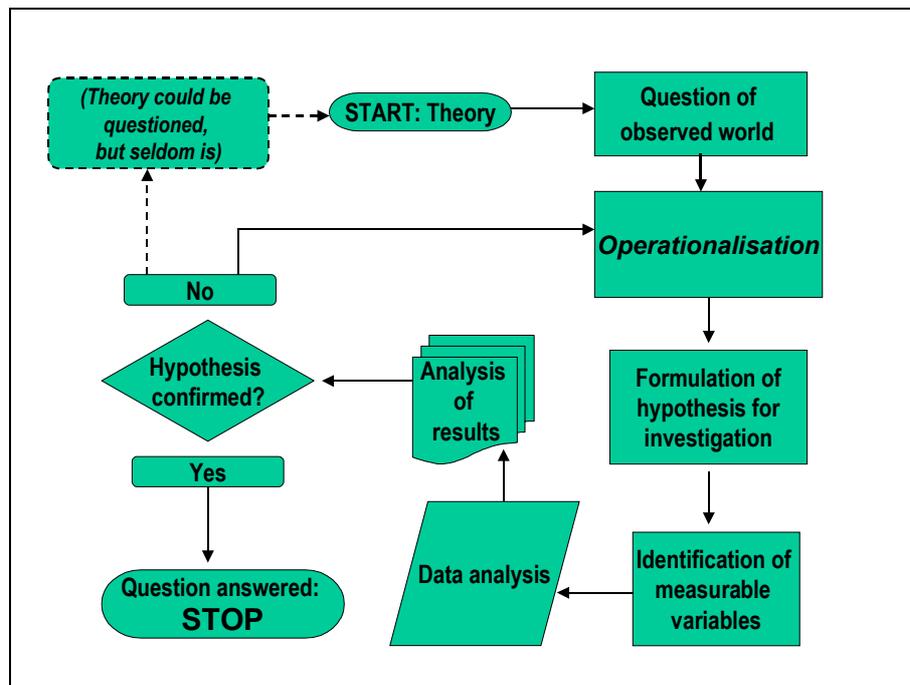

Figure 3.1 The process of operationalisation

Finding Questions

In Module 3 it was shown how data analysis is linked to theory. The first stage in analysis is linking this theory to data. This is known as *operationalisation*[1], the re-formulating of a theoretical question into an operational question that can be investigated and, for numerical data analysis, measured.

Figure. 3.1 illustrates the way questions are formulated and answered. Theory is always an input to question formulation. There are several steps between observing real-world issues and analysing data; the process of operationalisation turns a theoretical issue into a hypothesis which, if confirmed will answer the question, and possibly explain the issue.

In the previous module, for example, it was put forward that the fear of crime was affected by class. Class was then operationalised as being measured by income, and the hypothesis could then be formulated that fear of crime was affected by income, and that one would vary as the other varied. In fact, many commentators would argue that income is no longer a good measure of class.

In the British Crime Survey report[2], which was extensively cited in the previous module, the theoretical question was the concern over crime. The authors took answers to a series of questions of particular aspects of crime as measuring fear. Sometimes operationalisation is less straightforward than this. Consider for example newspaper readership. This may seem a simple notion but it is illuminating to investigate how this is operationalised in the British Crime Survey.


Web exercise
Tabloid paper readership

This exercise is designed to introduce some of the online material available, and also to consider the validity of some of the variables used in the British Crime Survey report.
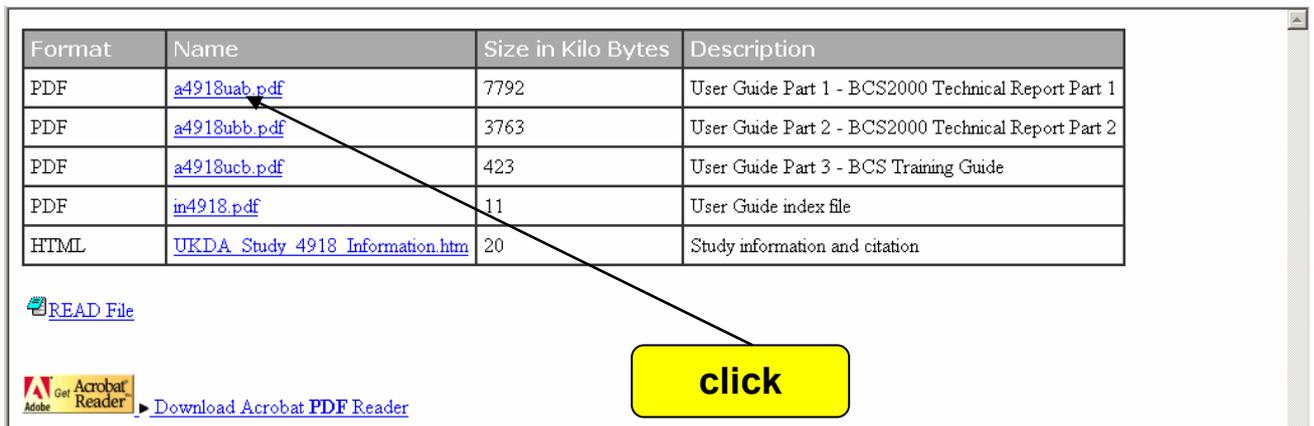
The exercise is to find the actual question that the report's author takes as a valid indicator of newspaper readership.

To do this you need to look in the *metadata* of the British Crime Survey. Metadata are data about data: it shows documentation about the survey. The British Crime Survey metadata can be found on the Nesstar web site (nesstar.esds.ac.uk/webview/)[3].

The actual questionnaire can be found in the *Technical Report* section of the *User Guide*. Click on the **Study Description** folder to get this. Several links will drop down in front of you, and you should click on the bottom link called **Other Study Description Materials**. The right-hand pane will then show a link to the **User Guide**. Click on this.
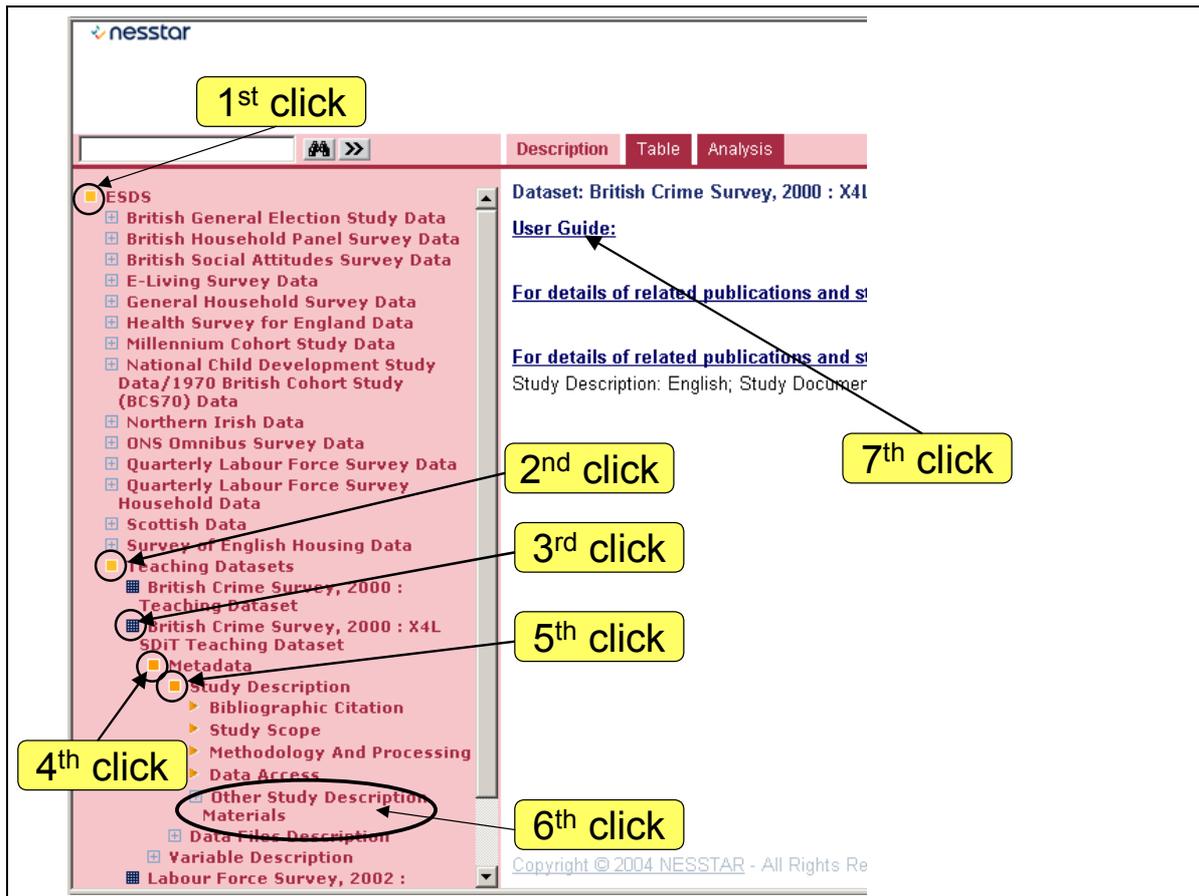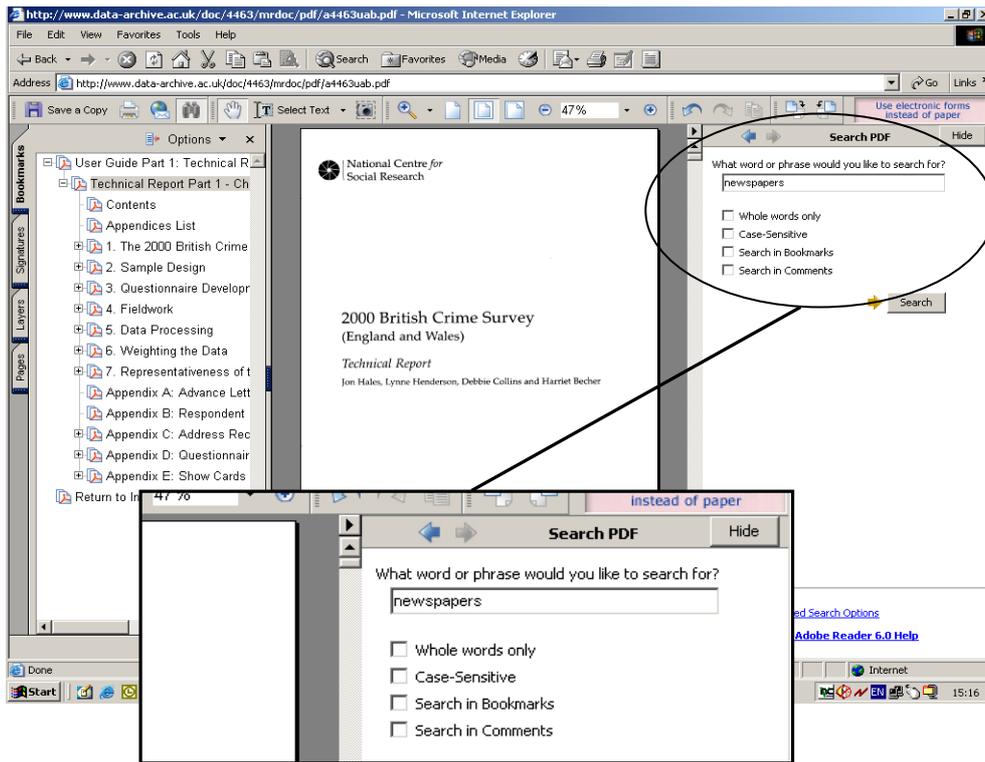
Then a new window pops up. Click on the top link to get the technical report with the questionnaire in it.

Format	Name	Size in Kilo Bytes	Description
PDF	a4918uab.pdf	7792	User Guide Part 1 - BCS2000 Technical Report Part 1
PDF	a4918ubb.pdf	3763	User Guide Part 2 - BCS2000 Technical Report Part 2
PDF	a4918ucb.pdf	423	User Guide Part 3 - BCS Training Guide
PDF	in4918.pdf	11	User Guide index file
HTML	UKDA_Study_4918_Information.htm	20	Study information and citation

READ File

Get Acrobat Reader ► Download Acrobat PDF Reader
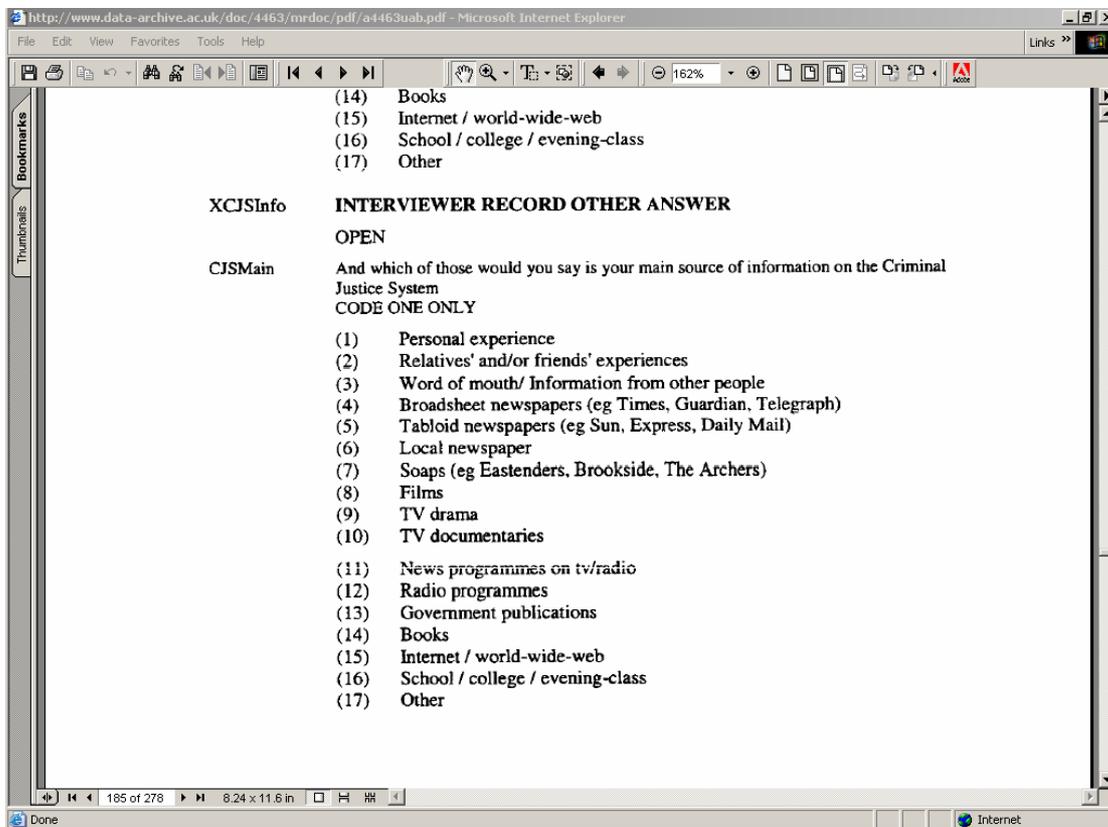
click

(You'll need the free Adobe Acrobat reader to read the file, but there's a link to download this onto your computer if you don't have it).

The technical report is very long and browsing through it is not very productive. Instead, use the search facility, which is started by clicking on the 'binoculars' button. Type in a suitable search term (maybe 'newspaper' or some such) and click on the ☐ Search ☐ button (note on older versions it might say ☐ Find ☐).The computer will find where that term is mentioned. One of these will be the actual question that was asked. Note that the word 'newspaper' is in the *answer*, not in the wording of the question.

Search PDF   Hide

What word or phrase would you like to search for?

newspapers

☐ Whole words only
☐ Case-Sensitive
☐ Search in Bookmarks
☐ Search in Comments

Search

Search PDF   Hide

What word or phrase would you like to search for?

newspapers

☐ Whole words only
☐ Case-Sensitive
☐ Search in Bookmarks
☐ Search in Comments

So what is the question that tells us which newspaper the survey respondents read?

(14)	Books
(15)	Internet / world-wide-web
(16)	School / college / evening-class
(17)	Other

XCJSInfo   **INTERVIEWER RECORD OTHER ANSWER**

OPEN

CJSMain   And which of those would you say is your main source of information on the Criminal Justice System
CODE ONE ONLY

(1)	Personal experience
(2)	Relatives' and/or friends' experiences
(3)	Word of mouth/ Information from other people
(4)	Broadsheet newspapers (eg Times, Guardian, Telegraph)
(5)	Tabloid newspapers (eg Sun, Express, Daily Mail)
(6)	Local newspaper
(7)	Soaps (eg Eastenders, Brookside, The Archers)
(8)	Films
(9)	TV drama
(10)	TV documentaries
(11)	News programmes on tv/radio
(12)	Radio programmes
(13)	Government publications
(14)	Books
(15)	Internet / world-wide-web
(16)	School / college / evening-class
(17)	Other

There are two questions that mention tabloid and broadsheet newspapers. We believe that the second of these (called CJS main) is the one that the report authors are drawing from when they compare newspaper readership with fear of crime. It's on page 7 of section 5, which is on page 185 of the report.

(?)	**What do you think?**
	Does this question measure newspaper readership well?

Examining the British Crime Survey dataset

The extent to which a measurement is adequate is called the *validity* of the measurement. The web exercise enabled you to obtain the actual question used in the British Crime Survey, which raised issues about the validity of the data as a measure of newspaper readership and therefore questions the Home Office analysis of press influence.

Quantitative studies usually involve identifying valid *variables*. A variable is simply something which can change. Each time a survey is filled in it is called a *case*. So the value of a variable can vary from case to case. The teaching database which this module uses (a part of the British Crime Survey) has 19,411 cases of 33 variables.

This means that each variable in this version of the BCS will be made up of a set of 19,411 numbers. The first step in examining data is therefore to find some way of summarising such a large set of numbers. This section will show you how to summarise a large quantity of numbers using only a couple of numbers, and to use graphs and tables to communicate data. It will also show how to use an easy computer program to achieve this.

While this can be done online using Nesstar, this module will show you how to use another program called NSDstat. This is because while Nesstar is a good program, its data manipulation facilities are more limited, which in later modules will become a problem. To find how Nesstar can be used to examine the British Crime Survey see Module 6.

NSDstat is a special data analysis program meant for teaching and learning, and is very simple to use. Starting up the program should also open up the special teaching version of the British Crime Survey that has been supplied. The list of variables should be visible in a separate window, otherwise click on the variable list button as shown, and the window will pop up on the screen.

To install NSDstat click here

The windows work in the same way as standard windows in any program[4]. To move between the different windows, simply click the mouse in the window. As can be seen, there are 33 variables in the dataset.

Summarizing variables

The first thing to do when investigating a dataset is to *describe* the relevant variables. In fact sometimes that is all that is required. However in many studies there are hundreds or even thousands of cases to describe. There are two ways to describe a large quantity of numbers, you can use a table or graph, or you can find a representative number.
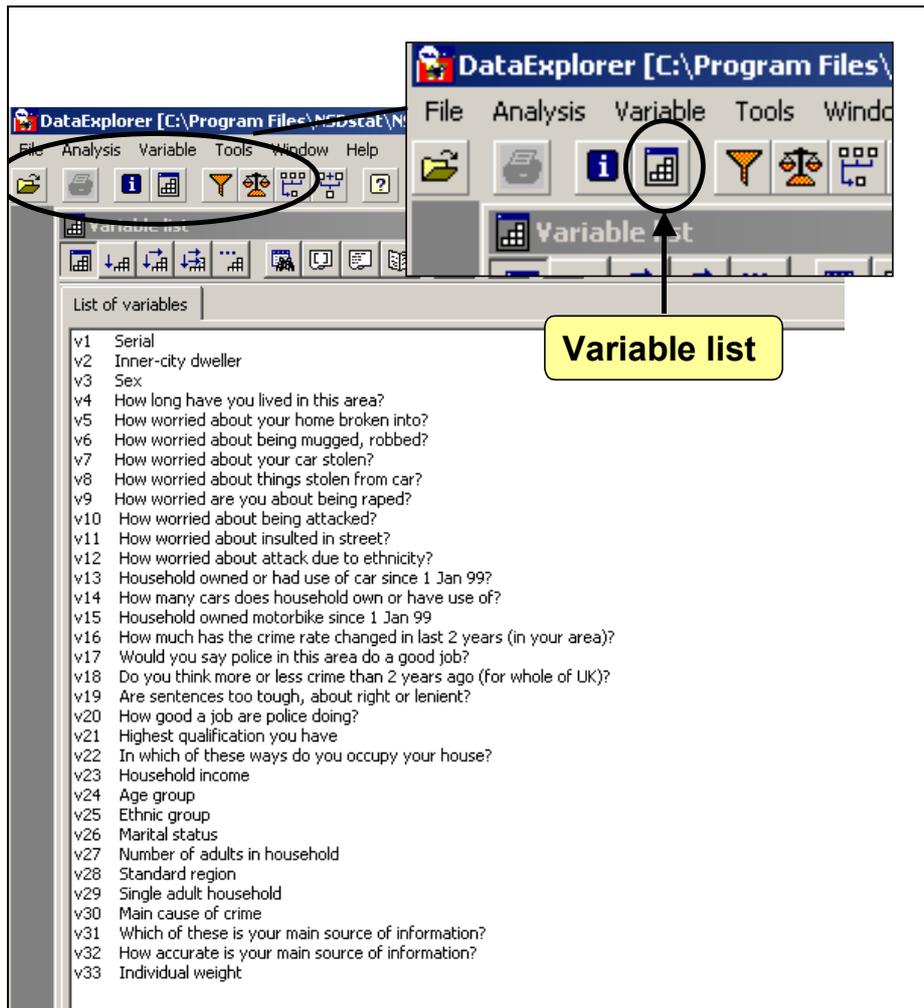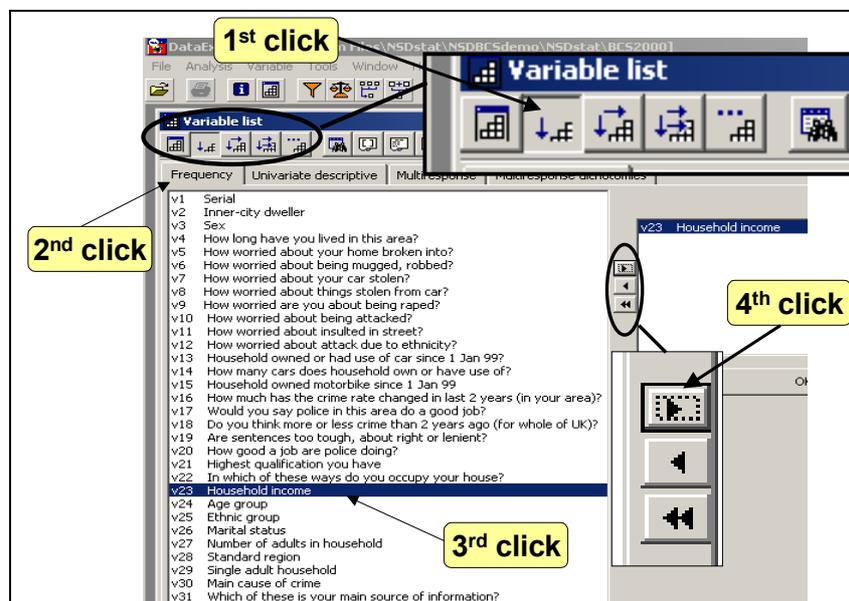
British Crime Survey 2000

Frequency Table

A frequency table is a common way of presenting data: the cases are put into a small number of groups and the data presented in a small table, such as Table 3.1. Table 3.1 is from the dataset supplied with this module. It has a title which enables you to refer to it in the text of your report (*always* refer in your text to any table you insert. If you ignore your own tables so will your readers). The title also tells the reader what the table is about. If the title does not mention the source, this is put elsewhere (often at the bottom of the table). The date of the survey and the geographic area it covers should also be in the table.

Table 3.1: Household Income, British Crime Survey, 2002		
Income	Number	%
Under £5,000	2157	12
£5,000-£14,999	5804	32.3
£15,000-£29,999	5747	32
£30,000-£49,999	2874	16
£50,000 or more	1383	7.7
TOTAL	19411	100

The table shows the groupings that the data have been put into, and the quantities or frequencies of each group. All of the data in the teaching database have already been put into categories. To generate a frequency table all that needs to be done is press the **Univariate** button, which has one vertical arrow (incidentally, all that the word 'univariate' means is one variable. In the next module we'll look at two variable – bivariate – analysis, which is the button next to it with two arrows).

British Crime Survey 2000

To make the table, click next on the **Frequency** tab, and on the particular variable of interest, which in this example is **v23 Household income**. Select this variable by clicking the right pointer key, which puts this variable in the box. (You can remove any variables already in the box by selecting them in the right hand frequency box and then hitting the left pointer key.) Click on the   **OK**   button to get the table.

The table produced by NSDstat shows the numerals attached to each category. These are called *codes*, and the process of dividing data into groups and attaching numerals is


v23 Household income

Value labels	Code	Number	% of all	% of valid
Under £5,000	1	2157	11.1	12.0
£5,000–£14,999	2	5804	29.9	32.3
£15,000–£29,999	3	5747	29.6	32.0
£30,000–£49,999	4	2874	14.8	16.0
£50,000 or more	5	1383	7.1	7.7
–	–	1446	7.4	–
Sum		19411	100.0	100.0

called *coding.* In the next module we'll find out how to change these categories, which is called *re*-coding. The table shows the frequencies and also the percentages. It also gives the number of 'invalid' responses, sometimes called *missing values*. These are where the respondent cannot or refuses to answer, or there is some other reason for not recording a valid reply to the question. These are on the bottom line of the table. (Note also how the variable list window has been moved out of the way).

investigating…

What happens if you put two variables in the box?
What do the left-hand buttons do?
Which menu choices do the same as the buttons?
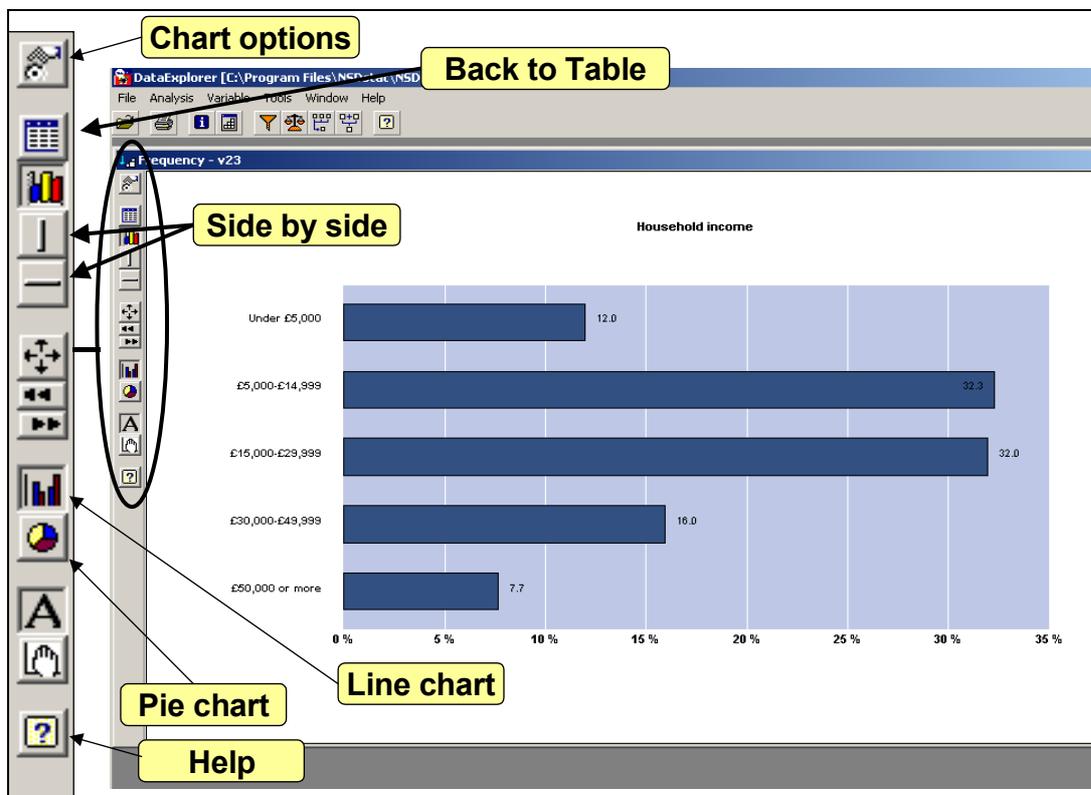
Note the buttons on the left hand side of the window. You can change the appearance of the table by clicking on the top button. You can also get a bar chart by clicking on the chart button. The help button will tell you about the other facilities, such as exporting your results and graphs into other programs, and printing them out.

To change the chart to vertical, click the chart options button and change the orientation to vertical.

Exercise

Produce a frequency table and graph to illustrate public views of police effectiveness [HINT- you first need to decide what variable measures this]
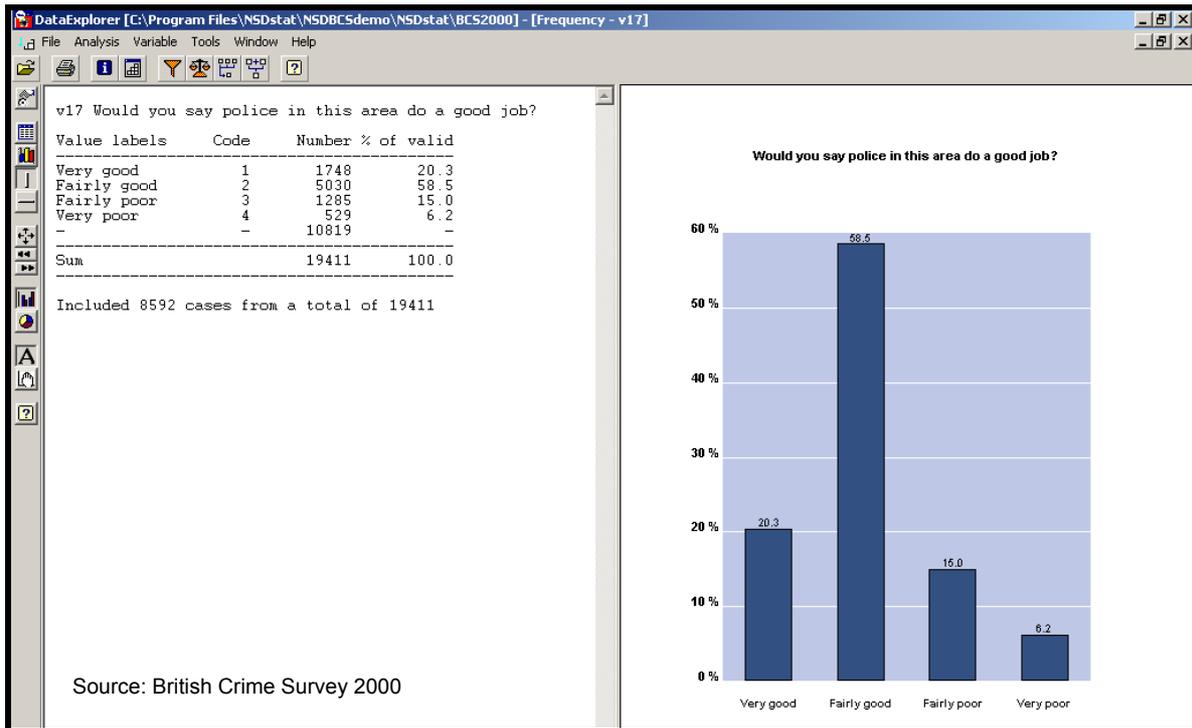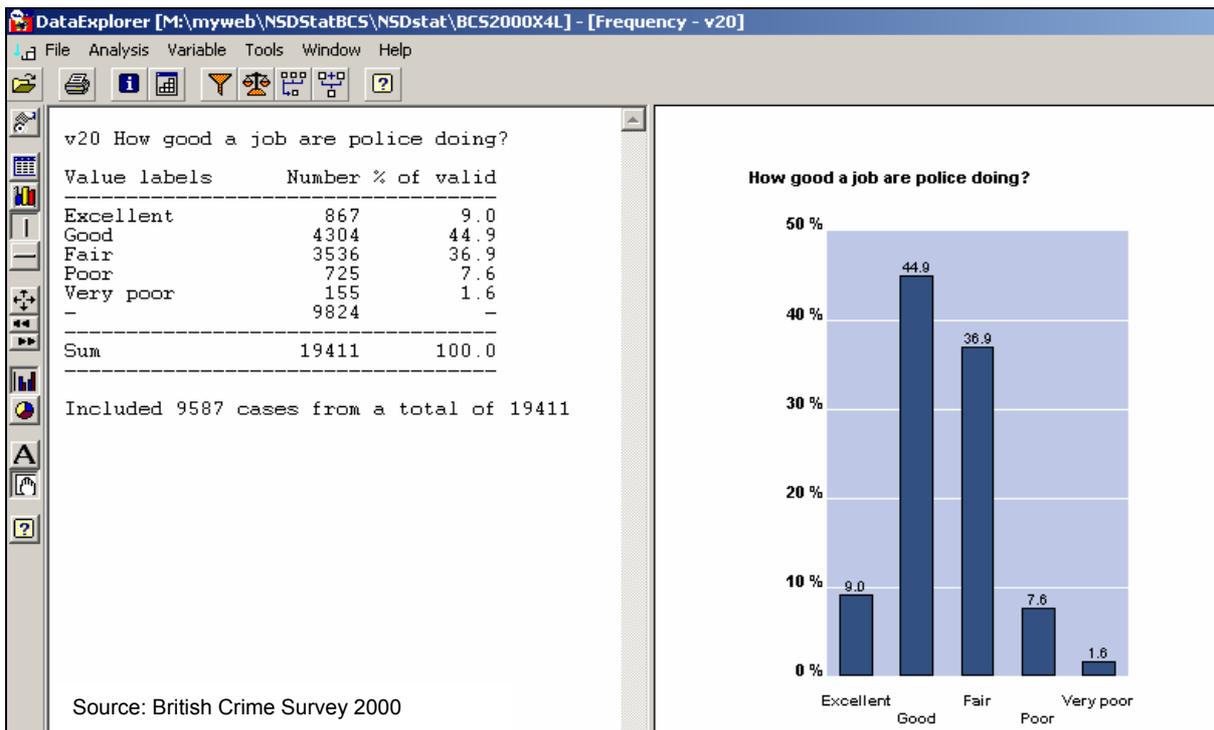
Fig. 3.2a: Views on local policing

Figure. 3.2b: Views on national policing

There are two measures of views on police effectiveness; a measure of local policing (variable v17) and national policing (v20). The frequency tables and graphs are reproduced as Figure.3.2a and Figure. 3.2b.

Representative numbers

Frequency tables are a useful method of summarising a large set of numbers, but it is also useful to get a single number to describe a set of numbers.
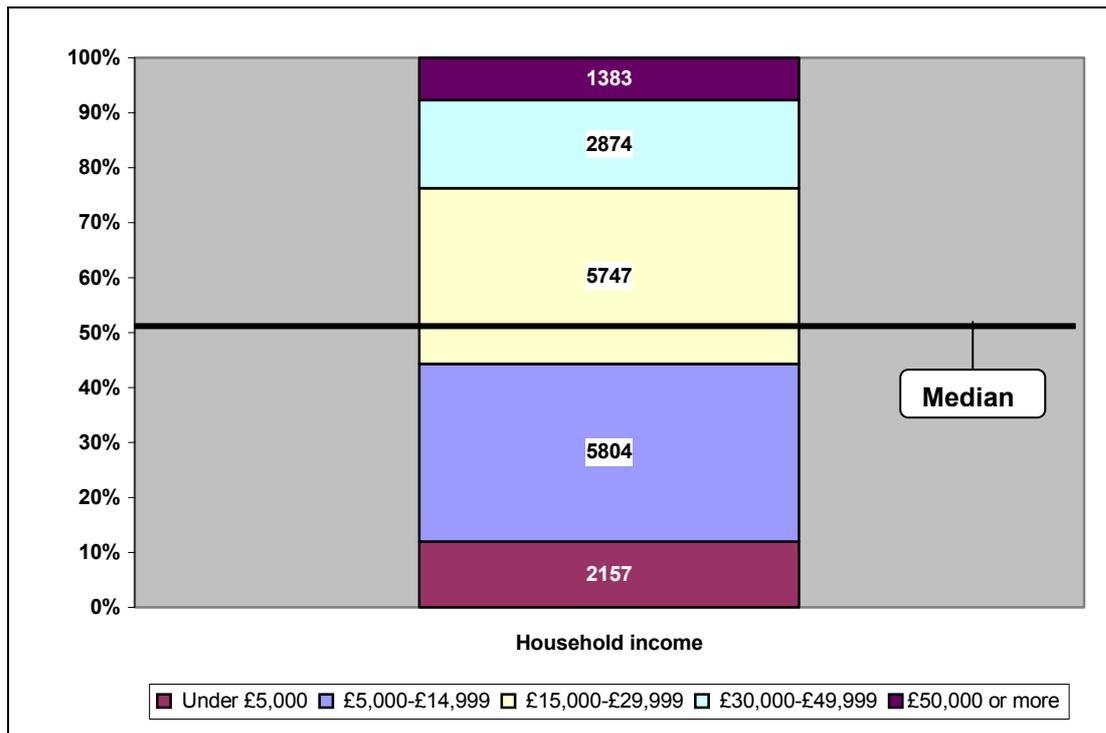
Figure 3.3 Household income, BCS 2000

Most people are familiar with the idea of the 'average', actually known as the *mean*. This is the aggregate of a set of numbers divided by the number of cases. However the mean is difficult or impossible to calculate from grouped data . An alternative is to use the *median*. If the data are ranked[5] in order from highest to lowest, the median would be the middle number. Half of the data are above the median. Mind you, half are below the median also. Fig. 3.3 shows a stacked graph[6] of the data in Table 1. The median is the 50% mark.

So to describe the income of the households in the survey, you could say that half the households earn more than £15,000 a year.

To get NSDstat to give you the percentages of the stacked graph in a frequency table, click on the table options button, and ask for accumulated valid under accumulated percentages. The percentages are then totted up in the right hand column (note that NSDstat does have a check box for the median. If you use this on grouped data, it will give you the code number of the category in which the median occurs).

Source: British Crime Survey 2000

From table 3.4 it can be seen that just over half the respondents were worried about burglary. The 50% mark in the right hand column would be reached just inside category 2, which represents 'fairly worried' respondents.

Table 3.4 Fear of burglary British Crime Survey 2000

Limitations of the median

So now you can present a large amount of data in a frequency table and describe them using the median. However a few notes of caution are in order. For example, try this trick question:
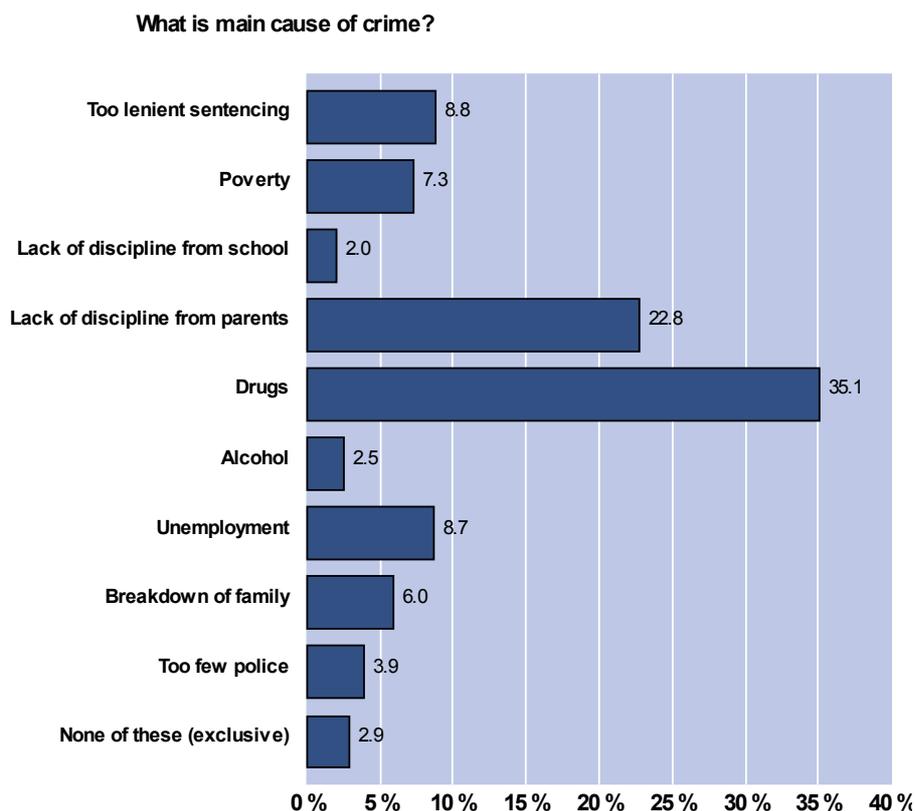
Trick Question

What is the median of variable 30 in the database?

This is a trick question because this variable is simply a category variable, sometimes called a *nominal* variable. It doesn't have any ranking or order at all. It doesn't have a median. To provide a representative number all you can do is present the most frequent category, which is called the *mode*. You can think of this as the 'typical' case.

    Figure. 3.5 shows a bar chart of the beliefs of the main causes of crime (although the chart is placed horizontally instead of vertically). The typical response in the survey is that the main cause of crime is drugs[7].

What is main cause of crime?

Category	Value
Too lenient sentencing	8.8
Poverty	7.3
Lack of discipline from school	2.0
Lack of discipline from parents	22.8
Drugs	35.1
Alcohol	2.5
Unemployment	8.7
Breakdown of family	6.0
Too few police	3.9
None of these (exclusive)	2.9

0 %   5 %   10 %   15 %   20 %   25 %   30 %   35 %   40 %

**Figure 3.5 Beliefs of main causes of crime,
British Crime Survey 2000**

So you have to work out if the categories can be ranked in order (highest to lowest or vice versa) or not. If they can, use the median, if not, the mode.

Exercise

Using NSDstat on the data in the 2000 British Crime Survey, pick your own topic and describe the data

---

You need to
- decide which variable measures the feature you're studying;
- generate a frequency table;
- decide if the variable can be ranked highest to lowest. if it can, describe the median, otherwise describe the two highest percentages;
- if you reported the median, report the middle half of the data also.

SUMMARY

- Data can be summarised by using a frequency table or graph
- They can also be communicated by using a summary number. Which summary number is best depends on the type of data being summarised.

NEXT UP….

Now you know how to use the computer to analyse variables and how to describe them. The next stage is to find out how to investigate the type of associations concerning the fear of crime you came up with in the exercises in Module 3.

---

[1] 'The process of operationalisation is specifying the procedure that will be employed to measure that concept… Operationalisation always encompasses measurement… the outcome of operationally defining a concept will be indicators that represent that concept' Carlson and Hyde, *Doing Empirical Political Research*, Milton Keynes:Open UP 2003 p.145.

[2] Simmonds J and Dodd T (eds) *Crime in England and Wales2002/2003*, Home Office Statistical Bulletin 07/03, London: HMSO, 2003 www.homeoffice.gov.uk/rds/crimeew0203.html

[3] Module 6 is a guide to using Nesstar

[4] To close any of the windows, click on the cross in the top right hand corner. To temporarily clear a window from the screen, click the horizontal bar in the corner. To move a window around on the screen, click and drag the blue bar running along the top.

[5] Data that can be ranked from highest to lowest are called *ordinal* data.

[6] A stacked graph is like a bar chart but with the frequencies stacked one on top of the other

[7] Note also that over half the respondents thought that either drugs or lack of parental discipline were the main causes of crime. These two categories formed over half of all responses.

References

British Crime Survey (2000): Home Office Research, Development and Statistics Directorate, National Centre for Social Research, *British Crime Survey 2000* [computer file], Colchester Essex,: UK Data Archive [distributor]. 16 January 2002, SN: 4463