MODULE 4

Examining Evidence:
How to Interrogate Crime Statistics
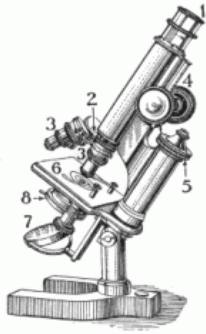
Survey Data in Teaching
enhancing critical thinking and
data numeracy

Version 1.0

Module 4
Examining Evidence:
How to Interrogate Crime Statistics

In this module:

- You learn how to alter data
- You learn how to examine associations between two variables
- You learn how to present and analyse data in tables

Overview

In the previous module you learned how to use computer programs to find out about the variables which measured concepts, and how to obtain summary numbers which described these variables. Before then, in Module 2, two conflicting models concerning the public perception of crime were presented. One was described as a 'moral panic' approach, whereby it was held that the media helped to amplify the portrayal of crime levels, which led to the public believing the level of crime to be much higher than it actually was. In contrast, a class-based model was also described, in which it was suggested that there was an inverse relationship[1] between the risk of crime and the fear of crime. The proposed explanation for this was that the effect of crime on poorer victims was higher, and therefore for these people it was more a cause for concern.

In this module you will find out how the BCS can be used to investigate ideas such as these yourself. You will see how to find out if variables are linked, and what effect one variable has on another.

Linking variables

Let's start by considering whether one variable is affected by another. This is sometimes called *bivariate analysis* in the textbooks, which just means analysing two variables (bi as in bicycle, which has two wheels). Examining more than two variables simultaneously is called *multivariate* analysis.

Bivariate analysis shows if two variables change in tandem; that is if one will alter as the other changes. The variable that alters first is called the independent variable, the variable that might follow is called the dependent variable. So while fear of crime may be affected by, say, gender, gender is unaffected by the fear of crime. Gender is therefore *independent* of the fear of crime.

Note that this is not the same as saying that gender *causes* the fear of crime to change. All that is being said is that fear of crime may be linked to gender. There may be many causes of the fear of crime, and establishing all of the possible links between gender and crime fear may be difficult.

This idea of changes in one variable being linked to changes in another is called *correlation*, the variables are said to be correlated if changes in the independent variable lead to changes in the dependent variable. To investigate correlations, a two-way or cross-

tabulated[2] table is used, such as Table 4.1

Table 4.1 Household owned or had use of car since 1 Jan 1999 by concern over car theft			
		British Crime Survey 2000	
	Car use		
Worried about theft	**No**	**Yes**	**TOTAL**
not worried	153	6289	6442
worried	173	7845	8018
TOTAL	326	14134	14460
Included 14460 cases from a total of 19411			

This type of table shows what happens to the dependent variable (Worried about theft) in Table 4.1 if changes in the independent variable (car use in Table 4.1) occur. That is, the table shows car ownership *by* concern over car theft. In the table the variable 'Worried about car theft' is split up by the categories in the variable 'car ownership/ use' to enable comparison of the two variables.

? A puzzle

Why are 173 respondents who do not have a car worried about their car being stolen?

In the next section you'll learn how to construct a two-way table like table 4.1 from a database using NSDstat. After this, you'll analyse tables like these and draw conclusions from them.

Constructing a table (i): Simplifying data

The usual first step in constructing a small two-way table is to simplify the data. This is also usual when constructing graphs. All the data in the teaching version of the BCS which accompanies these modules are already in categories. The easiest way of constructing tables is to either combine or omit some of the categories until only two categories are left. The process of altering categories is known as *recoding*.

There are several strategies for recoding. Omitting 'don't know' or 'not applicable' categories is an obvious idea. You can also merge categories together. In the BCS, if we were interested in the fear of car contents theft for example, we might merge the 'very worried' and 'slightly worried' categories together, and then merge the 'not very worried' and 'not at all worried' categories together, to form two categories; 'worried' and 'not worried'. Or we could (as the Home Office report does), simply take the first category – 'very worried' – and merge the other three together.

Click here

And this pops up
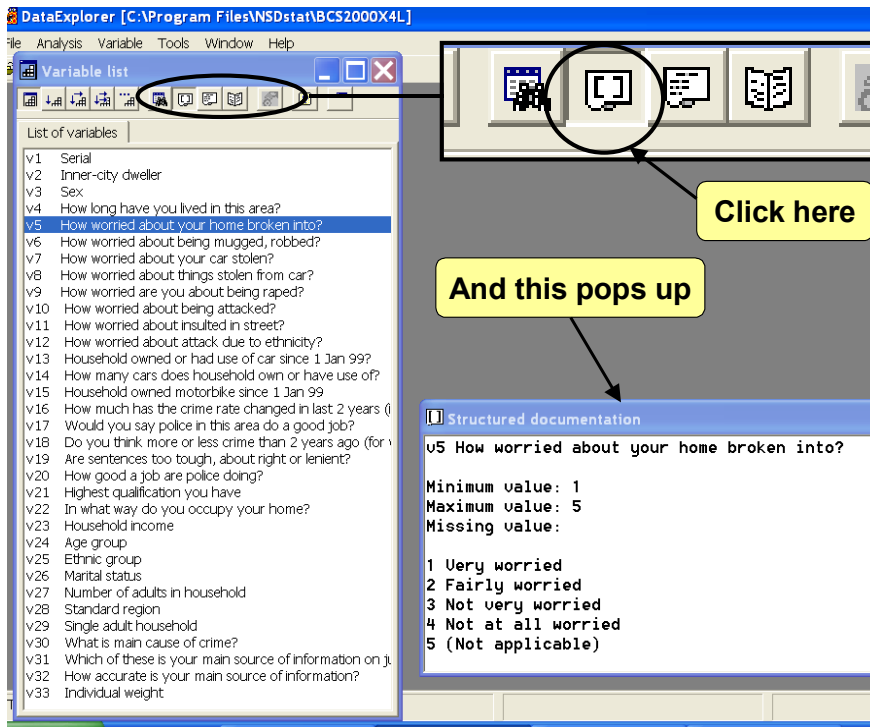
To recode using NSDstat[3], the first thing to do is to find out what the original codes were. NSDstat gives these codes the grandiose title of 'structured documentation'. To find them, highlight the variable you're going to recode, and then click on the 'view labels/ extreme values' button.

For example, let's look at the fear of burglary. The structured documentation shows that in this dataset, fear of burglary has four categories: 'Very worried' through to 'Not at all worried'. Each category has been assigned a numeral, which is called a *code*. To re-code, you get NSDstat to change these numerals. In this example, we want to combine categories 1 and 2 to form a new category of 'worried', and combine categories 3 and 4 to form a new category of 'not worried'.

The recode button in NSDstat is on the main window
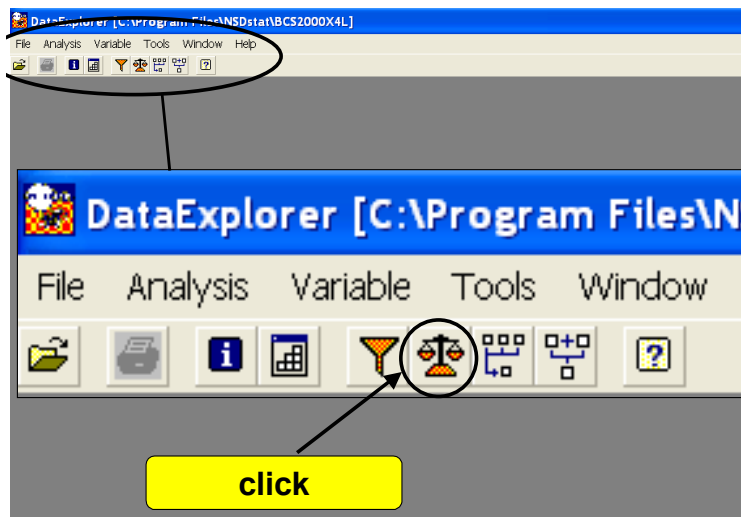
click

Clicking this button automatically creates a new variable (v34 in this database), which it asks you to describe. We might label it '**Recoded burglary fear**'. The next stage is to convert the data from the existing variable into a new format in the new variable.

To combine categories 1 and 2 from the original variable v5 we might enter

v5=1-2

in the **Selection** column and code this New Value '1'. An obvious description of the new category, to be entered as the Value Label into the right-hand column, is 'worried'. Similarly you could combine categories 3 and 4 into a new category, with a new code ('2' is obvious).

v5= 3-4

This you could call 'not worried'.
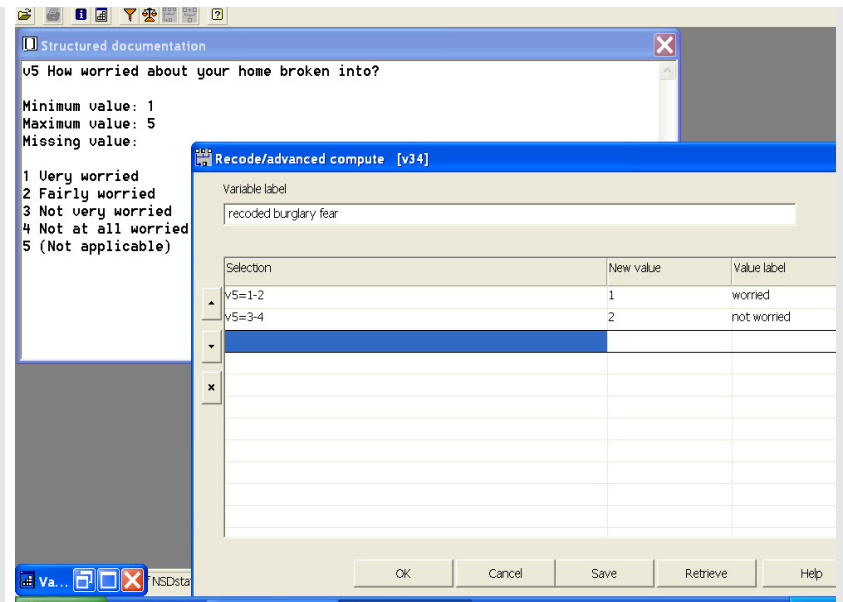
To move from one column to the next hit the <enter> ⬅ key on the keyboard

The following symbols can be used in recoding:

<	less than	>	greater than
<=	less than or equal to	>=	greater than or equal to
=	equal to	< >	unequal to
and	the conditions on both sides need to be met	or	the conditions on either side need to be met
( )	grouping of sub-conditions	-	group of numbers
,	several numbers		

Examples

v1<50	All cases where variable v1 is 49 or less
v1<30 or v1>50	All cases where variable v1 is *either* below 30 or above 50
v1>=30 and v1<50	All cases where variable v1 is between 30 and 49 inclusive
v1>50 and v2=2	All cases where variable v1 is above 50 *and also* variable v2 is equal to 2
v3=1,2,3	All cases where variable v1 is either 1 or 2 or 3
v1<10 and (v2=2 or v3=2)	All cases where variable v1 is below ten *and also* when *either* variable v2 or variable v3 is equal to 2
v3=1-3	All cases where variable v1 is between 1 and 3 inclusive
(v1=1 and v2=1) or v3<>4	All cases where *either* both variable v1 and variable v2 are equal to 1 *or* variable v3 is unequal to 4

When you're done with the recoding instructions, hit the OK button

The new variable v34 with only two categories will have been created from the data in variable v5, and this new variable can be examined just the same as the original variables, using frequency tables and graphs.

Exercise
Tabloid paper readership

See if you can recode the variable dealing with newspaper readership so that there are only two categories: tabloid newspaper readers and broadsheet newspaper readers.

The most relevant variable in the BCS is v31 concerning information about the criminal justice system. You need to make a new variable which only retains categories 1 and 2.

This will create a new variable with these frequencies.

Source: *British Crime Survey* 2000

Constructing a table (ii): Cross-tabulating

Now you are in a position to create the cross-tabulation. In this instance, we'll look at the fear of burglary by newspaper readership. To get NSDstat to do this, you click the bivariate button on the variable list, which has two arrows on it. Enter the independent variable (the new variable on newspaper readership) and the dependent variable (the new variable on recoded burglary fear) in the same way as you did for a frequency table, by highlighting the variable and clicking the right arrow.
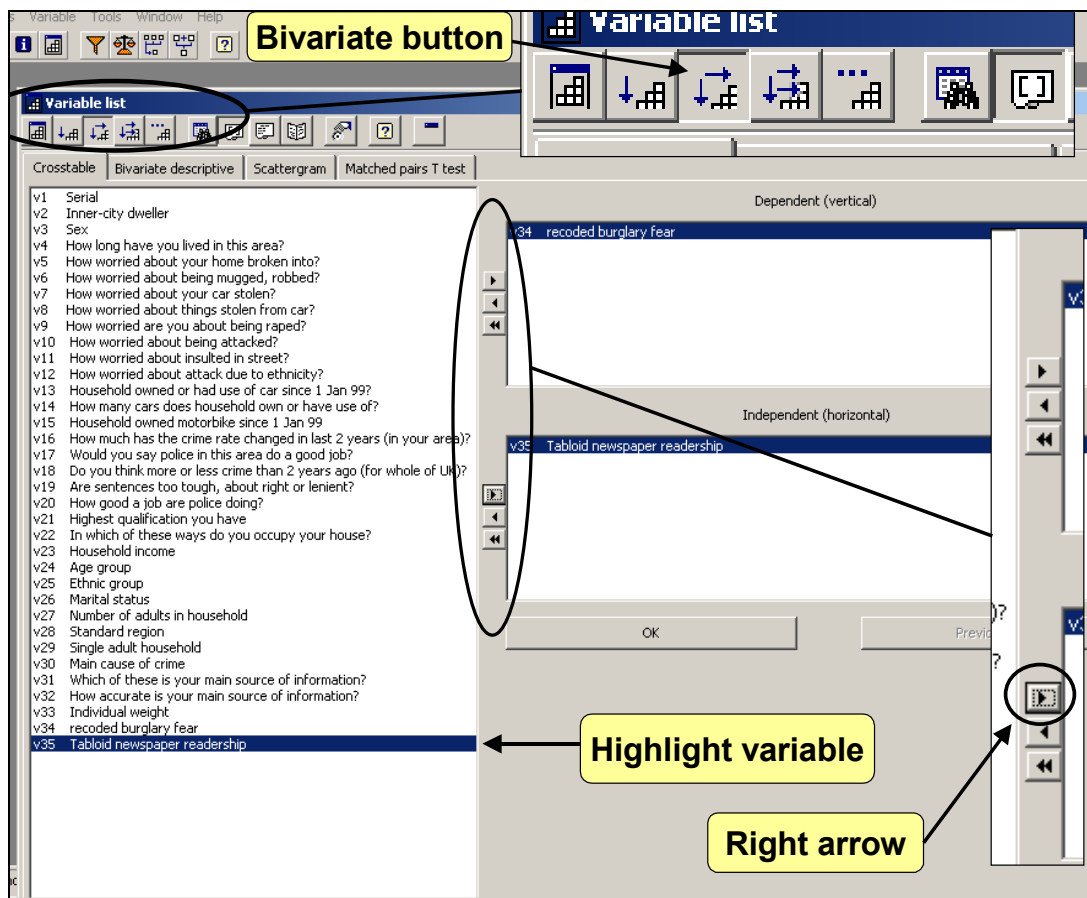
When you're done, click **OK**.

This produces the bivariate table, which tells you the level of association between the two variables. The NSDstat tables can be cut and pasted into other programs such as a word processor, or can be written to a file using tabs for import into a spreadsheet[4].

```
DataExplorer [C:\PROGRA~1\NSDstat\NSDBCS~1\NSDstat\COPY(2~1]
 File  Analysis  Variable  Tools  Window  Help

v35 Tabloid newspaper readership
v34 recoded burglary fear

Raw numbers
                    Broadsheet      Tabloid          Sum
-----------------------------------------------------------
worried                 588            846           1434
not worried             508            482            990
-----------------------------------------------------------
Sum                    1096           1328           2424
-----------------------------------------------------------

Included 2424 cases from a total of 19411
```

Source: *British Crime Survey* 2000

This table has two variables, each of which has two categories. It is therefore sometimes called a 2 x 2 table. A 2 x 2 table has four cells containing the data. In the next section you'll see how to examine tables such as these, and then how to analyse larger tables.

Analysing cross-tabs: small tables

The real advantage of simplifying two-way tables into a small number of categories is that it makes analysing the tables much easier.

The first step in analysing small tables is to decide which is the independent variable and which the dependent variable. When you use NSDstat it forces you to do this when constructing the table. In this example newspaper readership is the independent variable and fear of burglary the dependent variable, since it seems likely that the media affects the fear of burglary, rather than the other way round.

NSDstat puts the independent variable at the top of the table. The next step in analysing the table is therefore to get the percentages going down the columns of the table; the column

percentages. To do this, click on the table options button and choose Vertical percentaging from the pop-up box. This turns both columns into percentages going down, with both column percentages adding up to 100%. All the frequencies are then changed into a common scale of between one and one hundred, and making it easy to compare different tables.

All you do then is compare the two adjacent percentages in either row by subtracting one number from the other[5]. This shows the difference that the independent variable makes to the dependent variable. Ignore the sign, which is irrelevant for category data. (Note that since the sign is ignored it is irrelevant which row you compare, one row will be plus, and the other will be minus.)

So for these data the choice of tabloid or broadsheet newspaper makes a 10.1% difference (53.6 – 63.7) to the fear of burglary[6].

Source: *British Crime Survey* 2000

That is to say, we can report that in the BCS sample, reading tabloid newspapers as opposed to broadsheets made a 10.1% difference to whether or not the respondents were worried about crime.

Note that you only used vertical (column) percentages because NSDstat prompts you to put the independent variable in the columns. If you see a printed table with the independent variable going across the rows, you should use horizontal (row) percentaging.

Table 4.2: Fear of Attack by Gender British Crime Survey 2000			
			Percentages
	Not worried	**Worried**	**N=**
Female	53.5	46.5 =100%	*10542*
Male	28.8	71.2 =100%	*8803*
ALL	42.3	57.7 =100%	*19345*
Included 19345 cases from a total of 19411			

Table 4.2 shows the fear of being attacked broken down by gender. Since the independent variable – gender – is along the rows in the table you should percentage horizontally across the rows in this instance[7]. You then compare any two adjacent percentages in either column by subtracting one from the other[8].

A natural question is what percentage difference should be regarded as high. Prof. Davis[9] has suggested the following scale:

Value	Possible description
70% or above	A very strong association
50% to 69%	A substantial association
30% to 49%	A moderate association
10% to 29%	A low association
1% to 9%	A negligible association
0%	No association

(Note that 100% is a perfect association.)

However, don't feel too bound by these suggestions. The thing is to ask yourself what level of difference would you regard as noteworthy, or worthwhile following up. Associations in social science above 70% are very rare indeed; so even a substantial or moderate association may be noteworthy.

Analysing cross-tabs: large tables

Sometimes it is not possible or desirable to reduce the number of categories in both variables to two. Occasionally it may even be difficult to decide which of the variables is independent and which dependent. In this case, you can use a different measure, called 'V'[10]. Pick this from the table options, as well as vertical percentaging.

This is actually a more accurate measure but it tells you something different. This measure tells you how much of the dependent variable can be explained by the association with the independent variable, which is another way of thinking about association. However, while the measure is more accurate, it is harder to explain to non-specialists.

Source: *British Crime Survey* 2000

Note also that most computer programs print V (and most other measurements) as decimals rather than percentages. Just move the decimal point 2 places to convert one to the other. In the example above the association between household income and fear of burglary is 9.4%, a low to negligible association.

You are now in a position to start investigating the claims made in previous modules about the underlying trends in crime.

Exercise 1
The Home Office report on the British Crime Survey recoded the concern about crime into 'very worried' and all the others, rather than combining the two 'worried' and the two 'not worried' categories as we did above. Try recoding and cross-tabulating the concerns over crime with newspaper readership in the way the Home Office did, and see if you get similar results to the report, which were charted in Module 2.

Exercise 2
In Module 2 it was claimed that tabloid newspaper readership had an effect on the perception of crime rates.

Investigate this by
1. Considering which two variables in the British Crime Survey measure these variables
2. Recoding these variables
3. Cross-tabulating the variables, and analyse the table

Exercise 3
The alternative explanation in Module 2 was that income affected the fear of crime. Use the British Crime Survey to examine this explanation.

What do you think?

To what extent are either newspaper readership or income associated with views on crime?

SUMMARY

- An important question is whether variables are associated, so that one alters when another changes.
- To investigate this a cross-tabulation is used to produce a two-way table.
- This process often involves simplifying the data
- To analyse a two-way table turn the frequencies into percentages and compare them.

Notes

[1] An inverse relationship is one in which the variables change in opposite directions to eachother.

[2] This type of two variable table is sometimes called a contingency table.

[3] One very good reason for using NSDstat is that Nesstar does not support recoding.

[4] To do this, put the mouse pointer in the window and click the right hand mouse button. This will cause a range of choices to pop-up on the screen.

[5] Note that Windows has a calculator as an accessory program

[6] Some people prefer to do this another way for a 2 x 2 table: add together the top left percentage and the bottom right percentage and subtract 100. This gives the same answer, since you ignore the + or - sign.

[7] If you want a simple program to help you analyse a small table that you see printed in a book or newspaper try using the Figure-It program, available on Hwww.figuringfigures.comH.

[8] If you get confused as to which way to subtract try using the method in note 6 instead.

[9] Davis, J A, *Elementary Survey Analysis*, Englewood Cliffs NJ: Prentice Hall, 1971 p.49

[10] It is sometimes called 'Cramer's V', since it was developed by a statistician called Cramer.