Data Analysis workbook – Examining fear of crime in SPSS

This data analysis workbook shows how to use SPSS to explore data and conduct some analyses examining fear of crime.

In this worksheet, you will:

- use weighting variables
- examine missing values
- explore categorical variables using frequency tables and charts
- produce descriptive statistics for a continuous variable
- analyse the relationship between categorical variables using crosstabulations
- recode and compute variables

Task – Examine fear of crime by analysing the Crime Survey for England and Wales, 2013-2014

A research project aims to identify factors linked to the fear of crime. In particular, the research seeks to understand:

- how fear of crime varies by sex and age
- if a relationship exists between the fear and experience of crime.

The tasks in this workbook are to:

- get the data ready for analysis (including using a weighting variable) (section 1).
- explore the key variables (Section 2)
- examine relationships between variables (Section 3)
- manipulate data to aid data analysis (Section 3)
- introduce a third variable (Section 4)

Follow instructions marked with a ➔ on your computer.
Answers to the numbered questions can be found at the end.

This workbook is supported by our guide Using Survey Data

1. Get ready

1.1. Open data in SPSS

The worksheet uses the Crime Survey for England and Wales, 2013-2014, which registered users can download from the UK Data Service.

For a guide to registering, downloading and opening the data in SPSS, see the worksheets:

- Register and download UK Data Service data
- Getting started with SPSS

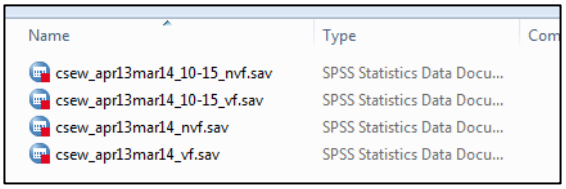
Which file?

When you download the Crime Survey for England and Wales, 2013-2014, the zip file includes four data files.

There are different files for two reasons:

1. data for young people (aged 10-15) is kept separate from the data for adults (marked with 10-15 on the file name)
2. in addition to data relating to individuals, there are records of data relating to incidences of crime (marked with a vf in the file name: vf standards for victim format).

For the research project here, the focus is on individuals over the age of 15. We therefore, need the file csew_apr13mar14_nvf.sav.



1.2. Apply the weight

Survey weights make the sample data more representative of the population. It is therefore important to use weighted results in your work.

- Find out more about weights in the accompanying guide 'Using Survey data'

You need to apply the weight before you start your analysis; the process is, however, fairly simple. Follow the instructions below.

Find the weight in the documentation

First, find out the name of the weight(s) in your dataset. Details of weighting variables are usually contained within a User Guide accompanying the survey data. For data from the UK Data Service, the documentation is available through:

- the catalogue record on the UK Data Service website (ukdataservice.ac.uk/get-data)
- the zip file you downloaded from the UK Data Service

To access the documentation and find details of the weights:

➔ go to the documentation for Crime Survey for England and Wales, 2013-2014
➔ open the user guide for the main adult dataset (7619_csew_adult_userguide.pdf)
➔ find the section on weights *(Tip: when looking at PDFs, press Control-F to search for 'weight')*

Which weight do I use?

Like many datasets, you will find that the Crime Survey for England and Wales, 2013-2014 contains more than one weight. Why? Different weights are for different types of analysis. In this case, there are weights to use when analysing individuals, households or incidents of crime.

For this research, our focus is on individuals, we therefore need to select the weight for individual level analyses.

Question 1 What is the name of the weight for individual level analyses?

Apply the weights in SPSS

The next step is to apply the weight in SPSS.

You can either:

- ➔ use the menus to select Data > Weight Cases
- ➔ press the weighting icon  on the tool bar

This will open the 'Weight Cases' Dialogue box.

In the 'Weight Cases' Dialogue box,

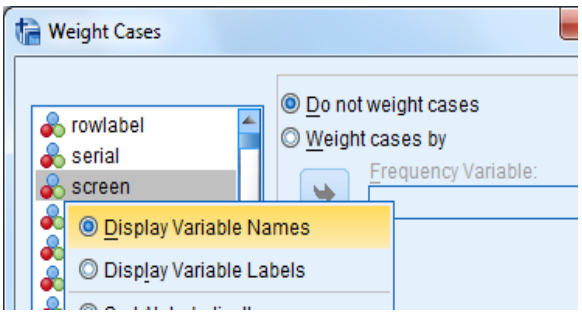
- ➔ find the weighting variable in the list on the left (In the CSEW, 2013-2014, the weight for individual level analyses is called c11indivwgt)
- ➔ select the variable and move it into the box on the right > click OK.

Tip: Names and Labels

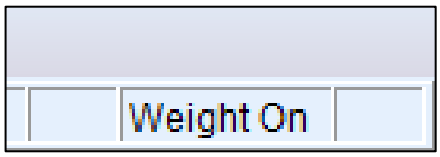
Variables tend to have short (often incomprehensible) names and a more descriptive label.

SPSS lists variables using either names or labels. To change from Names or Labels, right click anywhere on a variable list. This will open a new menu allowing you to choose either 'Display Variable Names or Display Variables Labels.

In this worksheet, we refer to variables by their name so you may find it easier if you select 'Display Variable Names'.



- ➔ check the weight is applied by looking at the bottom right corner of the Data Editor. It should display 'Weight On'.



2. Explore key variables

With the weight applied, start examining the variables.

2.1. Run frequency tables to explore fear of crime

First, consider the variables measuring our main concept of interest, 'fear of crime', which are:

- walkdark
- homealon

Check the question wording

Before running analyses, it is important to check how questions were asked in the questionnaire (which will usually be included in the survey documentation).

Note: you usually would have done this when deciding if the data meets your needs.

For the variable walkdark, the question is shown below.

WALKDARK [ASK ALL MODULE D RESPONDENTS]

How safe do you feel walking alone in this area after dark? Would you say you feel...READ OUT

NOTE: IF RESPONDENT NEVER GOES OUT ALONE AT NIGHT, PROBE: How safe WOULD you feel?

1. Very safe
2. Fairly safe
3. A bit unsafe
4. or very unsafe?

- Can you find the question wording for homealon? (tip: remember in a PDF you can use Ctrl+F and search for the variable name).

Make a frequency table

The variables measuring fear of crime are categorical variables. Categorical variables are common in survey data because questionnaires include lots of tick-box type questions. A good way to explore categorical variables is to use frequency tables.

Run a frequency table in SPSS,

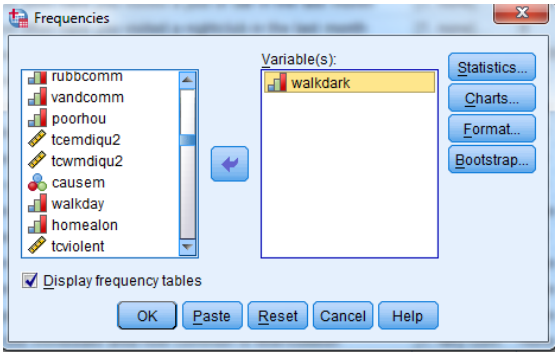
➔ using the menus, select Analyze > Descriptive Statistics > Frequencies...

In the 'Frequencies' dialog box,

➔ select the variable walkdark.

➔ press the arrow to move the variable into the right hand box > click OK

The table will appear in the *Output Viewer* window.



Tip: Clicking on any variable in the list and typing 'w' will allow you to search variables beginning with w.

Inspect the frequency table

The small table that appears first summarises the number of valid and missing responses. The larger table gives the frequencies and percent for each category.

Statistics

How safe do you feel walking alone after dark

N	Valid	10741299
	Missing	34541128

How safe do you feel walking alone after dark

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very safe	3402248	7.5	31.7	31.7
	Fairly safe	4250595	9.4	39.6	71.2
	A bit unsafe	2088792	4.6	19.4	90.7
	Very unsafe	999664	2.2	9.3	100.0
	Total	10741299	23.7	100.0	
Missing	Refusal	2991	.0		
	Don't know	71915	.2		
	System	34466222	76.1		
	Total	34541128	76.3		
Total		45282427	100.0		

Frequency tables are fairly straightforward ways of representing data; however, it is important to examine them thoroughly to ensure you understand the data.

Question 2 Looking at the table for Walkdark, can you identify any (odd) features of the data you may need to understand before interpreting the figures?

2.2. Rescale a grossing weight

A feature of the table above you may have noticed is that the frequencies are very large.

- According to the catalogue record, the sample size in the CSEW 2013-14 is 35,371. In this table, the number of cases appears to be over 45.5 million!

Grossing weights

The reason why the frequencies in the table above are much larger than the sample size is the weighting variable.

The weight in the CSEW is designed to gross the frequencies to the population of England and Wales. A 'grossing weight' is useful for calculating the prevalence of phenomena at the nation level. However, because of the large numbers the data appears overly reliable.

We need to weight the data to get accurate results; but, to get a clear idea of the variables it is useful to amend the weighting variable to avoid inflating the frequencies.

Weighting can be a confusing aspect of using survey data but the following instructions will guide you through the process of adjusting the weight variable.

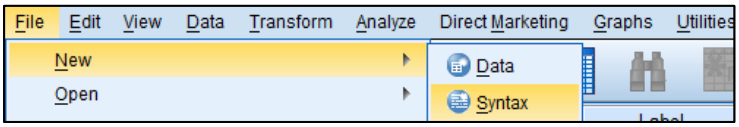
Create a rescaled weighting variable

SPSS syntax

The instructions below use SPSS syntax, a language for commanding SPSS.

In SPSS, you can run analyses using either menus with dialogue boxes or SPSS syntax. Menus and dialogue boxes are a good way to get started with SPSS. But, syntax is often a more efficient way of running and recording analyses. The efficiency of syntax is the reason it is used here in this example.

The aim is to rescale the weight to stop inflating frequencies from the sample counts to the population. To do this, we divide the weighting values for all cases by the mean of the weight.



➔ Open the syntax editor window using the menus
File > New > Syntax

➔ Copy the text below in to the Syntax window

```
COMPUTE rescaledweight=C11IndivWgt/1280.21. EXECUTE.
```

Tip: you need to be precise. Make sure the spelling is correct and include the full stops.

This syntax commands SPSS to create a new variable called 'rescaledweight'. Values on this new variable correspond to values on the old weight (C11IndivWgt) divided by 1280.21, which is the mean of the original weight.

➔ Highlight the commands and press the green arrow on the tool bar

Tip: How to get SPSS to calculate the mean of the weight
To use the same procedure for rescaling a grossing weight on other datasets (such as the CSEW for other years), you will need to identify the mean of the weight. To find the mean of a variable in SPSS:

➔ *Analyze > Decriptive Statistics > Descriptives*

In the 'Descriptives' dialogue box.

➔ Select and move the variable into the box on the right > click OK.

The mean will be reported in the output window.

Change the weight and re-run the frequency table

➔ Apply the new weight using either the menus or 

➔ Re-run the frequency table for Walkdark in SPSS

How safe do you feel walking alone after dark

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very safe	2658	7.5	31.7	31.7
	Fairly safe	3320	9.4	39.6	71.2
	A bit unsafe	1632	4.6	19.4	90.7
	Very unsafe	781	2.2	9.3	100.0
	Total	8390	23.7	100.0	
Missing	Refusal	2	.0		
	Don't know	56	.2		
	System	26922	76.1		
	Total	26981	76.3		
Total		35371	100.0		

2.3. Examine missing values

The next feature to consider is the missing values. Understanding the reasons for any missing values is important for ensuring you interpret figures correctly.

In the frequency table for Walkdark there are three categories of missing data. In particular, it is notable that the category 'System' has 26,922 missing values!

- Do you know why there are so many missing values?

If attentive when looking at the question wording, you may know the reason why. If not, try looking at the question again.

Not everyone was asked the question

In the questionnaire, you can see the text 'ASK ALL MODULE D RESPONDENTS'. This text indicates that only some survey respondents (those given Module D) were asked the question.

- Why? Read the following text from the User Guide, which explains the structure of the CSEW questionnaire.

"The main CSEW questionnaire has a complex structure consisting of a core set of modules asked of the whole sample, a set of modules asked only of different sub-samples, and self-completion modules asked of all respondents aged 16 to 59. Modules include, for example: victimisation; performance of the criminal justice system (CJS); contact with and attitudes to the police and the CJS; mobile phone theft; anti-social behaviour; plastic card fraud; and demographic characteristics of the respondent and household...Respondents are randomly allocated into one of four sub-samples, A, B, C or D which each represent around a quarter of the overall sample. When a question is only asked of a sub-sample of respondents this is indicated on the questionnaire' (p.8)

Question 3 Why do you think this structure is commonly used in surveys?

Refusal and don't know responses in attitudes questions

We also need to consider 'refusals' and 'don't know' responses.

If the number is large, it could indicate a problem with the question and for attitudinal questions, especially, you might want to retain 'don't know' or 'refusal' as valid responses.

For Walkdark, there are only a small number of don't know responses and refusals. We therefore do not need to be too concerned about treating these as missing values in our analysis. Cases with missing values will be excluded from the analyses.

Tip: It is good practice to report information on the missing values when writing up your research. When possible explain why data might be missing and consider the potential effects on the results.

2.4. Interpret the frequencies

With a good understanding of the missing values you are now able to focus on interpreting the data in the frequency table.

Examine the frequency table for Walkdark

Question 4 What percentage feels a bit or very unsafe walking alone after dark? Is this higher or lower than you would have expected?

Percent and valid percent

In the frequency table,

- 'Percent' displays the percent of cases in each category.
- 'Valid Percent' displays the percent of cases in each category excluding the missing values.

Usually, it will make more sense to refer to the valid percent when interpreting the data.

Run and interpret a frequency table for homealon

➔ Run a frequency table for homealon.

How safe do you feel when alone in home at night

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very safe	5458	15.4	64.6	64.6
	Fairly safe	2376	6.7	28.1	92.8
	A bit unsafe	494	1.4	5.8	98.6
	Very unsafe	117	.3	1.4	100.0
	Total	8444	23.9	100.0	
Missing	Don't know	5	.0		
	System	26922	76.1		
	Total	26927	76.1		
Total		35371	100.0		

Question 5 What percentage feels either a bit or very unsafe when home alone?

Is this higher or lower than the proportion feeling unsafe walking alone after dark?

2.5. Use charts to show fear of crime

We can also present the information from the previous table using bar or pie charts.

Make a bar chart

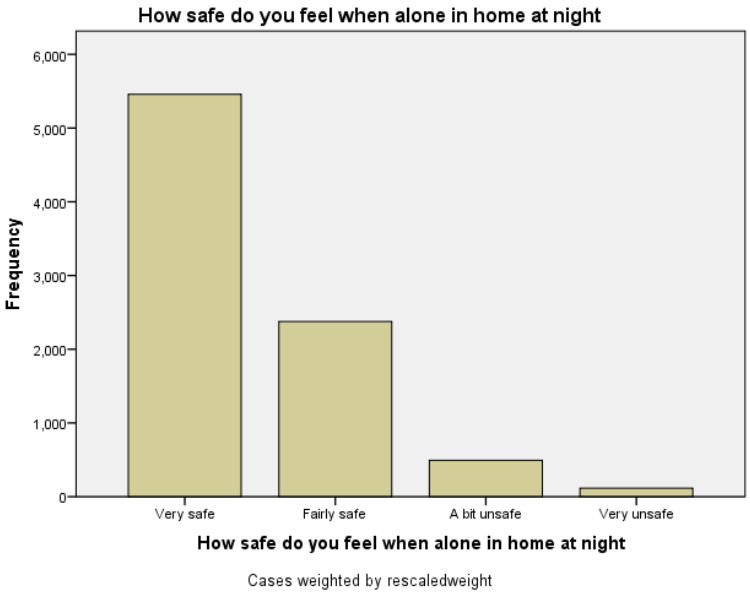
To get a bar chart,

- ➔ open the Frequencies dialog box

In the Frequencies dialog box,

- ➔ click Charts
- ➔ Select Bar charts and Select Frequencies > press Continue
- ➔ In the Frequencies dialog box, click OK

You should get the chart on the right.



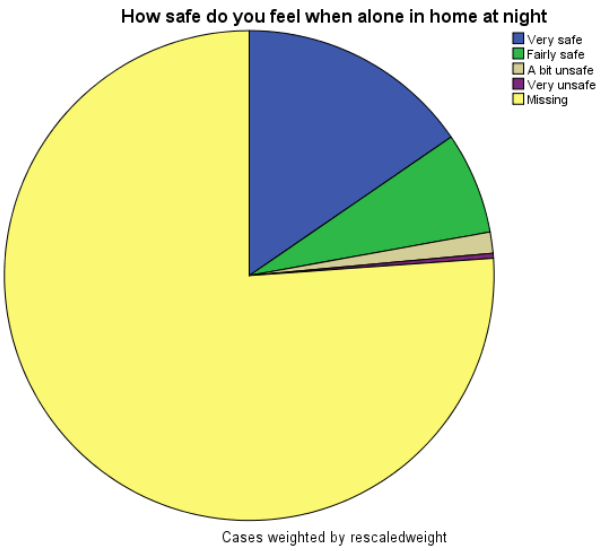
The bar chart shows the frequencies for the valid categories. From the chart, 'very safe' is clearly the largest category.

- Try copying and pasting the chart into a Microsoft Word document

Pie chart

You can also use a pie chart to show the percentages in each category. To create a pie chart,

- ➔ open the Frequencies dialog box
- ➔ select Charts
- ➔ Select Pie Charts and select Percentages > press Continue
- ➔ In the Frequencies dialog box, click OK



The pie chart here shows a high proportion of missing values. The number of missing values is high because not all respondents were asked this question (as discussed on page 8). Within this in mind, do you think this pie chart is easy to interpret?

Which chart to use?

There is not one correct chart for exploring variables. For example,

- a pie chart can highlight either a notably large or small proportion of responses
- a bar chart of the raw counts can usefully show how many people answered the question

2.6. Use descriptive statistics and a histogram to explore experience of crime

Next, examine the variables measuring experience of crime, which are:

- delibvio
- ndelibv

Check the question wording

Like with the questions on fear of crime, it is important to first check the questionnaire to understand how the questions were asked in the survey.

- Can you find the question wording for delibvio and ndelibv?

DELIBVIO [ASK ALL]
22

And again, [apart from anything you have already mentioned], since the first of [^DATE^] has anyone, including people you know well, DELIBERATELY hit you with their fists or with a weapon of any sort or kicked you or used force or violence in any other way?

1. Yes
2. No

NDELIBV [ASK IF DELIBVIO = YES]

How many times has this happened?

NOTE: 97 = 97 OR MORE/TOO MANY TO REMEMBER. USING CODE 97 CAN CAUSE PROBLEMS IN SEPARATING SINGLE AND SERIES INCIDENTS, SO PROBE FOR BEST ESTIMATES WHERE POSSIBLE

1..96
97 More/too many to remember

The first question refers to a date. Find out more about the timeframe in the question by scrolling further up the document until you find the start of this question module.

Question 6 The questions ask about experiences over the .....

What else is referred to in the framing of the question?
How might these references affect responses?

Question routing

The second question of 'how many times this happened?' is understandably only asked of participants who have experienced deliberate violence. This is a form of question routing.

Question 7 How might the question routing affect the number of valid and missing cases on ndelibv?

Run frequency tables

Make frequency tables for delibvio and ndelibv

➔ Analyze> Descriptive Statistics> Frequencies

If anyone has deliberately used force/violence on adult respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	647	1.8	1.8	1.8
	No	34721	98.2	98.2	100.0
	Total	35368	100.0	100.0	
Missing	Refusal	0	.0		
	Don't know	3	.0		
	Total	4	.0		
Total		35371	100.0		

How many times has this happened (Delibvio)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	466	1.3	72.0	72.0
	2	90	.3	13.9	85.9
	3	26	.1	4.0	89.9
	4	20	.1	3.0	93.0
	5	14	.0	2.1	95.1
	6	4	.0	.6	95.7
	8	3	.0	.5	96.2
	9	1	.0	.1	96.3
	10	4	.0	.6	96.9
	11	1	.0	.1	97.0
	12	0	.0	.1	97.1
	15	3	.0	.5	97.6
	20	1	.0	.1	97.7
	24	4	.0	.7	98.4
	40	2	.0	.4	98.7
	50	1	.0	.2	98.9
	97	7	.0	1.1	100.0
	Total	646	1.8	100.0	
Missing	Don't know	0	.0		
	System	34724	98.2		
	Total	34725	98.2		
Total		35371	100.0		

Question 8
What proportion has experienced deliberate violence?

Out of those who have experienced violence, how many times has it usually happened?

Obtain descriptive statistics

The variable ndelibv counts the number of times a respondent experienced force/violence used against them in the previous year.

Frequency tables are not always a helpful way to examine the distribution of variables with quantitative measurement scales as often there are too many possible values.

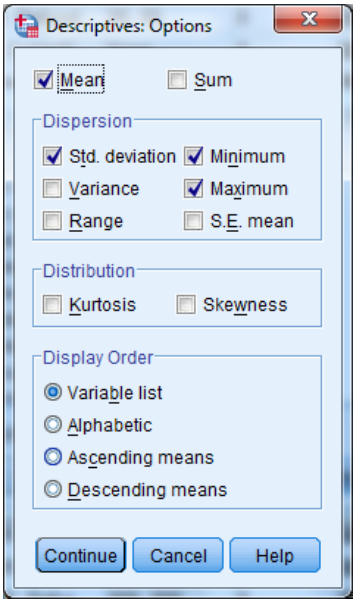
We can summarise interval variables using summary statistics such as the:

- minimum and maximum values reported
- mean (average) and median value
- the standard deviation (a measure of the spread of the data)

To get descriptive statistics for a variable using the menus,

➔ select Analyze> Descriptive Statistics> Descriptives
➔ select ndelibv from the list
➔ click OK

Tip: Select Options to choose which summary statistics SPSS reports



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
How many times has this happened (Delibvio)	646	1	97	3.06	10.659
Valid N (listwise)	646				

Warning: Always refer to the question wording

If we examine the question wording (shown on page 11) we find that "97" stands for "97 or more /too many to remember". This feature of the data is important. If the seven cases of "97" mean "too many to remember" rather than "97 or more" our interpretation of the mean and maximum values is incorrect. (As the next largest frequency is 50, the frequency table suggests that the 97's most likely mean "too many to remember).

What do we do?

An appropriate response may be to treat these seven values as missing (section 4.2 later in the guide includes instructions on how to indicate missing values).

The table below gives the descriptive statistics based on the remaining 639 cases.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
How many times has this happened (Delibvio)	639	1	50	2.02	3.965
Valid N (listwise)	639				

Question 9 Complete the text to summarise the variable?

From the people who have experienced violence against them in the last year, the number of incidences were between____ and ____. The average was____.

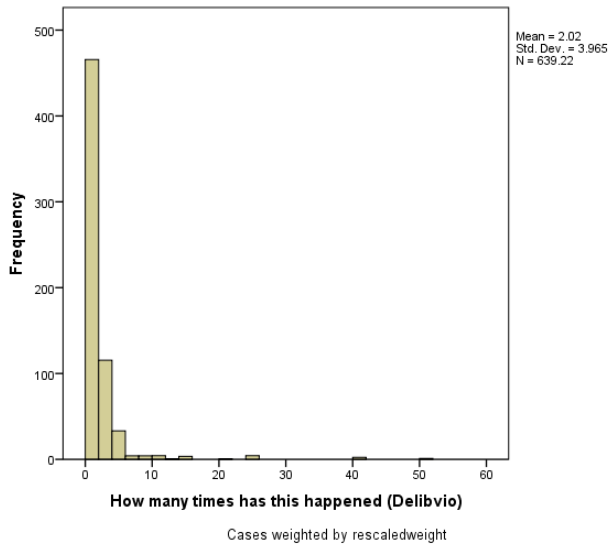
Question 10 Why are there no cases reporting 0 incidents of violence?

2.7. Make a histogram

A good way to look at the distribution of an interval variable is to use a histogram. A histogram groups values into 'bins' with bars to show the distribution of responses across the values.

One way to produce a histogram is to use the legacy dialog box:

- ➔ Select Graphs> Legacy dialogs> Histogram,
- ➔ Select ndelibv from the variable list.



The histogram for ndelibv, above, shows that most people who have been victims of violence, have only been victims once or a small number of times and that there are a smaller group of victims who have suffered many repeated violent attacks.

3. Examine relationships between variables: how fear of crime relates to sex and age

The previous sections have given us a good understanding of the key variables.

For example, in relation to the fear of crime, analysing the variables Walkdark and homealon suggests that:

- a majority of adults feel at least fairly safe walking alone and at home alone after dark
- nearly 30 percent of adults feel either a bit or very unsafe walking alone after dark
- a much greater proportion feel unsafe walking alone than home alone after dark (nearly 30 percent compared to less than 10 percent)

With this understanding of the key variables, you can start to examine the relationships between variables. In this section, we use bivariate analyses to establish if fear of crime relates to a person's sex and age.

In both cases, the analysis will relate to the relationship between two categorical variables. We will therefore use

- crosstabulations
- chi-square test
- stacked bar charts

3.1. Hypotheses

Before starting analyses, it is useful to consider what types of patterns you might expect to find. Thinking about the following questions:

- will fear of crime vary by sex? If so, are males or females more likely to report feeling unsafe?
- Are younger or older adults more likely to be concerned about their safety?
- will sex and age differences vary in relation to 'walking alone' or being 'home alone'?

....what patterns might you expect to find and why? (make some notes)

3.2. Crosstabulation of walkdark by sex

Before bivariate and multivariate analyses, you should examine each of the variables individually. We have examined the dependent variables walkdark and homealon but not the 'explanatory' variable sex.

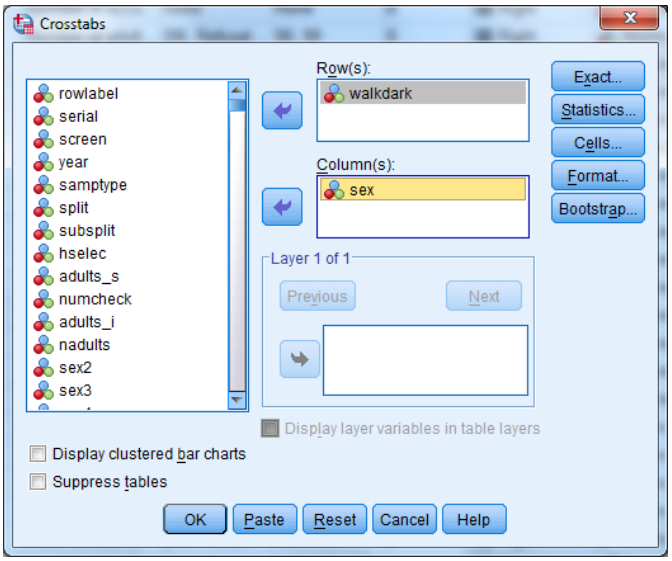
Start by examining sex:

➔ Select Analyze> Descriptives> Frequencies

Run a crosstabulation

Using the variable walkdark, examine whether feeling unsafe varies between males and females:

➔ select Analyze> Descriptives> Crosstabs...
➔ Put Walkdark in the Row(s)
➔ Put sex in the Column(s)



What goes in the row and what in the column?

It does not really matter which variable goes in the row and which in the column.

Though, in the social sciences it is common to

- put the output variable (the one you are interested in explaining) in the row
- the explanatory variable (the one that might influence the other) in the column

Ask for percentages

In this example, we are interested in looking at the effect of sex on fear of crime. When comparing groups, percentages for each group should add up to 100 percent. Sex is in the columns, we therefore need to ask for column percentages.

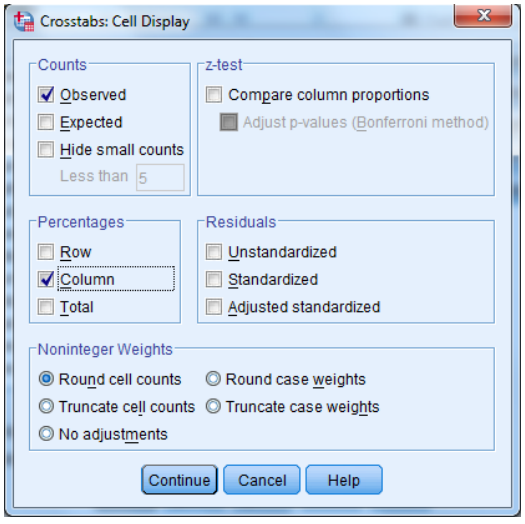
➔ Select the box Cells from the right

In the Crosstabs: Cell Display dialogue box

➔ Select column percentages > click Continue

You are now ready to make your table. In the crosstabs dialogue box

➔ Click OK



Interpret the two-way table (crosstab)

You should see the following table:

How safe do you feel walking alone after dark * Adult number 1 (respondent): Sex Crosstabulation

			Adult number 1 (respondent): Sex		
			Male	Female	Total
How safe do you feel walking alone after dark	Very safe	Count	1868	790	2658
		% within Adult number 1 (respondent): Sex	45.4%	18.5%	31.7%
	Fairly safe	Count	1630	1690	3320
		% within Adult number 1 (respondent): Sex	39.6%	39.5%	39.6%
	A bit unsafe	Count	475	1157	1632
		% within Adult number 1 (respondent): Sex	11.6%	27.0%	19.4%
	Very unsafe	Count	138	643	781
		% within Adult number 1 (respondent): Sex	3.4%	15.0%	9.3%
Total		Count	4111	4280	8391
		% within Adult number 1 (respondent): Sex	100.0%	100.0%	100.0%

What does this table tell us?

31.7 percent of adults report feeling 'very safe' walking alone after far (see the Total column on the right) but the proportion for women is only 18.5 percent.

Question 10 What proportion of females feel 'very unsafe' walking alone after dark? How does this compare to males?

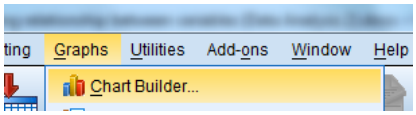
3.3. Make a stacked bar chart

A stacked bar chart is a good way to represent the differences by sex.

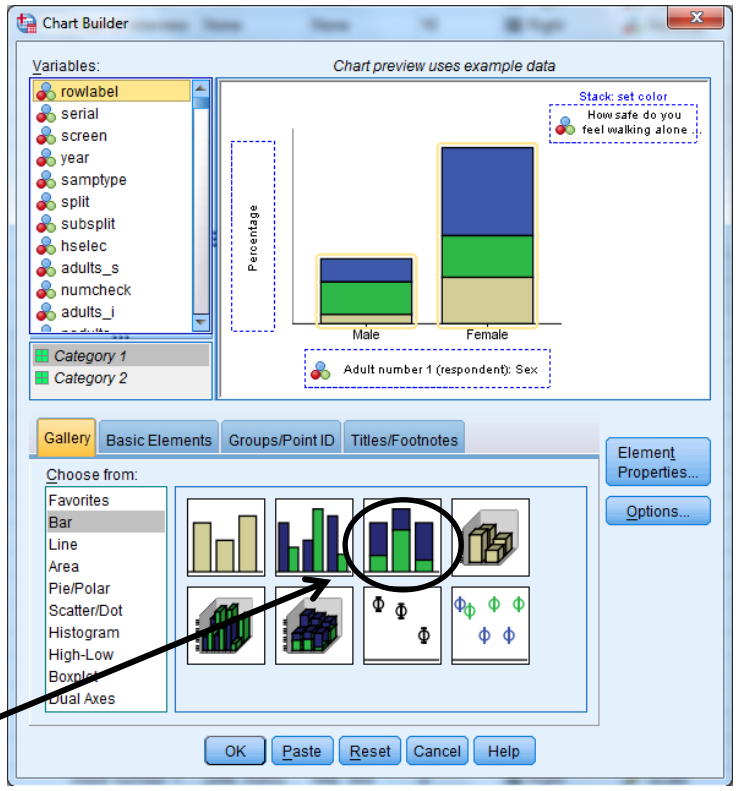
Use Chart Builder to make a stacked bar chart

To make a stacked bar chart, using the menus,

➔ choose Graphs> Chart Builder...



In the Chart Builder dialogue box,



In the preview box

➔ drag sex to the x axis

➔ drag walkdark to the Stack

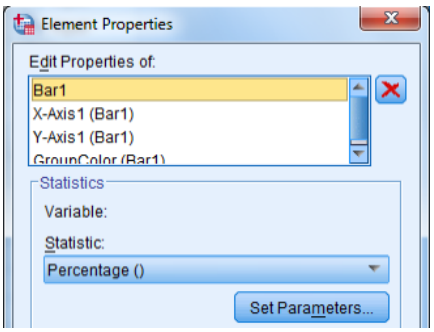
Under 'Choose from:'

➔ select Bar

➔ select the picture of the stacked bar chart

Use percentages to compare groups

Then, in Element Properties



➔ Select Percentage... as the statistic

➔ Press 'Set Parameters...'

➔ Select 'Total for Each X-Axis Category for the denominator'.

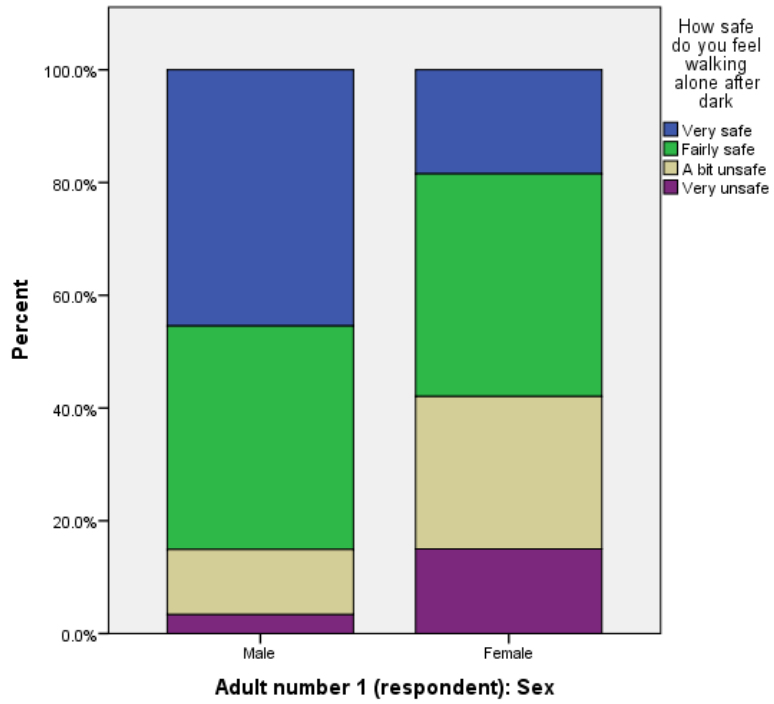
➔ Click Apply

Back in Chart Builder,

➔ Click OK

Examine your chart

You should have the following graph.

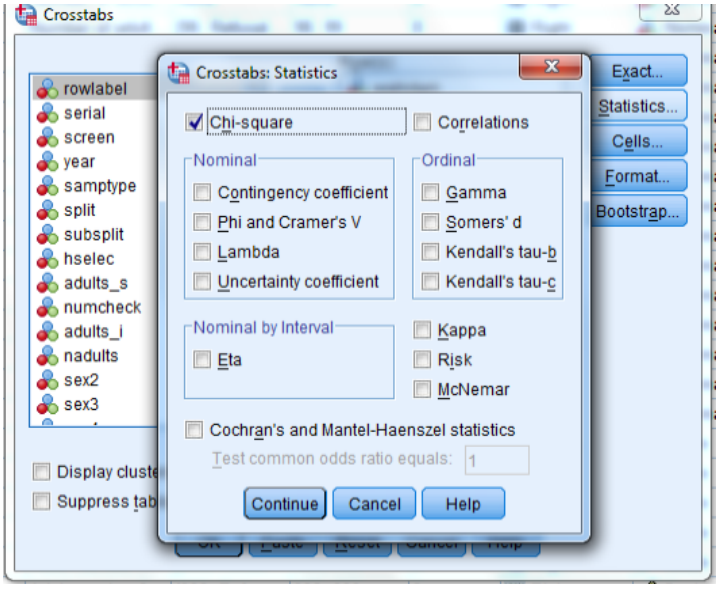


The graph shows how feelings of safety walking alone after dark vary between males and females. For example, we can see that the proportion of females that feel either a bit or very unsafe is about 40 percent compared to less than 20 percent of males.

3.4. A chi-square test

When examining relationships between categorical variables, a Pearson's chi-square test is commonly used.

Pearson's chi-square is a test of statistical significance. It is used to evaluate how likely it is that any observed difference between groups have occurred by chance.



Get the Chi-square

In SPSS, you can get a chi-square test through the crosstab menu.

➔ Analyze>Descriptive Statistics>Crosstab

In the Cross-tabs dialogue box,

➔ select Statistics...(on the right)

This will open a new dialogue box,

➔ select Chi-square
➔ press Continue

In the Cross-tabs dialogue box,

➔ press OK

Inspect and interpret the result

The results appear in a table where the first row reports the Pearson Chi-square Value, degrees of freedom and the significance value.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1046.847[a]	3	.000
Likelihood Ratio	1095.934	3	.000
Linear-by-Linear Association	1031.929	1	.000
N of Valid Cases	8391		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 382.64.

Question 11 Does the test suggest there is a statistically significant relationship between sex and fear of crime?

Assess reliability

For chi-square test results to be reliable there needs to be

- no more than 20% of the cells with expected counts of less than 5
- no cells with a count of less than 1.

Question 12 Check the expected cell counts given below the Chi-square results. Does the data meet the requirements of the chi-square test?

3.5. Crosstab sex and homealon

We have found sex differences in fear of crime using the variable walkdark. However, do we still find differences by sex when we examine feelings of safety when home alone at night.

➔ Follow previous steps to run a crosstab with a chi-square test to examine the relationship between sex and feelings of safety when home alone at night (homealon).

Add the missing information to the table below and answer the questions.

Table 1 How safe do you feel when alone in home at night by sex

	Male	*Female*	*Total*
Very safe	76.7%		64.6%
Fairly safe			28.1%
A bit unsafe		9.0%	
Very unsafe	.4%	2.3%	
Total	100.0%	100.0%	100.0%
n=		4317	

Pearson Chi-Square (χ^2)=561.883, p<0.001

Question 13 Are males or females more likely to feel safe at home at night?

According to the Pearson chi-square test, can we say there is a statistically significant relationship between sex and feelings of safety at home at night?

3.6. Differences by age: crosstabulations using a categorical age variable

The next step is to examine how fear of crime relates to age.

Select your variables

We will continue to use walkdark and homealon to measure fear of crime but how will we measure age?

If we measure age in age-bands we can use crosstabulations to explore the relationship between age and fear of crime. Fortunately, the dataset already includes three variables that measure age in bands.

Variable Name	Variable label
agegrp	Age group (3 bands
ageshort	Age group (5 bands)
agelong	Age group (9 bands)

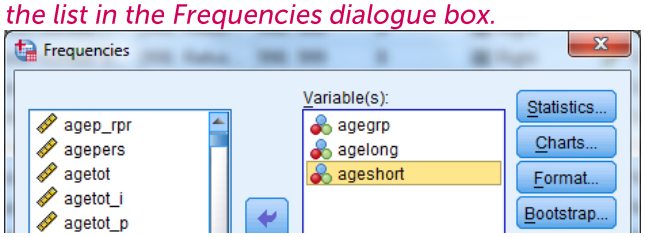
From the labels, we see that the first variable consists of 3 age bands, the second 5 age bands and the third 9 age bands.

Examine the variables to find out the age bands.

➔ Run frequency tables for agegrp, ageshort, agelong

Select an age variable to use.

Tip: you can ask SPSS to run tables for multiple variables simultaneously. Simply add all variables to the list in the Frequencies dialogue box.



Which one to use?

There is no 'correct' number of groups. Factors to consider include:

- categories need to be meaningful (keep together groups likely to share common experiences)
- too few categories can conceal important differences
- too many categories makes interpretation difficult and some percentages are based on only a very small number of cases

Run and interpret crosstabulations

➔ Using the same methods as for sex, produce two-way tables to examine if fear of crime varies across age groups.

Question 14 Is age related to feeling unsafe walking alone after dark? Summarise the pattern with reference to the percentages in the table.

4. Manipulate data to aid data analysis

In this section, you will create new variables to aid analysis.

4.1. Recode to create binary variables

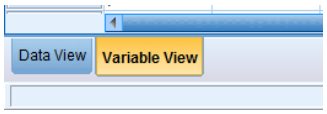
For certain analyses, it can be easier to simplify variables in a dataset by combining response categories.

For instance, we can create simpler measures of the fear of crime by recoding walkdark:

- the original variable has four main categories (Very safe, Fairly safe, A bit unsafe and Very safe)
- combining categories, the simpler measure will indicate whether someone feels unsafe (yes/no) walking alone after dark
- these binary variables will take the value 1 if the respondent feels either 'a bit' or 'very' unsafe and 0 if 'very' or 'fairly' safe

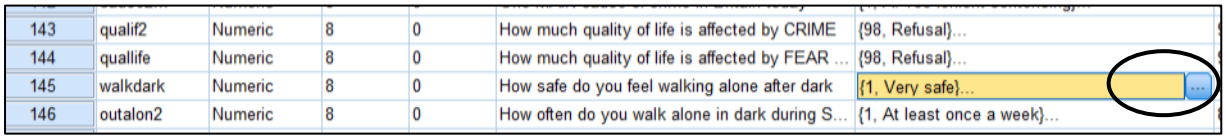
Check how variables are coded

Before we can recode an existing variable, we need to know how it is coded. You can check how variables are coded in the Variable View window.



In the Variable View window,

➔ identify the row for the variable of interest and in that row, click in the Values column

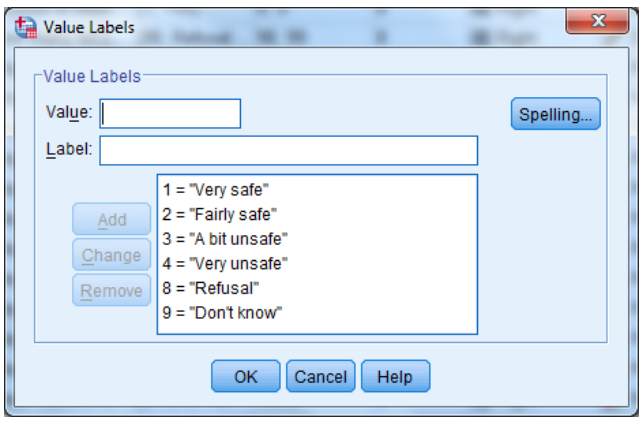


143	qualif2	Numeric	8	0	How much quality of life is affected by CRIME	{98, Refusal}...
144	quallife	Numeric	8	0	How much quality of life is affected by FEAR ...	{98, Refusal}...
145	walkdark	Numeric	8	0	How safe do you feel walking alone after dark	{1, Very safe}...
146	outalon2	Numeric	8	0	How often do you walk alone in dark during S...	{1, At least once a week}...

➔ Click the box with '...' to open the Value Labels dialogue box.

In the Value Labels dialogue box, you can see the value for each response category.

For instance, in this case those cases responding "Very Safe" will be assigned the value 1.



Plan the recode

The next step is to identify how to recode the existing variable into your new variable.

It is useful to use pen and paper to write down a list of the old and new values and their labels.

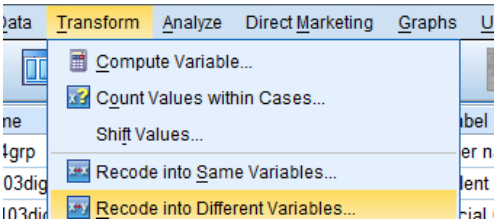
- Question 15 using the information above, can you plan the recode for the new variable identifying those who feel unsafe walking alone after dark?

Existing Variable = Walkdark			New Variable= unsafedark	
Value	**Label**		**Value**	**Label**
1	Very safe	}	0	No
2	_____			
__	_____	}	1	Yes
__	_____			
__	_____		8	Refusal
__	_____		9	Don't know

Recode into different variables…

With a plan, you can start the recode.

➔ select the menu Transform > Recode into Different Variables…

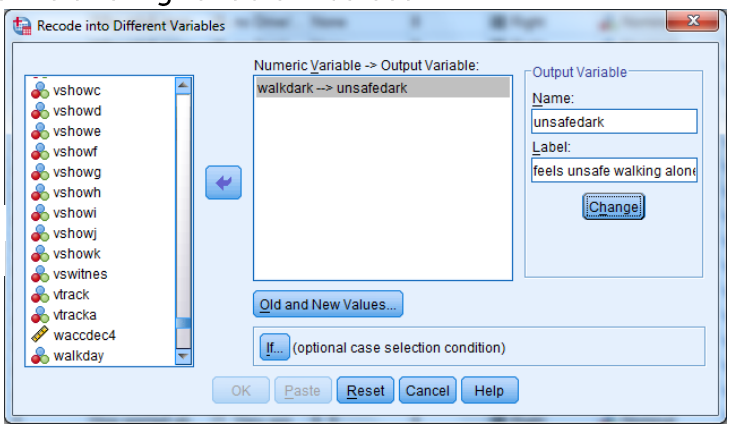


Rather than changing the original variable, this command creates a new variable as part of the recode.

- It is generally better to create a new variable and keep the original variable intact in case you make mistakes or change your mind.

Name the old and new variable

In the dialogue box, first indicate the existing variable in dataset:



➔ from the Variable list, select walkdark

➔ move the variable into the 'Numeric Variable __>Output Variable' box

Name and label the new variable using the Output Variable box,

➔ type 'unsafedark' under 'Name'

➔ type 'feels unsafe walking alone after dark (recoded binary measure)' under 'Label'

➔ click Change

Specify the values

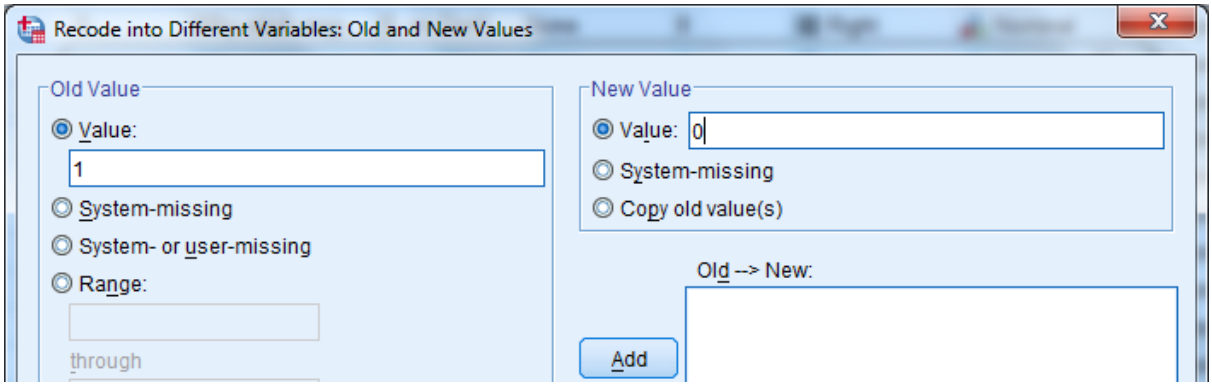
Now tell SPSS how to recode the existing variable into the new variable.

➔ click the Old and New values button

This opens a dialog box. To recode old to new values, you can:

➔ click Value on the left and add the value 1

➔ In the New Value area type 0 and click Add.



You have now defined the first new value. The new variable will take the value 0 when walkdark= 1.

Follow the same procedure to indicate values of the new variable when walkdark=2.

➔ click Value on the left and add the value 2, in the New Value area type 0

➔ click Add

An alternative way to recode consecutive values of an existing variable is to use the range option.

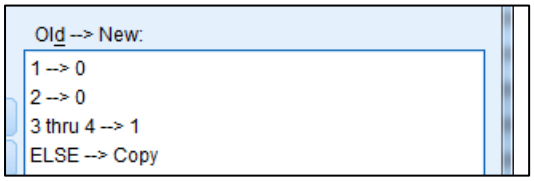
➔ on the left, click 'Range:' and insert 3 in the upper box and 4 in the lower box

➔ in the New Value area type 1 and click Add.

Finally, the code of the missing values is staying the same, so you can use the follow option

➔ On the left, click 'All other values'. In the New Value area click 'Copy old Value(s)' and click 'Add'

You should have following in the Old → New box.



➔ click *Continue* to return to the *Recode into Different Variables* dialog box

➔ click *OK*

The new variable unsafedark should now be visible at the end of your data in the Data Editor.

2645	bcageband02	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2646	bcageband03	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2647	bcageband04	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2648	bcageband05	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2649	bcageband06	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2650	bcageband07	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2651	bcageband08	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2652	bcageband09	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2653	bcageband10	Numeric	8	0	Household chil...	{1, 0-4}...	None	13
2654	rural3	Numeric	8	2	Type of area 20...	{1.00, Urban...	None	10
2655	rescaledwei...	Numeric	8	2		None	None	16
2656	unsafedark	Numeric	8	2	feels unsafe wa...	None	None	12
2657								
2658								
2659								

Check the distribution of the new variable

After any recode, it is useful to check the distribution of the new variable.

➔ create a frequency table for the new variable (*Analyze> Descriptive Statistics> Frequencies...*)

In the frequency table (below), the new variable looks as expected. The variable takes the values of 0, 1, 8 or 9. But, the table highlights how to complete the recode we still need to:

- add value labels to show what each value represents
- indicate that responses 8 or 9 should be classed as missing values

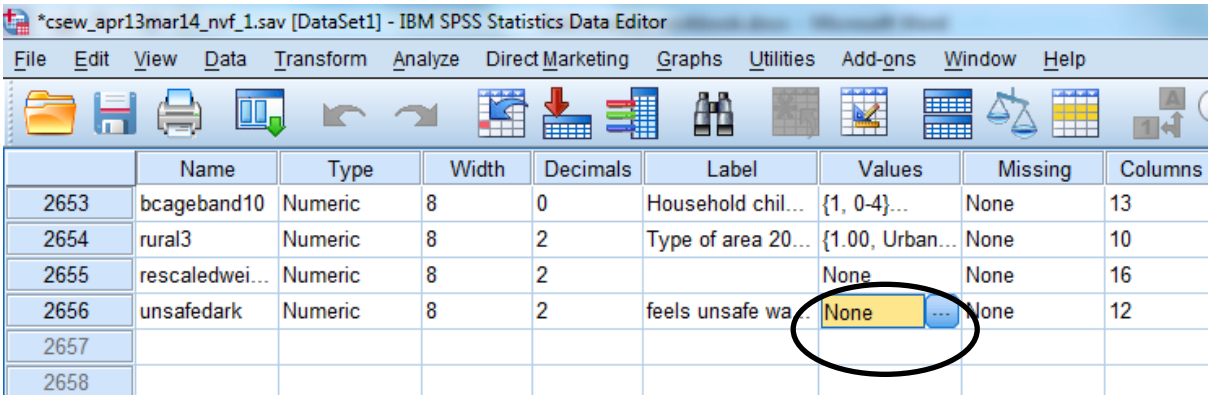
feels unsafe walking alone after dark (recoded binary measure)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	5978	16.9	70.8	70.8
	1.00	2412	6.8	28.6	99.3
	8.00	2	.0	.0	99.3
	9.00	56	.2	.7	100.0
	Total	8449	23.9	100.0	
Missing	System	26922	76.1		
Total		35371	100.0		

Label the categories in your new variable

Assigning value labels to the new variable will show what each value represents. To add value labels:

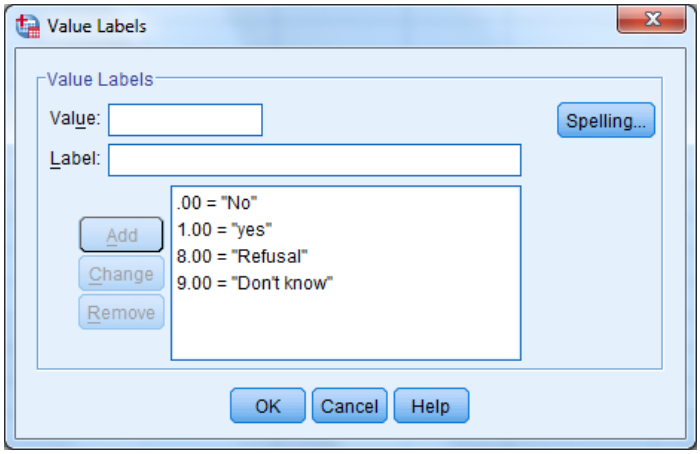
- ➔ find the variable in the Variable view in the Data Editor
- ➔ click on the cell in the Values column
- ➔ click on the box with '...' that appears when the cell is highlighted



The value labels dialog box will appear. Using this box, you attach value labels by indicating the 'Value' and corresponding 'Label'.

To label the new variable,

- ➔ type 0 in the 'Value' field and 'No' in the 'Label' field. Click Add.
- ➔ repeat this process for 1 'Yes', 8 'Refusal' and 9 Don't know' to get the following:



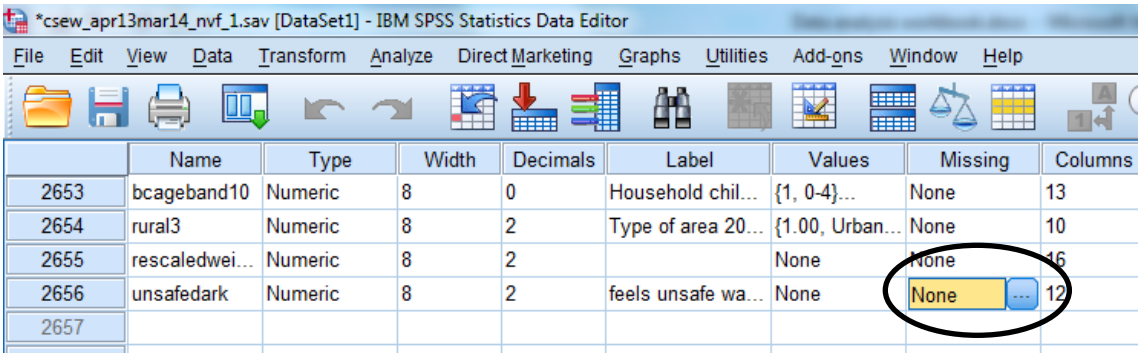
If you make any mistakes, use the *Change* and *Remove* buttons.

- ➔ select OK when finished

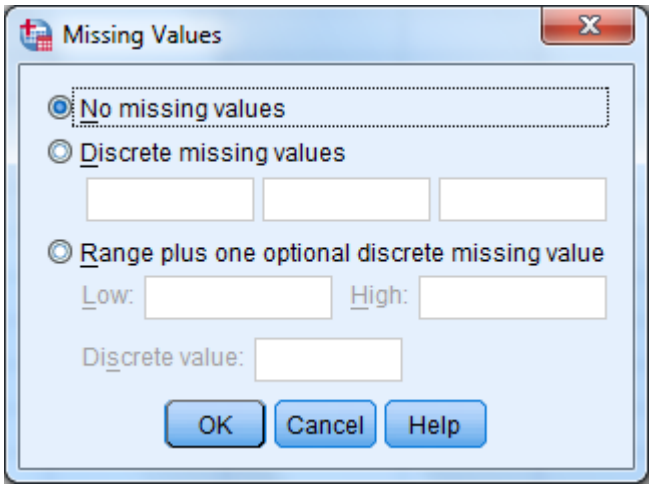
Missing values

You also use the variable view to indicate that certain response categories should be classed as missing values.

- ➔ find the variable in the Variable view in the Data Editor
- ➔ click on the cell in the Missing column
- ➔ click on the box with '...' that appears when the cell is highlighted



The Missing Values dialog box will appear with different options for indicating values.



In this case you can either,

- ➔ select 'Discrete missing values' and typing 8 in the first box and 9 in the second

or

- ➔ select 'Range plus one optional discrete missing value' and type 8 in the 'Low' field, 9 in the 'High' field (leaving the 'Discrete value' field empty.

- ➔ select OK when finished

To check the labels and indicating of missing values has work, we can create another frequency table

- ➔ Analyze> Descriptive Statistics> Frequencies...

4.2. Crosstabulate new variable with original to check recoding

When you create a new variable you should check the coding is as you expect. To check a straightforward recode, you can use a crosstabulation of the original and new variable.

➔ Analyze > Descriptive Statistics > Crosstabs

➔ move the variable walkdark to the Row(s) and unsafedalk into the Column(s)

The crosstabulation of the variable (shown below) confirms that those who responded 'Very safe' or 'Fairly safe' fall into the 'No' group on the new variable. Conversely, those who indicated that they feel either 'a bit unsafe' or 'very unsafe' are in the 'Yes' group.

How safe do you feel walking alone after dark * feels unsafe walking alone after dark (recoded binary measure Crosstabulation

Count

		feels unsafe walking alone after dark (recoded binary measure		Total
		No	yes	
How safe do you feel walking alone after dark	Very safe	2658	0	2658
	Fairly safe	3320	0	3320
	A bit unsafe	0	1632	1632
	Very unsafe	0	781	781
Total		5978	2413	8391

4.3. Using the recoded variable

You can now use the newly coded variable in your analyses.

Creating simple measures such as this binary measure indicating whether someone does or does not feel unsafe walking alone after dark can be useful in data analysis. For example,

- reducing the number of categories can make results easier to present and interpret and increase the number of cases within the different categories
- compared to ordinal measures, binary dependent variables are more easily analysed in multivariate analyses

But, thinking about the differences between Walkdark and Unsafedark, identify any disadvantages in recoding variables in this way?

5. Introduce a third variable

The previous analyses have been either univariate (looking at one variable) or bivariate (looking at the relationship between two variables). This sections, extends into more complex multivariate analyses.

Using three variables, the aim is to establish if:

- there is a relationship between experience and the fear of crime
- the relationship is the same between males and females

To address these aims, we will:

- use a crosstabulation to examine the relationship between the experience and fear of crime
- include sex as a third variable to make a three-way crosstabulation
- conduct and interpret a chi-square test for a three-way table

5.1. How does fear of crime relate to experience of violence?

Delibvio measures experiences of violence and the use of force. Exploring this variable in section 2, we found that just under 2 percent had experienced violence or the use of force during the 12 months prior to the survey.

Run a crosstabulation

Using a crosstabulation, we can examine if such experiences of crime are linked to the fear of crime.

➔ analyze > Descriptive Statistics > Crosstab
➔ put walkdark in rows, delibvio in columns
➔ in 'cells' choose column percentages
➔ in 'statistics' choose Chi-square

- You could also try using the simpler measure unsafedark derived in the previous section

Interpret the table and the chi-square test (see over page).

Question 16 Do feelings of safety walking alone after dark vary according to experience of violence? Does there appear to be a statistically significant relationship between experience of violence and the fear of crime?

How safe do you feel walking alone after dark * If anyone has deliberately used force/violence on adult respondent Crosstabulation

			If anyone has deliberately used force/violence on adult respondent		
			Yes	No	Total
How safe do you feel walking alone after dark	Very safe	Count	42	2615	2657
		% within If anyone has deliberately used force/violence on adult respondent	28.2%	31.7%	31.7%
	Fairly safe	Count	61	3260	3321
		% within If anyone has deliberately used force/violence on adult respondent	40.9%	39.6%	39.6%
	A bit unsafe	Count	31	1601	1632
		% within If anyone has deliberately used force/violence on adult respondent	20.8%	19.4%	19.4%
	Very unsafe	Count	15	766	781
		% within If anyone has deliberately used force/violence on adult respondent	10.1%	9.3%	9.3%
Total		Count	149	8242	8391
		% within If anyone has deliberately used force/violence on adult respondent	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.888[a]	3	.828
Likelihood Ratio	.903	3	.825
Linear-by-Linear Association	.696	1	.404
N of Valid Cases	8391		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.87.

5.2. Introducing a sex variable

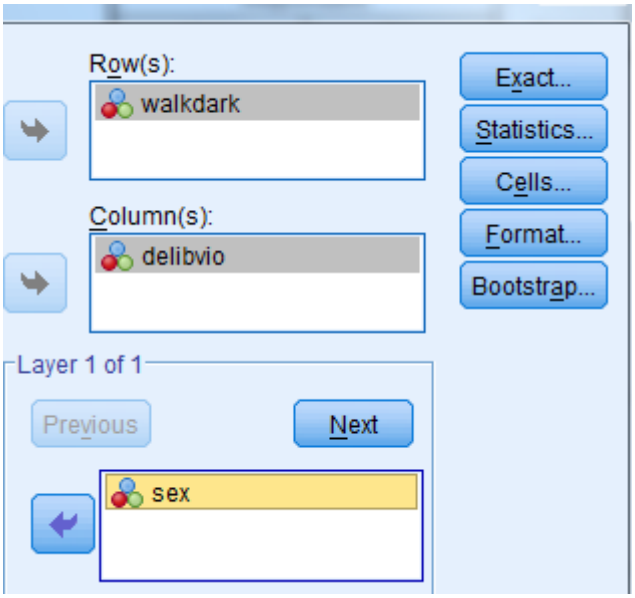
Exploring the key variables in section 2, we found females tend to be more afraid of crime than males. This increased level of fear may result from general socialisation processes rather than actual experiences of crime. Thus, the relationship between the experience and fear of crime could be different for males and females.

A three-way table, allows us to examine the effect of the fear of crime among males and females separately.

Run a three-way table and chi-square test

To create a three-way table to examine gender differences in the relationship between experience and fear of crime, we need to add Sex as a layer in the crosstabulation.

- ➔ Analyze> Descriptive Statistics> Crosstab...
- ➔ Put walkdark in the rows, delibvio in the columns
- ➔ Add sex as a layer
- ➔ Click on Cells and choose column percentages
- ➔ Click on Statistics and select Chi-square
- ➔ Press OK



Interpret the results

The three-way table presents values for males only, females only and the overall results (which are the same as earlier). Examining the differences, we can establish if the pattern is the same for males and females.

Question 17 Among males, which of the following statements applies?

- ☐ Those who have and have not recently experienced violence show not difference is feelings of safety walking alone after dark
- ☐ Those who have recently experienced violence feel less safe walking alone after dark
- ☐ Those who have recently experience violence feel more safe walking along after dark

Is the pattern the same for females?

How safe do you feel walking alone after dark * If anyone has deliberately used force/violence on adult respondent * Adult number 1 (respondent): Sex Crosstabulation

Adult number 1 (respondent): Sex		If anyone has deliberately used force/violence on adult respondent		Total
		Yes	No	
Male	Very safe	31	1837	1868
		32.0%	45.8%	45.4%
	Fairly safe	45	1585	1630
		46.4%	39.5%	39.6%
	A bit unsafe	15	460	475
		15.5%	11.5%	11.6%
	Very unsafe	6	132	138
		6.2%	3.3%	3.4%
	Total	97	4014	4111
		100.0%	100.0%	100.0%
Female	Very safe	12	778	790
		23.1%	18.4%	18.5%
	Fairly safe	15	1675	1690
		28.8%	39.6%	39.5%
	A bit unsafe	16	1141	1157
		30.8%	27.0%	27.0%
	Very unsafe	9	633	642
		17.3%	15.0%	15.0%
	Total	52	4227	4279
		100.0%	100.0%	100.0%
Total	Very safe	43	2615	2658
		28.9%	31.7%	31.7%
	Fairly safe	60	3260	3320
		40.3%	39.6%	39.6%
	A bit unsafe	31	1601	1632
		20.8%	19.4%	19.5%
	Very unsafe	15	765	780
		10.1%	9.3%	9.3%
	Total	149	8241	8390
		100.0%	100.0%	100.0%

Tip: Three-way tables tend to be fairly large. To help the table above fit on a page, some less necessary information has been removed. Reformatting output from SPSS can help to present results better. (you could also try examining a recoded variable such as Unsafedark from the previous section).

Chi-square test for a three way table

For the three-way table, there are separate chi-square tests for

- the relationship between experience of crime and fear of crime for men
- the relationship between experience of crime and fear of crime for women
- the relationship between experience of crime and fear of crime overall

Chi-Square Tests

Adult number 1 (respondent): Sex		Value	df	Asymp. Sig. (2-sided)
Male	Pearson Chi-Square	8.794[b]	3	.032
	Likelihood Ratio	8.600	3	.035
	Linear-by-Linear Association	8.369	1	.004
	N of Valid Cases	4111		
Female	Pearson Chi-Square	2.576[c]	3	.462
	Likelihood Ratio	2.667	3	.446
	Linear-by-Linear Association	.081	1	.777
	N of Valid Cases	4279		
Total	Pearson Chi-Square	.640[a]	3	.887
	Likelihood Ratio	.645	3	.886
	Linear-by-Linear Association	.564	1	.453
	N of Valid Cases	8390		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.85.

b. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 3.26.

c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.80.

Question 18 According to the Pearson Chi-square test, is there a significant association between fear and experience of crime...
...among males?

...among females?

Answers

Question 1 What is the name of the weight for individual level analyses?
c11indivwgt

Question 2 Looking at the table for Walkdark, can you identify any (odd) features of the data you may need to understand before interpreting the figures?
a) the very large number of cases
b) the high number of missing responses

Question 3 Why do you think this structure is commonly used in surveys?
If we ask different sets of questions to different sub-samples, we can increase the overall number of questions that can be asked without adding to the length of the questionnaire.

Question 4 What percentage feels a bit or very unsafe walking alone after dark? Is this higher or lower than you would have expected?
Around 29 percent

Question 5 What percentage feels either a bit or very unsafe when home alone?
Around 7.2 percent feel either a bit or very unsafe when home alone. A much lower than the proportion unsafe walking alone after dark

Question 6 The questions ask about experiences over the last 12 months. What else is referred to in the framing of the question? The question refers to places ('at home, in the street, at work, in a shop, in a park, on a train or anywhere else')
How might these references affect responses? The references to time should mean all responses relate to the same timeframe, which is the last year. Mentioning of places can help respondents recall incidences and clear any ambiguity about whether out or at home etc.

Question 7 How might the question routing affect the number of valid and missing cases on ndelibv?
A lower number of valid cases
A higher number of missing cases

Question 8 What proportion has experienced deliberate violence? Less than 2 percent
Out of those who have experienced violence, how many times has it usually happened? Only once

Question 9 Complete the text to summarise the variable?
From the people who have experienced violence against them in the last year, the number of incidences were between 1 and 50. The average was 2.02. (Note:statistics calculated excluding those responding "97 or more / too many to remember).

Question 10 Why are there no cases reporting 0 incidents of violence?
Those not experiencing incidence of violence were not asked the question

Question 11 What proportion of females feel 'very unsafe' walking alone after dark? How does this compare to males? 15 percent of females report feeling very unsafe walking alone after dark: a much higher proportion than for males (3.4 percent).

Question 12 Does the test suggests there is a significant relationship between sex and fear of crime? Yes, the test suggests the difference between males and females is statistically significant at the $p<0.001$ level.

Question 13 Check the expected cell counts given below the Chi-square results. Does the data meet the requirements of the chi-square test? In this case, the results are reliable

as fewer than 20% of cells have expected counts of less than 5 and no cells have less than 1.

Question 14 Are males or females more likely to feel safe at home at night?
Males are more likely to report feeling safe at home at night. Around 97 percent of males reporting feeling wither very or fairly safe compared to about 89 percent of females.
According to the Pearson chi-square test, there appears to be a significant association between sex and feeling safe when home alone at night.

Question 15 Is age related to feeling unsafe walking alone after dark? Summarise the pattern with reference to the percentages in the table.
The results suggest there are age differences in the fear of crime. Those in middle age bands are more likely to report feeling safe both at home alone at night and walking alone after dark. Younger and older adults are more likely to report feeling unsafe. In particular, younger adults are more likely to report feeling unsafe when at home alone at night and older adults are most likely to report feeling unsafe walking alone after dark. For example, nearly 20 percent of those aged 75+ report feeling very unsafe walking alone after dark compared to just under 8 percent of those aged 24-44 and 45-64.

Question 16 Using the information above, can you plan the recode for the new variable identifying those who feel unsafe walking alone after dark?

Existing Variable = Walkdark			New Variable= unsafedark	
Value	Label		Value	Label
1	Very safe	}	0	No
2	Safe			
3	A bit unsafe	}	1	Yes
4	Very unsafe			
8	Refusal		8	Refusal
9	Don't know		9	Don't know

Question 17 Do feelings of safety walking alone after dark vary according to experience of violence?
Examining these variables, the relationship between experience and fear of crime does not appear to be strong. On average, about 32 percent of people very safe walking alone after dark; among those who have experienced violence it reduces to 28 percent. The difference is then smaller looking at those who feel fairly safe, which is 40 percent for those who have not experienced violence, compared to 41 percent for those who have.
Does there appear to be a statistically significant relationship between experience of violence and the fear of crime?
Pearson Chi-Square (0.888, p=0.82) suggests the relationship is non-significant.

Question 18 Among males, which of the following statements applies? Those who have recently experienced violence feel less safe walking alone after dark.
Is the pattern the same for females? No, among females, feeling unsafe walking alone after dark is not related to recent experience of violence.

Question 19 According to the Pearson Chi-square test, is there a significant association between fear and experience of crime...
...among males? Yes, at the 5 percent level of significance, the chi-square test indicates there is an association between fear and experience of crime among males
...among females? No, there does not appear to be a significant association among the fear and experience of crime among females.