2.1.2.1 Tutorial - Frequencies for nominal and ordinal variables [Updated 1 Oct 2010]

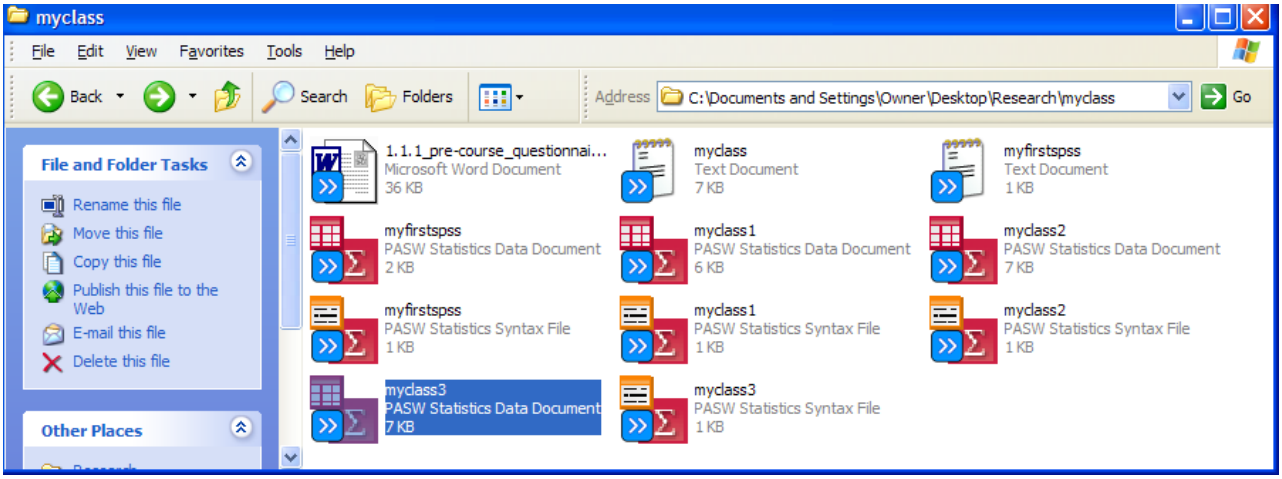
Exemplar: Pre-course survey of interests and experience.

Documentation: Questionnaire

File: myclass3.sav

If you do not already have these, download and save them now in a folder on your computer called **myclass** or copied to a floppy or CD. The files from this survey are small enough to go on a floppy in drive **a:** but files from other surveys are larger than 1.4 mb and will need to be kept in a folder on your computer, burned to a CD or stored on a remote server. If you don't have SPSS on your computer, you should still be able to follow the tutorial even if you can't do the exercises.

Your **myclass** folder should look like this if you have done all the exercises and downloaded the questionnaire. In this tutorial we only need the questionnaire and the SPSS saved file **myclass3.sav**



Using the SPSS saved file **myclass3.sav** we created as an exercise in session 1.3.3.7 (SPSS for real - my first saved file) and using the **Questionnaire** as user-documentation, let's have a look at a **nominal**[1] variable:

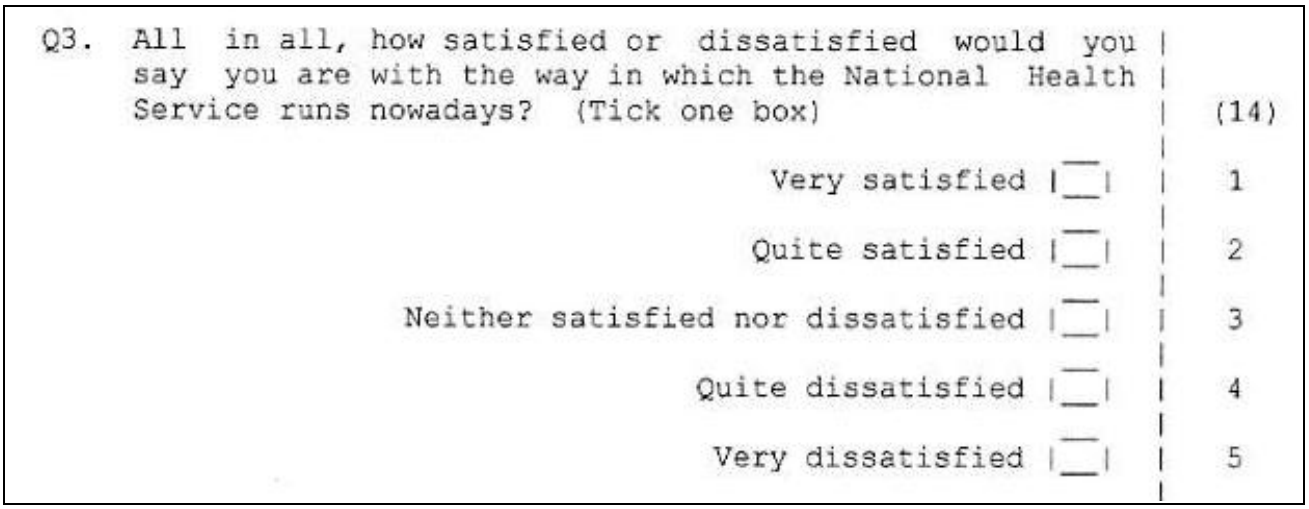
Main mode of travel (to campus)

```
Q6.  How will you be travelling to the Polytechnic?          |  (24)
                                                             |
     (Tick main mode only)          Public Transport  |__|  |   1
                                                             |
                                    Car               |__|  |   2
                                                             |
                                    M'cycle/Cycle     |__|  |   3
                                                             |
                                    Walking           |__|  |   4
```

[1] Nominal variables have distinct and exclusive categories with no inherent properties other than their names. They can be jumbled up into any order without losing any information.

... and an **ordinal**[2] variable:

Satisfaction with the National Health Service



There is only one line of data for each case: the data for question Q.5 are stored on column 24 and for Q.3 on column 14. We used **positional names**[3] for these when we read the data into SPSS to create the file, so the variables will be called **v24** and **v14**.

In my SPSS examples I use **UPPER CASE** for presenting general **FORMATS** and **lower case** for the actual **syntax** needed to run analyses. I also use the same colour-coding as SPSS/PASW 18 for **commands**, **sub-commands** and **keywords**. SPSS 15 and earlier versions do not have this facility, but I use it because, in addition to laying syntax out with longer indents and blank lines, it's an additional aid to learning and understanding.

FREQUENCIES

The **FREQUENCIES** command produces frequency tables, descriptive statistics and graphics for one or more variables.

General format: (You need to get used to this terminology: I use it a lot!)

```
FREQUENCIES <varlist>
    /STATISTICS <statistics list>
    / <graphics options>
    / <other options>
```

Within the **FREQUENCIES** procedure there are a number of sub-commands available for statistics, graphics and printed output. The most common options used are for **descriptive statistics** and for printing **charts**. Others allow you to change the order in which the categories appear in tables, suppress tables whilst printing statistics or charts etc.

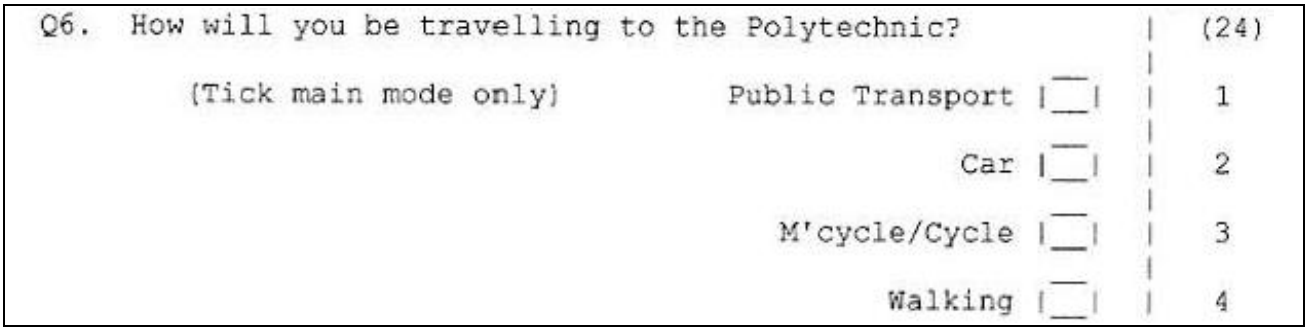
The simplest form of the command is:

```
FREQUENCIES  <varlist>
```

[2] Ordinal variables have distinct and exclusive categories, plus one additional property: they have an inherent order which can be reversed, but not jumbled up without losing important information.

[3] See 1.3.1 Conventions for Naming Variables in SPSS

Let's do the **nominal** variable first.



The SPSS command is:

frequencies v24 .

. . . which produces the following table:

v24 Q6: Main transport mode

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Public transport	115	68.0	68.0	68.0
	Car	26	15.4	15.4	83.4
	Motor cycle or cycle	15	8.9	8.9	92.3
	Walk	13	7.7	7.7	100.0
	Total	169	100.0	100.0	

The **Frequency** column contains the **count** (actual number) of cases in each category.

The **Percent** column displays these as a **percentage** (%) of the **whole sample** (base N = 169)

The **Valid Percent** column displays them as a percentage of the **valid** (non-missing) **cases** (there aren't any so N = 169 again).

The **Cumulative Percent** column displays them as the **percentage** so far included in the table, starting with (0% before) the first row and ending with 100% in the last, by adding the % for each successive category to the % for the previous category (base N = 169).

Cumulative Percent can be useful for deciding on cutting points for grouping interval or ordinal scale variables into fewer categories. It seems pretty pointless for nominal variables, but it's been there since 1968 so perhaps it's sacred.

The above table is in the **default** output format with rows in **ascending** order of the values used for coding the replies (**/FORMAT AVALUE**) but you can also have **descending** order (**/FORMAT DVALUE**). It can be edited to get rid of unwanted columns by manipulating it using **pivot** tables in the output, but that just complicates things and is a bit advanced for beginners.

You can change the order in which the categories appear, but there is no inherent meaning to the coded values. However, with nominal variables iit is sometimes useful to present tables in which the categories are arranged in ascending or descending order of the number of cases. This table can also be presented in ascending or descending order of size of category (**/FORMAT AFREQ** or **DFREQ**).

This would normally be in **descending** order of number of cases in the category, so:

```
frequencies v24
        /format dfreq .
```

v24 Q6: Main transport mode

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Public transport	115	68.0	68.0	68.0
	Car	26	15.4	15.4	83.4
	Motor cycle or cycle	15	8.9	8.9	92.3
	Walk	13	7.7	7.7	100.0
	Total	169	100.0	100.0	

Graphics and statistics options

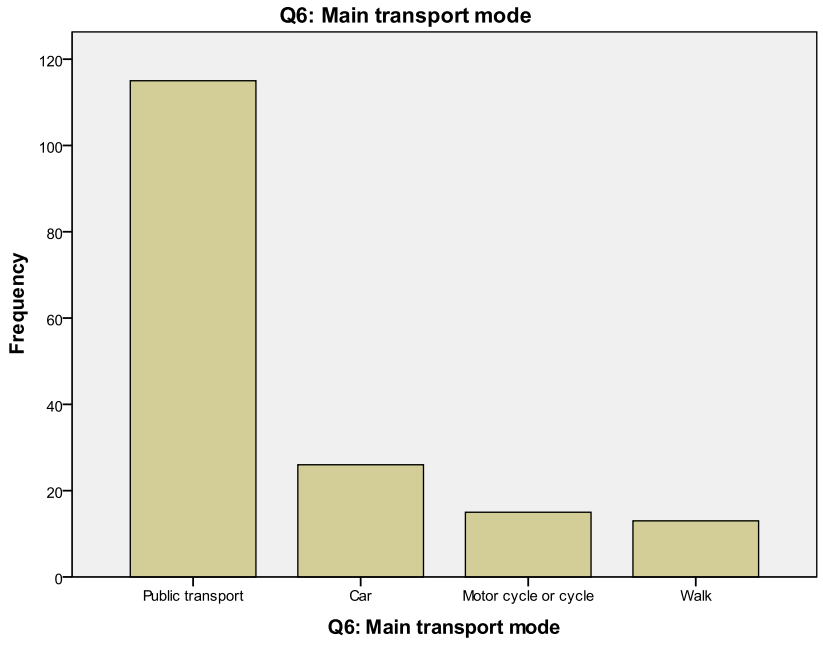
As well as frequency tables you can also specify appropriate graphic displays and descriptive statistics. If you only want the chart or the statistics, you can suppress the table with:

~ ~ ~ **/FORMAT NOTABLE**.

For nominal variables, the only legitimate charts are a **barchart** (**/BARCHART**) or a **piechart** (**/PIECHART**) Barcharts are used when there is no intrinsic order to the categories or when there is an order, but you don't know what the assumed underlying metric is (and therefore how wide the bars should be) so the bars **must never touch**.

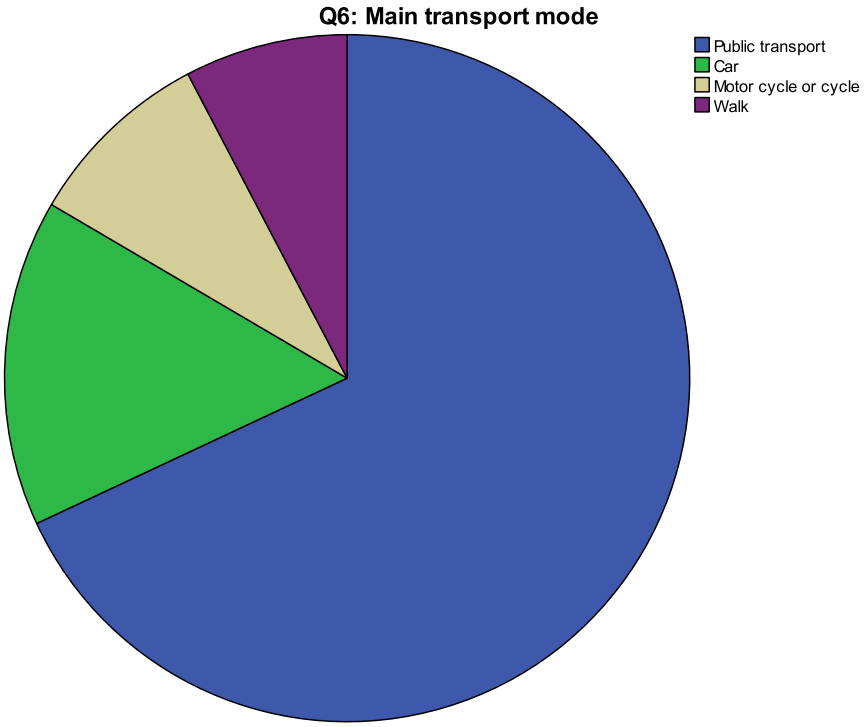
To get a **barchart** without the table:

```
frequencies v24
        /barchart
        /format notable .
```



To get a pie chart (without the table):

```
frequencies v24
        /piechart
        /format notable .
```



[**NB:** A piechart is normally only used for the whole sample: make sure you have plenty of ink!]

The only permissible **statistic** for nominal variables is the **mode** (the most commonly occurring value) but you can also legitimately request the lowest and highest code values (**/STATISTICS** = followed by one or more of **MINIMUM**, **MAXIMUM**, or **MODE**). You never know if there are values way outside the expected range!

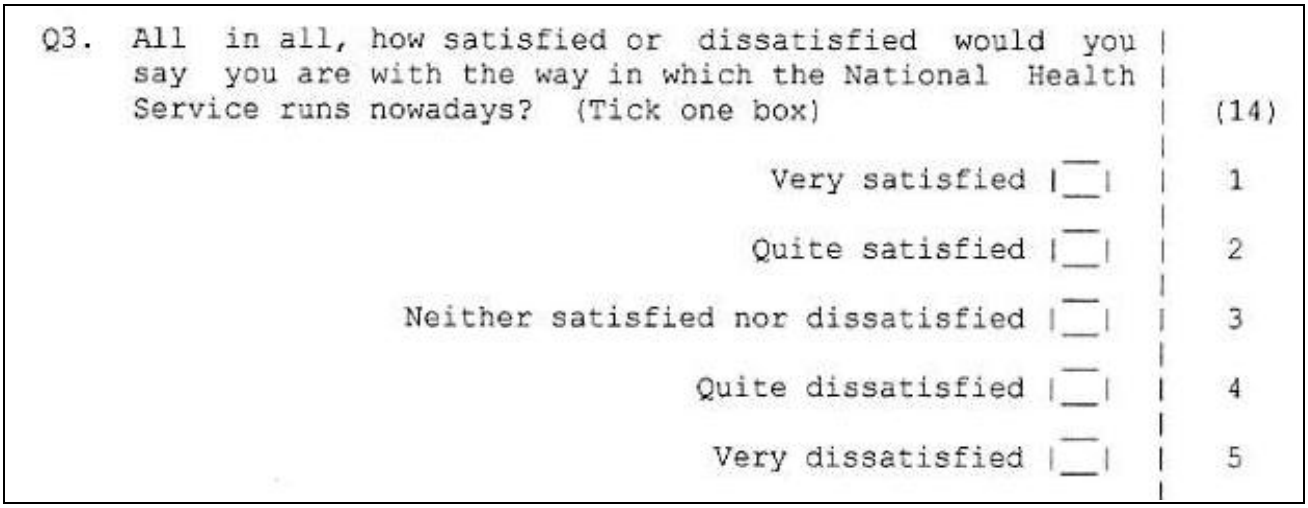
```
frequencies v24
        /statistics minimum maximum mode
        /format notable .
```

Statistics

v24 Q6: Main transport mode

N	Valid	169
	Missing	0
Mode		1
Minimum		1
Maximum		4

Now let's have a look at the **ordinal** variable



Q3. All in all, how satisfied or dissatisfied would you say you are with the way in which the National Health Service runs nowadays? (Tick one box) (14)

Very satisfied __ 1

Quite satisfied __ 2

Neither satisfied nor dissatisfied __ 3

Quite dissatisfied __ 4

Very dissatisfied __ 5

The SPSS command is:

frequencies v14 .

v14 Q3: Satisfaction with running of NHS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very satisfied.	2	1.2	1.2	1.2
	Quite satisfied.	13	7.7	7.8	9.0
	Neither	27	16.0	16.3	25.3
	Quite dissatisfied	64	37.9	38.6	63.9
	Very dissatisfied	60	35.5	36.1	100.0
	Total	166	98.2	100.0	
Missing	System	3	1.8		
Total		169	100.0		

The **Frequency** column contains the **count** (actual number) of cases in each category.

The **Percent** column displays these as a **percentage** (%) of the **whole sample** (base N = 169),

The **Valid Percent** column displays them as a percentage of the **valid** (non-missing) **cases** (N = 166)

The **Cumulative Percent** column displays them as the **percentage** so far included in the table, starting with (0% before) the first row and ending with 100% in the last, by adding the % for each successive category to the % for the previous category (base N = 166).

This is a case where **Cumulative Percent** can be useful for deciding on cutting points for grouping ordinal (and also interval) scale variables into fewer categories.

The default is **ascending** order of numeric values (here values 1 to 5) used for coding the replies (**/FORMAT AVALUE**), but you can also have **descending** order (**/FORMAT DVALUE**).

```
frequencies v14
        /format dvalue .
```

v14 Q3: Satisfaction with running of NHS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very dissatisfied	60	35.5	36.1	36.1
	Quite dissatisfied	64	37.9	38.6	74.7
	Neither	27	16.0	16.3	91.0
	Quite satisfied.	13	7.7	7.8	98.8
	Very satisfied.	2	1.2	1.2	100.0
	Total	166	98.2	100.0	
Missing	System	3	1.8		
Total		169	100.0		

Tables can also be presented in ascending or descending order of size of category (**/FORMAT AFREQ** or **DFREQ**) but this is pointless for ordinal variables as it will jumble up categories for which the order needs to be retained.

Because ordinal variables do not have a known and fixed metric, there is no way of knowing where any coded value should appear on the assumed underlying metric, or what its limits are. The only legitimate chart you can use is a barchart and the only legitimate statistics you can calculate are the cutting points to divide the sample into groups.

To find out where the centre of the distribution is we can't calculate an arithmetic mean but we can find the **median** (the point at which the sample can be divided into two equal sized groups: **/MEDIAN**) or any other cutting points. These latter are known as **percentiles** (**/PERCENTILES**) with subcategories of **quartiles** and **deciles**, but these are best kept for interval and ratio scales.

With ordinal scales the cutting points are likely to land **on** an actual value, not **between** two values, so it's better to choose cutting points by inspecting the cumulative frequencies.

Abbreviated syntax

As we have already seen, SPSS is case insensitive for syntax, and I prefer to work with abbreviated syntax in lower case as this not only saves time, but is much easier on the eye.

However, if you start typing direct into the syntax window, SPSS prompts with menu suggestions in UPPER CASE. If you click on these, the full command appears in UPPER CASE in colour-coded text. If the command is incomplete or SPSS can't make sense of the syntax, the text will appear in **red**. If you use abbreviated syntax, all or part of the command may not be colour-coded. The examples above follow the colour-coding of SPSS and, for now, I've typed them out in full. Variable names always have to be written in full.

You may also find it useful to put a space between commands to keep them visually separate, and also before the full stops as this makes them easier to check if they are present at the end of each command (or not, a common source of errors).

Once you get used to writing SPSS syntax, you can begin to use the abbreviated forms. SPSS only reads the first 3 or 4 characters of commands, sub-commands and keywords anyway, so the following will work just as well, but not everything will be colour-coded in the syntax window.

Thus the following examples will also work.

```
freq v24 .

freq v14 v24 /bar .

freq v24 /sta min max med .

freq v14 v24 /for not /bar .

freq v24 /for not /pie .
```

End of tutorial.

Now repeat the above analysis yourself by working through the following exercise 2.1.2.2.

Next session: 2.1.2.2 Exercise - Frequencies for nominal and ordinal variables

[Back to Block 2 menu]