

# Using a computer to perform statistical analysis

*In this article, using the clinical research question: 'what factors seem to predict who will take an overdose?', Denis Anthony shows how a statistical computer package can make research easier and more efficient*

## key words

- ▶ statistics
- ▶ research methods
- ▶ outcomes and prognoses
- ▶ information technology

## Introduction

This paper shows how to use a statistical computer package for research purposes, using a clinical research example for illustration, namely: 'what factors seem to predict who will take an overdose?'

It is clearly not possible to show all types of statistical analysis for even one package, and you will want to look at recent texts on your preferred package (there are many for SPSS alone). My particular favourite is *Quantitative Data Analysis with SPSS* (Bryman 2001), which, while aimed at social science students, is understandable by and relevant to health care students. Other good recent books include *SPSS Survival Manual* (Pallant 2001), and I have a book of my own (Anthony 1999). However, you should look at a few books in your library, and use one that is written in language you can comprehend, and at an appropriate level and pace. In my own library, I located 74 texts having SPSS as a word in the title. If you fail to understand the text of the first one you select, do not struggle; just pick an alternative that you can understand. Statistics are notoriously difficult to teach, and strategies that work with one student do not necessarily work with others. I tend to teach using clinical examples, as this seems to work better with healthcare students. I do not show

students the formulae behind the statistics, as they are most unlikely to need them, but concentrate instead on setting up the computer to produce the right analysis, and on how to interpret the results.

### Pros and cons of statistical software

There are two ways to conduct statistical analysis: manually and using a computer. The former is simple for very small problems, but becomes increasingly impractical as the data grow larger or the analysis more complex. However, using software has its own drawbacks. The advantages and disadvantages of using statistical software are shown in the Table 1.

**Table 1. Advantages and disadvantages of using statistical computer software**

Problem	Manual (pen and paper) statistics	Statistical software
Inputting data	You will want to have data in an optimal form for conducting analysis. This may mean having several forms of the data on paper	Input data only once
Statistical formulae	You need to know these, and be able to manipulate them	You only need to know the tests you want to conduct, not how to compute them
Errors in calculation	You may not know you have made an error, and if you do you will need to recalculate	The software should not make errors of this type
New data, or input data corrected	You will need to recalculate	The computer will recalculate when you ask for the analysis again. You can save analyses into a special file and simply call it up
Inappropriate test used	You can use the wrong test, but if you know statistics well enough to compute manually, it is likely you will know when to use a given test	You can ask the computer to make almost any test on almost any data. The software is in general not able to detect that you are doing the wrong thing
Many different statistical approaches needed on data	You will need to recalculate for each test, or plot each graph or produce each table individually	You can ask the software to conduct many tests, plot graphs and produce tables once the data are input.

## What software?

You can write your own software in many computer languages. I have written statistical programs in BBC Basic, C, C + +, Pascal, and dBase II. However this is time consuming, tedious, and unless your needs are highly specialised, almost certainly not the optimal method.

You can use many packages for statistical analysis. For example, many spreadsheets such as Microsoft Excel have statistical functions, and it is perfectly possible to conduct simple statistics and produce tables and graphs from Microsoft Excel and the open source spreadsheets. This may be cheap, as you probably have such a product already. However, if you intend to do a lot of analysis, you will get frustrated, as these packages are not designed with statistical analysis in mind. Specialised mathematical packages such as Matlab and Mathematica may also be used, and do have statistical and graphical capabilities, but are better suited to engineering and physics than to the sort of statistical analyses typically employed in health care.

If you are engaged in a lot of analysis, it is well worth investing the money and effort to purchase and learn to use a statistical package. You may find that your university, NHS trust, or employer has a licence for such a package, making it a free option for you.

There are dozens of statistics packages, some of them highly specialised. For example, if your work is in epidemiology you may prefer to use a package specifically designed to conduct epidemiological analysis, such as epi-info. However, for general purpose statistics, including those required by the overwhelming majority of academic nursing papers, a generic statistical package is ideal. Examples include SPSS (Statistical Package for the Social Sciences), SAS, Minitab, S, and R. These all run on several operating systems, and it should be possible to get a version to run on your computer.

Open source software is that which is freely available (though not necessarily free), and others are able to adapt such software for their own purposes (with acknowledgement). If you are interested in open source products, R, which is loosely based on a commercial package called S, will run on both Windows (Windows 95 and later) MacOS (System 8.6 to 9.1), MacOS X (Darwin/X11) and Linux. The R site, which allows free downloads, is at <http://www.r-project.org/>

	gender	psychhis	prevod	employ	assessor	repeat	maitsta	age	followup	var	var	var
81	1	2	2	2	1	2	2	24	2			
82	2	2	1	1	2	2	2	18	1			
83	1	2	2	4	2	2	2	20	2			
84	2	2	1	2	2	2	2	22	1			
85	2	1	1	2	2	1	2	23	1			
86	2	1	1	2	2	1	2	23	1			
87	2	1	1	2	1	1	2	23	1			
88	2	1	1	2	2	1	2	23	1			
89	1	1	1	2	2	2	2	60	1			
90	1	2	1	2	2	1	3	44	1			
91	1	1	1	2	2	1	3	43	1			
92	1	1	1	2	2	1	3	44	1			
93	1	1	1	2	1	1	3	43	1			
94	2	2	2	1	1	2	1	27	1			
95	1	2	2	2	1	1	2	17	1			
96	1	1	1	2	1	1	2	18	1			
97	1	2	2	1	2	2	3	40	1			
98	2	2	2	2	2	2	3	42	2			
99	2	1	1	1	1	1	1	32	2			
100	2	2	1	1	1	1	2	27	1			
101	2	1	2	2	2	.	1	27	1			
102	2	2	2	2	2	2	2	18	2			
103	1	2	2	1	2	2	1	41	2			
104	2	2	2	2	2	2	1	38	1			
105	1	2	2	1	2	2	1	17	2			
106	1	2	2	2	2	2	3	34	2			
107	1	2	1	2	1	.	1	34	1			
108	1	2	2	2	2	2	2	56	1			
109	1	3	3	5	2	2	2	18	2			

**Figure 1. Data entry**

A well-known package, and in my experience the most commonly used in UK universities and NHS trusts, is SPSS. It runs on Microsoft Windows, Macintosh computers, and UNIX (including the open source microcomputer version Linux). I will use SPSS in this article to demonstrate how a package makes statistical analysis simpler, and to illustrate I use data collected by a former student of mine, Jenny Cook.

## Data

Jenny collected data on people referred to her clinic who had taken an overdose, having previously been admitted to the accident and emergency department. Her clinic assessed patients and followed some of them up.

## Data entry

You cannot do anything until you have entered the data. In SPSS, this is done via a spreadsheet, in which rows are subjects (in Jenny's study, people who have taken an overdose) and columns are variables or attributes of the subject.

	gender	psychhis	prevod	employ	assessor	repeat	maitsta	age	followup	var	var	var
81	male	no	no	employed	doctor	no	single	24	no			
82	female	no	yes	employed	doctor	no	single	18	yes			
83	male	no	no	student	doctor	no	single	20	no			
84	female	no	yes	unemploy	doctor	no	single	22	yes			
85	female	yes	yes	unemploy	doctor	yes	single	23	yes			
86	female	yes	yes	unemploy	doctor	yes	single	23	yes			
87	female	yes	yes	unemploy	nurse	yes	single	23	yes			
88	female	yes	yes	unemploy	doctor	yes	single	23	yes			
89	male	yes	yes	unemploy	doctor	no	single	60	yes			
90	male	no	yes	unemploy	doctor	yes	div/sep	44	yes			
91	male	yes	yes	unemploy	doctor	yes	div/sep	43	yes			
92	male	yes	yes	unemploy	doctor	yes	div/sep	44	yes			
93	male	yes	yes	unemploy	nurse	yes	div/sep	43	yes			
94	female	no	no	employed	nurse	no	marr/coha	27	yes			
95	male	no	no	unemploy	nurse	yes	single	17	yes			
96	male	yes	yes	unemploy	nurse	yes	single	18	yes			
97	male	no	no	employed	doctor	no	div/sep	40	yes			
98	female	no	no	unemploy	doctor	no	single	42	no			
99	female	yes	yes	employed	nurse	yes	marr/coha	32	no			
100	female	no	yes	employed	nurse	yes	single	27	yes			
101	female	yes	no	unemploy	doctor	.	marr/coha	27	yes			
102	female	no	no	unemploy	doctor	no	single	18	no			
103	male	no	no	employed	doctor	no	marr/coha	41	no			
104	female	no	no	unemploy	doctor	no	marr/coha	38	yes			
105	male	no	no	employed	doctor	no	marr/coha	17	no			
106	male	no	no	unemploy	doctor	no	div/sep	34	no			
107	male	no	yes	unemploy	nurse	.	marr/coha	34	yes			
108	male	no	no	unemploy	doctor	no	single	56	yes			
109	male	not record	not record	not record	doctor	no	single	18	no			

**Figure 2: Value labels of input numbers**

In this case, gender is one of many variables collected on each subject. You can see this in the first column in Figure 1, where we coded gender as either 1 or 2. Note all the other variables have been given numerical values. (Figure 1 shows only part of the complete spreadsheet).

To see what the numbers refer to we can ask SPSS to show the value labels of the input numbers, as shown in Figure 2 (I do not show you how to do this here, but it is simple to give labels to numbers in SPSS; for gender I have given the label male to the number 1, and female to 2). Note that age is still a number, because it is the number of years the person has been alive. Thus age is different from all the other variables that are nominal (sometimes called categorical) data, meaning you can give them a number, but it is purely a label. For age, though, the number has meaning. So while female (2) is not twice male (1), an age of 40 is twice as old as an age of 20.

### **Descriptive analysis: frequencies and cross tabulation**

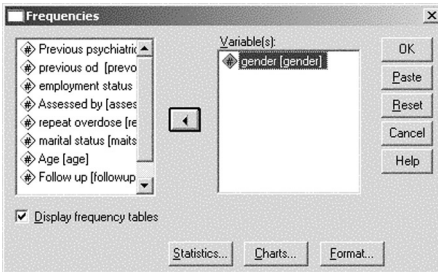
Now suppose we want to perform a very simple analysis: we want to see

gender	gender	employment	assessor	repeat	marriage	age	followup	var	var	
8	81	ma	no	employed	doctor	no	single	24	no	
8	82	fema	yes	employed	doctor	no	single	18	yes	
8	83	ma	no	student	doctor	no	single	20	no	
8	84	fema	yes	unemploy	doctor	no	single	22	yes	
8	85	fema	yes	unemploy	doctor	yes	single	23	yes	
8	86	fema	yes	unemploy	doctor	yes	single	23	yes	
8	87	female	yes	yes	unemploy	nurse	yes	single	23	yes
8	88	female	yes	yes	unemploy	doctor	yes	single	23	yes
8	89	male	yes	yes	unemploy	doctor	no	single	60	yes
9	90	male	no	yes	unemploy	doctor	yes	div/sep	44	yes
9	91	male	yes	yes	unemploy	doctor	yes	div/sep	43	yes
9	92	male	yes	yes	unemploy	doctor	yes	div/sep	44	yes
9	93	male	yes	yes	unemploy	nurse	yes	div/sep	43	yes
9	94	female	no	no	employed	nurse	no	marr/coha	27	yes
9	95	male	no	no	unemploy	nurse	yes	single	17	yes
9	96	male	yes	yes	unemploy	nurse	yes	single	18	yes
9	97	male	no	no	employed	doctor	no	div/sep	40	yes
9	98	female	no	no	unemploy	doctor	no	div/sep	42	no
9	99	female	yes	yes	employed	nurse	yes	marr/coha	32	no
10	100	female	no	yes	employed	nurse	yes	single	27	yes
10	101	female	yes	no	unemploy	doctor	no	marr/coha	27	yes
10	102	female	no	no	unemploy	doctor	no	single	18	no
10	103	male	no	no	employed	doctor	no	marr/coha	41	no
10	104	female	no	no	unemploy	doctor	no	marr/coha	38	yes
10	105	male	no	no	employed	doctor	no	marr/coha	17	no
10	106	male	no	no	unemploy	doctor	no	div/sep	34	no
10	107	male	no	yes	unemploy	nurse	no	marr/coha	34	yes
10	108	male	no	no	unemploy	doctor	no	single	56	yes
10	109	male	not record	not record	not record	doctor	no	single	18	no

**Figure 3: Performing simple analyses**

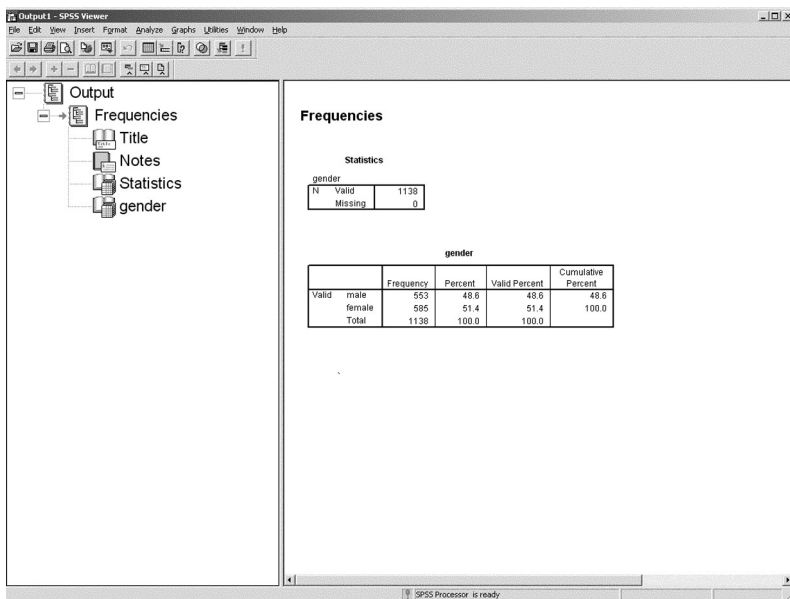
whether it is women or men who are most prominent in our sample. I can ask for this using the drop-down menu 'Analyze'. In the 'analyze' menu, one form of analysis we can select is 'Descriptive Statistics', and one method of descriptive analysis is 'Frequencies', as shown in Figure 3.

I then get the dialogue box shown in Figure 4, and ask for 'gender' to be the variable for frequencies by clicking on gender and then clicking on the arrow button to take it into the variables box:



**Figure 4: Frequencies**

Pressing OK, I get the output as shown in Figure 5:



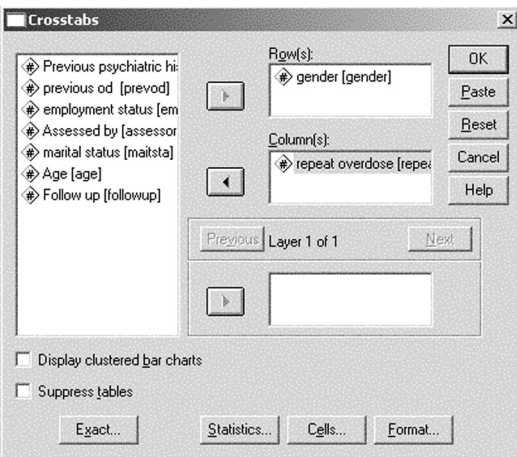
**Figure 5: Output for frequencies**

Note I am now given a new window, the Output window. SPSS allows several windows, one for input data, and another for displaying results (the Output window). Later we will see another one. The Output window shows me that marginally more women than men are in the sample. What I now want to know is, are males or females the ones most likely to take another overdose? I have a variable called repeat which is either yes (coded as 1) or no (coded as 0). These two variables, gender and repeat can be used to perform a 'cross tabulation', which means a table showing all possible values of both variables. There are only two values for each here, one can be male or female, and one can be repeat overdose or not repeat.

	gender	employment	assessor	repeat	marital status	age	followup	VAR	VAR	VAR
81	ma	employed	doctor	no	single	24	no			
82	fema	yes	employed	doctor	no	single	18	yes		
83	ma	no	student	doctor	no	single	20	no		
84	fema	yes	unemploy	doctor	no	single	22	yes		
85	fema	yes	unemploy	doctor	yes	single	23	yes		
86	fema	yes	unemploy	doctor	yes	single	23	yes		
87	female	yes	yes	unemploy	nurse	yes	single	23	yes	
88	female	yes	yes	unemploy	doctor	yes	single	23	yes	
89	male	yes	yes	unemploy	doctor	no	single	60	yes	
90	male	no	yes	unemploy	doctor	yes	div/sep	44	yes	
91	male	yes	yes	unemploy	doctor	yes	div/sep	43	yes	
92	male	yes	yes	unemploy	doctor	yes	div/sep	44	yes	
93	male	yes	yes	unemploy	nurse	yes	div/sep	43	yes	
94	female	no	no	employed	nurse	no	marr/coha	27	yes	
95	male	no	no	unemploy	nurse	yes	single	17	yes	
96	male	yes	yes	unemploy	nurse	yes	single	18	yes	
97	male	no	no	employed	doctor	no	div/sep	40	yes	
98	female	no	no	unemploy	doctor	no	div/sep	42	no	
99	female	yes	yes	employed	nurse	yes	marr/coha	32	no	
100	female	no	yes	employed	nurse	yes	single	27	yes	
101	female	yes	no	unemploy	doctor	no	marr/coha	27	yes	
102	female	no	no	unemploy	doctor	no	single	18	no	
103	male	no	no	employed	doctor	no	marr/coha	41	no	
104	female	no	no	unemploy	doctor	no	marr/coha	38	yes	
105	male	no	no	employed	doctor	no	marr/coha	17	no	
106	male	no	no	unemploy	doctor	no	div/sep	34	no	
107	male	no	yes	unemploy	nurse	no	marr/coha	34	yes	
108	male	no	no	unemploy	doctor	no	single	56	yes	
109	male	not record	not record	not record	doctor	no	single	18	no	

**Figure 6: Selecting cross tabulation**

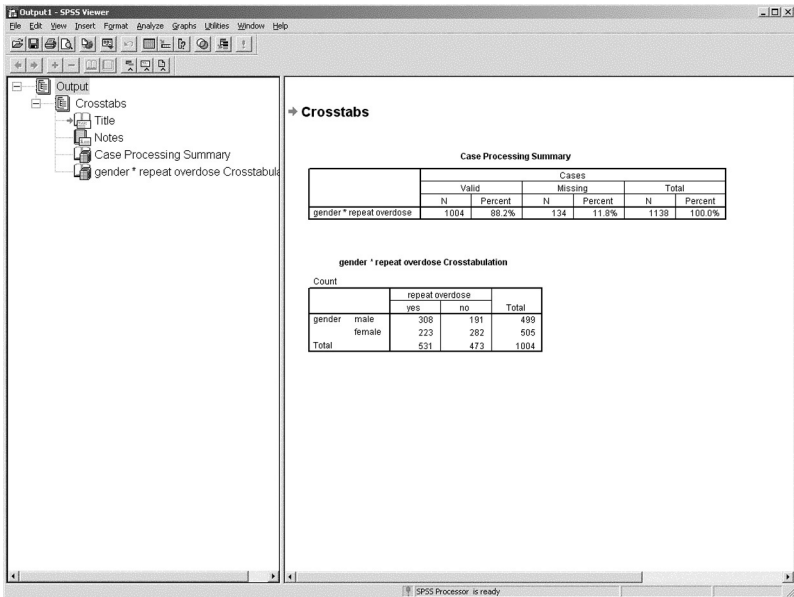
In the subsequent dialogue box I now ask for the two variables of interest (Figure 7):



**Figure 7: Selecting variables for cross tabulation**



If I press the OK button I get the following cross tabulation result:



**Figure 8: Cross tabulation result for 'gender' by 'repeat'**

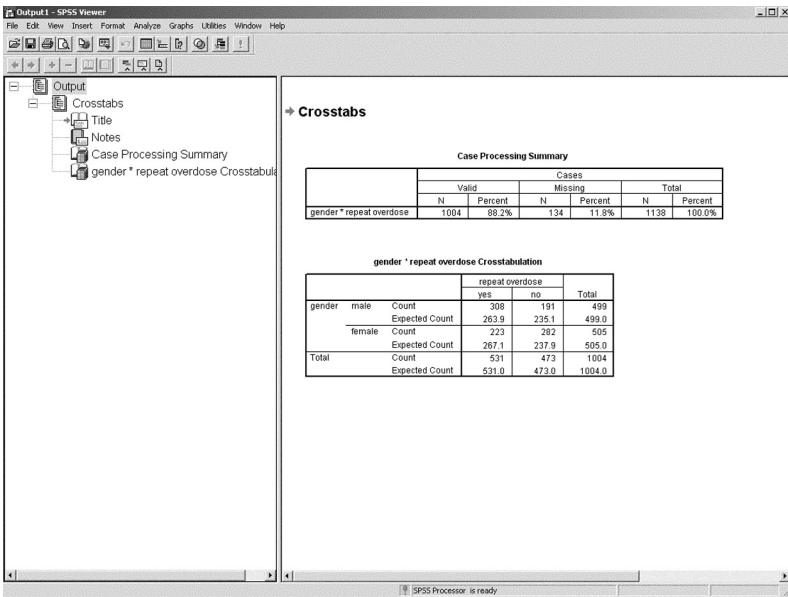
### Inferential statistics: Chi square

Inferential statistics test a hypothesis. Here I make the hypothesis that repeat overdose is independent of gender. I do not necessarily believe this, but it acts as a statement I can test with statistics.

From the cross tabulation result shown in Figure 8 it looks as though more males take repeat overdoses than females. Is this what I would have expected? Well, since slightly fewer men are in the sample I would have expected fewer also to repeat. Even if I accept that men and women are roughly equal in number in the sample (which happens to be true here), I would expect half of the repeaters to be male, and half female, or half of 531 (total in first column) or 265.5, which I round down to 265 (people either overdose or they do not, so 0.5 is meaningless here).

I can get SPSS to work out how many one would expect, if gender had nothing to do with (was independent of) repeating overdose (Figure 9).

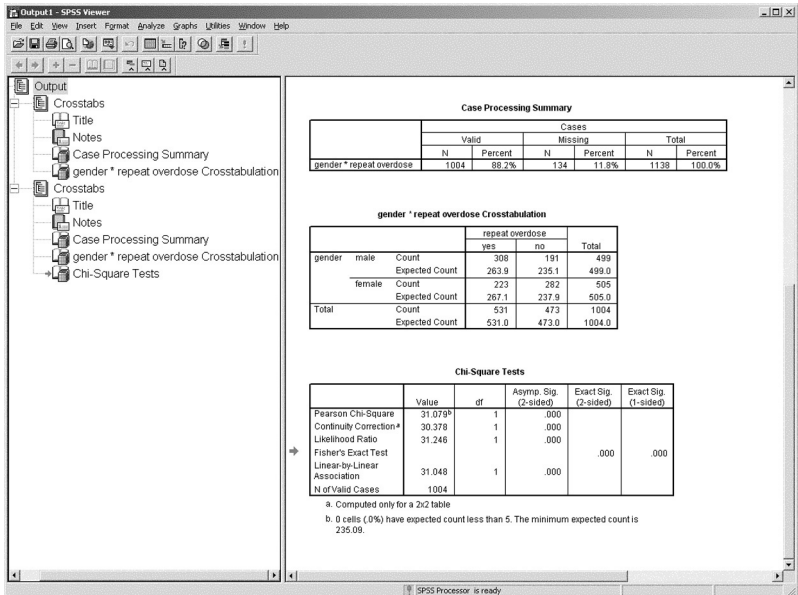
Indeed, we can see there is a difference between the actual count for male repeaters (308) and what we would have expected (263.9, or rounded to 264, my rough calculation is made more exact by SPSS).



**Figure 9: Expected frequencies**

But could these data have occurred by chance, it just so happening that my sample had a lot of men who repeated overdose? The answer is always yes, though the probability may be very small. Imagine if all the men took another overdose and none of the women did, it is still possible that we just had a very strange sample; however this would be very unlikely in a large sample.

I can get SPSS to work out the probability using a test called chi square. Chi square looks at the difference between actual and expected frequencies of two nominal variables, and returns the probability of it happening by chance alone. It gives higher values to bigger samples with the same proportions. Clearly, a very small sample of (say) ten people could more easily have all the men repeat and none of the women than a big sample of (say) 10,000.



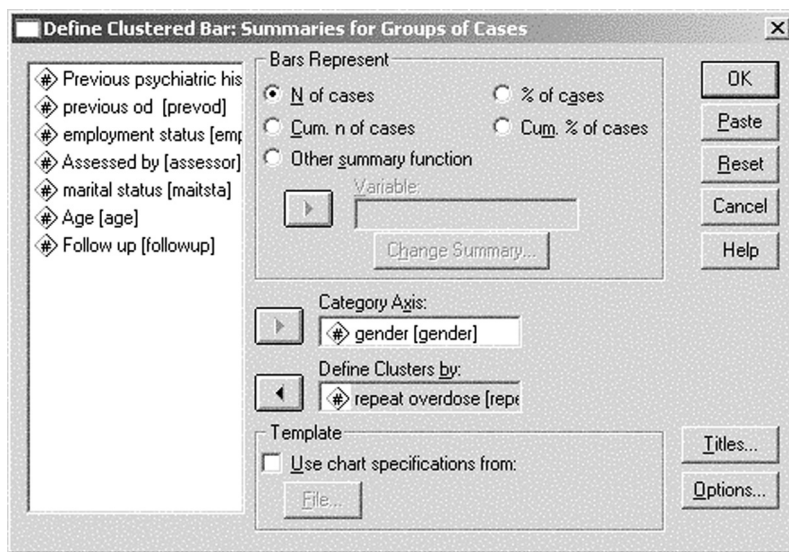
**Figure 10: Chi square test**

SPSS gives several results, but in this case, we are interested in the first line, which is the standard chi square, also called Pearson chi square (other values are given which are used in unusual circumstances, we will not worry about these here). We see above that Pearson chi square Asymp. sig (the significance or probability, usually called p value) is 0.000. P values are always between 0 and 1.0. It is customary to consider a test with a p value of less than 0.05 or 5 per cent to be significant; higher values are assumed to be explicable by chance alone. Exactly zero is clearly impossible, as this says there is no possibility that the data could have occurred by chance. However unlikely it may be, there is always some possibility that data occurred by chance. What SPSS is really saying is that the probability is less than 0.001 (in fact it has truncated the real p value of 0.00000002, which can be found in SPSS by asking for more decimal places to be shown, and left the first three decimal places 0.000). A very common error in papers I review is that the authors have quoted a p value of 0.000, showing they have used a statistics package, but do not understand one of the basics of statistics.

We conclude that while in our population males and females are roughly equal in number, it is males who are more likely to repeat overdose.

### Graphs: bar charts

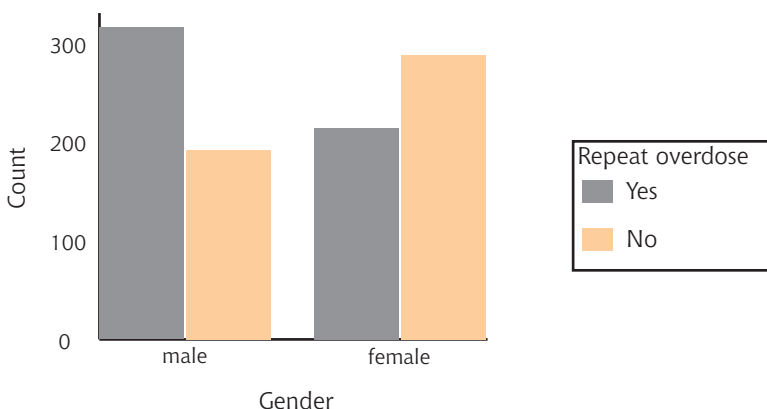
Now suppose I want to show this to my chief executive (CE). As she is an intelligent person who knows nothing about statistics, I might prefer to show her a graph. Here I ask for a clustered bar chart by choosing the 'Graphs' pull-down menu and then selecting 'Bar', and obtain the screen shown in Figure 11. I then put in the two variables.



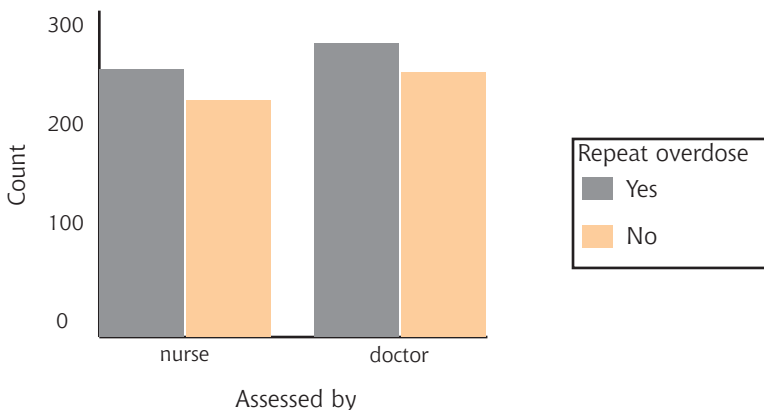
**Figure 11: Producing bar charts**

The resultant bar chart shows the relationship clearly, with more men repeating than not, and fewer women repeating than not.

The CE wants to know whether assessment by nurses is as effective as assessment by doctors, and also whether the money spent on follow-up is effective. I do similar graphs for her to show the answers to these questions (Figures 13 and 15):



**Figure 12: Bar chart for gender by repeat overdose**



**Figure 13: Bar chart for assessor by repeat overdose**

It looks as though the proportions are similar for doctors and nurses with respect to whether a patient has a further overdose. There is a slight difference in observed and expected frequencies, where nurses have about two more patients going on to overdose again than might be expected, and doctors about two fewer. However I can reassure the CE that the statistical test shows this small difference between doctors and nurses is not significant (chi square gives a p value of 0.776) and that the differences are obtainable by chance 77.6 per cent of the time, i.e. much more than half of the time, as shown in Figure 14.

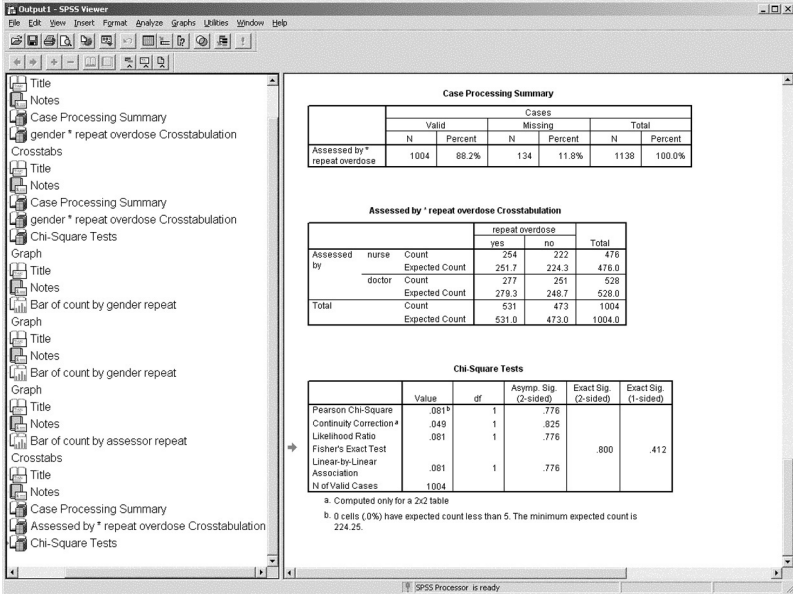


Figure 14: Chi square for assessor by repeat overdose

However, when I do the chart for follow up, the CE is very bothered. Can you see why, from the bar chart in Figure 15?

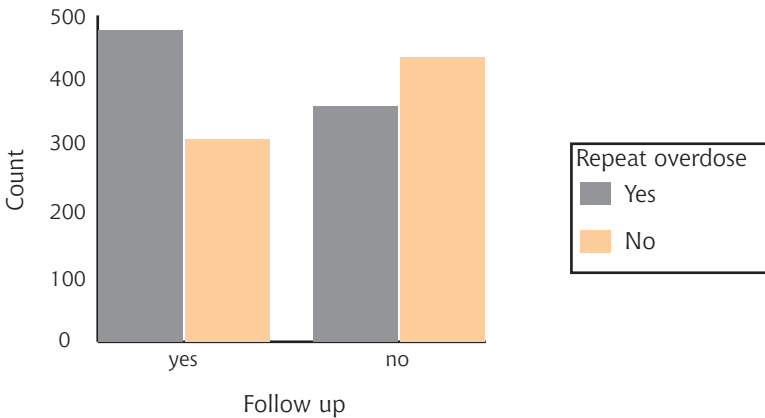
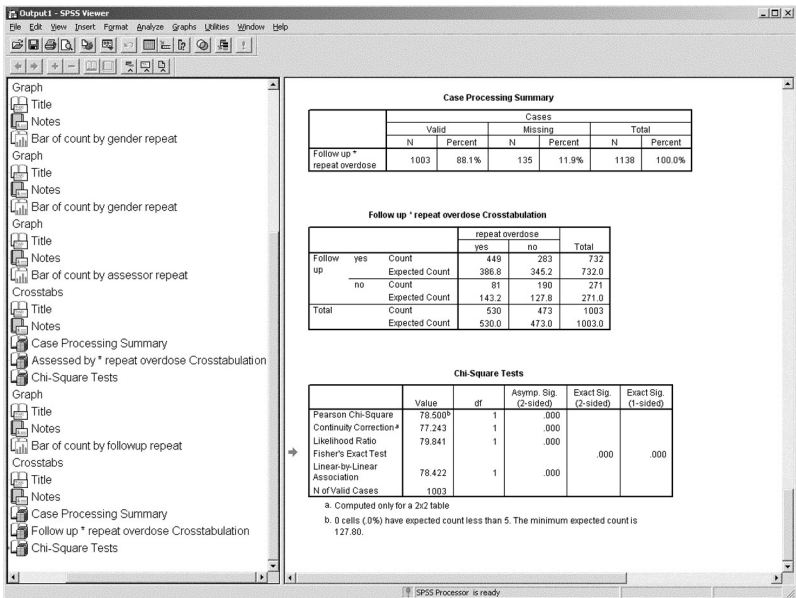


Figure 15: Bar chart for follow-up by repeat overdose

It looks as though followed-up patients have more overdoses! I do the chi square statistic, wondering if this could be explicable by chance, though not really believing this could be true as the difference is very marked. I get the results shown in Figure 16:



**Figure 16: Chi square for follow up by repeat overdose**

This shows that the p value (shown as 0.000) is less than 0.001, meaning this could occur by chance 0.1 per cent of the time, so I am 99.9 per cent sure the data did not occur by chance. Before the CE decides to ban follow up as a dangerous practice that causes patients to overdose again, let us stop and consider.

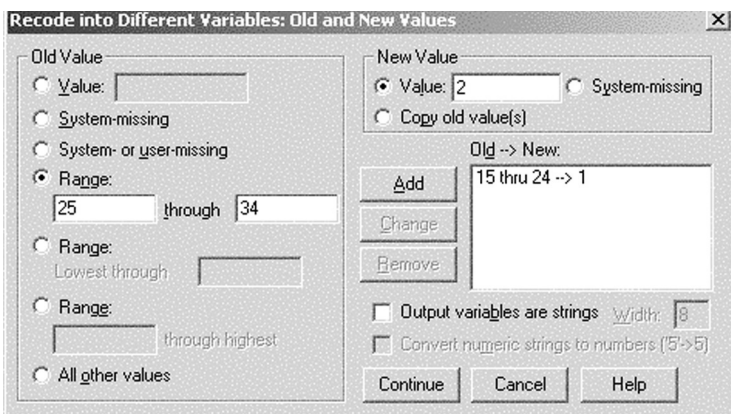
Were the patients allocated randomly to a follow up or not follow up regime? If they had been, then these results may well make us question the utility of follow up. However, this is not what happened. We just collected data from the normal clinical practice. What type of patient would be followed up? If the doctor or nurse thought someone was low risk, he or she may decide it was not necessary, but he or she probably will follow up someone about whom they have concern. Thus, all we may be showing is that clinicians are good at

spotting patients at risk, and while they are seemingly not always able to stop them repeating overdose, it is not necessarily the case that follow-up causes repeat overdose. Thus, one must interpret the statistics with care. Cause and effect are very difficult to establish where the study design is not randomised.

Could we make similar comments about gender? Well only if in the course of the study gender could be changed, which is unlikely! However, we might find that the effect of gender (males having more repeat overdoses) is not to do with gender as such, but with some other factor associated with gender. Possibly culturally determined gender roles make it more difficult to stop males repeating overdoses. It is not possible to say from this study. However, in a further study one could collect new data to explore such a relationship. Sometimes, therefore, statistics suggest further questions that may be answered, possibly in other ways. For example, having found males do repeat overdoses more than females, perhaps we could interview a sample of males and females to find out from them why they take overdoses.

### Re-coding and computing

Suppose you wanted to compare this study with another that collected age in bands, say 15-24, 25-34 and so on. You can save yourself the effort of recoding with SPSS's sophisticated 'recode' function, as shown in Figure 17.



**Figure 17: Recoding variables**

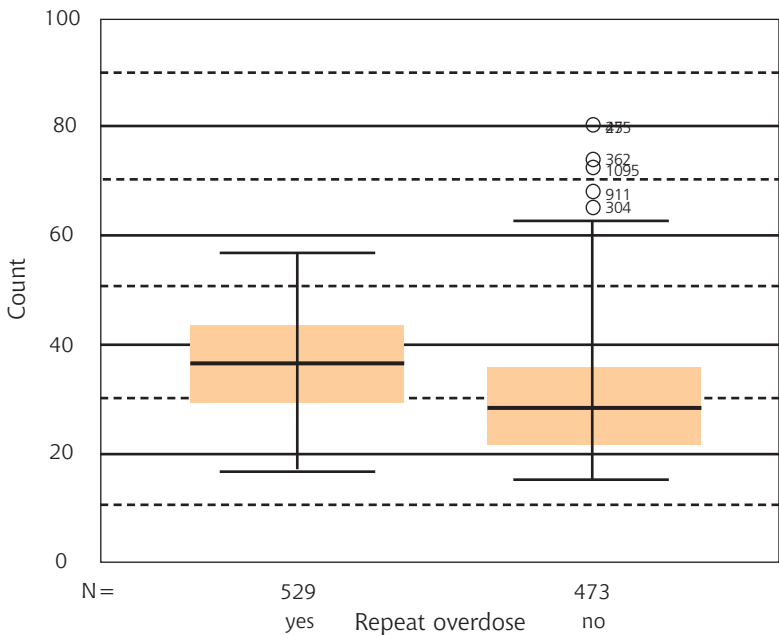


This will create a new variable that has age split into bands. Here I am part way through defining the bands, having done one band already (15-24), and am working on the second (25-34).

### Continuous data

Thus far, we have concentrated on nominal data. For age, the number is not just a label, it is a meaningful number, where 20 is more than 19 and less than 21, and these are not simply three different values. Age cannot be less than zero, and it is possible to talk of ratios, for example a person of 40 is twice the age of someone who is 20, and not simply older. Age is therefore ratio data.

I end with one more analysis, this time of age and overdose repetition. I am not doing a statistical test (though one can be done) but again I will use a type of graph, a boxplot, suitable for ratio data, available from the graphs pull-down menu in SPSS (Figure 18).



**Figure 18: Boxplot for age by repeat overdose**

This boxplot shows the range of values with a vertical line ended by two horizontal lines at the bottom and top of the range. Thus you can see that there is little difference between the yes (529 of them as seen just below the x axis) and no (473 of them), though the range for no is slightly wider with the bottom horizontal line lower and the top one higher than for yes. You can see a few circles above the top of the range for no; these are outliers. SPSS calculates that some values are extreme and not very common, so does not include them in the range, but allows you to see them. You will also see orange boxes with a black line in the middle. Half of all data lie in the orange boxes, so about half of people who repeated are from their late twenties to mid forties, and non-repeaters are from early twenties to late thirties. The black line is the median or middle value, so here the median for repeaters is in the mid thirties, and for non-repeaters late twenties. So a large amount of information can be found in the one plot.

If you want the exact median values, plus other information, this can be found via the Explore facility in SPSS. This gives you more precise data, as shown in Table 2. For example, the range of those who overdosed again was from 16 to 59, the median was 36, and the mean was quite similar, 35.83.

Thus, while all people taking overdoses tend to be young, repeaters tend to be a little older. Of course, the first time you overdose you will be younger than the second time, so we will need to explore this relationship carefully. Not now though.

### **Saving your analysis**

You can save your data for later use, just as you can save a word processing file. You can also save the output to a separate output file, so you can get back graphs and tables etc.

Now, suppose you obtained some more data for your study. You can of course fairly easily re-run all the tables, graphs, and tests above. However, it is even easier than that in SPSS. Each time you perform analysis or create graphs you can paste the request you are making to a syntax window. You will note that some of the dialogue boxes shown above have a 'paste' button. The syntax window may be saved, just as the data and output may be, but into a syntax file. Much of the above analysis and graphs are shown

**Table 2: Using 'Explore' to obtain precise data**

Repeat overdose			Statistic	Std. Error	
Age	Yes	Mean		35.83	.480
		95% Confidence Interval for Mean	Lower Bound	34.89	
			Upper Bound	36.77	
		5% Trimmed Mean		35.82	
		Median		36.00	
		Variance		121.902	
		Std. Deviation		11.041	
		Minimum		16	
		Maximum		59	
		Range		43	
		Interquartile Range		17.00	
		Skewness		-.011	.106
		Kurtosis		-1.034	.212
	No	Mean		31.83	.557
		95% Confidence Interval for Mean	Lower Bound	30.73	
			Upper Bound	32.92	
		5% Trimmed Mean		30.92	
		Median		29.00	
		Variance		146.639	
		Std. Deviation		12.109	
		Minimum		14	
		Maximum		80	
		Range		66	
		Interquartile Range		17.00	
		Skewness		1.052	.112
		Kurtosis		1.041	.224

below in the syntax window. You can if you prefer write all your analysis, graphics, recoding, and other functions in the SPSS language, though most users find it much easier to use the graphical interface. Sometimes it is convenient to edit the syntax, once you are familiar with the language, though most users never bother to learn it, or are even aware it exists.

Below (Figure 19) I am re-running only a selection of the saved syntax, that which deals with recoding of the age data.

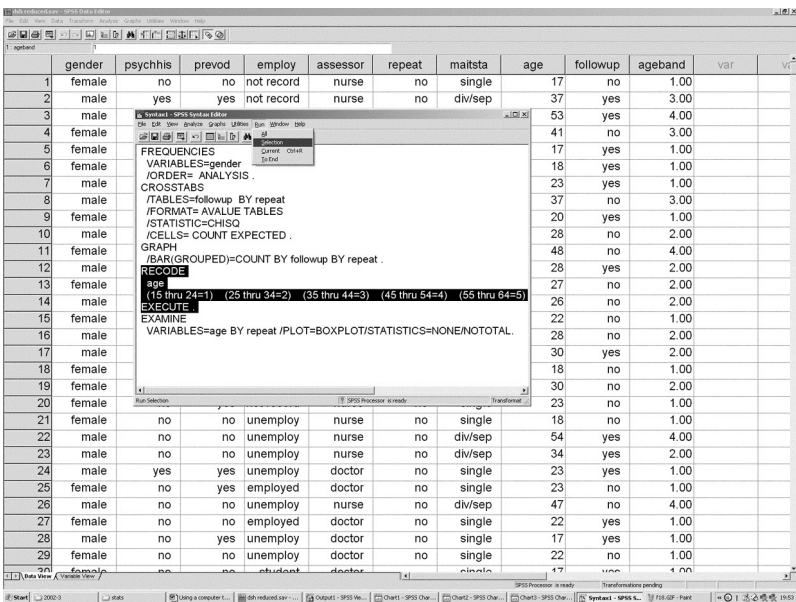


Figure 19: Saved syntax

This will have the effect of re-coding again, taking account of any data changes I may have made since the first re-coding.

## Summary

SPSS, and other statistical packages, make it easy to perform complex statistical analysis, but even simple analysis such as tables and graphical output are much simpler using such a package. Use of such a package does, unfortunately, also allow you to perform meaningless statistics and incorrect sta-

tistical tests, and give misleading or wrong interpretations. You will still need to understand some statistics, but you will not need to be able to compute the results yourself. A statistics package allows you to concentrate on the appropriateness of a test and interpretation of the results. It does not do the whole job for you.

**Denis Anthony RMN, RGN, RN(Canada), BA(Hons), MSc, PhD, ILTM, Professor of Nursing Informatics, School of Nursing & Midwifery, De Montfort University, UK**

**This article has been subjected to a double-blind review**

**references**

Anthony D (1999) *Understanding Advanced Statistics: a Guide for Nurses and Health Care Researchers*. Edinburgh, Churchill Livingstone.

Bryman A (2001) *Quantitative Data Analysis with SPSS Release 10 for Windows*. London, Routledge.

Pallant J (2001) *SPSS Survival Manual*. Buckingham, Open University Press.

**Online archive**

For related articles and author guidelines visit our online archive at [www.nursere searcher.co.uk](http://www.nursere searcher.co.uk)