

SPSS Lesson 4

Creating Data Summaries For Your Research Report

The user-friendliness of SPSS makes it easy to compute a million statistics and/or generate a gazillion charts and tables. Further, the windows compatibility of SPSS means that any charts, tables and/or figures you generate can easily be cut and pasted directly from SPSS into Microsoft Word and other word processing application. But will either your professor or TA be impressed if you do so and then dump them en masse into your research report? *Not.*

As is the case with every other step of the process, you need to know the proper way to represent your data, and understand that the responsibility is yours to present information that is meaningful, on-topic, non-redundant, and to do so in as succinct and clear a manner as possible. Given that objective, you will find there are some ways in which SPSS is very helpful in allowing you to just run procedures and cut and paste the elements directly into your report. Unfortunately, there are at least as many other types of output that contain the information you want along with a bunch of other clutter that you *don't* want, so you end up having to extract the information you need for a table (or whatever) that you must construct yourself.

In this lesson we'll go through the various statistical procedures we worked with in Lesson 3 and focus on the way the information you gain through those procedures might appear in your research report.

FREQUENCIES

The FREQUENCIES procedure in SPSS creates frequency distributions and will compute descriptive summary statistics (e.g., mean, median, mode) for each of your variables. There are basically two different kinds of information that will be useful for you via this procedure.

First is the situation where you have a bunch of variables you want to summarize. For example, you might have given 7 different Likert-type (agree-disagree) items in your survey, as I did in the two versions of the Crim 320 "QuickSurvey" data base, and you may want to show summary information for all 7 in your report. In the SPSS printout, *just one* of the seven items produced the following information:

Statistics

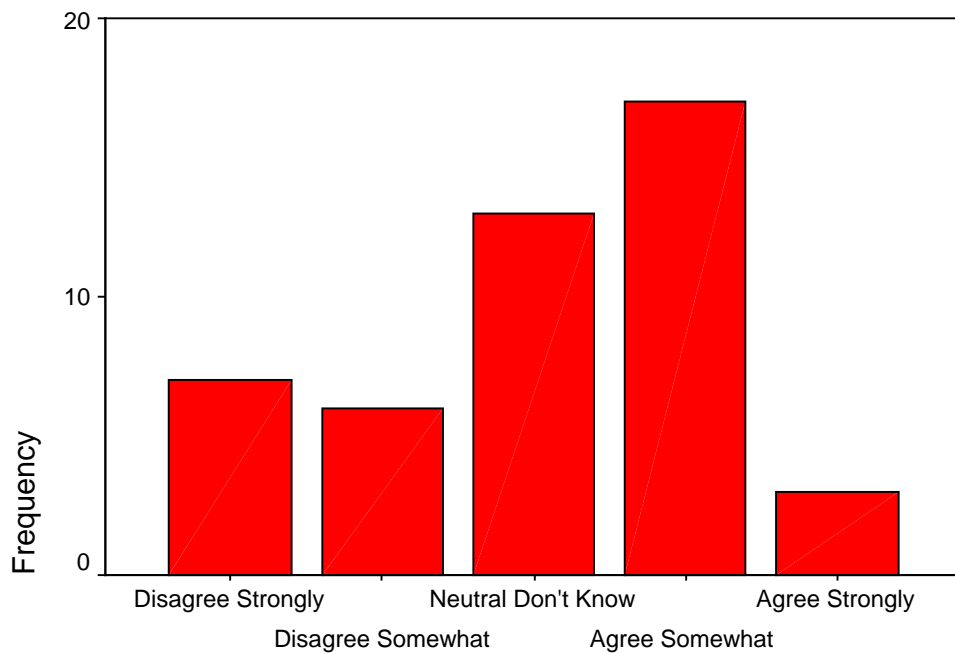
Traditional values undermined

N	Valid	46
	Missing	47
Mean		3.07
Median		3.00
Mode		4
Std. Deviation		1.18

Traditional values undermined

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree Strongly	7	7.5	15.2	15.2
	Disagree Somewhat	6	6.5	13.0	28.3
	Neutral Don't Know	13	14.0	28.3	56.5
	Agree Somewhat	17	18.3	37.0	93.5
	Agree Strongly	3	3.2	6.5	100.0
	Total	46	49.5	100.0	
Missing	9	47	50.5		
Total		93	100.0		

Traditional values undermined



Traditional values undermined

Wow. A whole page of information for just one item. And just think – we have 7 items! Will we be impressed if you give us 7 pages of charts and tables including all of that information? Not at all; your task is to take it all and figure out a way to put it on one page, preferably in one table. Here is one way to do so:

Table 1
Overall Distributions and Summary Statistics for
“Internationalist” and “Protectionist” Survey Items

Item	N	Mean	Distribution of Responses ^x				
			1	2	3	4	5
			DSt N (%)	DSo N (%)	N/DK N (%)	ASo N (%)	ASt N (%)
Internationalist Items							
Canada's military should continue primarily in a "peacekeeping" rather than "military" role.	47	4.11	1 (2.1)	3 (6.4)	6 (12.8)	17 (36.2)	20 (42.6)
Canada should do everything it can to foster its image as an open country that values human rights.	47	4.60	0 (0.0)	2 (4.3)	1 (2.1)	11 (23.4)	33 (70.2)
Canadians, in general, are a tolerant people who value cultural diversity.	48	3.73	0 (0.0)	7 (14.6)	5 (10.4)	30 (62.5)	6 (12.5)
Protectionist Items							
It seems that many traditional Canadian values are being undermined these days.	46	3.07	7 (15.2)	6 (13.0)	13 (28.3)	17 (37.0)	3 (6.5)
The unemployment level in Canada is far too high..	45	3.80	0 (0.0)	9 (20.0)	3 (6.7)	21 (46.7)	12 (26.7)
There is far too much money spent these days on social welfare services.	45	2.87	6 (13.3)	13 (28.9)	11 (24.4)	11 (24.4)	4 (8.9)
Criterion Item							
Current immigration levels are too high; fewer immigrants should be let in the country each year.	92	2.78	18 (19.6)	23 (25.0)	24 (26.1)	15 (16.3)	12 (13.0)

^x Dst = Disagree Strongly; Dso = Disagree somewhat; N/DK = Neutral or Don't Know; Aso = Agree Somewhat; Ast = Agree Strongly

Note that what started off as a tonne of information was boiled down to less than one page, largely by deciding what was most important to me and guided by a desire to place all the information in one place so that the reader could easily look at the distributions, compare means, etc. The resulting table was given a number, a name that reflected what was being summarized inside it, and expressed the information far more accessibly than page after page of printout would have done.

Of course not everything has to go in a table. There are times where you may want to show the distribution of a single “key” variable, for example, which you might do in a table, or could do in a chart or figure. Charts and figures, I should note, are quite amenable to straight cut and pasting if you take the time to set them up well. This starts with doing a good job of defining your variables (e.g., with informative variable and value labels), but needs to be supplemented by some planning at the point where you want to generate the chart.

For example, let’s say that you want to show people the distribution of scores on the item that dealt with criminology students’ career preferences. To do it as a bar graph, you would follow the following procedures:

1. Open your SPSS data file and click **graphs** on the menu at the top of the page.
2. A dropdown menu appears; click **bar** on that drop down menu in order to create a bar graph.
3. A **bar chart** dialogue box appears. Decide at left what sort of graph you have – for a single variable like this we would simply check **simple**. At right, click **define** in order to tell SPSS what you want a graph of.
4. A **define simple bar** dialogue box appears. Fill in the information the box requires:
 - a. At left, click whatever variable you want to graph, and then click the arrow that will take it to the “category axis” box.
 - b. At the top, click what you want the bars to represent – I want it to show the percentage of students falling in each group, so I click “% of cases.”
5. Note that there are also places in that dialogue box for you to do a title for your graph. I do so by clicking the **title** button. On “Line 1” I place a Figure number – “Figure 1” – because graphs are known as “figures.” On Line 2, I place a title: “Career aspirations of Crim 320 Students” to make clear what is being represented. Click **continue** to get back to the **Define simple bar** dialogue box.
6. Click **OK**

Upon following these procedures, I get the nicely labelled graph below, which can be cut and paste directly into your word processing programme (MSWord in my case) as follows:

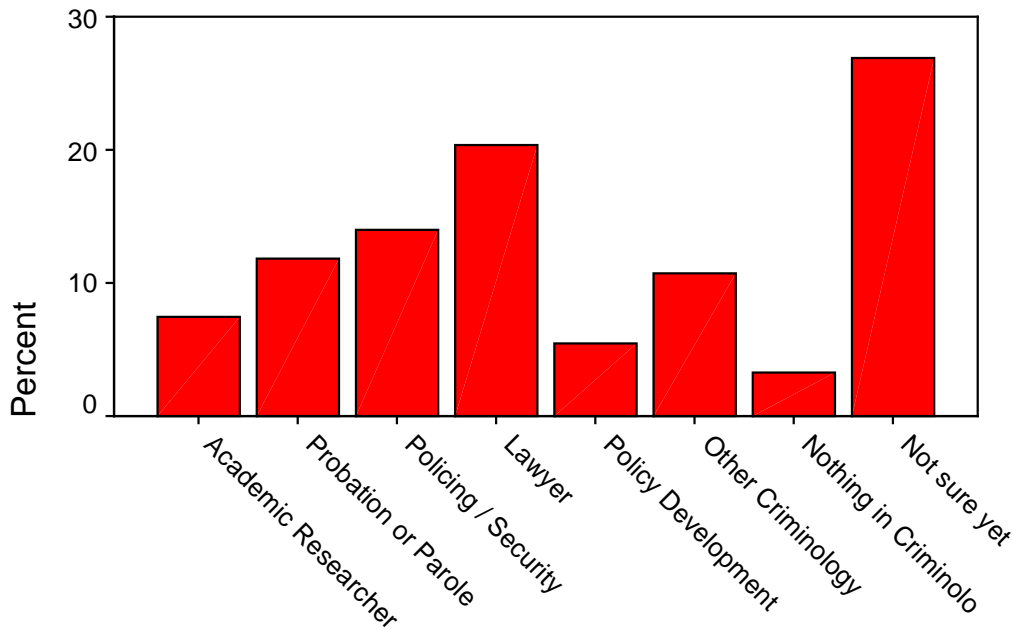
1. Place the cursor on top of the graph and right-click so that a dialogue box opens;
2. Click on **copy objects**
3. Go to your document and place the cursor in the location you want the graph inserted and either right-click on that spot and then click **paste** or;

4. Simply click paste.

If it looks a little too small or too big, click on the graph, at which point you will see the outline of the figure, including some black squares at the corners. To make the graph bigger (or smaller), just drag on a corner and watch it grow to whatever size you want.

Figure 1

Career Aspirations of Crim 320 Students



Current career preference

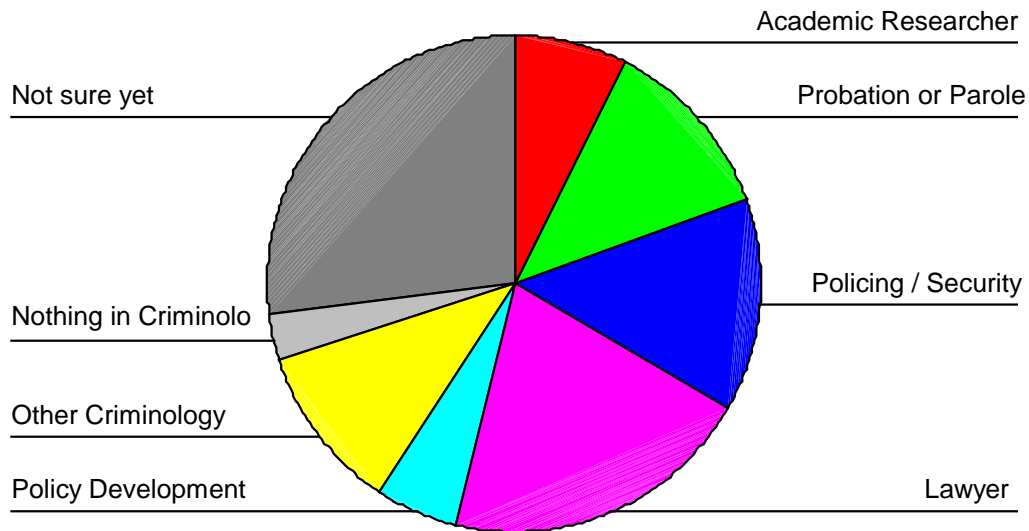
Note that you would follow essentially the same procedures to do a pie chart, except that you pick pie instead of bar when following the procedures outlined above. When you open the pie dialogue box, you will see that it looks pretty much identical to the bar dialogue box, including giving you the possibility of doing a title.

One word to the wise: Pie charts can be pretty spiffy as long as you do not have too many categories and/or have too many small slivers.

The “career aspirations” pie chart is pushing it a bit in terms of the number of categories, but still looks pretty good:

Figure 2

Career Aspirations of Crim 320 Students



CROSS-TABULATIONS AND CHI-SQUARE

Cross-tabulations allow you to look at how two variables interact while the chi-square statistic is the measure of whether the association between the two variables represented is statistically significant ($p < .05$), or within the realm of what you'd expect on the basis of chance variation alone ($p > .05$, or simply $p = ns$).

The procedures for running a chi-square were discussed in lesson 3, and what you should have seen at that time is that the cross-tabulation procedure allows you considerable flexibility in how you set up the table in the first place. For example, you can determine which is the row and which is the column variable, you can have the categories in descending or ascending order, and you can control which of many elements appears within each cell. Because of this, you can get a cross-tabs table pretty close to what you would want it to look like in your report. The one unfortunate aspect is that SPSS does not allow you to place a title on it like it does for the charts/graphs – it does a default version that is a bit too cryptic, so that you would surely have to redo it if you were submitting it for a journal, but it's close enough that I would allow you to include it, with a couple of small adaptations, in your research report.

Let's look again at the crosstabs we generated involving gender and involvement in team-based sports. You've seen the table and chi-square statistic in Lesson 3, here's what I would make it look like for a research report:

Table 3
Cross-Tabulation of Gender of Respondent and Whether They have Engaged Recently in a Team-Based Sports Activity¹

Gender of Respondent * Played team-based sport Crosstabulation

			Played team-based sport		Total
			No	Yes	
Gender of Respondent	Men	Count	19	15	34
		% within Gender of Respondent	55.9%	44.1%	100.0%
	Women	Count	49	10	59
		% within Gender of Respondent	83.1%	16.9%	100.0%
Total		Count	68	25	93
		% within Gender of Respondent	73.1%	26.9%	100.0%

¹ Chi-square = 8.10; df=1; $p < .05$

Note the two things I've done: (1) assigned a table number and title; and (2) placed the appropriate statistic – the chi-square value, degrees of freedom, and an indication of whether the statistic is significant or not. Note by the way, that I did *not* insert the exact probability level that you can find on the printout (which is .004; see notes for Lesson 3). Some people have the erroneous idea that the p value is a measure of how significant a statistic is, but that is not the case. Statistical significance is like a goal in hockey or soccer. Any shot is either a goal or not, but it is not any more or less of a goal depending on where in the net it hits. Similarly, you establish a significance level -- .05 normally – and then determine whether the statistic you observe is significant or not, at that level, period.

THE T-TEST

T-tests are used to look at the difference between two groups, and normally these are just reported in the text of your report, and not presented in a table, because there is really not all that much to present. Graphs of group means will sometimes appear in reports (recall, for example, the graph of mean for the groups in the Donnerstein and Berkowitz sex and aggression study), but not often, and not likely in your projects.

But let's take a couple of examples and work them through. An easy example is the one that was done in SPSS Lesson 3, in which we compared the means scores on the

“immigration” item for the groups of people who received the two different versions of the questionnaire. You would report something like that in your research papers this way:

The group who had been randomly assigned to the “protectionist” group had a mean rating of 2.83 on the 5-point scale, while the group who received the “internationalist” version had a mean rating of 2.74. Although these differences were in the predicted direction – the “protectionist” group indicated a higher level of agreement than the “internationalist” group to an item that stated that immigration levels were “too high” – a t-test showed that the differences were no greater than one would expect on the basis of chance variation alone ($t = 0.34$; $df = 90$; $p=ns$).

On some occasions, however, particularly where you are comparing two groups on multiple identical or similar dependent variables, it may well be more expedient to place the information in a table. For example, one of the survey items asked participants to rate a number of different attributes in terms of how important they considered each to be in the selection of a mate.

Using the scale at right, indicate how essential it is to you that your mate/partner have each of the following attributes.	1	2	3	4	5	6	7
	Not essential			Essential			
Well-educated:	1	2	3	4	5	6	7
Physically attractive:	1	2	3	4	5	6	7
Financially secure:	1	2	3	4	5	6	7
Good sense of humour:	1	2	3	4	5	6	7
Strong sex drive:	1	2	3	4	5	6	7
Intelligent:	1	2	3	4	5	6	7

This survey item was included because of a study reported in a book I had written (called *Research Decisions*) that included a study showing that men and women valued different attributes for a mate, with men emphasizing physical appearance while women emphasized education and income. Rather than reporting each and every comparison, you could do a whole series of t-tests (this is not the best statistical test in such a situation, but we can live with this option for this course), and then summarize the results in a table.

The fact that all the statistics are in the table means that you do not have to repeat them again in the narrative part of your paper, but can just talk about them on the assumption that your reader can see the table, and all you need to do is draw their attention to the highlights. Your write-up might look something like this:

As Table 4 reveals, there were numerous differences between the men and women in the degree of importance they attached to the different attributes, but

only two of these were statistically significant: the importance attached to “physical attractiveness” and “strong sex drive.” In both cases, as predicted, the ratings by men were significantly higher (at $p < .05$) than those by females on those attributes.

Table 4
Comparison of Men’s and Women’s Ratings Regarding the Desirability of Six Different Attributes in a Mate

Attribute	Mean Rating ²		Comparison		
	Men	Women	<i>t</i>	<i>df</i>	<i>p</i>
Physically Attractive	5.76	4.75	3.74	91	<.05
Strong Sex Drive	5.94	4.86	2.28	42	<.05
Financially Secure	4.50	4.88	1.18	91	ns
Well-Educated	5.50	5.12	1.13	91	ns
Sense of Humour	6.24	6.34	0.49	91	ns
Intelligent	6.15	6.03	0.43	91	ns

² The scale ran from 1 (Not essential) to 7 (Essential)

What you should see from this example is that the time you want to use a table (or figure) is when they can help you summarize a lot of information in one place that, if you had to report it narratively in the body of your report, would be about as exciting to read as a laundry list. And once again, note that Table 4 does not come directly from SPSS – I had to put it together on the basis of information extracted from six different analyses that originally consumed several pages of printout, but was able to boil it down into one small and easy to read table.

Final Admonitions

There are several themes that go through this lesson regarding the preparation of tables, charts and graphs:

- There are times when SPSS produces great output that can easily be cut and pasted into your research report (most notably figures/graphs), but many other times where you have to extract the information from multiple parts of SPSS printouts.
- The time to create a table or figure is when using one helps you represent a lot of information in one economical table or picture. The name of the game is data *summary* and *reduction*.
- Every table and figure will have a number (tables and figures are numbered separately) and a title; the title will make clear what information is in the table.