

Tabulating variables in SPSS (Practical)

Tabulation practical

In this practical we will take a first look at the cross-tabulation options in SPSS. Cross tabulations are useful for summarising the relationships between pairs of categorical variables.

This example using the PISA data explores the relationship between the highest educational qualification held by a parent of the student (PAREDU) and the extent to which the student agrees that they are "Interested in learning about science" (INTEREST_4).

We will begin by performing the simple cross-tabulation of the variables **PAREDU** and **INTEREST_4** using the SPSS instructions shown below:

- Select **Crosstabs...** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- Copy the **Highest qualification of parent[PAREDU]** variable into the **Row(s)** box.
- Copy the **Interested in learning science[INTEREST_4]** variable into the **Column(s)** box.
- Click on the **OK** button to produce the table as shown below.

Highest qualification of parent * Interested in learning science Crosstabulation

Count		Interested in learning science				Total
		Strongly disagree	Disagree	Agree	Strongly agree	
Highest qualification of parent	Low: GCSE or equiv	113	178	425	118	834
	Medium: A-level or equiv	156	343	835	278	1612
	High: University degree	165	384	1035	480	2064
Total		434	905	2295	876	4510

This simple cross tabulation simply looks at the variables in the full dataset and looks at how many times each combination of our two categorical variables appears. So we can see that there are 156 occurrences where **PAREDU** takes value Medium: A-level or equiv and **INTEREST_4** takes value Strongly disagree. This is out of 1612 occurrences in total where **PAREDU** takes value Medium: A-level or equiv and 434 occurrences in total where **INTEREST_4** takes value Strongly disagree.

From this table we can see that when **PAREDU** takes value Medium: A-level or equiv there are fewer instances of **INTEREST_4** taking value Strongly disagree than value Strongly agree. This may be because overall there are fewer instances of **INTEREST_4** taking value Strongly disagree than value Strongly agree.

It can be helpful in understanding the individual cell counts to look at how their relative frequencies within a particular row or column compare to overall relative frequencies for all rows and columns. To do this we look at percentage values as well as counts and we will do this first for rows by following the instructions below:

- Select **Crosstabs...** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- The Row and Column variables should be already chosen so Click on the **Cells...** button.
- Click on the **Row** tickbox found under the **Percentages** section.
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to produce the table as shown below.

Highest qualification of parent * Interested in learning science Crosstabulation

		Interested in learning science				Total	
		Strongly disagree	Disagree	Agree	Strongly agree		
Highest qualification of parent	Low: GCSE or equiv	Count	113	178	425	118	834
		% within Highest qualification of parent	13.5%	21.3%	51.0%	14.1%	100.0%
	Medium: A-level or equiv	Count	156	343	835	278	1612
	% within Highest qualification of parent	9.7%	21.3%	51.8%	17.2%	100.0%	
	High: University degree	Count	165	384	1035	480	2064
	% within Highest qualification of parent	8.0%	18.6%	50.1%	23.3%	100.0%	
Total		Count	434	905	2295	876	4510
		% within Highest qualification of parent	9.6%	20.1%	50.9%	19.4%	100.0%

This new table contains the same information as the last table but with the addition of percentages underneath each count. These percentages represent what percentage of the observations in each row are found in each cell. So we see that when **PAREDU** takes value Medium: A-level or equiv then for **INTEREST_4** 9.7% (or 156) take value Strongly disagree, 21.3% (or 343) take value Disagree, 51.8% (or 835) take value Agree and 17.2% (or 278) take value Strongly agree.

This compares with the overall distribution of **INTEREST_4** where 9.6% (or 434) take value Strongly disagree, 20.1% (or 905) take value Disagree, 50.9% (or 2295) take value Agree and 19.4% (or 876) take value Strongly agree. So for example we see that when **PAREDU** takes value Medium: A-level or equiv there is a greater percentage (21.3%) taking category Disagree than on average (20.1%). We can now look at columns instead of rows via the following:

- Select **Crosstabs...** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- The Row and Column variables should be already chosen so Click on the **Cells...** button.
- Remove the **Row** tickbox and instead click on the **Column** tickbox under the **Percentages** section.
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to produce the table as shown below.

Highest qualification of parent * Interested in learning science Crosstabulation

			Interested in learning science				
			Strongly disagree	Disagree	Agree	Strongly agree	Total
Highest qualification of parent	Low: GCSE or equiv	Count	113	178	425	118	834
		% within Interested in learning science	26.0%	19.7%	18.5%	13.5%	18.5%
	Medium: A-level or equiv	Count	156	343	835	278	1612
		% within Interested in learning science	35.9%	37.9%	36.4%	31.7%	35.7%
	High: University degree	Count	165	384	1035	480	2064
		% within Interested in learning science	38.0%	42.4%	45.1%	54.8%	45.8%
Total		Count	434	905	2295	876	4510
		% within Interested in learning science	100.0%	100.0%	100.0%	100.0%	100.0%

This new table contains the same information as the last table but this time the percentages underneath each count are column rather than row percentages. These percentages represent what percentage of the observations in each column are found in each cell. So we see that when **INTEREST_4** takes value Strongly disagree then for **PAREDU** 26.0% (or 113) take value Low: GCSE or equiv, 35.9% (or 156) take value Medium: A-level or equiv and 38.0% (or 165) take value High: University degree.

This compares with the overall distribution of **PAREDU** where 18.5% (or 834) take value Low: GCSE or equiv, 35.7% (or 1612) take value Medium: A-level or equiv and 45.8% (or 2064) take value High: University degree. So for example we see that when **INTEREST_4** takes value Strongly disagree there is a greater percentage (35.9%) taking category Medium: A-level or equiv than on average (35.7%). Finally for completeness we can look at total percentages as follows:

- Select **Crosstabs...** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- The Row and Column variables should be already chosen so Click on the **Cells...** button.
- Remove the **Column** tickbox and instead click on the **Total** tickbox under the **Percentages** section.
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to produce the table as shown below.

Highest qualification of parent * Interested in learning science Crosstabulation

			Interested in learning science				
			Strongly disagree	Disagree	Agree	Strongly agree	Total
Highest qualification of parent	Low: GCSE or equiv	Count	113	178	425	118	834
		% of Total	2.5%	3.9%	9.4%	2.6%	18.5%
	Medium: A-level or equiv	Count	156	343	835	278	1612
		% of Total	3.5%	7.6%	18.5%	6.2%	35.7%
	High: University degree	Count	165	384	1035	480	2064
		% of Total	3.7%	8.5%	22.9%	10.6%	45.8%
Total		Count	434	905	2295	876	4510
		% of Total	9.6%	20.1%	50.9%	19.4%	100.0%

In this final table the percentages represent percentages out of all the observations. So for example the 178 observations where **PAREDU** takes value Low: GCSE or equiv and **INTEREST_4** takes value Disagree represent 3.9% of all the observations.

These percentages are probably the least useful of the three options as they do not really aid in comparisons within the table but might aid comparisons between different tables. This ends the practical on tabulations.

It seems that there is a relationship between parental educational background and self-reported interest in learning science, with stronger agreement expressed by those from more educated families. In this example, it arguably makes sense to focus on the first set of row percentages, which show how the distribution of attitudes to science differs between educational groups. The alternative column percentages, in contrast, reveal how the distribution of educational background differs across attitudinal groups. It is always necessary to think carefully about which representation is the most informative one for the question at hand.