

Block 3: Analysing two variables (and sometimes three)

3.1.3 Contingency tables from SPSS

[updated 28 June 2011]

In survey research, contingency tables are used in two ways. The first is **descriptive** by which we mean things like checking the sex and age-group structure of your sample or checking ownership of consumer goods by housing tenure, household income etc.

The second is **explanatory**. As well as **describing** combinations of two or more variables, social researchers often seek to **account** for **variation** in **dependent** variables (the ones being explained) from variation in **independent** variables (the ones being used to do the explaining). Between any two variables it's up to you to decide which way you think the effect flows!

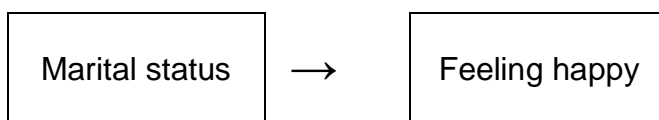
To do this with contingency tables involves comparing percentages or calculating statistical measures of (non-) association. (I'll deal with the latter later in the course, but they are discussed at length in section 6 of Jim Ring's accompanying [Statistical Notes](#) and there are plenty of on-line [tutorials for SPSS](#) around, far better than anything I am qualified to write.)

For instance, "Is there a relationship between marital status and feeling happy" or "What effect does marital status have on self-reported happiness" ?

The following question has been asked in countless surveys in many countries over the years. This is an extract from the questionnaire for the 1973 Quality of Life in Britain survey: the question was:

Q.J6	Taking all things together, how would you say things are these days? Would you say you are:			
		<u>READ</u>	VERY HAPPY	3
		<u>OUT</u>	FAIRLY HAPPY	2
			NOT TOO HAPPY	1
				(37)

We'll be using data from this survey to explore the relationship:



... or is it the other way round?

Is this the true story, or are there any other variables, (related or unrelated to marital status) which might influence feeling happy? What might they be? How do they affect the relationship between marital status and feeling happy? Thus as well as **dependent** and **independent** variables, we also need to think of **test** variables to examine the initial relationship between marital status and feeling happy by **controlling** for the **test** variables. Does marital status affect feeling happy at all? These are the kind of questions which make survey research so interesting.

When looking at contingency tables for only two variables (usually, but not always, dependent and independent) we use the term **zero order** tables (ie no control variables): when we introduce one test variable, we use the term **1<sup>st</sup> order** tables; two test variables, **2<sup>nd</sup> order** tables and so on. During this process we look at what happens to a zero order statistic (eg percent of married people saying they feel "very happy" these days) when we control for a test variable (eg gender).

Here is the table of counts from the survey. We shall use SPSS to investigate the relationship (if any) between these two variables.

marital \* happy Crosstabulation

Count		happy			Total
		Not too happy	Fairly happy	Very happy	
marital	Married or cohabiting	10	105	38	153
	Single	22	321	342	685
	Widowed	11	59	24	94
	Separated or divorced	12	19	3	34
Total		55	504	407	966

[Source: Quality of Life in Britain, 1973 wave]

## CROSSTABS

The SPSS command **CROSSTABS** creates 2-way contingency tables for two variables, and nested sub-tables when controlling for one or more test variables. It also provides percentages based on row, column and global totals, a range of measures of statistical (non-) association and various controls over table content and format.

### General format:

```

CROSSTABS [ TABLES =                                [optional only]
              <row varlist> BY <column varlist>         [ BY <control varlist>]
              / <row varlist> BY <column varlist>         [ BY <control varlist>]
              ~ ~ ~ ~ ~
              /CELLS <cell content options>
              /STATISTICS <statistics options>
              / <other options>
    
```

The variable(s) specified **before** the keyword **BY** will form the **rows** of the tables, and the variables specified **after** keyword **BY** will form the **columns**.

### Examples:

```

CROSSTABS      sex BY agegroup .
    
```

sex \* agegroup Crosstabulation

Count		agegroup				Total
		18-29	30-44	45-59	60+	
sex	Men	113	112	100	115	440
	Women	126	118	131	148	523
Total		239	230	231	263	963

[Source: Quality of Life in Britain, 1973 wave]

The first table above was produced by:

```

CROSSTABS      marital BY happy .
    
```

You can specify multiple tables in the same command, so both tables above can be produced by:

```
CROSSTABS          SEX BY AGEGROUP  
                   / MARITAL BY HAPPY .
```

**[NB]** It is **crucial** to remember to separate multiple table specifications with a forward slash /. If you fail to do this, you will produce dozens of useless tables of variables cross-tabulated by themselves, get RSI<sup>1</sup> from scrolling and possibly also waste lots of paper and trees!

As with the **FREQUENCIES** command there is a range of **sub-commands** and **keywords** to control the format and content of your tables and also to request appropriate statistics, and are used in the same way. The most commonly used sub-commands are:

**/CELLS =** to specify the content of the table cells, and  
**/STATISTICS =** to specify the statistics to be applied,

It is also possible to ask for missing values to be tabulated ( **/MISSING = INCLUDE** ) for labels to be omitted, tables suppressed or values tabulated in descending rather than ascending order (**/FORMAT =** ) and for cells to be written to disk for later statistical processing without having to read the data all over again ( **/WRITE =** )

The only ones you need to know about for now are:

<b>Sub-command</b>	<b>Keyword</b>	<b>Meaning</b>
<b>/CELLS</b>	<b>COUNT</b>	Absolute cell frequency
	<b>ROW</b>	Row percent
	<b>COLUMN</b>	Column percent
	<b>TOTAL</b>	Total percent
<b>/STATISTICS</b>	<b>CHISQ</b>	Chi-squared test (and I'm leaving that for later!)

---

<sup>1</sup> Repetitive strain injury, common in typist's wrists and fingers

Simple counts aren't much use in answering our research question, so we need to standardise the distributions by converting the counts into percentages. This makes comparisons easier.

As well as counts SPSS can produce three sets of percentages based on row totals, column totals and table totals.

**marital \* happy Crosstabulation**

			happy			Total
			Not too happy	Fairly happy	Very happy	
marital	Married or cohabiting	Count	10	105	38	153
		% within marital	6.5%	68.6%	24.8%	100.0%
		% within happy	18.2%	20.8%	9.3%	15.8%
		% of Total	1.0%	10.9%	3.9%	15.8%
Single	Single	Count	22	321	342	685
		% within marital	3.2%	46.9%	49.9%	100.0%
		% within happy	40.0%	63.7%	84.0%	70.9%
		% of Total	2.3%	33.2%	35.4%	70.9%
Widowed	Widowed	Count	11	59	24	94
		% within marital	11.7%	62.8%	25.5%	100.0%
		% within happy	20.0%	11.7%	5.9%	9.7%
		% of Total	1.1%	6.1%	2.5%	9.7%
Separated or divorced	Separated or divorced	Count	12	19	3	34
		% within marital	35.3%	55.9%	8.8%	100.0%
		% within happy	21.8%	3.8%	.7%	3.5%
		% of Total	1.2%	2.0%	.3%	3.5%
Total	Total	Count	55	504	407	966
		% within marital	5.7%	52.2%	42.1%	100.0%
		% within happy	100.0%	100.0%	100.0%	100.0%
		% of Total	5.7%	52.2%	42.1%	100.0%

[Source: Quality of Life in Britain, 1973 wave]

...but this makes for a very cluttered presentation, so we need to simplify it and keep only the figures that we really need.

To answer our research question, the dependent variable is "feeling happy" and the independent variable is marital status. We therefore need to look at the distribution of percentages feeling very happy, fairly happy or not too happy for each marital status category. This means we want percentages to add up to 100% within each category of marital status.

Totalling to 100% in columns is the way accountants and statisticians do it,

```

CROSSTABS <depvar> BY <indvar1> .
CROSSTABS HAPPY BY MARITAL
/CELLS COUNT COL .
  
```

happy \* marital Crosstabulation

			marital				Total
			Married or cohabiting	Single	Widowed	Separated or divorced	
happy	Very happy	Count	38	342	24	3	407
		% within marital	24.8%	49.9%	25.5%	8.8%	42.1%
	Fairly happy	Count	105	321	59	19	504
		% within marital	68.6%	46.9%	62.8%	55.9%	52.2%
	Not too happy	Count	10	22	11	12	55
		% within marital	6.5%	3.2%	11.7%	35.3%	5.7%
Total		Count	153	685	94	34	966
		% within marital	100.0%	100.0%	100.0%	100.0%	100.0%

. . . but it's much easier to compare figures visually down columns rather than across rows.

I prefer to have the dependent variable across the top of the table and the independent variable(s) down the side, viz:

**CROSSTABS** <indvar1> **BY** <depvar> .

To repeat the analysis for the tables above the **dependent** variable is **happy** and the **independent** variable is **marital** so we want **marital** to be the row variable and **happy** to be the column variable.

We need to be able to compare marital groups so we ask for percent happy within the categories of marital status (along the rows) therefore we need to tabulate **marital** by **happy** and print row percent, thus for the zero order table:

**CROSSTABS**      **MARITAL BY HAPPY**  
**/CELLS**            **COUNT ROW .**

marital \* happy Crosstabulation

			happy			Total
			Not too happy	Fairly happy	Very happy	
marital	Married or cohabiting	Count	10	105	38	153
		% within marital	6.5%	68.6%	24.8%	100.0%
	Single	Count	22	321	342	685
		% within marital	3.2%	46.9%	49.9%	100.0%
	Widowed	Count	11	59	24	94
		% within marital	11.7%	62.8%	25.5%	100.0%
	Separated or divorced	Count	12	19	3	34
		% within marital	35.3%	55.9%	8.8%	100.0%
Total		Count	55	504	407	966
		% within marital	5.7%	52.2%	42.1%	100.0%

[Source: Quality of Life in Britain, 1973 wave]

This is still a bit cluttered, so a simpler table can be produced by:

**CROSSTABS** <sup>2</sup> **MARITAL BY HAPPY /CELLS = ROW .**

Row percentages are asked for because we are treating **happy** as the dependent variable and its marginal frequencies will total to 100% at the end of each row.

marital \* happy Crosstabulation

		happy			Total
		Not too happy	Fairly happy	Very happy	
% within marital	marital Married or cohabiting	6.5%	68.6%	24.8%	100.0%
	Single	3.2%	46.9%	49.9%	100.0%
	Widowed	11.7%	62.8%	25.5%	100.0%
	Separated or divorced	35.3%	55.9%	8.8%	100.0%
	Total	5.7%	52.2%	42.1%	100.0%

[Source: Quality of Life in Britain, 1973 wave]

This table is easier to interpret, but we have lost the base for percentaging at the end of each row. Without special programming beyond the scope of this tutorial, SPSS cannot produce a table with n instead of 100%.

These are known as **zero order** tables (ie no other variable apart from the two being tabulated). If you wish to introduce a single **control** or **test** variable to see if the distribution changes within each sex (known as a **1<sup>st</sup> order** table) there will be nested sub-tables of row-variable(s) by column-variable(s) for each value, or combination of values, of the control variable(s) after the second **BY**.

**CROSSTABS** **MARITAL BY HAPPY BY SEX /CELL ROW.**

marital \* happy \* sex Crosstabulation

			happy			Total
			Not too happy	Fairly happy	Very happy	
% within marital	Men	marital Married or cohabiting	4.8%	76.2%	19.0%	100.0%
		Single	3.6%	53.5%	42.9%	100.0%
		Widowed	6.7%	66.7%	26.7%	100.0%
		Separated or divorced	50.0%	41.7%	8.3%	100.0%
		Total	5.2%	57.9%	36.9%	100.0%
Women	marital	Married or cohabiting	8.7%	59.4%	31.9%	100.0%
		Single	2.8%	40.7%	56.5%	100.0%
		Widowed	12.7%	62.0%	25.3%	100.0%
		Separated or divorced	27.3%	63.6%	9.1%	100.0%
		Total	6.1%	47.3%	46.6%	100.0%

[Source: Quality of Life in Britain, 1973 wave]

<sup>2</sup> SPSS reads only the first three or four characters of subcommands and keywords: you don't need the = sign either and lower case is easier to read so (less typing and easier to read): **cro marital by happy /cel row .**

From these tables we can see that, for people who are (living as) married, the figure of 24.8% "very happy" in the zero-order table is actually composed (a weighted average) of 19% for men and 31.9% for women in the 1<sup>st</sup> order table. Among separated or divorced respondents, men (50%) are almost twice as likely as women (27%) to report feeling "not too happy". Now why might that be? Is it the same for the 1975 wave (different sample from the same primary sampling units) or for more recent UK surveys, or for people in other countries? Is the sample big enough to be confident within certain limits?

This is what **analysis** (from the Greek for "break up") is all about. Once you start hunting for the truth in this way, hopefully guided by reasonable theory or previously formulated hypotheses, you will find it as addictive as it is challenging and also a lot fun, especially once you get the hang of SPSS syntax.

**End of tutorial**

**Next session:**        **3.1.4 Exercise: Contingency tables**

[\[Back to Block 3 menu\]](#)