

CROSSTABS

GET

FILE='U:\)Research\ISSP20072008\issp_2007_2008nl_def.sav'.

CROSS z06a1 by gender.

		gender Gender with imputation		Total
		1 Men	2 Women	
Z06a1 Resp: Highest completed education	1 Basisonderwijs	74	74	148
	2 LBO-VBO-VMBOb	219	239	458
	3 MAVO-MULO-VMBOt	153	214	367
	4 HAVO-MMS	58	89	147
	5 VWO-HBS-Athen-Gymn	55	53	108
	6 KMbo	24	39	63
	7 MBO	278	297	575
	8 HBO	303	332	635
	9 Universiteit	153	123	276
Total	1317	1460	2777	

- These are the raw frequencies that show how nine levels of education are distributed for men and women.
- We can ask many questions about these data:
 - How are education and gender associated?
 - How does gender influence education?
 - Do women and men have different educationa distributions?
 - Have men higher levels of education than women?
- Note that from the second question onwards there is a causal assumption. This directs how we compute percentages in the table.

CROSS z06a1 by gender /cell=count row col.

Table 2: Highest completed education by gender, counts, row and column percentages

		gender Gender with imputation		Total
		1 Men	2 Women	
Z06a1 Resp: 1		74	74	148
Highest completed education	Basisonderwijs	50.0%	50.0%	100.0%
		5.6%	5.1%	5.3%
	2 LBO-VBO-VMBOb	219	239	458
		47.8%	52.2%	100.0%
		16.6%	16.4%	16.5%
	3 MAVO-MULO-VMBOt	153	214	367
		41.7%	58.3%	100.0%

	11.6%	14.7%	13.2%
4 HAVO-MMS	58	89	147
	39.5%	60.5%	100.0%
	4.4%	6.1%	5.3%
5 VWO-HBS- Athen-Gymn	55	53	108
	50.9%	49.1%	100.0%
	4.2%	3.6%	3.9%
6 KMbo	24	39	63
	38.1%	61.9%	100.0%
	1.8%	2.7%	2.3%
7 MBO	278	297	575
	48.3%	51.7%	100.0%
	21.1%	20.3%	20.7%
8 HBO	303	332	635
	47.7%	52.3%	100.0%
	23.0%	22.7%	22.9%
9 Universiteit	153	123	276
	55.4%	44.6%	100.0%
	11.6%	8.4%	9.9%
Total	1317	1460	2777
	47.4%	52.6%	100.0%
	100.0%	100.0%	100.0%

- This is a table that should almost never present. It looks very confusing.
- It is also possible to request TOTAL percentages. That would be even more confusing. Unlike row and column percentages, total percentages hardly have an application.
- We can make two kinds of comparisons in the table:
 - Between column-percentages: compare within rows
 - Between row-percentages: compare within column
- Treiman seems to imply that one is right and one is wrong, but if you do it right, they give the same information. It is customary and useful to compare between categories of the X-variable (in this case gender), mainly because it leads to a linear (regression) model.

We can reduce the table much more, by computing counts and column percentages separately and combine these in excel.

Table 3: Highest completed education by gender, percentage distribution

	1 Men	2 Women
1 Basisonderwijs	5.6%	5.1%
2 LBO-VBO- VMBOb	16.6%	16.4%
3 MAVO-MULO- VMBOt	11.6%	14.7%

4 HAVO-MMS	4.4%	6.1%
5 VWO-HBS- Athen-Gymn	4.2%	3.6%
6 KMbo	1.8%	2.7%
7 MBO	21.1%	20.3%
8 HBO	23.0%	22.7%
9 Universiteit	11.6%	8.4%
	1317	1460
	100.0%	100.0%

- Note that I include only the N for the column marginal and have dropped alle other raw counts, as wel as the marginal counts and percentages. At this point no information is lost, as the original table can still be reconstructed.
- Treiman advises to drop the decimal number, but I disagree. I makes it possible to check you table with the original dataset.
- The distribution of men and women look remarkably similar. But how can we now precisely?

Run a statistical test. Here are some results:

Pearson Chi-square	19.1	ndf=8	p < .015
Somers D	-.041		p < .056
Pearson's R	-.036		p < .061

This suggest marginally significant differences between men and women. Somers D and Pearson's R assume ranko order and suggest that women are lower educated than men, give the current ordering / scaling of categories.

In table analysis we would simplify by dichotomizing the data:

Table 4: Highest completed education by gender, dichotomized.				
% within gender Gender with imputation				
		gender Gender with imputation		Total
		1 Men	2 Women	
HighEducated	0	65,4%	68,8%	67,2%
	1	34,6%	31,2%	32,8%
Total		100,0%	100,0%	100,0%

A further simplification without los of information is:

```
recode Z06a1 (1 thru 7=0)(8 9=1) into HighEducated.
cross Higheducared by gender /cel=col.
```

Table 5: Having High Education by Gender

1 Men	2 Women
-------	---------

34.6% 31.2%
1317 1460

The difference is not significant ($P < .052$)

However, a usefull alternative would be:

```
means z06a1 by gender /stat=all.
```

**Table 6: Mean Education
by Gender**

1 Men	2 Women
5.60	5.40
N=1317	N=1460

But again the difference is not significant ($P < .061$). Note that the test of differences of means is the same as that with pearson's R. These tests are in fact identical.