

**TEESSIDE UNIVERSITY
SCHOOL OF HEALTH & SOCIAL CARE**

SPSS Workbook 3 – Chi-squared & Correlation

**Research, Audit and data
RMH 2023-N**

Module Leader: Sylvia Storey
Phone: 01642 0384969
s.storey@tees.ac.uk

SPSS Workbook 3 – Inferential Statistics 1 (Chi Square & Correlation)

Last week you created a database and entered data into SPSS. This is something you need to become familiar with (especially as it will form part of your SPSS exam). This week you will again be entering data into SPSS but the data files will be much simpler.

1. Chi-square (χ^2)

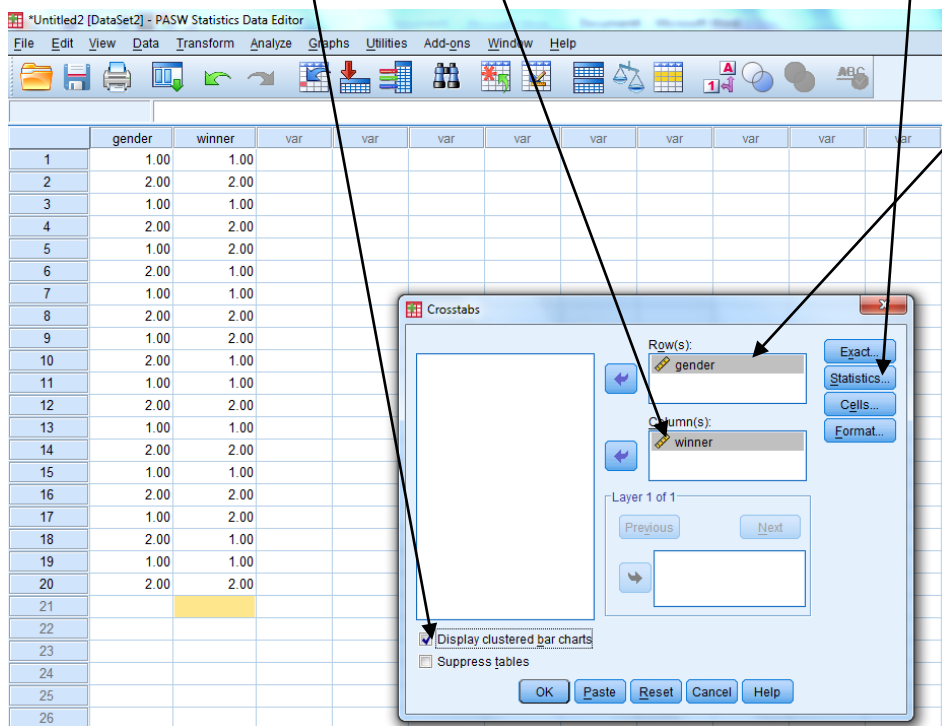
Chi-square is a test of association between 2 categorical (nominal) variables (*remember last week when we looked to see if there was any relationship between gender and smoking status?*).

The data below looks at the results of playing a Wii tennis game. All participants are paired male/female and we are looking to see if there is a relationship between the variables: gender and winner. Please set up your database accordingly and enter the data below (the data tables will not always be presented in the same format as they need to be entered into SPSS so you need to think about this):

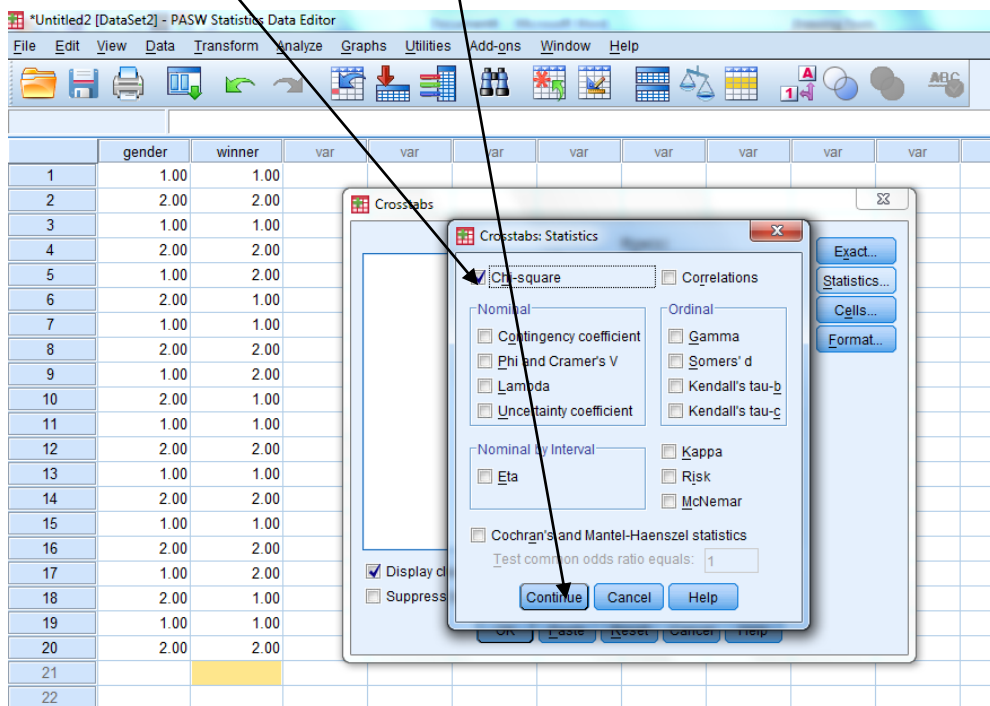
Part	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Sex	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f
Win?	y	n	y	n	n	y	y	n	n	y	y	n	y	n	y	n	n	y	y	n

To carry out the chi-square test select:

Analyze- Descriptive statistics – Crosstabs: then move **gender** into the **Row(s)** box and **winner** into the **Column(s)** box. Selected **Display clustered bar charts** and then click on **Statistics**.

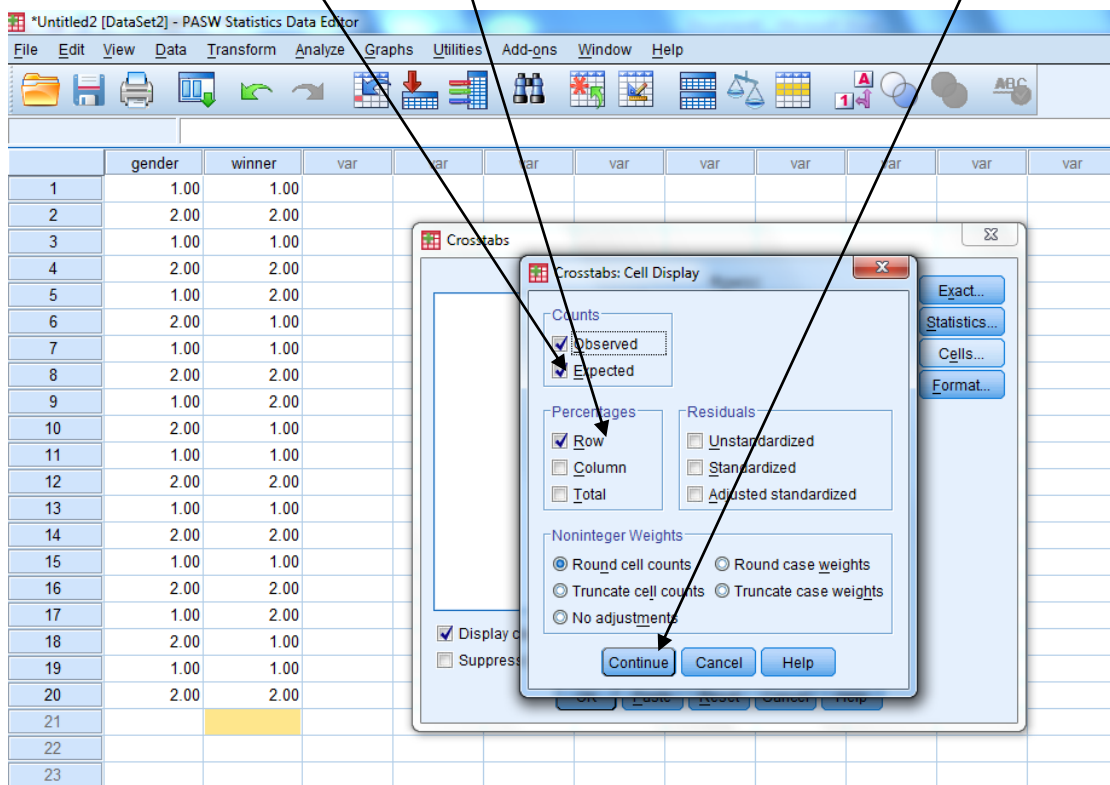


Click on **Chi-square** and **Continue**.



Click on **Cells**

Now select **Expected** and **Row** as detailed bellowed and the **Continue** and **OK** to finish.



The output from this test should appear as:

Crosstabs

[DataSet2]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * winner	20	100.0%	0	.0%	20	100.0%

gender * winner Crosstabulation

			winner		Total
			yes	no	
gender male	Count	7	3	10	
	Expected Count	5.0	5.0	10.0	
	% within gender	70.0%	30.0%	100.0%	
female	Count	3	7	10	
	Expected Count	5.0	5.0	10.0	
	% within gender	30.0%	70.0%	100.0%	
Total	Count	10	10	20	
	Expected Count	10.0	10.0	20.0	
	% within gender	50.0%	50.0%	100.0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.200 ^a	1	.074		
Continuity Correction ^b	1.800	1	.180		
Likelihood Ratio	3.291	1	.070		
Fisher's Exact Test				.179	.089
Linear-by-Linear Association	3.040	1	.081		
N of Valid Cases	20				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.00.
b. Computed only for a 2x2 table

Bar Chart

Annotations:

- This table details the number of cases and would identify any missing values
- This is the same as the table produced last week but we have added additional information. We can now see the % of males and females who won. This is particularly helpful when the numbers are uneven but for our data we had equal numbers of males and females.
- This table reports the chi-squared and p-values for the test, ie it tells us whether there is any significant difference between the groups.

What do the results tell you?

.....

.....

.....

.....

.....

.....

.....

2. Correlation

Correlation looks at the relationship between 2 variables eg weight & height. The type of correlation we are going to look at is Pearson's correlation (a non-parametric correlation called Spearman's correlation is also possible).

We need to determine if the data we have meets certain "parametric assumptions". These are :

- The data is at least interval level
- The data are drawn from normally distributed populations
- The relationship is linear and points are evenly distributed along the straight line (we will look at this later).

Now open the data set **POMS data** – you will need to download this from BB and save it.

A copy of the questionnaire can be found in the appendices and scores have been calculated for the following subscales of the questionnaire:

The Profile of Mood State (POMS) Questionnaire

This questionnaire looks at six aspects of mood state, 5 having a negative effect on performance and one, vigour/activity having a positive effect (McNair et al, 1971)

- **Tension-Anxiety(T)**: Relates to increased body tension, some which may not be readily observable (feeling tense, anxious, on edge) and some which are observable (feeling shaky, being restless). Related to the ability of the body to behave in a controlled and co-ordinated manner.
- **Depression-Dejection(D)**: Represents a mood of depression accompanied by feelings of personal inadequacy, difficulty in adjusting to requirements of the situation.
- **Anger-Hostility(A)**: Characterised by feelings of intense, overt anger (bad-tempered, rebellious), sullenness and milder feelings of hostility.
- **Vigour-Activity(V)**: Represents vigorousness and high energy. Negatively related to the other mood state measures in that a high V score is likely to be associated with lower scores on the other measures.
- **Fatigue-Inertia(F)**: Represents a mood of weariness, inertia and low energy level. May be related to physical and psychological stress.
- **Confusion-Bewilderment(C)**: MAY represent some cognitive (thinking) inefficiency as well as a mood state. Often the result of negative forms of arousal (anxiety, worry).

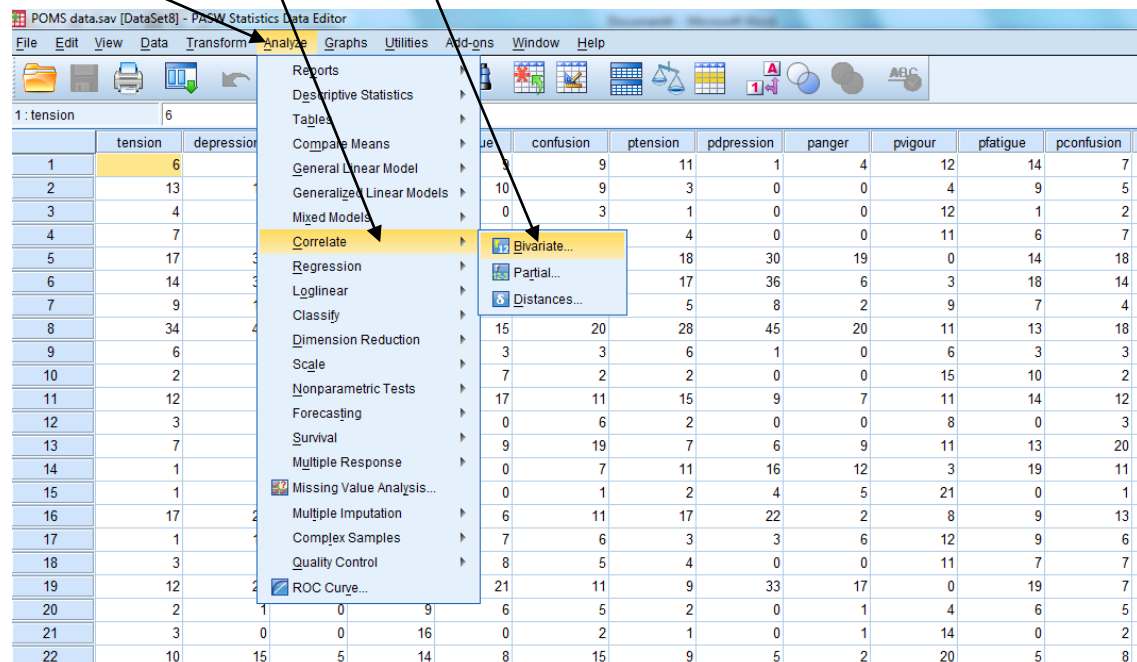
From the information identify 2 sets of variables where you would expect to see the following:

A negative correlation
A positive correlation

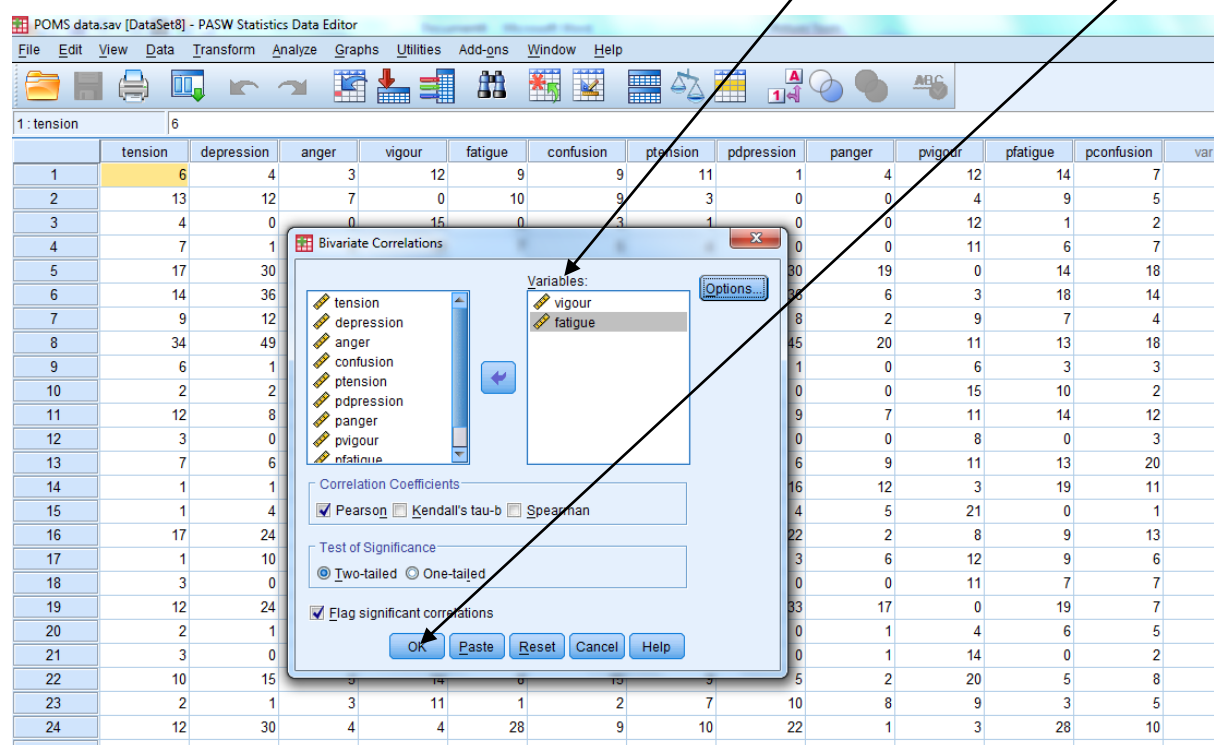
You will need to test the variables to see if they are normally distributed. This can be done in a number of ways and we will look at this in a later session.

Then carry out a correlation by selecting:

Analyse – Correlate – Bivariate



And move the variables you wish to compare into the **Variables** box and click on **OK** to finish.



The output from the test will appear as:

Correlations				
		vigour	fatigue	
vigour	Pearson Correlation	1	-.458**	The correlation coefficient (<i>r</i>) for this test is -0.458.
	Sig. (2-tailed)		.002	
	N	42	42	The p-value is 0.002, which is significant at the 0.01 level.
fatigue	Pearson Correlation	-.458**	1	
	Sig. (2-tailed)	.002		
	N	42	42	

** . Correlation is significant at the 0.01 level (2-tailed).

What does this mean?

.....

.....

.....

Now repeat this with the other set of variables.

Correlations				
		tension	fatigue	
tension	Pearson Correlation	1	.561**	The correlation coefficient (<i>r</i>) for this test is 0.561.
	Sig. (2-tailed)		.000	
	N	42	42	The p-value is 0.001 (ie you need to round up the value as you cannot quote it as being 0.000)
fatigue	Pearson Correlation	.561**	1	
	Sig. (2-tailed)	.000		
	N	42	42	

** . Correlation is significant at the 0.01 level (2-tailed).

Now produce a graph for these by selecting:

Graphs – Legacy Dialogs – Scatter/dot

Select **Simple** and click on **Ok** to continue.

Move the 2 variables from your first correlation into the boxes marked **Y-axis** and **X-axis**. Click on **OK** to continue.

The screenshot shows the PASW Statistics Data Editor interface. The background is a data table with 30 rows and 12 columns. The columns are: tension, depression, anger, vigour, fatigue, confusion, ptension, pdpression, panger, pvigour, pfatigue, pconfusion. The first row of data is highlighted. Overlaid on the table is the 'Simple Scatterplot' dialog box. In the dialog, the 'Y Axis' is set to 'vigour' and the 'X Axis' is set to 'fatigue'. There are buttons for 'Titles...', 'Options...', 'Set Markers by:', and 'Label Cases by:'. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

	tension	depression	anger	vigour	fatigue	confusion	ptension	pdpression	panger	pvigour	pfatigue	pconfusion
1	6	4	3	12	9	9	11	1	4	12	14	7
2	13	12							0	4	9	5
3	4	0							0	12	1	2
4	7	1							0	11	6	7
5	17	30							9	0	14	18
6	14	36							6	3	18	14
7	9	12							2	9	7	4
8	34	49							0	11	13	18
9	6	1							0	6	3	3
10	2	2							7	15	10	2
11	12	8							0	11	14	12
12	3	0							0	8	0	3
13	7	6							9	11	13	20
14	1	1							2	3	19	11
15	1	4							5	21	0	1
16	17	24							2	8	9	13
17	1	10							6	12	9	6
18	3	0							0	11	7	7
19	12	24							7	0	19	7
20	2	1							1	4	6	5
21	3	0							1	14	0	2
22	10	15							2	20	5	8
23	2	1							8	9	3	5
24	12	30							1	3	28	10
25	5	1							7	20	6	10
26	5	0							1	0	22	9
27	3	2							3	17	0	4
28	13	4								6	24	15
29	12	14								14	6	13
30	8	0								6	6	3

Now repeat for the 2nd set of variables and compare the 2 graphs.

ANSWERS

Appendices.

Data set 1:

*Untitled2 [DataSet2] - PASW Statistics Data Editor

	gender	winner	var	var
1	1.00	1.00		
2	2.00	2.00		
3	1.00	1.00		
4	2.00	2.00		
5	1.00	2.00		
6	2.00	1.00		
7	1.00	1.00		
8	2.00	2.00		
9	1.00	2.00		
10	2.00	1.00		
11	1.00	1.00		
12	2.00	2.00		
13	1.00	1.00		
14	2.00	2.00		
15	1.00	1.00		
16	2.00	2.00		
17	1.00	2.00		
18	2.00	1.00		
19	1.00	1.00		
20	2.00	2.00		
21				

Output from Chi-squared:

			winner		Total
			yes	no	
gender	male	Count	7	3	10
		Expected Count	5.0	5.0	10.0
		% within gender	70.0%	30.0%	100.0%
	female	Count	3	7	10
		Expected Count	5.0	5.0	10.0
		% within gender	30.0%	70.0%	100.0%
Total	Count	10	10	20	
	Expected Count	10.0	10.0	20.0	
	% within gender	50.0%	50.0%	100.0%	

From this table we can see that 70% males (compared to 30% females) were winners. But remember this is not cause and effect ie being male does not make you a winner.

The table below tells us that the relationship is not significant ie $p > 0.05$

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.200 ^a	1	.074	.179	.089
Continuity Correction ^b	1.800	1	.180		
Likelihood Ratio	3.291	1	.070		
Fisher's Exact Test					
Linear-by-Linear Association	3.040	1	.081		
N of Valid Cases	20				

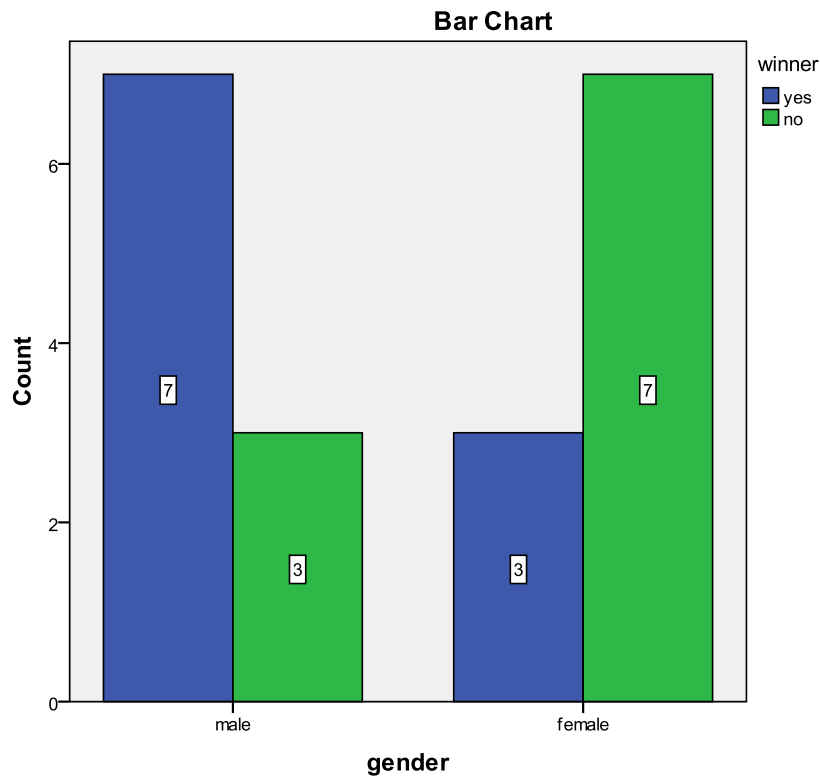
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.00.

b. Computed only for a 2x2 table

In a report this would be expressed as ($\chi^2 = 3.2, df = 1, p = 0.074$).

The p value is however for a 2-tailed test and if we had a directional hypothesis this value would be halved and at $p = 0.037$ the results would be significant.

The clustered bar chart below shows the results in visual format



2. Correlation.

Correlations		vigour	fatigue
vigour	Pearson Correlation	1	-.458**
	Sig. (2-tailed)		.002
	N	42	42
fatigue	Pearson Correlation	-.458**	1
	Sig. (2-tailed)	.002	
	N	42	42

** . Correlation is significant at the 0.01 level (2-tailed).

The table shows that there is a relationship between the fatigue and vigour scores and that this relationship is a negative one (ie the r value is negative). If you look back at the explanation of the scales it says that of the 6 sub-scales, 5 represent negative aspects of mood and that the vigour scale represents a positive aspect of mood. We would therefore expect to see this scale negatively correlated with the other mood subscales.

Compare this to the results below which show a positive correlation ie the r value is +0.561.

Correlations

		tension	fatigue
tension	Pearson Correlation	1	.561**
	Sig. (2-tailed)		.000
	N	42	42
fatigue	Pearson Correlation	.561**	1
	Sig. (2-tailed)	.000	
	N	42	42

The p value identifies that the relationship is significant and the r value indicates the strength of the relationship (range from -1 to +1). The classification of these ranges vary between text books.

You should have a table for the correlation values from your own search.

Scattergrams

