

BMPF7102

Quantitative Research Methods:

Using SPSS for Inferential statistics: some introductory points

Christine Davies

What can SPSS be used for?

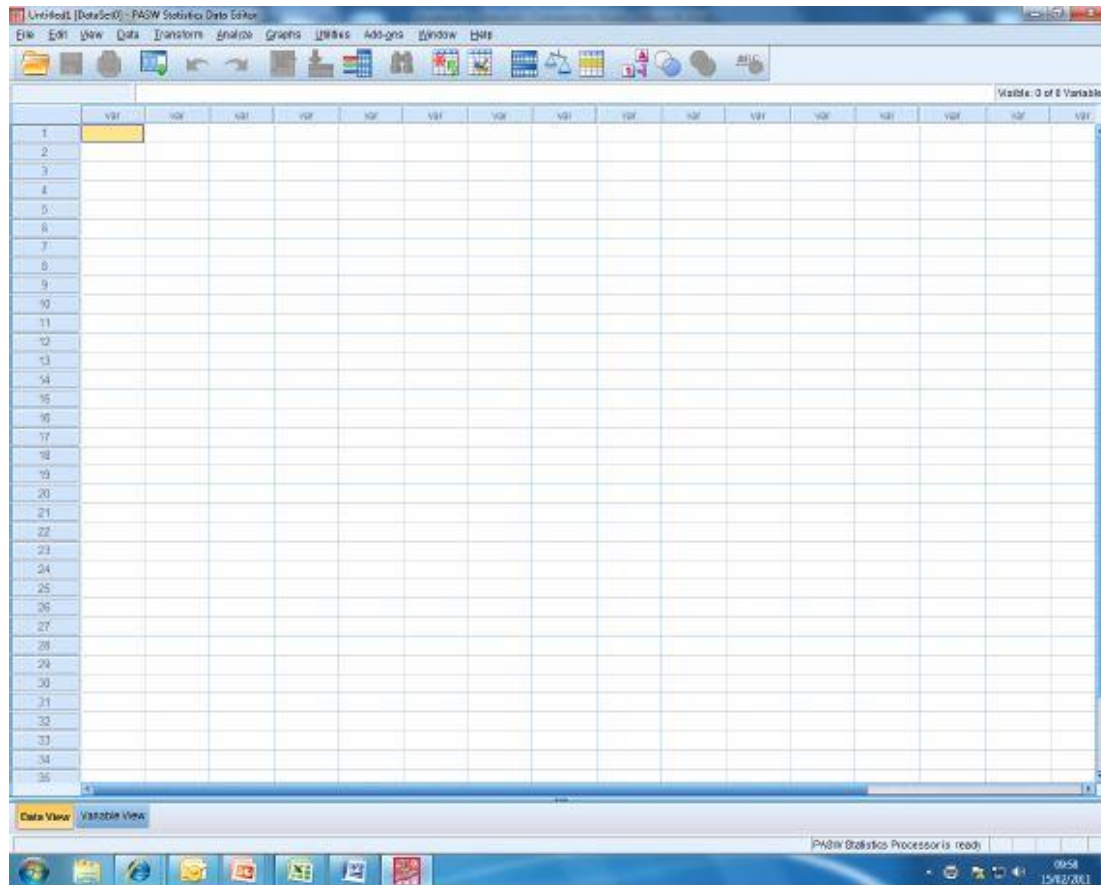
- Statistical Package for the Social Sciences
- Large number of different analyses (descriptive and inferential)
- Categorical, ordinal, and interval/ratio data can be used
- Many analyses can only use continuous interval/ratio data

Preparing for analysis: what should your data look like?

age	Weekly Hrs on internet	Weekly income, £	M/F
22	10	200	M
44	8	400	F
66	7	200	F
18	12	20	F
65	10	150	M
77	6	250	M
33	4	350	M
82	4	150	F
46	7	450	F
49	9	550	M
52	6	300	M

- You will need data *per individual*, not group
- SPSS will provide summaries (though you may need to generate your own version for reports, etc)

SPSS: data entry



- The SPSS screen is like a spreadsheet
- The 'data' view allows for entry of data (copy & paste is possible)
- The 'variable' view let us define variables

demo.sav - Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

20 : age 40

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

Data View Variable View

- The functions in the toolbar allow for summarising or analysing data
- SPSS can undertake almost all forms of analysis

T tests in SPSS

What do you do??

- Add data to SPSS
- Click on 'Analyse', 'compare means', then select the best t test option, such as independent/paired (you will be asked to select the groups you wish to use)

What do you get out?

- Some descriptive statistics (eg. mean, SD) per groups
- A value of 't', the test statistic
- A value of p, the probability ('significance')
- You then have to decide significance levels, and whether to accept or reject the null hypothesis

T test: an SPSS output

Group Statistics

group	N	Mean	Std. Deviation	Std. Error Mean
time male	11	378.6182	44.01930	13.27232
female	11	381.4564	61.29437	18.48095

Independent Samples Test

t-test for Equality of Means					
t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Cor
					Lower
-.125	20	.902	-2.83818	22.75302	-50.3
-.125	18.148	.902	-2.83818	22.75302	-50.6

SPSS starts by giving a table of means and other descriptive statistics.

Another table shows the value of **t**, **df**, and the 'significance' (actually, probability), **p** (generally 2 tailed unless you direct differently).

- Here, the 'significance', ie. p value is 0.902
- This means we need to accept the null hypothesis

The T test: parametric and non-parametric versions

parametric

- Independent T test
- Dependent/paired T test

Non-parametric

- Mann-Whitney test
- Wilcoxon test

One-way ANOVA in SPSS

What do you do??

- Add data to SPSS
- Click on 'Analyse', 'compare means', then select 'one-way anova' (you will be asked to select the groups you wish to use)

What do you get out?

- A value of **F**, the test statistic
- A value of p , the probability ('significance')
- You then have to decide significance levels, and whether to accept or reject the null hypothesis

One-way ANOVA: an SPSS output

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1294.481	5	258.896	6.993	.000
Within Groups	2295.532	62	37.025		
Total	3590.013	67			

SPSS shows the value of 'sum of squares', df, F statistic, and the significance/probability, p

- The significance (ie. p value) here is smaller than 0.001 (it is very small!)
- We would accept the alternate hypothesis

Chi-squared in SPSS

What do you do??

- Add data to SPSS
- Click on 'Analyse', 'descriptive statistics', then select 'crosstabs' (you will be asked to select the variables you wish to use, and tick Chi squared)

What do you get out?

- A table summarising the data
- A value of χ^2
- A value of p, the probability ('significance')
- You then have to decide significance levels, and whether to accept or reject the null hypothesis

Chi-squared: an SPSS output

personality type * favourite colour Crosstabulation

			favourite colour				Total
			red	yellow	green	blue	
personality type	introvert	Count	20	6	30	44	100
		% within favourite colour	10.0%	15.0%	37.5%	55.0%	25.0%
	extrovert	Count	180	34	50	36	300
		% within favourite colour	90.0%	85.0%	62.5%	45.0%	75.0%
Total		Count	200	40	80	80	400
		% within favourite colour	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	71.200 ^a	3	.000
Likelihood Ratio	70.066	3	.000
Linear-by-Linear Association	69.124	1	.000
N of Valid Cases	400		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.00.

SPSS shows a summary table, and a table giving Chi-squared value, df, and the significance, p (2-sided)

- Here, the significance/probability is smaller than 0.001 (ie. very small!)
- We would accept the alternate hypothesis

Correlation in SPSS

What do you do??

- Add data to SPSS
- Click on 'Analyse', 'correlate', then select 'bivariate' (you will be asked to select the variables you wish to use). You also need to indicate Pearson(parametric) or Spearman (non-parametric)

What do you get out?

- A table summarising descriptive statistics (if required)
- A value of 'R' the correlation coefficient
- A value of p, the probability ('significance')
- You then have to decide significance levels, and whether to accept or reject the null hypothesis

Correlation: an SPSS output

Descriptive Statistics

	Mean	Std. Deviation	N
exam	61.0000	10.96406	20
study	39.1500	9.07440	20

Correlations

		exam	study
exam	Pearson Correlation	1	.821**
	Sig. (2-tailed)		.000
	N	20	20
study	Pearson Correlation	.821**	1
	Sig. (2-tailed)	.000	
	N	20	20

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

SPSS shows descriptive stats, and then correlation coefficients, sample number, and the significance, p

- Here, the significance/ probability is smaller than 0.001
- We would accept the alternate hypothesis
- The correlation coefficient is 0.821, indicating a strong positive relationship between variables

Regression: the practicalities

What do you do??

- Check your data are related in a linear way, with no 'outliers', and with similar variances
- Add data to SPSS
- Click on 'Analyse', 'regression', then select 'linear'. You will be asked to select the independent and dependent variables

What do you get out?

- A table summarising some key statistics including R, the correlation coefficient
- An ANOVA table including a value of p, the probability ('significance')
- You then have to decide significance levels, and whether to accept or reject the null hypothesis

Regression: an SPSS output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.873 ^a	.762	.749	874.779

a. Predictors: (Constant), Income

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	44182633.37	1	44182633.37	57.737	.000 ^b
	Residual	13774291.07	18	765238.393		
	Total	57956924.44	19			

a. Dependent Variable: Price

b. Predictors: (Constant), Income

SPSS shows the correlation coefficient, R, and an ANOVA with significance, p

- Here, significance/probability is smaller than 0.001 (ie. very small!)
- We would accept the alternate hypothesis

Exploratory factor analysis

- Questions in surveys etc ask about particular issues but may not get at the main causal factor(s)
- For example, responses relating to lack of sleep, anxiety etc, may relate to a factor that wasn't asked about such as discontent at work
- Exploratory factor analysis looks at **correlations** between variables
- It effectively **reduces the number of factors/variables**, and identifies new 'components' or factors



Factor Analysis in SPSS

What do you do??

- Add data to SPSS
- Click on 'Analyse', 'dimension reduction', then select 'factor'. You will be asked to select variables
- You also need to select 'extraction method': 'principal components' is usually selected
- Other items are selected based on sample size, and presence of correlating factors
- See:
<https://www.youtube.com/watch?v=ajvplACCyd4>

What do you get out?

- Many things!
- A table summarising some key statistics including correlation coefficients
- A series of 'eigenvalues' which explain how much of the observed variance is due to a particular factor (also available as a 'scree plot') – these indicate which 'new' components are most important
- Match of question to component