Instructions for SPSS data analyses
3122 A2 data

Winnifred Louis
w.louis@psy.uq.edu.au
PSYC3122 Materials – April 2009 – Week 7

1. Double click on your file "family name 1.sav" to open it.

2. Save your file as "analyses.sav" (or another name) on your stick so that you have an intact copy of the original data and are working on a separate file which if you over-write it will not be a disaster. NEVER SKIP THIS STEP.

3. Go through the data analyses (below). You will first take a look at the data for errors, then conducting a series of correlation tests and examine the means and standard deviations for your preliminary analyses, and finally conduct a standard multiple regression for Hypotheses 1-4. It's all easy, but people tend to rush through and make errors – be slow & steady.

4. At the end of the day, do not forget to re-save the data file and syntax file and copy them to your stick and/or e-mail to yourself.

Data checking

1. Double click the file – it should open in SPSS. Click on the "Data View" tab at the bottom of the screen on the left hand side. The file should look something like this:

posted a2.sav [DataSet1] - SPSS Data Editor

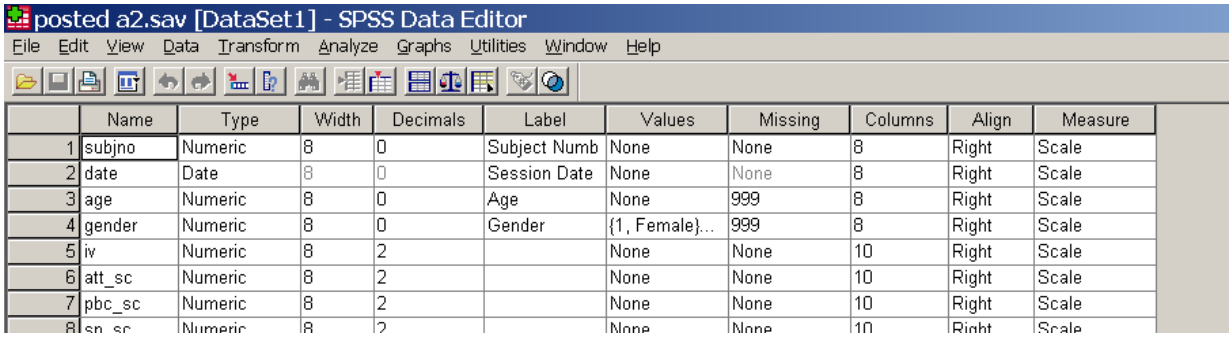
File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : subjno | 1

	subjno	date	age	gender	iv	att_sc	pbc_sc	sn_sc	int_sc	var
1	1	07.08.07	20	1	-2.60	2.67	2.50	2.40	1.40	
2	2	08.08.07	20	1	.20	2.83	1.50	2.37	2.60	
3	3	08.08.07	21	1	-1.93	2.83	1.00	.53	2.60	
4	4	08.08.07	19	1	1.20	2.67	2.25	2.37	2.60	
5	5	08.08.07	19	1	-1.20	2.83	1.75	1.20	1.00	
6	6	07.08.07	18	1	-1.60	.17	.25	.50	1.00	
7	7	07.08.07	19	2	-1.60	.17	.50	-.90	-1.80	
8	8	07.08.07	18	2	-1.47	3.00	3.00	2.60	3.00	
9	9	08.08.07	21	1	-1.53	2.83	2.75	1.37	2.60	
10	10	07.08.07	21	1	-2.27	2.33	-.25	-.43	.40	
11	11	07.08.07	24	2	-1.73	2.17	-.50	-.97	2.00	
12	12	07.08.07	20	2	-1.73	1.33	1.75	-.27	.00	
13	13	14.08.07	18	1	-.87	2.33	1.00	-.27	.60	
14	14	14.08.07	19	1	-.93	3.00	2.75	1.10	1.60	
15	15	07.08.07	20	2	-1.53	1.67	.25	-.50	-1.20	
16	17	08.08.07	19	2	-1.93	2.83	2.50	-.73	1.60	
17	18	08.08.07	20	1	-.33	2.50	2.00	.33	1.20	
18	19	01.08.07	21	2	-1.60	1.67	1.75	-.03	.60	
19	20	08.08.07	21	2	-1.20	2.00	.50	-.90	1.00	
20	21	08.08.07	21	1	-1.27	2.67	1.50	1.97	1.00	
21	22	06.08.07	19	1	-1.27	.33	-1.00	-.57	-2.00	
22	23	09.08.07	19	1	-.87	1.83	2.00	.50	1.00	
23	24	06.08.07	19	1	-1.27	2.50	.50	.97	1.20	
24	25	06.08.07	21	1	-1.93	2.00	.75	-1.07	.40	
25	26	06.08.07	19	2	-3.00	1.00	2.50	-.23	1.00	
26	27	06.08.07	20	1	-.80	2.83	1.25	.77	1.00	
27	28	06.08.07	18	2	-1.53	2.83	.50	2.03	-.40	
28	29	07.08.07	21	1	-1.60	2.83	2.50	.40	1.00	
29	30	06.08.07	42	1	-.87	2.50	1.75	.97	1.00	
30	31	06.08.07	20	1	.00	2.83	.75	1.97	2.00	
31	32	06.08.07	18	2	-1.07	2.50	2.25	1.10	1.40	
32	33	01.08.07	22	1	-.93	3.00	2.50	-.23	1.40	
33	34	30.07.07	21	1	-1.93	1.17	-.50	1.00	-1.20	
34	35	06.08.07	19	1	-1.07	2.33	1.50	.53	1.20	

2. At the top you have the name of the variables included in the data set ("subjno" "date" "age" "gender" etc.). Each participant's data is entered on a row. So in the screen above the first participant is aged 20 and is gender =1, with the data collected on August 8th, 2007, for example. Your data set will be slightly different, but you should be able to find participant #, date, age and gender info for the first person.

   Sometimes it's not clear what the variable name means or what the coding is. E.g., is gender=1 male or female? To find out, click on variable view on the second tab at the bottom of the screen.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	subjno	Numeric	8	0	Subject Numb	None	None	8	Right	Scale
2	date	Date	8	0	Session Date	None	None	8	Right	Scale
3	age	Numeric	8	0	Age	None	999	8	Right	Scale
4	gender	Numeric	8	0	Gender	{1, Female}...	999	8	Right	Scale
5	iv	Numeric	8	2		None	None	10	Right	Scale
6	att_sc	Numeric	8	2		None	None	10	Right	Scale
7	pbc_sc	Numeric	8	2		None	None	10	Right	Scale
8	sn_sc	Numeric	8	2		None	None	10	Right	Scale

Put your mouse over the box for the "values" column intersecting with the "gender" row. If you click you will see the values given for the codes (could be anything – often 1 and 2, or 0 and 1, or 1 and -1) for gender. In the example above it appears that 1 is female, 2 is male. What is it in your data? Make a note to yourself.

While you're here, you can notice that some slack person has failed to put in labels for the key variables in the study. Put your cursor in the box at the intersection of the "label" column and the iv row (row 5). Type the name of your IV. Then move the cursor to the row underneath for att_sc – enter "Attitudes". Repeat for perceived behavioural control, subjective norm, and intentions.

Optional: Show tutor.

Save the file!

3. Before doing ANYTHING in the way of statistics, you must always take the time to examine the data. Never skip this step! There's nothing more painful than doing analyses where your data have errors and discovering later the analyses are totally wrong.

Go through the menus to analyze > descriptive > frequencies to get descriptive statistics and histograms for the data.

- Select all the variables and put them in the window using the arrow
- Click on statistics and tick minimum and maximum. NB, for statistics normally you also ask for mean and standard deviation [useful descriptive info one always wants] and median, skewness, and kurtosis [to check for violations of assumptions of normality] but most of you won't know what that is, so for this assignment don't worry about anything but min and max).
- Click on chart and tick histogram
- Click on paste. This will open a syntax window which will look like this:

```
FREQUENCIES
  VARIABLES=iv1 iv2 iv3                    <- "iv1 iv2 iv3" = a list of all your variables
```

```
  /STATISTICS= MINIMUM MAXIMUM
  /HISTOGRAM
  /ORDER=  ANALYSIS .
```

The good thing about having a syntax file is you have a record, which you can save, of what you have done. You can easily go back to it later, if necessary, and it is easy for you or another person to check for errors. In contrast, if you are running commands just using windows, you can easily miss mistakes, and every time you re-run the command you have to start from scratch.

Run the command by highlighting it and clicking on the arrow in the tool bar (or going "control R"). Look at the histogram and at the minimum and maximum for each variable. You can see if you made a mistake and there are crazy #s there (e.g., a score of 77 on a 1-7 scale).

In this case, there are no errors, so you can go ahead and do the preliminary analyses below. But there's also a few things we can note.

1. Age and gender are unbalanced (lots of women and 19-20 year olds, as usual with psych data sets). We can't do anything about that but you could note it down for later discussion. (When I'm doing analyses, I often create a separate page of notes or word file of "things to talk about in the discussion"). The exact #s of men and women, etc. are also different than you reported in your method section originally – but that doesn't matter for Assignment 2.

2. In the method section of Assignment 1 we said the scales were measured with response scales from 1 to 7. But some data here might be measured using -3 to +3 response formats, or some other scale even (depending on which data set you have). The end points of the scale don't make any difference at all to correlations or betas (they are "scale free" statistics). But NB, when you are commenting on the means for the overall scales (see below), you'll want to interpret means below the midpoint of the variable as showing low levels of the variable, on average, and means above the midpoint as showing high levels of the variable, overall. If we are using 1 to 7 end points, 4 would be the midpoint of the scale. For scales -3 to +3, if your mean is over zero, most people are high on this variable overall.

3. Some of you might have imported or created scales for your IV that had other end points, e.g. 0 to 4, or 1 to 9. For Assignment 2, ignore those endpoints and assume you measured your IV on a scale as it is in your data set. Overall, are your participants low or high or around the middle on the IV? If they're low or high on any variable, you might want to make a note comment on that in the discussion.

4. Also, in the method section, we had multiple items for each scale. In a full data set, you would have separate columns for each item response (e.g., 6 6 5) plus a final column for the average score (e.g., 5.33). Here we just have the final average score for each scale. This is to minimise brain drain for people who haven't done stats by avoiding reliability checking, and scale creation / averaging.

Preliminary Data Analyses

Using the menus, go Analyze > Correlations > Bivariate

Enter the iv, attitude, norm, control and intentions variables in the box using the arrow. Click on "options" and tick "means and standard deviations". While you're there, change the missing values from pairwise to listwise. Then press continue and then press paste. In the syntax file, it should look like this:

```
CORRELATIONS
```

```
/VARIABLES=iv att_sc pbc_sc sn_sc int_sc
/PRINT=TWOTAIL NOSIG
/STATISTICS DESCRIPTIVES
/MISSING=LISTWISE .
```

If you now run this syntax, you will find the info needed to create Table 1 and fill in the preliminary analyses part of your results section.

Just a couple of points:

1. The full correlation table in SPSS has 2 diagonal triangles of #s that are exactly the same (e.g., the correlation between IV and attitudes is the same as the correlation between attitudes and IV). To avoid redundancy, in a report you only fill in one of the diagonals (usually the one below the diagonal – see Table 1, template results file).
2. Watch out about the order of the variables. To encourage thoughtful completion of Table 1 vs mindless cutting and pasting, you'll see that the order of the variables in Table 1 in the template results is slightly different than the one in the output. You can use either order, just be consistent.

Testing Hypotheses 1-4

As just revisited by your tutor, we hypothesise that:

- Hypothesis 1: The more positive someone's attitude to the behaviour, the more likely to intend to act
- Hypothesis 2: The more positive the subjective norm for the behaviour, the more likely to intend to act
- Hypothesis 3: The more positive the perceived control for the behaviour, the more likely to intend to act
- Hypothesis 4: The [higher or lower] the level of [your IV], the more likely to intend to act

We can test these hypotheses using a multiple regression analysis.

Go Analyze > Regression > Linear in the menus.
Enter the intentions scale as the dependent measure using the arrow and the attitudes, norm, and PBC scales as well as your IV scale as independent variables. Then click paste. It should look somewhat like this:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT int_sc
  /METHOD=ENTER iv att_sc pbc_sc sn_sc  .
```

Run the syntax.

It will come up like this:

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	iv_sc, att_sc, pbc_sc, sn_sc[a]	.	Enter

All 4 variables entered – good.

a. All requested variables entered.

b. Dependent Variable: int_sc

You now need to note the model R2 and the F test for the model for your results – the model fit should be significant. (The output you will have has different #s from this example though – and yours is a real data set from last year – so you never know what you'll find!) In addition, you will want to report each independent variable's beta and its p-value.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.974[a]	.950	.942	.09110

a. Predictors: (Constant), iv_sc, att_sc, pbc_sc, sn_sc

Things to report!

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.911	4	.978	117.807	.000[a]
	Residual	.207	25	.008		
	Total	4.119	29			

a. Predictors: (Constant), iv_sc, att_sc, pbc_sc, sn_sc

b. Dependent Variable: int_sc

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.854	.088		9.696	.000
	att_sc	.023	.023	.047	.979	.337
	pbc_sc	.245	.048	.360	5.105	.000
	sn_sc	.219	.066	.435	3.297	.003
	iv_sc	-.911	.069	-1.393	-13.177	.000

a. Dependent Variable: int_sc

Your output will look different from this example output, in that the #s will be different. But the columns should be the same. NB according to theory all the betas for the TPB variables should be positive and significant. The IV should also be significant – either positive or negative according to your hypothesis.

Congratulations ! You have now completed your analyses!

Optional: Show tutor.

- **Save the syntax.**
- **Save the data file.**
- **E-mail them BOTH to yourself or copy to your stick.**