SPSS Basics
A CFDR Tutorial

(Data used in this tutorial: General Social Survey 1998)

Suppose that you are interested in attitudes about cohabitation. You have chosen two variables from the GSS dataset to use as dependent variables. You have chosen to examine the relationship between a respondent's sex, education, and political views and his or her attitude about cohabiting before marriage.

Dependent Variables:

COHABOK: It is OK for couples to live together without intending to be married.

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

COHABBEF: It is a good idea to live together first before marriage.

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

Independent Variables:

SEX: Respondent's gender

1. Male
2. Female

EDUC: Years of formal education

0 to 20 (8 years of college)
98. Don't know
99. No answer

POLVIEWS: Scale for political view of respondent

1. Extremely liberal
2. Liberal
3. Slightly liberal
4. Moderate, middle of the road
5. Slightly conservative
6. Conservative
7. Extremely conservative
8. Don't know
9. No answer

Blank: Not applicable

SOME BASIC RECODING TASKS

It is often necessary to modify or recode your variables before you begin your analysis. For example, you will notice that responses 98 and 99 are invalid for education, and therefore should not be included in the analysis. The purpose of this tutorial is to explain how to perform some of these common tasks in SPSS.

To Eliminate Response Choices

Ex: Recode EDUC to eliminate responses 98 and 99

EDUC education level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	2	.1	.1	.1
	2.00	5	.2	.2	.2
	3.00	10	.4	.4	.6
	4.00	9	.3	.3	.9
	5.00	8	.3	.3	1.2
	6.00	23	.8	.8	2.0
	7.00	21	.7	.7	2.8
	8.00	82	2.9	2.9	5.6
	9.00	75	2.6	2.6	8.3
	10.00	113	4.0	4.0	12.3
	11.00	138	4.9	4.9	17.2
	12.00	851	30.0	30.0	47.2
	13.00	270	9.5	9.5	56.7
	14.00	350	12.4	12.4	69.1
	15.00	146	5.2	5.2	74.3
	16.00	412	14.5	14.5	88.8
	17.00	86	3.0	3.0	91.8
	18.00	109	3.8	3.8	95.7
	19.00	41	1.4	1.4	97.1
	20.00	69	2.4	2.4	99.6
	98.00	8	.3	.3	99.9
	99.00	4	.1	.1	100.0
	Total	2832	100.0	100.0	

A simple frequency indicates that educational attainment is coded as 98 or 99 for 12 respondents (see the table above). You want to recode these values as missing so that they are not included in statistical tests.

SPSS Command: Select Transform – Recode – Into Different Variable (to create new variable)

1. Select the variable EDUC and select the arrow to move it to the window
2. Under 'Output Variable' – 'Name' type a name for the new variable (Ex: REDUC)
3. Select 'Change' to change the name
4. Select 'Old and New Values'
5. Under 'Old Value' select 'Range'
6. Type '98' through '99'
7. Under 'New Value' select 'System Missing'
8. Select 'Add'
9. Under 'Old Value' select 'All Other Values'
10. Under 'New Value' select 'Copy Old Value(s)'
11. Select 'Add' – Select 'Continue' – Select 'OK'

REDUC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	2	.1	.1	.1
	2.00	5	.2	.2	.2
	3.00	10	.4	.4	.6
	4.00	9	.3	.3	.9
	5.00	8	.3	.3	1.2
	6.00	23	.8	.8	2.0
	7.00	21	.7	.7	2.8
	8.00	82	2.9	2.9	5.7
	9.00	75	2.6	2.7	8.3
	10.00	113	4.0	4.0	12.3
	11.00	138	4.9	4.9	17.2
	12.00	851	30.0	30.2	47.4
	13.00	270	9.5	9.6	57.0
	14.00	350	12.4	12.4	69.4
	15.00	146	5.2	5.2	74.6
	16.00	412	14.5	14.6	89.2
	17.00	86	3.0	3.0	92.2
	18.00	109	3.8	3.9	96.1
	19.00	41	1.4	1.5	97.6
	20.00	69	2.4	2.4	100.0
	Total	2820	99.6	100.0	
Missing	System	12	.4		
Total		2832	100.0		

Another frequency verifies that the coding is correct. The 12 responses of 98 and 99 are now categorized as missing.

To Combine Response Categories

Ex: Combine response categories of POLVIEWS to three categories:

1 = Conservative
2 = Moderate
3 = Liberal

POLVIEWS political view

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	64	2.3	2.3	2.3
	2.00	357	12.6	12.6	14.9
	3.00	351	12.4	12.4	27.3
	4.00	986	34.8	34.8	62.1
	5.00	432	15.3	15.3	77.3
	6.00	415	14.7	14.7	92.0
	7.00	86	3.0	3.0	95.0
	8.00	133	4.7	4.7	99.7
	9.00	8	.3	.3	100.0
	Total	2832	100.0	100.0	

The original 9 response categories may make it difficult to describe how cohabitation attitudes differ by political view. Combining these responses into 3 categories may be helpful.

SPSS Command: Transform – Recode – Into Different Variable

1. Select POLVIEWS from variable list and add to window
2. Under 'Output Variable' – 'Name' type a name for the new variable (Ex: RPOLVIEW)
3. Select 'Change'
4. Select 'Old and New Values'
5. Under 'Old Value' select 'Range' – type 1 through 3
6. Under 'New Value' select 'Value' – type 3 – select 'Add'
7. Under 'Old Value' select 'Value' – type 4
8. Under 'New Value' select 'Value' – type 2 – select 'Add'
9. Under 'Old Value' select 'Range' – type 5 through 7
10. Under 'New Value' select 'Value' – type 1 – select 'Add'
11. Under 'Old Value' select 'All Other Values'
12. Under 'New Value" select 'System Missing' – select 'Add'
13. Select 'Continue" – Select 'OK'

RPOLVIEW

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	933	32.9	34.7	34.7
	2.00	986	34.8	36.6	71.3
	3.00	772	27.3	28.7	100.0
	Total	2691	95.0	100.0	
Missing	System	141	5.0		
Total		2832	100.0		

As you can see, the original response categories were correctly combined into three new categories.

To Create a Dummy Variable

Dummy variables are dichotomous variables in which a value of 1 is assigned to the group of interest and a value of 0 to everyone who is in the other category.

Ex: Create a Dummy Variable for SEX where
1 = Females
0 = Males

Two Different SPSS Commands:

1. Transform – Recode – Into Different Variable

   This command will allow you to assign values of 0 and 1 to the appropriate response categories by indicating the "Old Value" and specifying the "New Value." (The steps will be similar to the EDUC recode example).

2. Transform – Count

   1. Under 'Target Variable' type the name of the new variable (Ex: FEMALES)
   2. Select the variable to change in the variable list and put it in the window by selecting the arrow (Ex: SEX)
   3. Select 'Define Values'
   4. Under 'Value' select 'Value'
   5. Type the value of the response category that you want to be valued as 1 in the dummy variable (Ex: 2 for Females). This tells SPSS to change the values 2 in SEX to values of 1 in the new variable FEMALES. All other values will be changed to 0.
   6. Select 'Add' – Select 'Continue' – Select 'Ok'

SEX gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1232	43.5	43.5	43.5
	2.00	1600	56.5	56.5	100.0
	Total	2832	100.0	100.0	

FEMALES

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1232	43.5	43.5	43.5
	1.00	1600	56.5	56.5	100.0
	Total	2832	100.0	100.0	

The original SEX variable classified males as 1 and females as 2. The created dummy variable (FEMALES) now classifies males as 0 and females as 1.

To Reverse Code

Reverse coding refers to 'flipping' the category values (i.e. 1 becomes 8, 2 becomes 7,…,etc.). Reverse coding is often performed to simplify the interpretation of the relationship between to variables or to prepare variables to be combined into a scale.

Ex: Reverse code COHABOK to match coding of COHABBEF

Two Different SPSS Commands:

1. Transform – Recode – Into Different Variable

   Enter old value and replace it with new value. For example, for five response categories, old value 5 would become new value 1, old value 4 would become new value 2, and value 3 would be copied to remain the same.

2. Transform – Compute

   1. Under 'Target Variable' type a name for the new variable (Ex: REVCHBOK)
   2. Determine the number of response categories in the variable and add one (Ex: 5+1 = 6)
   3. Under 'Numeric Expression' type 6 – select minus – select variable from list (COHABOK) and move it to the window
   4. Select 'OK'

COHABOK ok to live without intention of marriage

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	213	7.5	17.3	17.3
	2.00	329	11.6	26.7	44.0
	3.00	258	9.1	21.0	65.0
	4.00	201	7.1	16.3	81.3
	5.00	230	8.1	18.7	100.0
	Total	1231	43.5	100.0	
Missing	System	1601	56.5		
Total		2832	100.0		

REVCHOK

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	230	8.1	18.7	18.7
	2.00	201	7.1	16.3	35.0
	3.00	258	9.1	21.0	56.0
	4.00	329	11.6	26.7	82.7
	5.00	213	7.5	17.3	100.0
	Total	1231	43.5	100.0	
Missing	System	1601	56.5		
Total		2832	100.0		

These tables portray the frequency distributions of COHABOK before and after reverse coding.

To Create a Scale

Social science research often uses scales as independent and dependent variables. Scales are comprised of 2 or more variables that measure similar constructs. When creating scales make sure variable coding are in the same direction (reverse coding may be necessary) and be sure to set desired responses to system missing.

Ex: Make a cohabitation attitude scale that includes COHABBEF and REVCHBOK.

SPSS Command: Transform – Compute

1. Name the new variable (Ex: COHABSCL)
2. Scroll in the functions list to find the MEAN function – Select it and add it to the window
3. Select the variables that comprise the scale and add them to the window (one at a time) (Ex: REVCHBOK and COHABBEF)
4. Move cursor to end of parentheses – select multiplication sign
5. The expression is multiplied by the number of variables comprising the scale (Ex: 2)
6. Select 'OK'

COHABSCL

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	172	6.1	13.8	13.8
	3.00	44	1.6	3.5	17.3
	4.00	149	5.3	11.9	29.2
	5.00	79	2.8	6.3	35.5
	6.00	221	7.8	17.7	53.2
	7.00	144	5.1	11.5	64.8
	8.00	221	7.8	17.7	82.5
	9.00	96	3.4	7.7	90.2
	10.00	123	4.3	9.8	100.0
	Total	1249	44.1	100.0	
Missing	System	1583	55.9		
Total		2832	100.0		

Here is the frequency distribution of the final cohabitation attitude scale.

Standardizing Variables for a Scale

If the variables in the scale do not have the same response categories, they must first be standardized. To do this:

SPSS Command: Analyze – Descriptive Statistics – Descriptive

1. Move the variables of the scale into the window
2. Select the box at the bottom that says 'Save Standardized Values as Variables'
3. Select 'OK'
4. Now, use these standardized variables, instead of the original variables, in the Compute command to create the scale

---

SOME BASIC BIVARIATE AND MULTIVARIATE ANALYSES

Correlations

SPSS Command: Analyze – Correlate – Bivariate

Ex: Test the correlation between REDUC and COHABSCL

1. Select the variables in the variable list and move them to the window
2. You can choose the type of correlation to perform and whether to use a one or two tailed test of significance
3. Select 'OK'
4. The results appear in the output viewer

Correlations

		COHABSCL	REDUC
COHABSCL	Pearson Correlation	1	.029
	Sig. (2-tailed)	.	.304
	N	1249	1244
REDUC	Pearson Correlation	.029	1
	Sig. (2-tailed)	.304	.
	N	1244	2820

t-Test

Ex: Do attitudes about cohabitation differ for males and females? (Is the mean COHABSCL response different for males and females?)

SPSS Command: Analyze – Compare Means – One Sample t-test (to compare sample mean against a population mean) / Independent Samples t-test (to compare 2 sample groups) / Paired Samples t-test (for couples or for comparing the same people at two different times)

1. This is an independent samples test
2. Choose the dependent variable from the list and move it to the 'Test Variable' window (Ex: COHABSCL)
3. Move the independent variable to the 'Grouping Variable' window
4. Select 'Define Groups'
5. Select 'Use Specified Values'
6. Type the numerical value associated with each group (Ex: 0 and 1)
7. Select 'Continue'
8. Select 'Options'
9. Here you can change the confidence interval
10. Select 'Continue'
11. Select 'OK'
12. The results appear in the output viewer

Group Statistics

	FEMALES	N	Mean	Std. Deviation	Std. Error Mean
COHABSCL	.00	559	6.3238	2.48089	.10493
	1.00	690	5.9826	2.46596	.09388

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
COHABSCL	Equal variances assumed	.967	.326	2.425	1247	.015	.3412	.14071	.06514	.61723
	Equal variances not assumed			2.423	1190.870	.016	.3412	.14080	.06495	.61742

Regression

Ex: Examine the effect of the independent variables on the dependent variable

SPSS Command: Analyze – Regression – Linear

1. Move the dependent variable to the 'Dependent' window (Ex: COHABSCL)
2. Move the independent variables to the 'Independent(s)' window (Ex: FEMALES, REDUC, RPOLVIEW)
3. Select 'OK'
4. The results appear in the output viewer

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	RPOLVIEW, REDUC, FEMALES[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: COHABSCL

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.327[a]	.107	.105	2.33732

a. Predictors: (Constant), RPOLVIEW, REDUC, FEMALES

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	780.189	3	260.063	47.604	.000[a]
	Residual	6501.034	1190	5.463		
	Total	7281.223	1193			

a. Predictors: (Constant), RPOLVIEW, REDUC, FEMALES

b. Dependent Variable: COHABSCL

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.452	.362		12.301	.000
	FEMALES	-.465	.136	-.094	-3.407	.001
	REDUC	9.591E-04	.024	.001	.040	.968
	RPOLVIEW	.994	.085	.322	11.680	.000

a. Dependent Variable: COHABSCL

INPUTTING A RAW DATA FILE INTO SPSS

1. Open the SPSS Data Definition Statement in Notepad or Word
2. Open SPSS – File – New – Syntax
3. Copy the information from Notepad/Word into the SPSS Syntax window
4. In the File statement, type the pathway to the data
5. At the end of the statement type: EXECUTE.
6. Select: Run – All

Any errors will appear in the output window