SPSS BASICS

(Data used in this tutorial: General Social Survey 2000 and 2002)

How to do Recoding

Eliminating Response Categories

Ex: Mother's Education to eliminate responses 97,98, 99;

When we run a frequency table for highest year of school completed for mother

HIGHEST YEAR SCHOOL COMPLETED, MOTHER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	90	1.6	1.6	1.6
	1	3	.1	.1	1.7
	2	31	.6	.6	2.2
	3	67	1.2	1.2	3.4
	4	61	1.1	1.1	4.5
	5	46	.8	.8	5.3
	6	150	2.7	2.7	8.0
	7	73	1.3	1.3	9.3
	8	457	8.2	8.2	17.5
	9	145	2.6	2.6	20.1
	10	250	4.5	4.5	24.6
	11	165	3.0	3.0	27.6
	12	1936	34.7	34.7	62.2
	13	204	3.7	3.7	65.9
	14	406	7.3	7.3	73.2
	15	81	1.5	1.5	74.6
	16	424	7.6	7.6	82.2
	17	29	.5	.5	82.7
	18	93	1.7	1.7	84.4
	19	10	.2	.2	84.6
	20	45	.8	.8	85.4
	NAP	300	5.4	5.4	90.8
	DK	462	8.3	8.3	99.0
	NA	54	1.0	1.0	100.0
	Total	5582	100.0	100.0	

Here is How to Do it in SPSS

1. Select Transform-Recode-Into Different Variable.
2. Select the variable (ex/ maeduc) and select the arrow to move it to the window
3. Under "Output Variable"-"Name" type a name for the new variable (ex/ rmaeduc)

4. Select “Change” to change the name
5. Select “Old and New Values”
6. Under” old Value” select “Range”
7. Type “97 thru 99”
8. Under “New Value” select “System Missing”
9. Select “Add”
10. Under “Old Value” select “All Other Values”
11. Under “New Value” select “Copy Old Values”
12. Select “Add”-Select “Continue”- Select “OK”

recoded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	90	1.6	1.9	1.9
	1.00	3	.1	.1	2.0
	2.00	31	.6	.7	2.6
	3.00	67	1.2	1.4	4.0
	4.00	61	1.1	1.3	5.3
	5.00	46	.8	1.0	6.3
	6.00	150	2.7	3.1	9.4
	7.00	73	1.3	1.5	10.9
	8.00	457	8.2	9.6	20.5
	9.00	145	2.6	3.0	23.6
	10.00	250	4.5	5.2	28.8
	11.00	165	3.0	3.5	32.3
	12.00	1936	34.7	40.6	72.9
	13.00	204	3.7	4.3	77.2
	14.00	406	7.3	8.5	85.7
	15.00	81	1.5	1.7	87.4
	16.00	424	7.6	8.9	96.3
	17.00	29	.5	.6	96.9
	18.00	93	1.7	2.0	98.8
	19.00	10	.2	.2	99.1
	20.00	45	.8	.9	100.0
	Total	4766	85.4	100.0	
Missing	System	816	14.6		
Total		5582	100.0		

How to Combine Response Categories:

Combine response categories of POLVIEWS from 7 categories to 3:

1=Liberal
2=Moderate
3=Conservative

THINK OF SELF AS LIBERAL OR CONSERVATIVE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	EXTREMELY LIBERAL	154	2.8	3.9	3.9
	LIBERAL	451	8.1	11.3	15.2
	SLIGHTLY LIBERAL	444	8.0	11.2	26.4
	MODERATE	1576	28.2	39.6	66.0
	SLGHTLY CONSERVATIVE	599	10.7	15.1	81.1
	CONSERVATIVE	621	11.1	15.6	96.7
	EXTRMLY CONSERVATIVE	130	2.3	3.3	100.0
	Total	3975	71.2	100.0	
Missing	NAP	1393	25.0		
	DK	185	3.3		
	NA	29	.5		
	Total	1607	28.8		
Total		5582	100.0		

Here is how to do this in SPSS:

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

recode political views

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1049	18.8	26.4	26.4
	2.00	1576	28.2	39.6	66.0
	3.00	1350	24.2	34.0	100.0
	Total	3975	71.2	100.0	
Missing	System	1607	28.8		
Total		5582	100.0		

How to Create a Dummy Variable

Dummy variables are dichotomous variables in which the value of "1" is usually assigned to the groups of interest and a value "0" to anything else.

RESPONDENT'S SEX

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALE	2457	44.0	44.0	44.0
	FEMALE	3125	56.0	56.0	100.0
	Total	5582	100.0	100.0	

Example: Create a Dummy Variable for SEX where

1=Females

0=Males

Here is How to Do it in SPSS

1. Transform-Recode-Into Different Variable

   This way is similar to the recode command example above.

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 (in this example SEX)
   3. Select "Define Values"
   4. Under "Value" select "Values"
   5. Type the value of the response category that you want to be valued as 1 in the dummy variable (in this example 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"

Females

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	2457	44.0	44.0	44.0
	1.00	3125	56.0	56.0	100.0
	Total	5582	100.0	100.0	

How to Reverse Code

Sometimes when working with data there needs to be a flipping of categories where the response of "1" becomes and "8" and "2" becomes a "7" and so on. This is often done to a single variable that is going to be included into a scale.

Here is How to Do it in SPSS

We are going to recode the variable FECHLD so that "Strongly Disagree" is now "1" and so on.

MOTHER WORKING DOESN'T HURT CHILDREN

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	STRONGLY AGREE	594	10.6	21.6	21.6
	AGREE	1107	19.8	40.3	61.9
	DISAGREE	784	14.0	28.5	90.5
	STRONGLY DISAGREE	262	4.7	9.5	100.0
	Total	2747	49.2	100.0	
Missing	NAP	2797	50.1		
	DK	29	.5		
	NA	9	.2		
	Total	2835	50.8		
Total		5582	100.0		

<u>Example 1</u>

1. Transform-Recode-Into Different Variables
2. Enter old value and replace it with new value. The Old value of "1" will become a "4", etc…

<u>Example 2</u>

1. Transform-Compute
2. Under "Target Variable" type a name for the new variable (Ex: RFECHLD)
3. Think about the number of response categories in the variable and add one. In this case it would be 4+1=5.
4. Under "Numeric Expression" type 5-select the minus sign-select variable from list (FECHLD) and move it to the window
5. Select "OK"

RFECHLD

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	262	4.7	9.5	9.5
	2.00	784	14.0	28.5	38.1
	3.00	1107	19.8	40.3	78.4
	4.00	594	10.6	21.6	100.0
	Total	2747	49.2	100.0	
Missing	System	2835	50.8		
Total		5582	100.0		

How to Create a Scale

Often times when analyzing data in social science we create scales which are two or more variables that measure similar constructs. In this example there are three questions that tap into gender ideology. When creating scales make sure that the variable coding are in the same direction (see above example for reverse coding) and make sure that the desired responses are missing.

Here is How to Do it in SPSS

1. Transform-Compute
2. Name the new variable (GI)
3. Scroll in the functions list to find the mean function-Select it and add it to the window
4. Select the variable that comprise the scale and add them tot eh window
5. Move cursor to end of parentheses-select multiplication sign
6. The expression in multiplied by the number of variables comprising the scale
7. Select "OK"

GI

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	75	1.3	2.7	2.7
	4.00	68	1.2	2.5	5.2
	4.50	9	.2	.3	5.5
	5.00	141	2.5	5.1	10.6
	6.00	397	7.1	14.3	24.9
	7.00	423	7.6	15.3	40.2
	7.50	32	.6	1.2	41.3
	8.00	451	8.1	16.3	57.6
	9.00	580	10.4	20.9	78.5
	10.00	289	5.2	10.4	89.0
	10.50	14	.3	.5	89.5
	11.00	153	2.7	5.5	95.0
	12.00	139	2.5	5.0	100.0
	Total	2771	49.6	100.0	
Missing	System	2811	50.4		
Total		5582	100.0		

How to Standardize a Variable for a Scale

If the variables in the scale do not have the same response categories, they must first be standardized into z-scores.

How to Do this in SPSS

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

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
GI	2775	-7.66	6.23	-.0141	2.83231
Valid N (listwise)	2775				

SOME BASIC BIVAARIATE AND MULTIVARIATE ANALYSES

How to do Correlations

Sometimes there is a need to determine if two variables are correlated. There are three correlation coefficients that can be used doing this technique.

- **Pearson.** This coefficient is used to determine the strength of the linear association between interval level measures.
- **Kendall's tau-b.** This coefficient is used to determine the strength of the association between variables measured at the ordinal level
- **Spearman.** A rank-order correlation coefficient that also measures association a the ordinal level.

Here is how to Do it in SPSS

1. Analyze-Correlate-Bivariate
2. Select the variables in the variable list and move them to the window
3. Choose the type of correlation to perform and whether to use one or two tailed test of significance
4. Flag significant correlations

Correlations

		HIGHEST YEAR OF SCHOOL COMPLETED	HIGHEST YEAR SCHOOL COMPLETED, MOTHER
HIGHEST YEAR OF SCHOOL COMPLETED	Pearson Correlation	1	-.200**
	Sig. (2-tailed)		.000
	N	5561	5561
HIGHEST YEAR SCHOOL COMPLETED, MOTHER	Pearson Correlation	-.200**	1
	Sig. (2-tailed)	.000	
	N	5561	5582

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

How to do t-Tests

Independent t-tests are a way to compare means of two different samples. There are three assumptions for t-tests.

1. Independence
2. The distributions are approximately normal (histogram)
3. Equal population variances

In order to determine equal variance, SPSS employs the Levene's test which is found on the t-test printout. The Levene's test tests the null hypothesis of equal variance. If the Levene's test is significant then you reject the null that the two samples have equal variance. Fortunately, SPSS gives results for "equal variances assumed" and "equal variances not assumed."

How to Do it in SPSS

1. Analyze-Compare Means-Independent Samples
2. Choose the dependent variable from the list and move it to the "Test Variable" window
3. Move the independent variable tot eh "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"

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
RESPONDENT'S SOCIOECONOMIC INDEX	Equal variances assumed	.401	.527	1.582	5288	.114	.8452	.5341	-.2020	1.8923
	Equal variances not assumed			1.580	5055.991	.114	.8452	.5349	-.2034	1.8938

How to do a One-Way ANOVA (analysis of variance)

Sometimes there needs to be a comparison of three or more means. For example, if we wanted to test whether years of school completed (educ) were significantly different levels of happiness in marriage (hapmar).

How to Do it in SPSS

1. Analyze-Compare Means-One Way ANOVA
2. Choose the dependent variable and place it in the "Dependent Variable"
3. Choose the independent variable and place it in the "Factor List"
4. Click on "Post Hoc"

5. Check whatever statistic you need(ex. Tukey)
6. Click “Continue”
7. Click on “Options”
8. Check the box marked “Descriptives”
9. Click “Continue”
10. Click “OK”

Multiple Comparisons

Dependent Variable: HIGHEST YEAR OF SCHOOL COMPLETED

Tukey HSD

(I) HAPPINESS OF MARRIAGE	(J) HAPPINESS OF MARRIAGE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
VERY HAPPY	PRETTY HAPPY	.579*	.141	.000	.25	.91
	NOT TOO HAPPY	.769	.382	.110	-.13	1.67
PRETTY HAPPY	VERY HAPPY	-.579*	.141	.000	-.91	-.25
	NOT TOO HAPPY	.190	.390	.877	-.72	1.10
NOT TOO HAPPY	VERY HAPPY	-.769	.382	.110	-1.67	.13
	PRETTY HAPPY	-.190	.390	.877	-1.10	.72

*. The mean difference is significant at the .05 level.

How to do Linear Regression

Linear Regression can look at the effects of the independent variables on the dependent variable.

How to Do this in SPSS

1. Analyze-Regression-Linear
2. Move the dependent variable to the “Dependent” Window (ex/ GI)
3. Move the independent variables to the “Independent” Window (Females, childs, educ)
4. Select “OK”

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	HIGHEST YEAR OF SCHOOL COMPLETED, Rsex, NUMBER OF CHILDREN[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: GI

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.296[a]	.088	.087	2.70899

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED, Rsex, NUMBER OF CHILDREN

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1937.165	3	645.722	87.990	.000[a]
	Residual	20188.496	2751	7.339		
	Total	22125.662	2754			

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED, Rsex, NUMBER OF CHILDREN

b. Dependent Variable: GI

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.804	.272		-10.314	.000
	Rsex	1.003	.105	.175	9.579	.000
	NUMBER OF CHILDREN	-.190	.031	-.114	-6.066	.000
	HIGHEST YEAR OF SCHOOL COMPLETED	.194	.018	.200	10.658	.000

a. Dependent Variable: GI

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 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 error will appear in the output window