

The following slides are the property of their authors and are provided on this website as a public service. Please do not copy or redistribute these slides without the written permission of all of the listed authors.

Introduction to SPSS & Defining a Research Question

May 2012

Nancy Burns Ashley Jardina University of Michigan



What is SPSS?

- SPSS: Statistical Product and Service Solutions
- SPSS is a software package used for statistical analysis and data management.
- It can be used to analyze data and produce reports, charts, plots, and descriptive statistics for a wide range of data sources.
- The version of SPSS used in this presentation is SPSS 20.

The SPSS Statistics Main Window

ta Untitled1 [DataSet0] - IBM	SPSS Statistics	Data Editor		-		-		-			
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform /	<u>A</u> nalyze <u>G</u> rap	hs <u>U</u> tilities	s Add- <u>o</u> ns <u>W</u> in	dow <u>H</u> elp						
							💻 🖧 🖡			ABC		
	Nama	Ture	Width	Desimals			Missing	Columna	Alian	Maggurg	Dela	
1	Ivame	туре	vvidtri	Decimais	Laber	values	wissing	Columns	Align	weasure	Role	
2												-
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												-
10												-
17												
10												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
.31	4											▼
Data View	Variable View											
	Tanabic View										IRM SPSS Static	tics Processor is ready
											IDM OF OO OIdUS	acon rocessor is ready

SPSS – Data View and Variable View



"Data View" displays a spreadsheet of the data, with each row as a case and each column as a variable.

"Variable View" displays information about the variables in the dataset, including variable names, types, and any labels associated with the variables.

SPSS – Output View



SPSS Menus and Toolbar



Opening a Dataset

ta Ur	ntitled1	[DataSe	t0] - IBM	SPSS Statisti	cs Data Ed	itor		
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Add- <u>c</u>
	<u>N</u> ew				•		<u>+</u> =	
	<u>O</u> pen				•	🕞 D <u>a</u> t	а	—
	Open D	ata <u>b</u> as	е		•	🔁 Syn	tax	L
	Rea <u>d</u> T	ext Data	i			🕞 Out	put	<u> </u>
20	<u>C</u> lose			Ctrl+F4	4	🔒 S <u>c</u> ri	pt	<u> </u>
	Save			Ctrl+S				
	S <u>a</u> ve As	S						
	Save Al	l Data						
5	Expor <u>t</u> t	o Datab	ase					
	Mar <u>k</u> Fi	le Read	Only					
1	Rena <u>m</u>	e Datas	et					
	D <u>i</u> splay	Data Fi	ile Inform	ation	•			
1	Cac <u>h</u> e I	Data						
•	Stop Pr	oc <u>e</u> sso	Г	Ctrl+P	eriod			
-	S <u>w</u> itch	Server						
	<u>R</u> eposi	tory			•			
	Print Pr	eview						
	Print			Ctrl+P				
	Recent	l <u>y</u> Used	Data		•			
	Recent	ly Used	<u>F</u> iles		•			
	Exit							
	22					-		
	23	1						

Open Dataset in Variable View

					\square				-							
omnibus_20.1_weighte 1_UM.sav [DataSet1] - IBM SPSS Statistics Dat". Editor													on			
						ow <u>H</u> elp	Wind	s Add- <u>o</u> ns	iliti es	hs <u>U</u> ti	Grapi	Analyz	Transform	<u>V</u> iew <u>D</u> ata	Edit	File
												6				
Role	Measure	Λ	Align	Columns	Missing	Values		Labe	als	Decim	Width	1	Туре	Name		
🔪 Input	cale	So So	≡ Right	8	None	None	d	household i		0		12	Numeric	caseid	1	
🔪 Input	ominal	💰 No	≣ Right	8	None	1, Qatari}	ype	household t		0		14	Numeric	household	2	
🔪 Input	ominal	💰 No	≡ Right	8	None	{1, Doha}	1	municipality		0		8	Numeric	munid	3	
S Input	cale	Se Se	≣ Right	8	None	None				0		8	Numeric	zoneid	4	
	cale	So So	≣ Right	8	None	None				0		8	Numeric	zoneid	4	

There are several key pieces of information to note about your data under "variable view" in SPSS.

The "name" column specifies the name of the variable.

The "label" column provides a more detailed description of the variable.

The "missing" column indicates which values for each variable SPSS will ignore when it runs statistics or generates tables.

SPSS - types of variables

(🔄 Variable Type
	 Numeric Comma Characters: 2 Dot Scientific notation Date Dollar Custom currency String
	OK Cancel Help

"Variable View" also indicates the *type* of each variable. If you click on the "type" box in variable view, you will see this window, which allows you to alter a variable's type. The two basic types of variables are **numeric** and **string**. Numeric variables contain only numbers. String variables may contain letters or numbers.

Renaming and labeling a new variable

- Variable f13in in your dataset corresponds to the household income for each respondent. We will use the "Compute" command to create a new variable equal to f13in with a more intuitive name and a clear label.
- Use the *Transform / Compute Variable* menu option.
- Note that we could simply rename f13in in the "Variable View" window, but it is good practice to create new variables rather than manipulating the original variables.

c ΣZ Compute Variable The "Target Variable:" box is Target Variable: Numeric Expression: f13in hhincome where you want to Type & Label.. input the name of Very satisfied wit... your new variable. 🖉 Verv satisifed wit... Function group: 🖉 Very satisfied wit.. All 🖋 Stronaly aaree I h. 8 9 Arithmetic 🖉 Stronaly agree w/ CDF & Noncentral CDF In the example, we 🖉 Stronaly agree/le... 5 6 <= Conversion 🔗 Strongly agree m... used "hhincome" to Current Date/Time 2 3 🖋 Strongly agree w... = Date Arithmetic 🔗 Verv satisfied wit.. stand for household Date Creation & 0 🔗 Normalized weig... 🔗 Number of maids. income. Functions and Special Variables: Delete Nanny01 + cook01 🔗 driver01 Set your new Sear01 🔗 suv01 variable equal to Nruck01 🖉 boat01 the original 🔄 Compute Variable: Type 🖳 🗮 🏹 🔗 f13in variable, which in -Label: (optional case selection condition) this case is f13n. Label: Household income Paste Reset OK Cancel O Use expression as label Type Select "Type & Label" to give your variable a descriptive label, Numeric like "Household income". This label will appear in "Variable String Width: 8 View" as well as in output produced by SPSS when analyzing Continue Cancel Help

your data.

Renaming and labeling a new variable

String

Continue

Width: 8

Cancel

Help

Compute Variable	W
Larget Variable: Numeric Expression: education = Type & Label f15c	th
Function group:	re
Image: Constraint of the second se	
Image: Construction of the construc	re
 ♣ f18 ♣ f18a ♣ f18b ♣ f18c 	le
A f18c7 A f19 F19a Compute Variable: Type	ec
Label	education.
□ Use expression as □ Use expression as □ Use expression as	label

We can repeat the procedure to rename variable f15c, which corresponds to respondents' level of education.

Renaming and labeling a new variable

🕼 Compute Variable		
Target Variable: gender Type & Label Image: Constraint of the second	= Numeric Expression: f1 +<<<>> 7 8 +<<<>> 7 8 9 -<<=>< 4 5 6 = = 1 2 / & 1 0 . ** ~ () Delete ✓ Functions and ✓ Functions and .	ntral CDF Time ic Special Variables:
(optional case sele	ion condition) OK <u>Paste R</u> eset Cancel Help	© <u>L</u> abe © <u>U</u> se Type © <u>N</u> um © <u>S</u> trin

We can repeat the procedure to rename variable f1, which corresponds to the respondents' gender.



13

Frequencies

 The frequencies command provides a useful summary of a given variable. It indicates the number of responses that correspond to each value of a variable.

 Use the Analyze / Descriptive Statistics / Frequencies menu to find the frequencies of a variable.



Statistics

respondent's gender

N	Valid	2000
	Missing	854

Using the frequencies command, we can see that there are 1020 male and 980 female survey respondents in our dataset.

respondent's gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	1020	35.7	51.0	51.0
	female	980	34.3	49.0	100.0
	Total	2000	70.1	100.0	
Missing	System	854	29.9		
Total		2854	100.0		

Cross Tabulation

- A cross tab shows the relationship between two variables.
- It is a table representing the number of survey respondents falling into each possible pairing of responses to two survey questions.
- Use the Analyze / Descriptive Statistics / Crosstabs menu option to run a crosstab.

Cross Tabulation



	Cases									
	Va	lid	Miss	sing	Total					
	Ν	Percent	N	Percent	Ν	Percent				
gender * Highest level of education completed	1577	55.3%	1277	44.7%	2854	100.0%				

Case Processing Summary

gender * Highest level of education completed Crosstabulation

Count

			Highes	st level of ed	ucation com	pleted		
		1.00	2.00	3.00	4.00	5.00	6.00	Total
gender	1.00	248	53	66	379	52	18	816
	2.00	237	38	73	366	41	6	761
Total		485	91	139	745	93	24	1577

A cross tabulation of education by gender will show the number of men and women that correspond to each level of education in the dataset.

Cross Tabulation – Gender and Income



Case Pr	ocessing	Summary
---------	----------	---------

	Cases								
	Va	lid	Miss	sing	Total				
	Ν	Percent	Ν	Percent	Ν	Percent			
gender * Household income	1819	63.7%	1035	36.3%	2854	100.0%			

gender * Household income Crosstabulation

Count

		Household income									
		1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	Total
gender	1.00	223	200	128	63	62	79	71	56	60	942
	2.00	154	202	126	60	66	94	83	59	33	877
Total		377	402	254	123	128	173	154	115	93	1819

Exercise

- Rename variable f15c, which corresponds to the survey respondent's level of education.
- Rename variable f1, which indicates the gender of the survey respondent.
- Run a frequency on your new education variable. Run a frequency on f15c. Compare the two variables to make sure they are the same.
- Run two cross tabs: Gender and Education and Gender and Income.

The Process of Designing a Survey Instrument: Defining your question and outlining your analysis

A Good Research Question...

- Identifies both a cause and an effect.
- Makes an explicit comparison.
- States the nature of the relationship, not just that two things are related, but how those two things are related.
- Is specific.

Developing a strong question: From what your study is about to what you expect

- Bad Example: I want to study student achievement.
- Better: I want to study the relationship between parental expectations and student achievement.
- Even Better: Student achievement depends, in part, on parental expectations about academic achievement. The higher the parental expectations for achievement, the higher the student achievement.

What alternative – or rival – explanations do you need to consider?

--Teacher expectations?

--Teacher training?

--The number of students in the class?

--Others?

- Define the precise analysis you need to carry out to answer your research question.
- Define the precise analyses you will need to carry out to rule out rival or alternative explanations.
- Sketch all of the tables you will need to make your case.

Example Tables

Table 1. Percentage of Qataris and White Collar Expatriates who report having a driver's license

	Qataris	White Collar Expatriates
Have Driver's License	57.9%	65.6%
Do Not Have Driver's License	42.1%	34.4%
Total N	992	1008

 χ^2 (1, N=2000) = 12.6, p = .00

Source: 2011 SESRI Omnibus Survey

Table 2. Average number of days in a week respondents send text messages and emails by householdtype and gender

	Qatari Men	Qatari	t-test	White Collar	White Collar	t-test
	(N=469)	Women		Men	Women	
		(N=518)		(N=546)	(N=458)	
Text	3.72	3.28	2.38*	3.35	2.89	2.621**
Messages	(2.88)	(2.95)		(2.89)	(2.68)	
Emails	3.06	2.35	3.97**	4.48	3.27	6.63**
	(2.87)	(2.76)		(2.87)	(2.91)	

**P<.01 *p<.05

Note: Numbers in parentheses are standard deviations.

Source: 2011 SESRI Omnibus Survey

Qatari	No parent in household	Parent is head of household	Lives with Parent, but Parent is not head of household
Men	34%	47%	19%
Women	53%	31%	16%
Total	878	739	344

Chi-squared (2, 1961)=73.20; p<.000. Source: 2010 SESRI Omnibus Survey.

After doing all of these things very precisely...

Then you're ready to design items for an instrument.

These items will

measure the concepts in your analysis, and measure the concepts in the alternative explanations.

Thought Exercise

- Define a research question, with all of the elements we talked about.
- Specify the cause and the effect.
- What are plausible alternative explanations?
- What are the tables you'd like to be able to include? What variables are on those tables?
- What concepts will you need to measure in your survey, in order to carry out the analyses?