



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

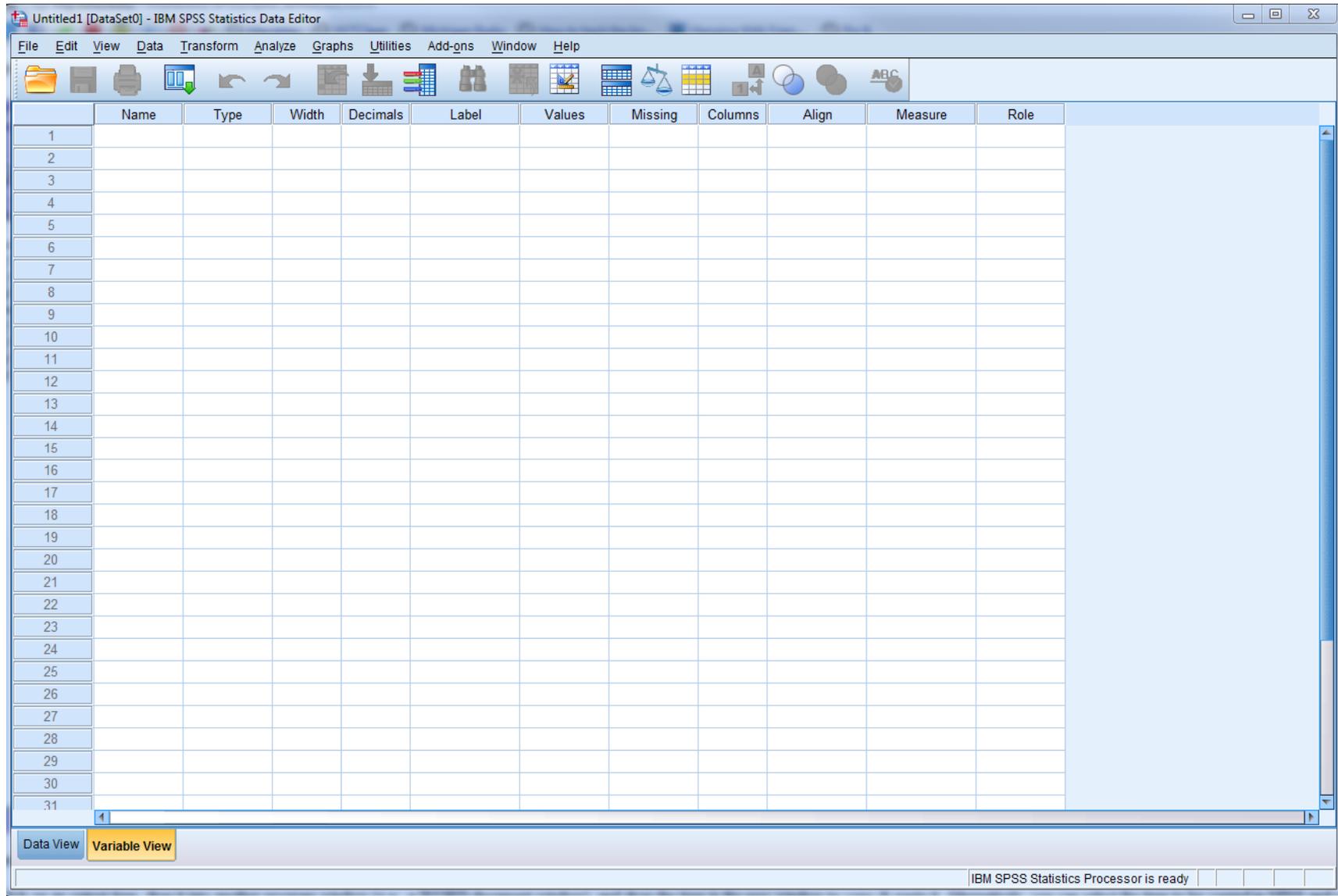


معهد البحوث الاجتماعية والاقتصادية المسحية
Social & Economic Survey Research Institute

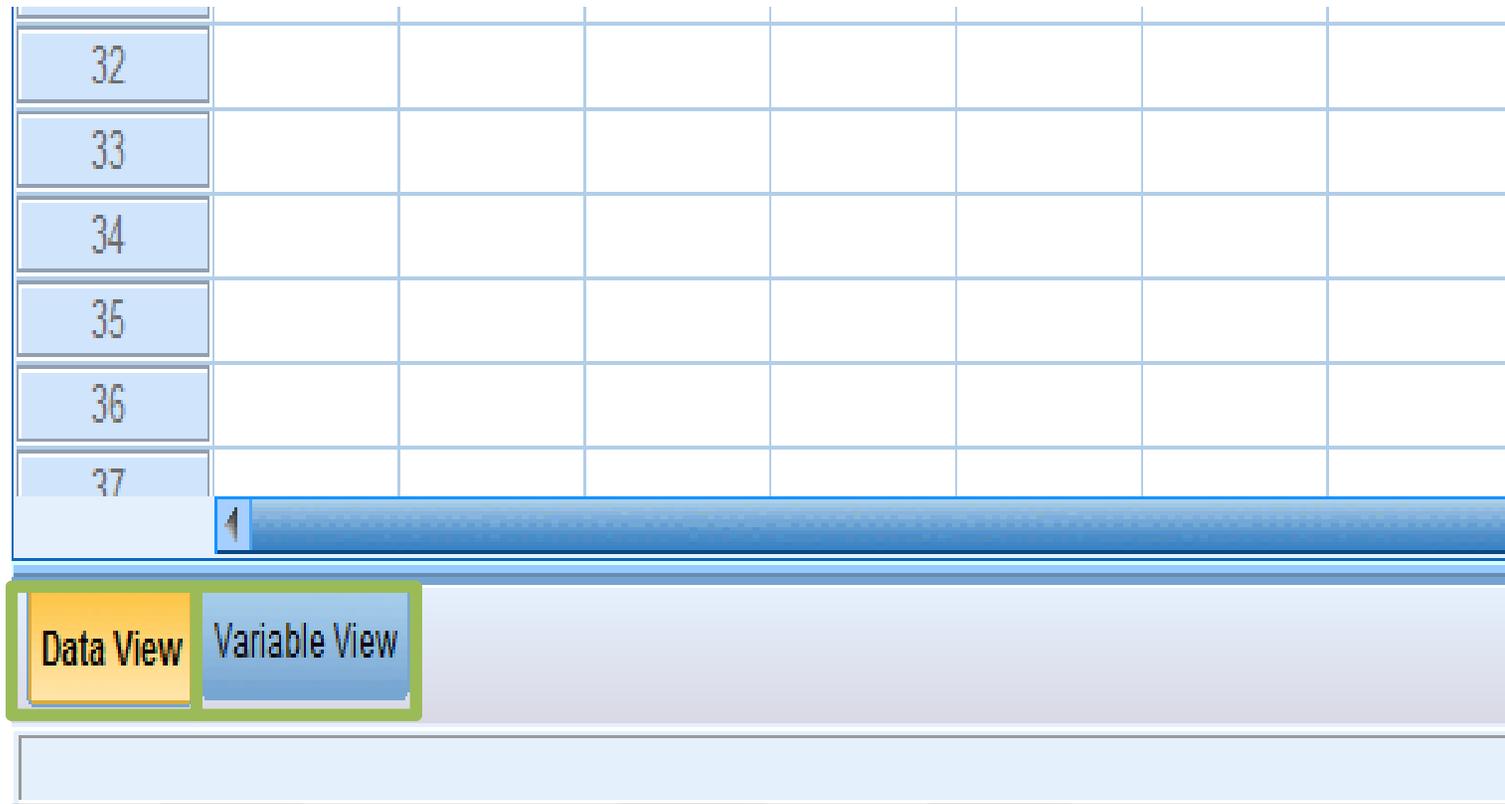
What is SPSS?

- 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



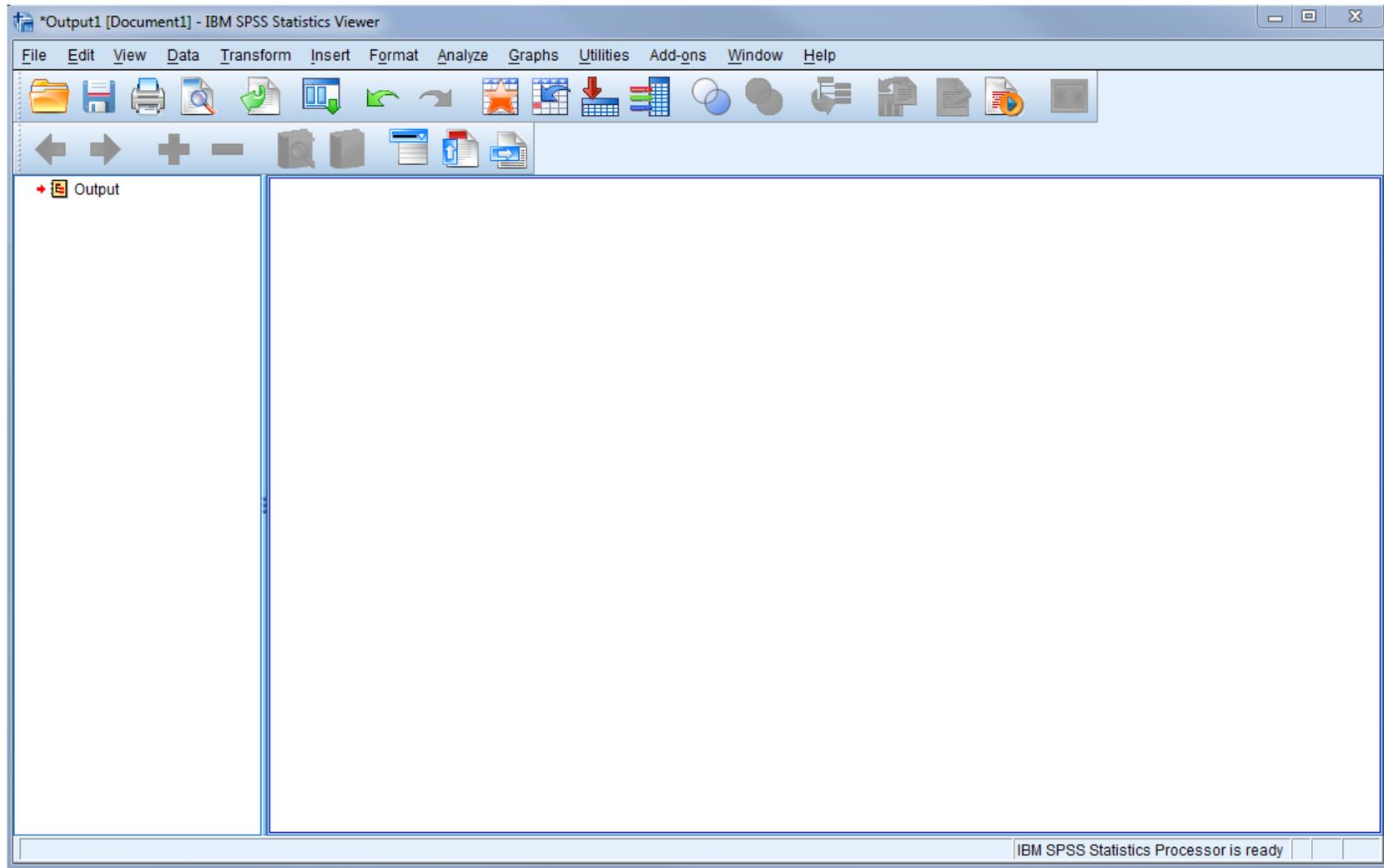
SPSS – Data View and Variable View



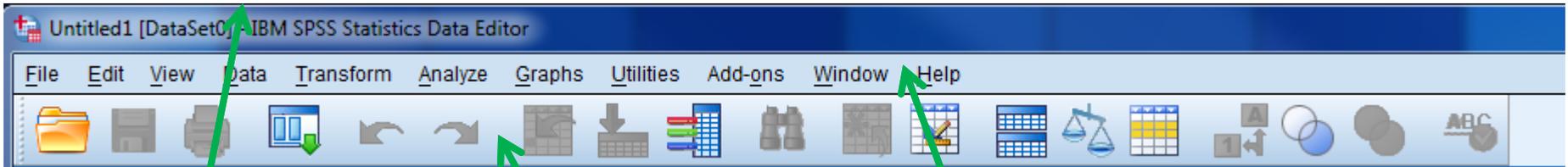
“Data View” displays a spreadsheet of the data, with each row as a case and each column as a variable.

“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

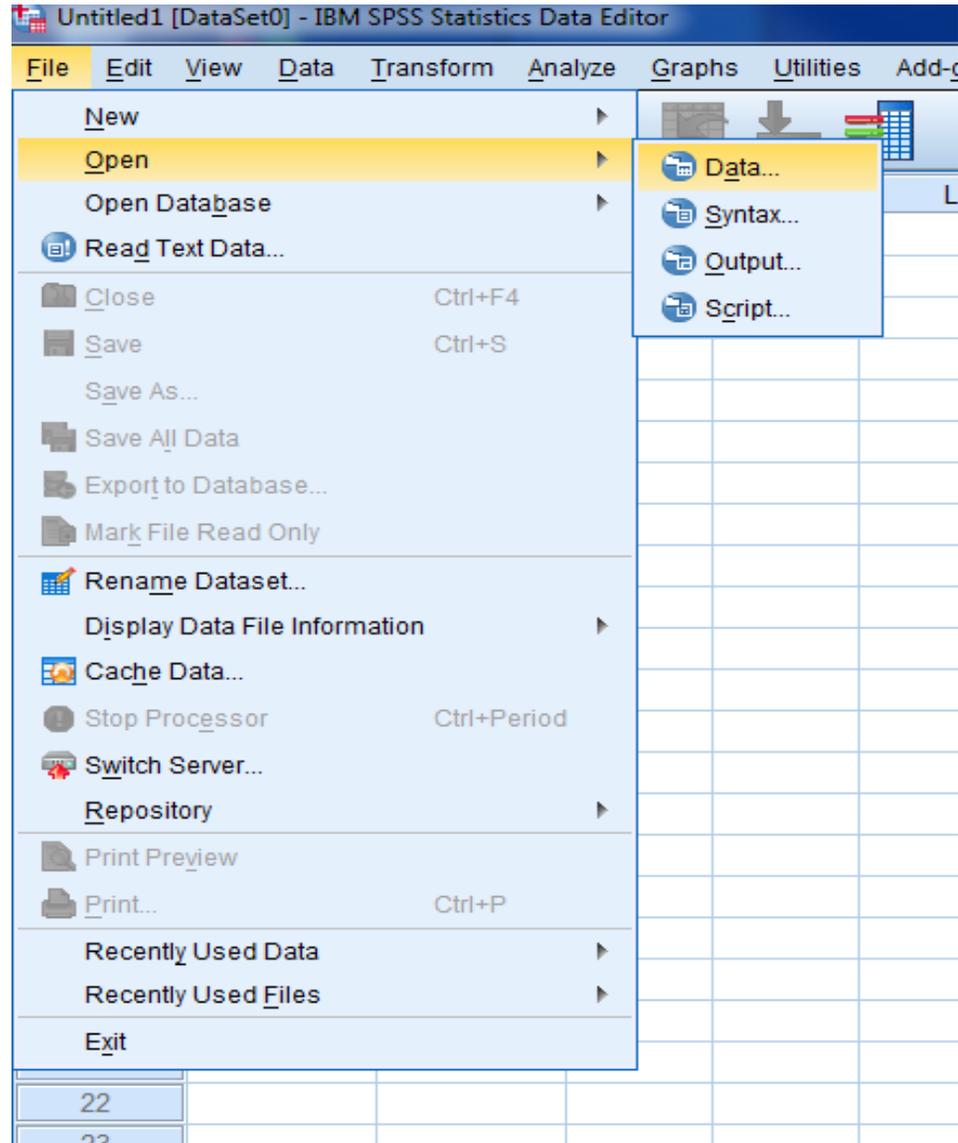


Title Bar

Tool Bar

Menu Bar

Opening a Dataset



Open Dataset in Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	caseid	Numeric	12	0	household id	None	None	8	≡ Right	Scale	Input
2	household	Numeric	14	0	household type	{1, Qatari}...	None	8	≡ Right	Nominal	Input
3	munid	Numeric	8	0	municipality	{1, Doha}...	None	8	≡ Right	Nominal	Input
4	zoneid	Numeric	8	0		None	None	8	≡ Right	Scale	Input

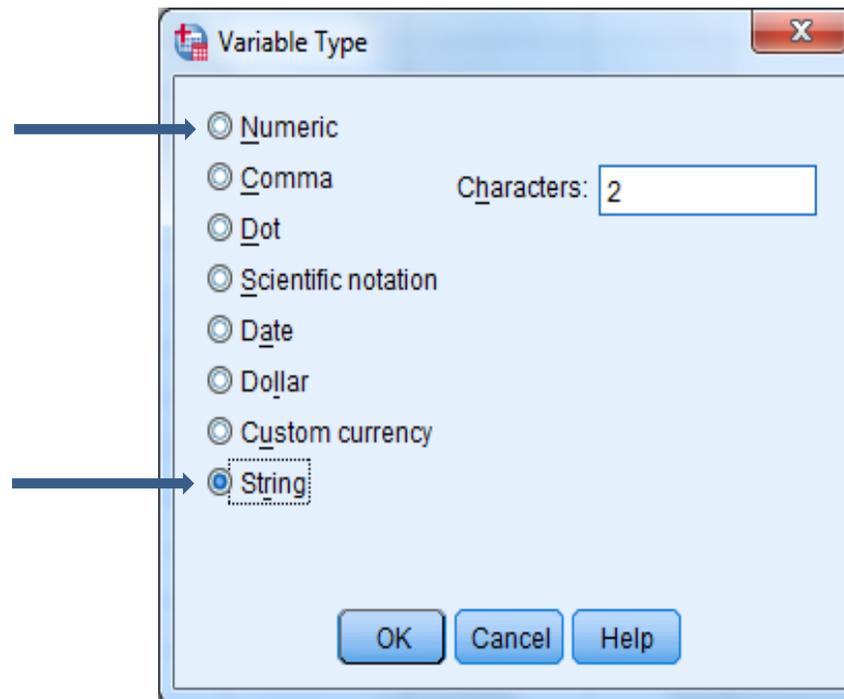
There are several key pieces of information to note about your data under “variable view” in SPSS.

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 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.

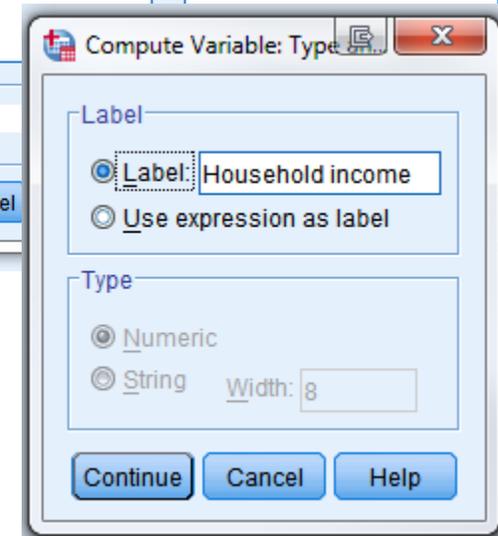
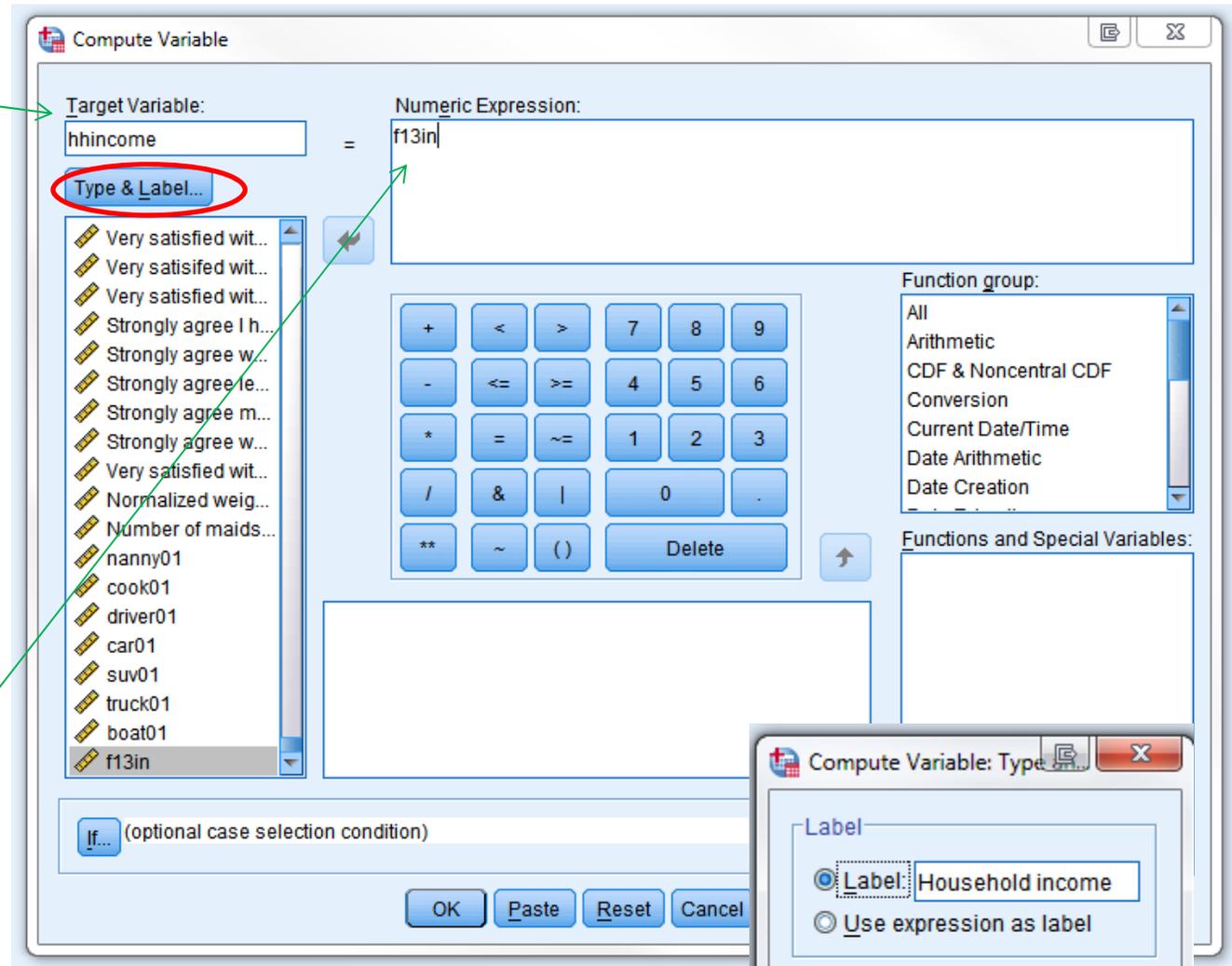
Renaming and labeling a new variable

- 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.

The “Target Variable:” box is where you want to input the name of your new variable.

In the example, we used “hhincome” to stand for household income.

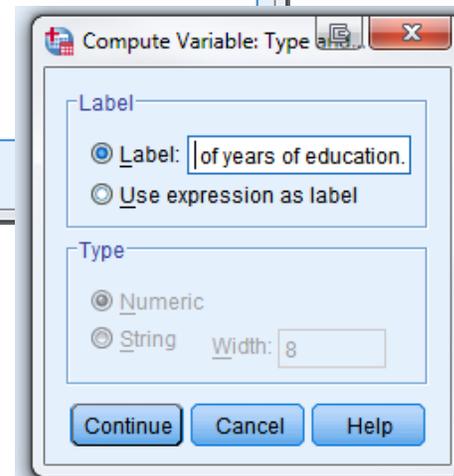
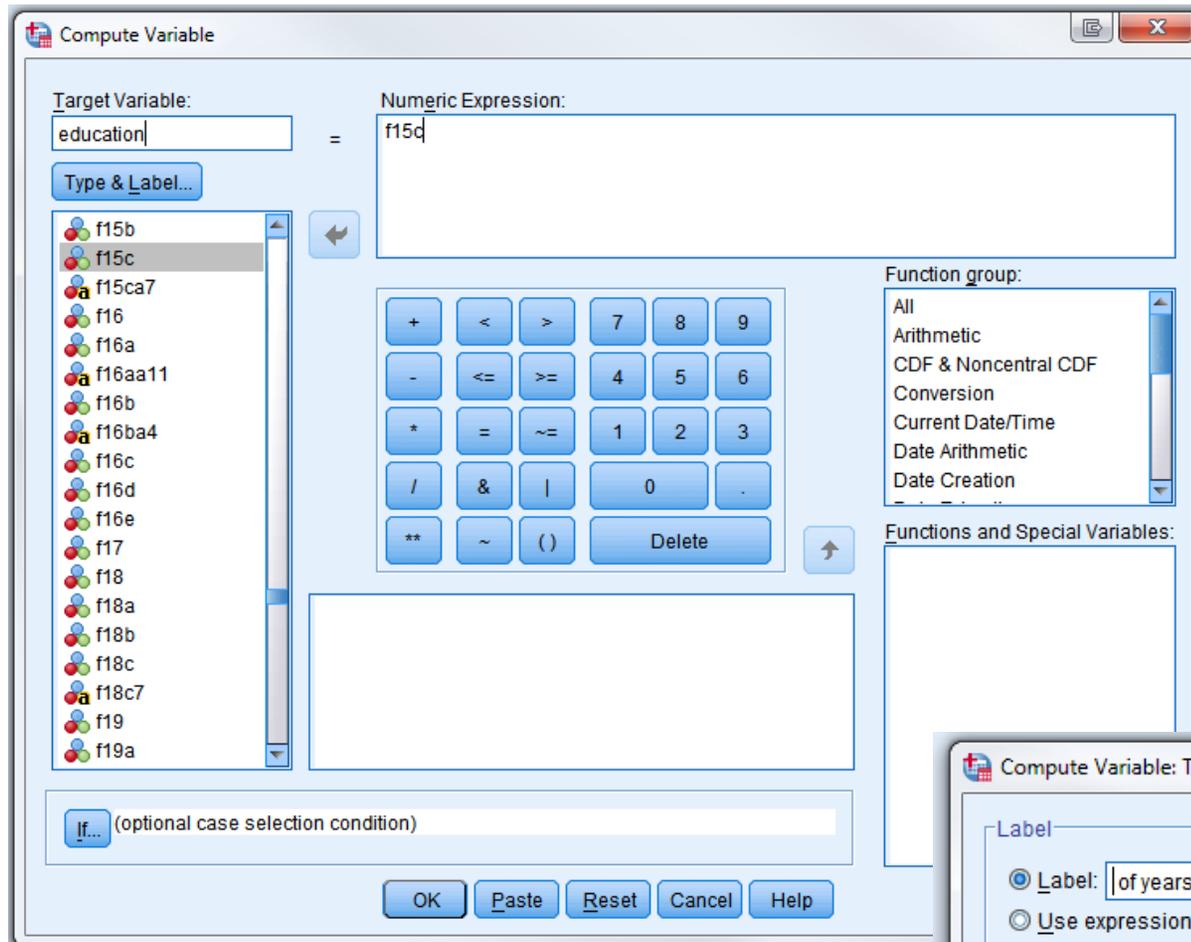
Set your new variable equal to the original variable, which in this case is f13n.



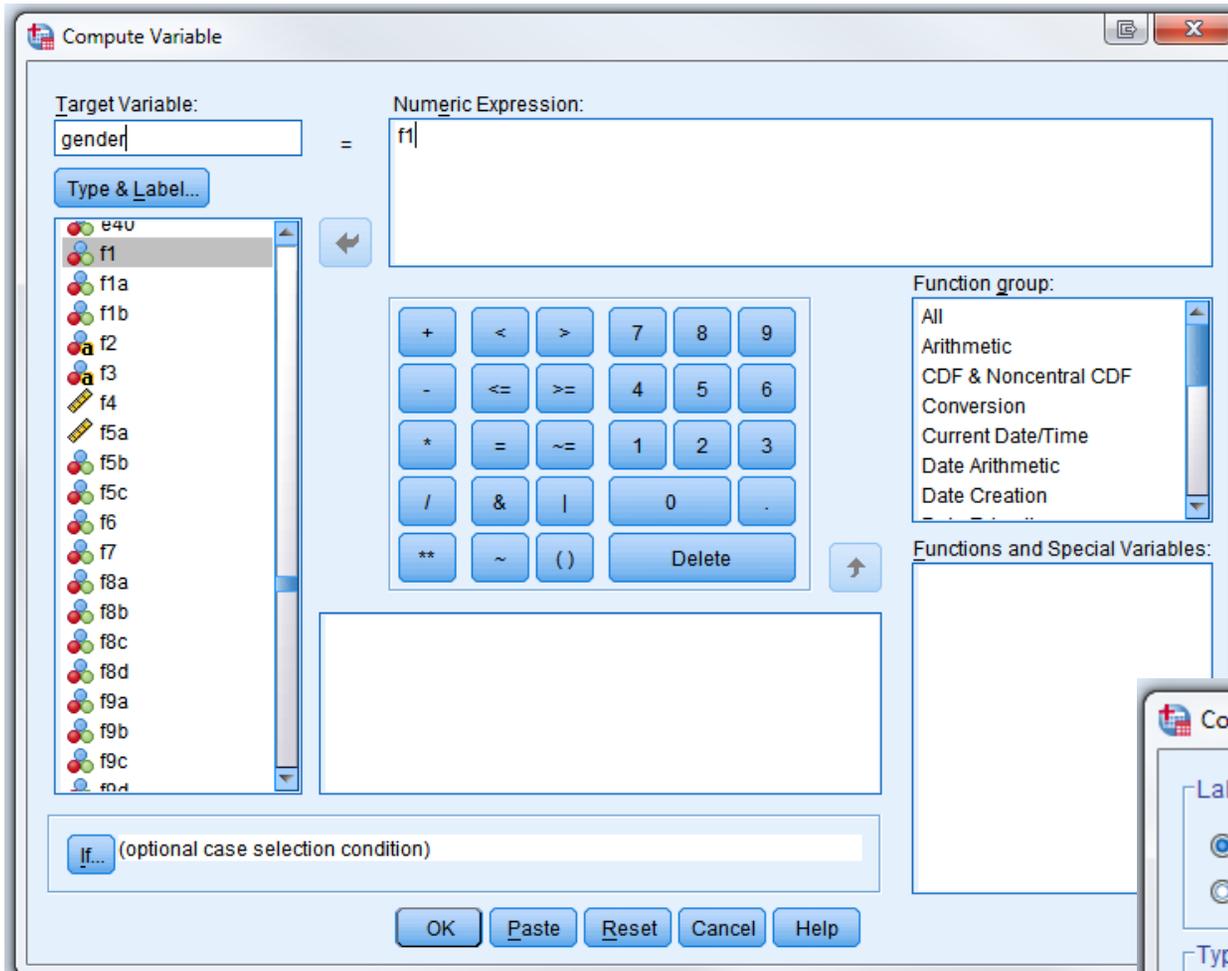
Select “Type & Label” to give your variable a descriptive label, like “Household income”. This label will appear in “Variable View” as well as in output produced by SPSS when analyzing your data.

Renaming and labeling a new variable

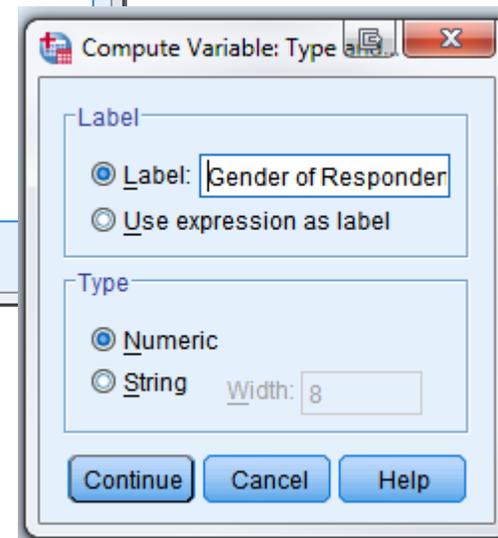
We can repeat the procedure to rename variable **f15c**, which corresponds to respondents' level of education.



Renaming and labeling a new variable

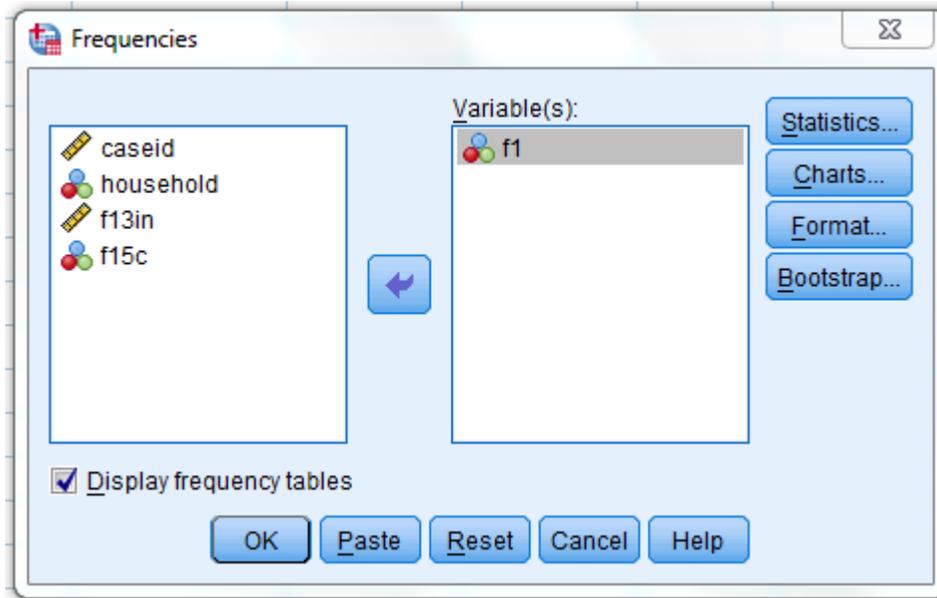


We can repeat the procedure to rename variable **f1**, which corresponds to the respondents' gender.



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.



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

Statistics

respondent's gender

N	Valid	2000
	Missing	854

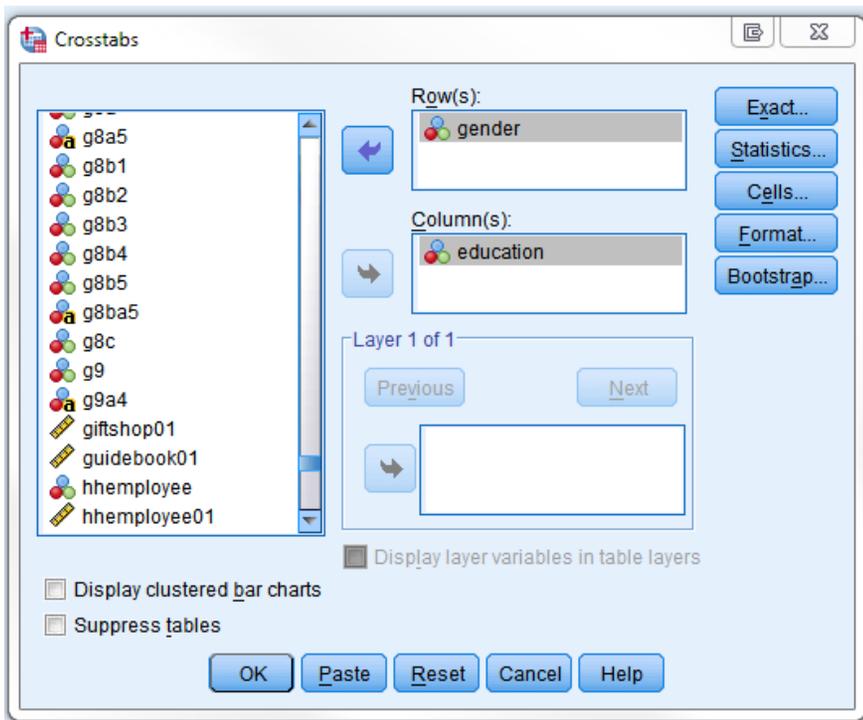
respondent's gender

		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



Case Processing Summary

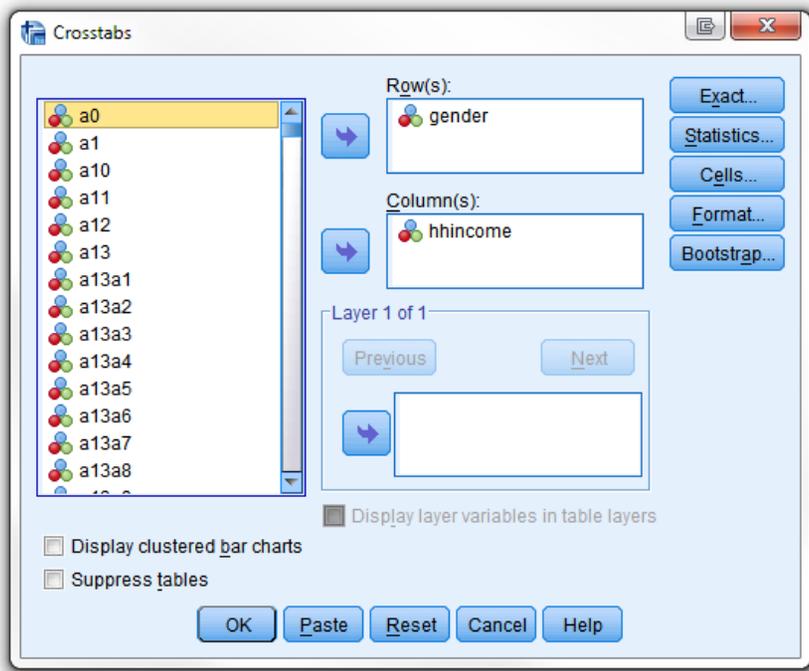
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * Highest level of education completed	1577	55.3%	1277	44.7%	2854	100.0%

gender * Highest level of education completed Crosstabulation

Count		Highest level of education completed						Total
		1.00	2.00	3.00	4.00	5.00	6.00	
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 Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * Household income	1819	63.7%	1035	36.3%	2854	100.0%

gender * Household income Crosstabulation

Count		Household income									Total
		1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	
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

$\chi^2 (1, N=2000) = 12.6, p = .00$

Source: 2011 SESRI Omnibus Survey

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

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

**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?