



# Thinking About Cross Tabulation

March 2013

Kenneth M Coleman



معهد البحوث الاجتماعية والاقتصادية المسحية  
Social & Economic Survey Research Institute

# Exploring a Data Set, I

- The initial stages of exploring a data set involve numerous activities, some of which have been covered in earlier training sessions.
  - Confirming the variables and number of cases.
  - What are the full texts of questions asked?
  - Is there a topline questionnaire with basic distributions of responses or is there a codebook?
  - How do these things compare to the Dictionary and Data File in the SPSS Data Set.

# Exploring a Data Set, II

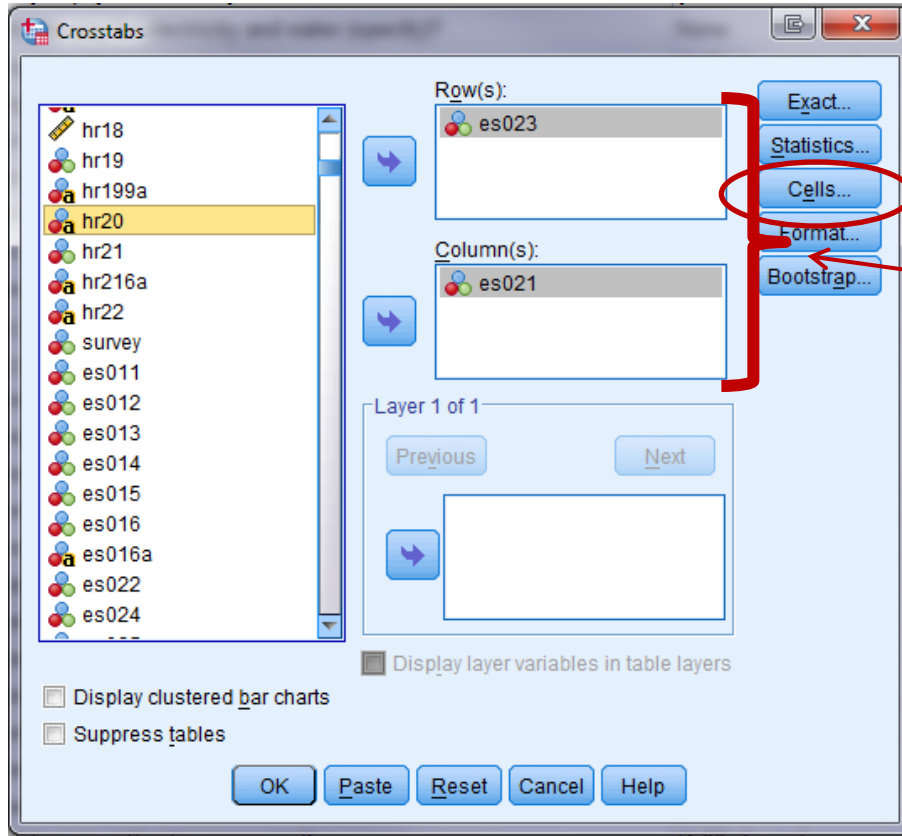
- Then, typically experienced researchers look at some basic descriptive statistics for key variables
  - ones of greatest theoretical importance in the study.
    - Key dependent variables (effects)
    - Key independent variables (causes)
- In both cases we hope to find (i) variation, and (ii) perhaps something resembling a normal distribution (although the latter may not always be expected).
- This will be done via the Frequencies procedure previously discussed.

# Exploring a Data Set via Looking at CrossTabs

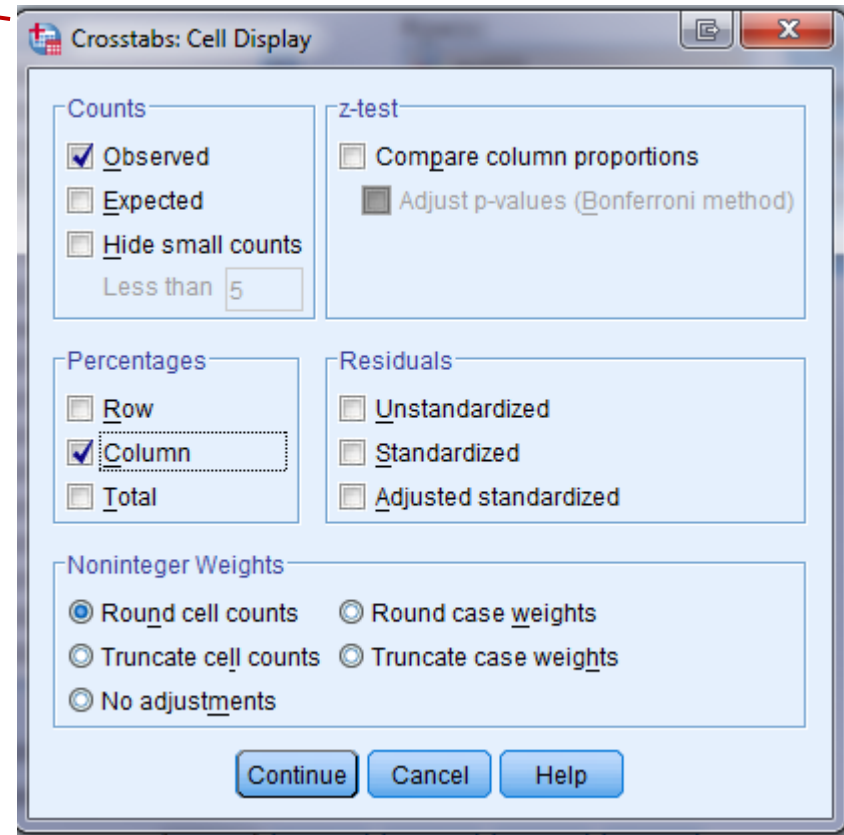
- Next, doing a number of quick cross tabulations via the Crosstabs procedure in SPSS is useful for a number of exploratory purposes that go beyond mere description.
- One of those exploratory purposes is to examine patterns of co-variation among possible indicators of the same concept, prior to attempts at scale construction and evaluating the reliability and validity of measurement scales. Illustrations of this use of Crosstabs can be found in Appendix B.

# Running a Crosstab in SPSS

The crosstabs command is available under the *Analyze / Descriptive Statistics / Crosstabs* menu.



Indicate in the two boxes which of your variables you would like to appear in the **rows** of your table, and which you'd like to appear in the **columns**.



The “cells” button allows you to add more information to your table than just the exact number of cases in each cell. **Selecting percentages by “column” is the generally accepted convention, rather than requesting a “total” percentage or doing percentages by rows.**

# Crosstabs Exercise [Revised]

As we know from our Mean, Median, and Mode exercise, there are several measures of wealth or family Affluence in the 2010 Omnibus dataset, including the following:

Variable Name	Variable Label
es011	Number of maids employed in household
es012	Number of nannies employed in household
es013	Number of drivers employed in household
es014	Number of gardeners employed in household
es015	Number of cooks employed in household
es04	Number of bedrooms in household
es05	Total monthly income of all household members

Run a series of crosstabs comparing each of the above variables. What can we learn from the crosstabs?

For example, run **es05** by all the other variables, i.e., place **es05** in the column box and all the other variables in the row box. You can do this all at once by including multiple variables in the Row box in SPSS. Remember, however, to percentage down the columns.

Does the number of employees of various types or number of bedrooms co-vary in any meaningful way with the total monthly income (**es05**)?

# Illustratively, es05 (columns, HH income) es011 (rows, number of maids).

Crosstab

		total monthly income of all hh members				Total
		1. less than qr50,000	2. qr50,000 to less than qr100,000	3. qr100,000 to less than qr150,000	4. qr150,000 or more	
number of maid employed in hh	.00	Count 31 6.1%	3 3.5%	0 0.0%	0 0.0%	34 5.6%
	1.00	Count 239 46.9%	21 24.4%	0 0.0%	0 0.0%	260 42.6%
	2.00	Count 179 35.1%	38 44.2%	6 75.0%	4 57.1%	227 37.2%
	3.00	Count 48 9.4%	16 18.6%	0 0.0%	1 14.3%	65 10.6%
	4.00	Count 5 1.0%	8 9.3%	1 12.5%	1 14.3%	15 2.5%
	5.00	Count 6 1.2%	0 0.0%	1 12.5%	0 0.0%	7 1.1%
	6.00	Count 2 0.4%	0 0.0%	0 0.0%	0 0.0%	2 0.3%
	9.00	Count 0 0.0%	0 0.0%	0 0.0%	1 14.3%	1 0.2%
	Total	Count 510 100.0%	86 100.0%	8 100.0%	7 100.0%	611 100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	154.325 <sup>a</sup>	21	.000
Likelihood Ratio	70.534	21	.000
Linear-by-Linear Association	46.521	1	.000
N of Valid Cases	611		

a. 23 cells (71.9%) have expected count less than 5. The minimum expected count is .01.

# Exploring a Data Set via Looking at CrossTabs

- Another possible use of crosstabs is to help one to begin to think about possible causal relationships. As noted earlier by Professor Hicken, there can be no causation without co-variation.
- To be sure, mere co-variation says nothing about causation. There has to be a plausible account of why variation in X might generate variation in Y, in short, there must be a theory to transform observed co-variation into a possible causal explanation.



# Cross Tabulation and Explorations of Causation

- One can begin to get a sense of the possibility of causal explanations emerging via elementary cross tabulation of variables (abbreviated in SPSS and elsewhere to Crosstabs).
- Earlier Professor Hicken argued that identification of causal relationships involves at least these three elements:
  - Co-variation (Crosstabs can help here)
  - Temporal order (Crosstabs cannot help here, but theory might)
  - Elimination of rival hypotheses (There are better tools than Crosstabs for this in most cases, but sometimes even Crosstabs can help eliminate rival hypotheses).

# Crosstabs as a Stimulus to Thought about Causation

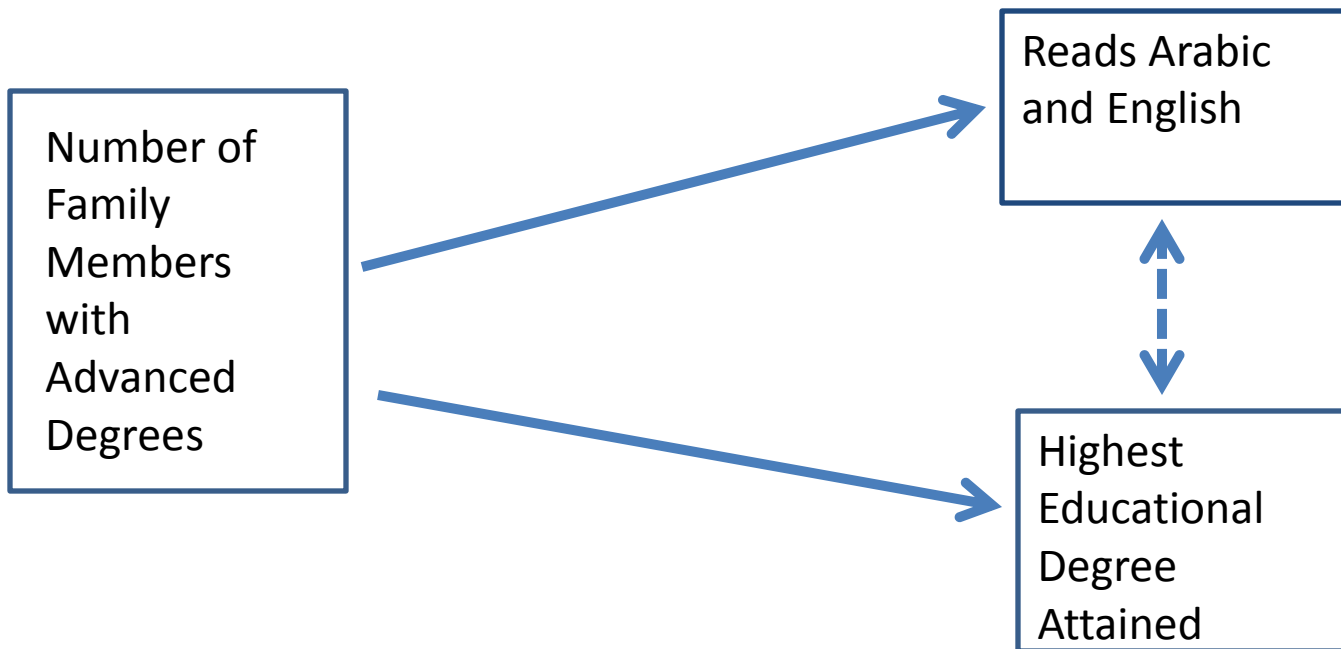
- In the most simple of causal accounts there is an independent variable (cause) and a dependent variable (effect), so co-variation can be explored.
- So, if theory makes it clear which is which, we can begin to explore relationships between independent and dependent variables.
  - Level of Education  $\rightarrow$  Languages in Which People Read
- Or the theory may envision more complicated causal sequences with reciprocal causation.
  - Level of Education  $\leftarrow \rightarrow$  Languages in Which People Read



# What Else Could It Be?

## Spurious Correlation

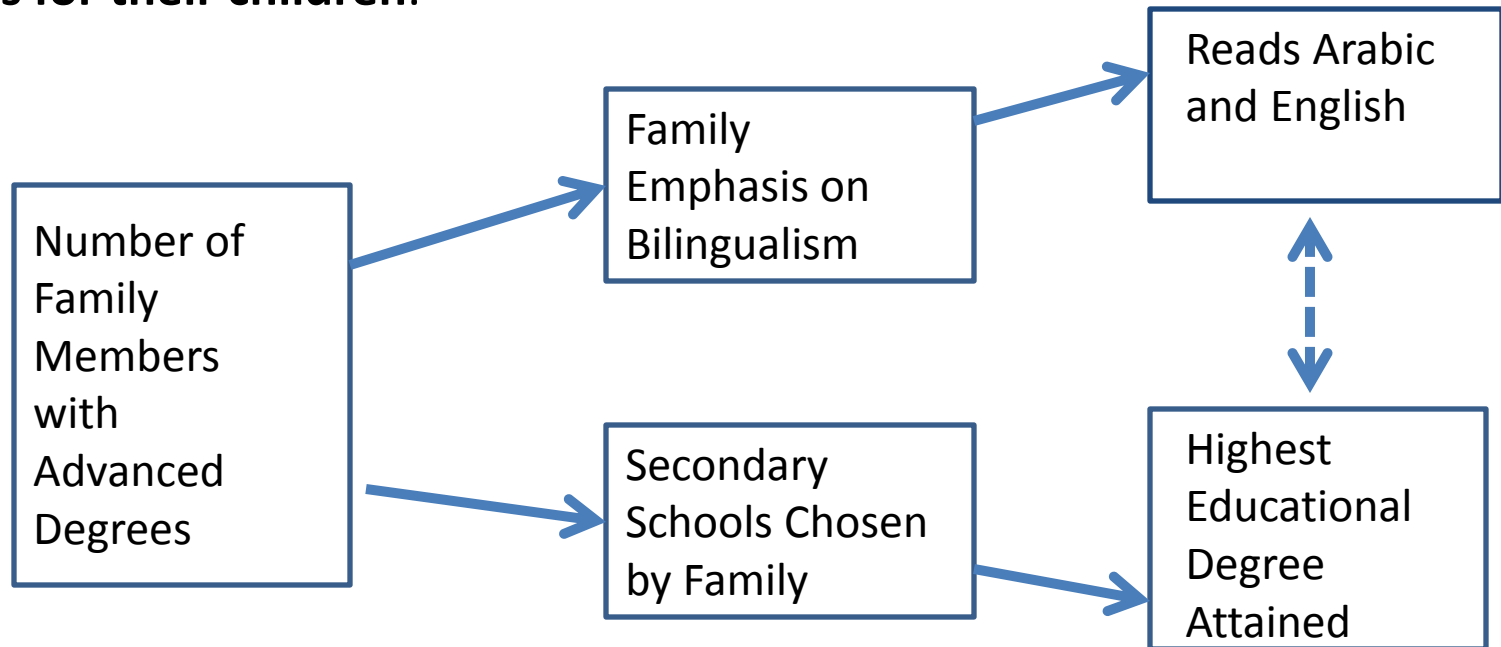
- One step in theorizing about this relationship is to envision alternative causal explanations. Perhaps Languages Read is related to the highest degree attained via **spurious correlation – variations in both being caused by variation in a third variable**, such as the number of advanced degrees held by family members.



# What Else Could It Be?

## More Complicated Causal Pathways

- But the possibility of spurious correlation lying behind an observed empirical relationship does not exhaust the possibilities of theoretical development. In this particular case we might also theorize that families holding many advanced degrees may behave differently – **emphasizing bilingualism to their children** and choosing **different types of secondary schools for their children**.



# Theorizing

- There is a long-standing debate, probably un-resolvable, about the merits of deductive theorizing versus inductive theorizing.
- In this particular illustration we started with an empirical observation of a relationship, and then considered theoretical options. This would be seen as **inductive** theorizing.
- However, in the process of considering “what else” might account for observable results, we might choose to consider theories of educational attainment, from which we could deduce certain propositions, such as the expected behavioral patterns of families with large numbers of advanced degree holders, an example of **deductive** theorizing.
- Much theory construction involves both inductive and deductive reasoning. That is not bad. Cross tabulation can help with both parts, but one should keep one’s eye open for theoretical puzzles when scanning exploratory results. This is the inductive part.

# Theorizing, II

- When seeking to test theories with survey data one confronts the potential dilemma of not having appropriate questions or items in the survey.
- For that reason alone, deductive theorizing can be helpful in survey construction. If we know in advance which theories we wish to test, we will be certain to include questions that can serve as indicators of key variables.
- But it is also the case that exploring a survey data set may generate some inductive theorizing. If fortunate, one may have items that allow one to test such inductive theories.

# What Else Could It Be Analysis?

- A “thought experiment” for participants in groups.
  - How might one measure the variables added to the “theory” suggested above?
  - Are there other variables that might help to account for the apparent relationship between reading multiple languages and attaining advanced degrees?
    - If so, how might we measure these additional variables?



# Crosstabs: Level of Measurement

- Often Crosstabs is used with two nominal or ordinal level variables, but:
  - Can be used with more than two variables.
    - One could for example, compare women and men in their views on the ideal family size, but one could also break that down by women 35 years of age and under, women over 35 years of age, men 35 years and under versus men over 35 years of age.
  - As noted, Crosstabs is used with nominal and/or ordinal variables, although for purposes of statistical testing, one should be cautious as to which tests are used in doing so. And, in some situations, it makes sense to run a nominal or an ordinal variable against an interval count variable, such as the number of maids employed in a household.

# Crosstabs: Looking for proximity to the diagonal

- One of the things one seeks in exploring co-variation via cross tabs is a measure of the strength of association. We will consider three.
  - Cramer's V ( $\phi$  or Phi); Tau  $\beta$ ; Eta ( $\eta$ )
- But before one even reads the statistics generated by SPSS, one can get a sense visually. Does the presumed dependent variable (effect) appear to move systematically when the independent variable (cause) increases or decreases? Illustratively, let's take the case of owning palaces and large TVs?

**own any tv bigger than 46' tv? ^ own a palace inside or outside qatar? Crosstabulation**

			own a palace inside or outside qatar?		Total
			1. yes	2. no	
own any tv bigger than 46' tv?	1. yes	Count	13	241	254
		% within own a palace inside or outside qatar?	68.4%	37.1%	38.0%
	2. no	Count	6	409	415
		% within own a palace inside or outside qatar?	31.6%	62.9%	62.0%
Total		Count	19	650	669
		% within own a palace inside or outside qatar?	100.0%	100.0%	100.0%

- So , in this case there is some co-variation, but it is not perfect, as at least 30% of those who own one of these things – a palace or a very large TV – do not own the other. So how do we make sense of this result? Have we found a meaningful degree of co-variation? Chi-Square will help us to assess that relationship.

# Chi-Square ( $\chi^2$ )

- This is a measure of the sum of the squared differences between observed frequencies in a contingency table (a two by two table, or two by three table, etc.) and the frequencies that would be expected if the null hypothesis were correct, i.e., if there were NO relationship between two variables, divided by the sum of expected frequencies.
- There is a known distribution of  $\chi^2$  for various degrees of freedom, such that statistical significance can be calculated.

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

# Illustrating a Chi-Square of Zero, Indicating No Relationship

Hypothetical Distribution of Highest Level of Education by Type of Household				
<u>Education</u>	Qataris	Ex-Patriots	Guest Workers	Total
Primary and Secondary	10	10	10	30
Some University	10	10	10	30
Complete University through Post-Graduate	10	10	10	30
Total	30	30	30	90

Here the Chi-Square value would be 0 because there is no difference between the observed frequencies and expected frequencies, so there is truly no relationship between the two variables and this result would **not be** statistically significant.

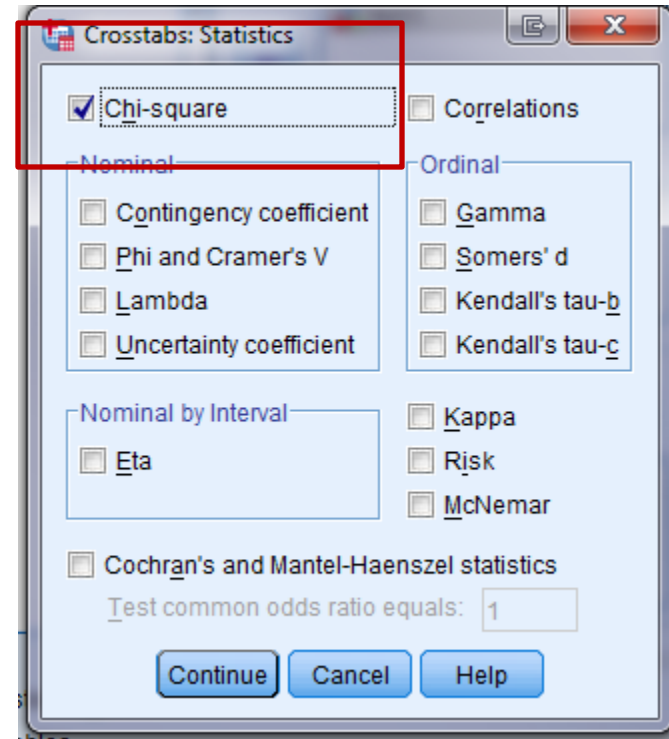
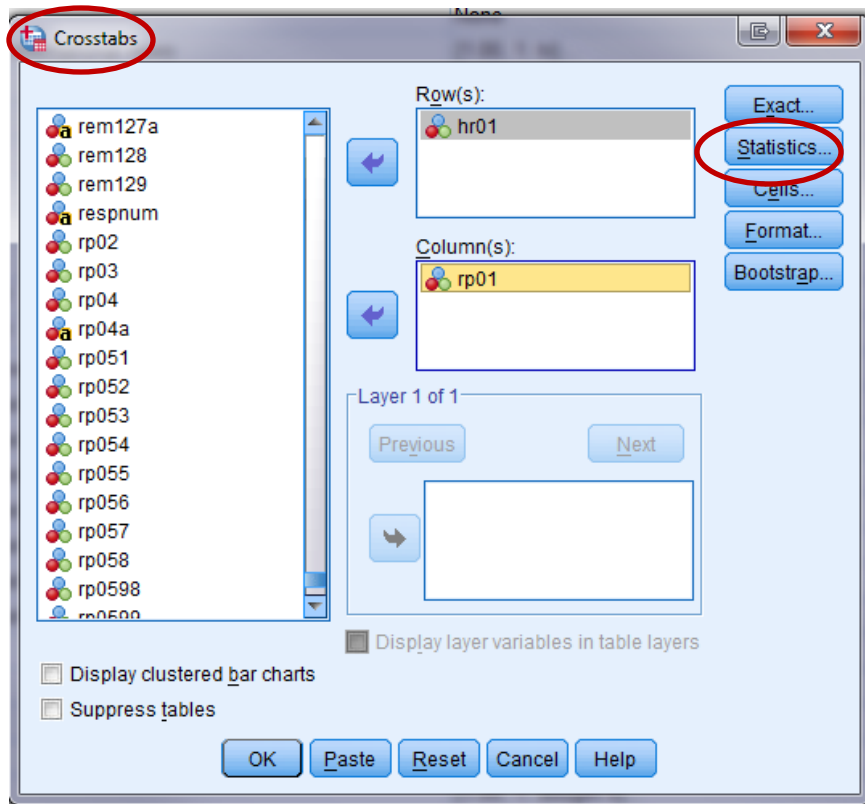
# Illustrating a Chi-Square Value for a Statistically Significant Relationship

Hypothetical Distribution of Highest Level of Education by Type of Household				
<u>Education</u>	Qataris	Ex-Patriots	Guest Workers	Total
Primary and Secondary	6	0	14	20
Some University	12	18	12	42
Complete University through Post-Graduate	12	12	4	28
Total	30	30	30	90

- Here the Chi-Square value would be 21.1, with four degrees of freedom, which indicates that the probability of making an error in rejecting the null hypothesis of no relationship would be less than one in a thousand ( $p < .001$ ).
- Degrees of freedom = (number of rows – 1) x (number of columns – 1).

# Performing a Chi-Square Test in SPSS

An option to run a Chi-Square test is available in the Crosstabs window under Analyze / Descriptive Statistics / Crosstabs. Got to the Statistics tab.



# An Example from the 2010 Omnibus Survey – All Respondents

Here is an example of SPSS output on a crosstab of household type (**hr01**) and use of the Internet (**rp01**). The Pearson Chi-Square is a highly significant 784.468 ( $p < .001$ ).

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	784.468 <sup>a</sup>	6	.000
Likelihood Ratio	788.824	6	.000
Linear-by-Linear Association	358.374	1	.000
N of Valid Cases	2084		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 33.94.

using the internet at home or at work? \* household type Crosstabulation

			household type			Total
			1. qatari	2. white collar	3. blue collar	
using the internet at home or at work?	1. yes, at home	Count	388	416	135	939
		% within household type	59.2%	55.5%	19.9%	45.1%
	2. yes, at work	Count	37	40	31	108
		% within household type	5.6%	5.3%	4.6%	5.2%
	3. yes, at both places	Count	139	233	43	415
		% within household type	21.2%	31.1%	6.3%	19.9%
	4. no	Count	91	60	471	622
		% within household type	13.9%	8.0%	69.3%	29.8%
Total	Count	655	749	680	2084	
	% within household type	100.0%	100.0%	100.0%	100.0%	

# Measures of Strength of Association

- While there are others, we will illustrate three measures of strength of association generated by the SPSS program Crosstabs.
  - Cramer's V (or  $\phi$ ): Appropriate for data at the nominal level of measurement, and varies between 0 and +1.0.  $\phi$  is used for 2 x 2 tables; Cramer's V for 2 x n tables. Cramer's V and  $\Phi$  are identical in 2 x 2 tables.
  - Kendall's Tau  $\beta$ : Appropriate for data at the ordinal level of measurement, varies between -1.0 and +1.0, and accounts for "ties" – as often occur in survey responses where there are a finite number of response categories and a large number of persons giving the same answer.
  - Eta ( $\eta$ ): Appropriate for one nominal or ordinal variable and a variable measured at the interval level.



# Illustration of Cramer's V (or $\phi$ )

Here we see a weak relationship between gender (**hr06**) and having visited the US (**pv016**), with males 5.5% more likely to have visited than females. While the difference is statistically significant ( $p = .006$ ) due to the large number of cases ( $n = 1,403$ ), Phi is only .074. Given that  $\phi$  does not approach 1.0, the relationship is quite weak.

have you ever visited usa \* gender of hh member Crosstabulation

			gender of hh member		Total
			1. male	2. female	
have you ever visited usa	1. yes	Count	134	102	236
		% within gender of hh member	19.6%	14.1%	16.8%
	2. no	Count	548	619	1167
		% within gender of hh member	80.4%	85.9%	83.2%
Total		Count	682	721	1403
		% within gender of hh member	100.0%	100.0%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.074	.006
	Cramer's V	.074	.006
N of Valid Cases		1403	

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.

# Illustration of Kendall's Tau $\beta$

If exploring a data set to see which indicators might reflect a propensity to support non-traditional roles for women, one might do a cross tab of these two indicators: “working outside the home” (**gen07b**) and “education helping women to have a career” (**gen06d**), both at the ordinal level.

The Tau  $\beta$  is .36, which is highly significant ( $p < .001$ ).

higher education is good for woman because it helps her to have a good career \* a married woman should be allowed to work outside the home if she wants to Crosstabulation

			a married woman should be allowed to work outside the home if she wants to				Total
			1. strongly agree	2. agree	3. disagree	4. strongly disagree	
higher education is good for woman because it helps her to have a good career	1. strongly agree	Count	229	199	29	2	459
		% within a married woman should be allowed to work outside the home if she wants to	90.9%	60.9%	49.2%	18.2%	70.7%
	2. agree	Count	22	105	27	0	154
		% within a married woman should be allowed to work outside the home if she wants to	8.7%	32.1%	45.8%	0.0%	23.7%
	3. disagree	Count	1	20	3	4	28
		% within a married woman should be allowed to work outside the home if she wants to	0.4%	6.1%	5.1%	36.4%	4.3%
	4. strongly disagree	Count	0	3	0	5	8
		% within a married woman should be allowed to work outside the home if she wants to	0.0%	0.9%	0.0%	45.5%	1.2%
Total	Count		252	327	59	11	649
	% within a married woman should be allowed to work outside the home if she wants to		100.0%	100.0%	100.0%	100.0%	100.0%

## Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.360	.031	10.561	.000
Measure of Agreement	Kappa	.214	.026	8.022	.000
N of Valid Cases		649			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

# Illustration of Eta ( $\eta$ ) [Revised]

Here are SPSS results for the relationship among Qataris between gender (**hr06**) and household income reported (**incomeQW**). In this subset of cases, the strength of the relationship is very weak, .087, not approaching  $\pm 1.0$ , so gender does not really affect income reported.

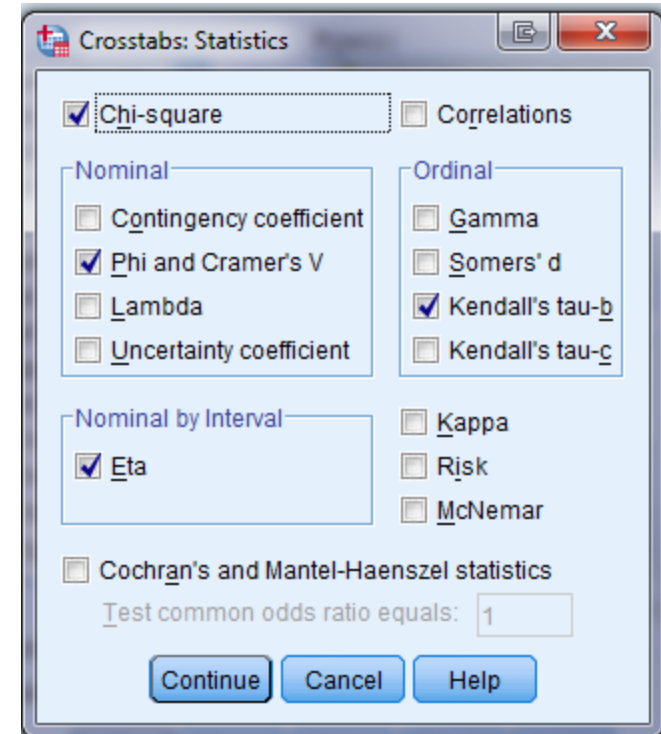
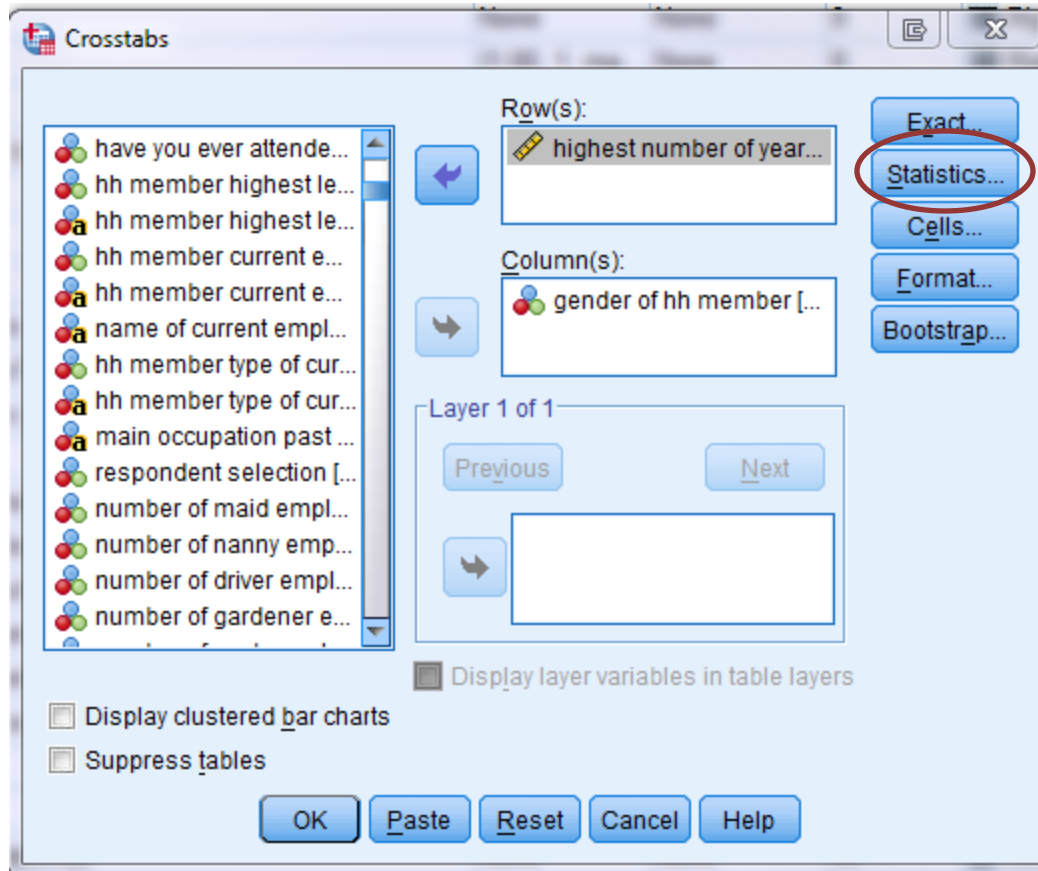
## Directional Measures

		Value
Nominal by Interval	Eta	Monthly household income for Qatari Citizens & White Collar Workers Dependent
		gender of hh member Dependent
		.087
		.164

Monthly household income for Qatari Citizens & White Collar Workers \* gender of hh member Crosstabulation

			gender of hh member		Total
			1. male	2. female	
Monthly household income for Qatari Citizens & White Collar Workers	Less than QR10,000	Count	37	49	86
		% within gender of hh member	13.0%	15.7%	14.4%
	QR10,000 to less than QR20,000	Count	77	93	170
		% within gender of hh member	27.0%	29.8%	28.5%
	QR20,000 to less than QR30,000	Count	60	64	124
		% within gender of hh member	21.1%	20.5%	20.8%
	QR30,000 to less than QR40,000	Count	29	35	64
		% within gender of hh member	10.2%	11.2%	10.7%
	QR40,000 to less than QR50,000	Count	29	27	56
		% within gender of hh member	10.2%	8.7%	9.4%
	QR50,000 to less than QR60,000	Count	18	12	30
		% within gender of hh member	6.3%	3.8%	5.0%
	QR60,000 to less than QR70,000	Count	14	12	26
		% within gender of hh member	4.9%	3.8%	4.4%
QR70,000 to less than QR80,000	Count	5	10	15	
	% within gender of hh member	1.8%	3.2%	2.5%	
QR80,000 to less than QR90,000	Count	3	2	5	
	% within gender of hh member	1.1%	0.6%	0.8%	
QR90,000 to less than QR100,000	Count	4	4	8	
	% within gender of hh member	1.4%	1.3%	1.3%	
QR100,000 to less than QR110,000	Count	1	3	4	
	% within gender of hh member	0.4%	1.0%	0.7%	
QR110,000 to less than QR120,000	Count	2	0	2	
	% within gender of hh member	0.7%	0.0%	0.3%	
QR140,000 to less than QR150,000	Count	2	0	2	
	% within gender of hh member	0.7%	0.0%	0.3%	
QR150,000 to less than QR160,000	Count	1	0	1	
	% within gender of hh member	0.4%	0.0%	0.2%	
QR160,000 to less than QR170,000	Count	0	1	1	
	% within gender of hh member	0.0%	0.3%	0.2%	
QR200,000 or more	Count	3	0	3	
	% within gender of hh member	1.1%	0.0%	0.5%	
Total	Count	285	312	597	
	% within gender of hh member	100.0%	100.0%	100.0%	

# Measures of Strength of Association in SPSS



Options for measures of strength of association are found under the "Statistics" button when you run a Crosstab.

# Chi-Square and Strength of Association Exercises [Distribute]

Using the 2010 Omnibus Survey [Data Set 2] determine whether men and women in Qatar are different in terms of their levels of education.

**Run a crosstab for gender (hr06) and the respondent's preference for gender of survey interviewer (rp03).** Specify gender of interviewer as the rows, and gender of respondent as the columns. Then, run a chi-square test, and direct SPSS to generate values for the three measures of strength of association (Cramer's V, Kendall's Tau B, and Eta).

Does the Chi-square test indicate whether men and women differ in terms of their preference for a male or female interviewer?

Which measure of the strength of association is appropriate to consider when comparing interviewer preference between men and women? To help you answer this question, determine whether interviewer preference is nominal, ordinal, or interval data. Is gender nominal, ordinal, or interval data?

# Chi-Square and Measures of Strength of Association Exercises [Distribute]

**Run a crosstab for respondent's highest level of education (hr17) and respondent's evaluation of personal financial situation (personalfin).** Specify the financial evaluation as the rows and levels of education as the columns.

Does the Chi-square test indicate whether individuals at different levels of education differ in their evaluations of their personal financial situation?

What type of variable (nominal, ordinal, or interval) is the evaluation of one's personal financial situation?

Which strength of association measure is appropriate in this case?

**Run a crosstab for gender (hr06) and respondent's belief about the ideal number of children to have (gen01a).** Specify gender as the columns and ideal number of children as the rows.

Do men and women differ in their belief about the ideal number of children to have?

What type of variable is the belief about the ideal number of children to have (nominal, ordinal, or interval)?

Which strength of association measure is appropriate in this case?

# Hr06 by rp03.

## Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.617	.000
	Cramer's V	.617	.000
N of Valid Cases		2070	

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.

**Nominal by nominal**

Reasonably strong relationship for attitudinal data.

## interviewer preferences (gender) \* gender of hh member Crosstabulation

			gender of hh member		Total
			1. male	2. female	
interviewer preferences (gender)	1. male	Count	597	19	616
		% within gender of hh member	45.6%	2.5%	29.8%
	2. female	Count	135	494	629
		% within gender of hh member	10.3%	64.9%	30.4%
	3. it does not matter	Count	577	248	825
		% within gender of hh member	44.1%	32.6%	39.9%
Total		Count	1309	761	2070
		% within gender of hh member	100.0%	100.0%	100.0%





# Hr06 by gen01a

Directional Measures

			Value
Nominal by Interval	Eta	what is the ideal number of children to have? Dependent	.023
		gender of hh member Dependent	.174

**Nominal (gender) by interval (number of desired children).**

No real diagonal here – in both genders, between 72% and 76% think that 4-6 children is the ideal number. No relationship.

what is the ideal number of children to have? \* gender of hh member Crosstabulation

			gender of hh member		Total
			1. male	2. female	
what is the ideal number of children to have?	.00	Count	5	5	10
		% within gender of hh member	1.7%	1.5%	1.6%
2.00	Count	16	15	31	
	% within gender of hh member	5.6%	4.5%	5.0%	
3.00	Count	16	9	25	
	% within gender of hh member	5.6%	2.7%	4.0%	
4.00	Count	81	132	213	
	% within gender of hh member	28.1%	39.3%	34.1%	
5.00	Count	49	47	96	
	% within gender of hh member	17.0%	14.0%	15.4%	
6.00	Count	79	75	154	
	% within gender of hh member	27.4%	22.3%	24.7%	
7.00	Count	9	15	24	
	% within gender of hh member	3.1%	4.5%	3.8%	
8.00	Count	13	18	31	
	% within gender of hh member	4.5%	5.4%	5.0%	
9.00	Count	1	3	4	
	% within gender of hh member	0.3%	0.9%	0.6%	
10.00	Count	15	11	26	
	% within gender of hh member	5.2%	3.3%	4.2%	
11.00	Count	0	1	1	
	% within gender of hh member	0.0%	0.3%	0.2%	
12.00	Count	2	5	7	
	% within gender of hh member	0.7%	1.5%	1.1%	
15.00	Count	1	0	1	
	% within gender of hh member	0.3%	0.0%	0.2%	
20.00	Count	1	0	1	
	% within gender of hh member	0.3%	0.0%	0.2%	
Total	Count	288	336	624	
	% within gender of hh member	100.0%	100.0%	100.0%	

# Two-Level Crosstabs [Revised]

Important insights can sometimes come via Two-Level Crosstabs in SPSS. For example, here we see **among Qataris** that the relationship between gender (**hr06**) and visits to the US (**pv016**) is a stronger among those who read both Arabic and English (**premed04**).

have you ever visited usa \* gender of hh member \* reading languages Crosstabulation

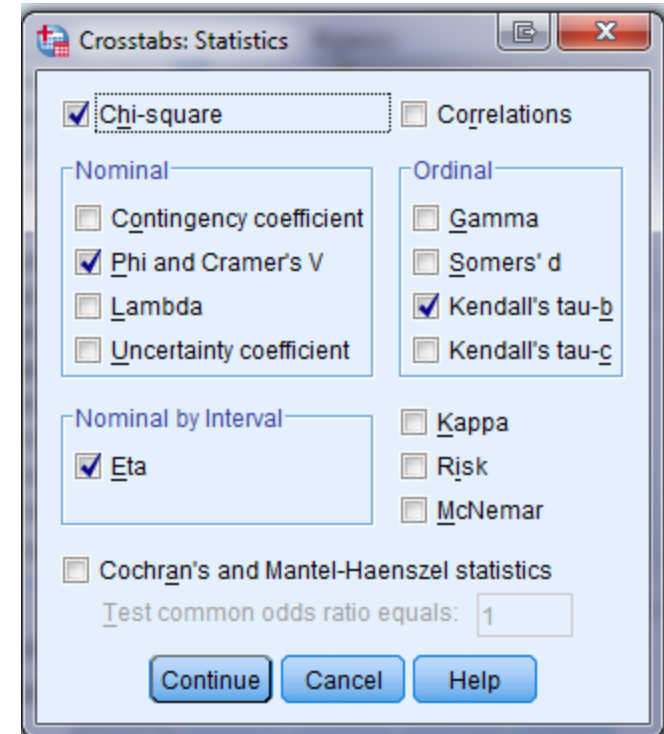
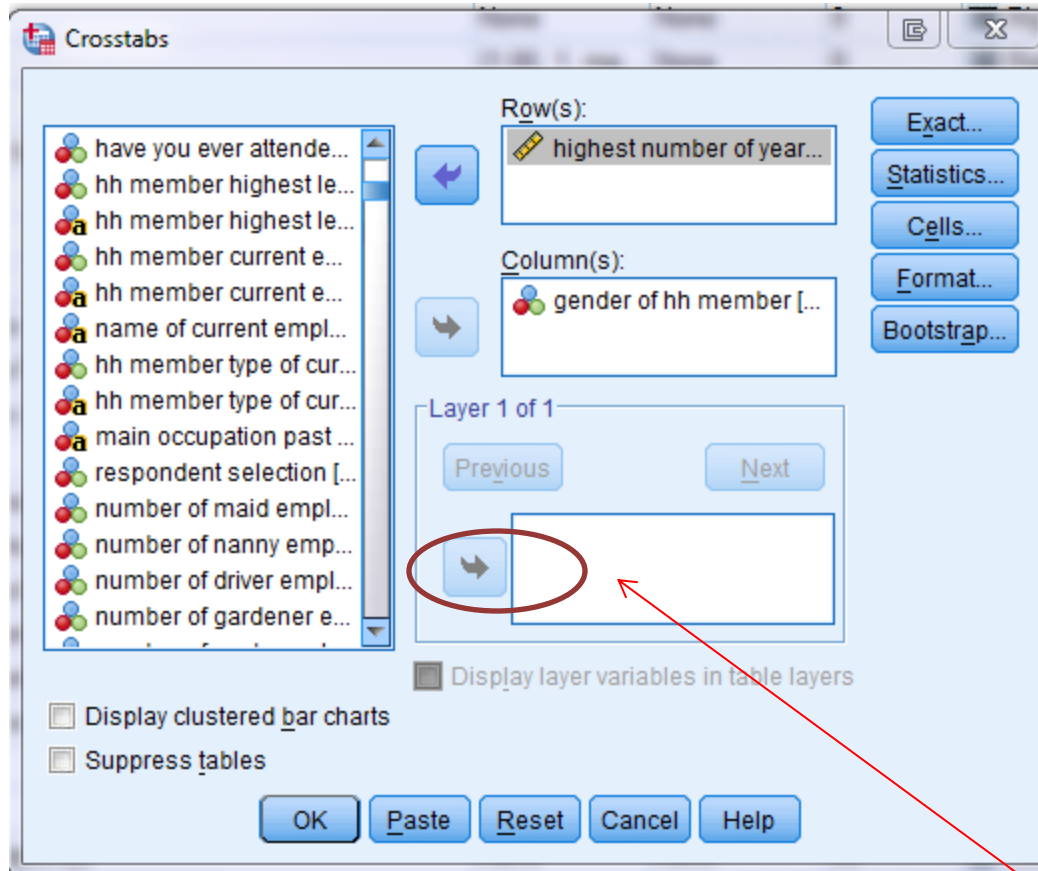
reading languages				gender of hh member		Total
				1. male	2. female	
1. arabic	have you ever visited usa	1. yes	Count	48	38	86
			% within gender of hh member	20.3%	13.7%	16.8%
		2. no	Count	188	239	427
			% within gender of hh member	79.7%	86.3%	83.2%
Total			Count	236	277	513
			% within gender of hh member	100.0%	100.0%	100.0%
2. english	have you ever visited usa	2. no	Count	4	3	7
			% within gender of hh member	100.0%	100.0%	100.0%
		Total	Count	4	3	7
			% within gender of hh member	100.0%	100.0%	100.0%
3. both	have you ever visited usa	1. yes	Count	28	14	42
			% within gender of hh member	45.2%	18.9%	30.9%
		2. no	Count	34	60	94
			% within gender of hh member	54.8%	81.1%	69.1%
Total			Count	62	74	136
			% within gender of hh member	100.0%	100.0%	100.0%
Total	have you ever visited usa	1. yes	Count	76	52	128
			% within gender of hh member	25.2%	14.7%	19.5%
		2. no	Count	226	302	528
			% within gender of hh member	74.8%	85.3%	80.5%
Total			Count	302	354	656
			% within gender of hh member	100.0%	100.0%	100.0%

Symmetric Measures

reading languages			Value	Approx. Sig.
1. arabic	Nominal by Nominal	Phi	.088	.045
		Cramer's V	.088	.045
	N of Valid Cases			513
2. english	Nominal by Nominal	Phi	<sup>a</sup>	
		N of Valid Cases		7
	3. both	Nominal by Nominal	Phi	.283
Cramer's V			.283	.001
N of Valid Cases			136	
Total	Nominal by Nominal	Phi	.132	.001
		Cramer's V	.132	.001
	N of Valid Cases			656

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.
- c. No statistics are computed because have you ever visited usa is a constant.

# Measures of Strength of Association in SPSS



Where one would put an additional control variable in a two-level cross tab.

# Appendix A: Other Measures of Strength of Association Found in SPSS, I

- **Contingency Coefficient [C]**: Based on Chi-Square, can be used with a crosstab of any size, has minimum value of 0 [no association] and a maximum value that depends on the size of the table. Best used with tables with same numbers of rows and columns.
- **Lambda [ $\lambda$ ]**: For crosstabs using nominal variables. Measures percentage improvement in predictions of value of dependent variable once we know value of independent variable. Varies between 0 [no improvement in predictions] to 1 [perfect ability to predict value of dependent variable].

# Appendix A: Other Measures of Strength of Association Found in SPSS, II

- **Uncertainty Coefficient [U]**: Also for use with two nominal variables. Estimates the “reduction in uncertainty” of the value of the dependent variable from knowing the value of the independent variable. Varies between 0 [no reduction in uncertainty] to 1 [total elimination of uncertainty].
- **Tau<sub>c</sub>**: Similar to Tau<sub>b</sub>, but more appropriate when the number of rows differs from the number of columns. Appropriate for ordinal level variables.

# Appendix A: Other Measures of Strength of Association Found in SPSS, III

- **Gamma**: For use with ordinal data, but makes no adjustment for table size (number of rows & columns) nor does it adjust for ties. As a result, will generally exhibit a higher value than will  $\text{Tau}_b$  or  $\text{Tau}_c$ . Can vary between -1 [perfect negative relationship] through 0 [no relationship] to +1 [perfect positive relationship].
- **Somer's d [d]**. Another measure for use with ordinal variables, in which ties are taken into account in the denominator of an equation but not in the numerator. Two separate versions are generated by SPSS, depending on which variable is assumed to be dependent.

# Appendix A: Other Measures of Strength of Association Found in SPSS, IV

- **Kappa**: Often used in healthcare research or in coding data to assess the extent to which two observers agree on the classification of an observation. Assumes nominal data.
- **McNemar test**: Particularly appropriate for pre-test/post-test situations using the same sample. Assumes nominal measurement. **Cochran's Q test**, extends the basic concept of McNemar to observations across 3+ time periods.
- More detailed descriptions of these measures of strength of association can be found in:
  - Norman H. Nie, C. Hadlai Hull, Jean G. Jenkins, Karin Steinbrenner, Dale H. Bent, Statistical Package for the Social Sciences [SPSS], second edition, McGraw Hill, 1975, section on Crosstabs; or,
  - Marjorie Pett, Nonparametric Statistics for Health Care Research, Sage Publications, 1997, passim.

# Appendix B: Exploring Possible Scale Items via Crosstabs

- As initially noted, doing a number of quick cross tabulations via the Crosstabs procedure in SPSS is useful for a number of exploratory purposes that go beyond mere description.
- First, one can examine if certain indicators of a general concept seem to be related. This may give one a sense of whether scales can be developed that will prove to be **reliable**.
  - If some degree of co-variation exists, items may be candidates for inclusion in a pool of items to be employed in developing a scale.



# Appendix B: Exploring Possible Scale Items via Crosstabs

- Scale-construction consists of identifying indicators that seem to co-vary sufficiently to be measuring an underlying concept, but which also measure differing facets of that concept.
- The whole point of building scales is an underlying assumption that no one indicator is going to measure the concept perfectly. Generally, that's not a problem – our indicators rarely measure concepts perfectly.

# Appendix B: Exploring Possible Scale Items via Crosstabs

- But to see how Crosstabs in SPSS can help, consider the following cross tabulation of 2 items from the 2010 omnibus survey. The variable in the columns deals with subjective economic status (**qol05**) which varies between one and four, with one being the high end, and the other variable deals with whether one's economic status has improved or worsened over the last two years (**qol06**). And we will do this among **Blue Collar guest workers**.

# Appendix B: Two Possible Indicators of Subjective Economic Status in 2010 Omnibus Survey

Here among **Blue collar workers**, as subjective economic status improves (1 = highest subjective assessment of own economic status) the sense of that one's economic status has become "much better" over the past two years is 86.7%, while among those whose sense of subjective economic status is lowest (4), 38.9% feel that their economic condition is somewhat worse or much worse than two years earlier.

Crosstab

			subjective economic status				Total
			1.00	2.00	3.00	4.00	
comparing your current personal financial situation to two years ago	1. much better	Count	13	32	24	3	72
		% within subjective economic status	86.7%	23.2%	7.0%	2.1%	11.3%
	2. somewhat better	Count	2	68	116	27	213
		% within subjective economic status	13.3%	49.3%	34.0%	18.8%	33.4%
	3. about the same	Count	0	27	111	58	196
		% within subjective economic status	0.0%	19.6%	32.6%	40.3%	30.7%
	4. somewhat worse	Count	0	11	74	44	129
		% within subjective economic status	0.0%	8.0%	21.7%	30.6%	20.2%
	5. much worse	Count	0	0	16	12	28
		% within subjective economic status	0.0%	0.0%	4.7%	8.3%	4.4%
Total	Count	15	138	341	144	638	
	% within subjective economic status	100.0%	100.0%	100.0%	100.0%	100.0%	

# Appendix B: Co-variation between possible indicators of affluence **among Qataris, 2010**

- These suggest a pattern of co-variation: those who own palaces (**es021**) tend also to own large TVs (**es02a**). We see that most Qataris who own a palace also own a large TV of over 46" (68.4%), while most Qataris who do not own a palace also do not own a large TV (62.9%). But co-variation is not perfect, as over 30% own one but not the other.

own any tv bigger than 46' tv? \* own a palace inside or outside qatar? Crosstabulation

			own a palace inside or outside qatar?		Total
			1. yes	2. no	
own any tv bigger than 46' tv?	1. yes	Count	13	241	254
		% within own a palace inside or outside qatar?	68.4%	37.1%	38.0%
	2. no	Count	6	409	415
		% within own a palace inside or outside qatar?	31.6%	62.9%	62.0%
Total		Count	19	650	669
		% within own a palace inside or outside qatar?	100.0%	100.0%	100.0%

# Appendix B: Co-variation of Possible Indicators of Objective Economic Status

- In contrast, owning a palace (**es021**) and owning a yacht (**es023**) co-vary strongly only in the sense that most Qataris own neither. Only 4 of 19 Qataris who own a palace also own a yacht (21%), while 648 of the 666 who do not own a palace do not own a yacht (97.3%). So these two items essentially measure the same thing – the **lack of** conspicuous wealth.

Crosstab

			own a palace inside or outside qatar?		Total
			1. yes	2. no	
own a yacht inside or outside qatar?	1. yes	Count	4	18	22
		% within own a palace inside or outside qatar?	21.1%	2.7%	3.2%
	2. no	Count	15	648	663
		% within own a palace inside or outside qatar?	78.9%	97.3%	96.8%
Total		Count	19	666	685
		% within own a palace inside or outside qatar?	100.0%	100.0%	100.0%

# Appendix B: Scale Development

- Ultimately, the best way to develop scales is to find a pool of plausible items and to test their reliability via procedures easily available in SPSS – which we have addressed in other sessions. Reliability basically addresses whether repeated measurements produce similar assessments, but capturing differing elements of the underlying concept.
- And, then, once a reliable scale has been developed, one needs to address its validity. Validity addresses whether we are measuring what we seek to measure, and is often assessed via whether a scale relates to other variables in an expected way.
- Cross tabulation can provide a modest first step in assessing the potential for reliability. It will be of less help in assessing validity. But there are even ways in which Crosstabs can give clues about validity.

# Additional References on Cross Tabulation

- Herbert F. Weisberg, Jon A. Krosnick, Bruce D. Bowen, An Introduction to Survey Research, Polling and Data Analysis, Sage Publications, 1996, especially chapters on measures of association and control tables.
- Schaeffer, Richard L. “Categorical Data Analysis” at this URL:  
[http://courses.ncssm.edu/math/Stat\\_Inst/PDFS/Categorical%20Data%20Analysis.pdf](http://courses.ncssm.edu/math/Stat_Inst/PDFS/Categorical%20Data%20Analysis.pdf)
- Reynolds, H. T. Analysis of Nominal Data. Beverly Hills, Calif.: Sage Publications, 1977.
- Shively, W. Phillips. The Craft of Political Research, Sixth Edition, Upper Saddle River, New Jersey, 2005, especially chapter on inference (or how to gamble on your research).

# Additional References on Cross Tabulation, II

- Reynolds, H. T. The Analysis of Cross-Classifications. New York: Free Press, 1977.
- McCutcheon, Allan L. Latent Class Analysis. Newbury Park, Calif.: Sage, 1987.
- And this URL is useful in calculating  $X^2$  from someone else's data, if the data are not in an SPSS data set.

[http://www.physics.csbsju.edu/stats/contingency\\_NROW\\_NCOLUMN\\_form.html](http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html)