Last week (Chapter 10)

Bivariate table, association and Chi square test of independence…

Why do we use Chi square?

To determine whether there is a "significant" association between variables.. (note: we are working with samples, not the full population)

Examples: Education & smoking?

Place of Study and employment status??

Month of birth & Success as an Athlete?

Today (Chapter 11)

More on:
Associations between Variables and the Bivariate Table (Crosstab)

Three fundamental questions that we ask in examining bivariate associations (significance? strength? pattern?)

A few measures of association Phi, Cramer's v and Lambda.. (nominal variables)..

Introduction to Bivariate Association

In a bivariate table:
Evidence for an association exists if the conditional distributions of one variable change across the values of the other variable.

Always useful to produce Column %'s

Interview 400 persons (Sample size)								
	Quarter of birth:							
	First (Jan-March)		Second (April-June)		Third (July-Sept)		Fourth (Oct-Dec)	
Universtiy Athlete	37	37%	30	30%	18	18%	15	15%
Non-Athlete	63	63%	70	70%	82	82%	85	85%
	100		100		100		100	

Note: To determine whether it is significant or not requires a "significance test" (chi square).

Interview 400 persons (Sample size)					
	Quarter of birth:				
	First (Jan-March)	Second (April-June)	Third (July-Sept)	Fourth (Oct-Dec)	TOTAL
Universtiy Athlete	37	30	18	15	100
Non-Athlete	63	70	82	85	300
TOTAL	100	100	100	100	400
Is there a significant relationship?					

Is there a relationship between "month of birth" and "success as an "athlete"..

Performing the Chi Square Test Using the Five-Step Model

Step 1: Make Assumptions and Meet Test Requirements

- Independent random samples

4 samples, by month of birth (First quarter, 2nd quarter, etc).

Level of measurement:

Nominal: University Athlete or not

Step 2: State the Null Hypothesis

- H_0: The variables are independent
  - Another way to state the H_0, more consistently with previous tests:
    - $H_0: f_o = f_e$
- H_1: The variables are dependent
  - Another way to state the H_1:
    - $H_1: f_o \neq f_e$

Step 3: Select Sampling Distribution and Establish the Critical Region

Interview 400 persons (Sample size)

	Quarter of birth:				
	First (Jan-March)	Second (April-June)	Third (July-Sept)	Fourth (Oct-Dec)	TOTAL
Universtiy Athlete	37	30	18	15	100
Non-Athlete	63	70	82	85	300
TOTAL	100	100	100	100	400

df = (4-1)(2-1) = 3

- Sampling Distribution = χ^2
- Alpha = .05
- df = (r-1)(c-1)
- χ^2 (critical) = ?

Appendix C Distribution of Chi Square

Critical values at alpha =.05

	.99	.98	.95	.90	.80	.70	.50	.30	.20	[illegible]	.05	.02	.01	.001
1	.000	.001	.004	.016	.064	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	[illegible]	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	[illegible]	[illegible]	7.815	9.837	11.341	16.268
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	[illegible]	9.488	11.668	13.277	18.465
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Step 3: Select Sampling Distribution and Establish the Critical Region

- Sampling Distribution = χ^2
- Alpha = .05
- df = (r-1)(c-1) = 1
- χ^2 (critical) = **7.851**

df = (4-1)(2-1) = 3

Interview 400 persons (Sample size)					
	Quarter of birth:				
	First (Jan-March)	Second (April-June)	Third (July-Sept)	Fourth (Oct-Dec)	TOTAL
Universtiy Athlete	37	30	18	15	100
Non-Athlete	63	70	82	85	300
TOTAL	100	100	100	100	400

Step 4: Calculate the Test Statistic

- fo

Interview 400 persons (Sample size)					
	Quarter of birth:				
	First (Jan-March)	Second (April-June)	Third (July-Sept)	Fourth (Oct-Dec)	TOTAL
Universtiy Athlete	37	30	18	15	100
Non-Athlete	63	70	82	85	300
TOTAL	100	100	100	100	400

$$f_e = \frac{\text{Row marginal X Column marginal}}{N}$$

fe

	First	second	third	fourth
Univ athlete	25	25	25	25
Non-athlete	75	75	75	75

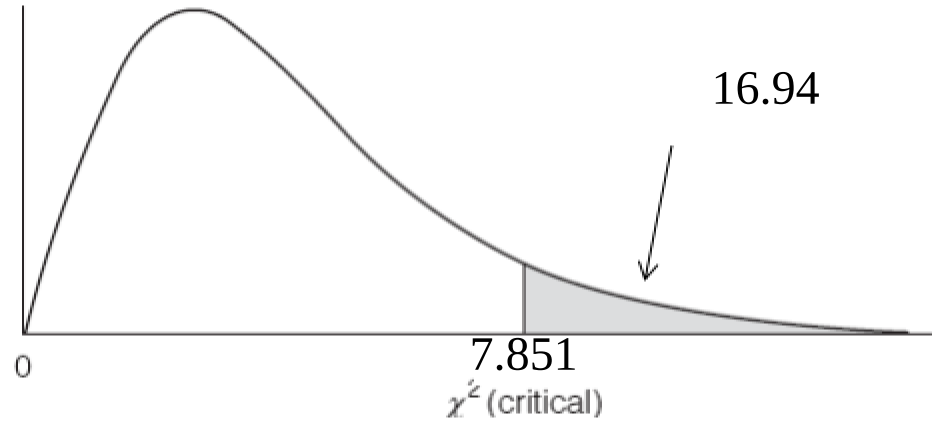
Create our corresponding Table for calculating chi square..

f0	fe	f0-fe	$(fo-fe)^2$	$(fo-fe)^2/fe$
37	25	12	144	5.76
63	75	-12	144	1.92
30	25	5	25	1.00
70	75	-5	25	0.33
18	25	-7	49	1.96
82	75	7	49	0.65
15	25	-10	100	4.00
85	75	10	100	1.33

$$\chi^2(\text{obtained}) = \sum \frac{(f_o - f_e)^2}{f_e} = 16.94$$

Step 5: Make Decision and Interpret Results

- χ^2 (critical) = 7.851
- χ^2 (obtained) = 16.94
- The test statistic is in the Critical (shaded) Region:
  - We reject the null hypothesis of independence.
  - Opinion on healthcare privatization is associated with political ideology.



- Bivariate association can be investigated by finding answers to three questions:
  1. Does an association exist (significance)?
  2. How strong is the association?
  3. What is the pattern or direction of the association?

1. Does an association exist?

- To detect association within bivariate tables:
  1. Calculate percentages within the categories of the independent variable.
  2. Compare percentages across the categories of the independent variable.
  3. Also: Chi Square test of Independence

     formally determines “statistical significance”

Careful!!!!!!!!!!! In setting up your crosstab!!!!

- When independent variable is the column variable (in this course):

1. Calculate percentages **within** the columns (vertically).
   Column percentages are conditional distributions of *Y* for each value of *X*.

2. Compare percentages **across** the columns (horizontally).

Follow this rule:

“**Percentage Down, Compare Across**”

Example: Does an association exist?

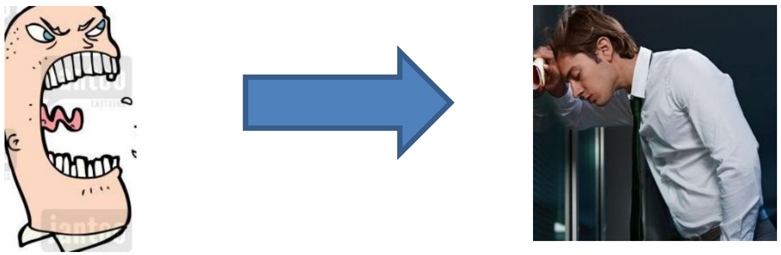
- Forty-four departments within a large organization have been sampled (N= 44)

- Each department has been rated:
- the extent to which the departmental supervisor practices "authoritarian style of leadership and decision making"
- the "efficiency (productivity) of workers within the department"

- Ask question: Does an association exist?

- Which is the likely dependent variable?
- Management style ⟶ efficiency

Does an association exist? Example



?

- The table below shows the relationship between:
- authoritarianism of supervisors (*X*) and
- the efficiency of workers (*Y*)
- Is there an association between these variables?

	Authoritarianism		
Efficiency	Low	High	Totals
Low	10	12	22
High	17	5	22
Totals	27	17	44

- An association exists if the conditional distributions of one variable change across the values of the other variable.
-

Efficiency by Authoritarianism, Frequencies (Percentages)

	Authoritarianism		
Efficiency	Low	High	Totals
Low	10 *(37.04%)*	12 *(70.59%)*	22
High	17 *(62.96%)*	5 *(29.41%)*	22
Totals	27 *(100.00%)*	17 *(100.00%)*	44

To calculate column percentages, each cell frequency is divided by the column total, then multiplied by 100:

- (10/27)*100 = 37.04%
- (12/17)*100 = 70.59%
- (17/27)*100 = 62.96%
- (5/17)*100 = 29.41%

Does an association exist?

Efficiency by Authoritarianism, Percentages

	Authoritarianism	
Efficiency	Low	High
Low	37.04%	70.59%
High	62.96%	29.41%
Totals	100.00%	100.00%

- The column percentages show efficiency of workers by authoritarianism of supervisor.
  - The column percentages do change (differ across columns), so these variables appear to be associated.
  - NOTE: FORMAL TEST OF STATISTICAL SIGNIFICANCE IS POSSIBLE (CHI SQUARE: Last week's lecture)

Reminder: 5 step procedure: Chi square test of independence

	Authoritarianism		
Efficiency	Low	High	Totals
Low	10	12	22
High	17	5	22
Totals	27	17	44

Performing the Chi Square Test Using the Five-Step Model

Step 1: Make Assumptions and Meet Test Requirements

- Independent random samples
- Level of measurement is nominal
- e.g. low or high on efficiency

Step 2: State the Null Hypothesis

- H_0: The variables are independent
  - Another way to state the H_0, more consistently with previous tests:
    - $H_0: f_o = f_e$

- H_1: The variables are dependent
  - Another way to state the H_1:
    - $H_1: f_o \neq f_e$

Step 3: Select Sampling Distribution and Establish the Critical Region

- Sampling Distribution = χ^2
- Alpha = .05
- df = (r-1)(c-1) = 1
- χ^2 (critical) = ?

Appendix C Distribution of Chi Square

Critical values at alpha =.05

df	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.000	.001	.004	.016	.064	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179

Step 3: Select Sampling Distribution and Establish the Critical Region

- Sampling Distribution = χ^2
- Alpha = .05
- df = (r-1)(c-1) = 1
- χ^2 (critical) = 3.841

In this case, χ^2 (critical) allows us to identify in our sampling distribution a value of χ^2 which is quite unlikely, i.e. less than a 5% chance of getting it if our null hypothesis is true

Step 4: Calculate the Test Statistic

- χ^2 (obtained) =

	Authoritarianism		
Efficiency	Low	High	Totals
Low	10	12	22
High	17	5	22
Totals	27	17	44

FORMULA 11.2

$$f_e = \frac{\text{Row marginal} \times \text{Column margi}}{N}$$

	Authoritariansim		
Efficiency			
	Low	High	Totals
Low	13.5	8.5	22
High	13.5	8.5	22
Totals	27	17	44

$\frac{(22*27)}{44}$ → Low/Low: 13.5

$\frac{(22*27)}{44}$ → High/Low: 13.5

$\frac{(22*17)}{44}$ → Low/High: 8.5

$\frac{(22*17}{44}$ → High/High: 8.5

Example *(continued)*

- A computational table helps organize the computations.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
10	13.5			
17	13.5			
12	8.5			
5	8.5			
TOTAL 44	44			

- Subtract each f_e from each f_o. The total of this column *must* be zero.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
10	13.5	-3.5		
17	13.5	3.5		
12	8.5	3.5		
5	8.5	-3.5		
44	44			

TOTAL

- Square each of these values

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
	10	13.5	-3.5	12.25	
	17	13.5	3.5	12.25	
	12	8.5	3.5	12.25	
	5	8.5	-3.5	12.25	
TOTAL	44	44			

Computation of Chi Square: An Example

(continued)

- Divide each of the squared values by the f_e for that cell. The sum of this column is chi square

f_0	f_e	f_0-f_e	$(f_0-f_e)^2$	$(f_0-f_e)^2/f_e$
10	13.5	-3.5	12.25	0.907407
17	13.5	3.5	12.25	0.907407
12	8.5	3.5	12.25	1.441176
5	8.5	-3.5	12.25	1.441176
44	44			4.697168

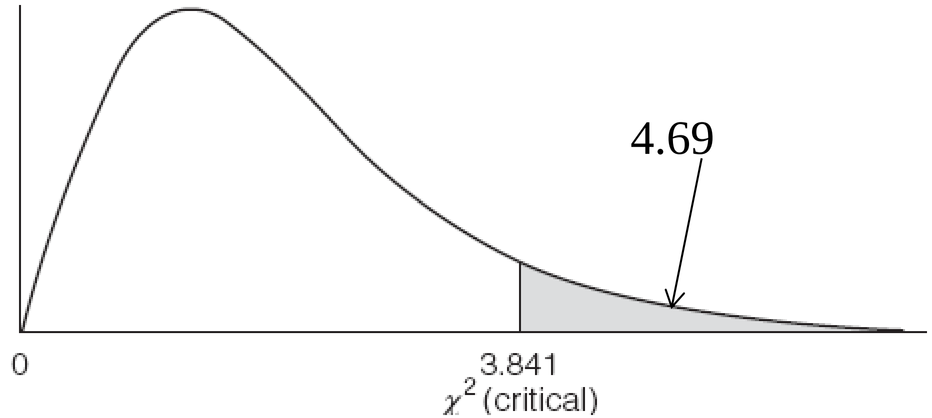
TOTAL

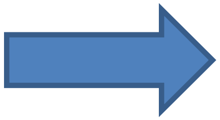
TEST STATISTIC -> 4.697

The larger the chi square, the more likely the association is significant

Step 5: Make Decision and Interpret Results

- χ^2 (critical) = 3.841
- χ^2 (obtained) = 4.69
- The test statistic is in the Critical (shaded) Region:
  - We reject the null hypothesis of independence.
  - Efficiency is associated with management style...





2. How Strong is the Association?

- NOTE: Chi square test of independence tells us "NOTHING" as to the strength of a relationship.. merely if there is a statistically significant association.. (yes or no)..

- The following two tables are of identical "strength".. (one has a sample which is merely 10X as large as the other's) -> would have identical column %'s

	Authoritarianism		
Efficiency	Low	High	Totals
Low	10	12	22
High	17	5	22
Totals	27	17	44

χ^2 (obtained) = 4.69

	Authoritarianism		
Efficiency	Low	High	Total
Low	100	120	220
High	170	50	220
Totals	270	170	440

χ^2 (obtained) = 46.97

The latter χ^2 (obtained) does not Imply that the association is 10 times as great!!!

2. How Strong is the Association?

- Previous example: identical % conditional distributions (column percentages), i.e. identical strength of association (the 2nd is merely with a larger sample and subsequently with a larger chi square)

- Differences in the strength of relationships are implied greater differences in percentages across columns (or conditional distributions).
  - In weak relationships, there is little or no change in column percentages.
  - In strong relationships, there is marked change in column percentages.

- One way to measure strength is to find the “maximum difference,” the biggest difference in column percentages for any row of the table.

  Note, the “maximum difference” method provides an easy way of characterizing the strength of relationships, but it is also limited.

Efficiency by Authoritarianism, Percentages

	Authoritarianism	
Efficiency	Low	High
Low	37.04%	70.59%
High	62.96%	29.41%
Totals	100.00%	100.00%

- The "Maximum Difference" is:
  - 70.59–37.04=33.55 percentage points.

The scale presented Table 11.5 can be used to describe (only arbitrary and approximately) the strength of the relationship”

TABLE 12.5 THE RELATIONSHIP BETWEEN THE MAXIMUM DIFFERENCE AND THE STRENGTH OF THE RELATIONSHIP

Maximum Difference	Strength
If the maximum difference is:	*The strength of the relationship is:*
between 0 and 10 percentage points	weak
between 11 and 30 percentage points	moderate
more than 30 percentage points	strong

Efficiency by Authoritarianism, Percentages

	Authoritarianism	
Efficiency	Low	High
Low	37.04%	70.59%
High	62.96%	29.41%
Totals	100.00%	100.00%

- The "Maximum Difference" is:
  - 70.59–37.04=33.55 percentage points.
  - Suggests is a strong relationship.

What if?

	Authoritarianism	
Efficiency	Low	High
Low	37.04%	40.59%
High	62.96%	59.41%
Totals	100.00%	100.00%

- The "Maximum Difference" is:
  - 62.59 – 59.04= 3.55 percentage points.
  - Suggests is a weak relationship.

  NOTE: OTHER POSSIBILITIES ->
  MEASURES OF ASSOCIATION ARE POSSIBLE that indicate "STRENGTH"!!
  (will return to this point later)

"Repeatedly concussed National Football League players," said the UNC report, "had five times the rate of mild cognitive impairment (pre-Alzheimer's) than the average population," while "retired NFL players suffer from Alzheimer's disease at a 37-per-cent higher rate than average." Then came the kicker. Two doctors determined "that the average life expectancy for all pro football players, including all positions and backgrounds, is 55. Several insurance carriers say it is 51 years."

NFL Linemen 1 in 5 will develop Alzheimer's in their lifetime..
Other men 1 in 9 develop Alzheimer's..

	Ex NFL Linemen		Other Americans	
Develops Alzheimer's	200	20.00%	111	11.10%
Does no develop Alzheimer's	800	80.00%	889	88.90%
Total Sample	1000		1000	

Do a chi square test (on your own time): Yes, it is significant!!

The Maximum Difference is:
88.90 – 80.00 -> 8.90.. So we'll consider this a relatively weak association..

- As mentioned earlier:
- Bivariate association can be investigated by finding answers to three questions:
  1. Does an association exist?
  2. How strong is the association?
  3. What is the pattern or direction of the association?

With regard to pattern??

Which scores of the variables tend to go together??

3. What is the Pattern of the Relationship?

- "Pattern" = which scores of the variables go together?

- Previous example:

Efficiency	Authoritarianism Low	Authoritarianism High
Low	37.04%	70.59%
High	62.96%	29.41%
	100.00%	100.00%

Question:
If someone scored "low" on authoritarianism: what would you predict on "efficiency"?
"High" (62.96% of cases)
"Low" on "Authoritarianism" tends to go with "High" on efficiency (62.96%)
If someone scored "high" on authoritarianism: what's your prediction?
"Low" (70.59% of cases)
High "Authoritarianism" tends to go with "Low" in efficiency (70.59%)

What is the Direction of the Relationship?

- If *both* variables are ordinal, we can discuss *direction* as well as *pattern*.
- In *positive* relationships, the variables vary in the same direction.
  - Low on *X* is associated with low on *Y*.
  - High on *X* is associated with high on *Y*.
  - As X increase, Y increases.
- In *negative* (inverse) relationships, the variables vary in opposite directions.
  - As one increases, the other decreases.

- Education and Income?
- Positive: As education goes up, we expect income to be higher (and vise versa)

- Hostile Parenting and Child Well-being
- Negative: Higher levels of hostile parenting is associated with “lower” levels of child well-being (and vise versa)

- Education of parents and academic success of children
- Positive: Better educated parents have more successful children (and vise versa)

- Number of hours work/weekly and time devoted to leisure activities/weekly
- Inverse: as hours of work increase, hours devoted to leisure decline (and vise versa)

- What about:
- “Religious affiliation and education”?
- If one or more variables is nominal., we can not speak of “direction”

Continuing with Chapter 11:

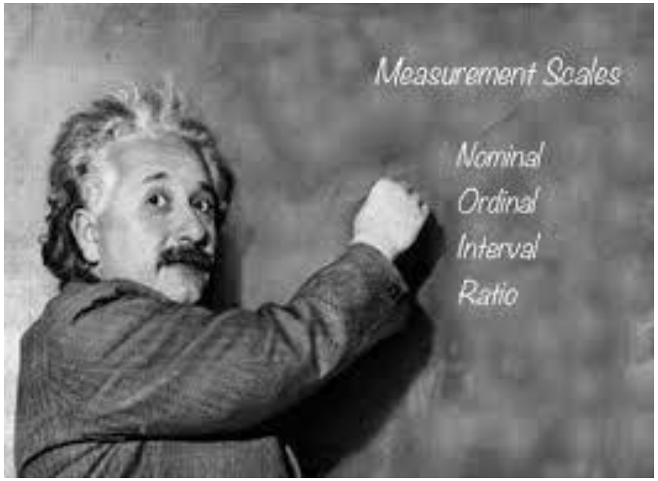
- Measures of association for nominal variables
- -> how strong is the relationship?
- (moving beyond comparing "column percentages")

It is also useful to have a summary measure
– a single number – to indicate the strength of the relationship.

For nominal level variables, there are two commonly used types of measures of association:

- Phi (φ) or Cramer's *V* (Chi square-based measures)
- Lambda (λ) (PRE measure)

Recall:
Nominal variable? You can merely classify cases, can’t rank order them..
Examples:
Religious affiliation
Country of Birth
Smoker/non-smoker,… etc..



Chi Square-Based Measures of Association

- Phi is used for 2x2 tables.
- Formula for phi:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

where the obtained chi square, χ^2, is divided by *N*, then the square root of the result taken.

Chi Square-Based Measures of Association *(continued)*

- Cramer's *V* is used for tables larger than 2x2.
- Formula for Cramer's *V*:

$$V = \sqrt{\frac{\chi^2}{(N)(\min r - 1, c - 1)}}$$

where $(\min r - 1, c - 1)$ = the minimum value of $r - 1$ (number of rows minus 1) or $c - 1$ (number of columns minus 1)

Chi Square-Based Measures of Association

• Phi and Cramer's *V* range in value from 0 (no association) to 1.00 (perfect association).

•Nothing on the "direction" of the relationship (why? Nominal)

• Phi and *V* are symmetrical measures; that is, the value of Phi and *V* will be the same regardless of which variable is taken as independent.

• General guidelines for interpreting the value of Phi and *V* are provided in Table 11.12

THE RELATIONSHIP BETWEEN THE VALUE OF NOMINAL-LEVEL MEASURES OF ASSOCIATION AND THE STRENGTH OF THE RELATIONSHIP

Value	Strength
If the value is	*The strength of the relationship is*
between 0.00 and 0.10	weak
between 0.11 and 0.30	moderate
greater than 0.30	strong

Chi Square-Based Measures of Association: An Example

The following problem is selected from Chapter 10 which was used to introduce the "chi square test" (pages 274-278)

Social Workers: Mobilizing Strengths in Individuals & Communities

A random sample of 100 social work graduates were classified in terms of whether the Canadian Association of Schools of Social Work (CASSW) accredited their undergraduate programs (independent variable) and whether they were hired in social work positions within three months of graduation (dependent variable).

	Accreditation Status		
	Accredited	**Not Accredited**	**Totals**
Employment Status			
Working as social worker	**30**	**10**	**40**
Not working as social worker	**25**	**35**	**60**
Totals	**55**	**45**	**100**

χ^2 (obtained) = 10.78

Example:

- We saw in Chapter 10 that this relationship was statistically significant:
- Chi square = 10.78, which was significant at the .05 level
- However, what about the strength of this association?

- To assess the strength of the association between CASSW accreditation and employment, phi is compute as:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

$$\phi = \sqrt{\frac{10.78}{100}}$$

$$\phi = 0.33$$

- A phi of .33 indicates what?
- Previous table,.. a strong relationship.., right?

Limitations of Chi Square-Based Measures of Association

- Phi is used for 2x2 tables only.
  - For larger tables, the maximum value of phi depends on table size and can exceed 1.0.
  - Use Cramer's *V* for larger tables.

Example: page 312 in text book

Academic Achievement by Student Club Memebership

Academic Achievement	Varisity	Club Membership Non-sports Club	No Membership	Totals
Low	4	4	17	25
Moderate	15	6	4	25
High	4	16	5	25
Totals	23	26	26	75

χ^2 (obtained) = 31.50

$$V = \sqrt{\frac{\chi^2}{(N)(\min r-1, c-1)}}$$

$$V = \sqrt{\frac{31.50}{(75)(2)}} = 0.46$$

Strong relationship between the two variables!!

- Phi (and Cramer’s*V*) are indices of the *strength* of the relationship *only*. They do *not* identify the pattern.
- With nominal:
- To analyze the pattern of the relationship, see the column percentages in the bivariate table.

Previous example

Academic Achievement by Student Club Memebership

Academic Achievement	Varisity	Club Membership Non-sports Club	No Membership	Totals
Low	4	4	17	25
Moderate	15	6	4	25
High	4	16	5	25
Totals	23	26	26	75

%

Academic Achievement by Student Club Memebership

Academic Achievement	Varisity	Club Membership Non-sports Club	No Membership	Totals
Low	17.39%	15.38%	65.38%	33.33%
Moderate	65.22%	23.08%	15.38%	33.33%
High	17.39%	61.54%	19.23%	33.33%
Totals	100.00%	100.00%	100.00%	100.00%

Lambda

- Lambda (λ) is a measure of association based on bivariate tables
- Like Phi (and *V*), Lambda (λ) is used to measure the strength of the relationship between nominal variables in bivariate tables.
- Like Phi (and *V*), the value of lambda ranges from 0.00 to 1.00.

- Unlike Phi (and *V*), Lambda has a more direct interpretation.
  - While Phi (and *V*) is only an **index** of strength, the value of Lambda tells us the **improvement** in predicting *Y* while taking *X* into account (PRE measure of association)

What is meant by Proportional Reduction in Error (PRE) Measure (of association)?

- Logic of PRE measures is based on two predictions:

1. **First prediction**: Ignore information about the independent variable, predict the score on the dependent variable, and inevitably make many errors (E_1)
2. **Second prediction**: Take into account information about the independent variable and on this basis, predict the value of the dependent. If the variables are associated we should make fewer errors (E_2).

Example: Assume you only had the following information on 50 Kings Students

50 Kings Students:	Frequency
Live on residence	10
Live off Campus (with roommate)	10
Live off Campus (with family)	30

The same 50 students are about to enter the room:
You only have the above information.

You had to predict the living arrangements for each student.

What would be your best guess?
Our best guess is "live off campus" with family..
We would be correct 30 times and wrong 20 times? $E_1 = 20$

What if you were given additional information on 50 Kings Students, i.e. Conditional distributions by year at Kings (1^{st}, 2^{nd} or 3^{rd})

50 Kings Students:	1st	2nd	3rd
Live on residence	10	0	0
Live off Campus (with roommate)	0	2	8
Live off Campus (with family)	20	6	4

The same 50 students are about to enter the room. You are told:

the first 30 are in Year 1. What would you predict?
-> "living off campus with family" (wrong 10 times, right 20)

the next 8 are second year? What would you predict?
-> "living off campus with family" (wrong 2 times, correct 6 times)

the next 12 are in 3^{rd} year? What would you predict?
Living off campus with roommate (wrong 4 times, correct 8)

Add the three together, we will be wrong 16 times, right?
This is better than how we did initially: we were wrong initially 20 times, right?
There is reduction in error when using information from another variable..

- Formula for Lambda:

FORMULA 13.3

$$\lambda = \frac{E_1 - E_2}{E_1}$$

Working with a bivariate table

E_1 = N – largest row total

E_2 = For each column, subtract the largest cell frequency from the col. total

Example (previous table)

Efficiency	Authoritarianism: Low	High	Totals
Low	10	12	22
High	17	5	22
Totals	27	17	44

$E_1 = 44 – 22 = 22$

$E_2 = (27 – 17) + (17 – 12) = 15$

$\lambda = (22- 15)/22 = .32$

Lambda: An Example *(continued)*

- A lambda of .32 means that authoritarianism (*X*) increases our ability to predict efficiency (*Y*) by 32%.
- According to the guidelines suggested in Table 11.12, a lambda of 0.32 indicates a strong relationship.

The Limitations of Lambda

1. Lambda is asymmetric: Value will vary depending on which variable is independent. Need care in designating independent variable.

2. When row totals are very unequal, lambda can be zero even when there is an association between the variables. For very unequal row marginals, better to use a chi-square based measure of association.

3. Lambda gives an indication of the *strength* of the relationship *only*.
   - It does *not* give information about pattern.
   - To analyze the pattern of the relationship, use the column percentages in the bivariate table.

One more example:
Is there a relationship between the status of women and the geographic region of a given country?

Logical dependent variable?

-> "status of women"…

Status of Women by Region

Women's Status	Africa	Latin Amer	Europe	Totals
Low	13	8	4	25
High	3	7	12	22
Totals	16	15	16	47

What of its strength??

Is there a significant relationship?
Chi square (obtained) = 10.17
5 step test of independence possible (skipped here)
This Chi square is much higher than critical value, hence: significant!!

$$V = \sqrt{\frac{\chi^2}{(N)(\min r-1, c-1)}}$$

$$V = \sqrt{\frac{10.17}{47}}$$

$$= 0.47$$

Cramer's V (=.47) suggests a strong relationship between the two variables

We can also calculate Lambda in this context…

Status of Women by Region

Women's Status	Africa	Latin Amer	Europe e	Totals
Low	13	8	4	25
High	3	7	12	22
Totals	16	15	16	47

$$\lambda = \frac{E_1 - E_2}{E_1}$$

Where:

E_1 = N – largest row total

E_2 = For each column, subtract the largest cell frequency from the col total & then add them up..

$E_1 = 47 – 25 = 22$

$E_2 = (16 – 13) + (15 – 8) + (16 – 12) = 14$

$\lambda = (22- 14)/22 = .36$

Lambda: 36% fewer errors of prediction using information from independent variable

Again: THIS IMPLIES A RELATIVELY STRONG RELATIONSHIP!!

Summary..

In this example:

Chi square tells us that it is significant!! i.e. association is not merely the by-product of sampling error

Cramer's V and Lambda both suggest a relatively strong relationship..

But what of the character of the relationship??

Status of Women by Region

Women's Status	Africa	Latin Amer	Europe	Totals
Low	13	8	4	25
High	3	7	12	22
Totals	16	15	16	47

Calculate Column Percentages:

Status of Women by Level of Development for 47 Nations

Women's Status	Africa		Latin Amer		Europe		Totals
Low	13	81.25%	8	53.33%	4	25.00%	25
High	3	18.75%	7	46.67%	12	75.00%	22
Totals	16	100.00%	15	100.00%	16	100.00%	47

Here we see the Status of women is Highest in Europe,…