Moving onto the next chapter:

- Bivariate (Cross tabulation) Tables
- The basic logic of Chi Square
- Perform the Chi Square test using the five-step model
- Limitations of Chi Square
- Measures of association (nominal level of measurement)

Why examine a “bivariate table”?
Example: We are conducting research on smoking & education..

Small sample (N=600), is there a significant association??

Bivariate Tables

- Bivariate tables: display the scores of cases on two different variables at the same time.

Cell Counts

Level of Education ← **INDEPENDENT VARIABLE**

DEPENDENT VARIABLE

Smoking Behavior		**< H.S.**	**H. School Grad**	**Some Post Sec**	
	No	60	100	300	460
	Yes	40	40	60	140
		100	140	360	600

Row marginals

column marginals

Total # of Cases (N)

Cell count for < HS and Non-smoker

More on Bivariate Tables

Cells are intersections of columns and rows.

- There will be as many cells as there are scores on the two variables combined.
- E.g. If 3 categories on dependent variable, and 5 categories on the indpendent, we have 3*5 = 15 cells

Marginals are the subtotals (either row or column)

N is the total number of cases in our cross tab..

- Crosstabs (or bivariate tables) provide evidence on potential "associations", i.e. two variables are said to be associated if the distribution of one variable changes for various categories of the other variable

For this course, we are following this convention:

- **Columns** will reflect different scores on the independent variable.
  - There will be as many columns as there are scores on the independent variable.
- **Rows** will reflect scores of the dependent variable.
  - There will be as many rows as there are scores on the dependent variable.

- Can calculate “column percentages”.

Cell Counts and Column % Level of Education

		< H.S.	H. School Grad	Some Post Sec	
Smoking	**No**	60	100	300	460
Behavior		60.00	71.43	83.33	
	Yes	40	40	60	140
		40.00	28.57	16.67	
		100	140	360	600

100/140*100 (→ 71.43)   60/360*100 (→ 16.67)

Interpretation:
40% of < HS smoke, in contrast to 28.57% among HS graduates
And 16.67% among those with some college

Note: When working with a bivariate table!!!

If dependent variable is in your rows..   USE column % in interpretation.. The row %'s can potentially be very misleading..

If dependent variable happened to be in your columns, you would have to use the "row %" in interpretation!!

What if?

Sample of 690 clerical workers (1980)

		Independent		
		Women	Men	total
Dependent				
	smokers	65	45	110
	non-smokers	500	80	580
	Total	565	125	690

Row % or Column %???

What if?

Sample of 690 clerical workers (1980)

Dependent		Independent		
		Women	Men	total
	smokers	65	45	110
	non-smokers	500	80	580
	Total	565	125	690

Row %

Dependent		Independent		
		Women	Men	total
	smokers	59.1%	40.9%	100.0%
	non-smokers	86.2%	13.8%	100.0%
	Total			

OR?

Column %

Dependent		Independent		
		Women	Men	total
	smokers	11.5%	36.0%	
	non-smokers	88.5%	64.0%	
	Total	100.0%	100.0%	

What if?

Sample of 690 clerical workers (1980)

	Independent		
Dependent	**Women**	**Men**	**total**
smokers	65	45	110
non-smokers	500	80	580
Total	565	125	690

Row %

	Independent		
Dependent	**Women**	**Men**	**total**
smokers	59[illegible]%	40.9%	100.0%
non-smokers	8[illegible]2%	13.8%	100.0%
Total			

OR?

Column %

	Independent		
Dependent	**Women**	**Men**	**total**
smokers	11.5%	36.0%	
non-smokers	88.5%	64.0%	
Total	100.0%	100.0%	

Cell Counts and Column %   Level of Education

Column %

		< H.S.	H. School Grad	Some Post Sec	
Smoking Behavior	No	60 60.00	100 71.43	300 83.33	460
	Yes	40 40.00	40 28.57	60 16.67	140
		100	140	360	600

OR (the exact same data) – both are okay, right?:

Row %

Smoking

Level of education		No	Yes	Total
	<H.S	60 60.0	40 40.0	100
	H. School Grad	100 71.4	40 28.6	140
	Some Post Sec.	300 83.3	60 16.7	360
	Total	460	140	600

- Interpret this table:

Independent variable

Incidence and % of Obesity by Province, 2008

Dependent variable

	Nfld	PEI	NS	NB	Quebec
Obese	173,298	36,998	230,913	229,299	1,739,628
Not Obese	336,402	105,302	711,588	522,501	6,167,772
Total	509,700	142,300	942,500	751,800	7,907,400

Interpretation

Not obvious with counts..

Can calculate column percentages to aid in interpretation since dependent variable is in the rows

Also: formal test of significance is possible... (chi square)

Interpretation?

Incidence and % of Obesity by Province, 2008

	Nfld	PEI	NS	NB	Quebec
Obese	173,298	36,998	230,913	229,299	1,739,628
	34.00%	26.00%	24.50%	30.50%	22.00%
Not Obese	336,402	105,302	711,588	522,501	6,167,772
	66.00%	74.00%	75.50%	69.50%	78.00%
Total	509,700	142,300	942,500	751,800	7,907,400
	100.00%	100.00%	100.00%	100.00%	100.00%

An association “appears to exist” between province of residence and obesity; the distribution of obese and non-obese vary across provinces e.g. 34% of Nfld are obese, as apposed to only 22% of Quebec residents

NOTE: VERY LARGE #s of cases in the study here: LIKELY REAL!!!

What if we are working with relatively small numbers?

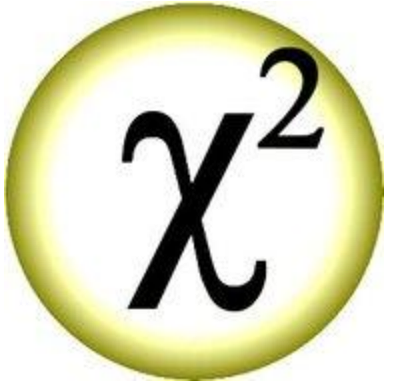
- Can we be sure an association (relationship) really exists for the larger population even if the %'s differ ???

Incidence and % of Obesity by Province, 2008

	Nfld	PEI	NS	NB	Quebec
Obese	17	4	23	23	17
	33.33%	26.67%	24.47%	30.67%	21.52%
Not Obese	34	11	71	52	62
	66.67%	73.33%	75.53%	69.33%	78.48%
Total	51	15	94	75	79

- Numbers here are quite small.. Might the variation merely be the by-product of sampling error?
- There is a formal test to see whether the differences are significant or not -> chi square test..

Our Chi Square test is also called, the Chi Square test of ***“Independence”***….



What do we mean by “Independence” in this context?

The opposite of having an “association between two variables”… i.e. an absence of any type of association or relationship

- With this table? Is there a relationship between the two variables??

TABLE 11.2 THE CELL FREQUENCIES THAT WOULD BE EXPECTED IF RATES OF PARTICIPATION AND SEX WERE INDEPENDENT

	Sex					
Participation Rates	Male		Female			%
High	50	66.67%	50	66.67%	100	66.7
Low	25	33.33%	25	33.33%	50	33.3
	75		75		150	100

Males are no more likely to participate than Females
NO RELATIONSHIP

"Independence"

- Two variables are **independent** if the classification of a case into a particular category of one variable has no effect on the probability that the case will fall into any particular category of the second variable.

- Let us return to our example with education and smoking…

Cell Counts and Column % Level of Education

		< H.S.	H. School Grad	Some Post Sec		
Smoking Behavior	**No**	60 60.00	100 71.43	300 83.33	460	77%
	Yes	40 40.00	40 28.57	60 16.67	140	23%
		100	140	360	600	100%

- Complete "Independence" would look like:

Expected frequencies, if we had independence..

Smoking behavior	< HS	H.School Grad	Some Post sec	
No	**77** **77%**	**107** **77%**	**276** **77%**	460 77%
Yes	**23** **23%**	**33** **23%**	**84** **23%**	140 23%
	100	**140**	**360**	600

Basic Logic of Chi Square TEST

- Again, a fundamental 5 step model!!!
- Question to answer:
- Does an "association" really exist? (given N)
- Or do we have "independence"?

- Chi Square, χ^2, is a test of significance based on bivariate, cross tabulation tables.
- Chi Square is a test for **independence**.

- Specifically, we are looking for significant differences between the *observed* cell frequencies in a table (f_o) and those that would be *expected* by random chance or if cell frequencies were **independent** (f_e):

Formulas for Chi Square

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

.. Gives us our "expected frequencies" under assumption of "independence"

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

Formal test statistic

Step 4!

where f_o = the cell frequencies observed in the bivariate table

f_e = the cell frequencies that would be expected if the variables were independent

Computation of Chi Square: An Example

- Is there a relationship between support for privatization of healthcare and political ideology? Are liberals significantly different from conservatives on this variable?
  - The table below reports the relationship between these two variables for a random sample of 78 adult Canadians.

	Political Ideology		
Support	Conservative	Liberal	Total
No	14	29	43
Yes	24	11	35
Total	38	40	78

How do we calculate our “test statistic” in our chi squared test of independence?

Must first use:

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

And then calculate:

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

where f_o = the cell frequencies observed in the bivariate table

f_e = the cell frequencies that would be expected if the variables were independent

An Example *(continued)*

Observed Frequencies (fo)

	Conservative	Liberal	Total
No	14	29	43
Yes	24	11	35
Total	38	40	78

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

Use Formula 10.2 to find f_e.

– To obtain fe multiply column and row marginals for each cell and divide by *N*.

- (38*43)/78 = 1634 /78 = 20.9
- (40*43)/78 = 1720 /78 = 22.1
- (38*35)/78 = 1330 /78 = 17.1
- (40*35)/78 = 1400 /78 = 17.9

Expected frequencies (f_e)

Political Ideology

Support	Conservative	Liberal	Total
No	20.9	22.1	43
Yes	17.1	17.9	35
Total	38	40	78

Example:

Observed: (f_0)

Support	Conservative	Liberal	Total
	Political Ideology		
No	14	29	43
Yes	24	11	35
Total	38	40	78

Expected frequencies (f_e)

Support	Conservative	Liberal	Total
	Political Ideology		
No	20.9	22.1	43
Yes	17.1	17.9	35
Total	38	40	78

OUR test statistic tells us whether these are Significantly different!!

Example *(continued)*

- A computational table helps organize the computations.

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

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
	14	20.9			
	29	22.1			
	24	17.1			
	11	17.9			
TOTAL	78	78			

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

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
14	20.9	-6.9		
29	22.1	6.9		
24	17.1	6.9		
11	17.9	-6.9		
78	78	0		

TOTAL

- Square each of these values

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
	14	20.9	-6.9	47.61	
	29	22.1	6.9	47.61	
	24	17.1	6.9	47.61	
	11	17.9	-6.9	47.61	
TOTAL	78	78	0		

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_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
	14	20.9	-6.9	47.61	2.28
	29	22.1	6.9	47.61	2.15
	24	17.1	6.9	47.61	2.78
	11	17.9	-6.9	47.61	2.66
TOTAL	78	78	0		$\chi^2 = 9.87$

What to do with this chi square? 9.87?

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

We need a formal test…

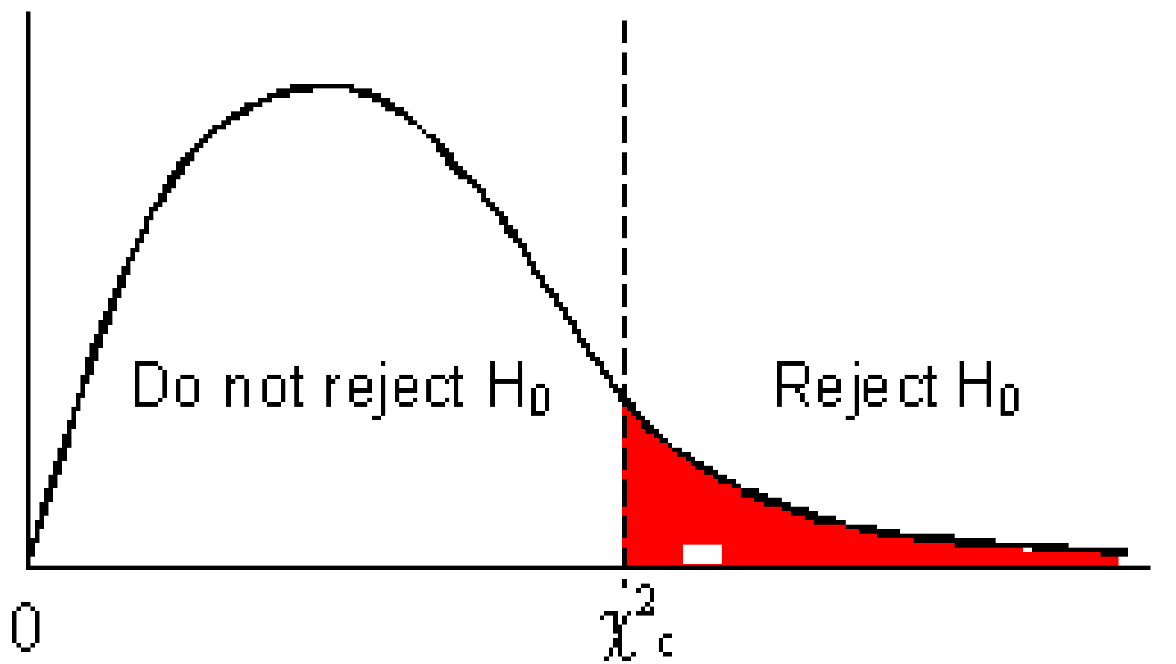
What about our “sampling distribution” and “critical score” in our Formal test?

Here, we use a sampling distribution called the

CHI square sampling distribution….

The Chi Square Distribution

- Type of sampling distribution
- The chi square distribution is asymmetric and its values are always positive (Appendix C).
- Its shape varies by the degrees of freedom involved in the test , which in turn is determined by the number of columns and rows in the table



Working with the chi square distribution

- χ^2 can be calculated for any bivariate table
- The shape of the χ^2 distribution is influenced by the number of rows and columns in the table df=(r-1)(c-1)
- The sampling distribution we are working with in this case (TABLE C) relates to all possible χ^2 under a hypothetical situation whereby we have independence with a table of given size (# of columns, # of rows)
- With our significance test, we work with this χ^2 distribution (with the null hypothesis that we have "independence"), and determine whether our test statistic χ^2 is likely or not,.. under this assumption
- If highly unlikely (we set our alpha at .05), we reject our null hypothesis, and conclude significance
- 95% confident that there is a relationship,.. If we set our alpha value at .05 and our test score falls within the critical area..

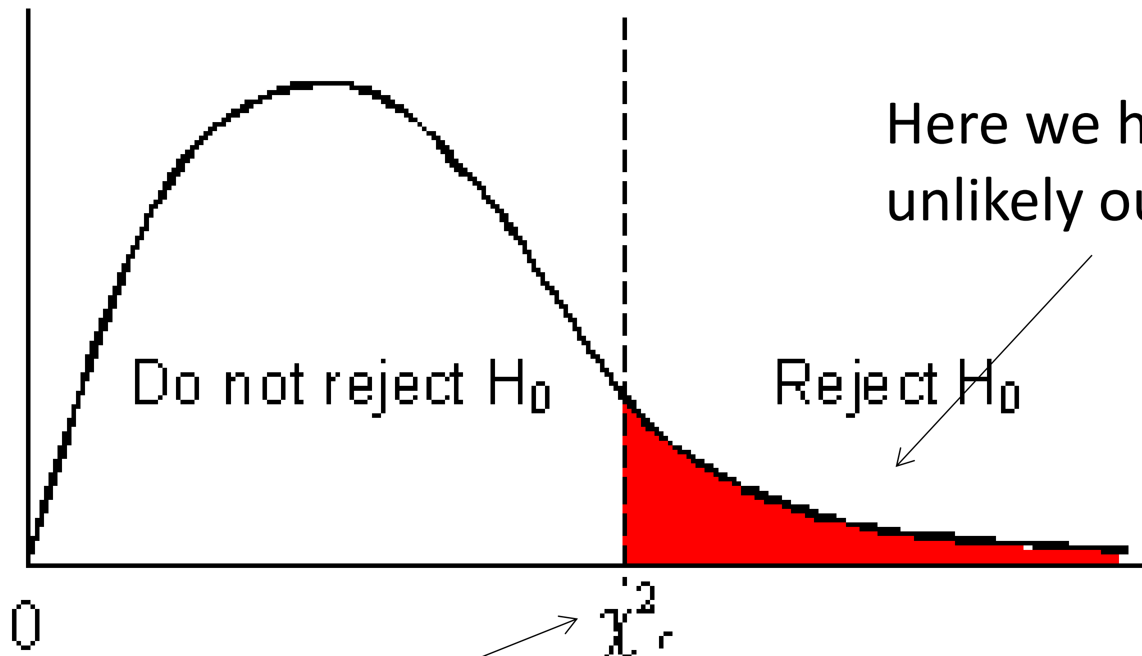
Appendix C Distribution of Chi Square

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

The Chi Square Distribution

- The chi square distribution is asymmetric and its values are always positive (Appendix C).
- Its shape varies by the degrees of freedom involved in the test



Here we have highly unlikely outcomes

Appendix provides us with critical values for our test

We use an alpha of .05 unless otherwise specified

Back to our example

- Is there a relationship between support for privatization of healthcare and political ideology? Are liberals significantly different from conservatives on this variable?
  - The table below reports the relationship between these two variables for a random sample of 78 adult Canadians.

	Political Ideology		
Support	Conservative	Liberal	Total
No	14	29	43
Yes	24	11	35
Total	38	40	78

Performing the Chi Square Test Using the Five-Step Model

Step 1: Make Assumptions and Meet Test Requirements

- Independent random samples
- e.g. independent samples of conservatives & liberals
- Level of measurement is nominal
- e.g. support for privatization

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) = ?

2 rows and 2 columns, hence: df = 1

	Political Ideology		
Support	Conservative	Liberal	Total
No	14	29	43
Yes	24	11	35
Total	38	40	78

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

Using Table C (page 510) in our appendix, we can indentify the χ^2 (critical) for alpha = .05
This χ^2 (critical) varies by the size of the table (# of rows/columns)

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. Get our test statisitc

Observed Frequencies (fo)

	Conservative	Liberal	Total
No	14	29	43
Yes	24	11	35
Total	38	40	78

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

Use Formula 10.2 to find f_e.

– To obtain fe multiply column and row marginals for each cell and divide by *N*.

- (38*43)/78 = 1634 /78 = 20.9
- (40*43)/78 = 1720 /78 = 22.1
- (38*35)/78 = 1330 /78 = 17.1
- (40*35)/78 = 1400 /78 = 17.9

Expected frequencies (f_e)

Political Ideology

Support	Conservative	Liberal	Total
No	20.9	22.1	43
Yes	17.1	17.9	35
Total	38	40	78

Step 4: Calculate the Test Statistic

As demonstrated earlier:

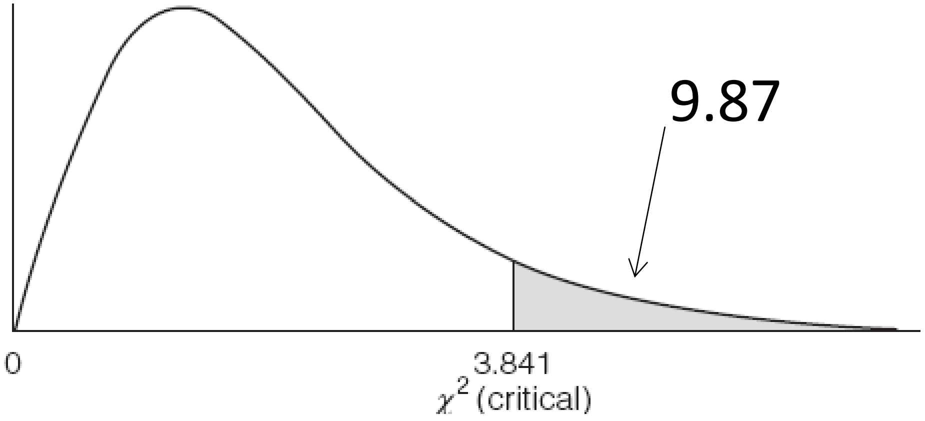
f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
14	20.9	-6.9	47.61	2.28
29	22.1	6.9	47.61	2.15
24	17.1	6.9	47.61	2.78
11	17.9	-6.9	47.61	2.66
78	78	0		$\chi^2 = 9.87$

Step 4: Calculate the Test Statistic

- χ^2 (obtained) = 9.87

Step 5: Make Decision and Interpret Results

- χ^2 (critical) = 3.841
- χ^2 (obtained) = 9.87
- 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.



- Another example:
- Is there a relationship between where one studies and whether or not one works while studying?
  - The table below reports the relationship between these two variables for a random sample of 1320 students at UWO.

Work Status (working or not?)

	Kings	UWO (main)	totals
Not working	420	660	1080
Working	120	120	240
	540	780	1320

Performing the Chi Square Test Using the Five-Step Model

Step 1: Make Assumptions and Meet Test Requirements

- Independent random samples
- UWO and Kings samples
- Level of measurement is nominal
- work status

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

Table was 2 X 2, so df=1
(r-1)(c-1)

Critical values at alpha =.05
With 1 degree of freedom

f	.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	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

Here we have our "observed cells".. f_0

Work Status (working or not?)

	Kings	UWO (main)	totals
Not working	420	660	1080
Working	120	120	240
	540	780	1320

An Example *(continued)*

- Use Formula 11.2 to find f_e.

FORMULA 11.2

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

Work Status (working or not?)

	Kings	UWO (main)	totals
Not working	420	660	1080
Working	120	120	240
	540	780	1320

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

Expected Work Status (assuming independence)

	Kings	UWO (main)	totals
Not working	441.82	638.18	1080
Working	98.18	141.82	240
	540	780	1320

$\frac{(1080*540)}{1320}$ → 441.82

$\frac{(1080*780)}{1320}$ → 638.18

$\frac{(240*540)}{1320}$ → 98.18

$\frac{(240*780)}{1320}$ → 141.82

Example *(continued)*

- A computational table helps organize the computations.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
420	441.82			
660	638.18			
120	98.18			
120	141.82			
TOTAL				
1320	1320			

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

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
	420	441.82	-21.82		
	660	638.18	21.82		
	120	98.18	21.82		
	120	141.82	-21.82		
TOTAL	1320	1320			

- Square each of these values

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
420	441.82	-21.82	476.03	
660	638.18	21.82	476.03	
120	98.18	21.82	476.03	
120	141.82	-21.82	476.03	
TOTAL 1320	1320			

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

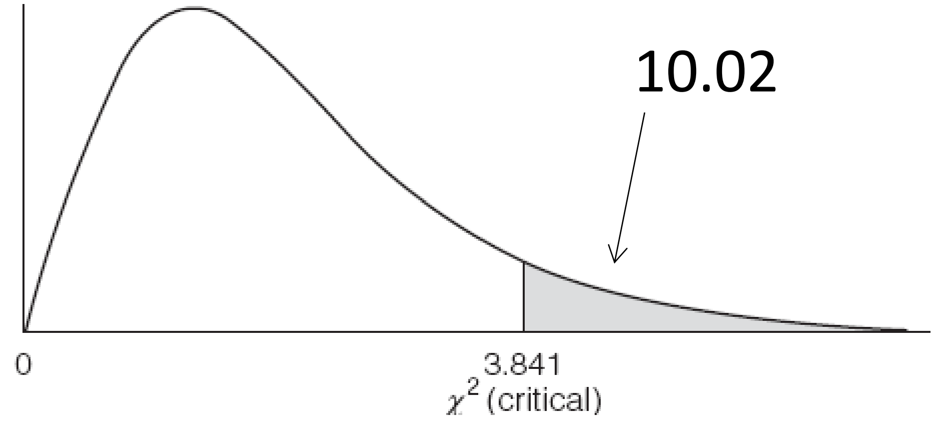
f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
420	441.82	-21.82	476.03	1.08
660	638.18	21.82	476.03	0.75
120	98.18	21.82	476.03	4.85
120	141.82	-21.82	476.03	3.36
TOTAL 1320	1320			10.02

TEST STATISTIC -> 10.02

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) = 10.02
- The test statistic is in the Critical (shaded) Region:
  - We reject the null hypothesis of independence.
  - Where one studies,.. Is associated with whether one works part time…



Interpreting Chi Square

- The chi square test tells us *only* if the variables are independent or not.
- It does not tell us the pattern or nature of the relationship.
    - To investigate the pattern, compute %'s within each column and compare across the columns.

Interpreting Chi Square *(continued)*

Work Status (working or not?)

	Kings	UWO (main)	totals
Not working	420	660	1080
	77.78%	84.62%	
Working	120	120	240
	22.22%	15.38%	
	540	780	1320

- This relationship has a clear pattern. Kings students are more likely to be working part time.
  - Chi square told us that this relationship is significant (unlikely to be caused by random chance) and now, with the aid of column percents, we know how the two variables are related.

Sociology of sport

Who’s likely to be “successful” with sport, who’s most likely to give up on it at a young age?

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

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

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

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
24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	23.337	[illegible]	29.553	33.196	36.415	40.270	42.980	51.179

Step 4. Calculate our test statistic

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

fo	fe	fo-fe	$(fo-fe)^2$	$(fo-fe)^2/fe$
37				
63				
30				
70				
18				
82				
15				
85				

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

where f_o = the cell frequencies observed in the bivariate table

f_e = the cell frequencies that would be expected if the variables were independent

From a sample of 400 students

fo

	First Quarter	Second Quarter	Third Quarter	Fourth Quarter	
	Jan-March	April-June	July-Sept	Oct-Dec	
Athlete	37	30	18	15	100
Non-Athlete	63	70	82	85	300
	100	100	100	100	400

fe

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

From a sample of 400 students

fo

	First Quarter	Second Quarter	Third Quarter	Fourth Quarter	
	Jan-March	April-June	July-Sept	Oct-Dec	
Athlete	37	30	18	15	100
Non-Athlete	63	70	82	85	300
	100	100	100	100	400

fe

	First Quarter	Second Quarter	Third Quarter	Fourth Quarter
	Jan-March	April-June	July-Sept	Oct-Dec
Athlete	25	25	25	25
Non-Athlete	75	75	75	75

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

Step 4. Calculate our test statistic

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

fo	fe	fo-fe	$(fo\text{-}fe)^2$	$(fo\text{-}fe)^2/fe$
37	25			
63	75			
30	25			
70	75			
18	25			
82	75			
15	25			
85	75			

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

where f_o = the cell frequencies observed in the bivariate table

f_e = the cell frequencies that would be expected if the variables were independent

Step 4. Calculate our test statistic

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

fo	fe	fo-fe	$(fo-fe)^2$	$(fo-fe)^2/fe$
37	25	12		
63	75	-12		
30	25	5		
70	75	-5		
18	25	-7		
82	75	7		
15	25	-10		
85	75	10		

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

where f_o = the cell frequencies observed in the bivariate table

f_e = the cell frequencies that would be expected if the variables were independent

Step 4. Calculate our test statistic

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

fo	fe	fo-fe	$(fo-fe)^2$	$(fo-fe)^2/fe$
37	25	12	144	
63	75	-12	144	
30	25	5	25	
70	75	-5	25	
18	25	-7	49	
82	75	7	49	
15	25	-10	100	
85	75	10	100	

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

where f_o = the cell frequencies observed in the bivariate table
f_e = the cell frequencies that would be expected if the variables were independent

Step 4. Calculate our test statistic

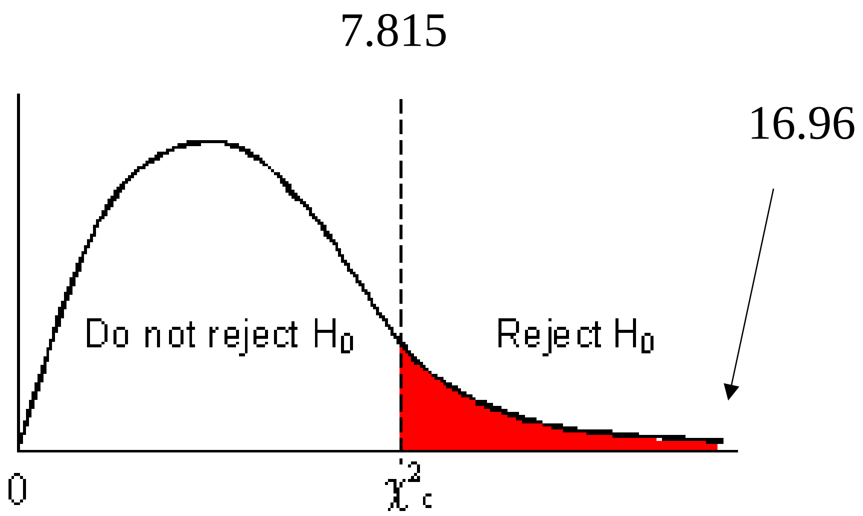
$$f_e = \frac{\text{Row marginal} \times \text{Column marginal}}{N}$$

fo	fe	fo-fe	$(fo-fe)^2$	$(fo-fe)^2/fe$
37	25	12	144	5.7600
63	75	-12	144	1.9200
30	25	5	25	1.0000
70	75	-5	25	0.3333
18	25	-7	49	1.9600
82	75	7	49	0.6533
15	25	-10	100	4.0000
85	75	10	100	1.3333
				16.96

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

where f_o = the cell frequencies observed in the bivariate table
f_e = the cell frequencies that would be expected if the variables were independent

Step 5. Make a decision, using this test statistic and our critical region.



Reject null hypothesis…

There is a significant association between the time of year in which one is born and whether or not one is very successful as an athlete

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)..

- Bivariate association can be investigated by finding answers to three questions:

1. Does an association exist (significance)?
2. What is the pattern or direction of the association?
3. How strong is the association?

1. Does an association exist?

- To detect association within bivariate tables:

    1. Chi Square test of independence, formally determines statistical significance.

2. What is the pattern?

- To examine pattern within bivariate tables:

  1. Calculate percentages within the categories of the independent variable.
  2. Compare percentages across the categories of the independent variable.

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**"

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

	First Q	Second Q	Third Q	Fourth Q
Athlete	37%	30%	18%	15%
Non-Athlete	63%	70%	82%	85%

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

- Evidence for 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%

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 NECESSARY TO DECISIVELY DETERMINE ASSOCIATION (CHI SQUARE)

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 ordinal
-     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

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

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

Low/Low: $\frac{(22*27)}{44}$

Low/High: $\frac{(22*17)}{44}$

High/Low: $\frac{(22*27)}{44}$

High/High: $\frac{(22*17}{44}$

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

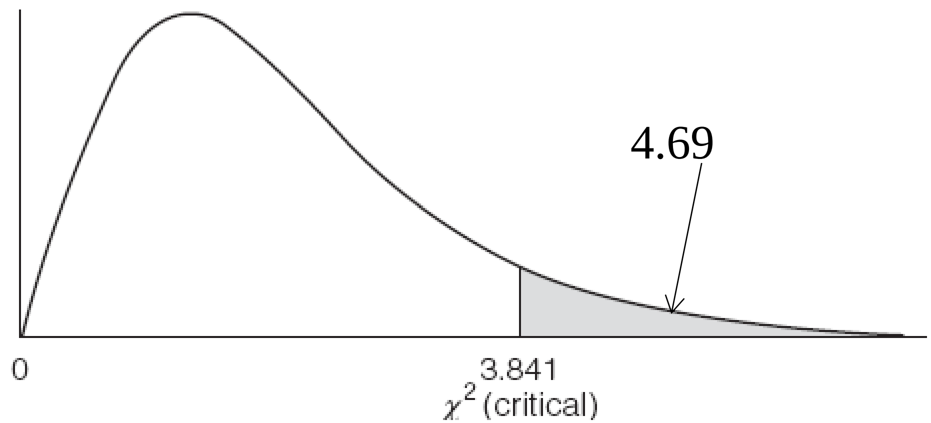
TEST STATISTIC -> 4.697

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. What is the pattern of the association?

	Authoritarianism				
Efficiency	Low		High		Totals
Low	10	37.0%	12	70.6%	22
High	17	63.0%	5	29.4%	22
Totals	27	100.0%	17	100.0%	44

In this example, among those who worked in workplaces with "low Authoritarian" management style, fully 63% ranked "high" on efficiency

Compare that with those who were in the "high authoritarianism" workplaces, where only 29.4% ranked "high" on efficiency.

3. How strong is the association?

To what extent do the conditional distributions of your "dependent variable" differ???

Sweden					
Smoking		Men		Women	
No		88%		89%	
Yes		12%		11%	
		100%		100%	

Japan					
Smoking	Men		Women		
No		65%		95%	
Yes		35%		5%	
		100%		100%	

Note: Sweden is close to "independence" on this, but it may still be significant if the sample size is large enough.. In Japan, we see a considerable departure from "independence" on these two variables (i.e. a stronger relationship)

3. 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	37.0%	12	70.6%	22
High	17	63.0%	5	29.4%	22
Totals	27	100.0%	17	100.0%	44

→ χ^2 (obtained) = 4.69

	Authoritarianism				
Efficiency	**Low**		**High**		**Total**
Low	**100**	37.0%	**120**	70.6%	**220**
High	**170**	63.0%	**50**	29.4%	**220**
Totals	**270**	100.0%	**170**	100.0%	**440**

→ χ^2 (obtained) = 46.97

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

3. 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”

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

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

THIS TABLE ISN'T EXHAUSTIVE AND MUTUALLY EXCLUSIVE

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

If the maximum difference is:

between 0 and 9.99%	**weak**	**USE**
between 10 and 29.99%	**moderate**	
30 percentage points or more	**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..