Chapter 10:
Relationships Between Two Variables

1. ***Constructing a Bivariate Table***
2. **Elaboration**
   - ***Spurious relationships***
   - ***Intervening relationships***
   - ***Conditional Relationships***

Introduction

- **Bivariate Analysis:** A statistical method designed to detect and describe the relationship between two nominal or ordinal variables (typically independent and dependent variables).
- **Cross-Tabulation:** A technique for analyzing the relationship between two or more nominal or ordinal variables. Allows for consideration of "control" variables.

Constructing a Bivariate Table

- ***Column variable:*** A variable whose categories are the columns of a bivariate table (from my experience it is usually the dependent variable).
- ***Row variable:*** A variable whose categories are the rows of a bivariate table (from my experience it is usually the independent variable).
- ***Marginals:*** The row and column totals in a bivariate table.

Example of Bivariate Crosstabulation:

Support for Abortion by Job Security
(absolute numbers provided)

Support for Abortion		Job Security: Can Find Job Easy	Job Security: Can Not Find Job Easy	Row Total
	Yes	24	25	49
	No	20	26	46
Column Total		44	51	95

What are the dependent and independent variables? What are the column and row variables? Marginals?
What is the disadvantage of providing only absolute numbers?
What is the advantage of providing percentages?

Constructing a Bivariate Table:

Percentages Can Be Computed in Different Ways:

1. *Column* Percentages: column totals as base
2. *Row* Percentages: row totals as base

We typically provide percentages for the independent variable

Column Percentages

Effect of Job Security on Support for Abortion
(absolute numbers in parentheses)

Abortion	Can Find Job Easy	Can Not Find Job Easy	Row Total
Yes	55% (24)	49% (25)	52% (49)
No	45% (20)	51% (26)	48% (46)
Column Total	100%	100%	

Questions to Answer When Examining a Bivariate Relationship

1. What are the **dependent** and **independent** variable?
2. Does there appear to be a **relationship**? (the chi square statistic is usually used with a crosstabulation)
3. How **strong** is it? (There are "measures of association" that will indicate the strength of the relationship. We will learn a few such as "lambda" and "gamma")
4. What is the **direction** of the relationship?

Direction of the Relationship

- ***Positive relationship***: A relationship between two variables (i.e., a bivariate relationship) measured at the ordinal level or higher in which the variables vary in the **same direction (both go up or both go down)**.
- ***Negative relationship***: A bivariate relationship measured at the ordinal level or higher in which the variables vary in **opposite directions (when one goes up the other goes down)**.

A Positive Relationship
(as class goes up "health" goes up)

Table 6.8 **Health Condition by Social Class: A Positive Relationship**

	CLASS		
HEALTH	Low	Middle	High
Poor	39%	12%	9%
Fair	36%	45%	28%
Good	25%	43%	63%
Total	100%	100%	100%
(N)	(39)	(254)	(202)

Source: General Social Survey, 1987 to 1992.

A Negative Relationship
(as "class" goes up "traumas" go down)

Table 6.9 **Frequency of Trauma by Social Class: A Negative Relationship**

	CLASS		
TRAUMA	Low	Middle	High
0	31%	41%	48%
1	22%	42%	20%
2+	47%	17%	32%
Total	100%	100%	100%
(N)	(48)	(220)	(180)

Source: General Social Survey, 1987 to 1992.

More Examples

Which are likely to be positive relationships and which negative relationships?

1. The relationship between hrs. studying and grades
2. The relationship between partying and grades
3. The relationship between "amount of sleep" and grades
4. The relationship between "color of shoes" and grades

Elaboration

- **Elaboration** is a process designed to further explore a bivariate relationship; it involves the introduction of a "control" variable (it's the process of "elaborating" on the relationship between two variables by considering a third variable).
- A **control variable** is an additional variable considered in a bivariate relationship. This third variable is "controlled for" when we examine the relationship between two variables.

Elaboration Tests

- Spurious relationships
- Intervening relationships
- Conditional Relationships

1. Testing for a spurious relationship

- A **Spurious relationship** is a relationship in which both the independent variable (IV) and the dependent variable (DV) are influenced by a third variable. The IV and DV are not causally linked, although it might appear so if one was unaware of the third variable.
- The relationship between the IV and DV is said to be "explained away" by the control variable.
- A **Direct causal relationship** is a relationship between two variables that cannot be accounted for (or explained away) by other variables. It is a "nonspurious" relationship.

Example of a Bivariate Relationship that appears to be spurious, prior to testing for spuriousness:

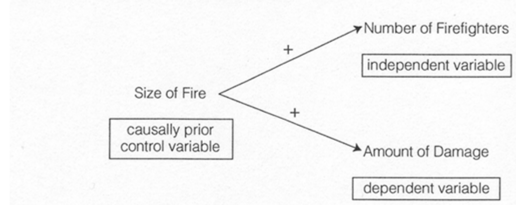
of Firefighters and Property Damage

Number of firefighters at the scene of a fire (IV)

↓ +

Property Damage (DV)

Figure 6.5 **Spurious Relationship**



A second example of a spurious relationship: A relationship between two variables prior to considering a third variable:

(that is, prior to elaboration)

(two variables appear related)

Sale of Ice Cream

↓ +

Number of Outdoor Crimes

Example of a third (control) variable causing a "spurious" relationship:

(elaboration considers control variables)

Outdoor temperature → (+) Sale of Ice Cream

Outdoor temperature → (+) Number of Outdoor Crimes

In-Class Assignment:

Write down an example of a spurious relationship (don't confer with your neighbor just do the best you can to think of one)

Identify the dependent and independent variable and the "control" variable that is causing the spurious relationship?

2. Elaboration can test for an <u>intervening</u> relationship

- ***Intervening relationship:*** a relationship in which the control variable intervenes between the independent and dependent variables.
- ***Intervening variable:*** a control variable that follows an independent variable but precedes the dependent variable in a causal sequence.

<u>Intervening Relationship</u>:
Examination of two variables prior to considering a third "intervening" variable

Attending Week Day Parties —(–)→ Grades

(independent variable) (dependent variable)

Examination of an intervening variable between two other variables

Attending Weekday Parties —(–)→ Hours Studying —(+)→ Grades

(IV) (Intervening Variable) (DV)

In-Class Assignment:

What is another example of an intervening relationship?

What is the dependent and independent variable and what is the "control" variable that is intervening between the two variables?

3. Elaboration tests for Conditional Relationships

- ***Conditional relationship:*** a relationship in which the independent variable's effect on the dependent variable depends on (or is conditioned by) a category of a control variable.
- The relationship between the independent and dependent variables will change according to the different conditions (or categories) of the control variable.

Example of a relationship between two variables prior to considering a "conditional" control variable

\# of Dolls owned —(+)→ Hours spent Playing with Dolls

Example of a Conditional Relationship

\# of Dolls owned —(+)→ Hours spent Playing with Dolls

↑

Gender of Child
(two categories: male and female)

Another Example of a Conditional Relationships

Empowering the Employee (letting him/her make decisions) —(+)→ Job Satisfaction

Another Example of a Conditional Relationships

Empowering the Employee (letting him/her make decisions) —(+)→ Job Satisfaction

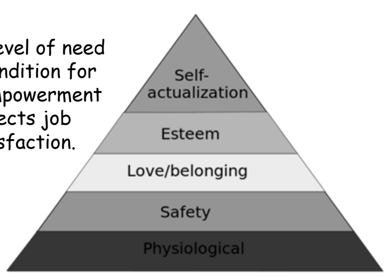
↑

Maslow's Hierarchy of Need

(categories range from basic "physical" needs to "self actualization" and "esteem" needs)

Maslow's Hierarchy of Needs, 1943

One's level of need is a condition for how empowerment affects job satisfaction.



Three Goals of Elaboration

1. Elaboration allows us to test for **spurious relationships**
2. Elaboration clarifies the causal sequence of bivariate relationships by introducing variables hypothesized to **intervene** between the IV and DV.
3. Elaboration specifies the different **conditions** under which the original bivariate relationship might hold.

Religion and Society (negative effects)

http://www.youtube.com/watch?v=YxTZv8c_GBM&feature=related

Chapter 11:

Test of Statistical Significance (such as t-test and Chi Square)

and

Measures of Association

A test of statistical significance, like the t-test and chi square, gives us the probability that the null hypothesis is correct.

If the t value shows us that there is a 5% (.05) or less chance that the null hypothesis is correct, we reject the null hypothesis and accept the research hypothesis.

What is the Chi Square?

Just like the t-test, the Chi Square is a test of statistical significance providing the level of probability that the null hypothesis is true.

If the probability of it being true is less than 5% (.05), we will reject the null hypothesis and accept the research hypothesis.

Why do we need a Chi Square test when we have the t-test?

The t-test requires interval level data. A chi square can be used with ordinal or nominal level data.

What is the difference between a test of statistical difference and measures of association?

1. A test of statistical significance tests whether we should reject the null hypothesis and accept the research hypothesis.
2. Measures of association (such as lambda and gamma) measure the "strength" of the relationship"

Measures of Association:
(such as Lambda and Gamma)

examine the <u>size</u> or <u>strength</u> of the association between two variables in a sample (not focused on whether or not the null-hypothesis can be rejected).

Typically, if the null hypothesis cannot be rejected (i.e., we assume there is no statistical association between the two variables), then we ignore the "strength" of the association found since whatever it is, we have determined it is due to sampling error.

Suppose we found that the strength/size of the association between two variables was large, BUT, the test of statistical significance indicated that we cannot reject the null hypothesis of no association.

How would we interpret these results?

<u>Answer</u>: in most cases, if Chi Square is not significant then we must assume that the two variables are not associated (we cannot reject the null hypothesis of "no association") even if the measure of association (e.g., lambda, gamma) is large.

If the two variables are not associated, then it doesn't matter how "strong" the relationship is in our <u>sample</u> since the probability is sufficiently high that the two variables are not associated with one another in the population.

How does Chi Square help us determine the level of probability that the null hypothesis is true?

That is, the probability that the association/relationship found between two variables in our sample is simply due to sampling error and not an association found in the population.

Answer: Chi Square <u>compares</u> the observed relationship, found in the sample, to a
<u>"table of no relationship."</u>

That is, it creates a table displaying the 2 variables and calculates the numbers that would be in each cell of the table if there were no relationship and then compares this table to the table of actual data <u>found from the sample.</u>

If the values in the 2 tables are similar, then there is a high probability that, whatever relationship is seen in the sample, is due to sampling error and not due to a real relationship in the population.

More Specifically:
How does Chi Square Work?

<u>First</u>, it creates a table of "no relationship". This is done by, first, creating a table that shows the two variables and their margin totals found from the sample (no numbers are placed in the cells of the table yet).

<u>Next</u>, it uses the marginal totals to determine what numbers will go in each cell of the table.

<u>Once this table is created</u>, it compares this "table of no relationship"
to the
table displaying the actual data found from the sample

The more similar the
table of no relationship is to the
actual-data table the more likely that there is NO association between the two variables in the population.

That is, whatever relationship is found in the sample, and subsequently shown in the actual-data table, is likely due to sampling error and not a true reflection of the population.

Table 11.1 below provides sample frequencies (taken from your book). What are the dependent and independent variables? What are the marginals?

In order to calculate a Chi Square for these data, what would you guess is the next step that should be taken?

Table 11.1 Percentage of Men and Women Who Are First-Generation College Students

First-Generation	*Men*	*Women*	*Total*
Firsts	35.4% (691)	46.6% (1,245)	41.9% (1,936)
Nonfirsts	64.6% (1,259)	53.4% (1,425)	58.1% (2,684)
Total (*N*)	100.0% (1,950)	100.0% (2,670)	100.0% (4,620)

Source: Adapted from W. Elliot Inman and Larry Mayes, "The Importance of Being First: Unique Characteristics of First

The next step is to calculate a table of "no association". That is, if this sample had been drawn and there was "no association" between the variables what would the table look like?

Table 11.1 provides actual frequencies from a sample.

Table 11.1 Percentage of Men and Women Who Are First-Generation College Students

First-Generation	*Men*	*Women*	*Total*
Firsts	35.4% (691)	46.6% (1,245)	41.9% (1,936)
Nonfirsts	64.6% (1,259)	53.4% (1,425)	58.1% (2,684)
Total (*N*)	100.0% (1,950)	100.0% (2,670)	100.0% (4,620)

Source: Adapted from W. Elliot Inman and Larry Mayes, "The Importance of Being First: Unique Characteristics of First Generation Community College Students," *Community College Review* 26, no. 3 (1999): 8. Reprinted with permission.

Table 11.3 provides the table of "no association". That is, what one would expect to find if these two variables were not associated.

Table 11.3 Expected Frequencies of Men and Women and First-Generation College Status

First-Generation	*Men*	*Women*	*Total*
Firsts	817.14	1,118.86	1,936
Nonfirsts	1,132.86	1,551.14	2,684
Total (*N*)	1,950	2,670	4,620

We need to calculate f_e = Expected Frequency if No Association.

That is, the cell frequencies that would be expected in a bivariate table if the two variables were unrelated (statistically independent)

For each cell in the table:

$$f_e = \frac{\text{(column marginal) (row marginal)}}{\text{Total N}}$$

Table 11.3 Expected Frequencies of Men and Women and First-Generation College Status

First-Generation	*Men*	*Women*	*Total*
Firsts	817.14	1,118.86	1,936
Nonfirsts	1,132.86	1,551.14	2,684
Total (*N*)	1,950	2,670	4,620

Once we have created our second table of "no association", how do we calculate Chi-Square?

$$X_2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Where:

fo = observed frequencies
fe = expected frequencies if no association

Chi Square compares the observed and the expected frequencies and from this comparison provides the probability that the null hypothesis of no association should be accepted. Typically, alpha is set at .05. If there is a 5% or less probability that the null hypothesis is true, then we reject the null hypothesis.

Table 11.5 calculates chi square for our example using the chi square formula.

Table 11.5 Calculating Chi-Square

First-Generation College Status and Gender	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Men firsts	691	817.14	–126.14	15911.2996	19.47
Men nonfirsts	1,259	1132.86	126.14	15911.2996	14.04
Women firsts	1,245	1118.86	126.14	15911.2996	14.22
Women nonfirsts	1,425	1551.14	–126.14	15911.2996	10.26

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 57.99$$

How do we interpret the Chi Square Statistic?

That is, in our example what does the number, 57.99 mean?

Answer: we use a Chi Square distribution table to locate 57.99 and it's associated probability (Appendix D in text). Or, we observe computer results such as from SPSS.

Appendix D
Distribution of Chi-Square

df	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.03157	.03628	.00393	.0158	.0642	.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
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

Is Chi Square significant when it equals 57.99 with one degrees of freedom (DF)?

(see Chi Square distribution table)

An examination of the Chi Square Distribution table, with a df of 1, shows us that:

the probability of obtaining a X^2 of 57.99, when the null hypothsis is true, is less than .001.

That is, if there is "no association" between the variables, then the chances of drawing a sample with the degree of association found in our sample is less than 1 in 10,000 samples. Therefore, we will reject the null-hypothesis of no difference and assume that the difference found in the sample is a difference existing in the whole population (we accept the research hypothesis).

Limitations of Chi Square:

1. Chi Square is sensitive to sample size. The larger the sample size the larger the chi square. Consequently, the null hypothesis is more likely be rejected with a large sample.

2. Chi Square is sensitive to small expected frequencies. Each cell should include at least 5 cases to be sure that chi square is accurate.

3. While Chi Square shows us statistical significance it does not give us information about the strength of the relationship or substantive significance. (This is left for measures of association and interpretation of the data)

So, given the limitations, it is useful to revisit our earlier question:

Does it make sense to report (or even examine) the measures of association if the test of statistical significance tells us that we should not reject the hypothesis of "no association"?

Answer: typically, if the chi square is not significant, then the measure of association (e.g., lambda) should not be considered since we must accept the null hypothesis of no difference.

However, because chi square is affected by the number of cases in the sample, if the sample is small, chi square is more likely to suggest no relationship between two variables (even if one exists).

Therefore, if one has a small sample it would be wise to examine the size of the measure of association (e.g., lambda) even if chi square is not significant.

冞訝

(see you later)

Table 11.1 provides actual frequencies from a sample.

Table 11.1 Percentage of Men and Women Who Are First-Generation College Students

First-Generation	*Men*	*Women*	*Total*
Firsts	35.4% (691)	46.6% (1,245)	41.9% (1,936)
Nonfirsts	64.6% (1,259)	53.4% (1,425)	58.1% (2,684)
Total (*N*)	100.0% (1,950)	100.0% (2,670)	100.0% (4,620)

Source: Adapted from W. Elliot Inman and Larry Mayes, "The Importance of Being First: Unique Characteristics of First Generation Community College Students," *Community College Review* 26, no. 3 (1999): 8. Reprinted with permission.

Table 11.3 provides the table of "no association". That is, what one would expect to find if these two variables were not associated.

Table 11.3 Expected Frequencies of Men and Women and First-Generation College Status

First-Generation	*Men*	*Women*	*Total*
Firsts	817.14	1,118.86	1,936
Nonfirsts	1,132.86	1,551.14	2,684
Total (*N*)	1,950	2,670	4,620

To read the table we need to know the degrees of freedom.

With cross-tabulation data we find the degrees of freedom by using the following formula:

$$df = (r - 1)(c - 1)$$

Where:

r = the number of rows
c = the number of columns

What are the degrees of freedom in our bivariate (2 X 2) table?