The Chi-Square Test and Measures of Association

Chi-Square as a Statistical Test

- ***Chi-square test:*** an inferential statistics technique designed to test for significant relationships between two variables organized in a bivariate table.
- Chi-square requires **no assumptions** about the shape of the population distribution from which a sample is drawn.
- It can be applied to **nominally** or **ordinally** measured variables.

Gender and Fear of Walking Alone at Night

		SEX		
Are You Afraid To Walk Alone		Male	Female	Total
	No	21(87%)	14(40%)	35(59%)
	Yes	3(13%)	21(60%)	24(41%)
		24(100%)	35(100%)	59(100%)

Gender and Fear are associated but can this relationship be generalized to the population?

Statistical Independence

- ***Independence (statistical):*** the absence of association between two cross-tabulated variables. The percentage distributions of the dependent variable within each category of the independent variable are identical.

Gender and Fear of Walking Alone at Night

		SEX		
Are You Afraid To Walk Alone		Male	Female	Total
	No	14(59%)	21(59%)	35(59%)
	Yes	10(41%)	14(41%)	24(41%)
		24(100%)	35(100%)	59(100%)

Hypothesis Testing with Chi-Square

Chi-square follows five steps:

1. Making assumptions
2. Stating the research and null hypothesis and selecting alpha
3. Selecting the sampling distribution and specifying the test statistic
4. Computing the test statistic
5. Making a decision and interpreting the results

The Assumptions

- The chi-square test requires **no assumptions** about the shape of the population distribution from which the sample was drawn.
- However, like all inferential techniques it assumes **random sampling**.
- It can be applied to variables measured at a **nominal** and/or **ordinal** level of measurement.

Stating Research and Null Hypotheses

- The **research hypothesis** (H_1) proposes that the two variables are related in the population.
- The **null hypothesis** (H_0) states that no association exists between the two cross-tabulated variables in the population, and therefore the variables are statistically independent.

H_1: The two variables are related in the population.

Gender and fear of walking alone at night are *statistically dependent*.

H_0: There is no association between the two variables.

Gender and fear of walking alone at night are *statistically independent*.

The Concept of Expected Frequencies

Expected frequencies f_e: the cell frequencies that would be expected in a bivariate table if the two tables were statistically independent.

Observed frequencies f_o: the cell frequencies actually observed in a bivariate table.

Calculating Expected Frequencies

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

To obtain the expected frequencies for any cell in any cross-tabulation in which the two variables are assumed independent, multiply the row and column totals for that cell and divide the product by the total number of cases in the table.

Gender and Fear -Observed and Expected Frequencies

		SEX		
Are You Afraid To Walk Alone		Male	Female	Total
	No	21(?)	14(?)	35
	Yes	3 (?)	21(?)	24
	Total	24(100%)	35(100%)	59(100%)

((24)*(35))/59=14

Calculating the Obtained Chi-Square

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

f_o = observed frequencies

f_e = expected frequencies

Calculating the Obtained Chi-Square

	Fo	Fe	Fo-fe	$(Fo\text{-}fe)^2$	$(fo\text{-}fe)^2/fe$
Male/No	21	14			
Male/Yes	3	10			
Female/No	14	21			
Female/Yes	21	14			

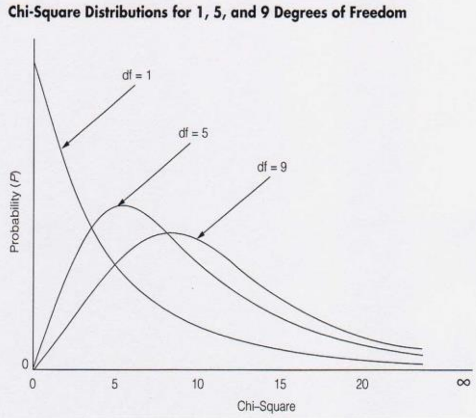
The Sampling Distribution of Chi-Square

- The sampling distribution of chi-square tells the **probability** of getting values of chi-square, assuming no relationship exists in the population.
- The chi-square sampling distributions depend on the degrees of freedom.
- The χ^2 sampling distribution is not one distribution, but is a family of distributions.

The Sampling Distribution of Chi-Square

- The distributions are **positively skewed**. The research hypothesis for the chi-square is always a one-tailed test.
- Chi-square values are always positive. The minimum possible value is zero, with no upper limit to its maximum value.
- As the number of degrees of freedom increases, the χ^2 distribution becomes more symmetrical.

Figure 14.1 **Chi-Square Distributions for 1, 5, and 9 Degrees of Freedom**



Determining the Degrees of Freedom

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

where

r = the number of rows

c = the number of columns

Calculating Degrees of Freedom

Are you Afraid to Walk Alone		SEX		
		Male	Female	Total
	No	14(59%)	21(59%)	35(59%)
	Yes	10(41%)	14(41%)	24(41%)
		24(100%)	35(100%)	59(100%)

The Sampling Distribution of Chi-square-Appendix D

df	.05	.01	.001
1	3.841	6.635	10.827
2	5.991	9.210	13.815
3	7.815	11.341	16.268

Limitations of the Chi-Square Test

- The chi-square test does not give us much information about the ***strength*** of the relationship or its ***substantive significance*** in the population.
- The chi-square test is sensitive to ***sample size***. The size of the calculated chi-square is directly proportional to the size of the sample, independent of the strength of the relationship between the variables.

Measures of Association

- The measure enables us to use a single summarizing measure or number for analyzing the relationship between two variables.
- The measure tell us the strength of the relationship and at times its direction (positive or negative).

Types of Measure of Association

- In this chapter we will talk about measures for nominal and ordinal variables only.
- Lambda and Cramer's V(measure of association for nominal variables).
- Gamma (measure of association for ordinal variables).

Measure of Association

- The measure of association is a single summarizing number that reflects the strength of the relationship, indicates the usefulness of predicting the dependent variable from the independent variable, and often shows the direction of the relationship.
- What does this mean?

Proportional Reduction of Error (PRE)

- **PRE**—the concept that underlies the definition and interpretation of several measures of association. PRE measures are derived by comparing the errors made in predicting the dependent variable while ignoring the independent variable, with errors made when making predictions that use information about the independent variable.

Proportional Reduction of Error

- All measures of association are based on the concept of proportional reduction of error (PRE).
- According to PRE, two variable are associated when information about one can help us improve our prediction of the other.

Proportional Reduction of Error

- The formula for PRE is:

$$PRE = \frac{E1-E2}{E1}$$

  - E1 = errors of prediction made when the independent variable is ignored.
  - E2 = errors of prediction made when the prediction is based on the independent variable.

Support for Abortion by Number of Children

	Number of Children		
Support Abortion for Any Reason	None or 1 child	2 or more children	Total
Yes	36	15	51
No	23	31	54
Total	59	46	105

PRE Measures of Association

- The PRE can range from 0.0 to $^{+}_{-}$1.0.
- A PRE of 0.0 indicates that the two variables are not associated; information about the independent variable will not improve prediction about the dependent variable.
- A PRE of $^{+}_{-}$ 1.0 indicates a perfect positive or negative association between the variables.

What is Strong? What is Weak?

-1.00	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	1.00
Perfect negative Relationship	Very Strong negative relationship	Strong negative relationship	Moderate negative relationship	Weak negative relationship	No relationship at all	Weak positive relationship	Moderate positive relationship	Strong positive relationship	Very Strong positive relationship	Perfect positive Relationship

Two PRE Measures: Lambda, Cramer's V & Gamma

		Appropriate for…
• *Lambda*	λ	NOMINAL variables
• *Cramer's V*		Chi-Square related measure
• *Gamma*	γ	ORDINAL & DICHOTOMOUS NOMINAL variables

Lambda: Measure of Association for Nominal Variables

- An asymmetrical measure of association, lambda is suitable for use with nominal variables and may range form 0.0 to 1.0.
- It provides us with an indication of the strength of an association between the independent and dependent variables.
- Lambda is a PRE and follows the basic formula.

Lambda λ

$$\text{Lambda} = \frac{\text{E1} - \text{E2}}{\text{E1}}$$

where:

$$\text{E1} = N_{\text{total}} - N_{\text{mode of dependent variable}}$$

$$\text{E2} = \sum_{\substack{\text{for all}\\ \text{categories}}} (N_{\text{category}} - N_{\text{mode for category}})$$

Financial Satisfaction by Home Ownership, Women Only

Financial Satisfaction	Home Ownership		
	Own	Rent	Row Total
Satisfied	104	18	122
More or Less	85	46	131
Not Satisfied	50	18	68
Column Total	239	82	321

Guidelines for Calculating Lambda

- When calculating Lambda the numbers will change when you change an independent variable to the dependent and vice versa.
- Make sure you know which variable is the independent and which variable is the dependent so that you have the correct calculation for lambda.

Cramer's V

- Based on the value of chi-square and ranges between 0 to 1
  - 0 indicating no association and 1 indicating perfect association
- Because it cannot take negative values, it is considered a nondirectional measure

Cramer's *V*

$$Cramer's\ V = \sqrt{\frac{\chi^2}{N \times m}}$$

where m = smaller of $(r \pm 1)$ or $(c \pm 1)$.

Symmetrical Measure of Association

- A measure whose value will be the same when either variable is considered the independent variable or the dependent variable.
- **Gamma** is a *symmetrical* measure of association…

Finding Gamma

- The purpose of gamma is to determine whether there is an association between the independent and dependent variables.
- We want to determine whether the association is positive or negative.
- Gamma can help us answer that question.

Symmetrical Measures of Association

- Gamma
  - suitable for use with ordinal variables or with dichotomous nominal variables
  - It can vary from 0.0 to ±1.0 and provides us with an indication of the strength and direction of the association between the variables

Measures of Association

- **Measures of association**—a single summarizing number that reflects the strength of the relationship. This statistic shows the **magnitude** and/or **direction** of a relationship between variables.
- **Magnitude**—the closer to the absolute value of 1 the stronger the association. If the measure equals 0, there is no relationship between the two variables.
- **Direction**—the sign on the measure indicates if the relationship is positive or negative. In a positive relationship, when one variable is high, so is the other. In a negative relationship, when one variable is high, the other is low.