Lecture Notes

# Chapter 10: The Chi-Square Test and Measures of Association

## Learning Objectives

10.1 Summarize the application of a chi-square test.

10.2 Calculate and interpret a test for the bivariate relationship between nominal or ordinal variables.

10.3 Determine the significance of a chi-square test statistic.

10.4 Explain the concept of proportional reduction of error.

10.5 Apply and interpret measures of association: lambda, Cramer’s *V*, gamma, and Kendall’s tau-*b*.

10.6 Interpret output for chi-square and measures of association.

## Chapter Outline

1. The Concept of Chi-square as a Statistical Test
	1. **Chi-square test** (pronounced kai-square and written as χ2) is an inferential statistical technique designed to test for significant relationships between two variables organized in a bivariate table.
	2. Chi-square test can also be applied to distribution of scores for a single variable.
	3. It is referred to as the goodness-of-fit test.
2. The Concept of Statistical Independence
	1. When two variables are not associated, one can say that they are **statistically independent**.
	2. Identify statistical independence in a bivariate table by comparing the distribution of the dependent variable in each category of the independent variable.
	3. When two variables are statistically independent, the percentage distributions of the dependent variable within each category of the independent variable are identical.
3. The Structure of Hypothesis Testing With Chi-square
	1. The Assumptions
		1. 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.
	2. Stating the Research and the Null Hypotheses
		1. Like all other tests of statistical significance, the chi-square is a test of the null hypothesis.
	3. The Concept of Expected Frequencies
		1. Assuming that the null hypothesis is true, we compute the cell frequencies that we would expect to find if the variables are statistically independent. These frequencies are called **expected frequencies** (and are symbolized as *f*e).
		2. The chi-square test is based on cell-by-cell comparisons between the expected frequencies (*f*e) and the frequencies actually observed (**observed frequencies** are symbolized as *fo*).
	4. Calculating the Expected Frequencies
		1. The difference between *f*o and *f*e will determine the likelihood that the null hypothesis is true and that the variables are, in fact, statistically independent.
		2. We can calculate the expected frequencies using this formula: 
	5. Calculating the Obtained Chi-Square
		1. The next step in calculating chi-square is to compare the differences between the expected and observed frequencies.
		2. **Obtained chi-square statistic**: 
			1. $f\_{o}$ = observed frequencies
			2. $f\_{e}$ = observed frequencies
	6. The Sampling Distribution of Chi-Square
		1. The sampling distribution of chi-square tells the probability of getting values of chi-square, assuming no relationship exists in the population.
		2. Chi-square distributions depend on degrees of freedom; it is a family of distributions.
		3. Chi-square values are always positive; the minimum possible value is zero, with no upper limit to its maximum value.
			1. A chi-square of zero means the variables are completely independent and the observed frequencies in every cell are equal to the corresponding expected frequencies.
	7. Determining the Degrees of Freedom
		1. With cross-tabulation data, we find the degrees of freedom by using the formula, *df* = (*r* – 1)(*c* – 1), where *r* = the number of rows, *c* = the number of columns.
		2. Yates’s correction for continuity: 
	8. Making a Final Decision
		1. With the Yates’s correction, the corrected chi-square is 57.54. Refer to Table 10.6 for calculations.
		2. We can establish that the probability of obtaining a chi-square of 57.54 is less than .001 if the null hypothesis were true.
4. Proportional Reduction of Error
	1. **Measures of association** enable one to use a single summarizing measure or number for analyzing the pattern of relationship between two variables.
	2. Four measures of association:
		1. Lambda (measures of association for nominal variables).
		2. Gamma.
		3. Kendall’s tau-*b* (measures of association between ordinal variables).
		4. Cramer’s *V* (a chi-square related measure of association).
	3. **Proportional reduction of error** often abbreviated as **PRE**.
		1. According to the concept of PRE, two variables are associated when information about one variable (an independent variable) can help us improve our prediction of the other variable (a dependent variable).
		2. The conceptual formula for allPRE measures of association is,  Where, *E*1 = errors of prediction made when the independent variable is ignored (Prediction 1) and *E*2 = errors of prediction made when the prediction is based on the independent variable (Prediction 2).
	4. PRE measures of association can range from 0.0 to ± 1.0.
5. Lambda: A Measure of Association for Nominal Variables
	1. **Lambda** is an **asymmetrical measure** used to determine the strength of the relationship between two nominal variables.
	2. An **asymmetrical measure** will vary depending on which variable is considered the independent variable and which the dependent variable.
	3. Lambda may range in value from 0.0 to 1.0; 0 indicates there is nothing to be gained by using the independent variable to predict the dependent variable, and 1.0 indicates that by using the independent variable as a predictor we are able to predict the dependent variable without any error.
		1. Lambda is always zero in situations in which the mode for each category of the independent variable falls into the same category of the dependent variable.
	4. A problem with interpreting lambda arises in situations in which lambda is zero, but other measures of association indicate that the variables are associated.
		1. To avoid this potential problem, examine the percentage differences in the table whenever lambda is exactly equal to zero.
		2. If the percentage differences are very small (usually 5% or less), lambda is an appropriate measure of association for the table.
		3. However, if the percentage differences are larger, indicating that the two variables may be associated, lambda will be a poor choice as a measure of association.
		4. In such cases, we may want to discuss the association in terms of the percentage differences or select an alternative measure of association.
6. Cramer’s *V*: A Chi-square–related Measure of Association for Nominal Variables
	1. **Cramer’s *V*** is an alternative measure of association that can be used for nominal variables; it’s based on the value of chi-square and ranges between 0 and 1.
		1. 0 indicates no association and 1 indicates perfect association.
	2. It is calculated using the formula:  Where *m* = smaller of (*r* – 1) or (*c* – 1)$.$
7. Gamma and Kendall’s tau-*b*: Symmetrical Measures of Association for Ordinal Variables
	1. **Gamma** and **Kendall’s tau-*b*** are **symmetrical measures of association** suitable for use with ordinal variables or with dichotomous nominal variables.
	2. Both gamma and Kendall’s tau-*b* can vary from 0.0 to ±1.0 and provide us with an indication of the strength and direction of the association between the variables.
	3. Gamma and Kendall’s tau-*b* can be positive or negative.
		1. A gamma or Kendall’s tau-*b* of 1.0 indicates that the relationship between the variables is positive and that the dependent variable can be predicted without any errors based on the independent variable.
		2. A gamma of –1.0 indicates a perfect, negative association between the variables.