**Section 11.2 (Part 2) - Inferences for Relationships**

**Comparing Several Proportions** - In chapter 10, we used the two-sample z-test to compare proportions from two samples from two populations. What if you want to compare proportions from more than two samples? The chi-square test for homogeneity can be used to do this.

|  |  |
| --- | --- |
| The example on p. 710 illustrates this.Are the differences between the groups statistically significant at the 1% level? | tableun_11_38.jpg |



**Relationships between Two Categorical Variables - Chi-Square Test for Association/Independence**

Another common situation that leads to a two-way table is when a *single* random sample of individuals is chosen from a *single* population and then classified according to two categorical variables. The goal is to analyze the relationship between the variables. This leads to the chi-square test for association/independence.

**Chi-Square Test for Association/Independence**

Suppose the Random, Large Sample Size, and Independent conditions are met. You can use the chi-square test for association/independence to test

*H0: There is no association between two categorical variables in the population of interest.
Ha: There is an association between two categorical variables in the population of interest.*

Or, alternatively

*H0: Two categorical variables are independent in the population of interest.
Ha: Two categorical variables are not independent in the population of interest.*

Start by finding the expected counts. Then calculate the chi-square statistic

$$χ^{2}=\sum\_{}^{}\frac{\left(Observed-Expected\right)^{2}}{Expected}$$

where the sum is over all cells (not including totals) in the two-way table. If *H0* is true, the chi-square statistic has approximately a chi-square distribution with degrees of freedom =
(number of rows - 1)(number of columns - 1). The *P*-value is the area to the right of the chi-square statistic under the corresponding chi-square density curve.

*Conditions*: Use this test when:

* **Random** - The data come from separate random samples from each population of interest or from the groups in a randomized experiment.
* **Large Sample Size** - All expected counts are at least 5.
* **Independent** - Both the samples or groups themselves and the individual observations in each sample or group are independent. When sampling without replacement, check that the individual populations are at least 10 times as large as the corresponding samples (10% condition).

**Example** - A study followed a random sample of 8474 people with normal blood pressure for about four years. All the individuals were free of heart disease at the beginning of the study. Each person took and anger scale test which measures how prone a person is to sudden anger. Researchers also recorded whether each individual developed coronary heart disease (CHD). This includes people who had heart attacks and those who needed medical treatment for heart disease.



Do the data provide convincing evidence of an association between anger level and heart disease in the population of interest?

**Using Chi-Square Tests Wisely**

Both the chi-square test for homogeneity and the chi-square test for association/independence start with a two-way table of observed counts. They even calculate the test statistic, degrees of freedom, and *P*-value in the same way. However, the questions that these two tests answer are different.

* A chi-square test for homogeneity tests whether the distribution of a categorical variable is the same for each of several populations or treatments.
* A chi-square test for association/independence tests whether two categorical variables are associated in some population of interest.

*Trick*: Look at how the data were produced to decide which test is appropriate.

* If the data come from two or more independent random samples or treatment groups in a randomized experiment, then do a test for *homogeneity*.
* If the data come from a single random sample, with the individuals classified according to two categorical variables, use a test for *association/independence*.

HW: Read Sec 11.2; problems 41, 43, 45, 47, 51-56