**Chapter 11.2 – Inference for Relationships between Categorical Variables**

*Chi-Square Test for Homogeneity*

Does a treatment have an effect on the distribution of a categorical variable?

*H0:* There is no difference in the distribution of a categorical variable for several populations or treatments.

*Ha:* There is a difference in the distribution of a categorical variable for several populations or treatments.

*Conditions:*

* Random – data must come from a random sample
* Large Sample Size – all EXPECTED cell counts must be at least 5
* Independent – check the 10% condition $10n\leq N$

Start by finding the expected counts. Then calculate the chi-square statistic where the sum is over all cells (not including totals) in the two-way table. If *H0* is true, the *χ2* 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 *χ*2 under the corresponding chi-square density curve.

**Example:** Market researchers suspect that background music may affect the mood and buying behavior of customers. One study in a supermarket compared three randomly assigned treatments: no music, French accordion music, and Italian string music. Under each condition, the researchers recorded the numbers of bottles of French, Italian, and other wine purchased. Does the type of music played affect the type of wine that’s purchased? **State:**

*H0:*

 *Ha:*

**Plan:**



df:

p-value:

Conditions: **Do:**

|  |
| --- |
| Expected counts |
| **Music** |
| **Wine** | None | French | Italian | **Total** |
| French |  |  |  |  |
| Italian |  |  |  |  |
| Other |  |  |  |  |
| **Total** |  |  |  |  |

**Conclude:**

*Chi-Square Test for Association/Independence*

Are two categorical variables related, or they independent of one another?

*H0:* There is no association between two categorical variables in the population of interest (independent)

*Ha:* There is an association between two categorical variables in the population of interest (not independent)

*Conditions:*

* Random – data must come from a random sample
* Large Sample Size – all EXPECTED cell counts must be at least 5
* Independent – check the 10% condition $10n\leq N$

Start by finding the expected counts. Then calculate the chi-square statistic where the sum is over all cells (not including totals) in the two-way table. If *H0* is true, the *χ2* 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 *χ*2 under the corresponding chi-square density curve.

**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 the Spielberger Trait 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 any people have more heart disease? **State:**

*H0:*

 *Ha:*

**Plan:**



df:

p-value:

Conditions: **Do:**

|  |
| --- |
| Expected counts |
|  | Low | Moderate | High | **Total** |
| CHD |  |  |  |  |
| No CHD |  |  |  |  |
| **Total** |  |  |  |  |

**Conclude:**