**Chapter 14.2**

1. **Goodness of Fit**

A goodness of fit test is used to help determine whether a population has a certain hypothesized distribution, expressed as proportions of individuals in the population falling into various outcome categories. A GOF Test is used when investigating multiple categories of **ONE** categorical variable.

1. Equal Proportions (all proportions are expected to be the same)
2. Fixed or Given Proportions (proportions are expected to follow given values)
3. **Chi-Square Test of Independence/Association**
	1. Data: Collected a single sample from the population and then classify based on **two** categorical variables.
	2. Question: Are the two variables independent?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Low Anger** | **Moderate Anger** | **High Anger** | **Total** |
| **CHD** | 53 | 110 | 27 | 190 |
| **No CHD** | 3057 | 4621 | 606 | 8284 |
| **Total** | 3110 | 4731 | 633 | 8474 |

**EX:** 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). The two-way table below summarizes the data.

Do these data provide convincing evidence of an association between anger level and heart disease?

**State:**

Ho:

HA:

**Plan:**

1)

2)

Expected counts: $\frac{row total x column total}{total}$

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Low Anger** | **Moderate Anger** | **High Anger** | **Total** |
| **CHD** |  |  |  | 190 |
| **No CHD** |  |  |  | 8284 |
| **Total** | 3110 | 4731 | 633 | 8474 |

**Calculate**:

χ2 = $Σ$ $\frac{(O-E)^{2}}{E}$=

**Interpret:**

1. **Chi-Square Test of Homogeneity**
	1. Data: take **two separate samples** from **two separate populations** and compare the same variable for each
	2. Question: Is that variable distributed the same way in both populations?

**EX:** We took a random sample of peanut and regular M&M’s and assessed the # of each color. The data is shown below. Is the distribution of colors in peanut M&M’s different than the distribution of colors in regular M&M’s?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Brown | Yellow | Red | Orange | Green | Blue |  |
| Peanut | 14 | 30 | 8 | 29 | 24 | 25 | 130 |
| Regular | 34 | 64 | 27 | 123 | 77 | 91 | 416 |
|  | 48 | 94 | 35 | 152 | 101 | 116 | 546 |

**State:**

Ho:

HA:

**Plan:**

Calculate the Expected Counts and verify that they are all at least 5

Expected counts: $\frac{row total x column total}{total}$

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Brown | Yellow | Red | Orange | Green |
| Peanut |  |  |  |  |  |
| Regular |  |  |  |  |  |

**Calculate**:

χ2 = $Σ$ $\frac{(O-E)^{2}}{E}$=

**Interpret:**

|  |  |  |
| --- | --- | --- |
| Goodness of Fit | Independence | Homogeneity |
| We want to know how closely our sample proportions fit the suggested sample proportions.(One variable) | Two variables measure on **one population**. Are the two variables somehow associated? | One variable measured independently on two (or more) populations. Are the proportions for the populations the same? |
| * Is the sample consistent with the stated proportions?
* Are the proportions the same?
 | * Are the variables independent?
* Is there an association between the two variables?
* Is \_\_ dependent on \_\_\_?
 | * Are the groups homogenous (the same)
* Does there appear to be a difference between the two groups?
 |