AP Statistics Chapter 11: Inference for Distributions of Categorical Data

11.1 – Chi-Square (χ^2) Goodness of Fit Test

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. There are two types of goodness of fit tests

1. Equal Proportions (all proportions are expected to be the same)
2. Fixed or Given Proportions (proportions are expected to follow given values)

Hypotheses for the Goodness of Fit Test

H_o: The stated distribution of the categorical variables in the population of interest is correct.
H_a: The stated distribution in the population of interest is **not** correct.

The Chi-Square Statistic

The formula is

$$\chi^2 = \sum \frac{(observed\ count - expected\ count)^2}{expected\ count} = \sum \frac{(O-E)^2}{E}$$

Expected Counts

- The expected counts for the equal proportions GOF test are all the same. They are found by dividing the total of the counts by the number of outcome categories.
- The expected counts for the given proportions GOF test are NOT all the same. They are found by multiplying the total of the counts by each given percentage.

Conditions for the Chi-Square Goodness of Fit Test

- **Random** – The data come from a well-designed random sample or randomized experiment.
- **Independent** – is the sample size less than 10% of the population size?
- **Large Counts** – All expected counts are at least 5.

Degrees of Freedom for the Goodness of Fit Test

The degrees of freedom for the GOF test are $n-1$ where *n* is the number of outcome categories.

Using the Calculator for the χ^2 Test on the TI-83 Calculator*

- The *observed* counts are to be stored in L_1
- The *expected* counts are to be stored in L_2
- Let $L_3 = (L_1-L_2)^2/L_2$
- The χ^2 is the sum of L_3 (which can be found by using 1-Var Stats)

***The TI-84 has a built-in test called χ^2GOF-Test. Find it in the STAT-TEST menu.**

11.2 – The χ^2 Test of Association/Independence for Two-Way Tables

Two-Way Tables

When there are two categorical variables, data can be arranged in a row and column format, called a Two-Way Table. Here is an example:

	Color Choice				
Grade	**blue**	**green**	**red**	**yellow**	**Totals**
1st	13	7	8	2	30
2nd	11	10	6	5	33
Totals	24	17	14	7	63

Hypotheses for the Test for Association between Two Categorical Variables

Ho: There is **no** association between the categorical variables
Ha: There is an association

OR

Ho: The categorical variables are independent (there is **no** association)
Ha: The variables are NOT independent (there is an association)

Degrees of Freedom for the Chi-Square Test of Association/Independence

The degrees of freedom for the Test of Association/Independence test are $(r-1)x(c-1)$ where *r* is the number of rows and *c* is the number of columns in the table.

Conditions for the Chi-Square Test of Association/Independence

The conditions are the same as for the goodness of fit test

The Chi-Square Statistic

The formula is the same as in the goodness of fit test.

Expected Counts

The expected counts for this test can be found as follows:

$$expected\ count = \frac{(row\ total)(column\ total)}{table\ total}$$

For example, for the expected count for 2nd grade/green in the table above, we would use the calculation

$$\frac{(33)(17)}{63} = 8.9$$