

AP Statistics Chapter 14: The Chi-Square (χ^2) Procedures

14.1 – Chi-Square 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

Ho: The actual population proportions are equal to the hypothesized proportions

Ha: The actual population proportions are different from the hypothesized proportions

The Chi-Square Statistic

The formula is

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

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 Test

- We have an **SRS**, as always, from the population of interest.
- None of the **expected** counts are 0
- No more than 20% of the expected counts are less than 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.

14.2 – Test of Association 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 (we first saw these in Chapter 4). 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

Test for Association between Two Categorical Variables

Ho: There is no association between the variables

Ha: There is an association

OR

Ho: The variables are independent (there is no association)

Ha: The variables are NOT independent (there is an association)

Degrees of Freedom for the Goodness of Fit Test

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

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:

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

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

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