11.1 - The Chi-Square Distribution

The chi-square distribution has only one parameter called the degrees of freedom.

The entire chi-square distribution curve lies to the right of the vertical axis. The chi-square distribution assumes nonnegative values only, and these are denoted by the symbol $\chi^2$ (read as "chi-square").

The shape of a chi-squared distribution curve is skewed to the right for small $df$ and becomes more symmetric for larger $df$. The peak of the curve occurs at $df-2$.

In this class we will use our TI83/84 calculators to find $\chi^2$ values corresponding to an area and the other way around.

To find $\chi^2$-value if given an area: CHINVR83, found under PRGM.
To find right tail area: CHI83, found under PRGM. Or if you don’t have this program you can use $\chi^2$ cdf (lowerbound, upperbound, df), found under DISTR (2ND VARS).

ex. Find the value of $\chi^2$ for 7 degrees of freedom and an area of 0.10 in the right tail of the chi-square distribution curve.

ex. Now work the previous problem backwards.
11.2 - A Goodness-of-Fit Test

In a goodness-of-fit test we test whether the observed frequencies for an experiment follow a certain pattern. That is, we test how good the observed frequencies fit a given pattern.

The frequencies obtained from the performance of an experiment are called the observed frequencies and are denoted by \( O \).

The expected frequency for a category, denoted by \( E \), are the frequencies that we would expect to obtain if the null hypotheses is true.

\[ E = np \], where \( n \) is the sample size and \( p \) is the probability that an element belongs to that category given that the null hypotheses is true.

Requirements:
1. The data have been randomly selected.
2. The expected frequency is at least 5 for each category.

The test statistics for a goodness-of-fit test is \( \chi^2 = \sum \left( \frac{O - E}{E} \right)^2 \)

\( H_0 : p_1 = \ldots , p_2 = \ldots , p_3 = \ldots , \ldots \), etc.

\( H_1 : \) at least one of the categories does not follow the distribution specified in \( H_0 \).

The goodness-of-fit tests are always right tailed.

The p-value is found using calculator program CHI83 or \( \chi^2 \text{cdf} \), with \( k-1 \) degrees of freedom, where \( k \) is the number of possible categories.

TI84 has a program for a Goodness of Fit tests called \( \chi^2 \text{GOF-Test} \), which can be found in the TESTS menu (STATS button first). Your observed frequencies have to be entered in one list and you expected frequencies entered in another list prior to running the test on the calculator.

If you don’t have this program, you may have downloaded a program from the MLC called GOODFT83 which works well also. When it asks you to enter number of outcomes you need to enter number of categories.
1. Over the last 3 years, Art’s Supermarket has observed the following distribution of modes of payment in the express lines: cash (C) 41%, check (CK) 24%, credit or debit card (D) 26%, and other (N) 9%. In an effort to make express checkout more efficient, Art’s has just begun offering a 1% discount for cash payment in the express line. The following table lists the frequency distribution of the modes of payment for a sample of 500 express-line customers after the discount went into effect.

<table>
<thead>
<tr>
<th>Mode of Payment</th>
<th>C</th>
<th>CK</th>
<th>D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Customers</td>
<td>240</td>
<td>104</td>
<td>111</td>
<td>45</td>
</tr>
</tbody>
</table>

Test at the 1% significance level whether the distribution of modes of payment in the express checkout line changed after the discount went into effect.

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed Frequency</th>
<th>p</th>
<th>Expected Frequency ( E = np )</th>
<th>( O - E )</th>
<th>( (O - E)^2 )</th>
<th>( \frac{(O - E)^2}{E} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ck</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( n = \) Sum =
2. Do “A” students tend to sit in a particular part of the classroom? A teacher recorded the locations of the students who received grades of A, with these results: 17 sat in the front, 9 sat in the middle, and 5 sat in the back of the classroom. Test at the 5% significance level the assumption that the “A” students are distributed evenly throughout the room.
11.3 – Test for Independence Using Contingency Tables

A contingency table, also called a two-way classification table or two-way frequency table, is a table in which frequencies correspond to two variables. One variable is used to categorize rows and a second variable is used to categorize columns. Contingency tables come in many sizes. An R x C table contains R rows and C columns.

In a Test of Independence we test the null hypothesis that there is no association between the row variable and the column variable in a contingency table. For the null hypothesis we will use the statement that “the row and column variables are independent”, and for the alternative hypothesis we will use the statement that “the row and column variables are dependent”.

The frequencies obtained from the performance of an experiment are called the observed frequencies and are denoted by $O$.

The expected frequency for a cell is denoted by $E$ and is calculated as shown below:

$$E = \frac{(\text{Row Total})(\text{Column Total})}{\text{Sample Size}}$$

Requirements:
1. The data have been randomly selected.
2. The expected frequency is at least 5 for each category.

The test statistics for a test of independence is

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

The p-value is found using the calculator and $df = (R-1)(C-1)$.

A test of independence is always right tailed.

When calculator use is permitted, you can use:

- $\chi^2$ - Test located in the TESTS menu. You will first have to create a matrix with the observed values. To do this, press 2nd, then MATRIX. Highlight EDIT and press ENTER to select 1:[A]. Enter size of the matrix and then enter each data. Use this matrix as the input for observed values. The default for the Expected value field is matrix [B], which will be calculated for you automatically.

- or

- CHITST83 located under PRGM. In this program you have to enter both observed and expected values.
1. In a test of the effectiveness of Echinacea, some test subjects were treated with Echinacea extracted with 20% ethanol, some were treated with Echinacea extracted with 60% ethanol, and others were given a placebo. All of the test subjects were then exposed to rhinovirus (common colds are typically caused by this virus). The results are summarized in the 2 x 3 table below.

Results from Experiment with Echinacea

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Placebo</th>
<th>Echinacea: 20% extract</th>
<th>Echinacea: 60% extract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infected</td>
<td>88</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>Not Infected</td>
<td>15</td>
<td>4</td>
<td>10</td>
</tr>
</tbody>
</table>

Using 5% significance level, test the claim that getting an infection (cold) is independent of the treatment group. What does the result indicate about the effectiveness of Echinacea as a treatment for colds?
2. A researcher wanted to study the relationship between gender and owning smart phones. She took a sample of 2000 adults and obtained the information given in the following table.

<table>
<thead>
<tr>
<th></th>
<th>Own Smart Phones</th>
<th>Do Not Own Smart Phones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>640</td>
<td>450</td>
</tr>
<tr>
<td>Women</td>
<td>440</td>
<td>470</td>
</tr>
</tbody>
</table>

At the 5% level of significance, can you conclude that gender and owning a smart phone are related for all adults?