Week 13 Lecture: Goodness of Fit Tests (Chapter 22)

These statistical methods test nominal scale (aka, categorical or attribute) data to determine if the observed distribution of counts (never percentages or ratios) fits a hypothesized distribution. The most notable analysis technique is the chi-square test.

**Example:** Testing the choice of salmon to select a certain home stream versus four nearby streams.

**Hypotheses:**

Ho: Home stream is chosen 75% of the time; remaining four streams 25% of the time (6.25% each).

(Note: The null hypothesis can be alternatively stated as: Ho: The sample came from a salmon population with a 12:1:1:1:1 ratio of choosing home and alternate streams.)

Ha: not Ho (note: either the data fits or not)

<table>
<thead>
<tr>
<th>N=200 fish</th>
<th>Home Stream</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>12 (0.75)</td>
<td>1 (0.0625)</td>
<td>1 (0.0625)</td>
<td>1 (0.0625)</td>
<td>1 (0.0625)</td>
</tr>
<tr>
<td>Expected Counts ((\hat{f}_i))</td>
<td>150</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Observed Counts ((f_i))</td>
<td>135</td>
<td>15</td>
<td>17</td>
<td>10</td>
<td>23</td>
</tr>
</tbody>
</table>

e.g.: for the home stream >>>> Ratio = 0.75 / 0.0625 = 12; %Ratio = 12 / 16 = 0.75; \(\hat{f}_i = 0.75 \ast 200 = 150\) fish
Test Statistic:  \[ \chi^2 = \sum_{i=1}^{k} \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} \]

\[ \chi^2 = \frac{(135 - 150)^2}{150} + \frac{(15 - 12.5)^2}{12.5} + \frac{(17 - 12.5)^2}{12.5} + \frac{(10 - 12.5)^2}{12.5} + \frac{(23 - 12.5)^2}{12.5} = 12.94 \]

Critical Value: compare to a chi-square distribution:  \[ \chi^2_{\alpha(k-1)} \], where k = the number of classes and \( \alpha \) = alpha level.

\[ \chi^2_{0.05(5-1=4)} = 9.488 \]

Decision Rule: reject Ho if \( \chi^2 \geq 9.488 \); otherwise, do not reject. Since 12.94 > 9.488 (0.01 < P < 0.025 – see Table B.1), we reject Ho and conclude that our observed frequency of returning salmon does not match our hypothesized distribution.

What does this probability mean? It is the probability of seeing a test statistic of at least 12.94 if Ho is true (with repeated experiments).
What’s the significance of the test? >>> The significance level must be specified. At the beginning of this hypothesis test, we specified the level of significance (alpha level or $\alpha$) of the test. The $\alpha$-level is the probability of falsely rejecting the null hypothesis, $H_0$, even though it is true. It is also the probability of committing a Type I error. A common level for $\alpha$ is 0.05; other common values are 0.10 and 0.01. For our example, we used $\alpha = 0.05$, which is a 1/20th chance of getting a bad data set that made us falsely reject our $H_0$.

**Subdividing the Chi-square Analysis**

In our salmon example, the number of fish returning to Stream 4 seems to have caused the $H_0$ rejection; so, we’ll subdivide it. Now, let’s test:

**Hypotheses:**

- $H_0$: The sample came from a salmon population with a 12:1:1:1 ratio of choosing home and alternate streams 1 - 3.
- $H_a$: not $H_0$.

<table>
<thead>
<tr>
<th>N=177 fish</th>
<th>Home Stream</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>12 (0.80)</td>
<td>1 (0.0667)</td>
<td>1 (0.0667)</td>
<td>1 (0.0667)</td>
</tr>
<tr>
<td>Expected Counts ($f_1^*$)</td>
<td>141.6</td>
<td>11.8</td>
<td>11.8</td>
<td>11.8</td>
</tr>
<tr>
<td>Observed Counts ($f_o$)</td>
<td>135</td>
<td>15</td>
<td>17</td>
<td>10</td>
</tr>
</tbody>
</table>

e.g.: for the home stream >>>> Ratio = 0.80 / 0.0667 = 12; %Ratio = 12 / 15 = 0.80; $f_1^* = 0.80 * 177 = 141.6$ fish
Test Statistic: \[
\chi^2 = \frac{(135 - 141.6)^2}{141.6} + \frac{(15 - 11.8)^2}{11.8} + \frac{(17 - 11.8)^2}{11.8} + \frac{(10 - 11.8)^2}{11.8} = 3.74
\]

Critical Value: \[
\chi^2_{0.05(3)} = 7.815
\]

Decision Rule: reject Ho if \(
\chi^2 \geq 7.815
\); otherwise, do not reject. Since \(3.74 < 7.815\) \((0.25 < P < 0.50 – see Table B.1)\), we do not reject Ho and conclude that our observed frequency of returning salmon does match our hypothesized distribution.

Now, test

Ho: The sample came from a salmon population with a 15:1 ratio of choosing the home stream and streams 1 – 3 versus stream 4 (Ho can alternatively be written as: \(H + S1 + S2 + S3 \mid S4\)).

Ha: not Ho.

<table>
<thead>
<tr>
<th>N=200 fish</th>
<th>Other Streams</th>
<th>Stream 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>15 (0.9375)</td>
<td>1 (0.0625)</td>
</tr>
<tr>
<td>Expected Counts ((\hat{f}_i))</td>
<td>187.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Observed Counts ((f_i))</td>
<td>177</td>
<td>23</td>
</tr>
</tbody>
</table>

Test Statistic: \[
\chi^2 = \frac{(177 - 187.5)^2}{187.5} + \frac{(23 - 12.5)^2}{12.5} = 9.41
\]

Critical Value: \[
\chi^2_{0.05(1)} = 3.841
\]
Decision Rule: reject Ho if $\chi^2 \geq 3.841$; otherwise, do not reject. Since $9.41 > 3.841$ ($0.001 < P < 0.005$ – see Table B.1), we reject Ho and conclude that the return rates of salmon to stream 4 is different from the other streams.

Ho was rejected because of Stream 4; the first subdivision was not rejected and this second subdivision was rejected. Note that the chi-square values from different tests are not algebraically equivalent: $12.94 \neq 3.74 + 9.41 = 13.15$ >>> however, they should be close.

**Continuity Correction**

Chi-square values calculated based on actual data belong to a discrete distribution. However, the theoretical chi-square distribution is continuous. Thus, we are not actually testing at the specified alpha-levels found in Table B.1. With an uncorrected chi-square test, you falsely reject Ho too often. This is not a problem except when df = 1, as in our second subdivision above.

When df = 1, the following formula should be used to calculate the chi-square value:

$$\chi^2 = \sum_{i=1}^{2} \left( \frac{|f_i - \hat{f}_i| - 0.5}{\hat{f}_i} \right)^2$$

The continuous distribution is also not a problem as long as “n” is large. However, when is “n” large enough? >>> no $\hat{f}_i < 1.0$ and no more than $\pm 20\%$ of the $\hat{f}_i$’s $< 5.0$ (the latter is a disputed rule).

**Heterogeneity Chi-Square**

You may find yourself in a situation where you want to determine if you can combine datasets to conduct a single chi-square analysis. This type of test is called a “heterogeneity chi-square
You are testing the null hypothesis, Ho: The samples all came from the same population. If you do not reject Ho, then you can combine the datasets and conduct an overall chi-square test. Zar describes this procedure in section 22.6 on page 474. He also provides a good example on page 475 (example 22.5), which uses classical data from Gregor Mendel’s pea breeding experiments.

**Kolmogorov – Smirnov Goodness of Fit Test for Discrete Data**

The KS test should be used for discrete data in ordered categories. The KS test is better for ordered data than the standard chi-square test because it takes into account the ordering of the categories when testing for a uniform distribution. The test for discrete data looks at the cumulative observed \((F_i)\) and expected frequencies \((\hat{F}_i)\), and takes the largest difference:

\[
d_{\text{max}} = \max |d_i| = \max |F_i - \hat{F}_i|
\]

This value is compared to a critical value (from Table B.8 in appendix):

\[
d_{\text{critical}} = d_{\text{max,}\alpha,k,n}
\]

where \(k = \text{number of categories}, n = \text{number of samples}, \alpha = \text{alpha-level}.

**Example:** We want to test if insect species are uniformly distributed along a light gradient.

Light gradients do not have a measurable difference, except that lower values represent darker light conditions than higher numbers.

Ho: Insect species uniformly distributed along a light gradient

Ha: Not Ho

\(\alpha = 0.05\)
### Test Statistic

\[ d_{\text{max}} = \max |d_i| = \max |F_i - \hat{F}_i| = 26 \]

### Critical Value

\[ d_{\text{critical}} = d_{\text{max}, 0.05, 65} = 10 \quad \text{(see Table B.8)} \]

### Decision Rule

Reject Ho if \( d_{\text{max}} \geq 10 \); otherwise, do not reject. Since 26 > 10 (P < 0.001), we reject Ho.

### Conclusion

The observed data do not follow a uniform distribution across the ordered light categories (P < 0.001).

**Kolmogorov – Smirnov Goodness of Fit Test for Continuous Data**

The KS test was originally designed for continuous data on ratio, ordinal, or interval scales of measurement. In the continuous case, the scale of measurement has a quantifiable interval that we’ll call \( X \) (e.g., temperature, heights along tree bole, water depth, etc. – thus, a measurement of sample location is performed). The test for continuous data looks at the maximum vertical distance between the two step functions (observed and predicted proportions – see page 482).

\[ D_i = |F_i - \hat{F}_i| \quad \text{and} \quad D_i' = |F_{i-1} - \hat{F}_i|, \]

### Table

<table>
<thead>
<tr>
<th>N=65</th>
<th>Dark – 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Light – 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_i )</td>
<td>0</td>
<td>7</td>
<td>6</td>
<td>38</td>
<td>14</td>
</tr>
<tr>
<td>( \hat{f}_i )</td>
<td>13.0</td>
<td>13.0</td>
<td>13.0</td>
<td>13.0</td>
<td>13.0</td>
</tr>
<tr>
<td>( F_i )</td>
<td>0</td>
<td>7</td>
<td>13</td>
<td>51</td>
<td>65</td>
</tr>
<tr>
<td>( \hat{F}_i )</td>
<td>13</td>
<td>26</td>
<td>39</td>
<td>52</td>
<td>65</td>
</tr>
<tr>
<td>(</td>
<td>d_i</td>
<td>=</td>
<td>F_i - \hat{F}_i</td>
<td>)</td>
<td>13</td>
</tr>
</tbody>
</table>
where \( \text{rel } F_i = \frac{F_i}{n} \) and \( \hat{F}_i = \frac{X_i}{\text{total } X \text{ under } Ho} \).

So, the test statistic is:

\[
D = \max [\max \{D_i\}, \max \{D_i'\}] = \max \{D_{\text{critical}}\}
\]

This value is compared to a critical value (from Table B.9 in appendix):

\[
D_{\text{critical}} = D_{\alpha, n},
\]

If \( D \geq D_{\alpha, n} \), then Ho is rejected.

Note that in the KS test for continuous data, the graph of the observed cumulative frequency (F) resembles a staircase (see graph on page 483). Thus, we need to examine the vertical distance on the left-hand (\( D_i \)) and right-hand (\( D_i' \)) ends of each step. Zar provides a good example of this test on page 481 (example 22.10).

**The G-test or the Log-Likelihood Ratio Test**

The G-test can be used in the same cases as the chi-square test. There is no general consensus on which test to use. Zar cites a study that recommends the G-test whenever any \( \left| \hat{f}_i - f_i \right| \geq \hat{f}_i \). So, either test will suffice; you can even run both tests to satisfy everybody. The G-statistic is based on the log-likelihood ratio, \( \sum f_i \ln f_i - \sum f_i \ln \hat{f}_i \). The G-statistic is twice this value,

\[
G = 2 \left[ \sum f_i \ln f_i - \sum f_i \ln \hat{f}_i \right] = 4.60517 \left[ \sum f_i \log f_i - \sum f_i \log \hat{f}_i \right].
\]

The G-statistic is distributed approximately \( \chi^2 \). When you have one degree of freedom, you should apply the Yates correction for continuity in the same fashion to a chi-square test. Zar provides a good example of a G-test on page 479 (example 22.8) that uses the same data and hypotheses from example 22.3.