Goodness of Fit & Contingency Tests

Brandan  Victor  Hasan
Outline:

• Goodness of fit test
• Binomial Test
• G-test
• Contingency test
• Fisher’s exact test
• Statistics programs coding
**Introduction: Goodness of Fit**

**Definition:** The goodness of fit test is used to determine whether sample data are consistent with a hypothesized distribution. Or simply used for categorical data when you want to see if your observations fits a theoretical expectation.

Pearson chi-squared $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$

- $o = \text{observed frequency}$
- $e = \text{expected frequency}$
Introduction: Goodness of Fit cont.

**Biological Significance:** Goodness of fit becomes useful when collecting data on age, sex, color morph, etc. and seeing if the collected distribution fits a expected distribution from some theory.

*Example:* Eye coloration in fruit flies.
Goodness of Fit: Assumptions

Non-parametric (does NOT assume normal distribution)

1. Random and Independent samples

2. \( \chi^2 \approx \chi^2 \)

3. No expected values < 1

4. No more than 20% of categories with expected value <5
Goodness of Fit Application

• Once chi-squared ($X^2$) is determined, degrees of freedom (df) is calculated: $df= \# \ of \ categories - 1$

• Critical value can then be found from a table, IF the critical value is less than the chi-squared value the null hypothesis can be rejected

• You can find the chi-squared distribution table through this link: https://www.medcalc.org/manual/chi-square-table.php
G-Statistic

• G-statistic is additive, so for elaborate experiments G-values add up to the overall G-value.

• Chi-squared values for parts of an experiment when added up come close to the overall chi-squared value but are not exact.

• Useful for large data sets; however, when observations are small becomes inaccurate.

• G-statistic: \( O = \) observed values, \( E = \) expected values, and \( \ln = \) natural logarithm.

\[
G = 2 \sum_i (O_i) \ln \left( \frac{O_i}{E_i} \right)
\]
Red Crossbill Example Using Chi-squared

<table>
<thead>
<tr>
<th></th>
<th>Left-billed</th>
<th>Right-billed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Frequency</td>
<td>1895</td>
<td>1752</td>
</tr>
<tr>
<td>Expected Frequency</td>
<td>1823.5</td>
<td>1823.5</td>
</tr>
</tbody>
</table>
Red Crossbill Example Using Chi-squared

<table>
<thead>
<tr>
<th></th>
<th>Left-billed</th>
<th>Right-billed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Frequency</td>
<td>1895</td>
<td>1752</td>
</tr>
<tr>
<td>Expected Frequency</td>
<td>1823.5</td>
<td>1823.5</td>
</tr>
</tbody>
</table>

$H_0$: Distribution of left and right-billed individuals is not significantly different.

$H_1$: Distribution of left and right-billed individuals is significantly different.

$\alpha = 0.05$ or 5%
Red Crossbill Example cont.

\[ X^2 = \frac{(1895 - 1823.5)^2}{1823.5} + \frac{(1752 - 1823.5)^2}{1823.5} \]

\[ X^2 = 5.61 \]

\( \text{df} = 1 \)
Interpreting $X^2$ for Red Crossbills

$X^2 = 5.61$
$\text{df} = 1$
$\alpha = 0.05$

Depends but researchers select significance level of 0.01, 0.05, or 0.10 to determine if the p-value is significant.

Find $X^2$ distribution of statistics in the chi-squared distribution table and compare it to the calculated one.

You can find the chi-squared distribution table through this link: https://www.medcalc.org/manual/chi-square-table.php

In our case we say if $X^2$ is greater than 3.84 we can reject the null hypothesis.

$X^2$ is greater than 3.84, so the null hypothesis is rejected. There are proportionately more left-billed individuals than right.
Goodness of Fit Example (Binomial):

Casino game:
Roll 3 dice; # of sixes determines how much money you win

Gambler plays 100 times. Are his dice rigged?

<table>
<thead>
<tr>
<th>Number of Sixes</th>
<th>Number of Rolls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
• If dice are fair, prob. of rolling 6 on any toss = 1/6
• Binomial Distribution (3, 1/6)

$$P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

Following binomial distribution probability:

Null Hypothesis:

$$p_1 = P(\text{roll 0 sixes}) = P(X=0) = 0.58$$
$$p_2 = P(\text{roll 1 six}) = P(X=1) = 0.345$$
$$p_3 = P(\text{roll 2 sixes}) = P(X=2) = 0.07$$
$$p_4 = P(\text{roll 3 sixes}) = P(X=3) = 0.005$$
$$p_1(0 \text{ sixes}) = 0.58$$
$$p_2(1 \text{ six}) = 0.345$$
$$p_3(2 \text{ sixes}) = 0.07$$
$$p_4(3 \text{ sixes}) = 0.005$$

<table>
<thead>
<tr>
<th>Number of Sixes</th>
<th>Expected Counts</th>
<th>Observed Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58</td>
<td>48</td>
</tr>
<tr>
<td>1</td>
<td>34.5</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>3</td>
</tr>
</tbody>
</table>
\[ X^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \]
<table>
<thead>
<tr>
<th>Number of Sixes</th>
<th>Expected Counts</th>
<th>Observed Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58</td>
<td>48</td>
</tr>
<tr>
<td>1</td>
<td>34.5</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \frac{(48-58)^2}{58} + \frac{(35-34.5)^2}{58} + \frac{(15-7)^2}{7} + \frac{(3-0.5)^2}{0.5} \]

\[ \chi^2 = 23.367 \]

- K=4,
- Degrees of freedom = K-1 = 3
Find the $X^2$ distribution of statistics from the chi-squared table and compare it to the calculated one.

You can find the chi-squared distribution table through this link: https://www.medcalc.org/manual/chi-square-table.php

23.67 > 7.81

So, reject the null hypothesis

Dice are not fair
A Brief Introduction Into Contingency Test:

• When analysis of categorical data is concerned with more than one variable, two way table (also known as contingency tables) are employed.

• These tables provide a foundation for statistical inference, where statistical tests question the relationship between the variables on the basis of the data observed.
Assumptions for Contingency Test:

1) Subjects are randomly sampled and independent

2) No expected value can be less than 1

3) Not more than 20% of expected can have a value less than 5
   • If there are more than 20%, then pooling of the category with less than 5 to the adjacent one
### Example:

<table>
<thead>
<tr>
<th>Goals</th>
<th>Grade</th>
<th>Grade</th>
<th>Grade</th>
<th>Grade</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
<td>168</td>
</tr>
<tr>
<td>Grades (marks)</td>
<td>49</td>
<td>50</td>
<td>69</td>
<td></td>
<td>168</td>
</tr>
<tr>
<td>Popular</td>
<td>24</td>
<td>36</td>
<td>38</td>
<td></td>
<td>98</td>
</tr>
<tr>
<td>Sports</td>
<td>19</td>
<td>22</td>
<td>28</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>108</td>
<td>135</td>
<td></td>
<td>335</td>
</tr>
</tbody>
</table>
• The expected values would be calculated based on the following:
  • Find the sum of each row, and each column
  • Find the total sum of all columns and rows
  • For each cell, multiply the row sum with the column sum and divide it by the total sum of all cells.
    \( \frac{\text{Row sum} \times \text{Column sum}}{\text{total sum}} \)
### Observed:

<table>
<thead>
<tr>
<th>Goals</th>
<th>Grade</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grades</td>
<td>49</td>
<td>50</td>
<td>69</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>Popular</td>
<td>24</td>
<td>36</td>
<td>38</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Sports</td>
<td>19</td>
<td>22</td>
<td>28</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>108</td>
<td>135</td>
<td>335</td>
<td></td>
</tr>
</tbody>
</table>

### Expected:

<table>
<thead>
<tr>
<th>Goals</th>
<th>Grade</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grades</td>
<td>46.1</td>
<td>54.2</td>
<td>67.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Popular</td>
<td>26.9</td>
<td>31.6</td>
<td>39.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports</td>
<td>18.9</td>
<td>22.2</td>
<td>27.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first cell in the expected values table, Grade 4 with "grades" chosen to be most important, is calculated to be \((92/335) \times 168 = 46.1\)
• The distribution of the statistic $X^2$ is chi-square with $(r-1)(c-1)$ degrees of freedom, where $r$ represents the number of rows in the contingency table and $c$ represents the number of columns.

• The P-value for the chi-square test is $P(\chi^2 \geq X^2)$, the probability of observing a value at least as extreme as the test statistic for a chi-square distribution with $(r-1)(c-1)$ df.

• Here in this example, the scientists set $P$ to 0.9
Once the expected value for each cell is found, chi squared formula would be used:

\[ \chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \]

\[ \chi^2 = \frac{(49 - 46.1)^2}{46.1} + \frac{(50 - 54.2)^2}{54.2} + \frac{(69 - 67.7)^2}{67.7} + \ldots + \frac{(28 - 27.8)^2}{27.8} \]
\[ = 0.18 + 0.33 + 0.03 + \ldots + 0.01 \]
\[ = 1.51 \]
Df is equal to \((3-1)(3-1) = 2*2 = 4\), so we are interested in the probability \(\chi^2 \geq 1.51\) on 4 degrees of freedom and P-value of 0.9.

- This value would be found in the chi-squared distribution table.
- \(\chi^2\) is 1.064, where 1.51 > 1.064.

You can find the chi-squared distribution table through this link: [https://www.medcalc.org/manual/chi-square-table.php](https://www.medcalc.org/manual/chi-square-table.php)
This indicates that there is no association between the choice of most important factor and the grade of the student -- the difference between observed and expected values under the null hypothesis is negligible and thus it’s rejected.
Fisher’s Exact Test of Independence

• Use the Fisher's exact test of independence when you have two nominal variables and you want to see whether the proportions of one variable are different depending on the value of the other variable. Use it when the sample size is small.
Assumptions of Fisher’s Test

• The number of samples should be less than 20.
• If N>20, no more than 80% of expected values greater than 5
• Individual observations are independent
• The test assumes that the row and column totals are fixed, or conditional but not random
  • If the totals are unconditioned, the test is not exact.
How the test works:

• Unlike most statistical tests, Fisher's exact test does not use a mathematical function that estimates the probability of a value of a test statistic; instead, you calculate the probability of getting the observed data, and all data sets with more extreme deviations, under the null hypothesis that the proportions are the same.
Example:

• In a van Nood et al. (2013) experiment, the scientists studied patients with *Clostridium difficile* infections, which cause persistent diarrhea. **One nominal variable was the treatment:** some patients were given the antibiotic vancomycin, and some patients were given a fecal transplant. **The other nominal variable was outcome:** each patient was either cured or not cured.
• The percentage of people who received one fecal transplant and were cured (13 out of 16, or 81%) is higher than the percentage of people who received vancomycin and were cured (4 out of 13, or 31%), which seems promising, but the sample sizes seem kind of small.

• Fisher's exact test will tell you whether this difference between 81% and 31% is statistically significant.
• Impractical to calculate by hand

<table>
<thead>
<tr>
<th></th>
<th>Fecal Transplant</th>
<th>Vancomycin</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sick</td>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Cured</td>
<td>13</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Totals</td>
<td>16</td>
<td>13</td>
<td>29</td>
</tr>
</tbody>
</table>

$H_0$ = Proportions of still sick and cured people are the same between the two treatments

$H_A$ = Proportions of still sick and cured people are not the same between the two treatments

(Two Tailed Test)
In order to calculate the probability:

Hypergeometric rule: $\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} = \text{Probability of any given Matrix}$

<table>
<thead>
<tr>
<th></th>
<th>Fecal Transplant</th>
<th>Vancomycin</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sick</td>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Cured</td>
<td>13</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Totals</td>
<td>16</td>
<td>13</td>
<td>29</td>
</tr>
</tbody>
</table>

$P = 0.00772$
Calculate the **Probabilities** of all other permutations of the observed values
Example of permutations who’s cell disproportions are greater than the observed matrix (more extreme distributions)

<table>
<thead>
<tr>
<th></th>
<th>Fecal Transplant</th>
<th>Vancomycin</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sick</td>
<td>2</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Cured</td>
<td>14</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>Totals</td>
<td>16</td>
<td>13</td>
<td>29</td>
</tr>
</tbody>
</table>

P=0.000661

<table>
<thead>
<tr>
<th></th>
<th>Fecal Transplant</th>
<th>Vancomycin</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sick</td>
<td>1</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Cured</td>
<td>15</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Totals</td>
<td>16</td>
<td>13</td>
<td>29</td>
</tr>
</tbody>
</table>

P=0.0000240

<table>
<thead>
<tr>
<th></th>
<th>Fecal Transplant</th>
<th>Vancomycin</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sick</td>
<td>0</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Cured</td>
<td>16</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Totals</td>
<td>16</td>
<td>13</td>
<td>29</td>
</tr>
</tbody>
</table>

P=0.000000251
• Add up the P of all permutations to get the total P-Value.

• For our Experiment, total **P-value = 0.00953**

• \(\alpha = 0.05\)

• The probability P calculated = 0.00953 <0.05, so we can reject \(H_0\)
• **For two tailed P test:**
  • You calculate the probabilities of getting deviations as extreme as the observed, but in the opposite directions.
    • There are several different techniques to calculate that probability, but the most common is to add together the probabilities of all combinations that have lower probabilities than that of the observed data.

• **For one-tailed P test:**
  • You would use a one-tailed test only if you decided, before doing the experiment, that your null hypothesis was that the proportion of sick fecal transplant people was the same as, or greater than, sick vancomycin people.
YATES CORRECTION

- Yates’ correction for continuity is used to correct the P-values of Chi-Square and G-test.
  - It subtracts 0.5 from each observed value that is greater than the expected, and add 0.5 to each observed value that is less than the expected, then chi-squared and G-test are done.
  - This only applies to tests with one df

- [http://www.biostathandbook.com/small.html](http://www.biostathandbook.com/small.html)
Contingency and Fisher’s Exact Tests on SAS:

```sas
data treatment;
input Outcome $ Treatment $ count;
datalines;
Sick Fecal 3
Sick Vancomycin 9
Cured Fecal 13
Cured Vancomycin 4
;
proc freq data=treatment;
tables Outcome*Treatment / chisq nocol norow nopercent expected;
weight count;
run;
```
## Treatment of Patients with C. difficile Infection

The FREQ Procedure

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Expected</th>
<th>Table of Outcome by Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fecal</td>
</tr>
<tr>
<td>Cured</td>
<td>13</td>
<td>9.3793</td>
</tr>
<tr>
<td>Sick</td>
<td>3</td>
<td>6.6207</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>13</td>
</tr>
</tbody>
</table>

### Statistics for Table of Outcome by Treatment

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>1</td>
<td>7.5350</td>
<td>0.0061</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>1</td>
<td>7.8454</td>
<td>0.0051</td>
</tr>
<tr>
<td>Continuity Adj. Chi-Square</td>
<td>1</td>
<td>5.5976</td>
<td>0.0180</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>7.2752</td>
<td>0.0070</td>
</tr>
<tr>
<td>Phi Coefficient</td>
<td></td>
<td>0.5097</td>
<td></td>
</tr>
</tbody>
</table>
## Statistics for Table of Outcome by Treatment

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>1</td>
<td>7.5350</td>
<td>0.0061</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>1</td>
<td>7.8454</td>
<td>0.0051</td>
</tr>
<tr>
<td>Continuity Adj. Chi-Square</td>
<td>1</td>
<td>5.5976</td>
<td>0.0180</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>7.2752</td>
<td>0.0070</td>
</tr>
<tr>
<td>Phi Coefficient</td>
<td></td>
<td>0.5097</td>
<td></td>
</tr>
<tr>
<td>Contingency Coefficient</td>
<td></td>
<td>0.4541</td>
<td></td>
</tr>
<tr>
<td>Cramer’s V</td>
<td></td>
<td>0.5097</td>
<td></td>
</tr>
</tbody>
</table>

### Fisher’s Exact Test

- Cell (1,1) Frequency (F): 13
- Left-sided Pr <= F: 0.9993
- Right-sided Pr >= F: 0.0084
- Table Probability (P): 0.0077
- Two-sided Pr <= P: **0.0095**

**Sample Size = 29**
R Example: Forest Bird Foraging

• Mannan & Meslow (1984) observed which trees Red-breasted Nuthatch prefered to forage in. Forest composition was 54% Douglas fir, 40% Ponderosa pine, 5% Grand fir, and 1% Western Larch.

• Null hypothesis- birds forage randomly without consideration to what tree they are on.

• Total of 156 observations were made with 70 in Douglas fir, 79 in Ponderosa pine, 3 in Grand fir, and 4 in Western larch.

• Are the differences in proportions significant?
R Example: Forest Bird Foraging cont.

Determining the chi-squared value:

observed = c(70, 79, 3, 4)
expected = c(0.54, 0.40, 0.05, 0.01)

chisq.test(x = observed, p = expected)

X-squared = 13.5934, df = 3, p-value = 0.0035
R Example: Forest Bird Foraging Graphing Data

```
Input="("Tree Value Count Total Proportion Expected"
'Douglas fir' Observed 70  156    0.4487    0.54
'Douglas fir' Expected 54  100    0.54      0.54
'Ponderosa pine' Observed 79  156    0.5064    0.40
'Ponderosa pine' Expected 40  100    0.40      0.40
'Grand fir'   Observed  3   156    0.0192    0.05
'Grand fir'   Expected  5   100    0.05      0.05
'Western larch' Observed  4   156    0.0256    0.01
'Western larch' Expected  1   100    0.01      0.01")"
```

Forage = read.table(textConnection(Input),header=TRUE)
R Example: Forest Bird Foraging Graphing Data cont.

```r
### Specify the order of factor levels. Otherwise R will alphabetize them.
library(dplyr)

Forage =
mutate(Forage,
  Tree = factor(Tree, levels=unique(Tree)),
  Value = factor(Value, levels=unique(Value)))
```
R Example: Forest Bird Foraging Graphing Data cont.

### Add confidence intervals

```r
Forage = mutate(Forage, low.ci = apply(Forage[c("Count", "Total", "Expected")], 1, function(x) binom.test(x["Count"], x["Total"], x["Expected"])$ conf.int[1]), upper.ci = apply(Forage[c("Count", "Total", "Expected")], 1, function(x) binom.test(x["Count"], x["Total"], x["Expected"])$ conf.int[2]))
Forage$ low.ci [Forage$ Value == "Expected"] = 0
Forage$ upper.ci [Forage$ Value == "Expected"] = 0
```
### R Example: Forest Bird Foraging Graphing Data cont.

**Forage**

<table>
<thead>
<tr>
<th>Tree</th>
<th>Value</th>
<th>Count</th>
<th>Total</th>
<th>Proportion</th>
<th>Expected</th>
<th>low.ci</th>
<th>upper.ci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Douglas fir</td>
<td>Observed</td>
<td>70</td>
<td>156</td>
<td>0.4487</td>
<td>0.54</td>
<td>0.369115906</td>
<td>0.53030534</td>
</tr>
<tr>
<td>Douglas fir</td>
<td>Expected</td>
<td>54</td>
<td>100</td>
<td>0.5400</td>
<td>0.54</td>
<td>0.0000000000</td>
<td>0.0000000000</td>
</tr>
<tr>
<td>Ponderosa pine</td>
<td>Observed</td>
<td>79</td>
<td>156</td>
<td>0.5064</td>
<td>0.40</td>
<td>0.425290653</td>
<td>0.58728175</td>
</tr>
<tr>
<td>Ponderosa pine</td>
<td>Expected</td>
<td>40</td>
<td>100</td>
<td>0.4000</td>
<td>0.40</td>
<td>0.0000000000</td>
<td>0.0000000000</td>
</tr>
<tr>
<td>Grand fir</td>
<td>Observed</td>
<td>3</td>
<td>156</td>
<td>0.0192</td>
<td>0.05</td>
<td>0.003983542</td>
<td>0.05516994</td>
</tr>
<tr>
<td>Grand fir</td>
<td>Expected</td>
<td>5</td>
<td>100</td>
<td>0.0500</td>
<td>0.05</td>
<td>0.0000000000</td>
<td>0.0000000000</td>
</tr>
<tr>
<td>Western larch</td>
<td>Observed</td>
<td>4</td>
<td>156</td>
<td>0.0256</td>
<td>0.01</td>
<td>0.007029546</td>
<td>0.06434776</td>
</tr>
<tr>
<td>Western larch</td>
<td>Expected</td>
<td>1</td>
<td>100</td>
<td>0.0100</td>
<td>0.01</td>
<td>0.0000000000</td>
<td>0.0000000000</td>
</tr>
</tbody>
</table>
R Example: Forest Bird Foraging Graphing Data cont.

```r
### Plot
library(ggplot2)
library(grid)
```
**R Example: Forest Bird Foraging Graphing**

Data cont.

```r
ggplot(Forage, aes(x = Tree, y = Proportion, fill = Value, ymax=upper.ci, ymin=low.ci)) +
  geom_bar(stat="identity", position = "dodge", width = 0.7) +
  geom_bar(stat="identity", position = "dodge",
            colour = "black", width = 0.7,
            show_guide = FALSE) +
  scale_y_continuous(breaks = seq(0, 0.60, 0.1),
                    limits = c(0, 0.60),
                    expand = c(0, 0)) +
  scale_fill_manual(name = "Count type",
                    values = c('grey80', 'grey30'),
                    labels = c("Observed value",
                               "Expected value")) +
  geom_errorbar(position=position_dodge(width=0.7),
                width=0.0, size=0.5, color="black") +
  labs(x = "Tree species",
       y = "Foraging proportion") +
  ## ggtitle("Main title") +
  theme_bw() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.major.y = element_line(colour = "grey50"),
        plot.title = element_text(size = rel(1.5),
                                face = "bold", vjust = 1.5),
        axis.title = element_text(face = "bold"),
        legend.position = "top",
        legend.title = element_blank(),
        legend.key.size = unit(0.4, "cm"),
        legend.key = element_rect(fill = "black"),
        axis.title.y = element_text(vjust = 1.8),
        axis.title.x = element_text(vjust = -0.5))
```

References

• http://www.stat.yale.edu/Courses/1997-98/101/chisq.htm
• http://www.biostathandbook.com/fishers.html
• http://www.biostathandbook.com/gtestgof.html