Section 7.1

Testing Goodness-of-Fit for a Single Categorical Variable

Kari Lock Morgan

Multiple Categories

- So far, we’ve learned how to do inference for categorical variables with only two categories
- Today, we’ll learn how to do hypothesis tests for categorical variables with multiple categories

Rock-Paper-Scissors (Roshambo)

How would we test whether all of these categories are equally likely?

Hypothesis Testing

1. State Hypotheses
2. Calculate a statistic, based on your sample data
3. Create a distribution of this statistic, as it would be observed if the null hypothesis were true
4. Measure how extreme your test statistic from (2) is, as compared to the distribution generated in (3)

Hypotheses

Let \( p_1 \) denote the proportion in the \( i \)th category.

\( H_0: \) All \( p_i \) s are the same
\( H_1: \) At least one \( p_i \) differs from the others

OR

\( H_0: \) Every \( p_i = 1/3 \)
\( H_1: \) At least one \( p_i \neq 1/3 \)
Test Statistic

Why can’t we use the familiar formula

\[
\frac{\text{sample statistic} - \text{null value}}{\text{SE}}
\]

to get the test statistic?

- More than one sample statistic
- More than one null value

We need something a bit more complicated…

Expected Counts

- The *expected counts* are the expected counts if the null hypothesis were true
- For each cell, the expected count is the sample size (n) times the null proportion, \( p_i \)

\[
\text{expected} = np_i
\]

Chi-Square Statistic

- A **test statistic** is one number, computed from the data, which we can use to assess the null hypothesis
- The **chi-square statistic** is a test statistic for categorical variables:

\[
\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}
\]
What Next?

We have a test statistic. What else do we need to perform the hypothesis test?

**A distribution of the test statistic assuming H₀ is true**

How do we get this?

Two options:
1) Simulation
2) Distributional Theory

Upper-Tail p-value

- To calculate the p-value for a chi-square test, we always look in the *upper tail*

- Why?
  - Values of the χ² are always positive
  - The higher the χ² statistic is, the farther the observed counts are from the expected counts, and the stronger the evidence against the null

Chi-Square (χ²) Distribution

- If each of the expected counts are at least 5, AND if the null hypothesis is true, then the χ² statistic follows a χ² –distribution, with degrees of freedom equal to

  \[ df = \text{number of categories} - 1 \]

- Rock-Paper-Scissors:

  \[ df = 3 - 1 = 2 \]
**Goodness of Fit**

- A *chi-square test for goodness of fit* tests whether the distribution of a categorical variable is the same as some null hypothesized distribution.
- The null hypothesized proportions for each category do not have to be the same.

**Chi-Square Test for Goodness of Fit**

1. State null hypothesized proportions for each category, \( p_i \). Alternative is that at least one of the proportions is different than specified in the null.
2. Calculate the expected counts for each cell as \( np_i \).
3. Calculate the \( \chi^2 \) statistic:
   \[
   \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}
   \]
4. Compute the p-value as the proportion above the \( \chi^2 \) statistic for either a randomization distribution or a \( \chi^2 \) distribution with \( df = (\# \text{ of categories} - 1) \) if expected counts all \( > 5 \).
5. Interpret the p-value in context.

**Mendel’s Pea Experiment**

- In 1866, Gregor Mendel, the “father of genetics” published the results of his experiments on peas.
- He found that his experimental distribution of peas closely matched the theoretical distribution predicted by his theory of genetics (involving alleles, and dominant and recessive genes).


**Mendel’s Pea Experiment**

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Theoretical Proportion</th>
<th>Observed Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round, Yellow</td>
<td>9/16</td>
<td>315</td>
</tr>
<tr>
<td>Round, Green</td>
<td>3/16</td>
<td>101</td>
</tr>
<tr>
<td>Wrinkled, Yellow</td>
<td>3/16</td>
<td>108</td>
</tr>
<tr>
<td>Wrinkled, Green</td>
<td>1/16</td>
<td>32</td>
</tr>
</tbody>
</table>

Let’s test this data against the null hypothesis of each \( p_i \) equal to the theoretical value, based on genetics.

\[
H_0: p_1 = 9/16, p_2 = 3/16, p_3 = 3/16, p_4 = 1/16
\]

\[H_1: \text{At least one } p_i \text{ is not as specified in } H_0\]
Mendel’s Pea Experiment

• \( \chi^2 = 0.47 \)
• Two options:
  o Simulate a randomization distribution
  o Compare to a \( \chi^2 \) distribution with \( 4 - 1 = 3 \) df

\[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]

Mendel’s Pea Experiment

p-value = 0.925

Does this prove Mendel’s theory of genetics?
Or at least prove that his theoretical proportions for pea phenotypes were correct?

a) Yes
b) No

Chi-Square Goodness of Fit

• You just learned about a chi-square goodness of fit test, which compares a single categorical variable to null hypothesized proportions for each category:
  1. Find expected (if \( H_0 \) true) counts for each cell: \( np \),
  2. Compute \( \chi^2 \) statistic to measure how far observed counts are from expected:
  \[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]
  3. Compare \( \chi^2 \) statistic to \( \chi^2 \) distribution with \( (df = \# \text{ categories} - 1) \) or randomization distribution to find upper-tailed p-value

ADHD or Just Young?

• In British Columbia, Canada, the cutoff date for entering school in any year is December 31\(^{st}\), so those born late in the year are younger than those born early in the year
• Are children born late in the year (younger than their peers) more likely to be diagnosed with ADHD?
• Study on 937,943 children 6-12 years old in British Columbia


<table>
<thead>
<tr>
<th>Birth Date</th>
<th>Proportion of Births</th>
<th>ADHD Diagnoses</th>
<th>Expected Counts</th>
<th>Contribution to ( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-Mar</td>
<td>0.244</td>
<td>6880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr-Jun</td>
<td>0.258</td>
<td>7982</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul-Sep</td>
<td>0.257</td>
<td>9161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct-Dec</td>
<td>0.241</td>
<td>8945</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ADHD or Just Young (Boys)

We have VERY (!!!) strong evidence that boys who are born later in the year, and so are younger than their classmates, are more likely to be diagnosed with ADHD.

p-value ≈ 0

ADHD or Just Young? (Girls)

• Want more practice?
• Here is the data for girls. \((\chi^2 = 236.8)\)

<table>
<thead>
<tr>
<th>Birth Date</th>
<th>Proportion of Births</th>
<th>ADHD Diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-Mar</td>
<td>0.243</td>
<td>1960</td>
</tr>
<tr>
<td>Apr-Jun</td>
<td>0.258</td>
<td>2358</td>
</tr>
<tr>
<td>Jul-Sep</td>
<td>0.257</td>
<td>2859</td>
</tr>
<tr>
<td>Oct-Dec</td>
<td>0.242</td>
<td>2904</td>
</tr>
</tbody>
</table>

Section 7.2

Testing for an Association between Two Categorical Variables

Kari Lock Morgan

Review: Chi-Square Goodness of Fit

• In the last section, a chi-square goodness of fit test, which compares a single categorical variable to null hypothesized proportions for each category:
  1. Observed counts from data
  2. Find expected (if \(H_0\) true) counts for each cell
  3. Compute \(\chi^2\) statistic to measure how far observed counts are from expected:
     \[
     \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}
     \]
  4. Compare \(\chi^2\) statistic to \(\chi^2\) distribution to find upper-tailed p-value

Two Categorical Variables

• The statistics behind a \(\chi^2\) test easily extends to two categorical variables
• A \(\chi^2\) test for association (often called a \(\chi^2\) test for independence) tests for an association between two categorical variables
• Everything is the same as a chi-square goodness-of-fit test, except:
  • The hypotheses
  • The expected counts
  • Degrees of freedom for the \(\chi^2\)-distribution

Award Preference & SAT

The data in StudentSurvey includes two categorical variables:

\(\text{Award} = \text{Academy, Nobel, or Olympic}\)

\(\text{HigherSAT} = \text{Math or Verbal}\)

Do you think there is a relationship between the award preference and which SAT is higher? If so, in what way?
**Statistics: Unlocking the Power of Data**

### Award Preference & SAT

<table>
<thead>
<tr>
<th>HigherSAT</th>
<th>Academy</th>
<th>Nobel</th>
<th>Olympic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>21</td>
<td>68</td>
<td>116</td>
<td>205</td>
</tr>
<tr>
<td>Verbal</td>
<td>10</td>
<td>79</td>
<td>61</td>
<td>150</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31</td>
<td>147</td>
<td>177</td>
<td>355</td>
</tr>
</tbody>
</table>

Data are summarized with a 2×3 table for a sample of size n=355.

**H₀**: Award preference is not associated with which SAT is higher

**H₁**: Award preference is associated with which SAT is higher

If H₀ is true ⇒ The award distribution is expected to be the same in each row.

### Expected Counts

\[ \text{Expected Count} = \frac{\text{row total} \times \text{column total}}{n} \]

<table>
<thead>
<tr>
<th>HigherSAT</th>
<th>Academy</th>
<th>Nobel</th>
<th>Olympic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td>205</td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td></td>
<td></td>
<td>150</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31</td>
<td>147</td>
<td>177</td>
<td>355</td>
</tr>
</tbody>
</table>

Notes: The expected counts maintain row and column totals, but redistribute the counts as if there were no association.

### Chi-Square Statistic

\[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]

<table>
<thead>
<tr>
<th>HigherSAT</th>
<th>Academy</th>
<th>Nobel</th>
<th>Olympic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31</td>
<td>147</td>
<td>177</td>
</tr>
</tbody>
</table>

**Randomization Test**

www.lock5stat.com/statkey

p-value=0.001 ⇒ Reject H₀
We have evidence that award preference is associated with which SAT score is higher.

### Chi-Square (χ²) Distribution

- If each of the expected counts are at least 5, AND if the null hypothesis is true, then the χ² statistic follows a χ² distribution, with degrees of freedom equal to

\[ df = (\text{number of rows} - 1)(\text{number of columns} - 1) \]

- Award by HigherSAT:

\[ df = (2 - 1)(3 - 1) = 2 \]

We have evidence that award preference is associated with which SAT score is higher.
Chi-Square Test for Association

Note: The $\chi^2$-test for two categorical variables only indicates if the variables are associated. Look at the contribution in each cell for the possible nature of the relationship.

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

1. State Hypotheses

$H_0$: Type of tag and survival are not associated
$H_A$: Type of tag and survival are associated

2. Create two-way table of observed counts

<table>
<thead>
<tr>
<th>Tag</th>
<th>Survival</th>
<th>Died</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal</td>
<td>Survived</td>
<td>Died</td>
<td>Total</td>
</tr>
<tr>
<td>Electronic</td>
<td>Survived</td>
<td>Died</td>
<td>Total</td>
</tr>
<tr>
<td>Total</td>
<td>Survived</td>
<td>Died</td>
<td>Total</td>
</tr>
</tbody>
</table>

3. Calculate expected counts

$$\text{Expected Count} = \frac{\text{row total} \times \text{column total}}{N}$$

4. Compute the $p$-value as the area in the tail above the $\chi^2$ statistic using either a randomization distribution, or a $\chi^2$ distribution with $df = (r - 1)(c - 1)$ if all expected counts $> 5$

5. Interpret the $p$-value in context.

Metal Tags and Penguins

Are metal tags detrimental to penguins? A study looked at the 10 year survival rate of penguins tagged either with a metal tag or an electronic tag. 20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins.

Is there an association between type of tag and survival?

Metal Tags and Penguins

4. Calculate chi-square statistic

<table>
<thead>
<tr>
<th></th>
<th>Survived</th>
<th>Died</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal Tag</td>
<td>33 (47.4)</td>
<td>134 (119.6)</td>
</tr>
<tr>
<td>Electronic Tag</td>
<td>68 (53.6)</td>
<td>121 (135.4)</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]

5. Compute p-value and interpret in context.

\[ p-value = 0.0007 \] (using a \( \chi^2 \)-distribution with 1 df)

There is strong evidence of an association between type of tag and survival of penguins, with electronically tagged penguins having better survival than metal tagged.

2 x 2 Table

- Note: because each of these variables only has two categories, we could have done this as a difference in proportions test
- We would have gotten the exact same p-value!
- For two categorical variables with two categories each, can do either difference in proportions or a chi-square test

Summary: Chi-Square Tests

- The \( \chi^2 \) goodness-of-fit tests if one categorical variable differs from a null distribution
- The \( \chi^2 \) test for association tests for an association between two categorical variables
- For both, you compute the expected counts in each cell (assuming \( H_0 \)) and the \( \chi^2 \) statistic:
  \[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]
- Find the proportion above the \( \chi^2 \) statistic in a randomization or \( \chi^2 \)-distribution (if all expected counts > 5)