Chi-Square Tests

SECTIONS 7.1, 7.2
- Testing the distribution of a single categorical variable: $\chi^2$ goodness of fit (7.1)
- Testing for an association between two categorical variables: $\chi^2$ test for association (7.2)

Comments on Projects
- "Random sampling" is different than convenience sampling
- Even if your initial sample selected is representative, if people choose not to respond this introduces bias
- Generally a good idea to present the response rate

Sample and Population
Was your sample the same as your population? (Did you have data on all the cases in your population?)

a) Yes
b) No

The Big Picture
Population
Sampling
Statistical Inference
Sample

Comments on Projects
- Sample is whatever your variable is measured on (rows of your dataset)
- The type of inference we have learned can only generalize from sample to population (and only if a random sample)
- We have not learned how to account for uncertainty in the data values themselves

Comments on Projects
- If you have data on the whole population, inference is not necessary!
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

Which did you throw first?

a) Rock
b) Paper
c) Scissors

Rock-Paper-Scissors (Roshambo)

Play a game!
Can we use statistics to help us win?

Hypothesis Testing

1. State Hypotheses
2. Calculate a statistic, based on your sample data
   \[ \text{test statistic} \]
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 \( p_i \) denote the proportion in the \( i^{th} \) category.

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

OR

\( H_0 \) : Every \( p_i = 1/3 \)
\( H_a \) : 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...

---

**Observed Counts**

- The *observed counts* are the actual counts observed in the study

<table>
<thead>
<tr>
<th></th>
<th>ROCK</th>
<th>PAPER</th>
<th>SCISSORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>36</td>
<td>12</td>
<td>27</td>
</tr>
</tbody>
</table>

---

**Expected Counts**

- The *expected counts* are the expected counts if the null hypothesis were true

<table>
<thead>
<tr>
<th></th>
<th>ROCK</th>
<th>PAPER</th>
<th>SCISSORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>36</td>
<td>12</td>
<td>27</td>
</tr>
<tr>
<td>Expected</td>
<td>28.3</td>
<td>28.3</td>
<td>28.3</td>
</tr>
</tbody>
</table>

- 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:

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

---

**Rock-Paper-Scissors**

<table>
<thead>
<tr>
<th></th>
<th>ROCK</th>
<th>PAPER</th>
<th>SCISSORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>36</td>
<td>12</td>
<td>27</td>
</tr>
<tr>
<td>Expected</td>
<td>28.3</td>
<td>28.3</td>
<td>28.3</td>
</tr>
</tbody>
</table>

---

**What Next?**

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

*An A distribution of the test statistic assuming \( H_0 \) is true*

How do we get this? Two options:
1) Simulation
2) Distributional Theory
Simulation

Let’s see what kind of statistics we would get if rock, paper, and scissors are all equally likely.

1) Take 3 scraps of paper and label them Rock, Paper, Scissors. Fold or crumple them so they are indistinguishable. Choose one at random.

2) What did you get? (a) Rock (b) Paper (c) Scissors

3) Calculate the $\chi^2$ statistic.

4) Form a distribution.

5) How extreme is the actual test statistic?

Chi-Square ($\chi^2$) Distribution

- If each of the expected counts are at least 5, AND if the null hypothesis is true, then the $\chi^2$ statistic follows a $\chi^2$ -distribution, with degrees of freedom equal to

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

- Rock-Paper-Scissors:

  \[ df = 3 - 1 = 2 \]

p-value using $\chi^2$ distribution

Visit [www.lock5stat.com/statkey](http://www.lock5stat.com/statkey) for statistical calculations.

Chi-Square p-values

1. StatKey – simulation or theoretical
2. RStudio: `tail.p("chisquare", stat, df, tail="upper")`
3. TI-83:

   - \[ 2^{nd} \rightarrow \text{DISTR} \rightarrow 7: \chi^2CDF \rightarrow \]
   - lower bound, upper bound, df

Because the $\chi^2$ is always positive, the p-value (the area as extreme or more extreme) is always the upper tail.
**Goodness of Fit**

- A \( \chi^2 \) 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 \( np_i \). Make sure they are all greater than 5 to proceed.
3. Calculate the \( \chi^2 \) statistic:
   \[
   \chi^2 = \sum \frac{(observed - expected)^2}{expected}
   \]
4. Compute the p-value as the area in the tail above the \( \chi^2 \) statistic, for a \( \chi^2 \) distribution with \( df = (# \text{ of categories} - 1) \)
5. Interpret the p-value in context.

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

**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).


<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.

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

\[ H_a : \text{At least one } p_i \text{ is not as specified in } H_0 \]

The \( \chi^2 \) statistic is made up of 4 components (because there are 4 categories). What is the contribution of the first cell (315)?

(a) 0.016  (b) 1.05  (c) 5.21  (d) 107.2
Mendel's Pea Experiment

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>$p_i$</th>
<th>Observed Counts</th>
<th>Expected Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round, Yellow</td>
<td>$9/16$</td>
<td>315</td>
<td>$556 \times 9/16 = 312.75$</td>
</tr>
<tr>
<td>Round, Green</td>
<td>$3/16$</td>
<td>101</td>
<td>$556 \times 3/16 = 104.25$</td>
</tr>
<tr>
<td>Wrinkled, Yellow</td>
<td>$3/16$</td>
<td>108</td>
<td>$556 \times 3/16 = 104.25$</td>
</tr>
</tbody>
</table>
| Wrinkled, Green          | $1/16$| 32              | $556 \times 1/16 = 34.75$ 

$$n = 556 \quad \sum \text{expected} = 556$$

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

$$= \frac{(315 - 312.75)^2}{312.75} + \frac{(101 - 104.25)^2}{104.25} + \frac{(108 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} = 0.47$$

Mendel's Pea Experiment

- $\chi^2 = 0.47$
- Two options:
  - Simulate a randomization distribution
  - Compare to a $\chi^2$ distribution with $4 - 1 = 3 \text{ df}$

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

You can never accept the null hypothesis!!! The null hypothesis can only be rejected with statistically significant results. A high $p$-value doesn’t let you conclude anything.

Rock-Paper-Scissors

- Did reading the example on Rock-Paper-Scissors in the textbook influence the way you play the game?
  - Did you read the Rock-Paper-Scissors example in Section 6.3 of the textbook?
    - a) Yes
    - b) No

Rock-Paper-Scissors

<table>
<thead>
<tr>
<th></th>
<th>Rock</th>
<th>Paper</th>
<th>Scissors</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Example</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Did Not Read Example</td>
<td>26</td>
<td>9</td>
<td>31</td>
<td>66</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>11</td>
<td>37</td>
<td>82</td>
</tr>
</tbody>
</table>

H$_0$: Whether or not a person read the example in the textbook is not associated with how a person plays rock-paper-scissors

H$_1$: Whether or not a person read the example in the textbook is associated with how a person plays rock-paper-scissors
**Expected Counts**

\[ \text{expected} = \frac{\text{row total} \times \text{column total}}{\text{sample size}} \]

Calculate the expected count for your cell. Note: Maintains row and column totals, but redistributes the counts as if there were no association.

**Chi-Square Statistic**

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

\[ \chi^2 = 0.617 \]

If the expected counts are all at least 5, this can be compared to a \( \chi^2 \) distribution with \((r - 1)(c - 1)\) degrees of freedom, where \( r \) = number of rows, \( c \) = number of columns.

**Chi-Square Distribution**

\( df = (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2 \)

**Chi-Square Test for Association**

1. \( H_0 \): The two variables are not associated
   \( H_1 \): The two variables are associated
2. Calculate the expected counts for each cell:
   \[ \text{expected} = \frac{\text{row total} \times \text{column total}}{\text{sample size}} \]
   Make sure they are all greater than 5 to proceed.
3. Calculate the \( \chi^2 \) statistic:
   \[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]
4. Compute the \( p \)-value as the area in the tail above the \( \chi^2 \) statistic, for a \( \chi^2 \) distribution with \( df = (r - 1) \times (c - 1) \)
5. Interpret the \( p \)-value in context.
Metal Tags and Penguins

Let's return to the question of whether metal tags are detrimental to penguins.

<table>
<thead>
<tr>
<th></th>
<th>Survived</th>
<th>Died</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal Tag</td>
<td>33</td>
<td>134</td>
<td>167</td>
</tr>
<tr>
<td>Electronic Tag</td>
<td>68</td>
<td>121</td>
<td>189</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>255</td>
<td>356</td>
</tr>
</tbody>
</table>

Is there an association between type of tag and survival?  
(a) Yes  
(b) No


Metal Tags and Penguins

<table>
<thead>
<tr>
<th></th>
<th>Survived</th>
<th>Died</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal Tag</td>
<td>33 (47.38)</td>
<td>134 (119.62)</td>
<td>167</td>
</tr>
<tr>
<td>Electronic Tag</td>
<td>68 (53.62)</td>
<td>121 (135.38)</td>
<td>189</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>255</td>
<td>356</td>
</tr>
</tbody>
</table>

Calculate expected counts (shown in parentheses). First cell:

\[ \text{expected} = \frac{\text{row total} \times \text{column total}}{\text{sample size}} = \frac{167 \times 101}{356} = 47.38 \]

Calculate the chi-square statistic:

\[ \chi^2 = \frac{(33 - 47.38)^2}{47.38} + \frac{(134 - 119.62)^2}{119.62} + \frac{(68 - 53.62)^2}{53.62} + \frac{(121 - 135.38)^2}{135.38} = 11.48 \]

p-value: \( df = (2 - 1) \times (2 - 1) = 1 \) \( p-value = 0.0007 \)

There is strong evidence that metal tags are detrimental to penguins.

Two Categorical Variables with Two Categories Each

- When testing two categorical variables with two categories each, the \( \chi^2 \) statistic is exactly the square of the \( z \)-statistic
- The \( \chi^2 \) distribution with 1 df is exactly the square of the normal distribution
- Doing a two-sided test for difference in proportions or doing a \( \chi^2 \) test will produce the exact same p-value.

Summary

- The \( \chi^2 \) test for goodness of fit tests whether one categorical variable differs from a hypothesized distribution
- The \( \chi^2 \) test for association tests whether two categorical variables are associated
- For both, you calculate expected counts for each cell, compute the \( \chi^2 \) statistic as

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

and find the p-value as the area above this statistic on a \( \chi^2 \) distribution

To Do

- Read Chapter 7
- Do Homework 6 (due Tuesday, 11/6)