Chi-Squared ($X^2$) Analysis
What is Chi-Squared?

• In genetics, you can predict genotypes based on probability (expected results)

• Chi-squared is a form of statistical analysis used to compare the actual results (observed) with the expected results

• **NOTE:** $X^2$ is the name of the whole variable – you will never take the square root of it or solve for $X$
Chi-squared

• If the **expected** and **observed** (actual) values are the same then the $X^2 = 0$

• If the $X^2$ value is 0 or is small then the data fits your hypothesis (the expected values) well.

• By calculating the $X^2$ value you determine if there is a statistically significant difference between the expected and actual values.
Step 1: Calculating $X^2$

- First, determine what your expected and observed values are.
- Observed (Actual) values: That should be something you get from data—usually no calculations 😊
- Expected values: based on probability
- Suggestion: make a table with the expected and actual values
Step 1: Example

• Observed (actual) values: Suppose you have 90 tongue rollers and 10 nonrollers

• Expected: Suppose the parent genotypes were both Rr → using a punnett square, you would expect 75% tongue rollers, 25% nonrollers

• This translates to 75 tongue rollers, 25 nonrollers (since the population you are dealing with is 100 individuals)
Step 1: Example

- Table should look like this:

<table>
<thead>
<tr>
<th></th>
<th>Expected</th>
<th>Observed (Actual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue rollers</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td>Nonrollers</td>
<td>25</td>
<td>10</td>
</tr>
</tbody>
</table>
Step 2: Calculating $X^2$

- Use the formula to calculate $X^2$
- For each different category (genotype or phenotype) calculate
  
  \[(\text{observed} - \text{expected})^2 / \text{expected}\]

- Add up all of these values to determine $X^2$
Step 2: Calculating $\chi^2$

The formula for Chi-squared is:

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

where $o = \text{observed number of individuals}$

$e = \text{expected number of individuals}$

$\Sigma = \text{the sum of the values}$ (in this case, the differences, squared, divided by the number expected)
Step 2: Example

• Using the data from before:
• Tongue rollers
  \[(90 - 75)^2 / 75 = 3\]
• Nonrollers
  \[(10 - 25)^2 / 25 = 9\]
• \[X^2 = 3 + 9 = 12\]
Step 3: Determining Degrees of Freedom

• Degrees of freedom = # of categories – 1

• Ex. For the example problem, there were two categories (tongue rollers and nonrollers) → degrees of freedom = 2 – 1

• Degrees of freedom = 1
Step 4: Critical Value

- Using the degrees of freedom, determine the critical value using the provided table
- $Df = 1 \rightarrow$ Critical value $= 3.84$

**BIOLOGISTS GENERALLY REJECT THE NULL HYPOTHESIS IF THE VALUE OF P IS LESS THAN 0.05.**
STEP 4A: DETERMINE WHERE YOUR CHI SQUARE VALUE IS IN THE TABLE BELOW:

- To see what p value matches your Chi square value
- Compare your chi square value with those in the row that corresponds to one degree of freedom.

<table>
<thead>
<tr>
<th>Degrees of Freedom</th>
<th>Probability</th>
<th>0.95</th>
<th>0.90</th>
<th>0.80</th>
<th>0.70</th>
<th>0.50</th>
<th>0.30</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.004</td>
<td>0.02</td>
<td>0.06</td>
<td>0.15</td>
<td>0.46</td>
<td>1.07</td>
<td>1.64</td>
<td>2.71</td>
<td>3.84</td>
<td>6.64</td>
<td>10.83</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.10</td>
<td>0.21</td>
<td>0.45</td>
<td>0.71</td>
<td>1.39</td>
<td>2.41</td>
<td>3.22</td>
<td>4.60</td>
<td>5.99</td>
<td>9.21</td>
<td>13.82</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.35</td>
<td>0.58</td>
<td>1.01</td>
<td>1.42</td>
<td>2.37</td>
<td>3.66</td>
<td>4.64</td>
<td>6.25</td>
<td>7.82</td>
<td>11.34</td>
<td>16.27</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.71</td>
<td>1.06</td>
<td>1.65</td>
<td>2.20</td>
<td>3.36</td>
<td>4.88</td>
<td>5.99</td>
<td>7.78</td>
<td>9.49</td>
<td>13.28</td>
<td>18.47</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1.14</td>
<td>1.61</td>
<td>2.34</td>
<td>3.00</td>
<td>4.35</td>
<td>6.06</td>
<td>7.29</td>
<td>9.24</td>
<td>11.07</td>
<td>15.09</td>
<td>20.52</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.63</td>
<td>2.20</td>
<td>3.07</td>
<td>3.83</td>
<td>5.35</td>
<td>7.23</td>
<td>8.56</td>
<td>10.64</td>
<td>12.59</td>
<td>16.81</td>
<td>22.46</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>2.17</td>
<td>2.83</td>
<td>3.82</td>
<td>4.67</td>
<td>6.35</td>
<td>8.38</td>
<td>9.80</td>
<td>12.02</td>
<td>14.07</td>
<td>18.48</td>
<td>24.32</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>2.73</td>
<td>3.49</td>
<td>4.59</td>
<td>5.53</td>
<td>7.34</td>
<td>9.52</td>
<td>11.03</td>
<td>13.36</td>
<td>15.51</td>
<td>20.09</td>
<td>26.12</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>3.94</td>
<td>4.86</td>
<td>6.18</td>
<td>7.27</td>
<td>9.34</td>
<td>11.78</td>
<td>13.44</td>
<td>15.99</td>
<td>18.31</td>
<td>23.21</td>
<td>29.59</td>
</tr>
</tbody>
</table>

Nonsignificant | Significant
• Find your chi square value on the chart.

<table>
<thead>
<tr>
<th>Degrees of Freedom</th>
<th>Probability</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>1</td>
<td>0.004</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>0.35</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>0.71</td>
<td>1.06</td>
</tr>
<tr>
<td>5</td>
<td>1.14</td>
<td>1.61</td>
</tr>
<tr>
<td>6</td>
<td>1.63</td>
<td>2.20</td>
</tr>
<tr>
<td>7</td>
<td>2.17</td>
<td>2.83</td>
</tr>
<tr>
<td>8</td>
<td>2.73</td>
<td>3.49</td>
</tr>
<tr>
<td>10</td>
<td>3.94</td>
<td>4.86</td>
</tr>
</tbody>
</table>

Nonsignificant

Significant
• Suppose that your Chi square value is 0.25
• 0.25 is located in the body of the chart between 0.15 and 0.46.
DETERMINE THE P VALUE

• After determining the position of the number most nearly matching your value, look at the head of the column it is in. The number at the head of the column is the probability (P) that the results obtained in the experiment differ from the expected results by chance.

• BIOLOGISTS GENERALLY REJECT THE NULL HYPOTHESIS IF THE VALUE OF P IS LESS THAN 0.05.
0.25 falls between the columns headed by P values of 0.50 and 0.70.
DETERMINE WHETHER YOU WILL ACCEPT OR REJECT THE NULL HYPOTHESIS:

• In this illustration, the value of P falls between 0.50 and 0.70.

• Clearly, the experimenter must ACCEPT THE NULL HYPOTHESIS.
What does this mean?

• Accepting the null hypothesis means that the results of the experiment differ from the expected only by chance.

• Thus the experimenter can conclude that the subject did not exhibit psychic powers in this particular experiment.
• The smaller the p value the more significant the results are said to be!
Step 5: Conclusion

- If $X^2 >$ critical value…
  there is a statistically significant difference between the actual and expected values.
- If $X^2 <$ critical value…
  there is a NOT statistically significant difference between the actual and expected values.
Step 5: Example

• $X^2 = 12 > 3.84$

→ There is a statistically significant difference between the observed and expected population
Chi-squared and Hardy Weinberg

• Review: If the observed (actual) and expected genotype frequencies are the same then a population is in Hardy Weinberg equilibrium

• But how close is close enough?
  – Use Chi-squared to figure it out!
  – If there isn’t a statistically significant difference between the expected and actual frequencies, then it is in equilibrium
Example

Using the example from yesterday…

<table>
<thead>
<tr>
<th>Ferrets</th>
<th>Expected</th>
<th>Observed (Actual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>$0.45 \times 164 = 74$</td>
<td>78</td>
</tr>
<tr>
<td>Bb</td>
<td>$0.44 \times 164 = 72$</td>
<td>65</td>
</tr>
<tr>
<td>bb</td>
<td>$0.11 \times 164 = 18$</td>
<td>21</td>
</tr>
</tbody>
</table>
Example

• $X^2$ Calculation
  
  BB: \( \frac{(78 - 74)^2}{74} = 0.21 \)
  
  Bb: \( \frac{(72 - 65)^2}{72} = 0.68 \)
  
  bb: \( \frac{(18 - 21)^2}{18} = 0.5 \)
  
  \( \chi^2 = 0.21 + 0.68 + 0.5 = 1.39 \)

• Degrees of Freedom = 3 – 1 = 2

• Critical value = 5.99

• $\chi^2 < 5.99 \rightarrow$ there is not a statistically significant difference between expected and actual values $\rightarrow$ population DOES SEEM TO BE in Hardy Weinberg Equilibrium