Chi Squared test

You have a set of expected observations that are in line with a specific hypothesis

The hypothesis predicts the number (fraction) individuals in each category

The observed numbers don’t match (exactly) the predicted numbers

Is this due to a chance deviation?

Consider a simple experiment
10 fertilization events for Aa X aa
Or 10 flips of a coin

Expect 5 heads and 5 tails
or expect 5 Aa progeny and 5 aa progeny

Observed: 8 head and 2 tails
Is there something wrong with the coin?
How often do we see deviations of 8:2 (or greater)?

<table>
<thead>
<tr>
<th>heads</th>
<th>tails</th>
<th>numerator/1024</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>45</td>
<td>4.4</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>120</td>
<td>11.7</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>210</td>
<td>20.5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>252</td>
<td>24.6</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>210</td>
<td>20.5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>120</td>
<td>11.7</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>45</td>
<td>4.4</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
What if we flipped 100 coins and got  
80 heads and 20 tails

Chi Squared Test
• Is the observed ratio in satisfactory agreement with a theoretical prediction?

• The chi square test tells us whether the observed results deviate too far from a theoretical expectation

• The chi square test tells us the probability that the data we’ve collected deviate from the expected just due to chance

Roller Example question on Mendel handout

State a simple hypothesis that gives a precise prediction about predicted results

O = observed    E= expected

\[ \sum \frac{[O - E]^2}{E} \]

degrees of freedom =  
number of progeny classes - 1

if you have 400 progeny
progeny progeny
class 1    class 2

if class 1 is assigned 300 progeny, then class 2 must have 100 -- this is one degree of freedom

similar logic for multiple classes

if you have 400 progeny
progeny progeny progeny progeny
class 1    class 2    class 3    class 4

4-1 or three degrees of freedom
The Chi squared calculation does not tell you which interpretation about the roller mutant is correct:

It tells you that likelihood that you would get this set of data just by chance if RR is viable

Significant decimal places:

Express the calculated chi square to 3 decimal places, because that is the accuracy of the table of critical values (the chi square table)

To avoid rounding errors, all intermediate computations, including the expected values should be carried out to 4 decimal places