

## Chapter 12

### Quantitative Aside 12.1--Testing Goodness of Fit--Chi Square

Let's assume we have a hypothesis that two genes are assorting independently, and we have data that are close to the expected 9:3:3:1 ratio for a dihybrid cross done using Mendel's crossing scheme. How close to the exact ratio do the data have to be for us to conclude that independent assortment is operating? More generally, if our hypothesis for the mode of inheritance allows for a quantitative prediction, how close do the actual data have to be to accept the hypothesis?

We call this hypothesis testing, or a test of **goodness of fit**. It allows us to assess the probability of deviation from expectations due to chance alone. If the deviation is more than expected in 5% of cases, we suspect the hypothesis is incorrect, and if it would be expected in only 1% of cases, we reject the hypothesis. These levels are referred to as the 5% and 1% level of statistical significance.

The **chi square test** is a test of goodness of fit for a set of data versus expectations that can be used for many genetic hypotheses. We can do this systematically using the following steps:

1. State the hypothesis in detail, specifying genotypes and corresponding phenotypes of parents and possible progeny.
2. Use the rules of probability to make explicit predictions of the number and kind of progeny from the cross.
3. For each progeny class, subtract expected from observed and square this number. Divide this result by expected.
4. Sum all of the results for each progeny class to determine the chi square for the data set.

$$\text{So: Chi Square} = \chi^2 = \sum(\text{Observed} - \text{Expected})^2 / \text{Expected}$$

This chi square value is then compared with a table of values that are arranged based on the degrees of freedom for the number of observations. This is equal to number of

observations – 1. This is the number of independent observations you can make for the data set (the last observation can be determined by subtraction from the total).

We can apply this to the data from earlier in the chapter for the dihybrid cross of round and yellow (p. 228).

| <b>Progeny Class</b>                      | <b>Observed</b> | <b>Expected</b> | <b>(Obs.– Exp.)<sup>2</sup>/Obs.</b> |
|---|-----------------|-----------------|--------------------------------------|
| round yellow                              | 315             | 313             | 0.013                                |
| round green                               | 108             | 104             | 0.15                                 |
| wrinkled yellow                           | 101             | 104             | 0.087                                |
| wrinkled green                            | 32              | 35              | 0.26                                 |
| $\Sigma$ (Obs. – Exp.) <sup>2</sup> /Exp. |                 |                 | 0.51                                 |

We can then compare this  $\chi^2$  value with those in the table below:

| <b>Degrees<br/>of Freedom</b> | <b><u>p Values</u></b> |             |             |             |             |             |
|-------------------------------|------------------------|-------------|-------------|-------------|-------------|-------------|
|                               | <b>0.99</b>            | <b>0.90</b> | <b>0.50</b> | <b>0.10</b> | <b>0.05</b> | <b>0.01</b> |
|                               | $\chi^2$ Calculations  |             |             |             |             |             |
| 1                             | –                      | 0.02        | .45         | 2.71        | 3.84        | 6.64        |
| 2                             | 0.02                   | 0.21        | 1.39        | 4.61        | 5.99        | 9.21        |
| 3                             | 0.11                   | 0.58        | 2.37        | 6.25        | 7.81        | 11.35       |
| 4                             | 0.30                   | 1.06        | 3.36        | 7.78        | 9.49        | 13.28       |
| 5                             | 0.55                   | 1.61        | 4.35        | 9.24        | 11.07       | 15.09       |

From this we can see that there is a more than 90% chance of this great a deviation based on chance alone.