Mendel's Data Falsified?

Because some of Mendel's data from his famous genetics experiments seemed too perfect to be true, statistician R. A. Fisher concluded that the data were probably falsified.

He used a chi-square distribution to show that when a test statistic is extremely far to the left and results in a *P*-value very close to 1, the sample data fit the claimed distribution almost perfectly, and this is evidence that the sample data have not been randomly selected. It has been suggested that Mendel's gardener knew what results Mendel's theory predicted, and subsequently adjusted results to fit that theory.

Ira Pilgrim wrote in *The Journal of Heredity* that this use of the chi-square distribution is not appropriate. He notes that the question is not about goodness-of-fit with a particular distribution, but whether the data are from a sample that is truly random. Pilgrim used the binomial probability formula to find the probabilities of the results obtained in Mendel's experiments. Based on his results, Pilgrim concludes that "there is no reason whatever to question Mendel's honesty." It appears that Mendel's results are not too good to be true, and they could have been obtained from a truly random process.

Interpretation

The sample of leading digits does not provide enough evidence to conclude that the Benford distribution is not being followed.

In Figure 11-3 we use a red line to graph the expected proportions given by Benford's law (as in Table 11-4) along with a green line for the observed proportions from Table 11-4. Figure 11-3 allows us to visualize the "goodness-of-fit" between the distribution given by Benford's law and the frequencies that were observed. In Figure 11-3, the red and green lines agree reasonably well, so it appears that the green line for the observed data fits the red line for the expected values reasonably well.

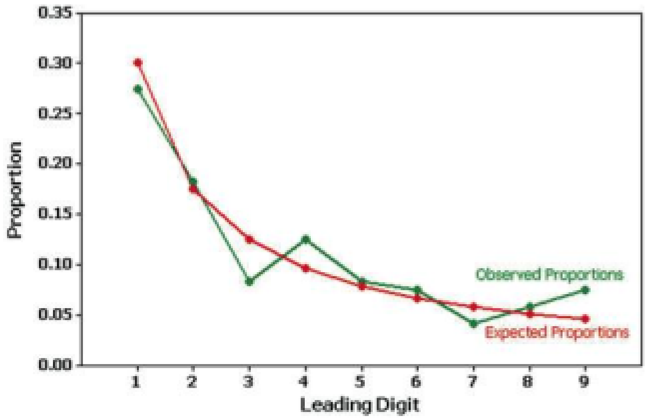

Figure 11-3 **Observed Proportions and Proportions Expected with Benford's Law**

Rationale for the Test Statistic Examples 1 and 2 show that the χ^2 test statistic is a measure of the discrepancy between observed and expected frequencies. Simply summing the differences $O - E$ between observed and expected values tells us nothing, because that sum is always 0. Squaring the $O - E$ gives us a better statistic. (The reasons for squaring the $O - E$ values are essentially the same as the reasons for squaring the $x - \bar{x}$ values in the formula for standard deviation.) The value of $\Sigma(O - E)^2$ measures only the magnitude of the differences, but we need to find the magnitude of the differences relative to what was expected. We need a type of average instead of a cumulative total. This relative magnitude is found through division by the expected frequencies, as in the test statistic.

The theoretical distribution of $\Sigma(O - E)^2/E$ is a discrete distribution because the number of possible values is finite. The distribution can be approximated by a chi-square distribution, which is continuous. This approximation is generally considered acceptable, provided that all expected values E are at least 5. (There are ways of circumventing the problem of an expected frequency that is less than 5, such as combining categories so that all expected frequencies are at least 5. Also, there are other methods that can be used when not all expected frequencies are at least 5.)

The number of degrees of freedom reflects the fact that we can freely assign frequencies to $k - 1$ categories before the frequency for every category is determined. (Although we say that we can "freely" assign frequencies to $k - 1$ categories, we cannot have negative frequencies, nor can we have frequencies so large that their sum exceeds the total of the observed frequencies for all categories combined.)