The test statistic allows us to measure the amount of disagreement between the frequencies actually observed and those that we would theoretically expect when the two variables are independent. Large values of the X^2 test statistic are in the rightmost region of the chi-square distribution, and they reflect significant differences between observed and expected frequencies. The distribution of the test statistic X^2 can be approximated by the chi-square distribution, provided that all expected frequencies are at least 5. The number of degrees of freedom $(r - 1)(c - 1)$ reflects the fact that because we know the total of all frequencies in a contingency table, we can freely assign frequencies to only $r - 1$ rows and $c - 1$ columns before the frequency for every cell is determined. However, we cannot have negative frequencies or frequencies so large that any row (or column) sum exceeds the total of the observed frequencies for that row (or column).

Polls and Psychologists

Poll results can be dramatically affected by the wording of questions. A phrase such as "over the last few years" is interpreted differently by different people. Over the last few years (actually, since 1980), survey researchers and psychologists have been working together to improve surveys by decreasing bias and increasing accuracy. In one case, psychologists studied the finding that 10 to 15 percent of those surveyed say they voted in the last election when they did not. They experimented with theories of faulty memory, a desire to be viewed as responsible, and a tendency of those who usually vote to say that they voted in the most recent election, even if they did not. Only the last theory was actually found to be part of the problem.

Finding Expected Values *E*

The test statistic X^2 is found by using the values of O (observed frequencies) and the values of E (expected frequencies). An individual expected frequency E can be found for a cell by simply multiplying the total of the row frequencies by the total of the column frequencies, then dividing by the grand total of all frequencies, as shown in Example 2.

Example 2 Finding Expected Frequency

Refer to Table 11-6 and find the expected frequency for the first cell, where the observed frequency is 54. (See Table 11-6 on page 577.)

Solution

The first cell lies in the first row (with a total frequency of 66) and the first column (with total frequency of 182). The "grand total" is the sum of all frequencies in the table, which is 253. The expected frequency of the first cell is

$$E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})} = \frac{(66)(182)}{253} = 47.478$$

Interpretation

We know that the first cell has an observed frequency of $O = 54$ and an expected frequency of $E = 47.478$. We can interpret the expected value by stating that if we assume that success is independent of the treatment, then we expect to find that 47.478 of the subjects would be treated with surgery and that treatment would be successful. There is a discrepancy between $O = 54$ and $E = 47.478$ and such discrepancies are key components of the test statistic that is a collective measure of the overall disagreement between the observed frequencies and the frequencies expected with independence between the row and column variables.

Rationale for Expected Frequencies To better understand expected frequencies, pretend that we know only the row and column totals in Table 11-6. Let's assume that the row and column variables are independent and that one of the 253 study subjects is randomly selected. The probability of getting someone counted in the first cell of Table 11-6 is calculated as follows:

$$P(\text{surgery treatment}) = 66/253 \text{ and } P(\text{success}) = 182/253$$

$$P(\text{surgery treatment } \textit{and} \text{ success}) = \frac{66}{253} \cdot \frac{182}{253} = 0.187661$$