Binary Outcomes: Chi-Squared Test of Independence or Goodness of Fit?

Sara E. Burke

Updated January 12, 2014

Let's say an experimenter has randomly assigned participants to two conditions (with equally many participants in each condition) and would like to determine whether the resulting groups differ on a binary outcome.

Let's call the conditions "treatment" and "control," and the outcomes "winning" and "losing." Each participant is placed in either the treatment condition or the control condition, and then takes a test resulting in a win or a loss. The null hypothesis states that condition is unrelated to outcome—wins and losses will be evenly distributed across the treatment and control groups.

The appropriate test to run is a chi-squared test of independence. However, some students might be curious why we cannot run a chi-squared test of goodness of fit. After all, we expect half of the "wins" to come from the treatment condition and half to come from the control condition—this sounds very much like an "expected distribution" of wins.

There are two heuristic responses to this source of confusion between the test of independence and the test of goodness of fit. First, we must consider the distribution of wins *and* the distribution of losses, and the only test that permits this is the one that examines all four cells. Second, the goodness of fit test forces us to hold constant the total number of wins. In the experimental design, the number of wins is permitted to vary, so the goodness-of-fit test ignores a source of variation in cell counts.

Let's name the cell counts in the 2×2 table as follows.

	Treatment	Control
Observed Wins	C_{WT}	C_{WC}
Observed Losses	C_{LT}	C_{LC}

Chi-Squared Test of Independence

To compute the X^2 statistic for the test of independence, we must add together four terms, one for each cell. Each term is an observed value minus an expected value, squared and then divided by the expected value. The expected count for a given cell can be computed by multiplying the row total and the column total and then dividing by the overall total.

Since we have specified in advance that equally many participants were assigned to each condition, we know that the column totals will be equal—half of the participants will be in the treatment condition and half in the control condition. Therefore, both column totals are equal to half of the overall total. The expected count for a given cell can be rewritten as follows.

$$\frac{RowTotal \times ColumnTotal}{OverallTotal} = \frac{RowTotal \times OverallTotal/2}{OverallTotal} = \frac{RowTotal}{2}$$

The expected counts for the four cells can be expressed as follows.

	Treatment	Control
Expected Wins	$E_{WT} = \frac{C_{WT} + C_{WC}}{2}$	$E_{WC} = \frac{C_{WT} + C_{WC}}{2}$
Expected Losses	$E_{LT} = \frac{C_{LT} + C_{LC}}{2}$	$E_{LC} = \frac{C_{LT} + C_{LC}}{2}$

And the chi-squared statistic can be computed as follows.

$$\mathrm{X}^2 = \frac{(C_{WT} - E_{WT})^2}{E_{WT}} + \frac{(C_{WC} - E_{WC})^2}{E_{WC}} + \frac{(C_{LT} - E_{LT})^2}{E_{LT}} + \frac{(C_{LC} - E_{LC})^2}{E_{LC}}$$

Chi-Squared Test of Goodness of Fit

If we were to try to do a chi-squared test of goodness of fit on the distribution of wins, our contingency table would be the same as the first row of the contingency table above.

	Treatment	Control
Observed Wins	C_{WT}	C_{WC}

If the treatment had no effect on the outcome, then we would expect equally many wins in the treatment and control conditions, so the expected win count for both cells would be half of the row total.

	Treatment	Control
Expected Wins	$E_{WT} = \frac{C_{WT} + C_{WC}}{2}$	$E_{WC} = \frac{C_{WT} + C_{WC}}{2}$

Notice that the expected values for this test would be exactly the same as the expected values for the first row of the table in the test of independence.

The chi-squared statistic for the goodness of fit test would be computed as follows. I've named it X^2_{gw} to distinguish it from the chi-squared statistic for the test of independence.

The g stands for goodness of fit and the w indicates that this test is being done on the distribution of wins.

$$\mathrm{X}^2_{gw} = \frac{(C_{WT} - E_{WT})^2}{E_{WT}} + \frac{(C_{WC} - E_{WC})^2}{E_{WC}}$$

This X^2_{gw} statistic is equal to the first two terms of the full X^2 statistic.

If we were to perform a goodness of fit test on the distribution of losses instead of wins, the result would be equal to the *last* two terms of the full X^2 statistic.

$$\mathrm{X}^2_{gl} = \frac{(C_{LT} - E_{LT})^2}{E_{LT}} + \frac{(C_{LC} - E_{LC})^2}{E_{LC}}$$

Therefore, the chi-squared statistics from the two goodness of fit tests that might *seem* appropriate for this design actually add up to the chi-squared statistic from the truly appropriate test of independence.

$$\mathrm{X}^2 = \mathrm{X}^2_{gw} + \mathrm{X}^2_{gl}$$