# Chapter 26 Summary
## *Comparing Counts*

*What have we learned?*

- We've learned how to test hypotheses about categorical variables.
- All three methods we examined look at counts of data in categories and rely on chi-square models.
    - Goodness-of-fit tests compare the observed distribution of a single categorical variable to an expected distribution based on theory or model.
    - Tests of homogeneity compare the distribution of several groups for the same categorical variable.
    - Tests of independence examine counts from a single group for evidence of an association between two categorical variables.
- Mechanically, these tests are almost identical.
- While the tests appear to be one-sided, conceptually they are many-sided, because there are many ways that the data can deviate significantly from what we hypothesize.
- When we reject the null hypothesis, we know to examine standardized residuals to better understand the patterns in the data.

Goodness-of-Fit

- A test of whether the distribution of counts in one categorical variable matches the distribution predicted by a model is called a goodness-of-fit test.
- As usual, there are assumptions and conditions to consider…

Assumptions and Conditions

- Counted Data Condition: Check that the data are *counts* for the categories of a categorical variable.
- Independence Assumption:
    - Randomization Condition: The individuals who have been counted and whose counts are available for analysis should be a random sample from some population.
- Sample Size Assumption: We must have enough data for the methods to work.
    - Expected Cell Frequency Condition: We should expect to see at least 5 individuals in each cell.
        - This is similar to the condition that *np* and *nq* be at least 10 when we tested proportions.

Calculations

- Since we want to examine how well the observed data reflect what would be expected, it is natural to look at the *differences* between the observed and expected counts (*Obs – Exp*).
- These differences are actually residuals, so we know that adding all of the differences will result in a sum of 0. That's not very helpful.
- We'll handle the residuals as we did in regression, by squaring them.
- To get an idea of the *relative* sizes of the differences, we will divide these squared quantities by the expected values.

Calculations (cont.)

- The test statistic, called the chi-square (or chi-squared) statistic, is found by adding up the sum of the squares of the deviations between the observed and expected counts divided by the expected counts: $\chi^2 = \sum_{\text{all cells}} \frac{(Obs - Exp)^2}{Exp}$
- The chi-square models are actually a family of distributions indexed by degrees of freedom (much like the *t*-distribution).
- The number of degrees of freedom for a goodness-of-fit test is $n - 1$, where $n$ is the number of categories.

One-Sided or Two-Sided?

- The chi-square statistic is used only for testing hypotheses, not for constructing confidence intervals.
- If the observed counts don't match the expected, the statistic will be large—it can't be "too small."
- So the chi-square test is always one-sided.
  - If the calculated value is large enough, we'll reject the null hypothesis.
- The mechanics may work like a one-sided test, but the interpretation of a chi-square test is in some ways *many*-sided.
- There are many ways the null hypothesis could be wrong.
- There's no direction to the rejection of the null model—all we know is that it doesn't fit.

The Chi-Square Calculation

1. Find the expected values:
   - Every model gives a hypothesized proportion for each cell.
   - The expected value is the product of the total number of observations times this proportion.
2. Compute the residuals: Once you have expected values for each cell, find the residuals, $Observed - Expected$.
3. Square the residuals.
4. Compute the components. Now find the components $\frac{(Observed - Expected)^2}{Expected}$ for each cell.
5. Find the sum of the components (that's the chi-square statistic).
6. Find the degrees of freedom. It's equal to the number of cells minus one.
7. Test the hypothesis.
   - Use your chi-square statistic to find the P-value. (Remember, you'll always have a one-sided test.)
   - Large chi-square values mean lots of deviation from the hypothesized model, so they give small P-values.

But I Believe the Model…

- Goodness-of-fit tests are likely to be performed by people who have a theory of what the proportions *should* be, and who believe their theory to be true.
- Unfortunately, the only *null* hypothesis available for a goodness-of-fit test is that the theory is true.
- As we know, the hypothesis testing procedure allows us only to reject or fail to reject the null.

But I Believe the Model… (cont.)

- We can never confirm that a theory is in fact true.
- At best, we can point out only that the data are consistent with the proposed theory.
    - Remember, it's that idea of "not guilty" versus "innocent."

Comparing Observed Distributions

- A test comparing the distribution of counts for two or more groups on the same categorical variable is called a chi-square test of homogeneity.
- A test of homogeneity is actually the generalization of the two-proportion $z$-test.
- The statistic that we calculate for this test is *identical* to the chi-square statistic for goodness-of-fit.
- In this test, however, we ask whether choices have changed (i.e., there is no model).
- The expected counts are found directly from the data and we have different degrees of freedom.

Assumptions and Conditions

- The assumptions and conditions are the same as for the chi-square goodness-of-fit test:
    - Counted Data Condition: The data must be counts.
    - Randomization Condition: As long as we don't want to generalize, we don't have to check this condition.
    - Expected Cell Frequency Condition: The expected count in each cell must be at least 5.

Calculations

- To find the expected counts, we multiply the row total by the column total and divide by the grand total.
- We calculated the chi-square statistic as we did in the goodness-of-fit test:

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp}$$

- In this situation we have $(R - 1)(C - 1)$ degrees of freedom, where $R$ is the number of rows and $C$ is the number of columns.
    - We'll need the degrees of freedom to find a P-value for the chi-square statistic.

Examining the Residuals

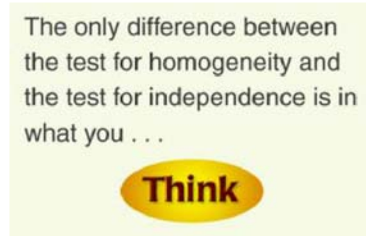
- When we reject the null hypothesis, it's always a good idea to examine residuals.
- For chi-square tests, we want to work with standardized residuals, since we want to compare residuals for cells that may have very different counts.
- To standardize a cell's residual, we just divide by the square root of its expected value:

$$c = \frac{(Obs - Exp)}{\sqrt{Exp}}$$

- These standardized residuals are just the square roots of the components we calculated for each cell, with the + or the – sign indicating whether we observed more cases than we expected, or fewer.
- The standardized residuals give us a chance to think about the underlying patterns and to consider the ways in which the distribution might not match what we hypothesized to be true.

Independence

- Contingency tables categorize counts on two (or more) variables so that we can see whether the distribution of counts on one variable is contingent on the other.
- A test of whether the two categorical variables are independent examines the distribution of counts for one group of individuals classified according to both variables in a contingency table.
- A chi-square test of independence uses the same calculation as a test of homogeneity.




Assumptions and Conditions

- We still need counts and enough data so that the expected values are at least 5 in each cell.
- If we're interested in the independence of variables, we usually want to generalize from the data to some population.
    - In that case, we'll need to check that the data are a representative random sample from that population.

Examine the Residuals

- It helps to examine the standardized residuals, just like we did for tests of homogeneity.

Chi-Square and Causation

- Chi-square tests are common, and tests for independence are especially widespread.
- We need to remember that a small P-value is *not* proof of causation.
    - Since the chi-square test for independence treats the two variables symmetrically, we cannot differentiate the direction of any possible causation even if it existed.
    - And, there's never any way to eliminate the possibility that a lurking variable is responsible for the lack of independence.
- In some ways, a failure of independence between two categorical variables is less impressive than a strong, consistent, linear association between quantitative variables.
- Two categorical variables can fail the test of independence in many ways.
    - Examining the standardized residuals can help you think about the underlying patterns.

What Can Go Wrong?

- Don't use chi-square methods unless you have counts.
    - Just because numbers are in a two-way table doesn't make them suitable for chi-square analysis.
- Beware large samples.
    - With a sufficiently large sample size, a chi-square test can always reject the null hypothesis.
- Don't say that one variable "depends" on the other just because they're not independent.
    - Association is not causation.