Agenda

1. Chi-Squared test for independence

Warmup Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.

1. The chi-square distribution, just like the normal distribution, has two parameters, mean and standard deviation.

2. The chi-square distribution is always right skewed, regardless of the value of the degrees of freedom parameter.

3. The chi-square statistic is always positive.

4. As the degrees of freedom increases, the shape of the chi-square distribution becomes more skewed.

5. As the degrees of freedom increases, the mean of the chi-square distribution increases.

6. When finding the p-value of a chi-square test, we always shade the tail areas in both tails.

7. As the degrees of freedom increases, the variability of the chi-square distribution decreases.

Independence Last time, we saw how to compare a sample from a categorical variable with multiple outcomes to a known distribution. This information could be summarized in a *one-way table*, and we had three methods of constructing the sampling distribution for the test statistic X^2:

1. Simulation
2. Multinomial Distribution
3. χ^2-approximation

A *two-way table* allows us to compare the relationship between two categorical variables, with at least two outcomes each. Here, one variable is broken down in terms of another, and we will again have three methods:

1. Simulation: Randomization Test (recall the mites and wilt disease, or the yawning experiment)

2. Hypergeometric Distribution: We won't discuss this, but it is a well-known discrete probability distribution. See the `R` command `dhyper` or Fisher's Exact Test for more information.

3. χ^2-approximation: this works basically the same as before. We compute the test statistic

$$X^2 = \sum_{i=1}^{k} \sum_{j=1}^{\ell} Z_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \left(\frac{observed_{ij} - expected_{ij}}{\sqrt{expected_{ij}}} \right)^2 = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{(observed_{ij} - expected_{ij})^2}{expected_{ij}},$$

   where i iterates over the k possible values of one variable, and j iterates over the ℓ possible values of another variable. X^2 will follow a χ^2 distribution, where the number of degrees of freedom is equal to $(k-1) \cdot (\ell - 1)$.

This type of analysis will allow us to evaluate the possibility that the two categorical variables are independent of one another.

In-Class Exercise, OI, 3.41 Offshore drilling, Part III The table below summarizes a data set we first encountered in Exercise 3.29 that examines the responses of a random sample of college graduates and non-graduates on the topic of oil drilling. Complete a chi-square test for these data to check whether there is a statistically significant difference in responses from college graduate and non-graduates.

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

1. State the null and alternative hypotheses.

2. Find the expected counts for all six cells under the null.

3. Compute the X^2 test statistic and carry out the hypothesis test.

In-Class Exercise, OI, 3.43 A 2011 survey asked 806 randomly sampled adult Facebook users about their Facebook privacy settings. One of the questions on the survey was, "Do you know how to adjust your Facebook privacy settings to control what people can and cannot see?"" The responses are cross-tabulated based on gender.

	Gender	
	Male	Female
Yes	288	378
No	61	62
Not sure	10	7
Total	359	447

1. State appropriate hypotheses to test for independence of gender and whether or not Facebook users know how to adjust their privacy settings.

2. Verify any necessary conditions for the test and determine whether or not a chi-square test can be completed.