Chi-Square Test of Homogeneity: The Good Samaritan

Learning outcomes

- Understand the differences between Chi-Square tests of independence and Chi-Square tests of homogeneity
- Construct the null and alternative hypotheses for a Chi-Square test of homogeneity in words and using appropriate statistical symbols.
- Describe and perform a simulation-based Chi-Square test of homogeneity.
  - Identify the steps required to simulate what could have happened if H_0 was true
  - Describe how these steps could be carried out with cards
- Describe the conditions required to use a χ^2 distribution to obtain a p-value for a Chi-Square test of homogeneity
- Interpret and evaluate the p-value obtained in the context of a Chi-Square test of homogeneity

The Good Samaritan

Researchers at the Princeton University wanted to investigate influences on behavior (Darley, 1973). The researchers randomly selected 67 students from the Princeton Theological Seminary to participate in a study. Only 47 students chose to participate in the study, and the data below includes 40 of those students (7 students were removed from the study for various reasons). As all participants were theology majors planning a career as a preacher, the expectation was that all would have a similar disposition when it comes to helping behavior. Each student was then shown a 5-minute presentation on the Good Samaritan, a parable in the Bible which emphasizes the importance of helping others. After the presentation, the students were told they needed to give a talk on the Good Samaritan parable at a building across campus. Half the students were told they were late for the presentation; the other half told they could take their time getting across campus (the condition was randomly assigned). On the way between buildings, an actor pretending to be a homeless person in distress asked the student for help. The researchers recorded whether the student helped the actor or not.

The researchers were interested in investigating whether these data provide evidence of a difference in how often someone will help help people in need.

A preview of the dataset is shown below:

```
good_sam
```

```
## # A tibble: 40 x 2
##    Condition Behavior
##    <chr>     <chr>
##  1 Hurry     Help
##  2 Hurry     Help
##  3 Hurry     No help
##  4 Hurry     No help
##  5 Hurry     No help
##  6 Hurry     No help
##  7 Hurry     No help
##  8 Hurry     No help
##  9 Hurry     No help
## 10 Hurry     No help
## # ... with 30 more rows
```

Vocabulary review

1. Based on the data preview, what is the name of the explanatory variable? What are its categories?

2. Based on the data preview, what is the response variable? What are its categories?

3. Fill in the blanks with one answer from each set of parentheses:

This is an ________________________ (experiment / observational study) because _____________________ (hurry or no hurry / help or no help) __________ (was / was not) randomly __________________ (assigned / selected).

4. Put an X in the box that represents the appropriate scope of inference for this study.

	Random / Representative Sample	No Random / Representative Sample
No Random Assignment of Explanatory Variable		
Random Assignment of Explanatory Variable		

Chi-Square Test of Homogeneity vs. Independence

Yesterday, we explored a Chi-Square Test of Independence. We compared what we saw in the data to what we expected if the null hypothesis was true—that there was no association / relationship between the two variables.

In the case that the explanatory variable in a study was randomly assigned, we are able to perform a test that is more specific than a test of independence. We are able to test if the proportion of successes are **homogeneous** across the levels of our categorical variable.

The Chi-Square test of homogeneity is more specific because our hypotheses are about **proportions**, rather than general relationships between two variables.

Setting Up the Research Question

The research question as stated by the researchers is: Do these data provide evidence that those in a hurry will be less likely to help people in need in this situation?

In order to set up our hypotheses, we need to express this research question in terms of parameters.

For a proportion, we write the sample proportion as $\hat{p}$ and the population proportion as π. In general, we will be comparing the proportion of observational units that are labeled as a "success" in the response variable. Here, a "success" would be a person choosing to "Help" the homeless person.

5. Fill in the details regarding the two parameters of interest for this study.

π______________ — The true proportion of people who are in a hurry who ______________

π______________ —

When comparing two groups, we assume the two parameters are equal in the null hypothesis.

6. Write the null hypothesis out in words using your answers to question 5.

7. Based on the research question, fill in the appropriate alternative hypothesis.

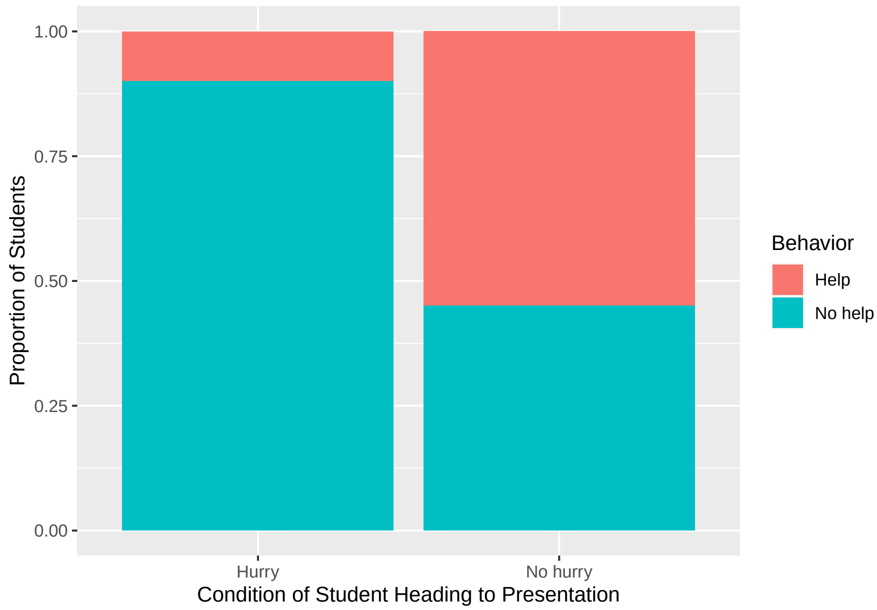
Visualize and Summarize the Data

Yesterday, we learned how to add a second categorical variable into our bar plot. We learned there are three orientations for bar plots with two categorical variables: stacked, dodged, and filled.

Generally, **dodged** barplots are used for Chi-Square tests of **independence**.

For Chi-Square tests of **homogeneity** we use **filled** bar plots instead. In a filled bar plot the proportions are on the y-axis, which connects with what we are testing in the null hypothesis—if the proportions are the same.

```
ggplot(data = good_sam,
       mapping = aes(x = `Condition`,
                     fill = Behavior)) +
  geom_bar(position = "fill") +
  labs(x = "Condition of Student Heading to Presentation",
       y = "Proportion of Students")
```



8. Based on the plot does there appear to be an association between the variables? Explain your answer.

Summarizing the Data

Similar to what we did yesterday, we can create a two-way table summarizing the number of students at each of the levels of the two variables. A two-way table of the results of this study is presented below.

Behavior	Hurry	No hurry	Total
Help	2	11	13
No help	18	9	27
Total	20	20	40

8. Using the two-way table, calculate the proportion of students who were told they were in a hurry who helped the actor.

9. Using the two-way table, calculate the proportion of students who were told they could take their time (no hurry) who helped the actor.

10. Based on these proportions, do you believe there is a relationship between being in a hurry and helping someone?

Carrying out a Chi-Square Test of Homogeneity

Similar to what we did last week with one categorical variable, we will be performing a Chi-Squared test to compare what we saw in the data to what we would have expected to see if the null hypothesis was true.

Expected Counts Under H_0

Remember from yesterday, to find the expected count for each cell in our two variable table, we use three pieces of information:

- the row total for that cell
- the column total for that cell
- the total sample size

We find the expected value of a cell using the following formula:

$$\frac{\text{row total} \times \text{column total}}{\text{total sample size}}$$

Once we have each of these, we can create a table of what frequencies we would have expected to see if H_0 was true.

11. Fill out the table of expected values below.

Behavior	Hurry	No Hurry
Help		
No Help		

Chi-Squared Statistic

Next, we compare each of our observed frequencies to what we would have expected if H_0 was true. We compare how far "off" our observed frequencies are from what was expected in a very specific way, using the following calculation:

$$\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

12. Using the formula above, calculate how far "off" each of the cells in our observed table is from what was expected under the null hypothesis.

13. Adding all of these differences together to obtain our observed X^2 statistic.

$X^2 =$

Sampling Distribution of X^2

In order for us to calculate our p-value—the probability of observing an X^2 statistic as or more extreme than what we got, if the null was true—we need a distribution of X^2 statistics that could have happened if H_0 was true.

Like we saw yesterday, there are two ways we can obtain this **sampling distribution**:

- using a χ^2 distribution
- using computer simulation

Let's see which option is a better choice.

14. Based on your table of expected counts, which method should you use for approximating the sampling distribution? Why?

Theory-based Null Distribution

If you chose a χ^2 distribution this is where you should be! Similar to what we did yesterday, we will use a χ^2 distribution with $(r-1)\times(c-1)$ degrees of freedom, where r is the number of rows and c is the number of columns in the table.

15. How many degrees of freedom would be use for our χ^2 distribution?

Using a χ^2 Distribution to Find the p-value

To find a p-value for a given X^2 statistic, we use the `pchisq()` function.

16. Using the values you calculated before, fill in the code below:

```
pchisq(_________, df = _____, lower.tail = FALSE)
```

Running the code you just wrote in `R` gave me a p-value of approximately 0.006934.

17. Based on this p-value what conclusion would you reach regarding the null hypothesis? *Hint:* Go back and see what you wrote for your null and alternative hypotheses in #11 and #12!

18. Write a paragraph summarizing the results of the study as if writing a press release. Be sure to describe:

- Summary statistic and interpretation
- P-value and interpretation
- Conclusion (written to answer the research question)
- Scope of inference

Simulated / Permuted Null Distribution

If you chose a simulation-based method this is where you should be! If there are not at least 5 expected counts in each cell of our table, we cannot use a χ^2 distribution to approximate what the true sampling distribution looks like. Instead, we need to use computer simulation to obtain our p-value.

Like all of the previous times, we can think about what the computer is doing using cards. Keep in mind, we are assuming that someone is equally likely to provide help to someone in need regardless of whether they are in a hurry or not.

To carry out **one** simulation we need to do the following steps:

Step 1: Write the response (____________________) and the explanatory variable (________________________) on ________ cards.

Step 2: Assume the null hypothesis is true and:

Step 3: Create a new dataset that could have happened if H_0 was true by:

Step 4: Create a table of the simulated counts for the shuffled cards.

Keep in mind the column and row totals will stay the **same** (there were only 20 people in a hurry and only 13 people who helped the stranger). However, the **cell values** will **change** (we won't always have 2 people who helped and were in a hurry).

Step 5: Calculate the X^2 statistic for the simulation.

Because the row and column totals stay the same, the table of expected counts will be the same for **every** simulation!

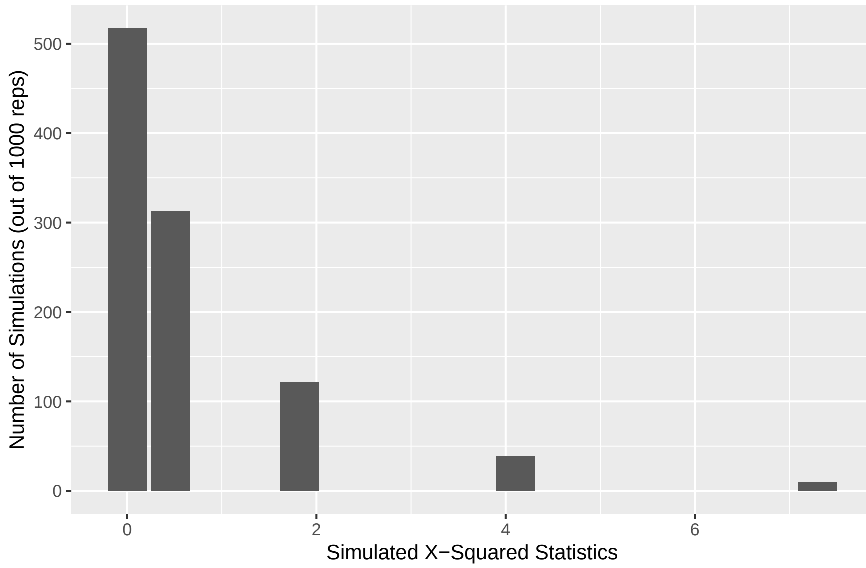
Step 6: Plot the simulated X^2 statistic on the distribution

Here's a table of one permuted / simulated dataset:

```
##  Behavior Hurry No hurry Total
##      Help     5        8    13
##   No help    15       12    27
##     Total    20       20    40
```

15. Calculate the X^2 statistic for this simulation.

Alright, after carrying out this process, I obtained the following distribution.



16. Draw a line where the observed X^2 statistic falls on this distribution.

17. Estimate the p-value for testing if the proportion of students who helped is the same between the two groups (hurry / no hurry).

18. Based on your p-value, what would you conclude for your null and alternative hypotheses? (Look back at what you said for #11 and #12)!

19. Write a paragraph summarizing the results of the study as if writing a press release. Be sure to describe:

- Summary statistic and interpretation
- P-value and interpretation
- Conclusion (written to answer the research question)
- Scope of inference