STAT 216 Coursepack

Spring 2023
Montana State University

Melinda Yager
Jade Schmidt
Stacey Hancock

This resource was developed by Melinda Yager, Jade Schmidt, and Stacey Hancock in 2021 to accompany the online textbook: Hancock, S., Carnegie, N., Meyer, E., Schmidt, J., and Yager, M. (2021). *Montana State Introductory Statistics with R*. Montana State University. https://mtstateintrostats.github.io/IntroStatTextbook/.

Contents

Preface

This coursepack accompanies the textbook for STAT 216: Montana State Introductory Statistics with R, which can be found at https://mtstateintrostats.github.io/IntroStatTextbook/. The syllabus for the course (including the course calendar), data sets, and links to D2L Brightspace, Gradescope, and the MSU RStudio server can be found on the course webpage: https://math.montana.edu/courses/s216/. Videos assigned in the course calendar and other notes and review materials are linked in D2L.

Each of the activities in this workbook is designed to target specific learning outcomes of the course, giving you practice with important statistical concepts in a group setting with instructor guidance. In addition to the in-class activities for the course, the coursepack includes reading guides to aid in taking notes while you complete the required readings and videos. Bring this workbook with you to class each class period, and take notes in the workbook as you would your own notes. A well-written completed workbook will provide an optimal study guide for exams!

The activities and labs in this coursepack will be completed during class time. Parts of each lab will be turned in on Gradescope. To aid in your understanding, read through the introduction for each activity before attending class each day.

STAT 216 is a 3-credit in-person course. In our experience, it takes six to nine hours per week outside of class to achieve a good grade in this class. By "good" we mean at least a C because a grade of D or below does not count toward fulfilling degree requirements. Many of you set your goals higher than just getting a C, and we fully support that. You need roughly nine hours per week to review past activities, read feedback on previous assignments, complete current assignments, and prepare for the next day's class. The following will give you an idea of what a typical week in the life of a STAT 216 student looks like.

- *Prior to class meeting*:
  - Read assigned sections of the textbook, using the provided reading guides to take notes on the material.
  - Watch assigned videos on that week's content, pausing to take notes and answer video quiz questions.
  - Read through the introduction to the day's in-class activity.
  - Read through the week's homework assignment and note any questions you may have on the content.
- *During class meeting*:
  - Work through the in-class activity or weekly lab with your classmates and instructor, taking detailed notes on your answers to each question in the activity.
- *After class meeting*:
  - Complete any parts of the activity you did not complete in class.
  - Review the activity solutions in the Math and Stat Center, and take notes on key points.
  - Finish watching any remaining assigned videos or readings for the week.
  - Complete the week's homework assignment.

WEEK 1

Inference for a Single Categorical Variable: Theory-based Methods + Errors and Power

1.1 Week 7 Reading Guide: Categorical Inference

Chapter 11 (Inference with mathematical models)

Videos

- Chapter11

Reminders from previous sections

n_1= sample size of group 1

n_2 = sample size of group 2

$\overline{x}$ = sample mean

s = sample standard deviation

μ = population mean

σ = population standard deviation

General steps of a hypothesis test:

1. Frame the research question in terms of hypotheses.
2. Collect and summarize data using a test statistic.
3. Assume the null hypothesis is true, and simulate or mathematically model a null distribution for the test statistic.
4. Compare the observed test statistic to the null distribution to calculate a p-value.
5. Make a conclusion based on the p-value and write the conclusion in context.

Parameter: a value summarizing a variable(s) for a population.

Statistic: a value summarizing a variable(s) for a sample.

Hypothesis test: a process to determine how strong the evidence of an effect is. Also called a 'significance test'.

Simulation-based method: Simulate lots of samples of size n under assumption of the null hypothesis, then find the proportion of the simulations that are at least as extreme as the observed sample statistic.

Theory-based method: Develop a mathematical model for the sampling distribution of the statistic under the null hypothesis and use the model to calculate the probability of the observed sample statistic (or one more extreme) occurring.

Null hypothesis (H_0): the skeptical perspective; no difference; no change; no effect; random chance; what the researcher hopes to prove is **wrong**.

Alternative hypothesis (H_A): the new perspective; a difference/increase/decrease; an effect; not random chance; what the researcher hopes to prove is **correct**.

Null value: the value of the parameter when we assume the null hypothesis is true (labeled as $parameter_0$).

P-value: probability of seeing the observed sample data, or something more extreme, assuming the null hypothesis is true.

$\implies$ Lower the p-value the stronger the evidence AGAINST the null hypothesis and FOR the alternative hypothesis.

Significance level (α): a threshold used to determine if a p-value provides enough evidence to reject the null hypothesis or not.

Common levels of α include 0.01, 0.05, and 0.10.

Statistically significant: results are considered statistically significant if the p-value is below the significance level.

Confidence interval: a process to determine how large an effect is; a range of plausible values for the parameter. Also called 'estimation'.

Vocabulary

Central Limit Theorem:

Sampling distribution:

Normal distribution (Also known as: normal curve, normal model, Gaussian distribution):

Notation:

Standard normal distribution:

Notation:

Z-score:

Xth percentile:

68-95-99.7 rule:

Standard error of a statistic:

Standard deviation of a statistic:

Margin of error:

Notes

The two general conditions for the sampling distribution for a sample proportion (or difference in sample proportions) to be approximately normally distributed are:

1)

2)

Interpretation of a Z-score:

True or False: The more unusual observation will be the observation with the largest Z-score.

Approximately what percent of a normal distribution is in the interval

(mean − standard deviation, mean + standard deviation):

(mean − 2×(standard deviation), mean + 2×(standard deviation)):

(mean − 3×(standard deviation), mean + 3×(standard deviation)):

Given a mean and standard deviation, what function in R would help us find the percent of the normal distribution above (or below) a specific value?

Given a mean and standard deviation, what function in R would help us find the value at a given percentile?

How is the standard deviation of a statistic ($SD(statistic)$) different from the standard error of a statistic ($SE(statistic)$)?

How is the standard deviation of a statistic ($SD(statistic)$) different from the standard deviation of a sample (s)?

Formulas

Z =

$SD(\hat{p}) =$

General form of a theory-based confidence interval =

General form for margin of error =

R coding

Calculating normal probabilities When using the `pnorm()` R function, you will need to enter values for the arguments `mean`, `sd`, and `q` to match the question.

```
pnorm(mean = mu, sd = sigma, q = x, lower.tail = TRUE)
```

This function will return the proportion of the N(`mu`,`sigma`) distribution which is *below* the value `x`.

Example: `pnorm(mean = 5, sd = 2, q = 3, lower.tail = TRUE)` will give us the proportion of a N(5,2) distribution which is below 3, which equals 0.159:

```
pnorm(mean = 5, sd = 2, q = 3, lower.tail = TRUE)
#> [1] 0.1586553
```

Changing to `lower.tail = FALSE` will give the proportion of the distribution which is *above* the value `x`.

```
pnorm(mean = 5, sd = 2, q = 3, lower.tail = FALSE)
#> [1] 0.8413447
```

Displaying normal probabilities When using the `normTail()` R function, you will need to enter values for the arguments `m`, `s`, and `L` (or `U`) to match the question.

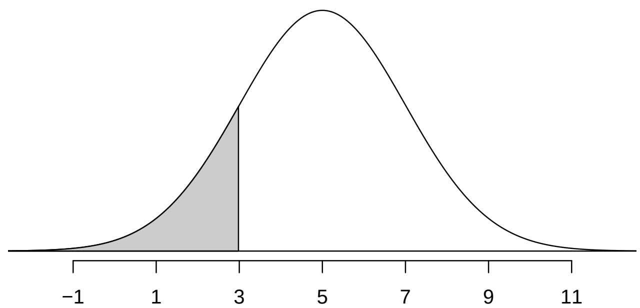
```
normTail(m = mu, s = sigma, L = x)
```

This function (in the `openintro` package) will plot a N(`mu`, `sigma`) distribution and shade the area that is below the value `x`.

Example: `normTail(m = 5, s = 2, L = 3)` creates the plot pictured below.

Changing `L` to `U` will shade the area *above* `x`.

Example: `normTail(m = 5, s = 2, U = 3)` plots a N(5,2) distribution with the area above 3 shaded.



Calculating normal percentiles When using the `qnorm()` R function, you will need to enter values for the arguments `mean`, `sd`, and `p` to match the question.

```
qnorm(mean = mu, sd = sigma, p = x, lower.tail = TRUE)
```

This function will return the value on the N(`mu`, `sigma`) distribution which has `x` area of the distribution *below* it.

Example: `qnorm(mean = 5, sd = 2, p = 0.159, lower.tail = TRUE)` will give us the value on a N(5,2) distribution which has 0.159 (15.9%) of the distribution below it, which equals 3 (from the R output above).

Changing to `lower.tail = FALSE` will give the value which has `x` area of the distribution *above* it.

We would recommend you work through each of the examples in Section 5.2.4 using R.

Section 14.3 (Theory-based inferential methods for π)

Videos

- 14.3TheoryTests
- 14.3TheoryIntervals

Vocabulary

Reminders from previous sections

n = sample size

$\hat{p}$ = sample proportion

π = population proportion

General steps of a hypothesis test:

1. Frame the research question in terms of hypotheses.
2. Collect and summarize data using a test statistic.
3. Assume the null hypothesis is true, and simulate or mathematically model a null distribution for the test statistic.
4. Compare the observed test statistic to the null distribution to calculate a p-value.
5. Make a conclusion based on the p-value and write the conclusion in context.

Parameter: a value summarizing a variable(s) for a population.

Statistic: a value summarizing a variable(s) for a sample.

Sampling distribution: plot of statistics from 1000s of samples of the same size taken from the same population.

Standard deviation of a statistic: the variability of statistics from 1000s of samples; how far, on average, each statistic is from the true value of the parameter.

Standard error of a statistic: estimated standard deviation of a statistic.

Hypothesis test: a process to determine how strong the evidence of an effect is.

Also called a 'significance test'.

Theory-based method: Develop a mathematical model for the sampling distribution of the statistic under the null hypothesis and use the model to calculate the probability of the observed sample statistic (or one more extreme) occurring.

Null hypothesis (H_0): the skeptical perspective; no difference; no change; no effect; random chance; what the researcher hopes to prove is **wrong**.

Alternative hypothesis (H_A): the new perspective; a difference/increase/decrease; an effect; not random chance; what the researcher hopes to prove is **correct**.

P-value: probability of seeing the observed sample data, or something more extreme, assuming the null hypothesis is true.

$\implies$ Lower the p-value the stronger the evidence AGAINST the null hypothesis and FOR the alternative hypothesis.

Decision: a determination of whether to 'reject' or 'fail to reject' a null hypothesis based on a p-value and a pre-set level of significance.

Significance level (α): a threshold used to determine if a p-value provides enough evidence to reject the null hypothesis or not.

Common levels of α include 0.01, 0.05, and 0.10.

Statistically significant: results are considered statistically significant if the p-value is below the significance level.

Central Limit Theorem: For large sample sizes, the sampling distribution of a sample proportion (or mean) will be approximately normal (bell-shaped and symmetric).

Confidence interval: a process to determine how large an effect is; a range of plausible values for the parameter; also called 'estimation'.

Margin of error: the value that is added to and subtracted from the sample statistic to create a confidence interval; half the width of a confidence interval.

Vocabulary

Null standard error:

Standardized statistic:

Confidence level:

Notes

Conditions for the Central Limit Theorem to apply (for the sampling distribution of $\hat{p}$ to be approximately normal)

Independence:

Checked by:

Success-failure condition:

Checked by:

How can we determine the value of $z^{\star}$ to use as the multiplier in a confidence interval?

In R, use `qnorm(mean = __, sd = __, p = __)`.

Select one answer in each set of parentheses: The higher the confidence level, the (larger/smaller) the multiplier, meaning the confidence interval will be (wider/narrower).

If the success-failure condition for the Central Limit Theorem is not met, what is the appropriate method of analysis? Select one: A. Theory-based approach B. Simulation based approach.

Formulas

$SD(\hat{p}) =$

Null standard error of the sample proportion:

$SE_0(\hat{p}) =$

Standardized statistic (in this case, standardized sample proportion):

$Z =$

Standard error of the sample proportion when we do not assume the null hypothesis is true:

$SE(\hat{p}) =$

Theory-based confidence interval for a sample proportion:

Margin of error of a confidence interval for a sample proportion:

Example: Payday loans

1. What is the parameter representing in the context of this problem? What notation would be used to represent this parameter?

2. Write the null and alternative hypotheses in words.

3. Write the null and alternative hypotheses in notation.

4. Are the conditions met to use theoretical methods to analyze these data? Show your calculations to justify your answer.

5. Calculate the null standard error of the sample proportion.

6. What is the sample statistic presented in this example? What notation would be used to represent this value?

7. Calculate the standardized sample proportion (standardized statistic).

8. How can we calculate a p-value from the normal distribution for this example?

9. What was the p-value of the test?

10. What conclusion should the researcher make?

11. Are the results in this example statistically significant? Justify your answer.

12. Calculate the standard error of the sample proportion when we do not assume the null hypothesis is true.

13. Calculate the margin of error for a 95% confidence interval for π using 1.96 as the multiplier.

14. Calculate a 95% confidence interval for π using your margin of error calculated above.

15. Interpret the 95% confidence interval provided in the textbook.

16. Does the 95% confidence interval support the same conclusion as the p-value from the hypothesis test? Justify your answer.

Chapter 12 (Errors, power, and practical importance)

Videos

- Chapter12

Reminders from previous sections

Significance level (α): a threshold used to determine if a p-value provides enough evidence to reject the null hypothesis or not.

Common levels of α include 0.01, 0.05, and 0.10.

Statistically significant: results are considered statistically significant if the p-value is below the significance level.

Vocabulary

Decision:

- If the p-value is small (less than or equal to the significance level), the decision will be to _______________ the null hypothesis.
- If the p-value is large (greater than the significance level), the decision will be to _______________ the null hypothesis.

Type 1 error:

Type 2 error:

Confirmation bias:

One-sided hypothesis tests:

Two-sided hypothesis tests:

Power:

Practical importance:

Notes

Fill in the following table with whether the decision was correct or not, and if not, what type of error was made.

	Test conclusion (based on data)	
Truth (unknown)	Reject null hyp.	Fail to reject null hyp.
H_0 is true		
H_A is true (H_0 is false)		

How are the significance level and type I error rate related?

How are the significance level and type II error rate related?

Explain the differences between a one-sided and two-sided hypothesis test.

How will the research questions differ?

How will the notation in the alternative hypothesis differ?

How does the p-value calculation differ?

How does the p-value in a two-sided test compare to the p-value in a one-sided test?

Should the default in research be a one-sided or two-sided hypothesis test? Explain why.

After collecting data, a researcher decides to change from a two-sided test to a one-sided test. Why is this a bad idea?

1. It __________________ (increases/decreases) the chance of a type I error.
2. This can result in ____________________________________.

How are power and type I error rate related?

How are power and type II error rate related?

How can we increase the power of a test?

1. ____________ (Increase/Decrease) the significance level
2. ____________ (Increase/Decrease) the sample size
3. Change from a _____ (one/two)-sided to a _____ (one/two)-sided test
4. Have a ____________ (larger/smaller) standard deviation of the statistic
5. Have the alternative parameter value ___________ (closer/farther) from the null value

Results are likely to be statistically significant (but may not be practically important) if the sample size is _______________(large/small).

Results are unlikely to be statistically significant (but may be practically important) if the sample size is _______________(large/small).

Examples:

1. In the Martian Alphabet study section 9.1 of the textbook,

   a. What was the p-value of the test?

   b. At the 5% significance level, what decision would you make?

   c. What type of error might have occurred in these data?

   d. Interpret that error in the context of the problem.

2. In the Medical Consultant study in section 10.1 of the textbook,

   a. What was the p-value of the test?

   b. At the 5% significance level, what decision would you make?

   c. What type of error might have occurred in these data?

   d. Interpret that error in the context of the problem.

3. In the Payday Loans study section 14.3 of the textbook,

   a. What was the p-value of the test?

   b. At the 5% significance level, what decision would you make?

   c. What type of error might have occurred in these data?

   d. Interpret that error in the context of the problem.

1.2 Activity 7A: Helper-Hinderer — Simulation-based Confidence Interval

1.2.1 Learning outcomes

- Use bootstrapping to find a confidence interval for a single proportion.
- Interpret a confidence interval for a single proportion.

1.2.2 Terminology review

In today's activity, we will introduce simulation-based confidence intervals for a single proportion. Some terms covered in this activity are:

- Parameter of interest
- Bootstrapping
- Confidence interval

To review these concepts, see Chapters 10 & 14 in your textbook.

1.2.3 Helper-Hinderer

In the last class, we found very strong evidence that the true proportion of infants who will choose the helper character is greater than 0.5. But what *is* the true proportion of infants who will choose the helper character? We will use this same study to estimate this parameter of interest by creating a confidence interval.

As a reminder: A study by Hamblin, Wynn, and Bloom reported in Nature (Hamblin, Wynn, and Bloom 2007) was intended to check young kids' feelings about helpful and non-helpful behavior. Non-verbal infants ages 6 to 10 months were shown short videos with different shapes either helping or hindering the climber. Researchers were hoping to assess: Are infants more likely to preferentially choose the helper toy over the hinderer toy? In the study, of the 16 infants age 6 to 10 months, 14 chose the *helper* toy and 2 chose the *hinderer* toy.

A **point estimate** (our observed statistic) provides a single plausible value for a parameter. However, a point estimate is rarely perfect; usually there is some error in the estimate. In addition to supplying a point estimate of a parameter, a next logical step would be to provide a plausible *range* of values for the parameter. This plausible range of values for the population parameter is called an **interval estimate** or **confidence interval**.

Activity intro

1. What is the value of the point estimate?

2. If we took another random sample of 16 infants, would we get the exact same point estimate? Explain why or why not.

In today's activity, we will use bootstrapping, sampling with replacement from the original sample, to find a 95% confidence interval for π, the parameter of interest.

Use statistical analysis methods to draw inferences from the data

To use the computer simulation to create a bootstrap distribution, we will need to enter the

- "sample size" (the number of observational units or cases in the sample),
- "number of successes" (the number of cases that choose the helper character),
- "number of repetitions" (the number of samples to be generated), and
- the "confidence level" (which level of confidence are we using to create the confidence interval).

3. What values should be entered for each of the following into the simulation to create the bootstrap distribution of sample proportions to find a 95% confidence interval?

- Sample size:

- Number of successes:

- Number of repetitions:

- Confidence level (as a decimal):

We will use the `one_proportion_bootstrap_CI()` function in R (in the `catstats` package) to simulate the bootstrap distribution of sample proportions and calculate a confidence interval. Using the provided R script file, fill in the values/words for each `xx` with your answers from question 3 in the one proportion bootstrap confidence interval (CI) code to create a bootstrap distribution with 1000 simulations. Then highlight and run lines 1–9.

```
one_proportion_bootstrap_CI(sample_size = xx, # Sample size
                    number_successes = xx, # Observed number of successes
                    number_repetitions = 1000, # Number of bootstrap samples to use
                    confidence_level = 0.95) # Confidence level as a decimal
```

4. Sketch the bootstrap distribution created below.

5. What is the value at the center of this bootstrap distribution? Why does this make sense?

6. Explain why the two vertical lines are at the 2.5th percentile and the 97.5th percentile.

7. Report the 95% bootstrapped confidence interval for π. Use interval notation: (lower value, upper value).

8. Interpret the 95% confidence interval in context.

Communicate the results and answer the research question

9. Is the value 0.5 (the null value) in the 95% confidence interval?

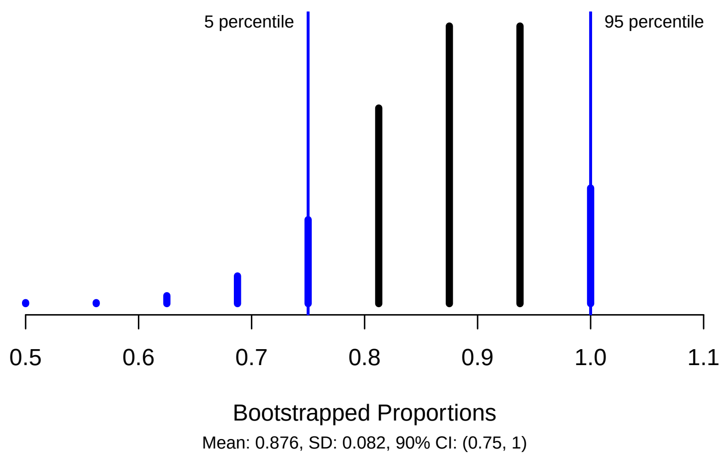
   Explain how this indicates that the p-value will provide similar evidence against the null hypothesis.

Effect of confidence level

10. Suppose instead of finding a 95% confidence interval, we found a 90% confidence interval. Would you expect the 90% confidence interval to be narrower or wider? Explain your answer.

11. The following R code produced the bootstrap distribution with 1000 simulations that follows. Circle the value that changed in the code.

```
one_proportion_bootstrap_CI(sample_size = 16, # Sample size
                            number_successes = 14, # Observed number of successes
                            number_repetitions = 1000, # Number of bootstrap samples to use
                            confidence_level = 0.90) # Confidence level as a decimal
```



12. Report both the 95% confidence interval (question 7) and the 90% confidence interval (question 12). Is the 90% confidence interval narrower or wider than the 95% confidence interval?

What does *confidence* mean?

In the interpretation of a 95% confidence interval, we say that we are 95% confident that the parameter is within the confidence interval. Why are we able to make that claim? What does it mean to say "we are 95% confident"?

For this part of the activity we will assume that the the true proportion of infants who choose the helper is 0.65. *Note: we are making assumptions about the population here. This is not based on our calculated data, but we will use this applet to better understand what happens when we take many, many samples from this believed population.*

13. Go to this website, http://www.rossmanchance.com/ISIapplets.html and choose 'Simulating Confidence Intervals'. In the input on the left-hand side of the screen enter 0.65 for π (the true value), 16 for n, and 100 for 'Number of intervals'. Click 'sample'.

a. In the graph on the bottom right, click on a green dot. Write down the confidence interval for this sample given on the graph on the left. Does this confidence interval contain the true value of 0.65?

b. Now click on a red dot. Write down the confidence interval for this sample. Does this confidence interval contain the true value of 0.65?

c. How many intervals out of 100 contain π, the true value of 0.65? *Hint*: This is given to the left of the graph of green and red intervals.

14. Click on 'sample' nine more times. Write down the 'Running Total' for the proportion of intervals that contain π.

15. Interpret the level of confidence in context of the problem. *Hint*: What proportion of samples would we expect to give a confidence interval that contains the parameter of interest?

1.2.4 Take-home messages

1. The goal in a hypothesis test is to assess the strength of evidence for an effect, while the goal in creating a confidence interval is to determine how large the effect is. A **confidence interval** is a range of *plausible* values for the parameter of interest.

2. A confidence interval is built around the point estimate or observed calculated statistic from the sample. This means that the sample statistic is always the center of the confidence interval. A confidence interval includes a measure of sample to sample variability represented by the **margin of error**.

3. In simulation-based methods (bootstrapping), a simulated distribution of possible sample statistics is created showing the possible sample-to-sample variability. Then we find the middle X percent of the distribution around the sample statistic using the percentile method to give the range of values for the confidence interval. This shows us that we are X% confident that the parameter is within this range, where X represents the level of confidence.

4. When the null value is within the confidence interval, it is a plausible value for the parameter of interest; thus, we would find a larger p-value for a hypothesis test of that null value. Conversely, if the null value is NOT within the confidence interval, we would find a small p-value for the hypothesis test and strong evidence against this null hypothesis.

5. To create one simulated sample on the bootstrap distribution for a sample proportion, label n cards with the original responses. Draw with replacement n times. Calculate and plot the resampled proportion of successes.

6. If repeat samples of the same size are selected from the population, approximately 95% of samples will create a 95% confidence interval that contains the parameter of interest.

1.2.5 Additional notes

Use this space to summarize your thoughts and take additional notes on today's activity and material covered.

1.3 Activity 7B: Handedness of Male Boxers — Theory-based Methods

1.3.1 Learning objectives

- Describe and perform a theory-based hypothesis test for a single proportion.
- Check the appropriate conditions to use a theory-based methods.
- Calculate and interpret the standardized sample proportion.
- Interpret and evaluate a p-value for a theory-based hypothesis test for a single proportion.
- Use the normal distribution to find the p-value.
- Calculate a theory-based confidence interval for a single proportion.
- Interpret a confidence interval for a single proportion.
- Use the normal distribution to find the multiplier needed for a confidence interval

1.3.2 Terminology review

In this activity, we will introduce theory-based hypothesis tests and confidence intervals for a single proportion. Some terms covered in this activity are:

- Parameter of interest
- Standardized Statistic
- Normal distribution
- p-value
- Multiplier
- Normal distribution

To review these concepts, see Chapters 11 & 14 in your textbook.

Activities 6A, 6B, and 7A covered simulation-based methods for hypothesis tests involving a single categorical variable. This activity covers theory-based methods for testing a single categorical variable.

1.3.3 Handedness of male boxers

Left-handedness is a trait that is found in about 10% of the general population. Past studies have shown that left-handed men are over-represented among professional boxers (Richardson and Gilman 2019). The fighting claim states that left-handed men have an advantage in competition. In this random sample of 500 male professional boxers, we want to see if there is an over-prevalence of left-handed fighters. In the sample of 500 male boxers, 81 were left-handed.

```
 # Read in data set
boxers <- read.csv("https://math.montana.edu/courses/s216/data/Male_boxers_sample.csv")
boxers %>% count(Stance)  # Count number in each Stance category
```

```
#>         Stance   n
#> 1  left-handed  81
#> 2 right-handed 419
```

Review of summary statistics

1. Write out the parameter of interest for this study.

2. Write out the null hypothesis in words.

3. Write out the alternative hypothesis in notation.

4. Give the value of the summary statistic (sample proportion) for this study. Use proper notation.

Theory-based methods

The sampling distribution of a single proportion — how that proportion varies from sample to sample — can be mathematically modeled using the normal distribution if certain conditions are met.

Conditions for the sampling distribution of $\hat{p}$ to follow an approximate normal distribution:

- **Independence**: The sample's observations are independent, e.g., are from a simple random sample. (*Remember*: This also must be true to use simulation methods!)
- **Success-failure condition**: We *expect* to see at least 10 successes and 10 failures in the sample, $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.

5. Verify that the independence condition is satisfied.

6. Is the success-failure condition met to model the data with the normal distribution? Show your work to support your answer.

To calculate the standardized statistic we use the general formula

$$Z = \frac{\text{point estimate} - \text{null value}}{SE_0(\text{point estimate})}.$$

For a single categorical variable the standardized sample proportion is calculated using

$$Z = \frac{\hat{p} - \pi_0}{SE_0(\hat{p})},$$

where the standard error is calculated using the null value:

$$SE_0(\hat{p}) = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

.

The standard error of the sample proportion measures the variability of possible sample proportions from the actual proportion. In other words, how far each possible sample proportion is from the actual proportion on average. For this study, the null standard error of the sample proportion is calculated using the null value, 0.1.

$$SE_0(\hat{p}) = \sqrt{\frac{0.1(1-0.1)}{500}} = 0.013$$

.

7. Interpret the null standard error of the sample proportion in context of the problem.

8. Label the standard normal distribution (figure 7.1) with the null value as the center value (below the value of zero). Label the tick marks to the right of the null value by adding 1 standard error to the null value to represent 1 standard error, 2 standard errors, and 3 standard errors from the null. Repeat this process to the left of the null value by subtracting 1 standard error for each tick mark.

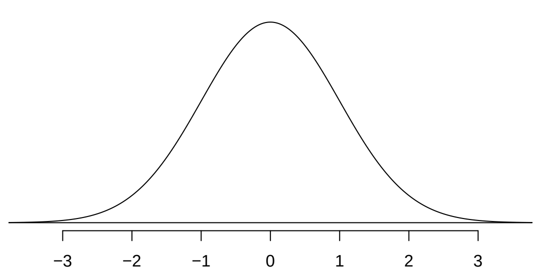


Figure 1.1: Standard Normal Distribution

9. Using the null standard error of the sample proportion, calculate the standardized sample proportion (Z). Mark this value on the standard normal distribution above.

The standardized statistic is used as a ruler to measure how far the sample statistic is from the null value. Essentially, we are converting the sample proportion into a measure of standard errors to compare to the standard normal distribution.

The standardized statistic measures the *number of standard errors the sample statistic is from the null value.*

10. Interpret the standardized sample proportion from question 9 in context of the problem.

We will use the `pnorm()` function in R to find the p-value. Use the provided R script file and enter the value of the standardized statistic calculated in question 8 at `xx` in line 7; highlight and run lines 7–9. Notice that in line 9 it says `lower.tail = FALSE`. R will calculate the p-value *greater* than the value of the standardized statistic.

Notes:

- Use `lower.tail = TRUE` when doing a left-sided test.
- Use `lower.tail = FALSE` when doing a right-sided test.
- To find a two-sided p-value, use a left-sided test for negative Z or a right-sided test for positive Z, then multiply the value found by 2 to get the p-value.

```
pnorm(xx, # Enter value of standardized statistic
      m=0, s=1, # Using the standard normal mean = 0, sd = 1
      lower.tail=FALSE) # Gives a p-value greater than the standardized statistic
```

11. Report the p-value obtained from the R output.

Theory-based confidence interval

To calculate a theory-based 95% confidence interval for π, we will first find the **standard error** of $\hat{p}$ by plugging in the value of $\hat{p}$ for π in $SD(\hat{p})$:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Note that we do not include a "0" subscript, since we are not assuming a null hypothesis.

12. Calculate the standard error of the sample proportion to find a 95% confidence interval.

To find the confidence interval, we will add and subtract the **margin of error** to the point estimate:

$$\text{point estimate} \pm \text{margin of error}$$
$$\hat{p} \pm z^* \times SE(\hat{p})$$
$$ME = z^* \times SE(\hat{p})$$

The z^* multiplier is the percentile of a standard normal distribution that corresponds to our confidence level. If our confidence level is 95%, we find the Z values that encompass the middle 95% of the standard normal distribution. If 95% of the standard normal distribution should be in the middle, that leaves 5% in the tails, or 2.5% in each tail.

13. Fill in the normal distribution shown in figure 7.2 to show how R found the z^* multiplier.

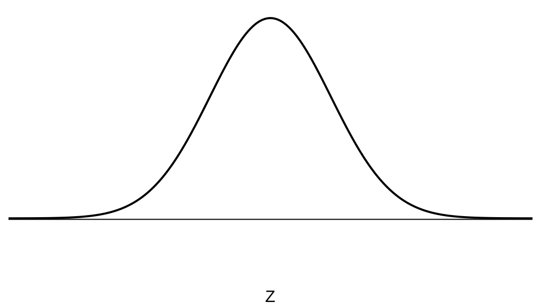


Figure 1.2: A standard normal curve.

The `qnorm()` function in R will tell us the z^* value for the desired percentile (in this case, 95% + 2.5% = 97.5% percentile). Enter the value of 0.975 for xx in the provided R script file. This will give the value of the multiplier for a 95% confidence interval.

```
qnorm(xx) # Multiplier for 95% confidence interval
```

14. Report the value of the multiplier (z^*) found from the R code needed to calculate the 95% confidence interval for the true proportion of male boxers that are left-handed?

15. Calculate the margin of error for the 95% confidence interval.

16. Calculate the 95% confidence interval for the parameter of interest.

17. Interpret the 95% confidence interval in the context of the problem.

18. Write a conclusion to the study in context of the problem.

19. Is the null value, 0.1, contained in the 95% confidence interval? Explain, based on the p-value from the earlier in activity, why you expected this to be true.

1.3.4 Take-home messages

1. Both simulation and theory-based methods can be used to find a p-value for a hypothesis test. In order to use theory-based methods we need to check that both the independence and the success-failure conditions are met.

2. The standardized statistic measures how many standard errors the statistic is from the null value. The larger the standardized statistic the more evidence there is against the null hypothesis.

3. In theory-based methods, we add and subtract a margin of error to the sample statistic. The margin of error is calculated using a multiplier that corresponds to the level of confidence times the variability (standard error) of the statistic.

4. The confidence interval calculated using theory-based methods should be similar to the confidence interval found using simulation methods provided the success-failure condition is met.

1.3.5 Additional notes

Use this space to summarize your thoughts and take additional notes on today's activity and material covered.

1.4 Week 7 Lab: Errors and Power

1.4.1 Learning outcomes

- Explain type 1 and type 2 errors in the context of a study.
- Explain the power of a test in the context of a study.
- Understand how changes in sample size, significance level, and the difference between the null value and the parameter value impact the power of a test.
- Understand how significance level impacts the probability of a type 1 error.
- Understand the relationship between the probability of a type 2 error and power.
- Be able to distinguish between practical importance and statistical significance.

1.4.2 Terminology review

In this activity, we will examine the possible errors that can be made based on the decision in a hypothesis test as well as factors influencing the power of the test. Some terms covered in this activity are:

- Significance level
- Type 1 error
- Type 2 error
- Power

To review these concepts, see Chapter 12 in the textbook.

1.4.3 ACL recovery

It is widely reported that the median recovery time for athletes who undergo surgery to repair a torn anterior cruciate ligament (ACL) is 8 months, indicating that 50% of athletes return to their sport within 8 months after an ACL surgery. Suppose a local physical therapy company hopes to advertise that their rehabilitation program can increase this percentage.

1. Write the parameter of interest (π) in words, in the context of this problem.

2. **Use proper notation to write the null and alternative hypothesis the company would need to test in order to check their advertisement claim.**

After determining hypotheses and prior to collecting data, researchers should set a **significance level** for a hypothesis test. The significance level, represented by α and most commonly 0.01, 0.05, or 0.10, is a cut-off for determining whether a p-value is small or not. The *smaller* the p-value, the *stronger* the evidence against the null hypothesis, so a p-value that is smaller than or equal to the significance level is strong enough evidence to *reject the null hypothesis*. Similarly, the *larger* the p-value, the *weaker* the evidence against the null hypothesis, so a p-value that is larger than the significance level does not provide enough evidence against the null hypothesis and the researcher would *fail to reject the null hypothesis*. Rejecting the null hypothesis or failing to reject the null hypothesis are the two **decisions** that can be made based on the data collected.

As you have already learned in this course, sample size of a study is extremely important. Often times, researchers will conduct what is called a power analysis to determine the appropriate sample size based on the goals of their research, including a desired **power** of their test. Power is the probability of correctly rejecting the null hypothesis, or the probability of the data providing strong evidence against the null hypothesis *when the null hypothesis is false.*

The remainder of this lab will be spent investigating how different factors influence the power of a test, after which you will complete a power analysis for this physical therapy company.

- Navigate to https://istats.shinyapps.io/power/. *Please note that this applet uses p_0 to represent the null value rather than π_0.*
- Use the scale under "Null Hypothesis value p_0" to change the value to your null value from question 2.
- Change the "Alternative Hypothesis" to the direction you wrote in question 2.
- Leave all boxes un-checked. Do not change the scales under "True value of p_0", "Sample size n", or "Type I Error α"

The red distribution you see is the scaled-Normal distribution representing the null distribution for this hypothesis test, if the sample size was 50 and the significance level was 0.05. This means the red distribution is showing the probability of each possible sample proportion of athletes who returned to their sport within 8 months ($\hat{p}$) if we assume the null hypothesis is true.

3. Based off this distribution and your alternative hypothesis, give one possible sample proportion which you think would lead to rejecting the null hypothesis. Explain how you decided on your value.

4. Check the box for "Show Critical Value(s) and Rejection Region(s)". You will now see a vertical line on the plot indicating the *minimum* sample proportion which would lead to reject the null hypothesis. What is this value?

5. Notice that there are some sample proportions under the red line (when the null hypothesis is true) which would lead us to reject the null hypothesis. Give the range of sample proportions which would lead to rejecting the null hypothesis when the null hypothesis is true. What is the statistical name for this mistake?

Check the "Type I Error" box under **Display**. This should verify (or correct) your answer to question 5! The area shaded in red represents the probability of making a **type 1 error** in our hypothesis test. Recall that a type 1 error is when we reject the null hypothesis even though the null hypothesis is true. To reject the null hypothesis, the p-value, which was found assuming the null hypothesis is true, must be less than or equal to the significance level. Therefore the significance level is the maximum probability of rejecting the null hypothesis when the null hypothesis is true, so the significance level IS the probability of making a type 1 error in a hypothesis test!

6. **Based on the current applet settings, what percent of the null distribution is shaded red (what is the probability of making a type 1 error)?**

Let's say this physical therapist company believes their program can get 70% of athletes back to their sport within 8 months of an ACL surgery. In the applet, set the scale under "True value of p" to 0.7.

7. Where is the blue distribution centered?

The blue distribution that appears represents what the company believes, that 0.7 (not 0.5) is the true proportion of its clients who return to their sport within 8 months of ACL surgery. This blue distribution represents the idea that the **null hypothesis is false**.

8. Consider the definition of power provided earlier in this lab. Do you believe the power of the test will be an area within the blue distribution or red distribution? How do you know? What about the probability of making a type 2 error?

- Check the "Type II Error" and "Power" boxes under **Display**. This should verify (or correct) your answers to question 8! The area shaded in blue represents the probability of making a **type 2 error** in our hypothesis test (failing to reject the null hypothesis even though the null hypothesis is false). The area shaded in green represents the power of the test. Notice that the type 1 and type 2 errors rates and the power of the test are provided above the distribution.

9. **Complete the following equation: Power + Type 2 Error Rate = . Explain why that equation makes sense.** *Hint: Consider what power and type 2 error are conditional on.*

Now let's investigate how changes in different factors influence the power of a test.

10. Using the same sample size and significance level, change the "True value of p" to see the effect on Power.

True value of p	0.60	0.65	0.70	0.75	0.80
Power					

11. What is changing about the simulated distributions pictured as you change the "True value of p"?

12. **How does increasing the distance between the null and believed true probability of success affect the power of the test?**

13. Using the same significance level, set the "True value of p" to 0.7 and change the sample size to see the effect on Power.

Sample Size	20	40	50	60	80
Power					

14. What is changing about the simulated distributions pictured as you change the sample size?

15. **How does increasing the sample size affect the power of the test?**

16. Using the same "True value of p", set the sample size to 50 and change the "Type I Error α" to see the effect on Power.

Type I Error α	0.01	0.03	0.05	0.10	0.15
Power					

17. What is changing about the simulated distributions pictured as you change the significance level?

18. **How does increasing the significance level affect the power of the test?**

19. **Complete the power analysis for this physical therapy company. The company believes 70% of their patients will return to their sport within 8 months of ACL surgery. They want to limit the probability of a type 1 error to 10% and the probability of a type 2 error to 15%. What is the minimum number of athletes the company will need to collect data from in order to meet these goals? Use the applet to answer this question, then download your image created and upload the file to Gradescope.**

20. Based on the goals outlined in question 19, which mistake below is the company more concerned about? In other words, which error were the researchers trying to minimize. Explain your answer.

- Not being able to advertise their ACL recovery program is better than average when their program really is better.
- Advertising their ACL recovery program is better even though it is not.

WEEK 2

Inference for Two Categorical Variables: Simulation-based Methods

2.1 Week 8 Reading Guide: Hypothesis Testing for a Difference in Proportions

Section 15.1 (Randomization test for $H_0 : \pi_1 - \pi_2 = 0$) and Section 15.2 (Bootstrap confidence interval for $\pi_1 - \pi_2$)

You may skip example 15.1.4, which discussed hypothesis testing for **relative risk**. We will discuss relative risk in Week 14.

Videos

- 15.1
- 15.2

Reminders from previous sections

n = sample size

$\hat{p}$ = sample proportion

π = population proportion

General steps of a hypothesis test:

1. Frame the research question in terms of hypotheses.
2. Collect and summarize data using a test statistic.
3. Assume the null hypothesis is true, and simulate or mathematically model a null distribution for the test statistic.
4. Compare the observed test (standardized) statistic to the null distribution to calculate a p-value.
5. Make a conclusion based on the p-value and write the conclusion in context.

Parameter: a value summarizing a variable(s) for a population.

Statistic: a value summarizing a variable(s) for a sample.

Sampling distribution: plot of statistics from 1000s of samples of the same size taken from the same population.

Standard deviation of a statistic: the variability of statistics from 1000s of samples; how far, on average, each statistic is from the true value of the parameter.

Standard error of a statistic: estimated standard deviation of a statistic.

Hypothesis test: a process to determine how strong the evidence of an effect is.

Also called a 'significance test'.

Simulation-based method: Simulate lots of samples of size n under assumption of the null hypothesis, then find the proportion of the simulations that are at least as extreme as the observed sample statistic.

Null hypothesis (H_0): the skeptical perspective; no difference; no change; no effect; random chance; what the researcher hopes to prove is **wrong**.

Alternative hypothesis (H_A): the new perspective; a difference/increase/decrease; an effect; not random chance; what the researcher hopes to prove is **correct**.

Null value: the value of the parameter when we assume the null hypothesis is true (labeled as $parameter_0$).

Null distribution: the simulated or modeled distribution of statistics (sampling distribution) we would expect to occur if the null hypothesis is true.

P-value: probability of seeing the observed sample data, or something more extreme, assuming the null hypothesis is true.

$\Longrightarrow$ Lower the p-value the stronger the evidence AGAINST the null hypothesis and FOR the alternative hypothesis.

Decision: a determination of whether to 'reject' or 'fail to reject' a null hypothesis based on a p-value and a pre-set level of significance.

Significance level (α): a threshold used to determine if a p-value provides enough evidence to reject the null hypothesis or not.

Common levels of α include 0.01, 0.05, and 0.10.

Statistically significant: results are considered statistically significant if the p-value is below the significance level.

Confidence interval: a process to determine how large an effect is; a range of plausible values for the parameter. Also called 'estimation'.

Margin of error: the value that is added to and subtracted from the sample statistic to create a confidence interval; half the width of a confidence interval.

Bootstrapping: the process of drawing with replacement n times from the original sample.

Bootstrapped resample: a random sample of size n from the original sample, selected with replacement.

Bootstrapped statistic: the statistic recorded from the bootstrapped resample.

Confidence level: how confident we are that the confidence interval will capture the parameter.

Vocabulary

Randomization test:

Notes

In a randomization test involving two categorical variables,

how many cards will you need and how will the cards be labeled?

Why, in the randomization test, are the cards all shuffled together and randomly dealt into two new groups?

After shuffling, how many cards are dealt into each pile?

To create a single bootstrap resample for two categorical variables,

how many cards will you need and how will the cards be labeled?

What is done with the cards once they are labeled?

Interpretations of confidence level must include:

How do you determine if the results of a hypothesis test agree with a confidence interval?

How are the confidence level and the significance level related (for a two-sided test)?

Notation

Sample size of group 1:

Sample size of group 2:

Sample proportion of group 1:

Sample proportion of group 2:

Population proportion of group 1:

Population proportion of group 2:

Example: Gender discrimination

1. What is the research question?

2. What are the observational units?

3. What type of study design was used? Justify your answer.

4. What is the appropriate scope of inference for these data?

5. What is the sample statistic presented in this example? What notation would be used to represent this value?

6. What is the parameter representing in the context of this problem? What notation would be used to represent this parameter?

7. Write the null and the alternative hypotheses in words.

8. Write the null and the alternative hypotheses in notation.

9. How could we use cards to simulate **one** sample *which assumes the null hypothesis is true*? How many blue cards — to represent what? How many red cards — to represent what? What would we do with the cards? What would you record once you have a simulated sample?

10. How can we calculate a p-value from the simulated null distribution for this example?

11. What was the p-value of the test?

12. At the 5% significance level, what decision would you make?

13. What conclusion should the researcher make?

14. Are the results in this example statistically significant? Justify your answer.

Example: Opportunity cost

1. What is the research question?

2. What are the observational units?

3. What type of study design was used? Justify your answer.

4. What is the appropriate scope of inference for these data?

5. What is the sample statistic presented in this example? What notation would be used to represent this value?

6. What is the parameter representing in the context of this problem? What notation would be used to represent this parameter?

7. Write the null and the alternative hypotheses in words.

8. Write the null and the alternative hypotheses in notation.

9. How could we use cards to simulate **one** sample *which assumes the null hypothesis is true*? How many blue cards — to represent what? How many red cards — to represent what? What would we do with the cards? What would you record once you have a simulated sample?

10. How can we calculate a p-value from the simulated null distribution for this example?

11. What was the p-value of the test?

12. Interpret the p-value in the context of the problem.

13. At the 5% significance level, what decision would you make?

14. What conclusion should the researcher make?

15. Are the results in this example statistically significant? Justify your answer.

Example: CPR and blood thinners

1. What is the research question?
2. What are the observational units?
3. What type of study design was used? Justify your answer.
4. What is the appropriate scope of inference for these data?
5. What is the sample difference in proportions presented in this example? What notation would be used to represent this value?
6. What is the parameter (using a difference in proportions) representing in the context of this problem? What notation would be used to represent this parameter?
7. Write the null and the alternative hypotheses in words.
8. Write the null and the alternative hypotheses in notation.
9. How could we use cards to simulate **one** sample *which assumes the null hypothesis is true*? How many blue cards — to represent what? How many red cards — to represent what? What would we do with the cards? What would you record once you have a simulated sample?
10. How can we calculate a p-value from the simulated null distribution for this example?
11. What was the p-value of the test?
12. Interpret the p-value in the context of the problem.
13. At the 5% significance level, what decision would you make?

14. What conclusion should the researcher make?

15. Are the results in this example statistically significant? Justify your answer.

16. How could we use cards to simulate **one** bootstrap resample? How many blue cards — to represent what? How many red cards — to represent what? What would we do with the cards? What would you record once you have a simulated sample?

17. How can we calculate a 90% confidence interval from the bootstrap distribution for this example?

18. What was the 90% confidence interval?

19. Interpret the confidence *interval* ((-0.03, 0.28)) in the context of the problem.

20. Interpret the confidence *level* (90%) in the context of the problem.

21. Does the conclusion of the hypothesis test match the confidence interval?

2.2 Activity 8A: The Good Samaritan — Simulation-based Hypothesis Test

2.2.1 Learning outcomes

- Given a research question involving two categorical variables, construct the null and alternative hypotheses in words and using appropriate statistical symbols.
- Describe and perform a simulation-based hypothesis test for a difference in proportions.
- Interpret and evaluate a p-value for a simulation-based hypothesis test for a difference in proportions.

2.2.2 Terminology review

In today's activity, we will use simulation-based methods to analyze two categorical variables. Some terms covered in this activity are:

- Conditional proportion
- Null hypothesis
- Alternative hypothesis

To review these concepts, see Chapter 15 in your textbook.

2.2.3 The Good Samaritan

Researchers at the Princeton University wanted to investigate influences on behavior (Darley and Batson 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 results of the study are shown in the table below. Do these data provide evidence that those in a hurry will be less likely to help people in need in this situation? Use the order of subtraction hurry – no hurry.

	Hurry Condition	No Hurry Condition	Total
Helped Actor	2	11	13
Did Not Help Actor	18	9	27
Total	20	20	40

These counts can be found in R by using the `count()` function:

```
# Read data set in
good <- read.csv("https://math.montana.edu/courses/s216/data/goodsam.csv")
good %>% group_by(Condition) %>% count(Behavior)
```

```
#> # A tibble: 4 x 3
#> # Groups:   Condition [2]
#>   Condition Behavior     n
#>   <chr>     <chr>    <int>
#> 1 Hurry     Help         2
#> 2 Hurry     No help     18
#> 3 No hurry  Help        11
#> 4 No hurry  No help      9
```

Vocabulary review

1. What is the name of the explanatory variable as it is written in the R output? What are its categories?

2. What is the response variable in the R output? 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.

		Study Type	
		Randomized Experiment	Observational Study
Selection of Cases	Random Sample		
	No Random Sample		

Ask a research question

The research question as stated above 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.

Remember, we define the parameter for a single categorical variable as the true proportion of observational units that are labeled as a "success" in the response variable.

5. Write the two parameters of interest for this study.

   π_{hurry} —

   $\pi_{\text{no hurry}}$ —

When comparing two groups, we assume the two parameters are equal in the null hypothesis—there is no association between the variables.

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

7. Based on the research question, fill in the appropriate sign for the alternative hypothesis ($<$, $>$, or $\neq$):

   $H_A : \pi_{\text{hurry}} - \pi_{\text{no hurry}}$ ______________ 0

Summarize and visualize the data

8. Using the two-way table given in the introduction, calculate the conditional proportion of students in the hurry condition who helped the actor.

9. Using the two-way table given in the introduction, calculate the conditional proportion of students in the no hurry condition who helped the actor.

10. Calculate the summary statistic (difference in sample proportion) for this study. Use Hurry - No hurry as the order of subtraction.

11. What is the notation used for the value calculated in question 10?

We will now simulate a **null distribution** of sample differences in proportions. The null distribution is created under the assumption the null hypothesis is true.

12. First, let's think about how one simulation would be created on the null distribution using cards.

    How many cards would you need?

    What would be written on each card?

13. Next, we would mix the cards together and shuffle into two piles.

    How many cards would be in each pile?

    What would each pile represent?

14. Once we have one simulated sample, what would we calculate and plot on the null distribution? *Hint*: What statistic are we calculating from the data?

15. Simulate one sample using the cards provided by your instructor. Write down the value of the simulated statistic. How does the value of your group's simulated statistic compare to the other groups at your table? Are the simulated values closer to the null value of zero than the actual calculated difference in proportions?

To create the null distribution of differences in sample proportions, we will use the `two_proportion_test()` function in R (in the `catstats` package). We will need to enter the response variable name and the explanatory variable name for the formula, the data set name (identified above as `good`), the outcome for the explanatory variable that is first in subtraction, number of repetitions, the outcome for the response variable that is a success (what the numerator counts when calculating a sample proportion), and the direction of the alternative hypothesis.

The response variable name is `Behavior` and the explanatory variable name is `Condition`.

16. What inputs should be entered for each of the following to create the simulation?

- First in subtraction (What is the outcome for the explanatory variable that is used as first in the order of subtraction? `"Hurry"` or `"No hurry"`):

- Number of repetitions:

- Response value numerator (What is the outcome for the response variable that is considered a success? `"Help"` or `"No help"`):

- As extreme as (enter the value for the sample difference in proportions):

- Direction (`"greater"`, `"less"`, or `"two-sided"`):

Using the R script file for this activity, enter your answers for question 16 in place of the `xx`'s to produce the null distribution with 1000 simulations; highlight and run lines 1–18.

```
two_proportion_test(formula = Behavior~Condition, # response ~ explanatory
    data = good, # Name of data set
    first_in_subtraction = "xx", # Order of subtraction: enter the name of Group 1
    number_repetitions = 1000, # Always use a minimum of 1000 repetitions
    response_value_numerator = "xx", # Define which outcome is a success
    as_extreme_as = xx, # Calculated observed statistic (difference in sample proportions)
    direction="xx") # Alternative hypothesis direction ("greater","less","two-sided")
```

17. Sketch the null distribution created here.

18. What value is the null distribution centered around? Explain why this makes sense.

19. What is the value of the p-value? *Remember*: This is the value given at the bottom of the null distribution.

20. Interpret the p-value in context of the study.

21. How much evidence does the p-value provide against the null hypothesis? *Hint*: Refer to the guidelines given in Activity 6A.

22. Write a conclusion to the test.

23. In the next activity we will find a 99% confidence interval. Based on the conclusion, do you expect the confidence interval to contain the null value of zero? Explain your answer.

2.2.4 Take-home messages

1. When comparing two groups, we are looking at the difference between two parameters. In the null hypothesis, we assume the two parameters are equal, or that there is no difference between the two proportions.

2. We use the same guidelines for the strength of evidence as we did in Activity 6A.

3. To create one simulated sample on the null distribution for a difference in sample proportions, label $n_1 + n_2$ cards with the response variable outcomes from the original data. Mix cards together and shuffle into two new groups of sizes n_1 and n_2, representing the explanatory variable groups. Calculate and plot the difference in proportion of successes.

2.2.5 Additional notes

Use this space to summarize your thoughts and take additional notes on today's activity and material covered.

2.3 Activity 8B: The Good Samaritan (continued) — Simulation-based Confidence Interval

2.3.1 Learning outcomes

- Identify the parameter of interest for a difference in proportions.
- Create and interpret a simulation-based confidence interval for a difference in proportions.

2.3.2 Terminology review

In today's activity, we will use simulation methods to estimate the difference in two proportions. Some terms covered in this activity are:

- Parameter of interest
- Bootstrapping
- Confidence interval
- Types of errors

To review these concepts, see Chapter 15 in your textbook.

2.3.3 The Good Samaritan

In the last activity, we found a small p-value for the hypothesis test for a difference in proportions. There was very strong evidence that those in a hurry will be less likely to help people in need. In today's activity, we will estimate the difference in true proportion of people who will help others for those in the hurry condition and those not in the hurry condition by finding a confidence interval.

Researchers at the Princeton University wanted to investigate influences on behavior (Darley and Batson 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 results of the study are shown in the table below. Do these data provide evidence that those in a hurry will be less likely to help people in need in this situation? Use the order of subtraction hurry – no hurry.

	Hurry Condition	No Hurry Condition	Total
Helped Actor	2	11	13
Did Not Help Actor	18	9	27
Total	20	20	40

Vocabulary review

1. Report the point estimate for this study.

Use the provided R script file to create a segmented bar plot of those who helped others for those in the hurry condition and those in the no hurry condition. Enter the name of the explanatory variable for `explanatory` and the name of the response variable for `response` in line 10. **Make sure to title your plot**. Highlight and run lines 1–15.

```
good <- read.csv("https://math.montana.edu/courses/s216/data/goodsam.csv")
good %>%
  ggplot(aes(x = explanatory, fill = response)) +   # This specifies the variables
  geom_bar(stat = "count", position = "fill") +  # Tell it to make a stacked bar plot
  labs(title = "Title",  # Make sure to title your plot
       x = "Condition",   # Label the x axis
       y = "") +  # Remove y axis label
  scale_fill_grey()  # Make figure black and white
```

2. Sketch the segmented bar plot created here.

3. Based on the segmented bar plot, does there appear to be an association between the condition assigned and the behavior? Explain.

4. Write out the conclusion you made in Activity 8A.

Use statistical analysis methods to draw inferences from the data

5. Write the parameter of interest in words, in the context of this study. What notation should be used to represent this parameter?

We will use the `two_proportion_bootstrap_CI()` function in R (in the `catstats` package) to simulate the bootstrap distribution of differences in sample proportions and calculate a confidence interval. We will need to enter the response variable name and the explanatory variable name for the formula, the data set name (identified above as `good`), the outcome for the explanatory variable that is first in subtraction, number of repetitions, the outcome for the response variable that is a success (what the numerator counts when calculating a sample proportion), and the confidence level as a decimal.

The response variable name is `Behavior` and the explanatory variable name is `Condition`.

6. What values should be entered for each of the following into the simulation to create a 99% confidence interval?

- First in subtraction (What is the outcome for the explanatory variable that is used as first in the order of subtraction? `"Hurry"` or `"No hurry"`):

- Response value numerator (What is the outcome for the response variable that is considered a success? `"Help"` or `"No help"`):

- Number of repetitions:

- Confidence level (entered as a decimal):

Using the R script file for this activity, enter your answers for question 6 in place of the `xx`'s to produce the bootstrap distribution with 1000 simulations; highlight and run lines 20–25.

```
two_proportion_bootstrap_CI(formula = Behavior ~ Condition,
        data=good, # Name of data set
        first_in_subtraction = "xx", # Order of subtraction: enter the name of Group 1
        response_value_numerator = "xx", # Define which outcome is a success
        number_repetitions = 1000, # Always use a minimum of 1000 repetitions
        confidence_level = xx) # Enter the level of confidence as a decimal
```

7. Where is the bootstrap distribution centered? Explain why.

8. Report the bootstrap 99% confidence interval.

9. What percentile of the bootstrap distribution does the upper value of the confidence interval represent?

10. Interpret the 99% confidence interval in context of the problem.

Table 2.3: Four different possible scenarios for hypothesis test decisions.

		Test conclusion	
		Fail to reject H_0	Reject H_0
	H_0 true	Good decision	Type 1 Error
Truth	H_A true	Type 2 Error	Good decision

Types of errors

Recall from a previous activity, hypothesis tests are not flawless. In a hypothesis test, there are two competing hypotheses: the null and alternative. We make a decision about which might be true, but we may choose incorrectly.

Shown in Table 2.3, a **Type 1 Error** happens when we reject the null hypothesis when H_0 is actually true. A **Type 2 Error** happens when we fail to reject the null hypothesis when the alternative is actually true.

11. Using a significance level of 0.01 and the simulation p-value found, what statistical decision would be made in regards to the null hypothesis?

12. What potential type of error could have been made?

13. Write this error in context of the problem.

2.3.4 Take-home messages

1. To create one simulated sample on the bootstrap distribution for a difference in sample proportions, label $n_1 + n_2$ cards with the outcomes for the original responses. Keep groups separate and randomly draw with replacement n_1 times from group 1 and n_2 times from group 2. Calculate and plot the resampled difference in the proportion of successes.
2. If the null value is not contained in a 99% confidence interval, then there is evidence against the null hypothesis and the p-value is less than the significance level of 0.01.

2.3.5 Additional notes

Use this space to summarize your thoughts and take additional notes on today's activity and material covered.

2.4 Week 8 Lab: Poisonous Mushrooms

2.4.1 Learning outcomes

- Given a research question involving two categorical variables, construct the null and alternative hypotheses in words and using appropriate statistical symbols.
- Describe and perform a simulation-based hypothesis test for a difference in proportions.
- Interpret and evaluate a p-value for a simulation-based hypothesis test for a difference in proportions.
- Interpret and evaluate a confidence interval for a simulation-based confidence interval for a difference in proportions.

2.4.2 Poisonous Mushrooms

Wild mushrooms, such as chanterelles or morels, are delicious, but eating wild mushrooms carries the risk of accidental poisoning. Even a single bite of the wrong mushroom can be enough to cause fatal poisoning. An amateur mushroom hunter is interested in finding an easy rule to differentiate poisonous and edible mushrooms. They think that the mushroom's gills (the part which holds and releases spores) might be related to a mushroom's edibility. They used a data set of 8124 mushrooms and their descriptions. For each mushroom, the data set includes whether it is edible (e) or poisonous (p) and the spacing of the gills (Broad (b) or Narrow (n)). Is there evidence gill size is associated with whether a mushroom is poisonous? PLEASE NOTE: According to The Audubon Society Field Guide to North American Mushrooms, there is no simple rule for determining the edibility of a mushroom; no rule like "leaflets three, let it be'' for Poisonous Oak and Ivy.

Upload and open the R script file for Week 8 lab. Upload and import the csv file, `mushrooms`. Click on the data set name to find the name of each variable and the level of each variable in the data set.

1. What is the explanatory variable? How are the two levels of the explanatory variable written in the data set?

2. What is the response variable? How are the two levels of the response variable written in the data set?

Enter the name of the data set for datasetname in the R script file in line 8. Highlight and run lines 1–9 to get the counts for each combination of categories.

```
poisonous <- datasetname # Read data set in
poisonous %>% group_by(gill.size) %>% count(class) #finds the counts in each group
```

3. Fill in the following two-way table using the R output.

	Gill Size		
Edible	Broad	Narrow	Total
Poisonous			
Edible			
Total			

4. Write the parameter of interest for this study.

5. **Calculate the difference in proportion of mushrooms that are poisonous for broad gill mushrooms and narrow gill mushrooms. Use broad - narrow for the order of subtraction. Use appropriate notation.**

6. Write the null hypothesis for this study in notation.

7. **Using the research question, write the alternative hypothesis in words.**

Fill in the missing values/names in the R script file for the `two-proportion_test` function to create the null distribution and find the p-value for the test.

```
two_proportion_test(formula = response~explanatory, # response ~ explanatory
    data= poisonous, # Name of data set
    first_in_subtraction = "xx", # Order of subtraction: enter the name of Group 1
    number_repetitions = 1000, # Always use a minimum of 1000 repetitions
    response_value_numerator = "xx", # Define which outcome is a success
    as_extreme_as = xx, # Calculated observed statistic (difference in sample proportions)
    direction="xx") # Alternative hypothesis direction ("greater","less","two-sided")
```

8. Report the p-value for the study.

9. **Do you expect that a 90% confidence interval would contain the null value of zero? Explain your answer.**

Fill in the missing values/names in the R script file in the two_proportion_bootstrap_CI function to create a simulation 90% confidence interval. **Upload a copy of the bootstrap distribution to Gradescope.**

```
two_proportion_bootstrap_CI(formula = response~explanatory,
        data=poisonous, # Name of data set
        first_in_subtraction = "xx", # Order of subtraction: enter the name of Group 1
        response_value_numerator = "xx", # Define which outcome is a success
        number_repetitions = 1000, # Always use a minimum of 1000 repetitions
        confidence_level = xx) # Enter the level of confidence as a decimal
```

10. Report the 90% confidence interval.

11. 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
  - Statement about probability or proportion of samples
  - Statistic (summary measure and value)
  - Direction of the alternative
  - Null hypothesis (in context)
- Confidence interval and interpretation
  - How confident you are (e.g., 90%, 95%, 98%, 99%)
  - Parameter of interest
  - Calculated interval
  - Order of subtraction when comparing two groups
- Conclusion (written to answer the research question)
  - Amount of evidence
  - Parameter of interest
  - Direction of the alternative hypothesis
- Scope of inference
  - To what group of observational units do the results apply (target population or observational units similar to the sample)?
  - What type of inference is appropriate (causal or non-causal)?

Upload your group's confidence interval interpretation and conclusion to Gradescope.

Paragraph (continued):

"Average Driving Distance and Fairway Accuracy." 2008. https://www.pga.com/ and https://www.lpga.com/.

Bulmer, M. n.d. "Islands in Schools Project." https://sites.google.com/site/islandsinschoolsprojectwebsite/home.

Darley, J. M., and C. D. Batson. 1973. ""From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior." *Journal of Personality and Social Psychology* 27: 100–108.

Education Statistics, National Center for. 2018. "IPEDS." https://nces.ed.gov/ipeds/.

Group, TODAY Study. 2012. "A Clinical Trial to Maintain Glycemic Control in Youth with Type 2 Diabetes." *New England Journal of Medicine* 366: 2247–56.

Hamblin, J. K., K. Wynn, and P. Bloom. 2007. "Social Evaluation by Preverbal Infants." *Nature* 450 (6288): 557–59.

Hirschfelder, A., and P. F. Molin. 2018. "I Is for Ignoble: Stereotyping Native Americans." Retrieved from https://www.ferris.edu/HTMLS/news/jimcrow/native/homepage.htm.

"IMDb Movies Extensive Dataset." 2016. https://kaggle.com/stefanoleone992/imdb-extensive-dataset.

Keating, D., N. Ahmed, F. Nirappil, Stanley-Becker I., and L. Bernstein. 2021. "Coronavirus Infections Dropping Where People Are Vaccinated, Rising Where They Are Not, Post Analysis Finds." *Washington Post*. https://www.washingtonpost.com/health/2021/06/14/covid-cases-vaccination-rates/.

Moquin, W., and C. Van Doren. 1973. "Great Documents in American Indian History." Praeger.

O'Brien, Lynch, H. D. 2019. "Crocodylian Head Width Allometry and Phylogenetic Prediction of Body Size in Extinct Crocodyliforms." *Integrative Organismal Biology* 1.

Porath, Erez, C. 2017. "Does Rudeness Really Matter? The Effects of Rudeness on Task Performance and Helpfulness." *Academy of Management Journal* 50.

Quinn, G. E., C. H. Shin, M. G. Maguire, and R. A. Stone. 1999. "Myopia and Ambient Lighting at Night." *Nature* 399 (6732): 113–14. https://doi.org/10.1038/20094.

"Rates of Laboratory-Confimed COVID-19 Hospitalizations by Vaccination Status." 2021. CDC. https://covid.cdc.gov/covid-data-tracker/#covidnet-hospitalizations-vaccination.

Richardson, T., and R. T. Gilman. 2019. "Left-Handedness Is Associated with Greater Fighting Success in Humans." *Scientific Reports* 9 (1): 15402. https://doi.org/10.1038/s41598-019-51975-3.

Stephens, R., and O. Robertson. 2020. "Swearing as a Response to Pain: Assessing Hypoalgesic Effects of Novel "Swear" Words." *Frontiers in Psychology* 11: 643–62.

Stewart, E. H., B. Davis, B. L. Clemans-Taylor, B. Littenberg, C. A. Estrada, and R. M. Centor. 2014. "Rapid Antigen Group a Streptococcus Test to Diagnose Pharyngitis: A Systematic Review and Meta-Analysis" 9 (11). https://doi.org/10.1371/journal.pone.0111727.

Stroop, J. R. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18: 643–62.

Sulheim, S., A. Ekeland, I. Holme, and R. Bahr. 2017. "Helmet Use and Risk of Head Injuries in Alpine Skiers and Snowboarders: Changes After an Interval of One Decade" 51 (1): 44–50. https://doi.org/10.1136/bjsports-2015-095798.

"Titanic." n.d. http://www.encyclopedia-titanica.org.

"US COVID-19 Vaccine Tracker: See Your State's Progress." 2021. Mayo Clinic. https://www.mayoclinic.org/coronavirus-covid-19/vaccine-tracker.

"Welcome to the Navajo Nation Government: Official Site of the Navajo Nation." 2011.Retrieved from https://www.navajo-nsn.gov/.

Wilson, Woodruff, J. P. 2016. "Vertebral Adaptations to Large Body Size in Theropod Dinosaurs." *PLoS ONE* 11(7).