STAT 216 Activity Coursepack

Fall 2020

Contents

Preface

This coursepack accompanies the textbook for STAT 216: Introduction to Statistics at Montana State University. 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. Bring this workbook with you to class each week, and take notes in the workbook as you would your own notes. A well-written complete workbook will provide an optimal study guide for exams!

ACTIVITY 1

Handedness of Male Boxers

1.1 Learning objectives

- Identify the two possible explanations (one assuming the null hypothesis, and one assuming the alternative hypothesis) for a relationship seen in sample data
- Given a research question, construct the null and alternative hypotheses in words and using appropriate statistical symbols
- Describe and perform simulation-based hypothesis tests for a single proportion
- Interpret and evaluate a p-value
- Use bootstrapping to find a confidence interval for a single proportion

1.2 Terminology review

These are a few of the terms we will cover in today's activity.

- Parameter of Interest
- Null Hypothesis
- Alternative Hypothesis
- Null distribuiton
- p-value
- Bootstrapping
- Confidence Interval

To review these concepts see Chapter 5 in your textbook

1.3 Steps of the statistical investigation process

We will work through a six step process to complete a hypothesis test for a single proportion.

- **Ask a research question** that can be addressed by collecting data. What are the researchers trying to show.
- **Design a study and collect data**. This step involves selecting the people or objects to be studied and how to gather relevant data on them.
- **Summarize and Visualize the data**. Calculate summary statistics and create graphical plots that best represent the research question.

- **Use statistical analysis methods to draw inferences from the data**. Choose an analysis technique appropriate for the data and identify the p-value. In this study, we will focus on using randomization.
- **Communicate the results and answer the research question**. Using the p-value and confidence interval from the analysis, determine whether the data provide statistical evidence against the null hypothesis.
- **Revisit and look forward** to point out limitations of the study and suggest new studies that could be performed to build on the findings of the study

1.4 Handedness of male boxers

Left-handedness is a trait that is found in about 10% of the population. The fighting claim states that left-handed men have an advantage in competition. Past studies have shown that left-handed men are over-represented among professional fighters. In this random sample of 500 male boxers we will see if there is an over-prevalence of left-handed fighters.

1.4.1 Summary statistics review

1. What are the observational units?

2. What variable are we testing? Is it categorical or quantitative?

3. What type of plot would be appropriate to visually display the data?

4. Write out in context the statistic will we calculate to summarize the data.

1.4.2 Ask a research question

5. Identify the research question for this study.

1.4.3 Design a study and collect data

6. What is the target population for this study?

7. Did the researchers use a biased or an unbiased method of selection? Explain your answer.

1.4.4 Summarize and visualize the data

```
#> Stance
#>  left-handed right-handed        Total
#>           81          419          500
```

8. Calculate the appropriate summary statistic that represents the research question. Use appropriate notation.

1.4.5 Use statistical analysis methods to draw inferences from the data

When testing data we must first identify the null hypothesis. The null hypothesis is written about the parameter of interest, the true value of interest.

9. Write out the parameter of interest. (Hint: the true proportion of….)

10. We will assume that the true proportion of male boxers who are left handed is the same as the general population, 0.1. Using the parameter of interest in question 9, write out the null hypothesis in words.

The notation used for a true proportion is π. Since this summarizes a population, it is a parameter. When writing the null hypothesis in notation we set the parameter equal to the null value, $H_0 : \pi = \pi_0$

11. Write the null hypothesis in notation using the null value of 0.1.

The alternative hypothesis is the claim to be tested and the direction is based on the research question.

12. Based on the research question from question 5, are we testing that the parameter is greater than 0.1, less than 0.1 or different than 0.1?

13. Write out the alternative hypothesis in words.

14. Write out the alternative hypothesis in notation.

Remember that when utilizing a hypothesis test, we are evaluating two competing possibilities. For this study the **two possibilities** are either…

- The true proportion of male boxers who are left handed is 0.1 and our results just occurred by random chance or
- The true proportion of male boxers who are left handed is greater than 0.1 and our results reflect this

Notice that these two competing possibilities represent the null and alternative hypotheses.

The null distribution is created under the assumption the null hypothesis is true. In this case, we assume the true proportion of male boxers who are left handed is 0.1 so we will create 1000 different simulations of 500 boxers under this assumption.

Let's think about how to use cards to create one simulation of 500 boxers under the assumption the null hypothesis is true. Suppose blue cards represents left-handed and red cards represents right-handed.

15. How many cards total do we need? How many blue ones? How many red ones?

16. Next, we would mix the cards together and draw 1 card, write down if it's red or blue, and replace the card. How many times would we need to repeat this process to simulate our sample?

17. What would we plot on the null distribution?

We will use the computer to simulate 1000 simulated proportions of male boxers who are left handed for a sample size of 500 based on the assumption that the true proportion of male boxers who are left handed is 0.1. This is called the null distribution because it is created based on the assumption that the null hypothesis is true.

To use the computer simulation, we will need to enter the "probability of success," "sample size," "number of repetitions," "as extreme as," and the "direction" (matches the direction of the alternative hypothesis).

18. What values should be entered into the simulation?

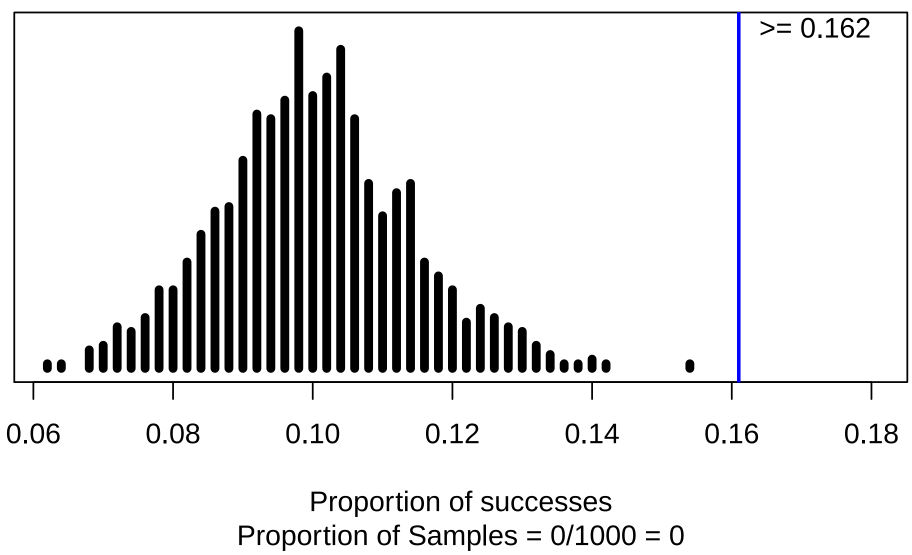
Probability of success:

Sample size:

Number of repetitions:

As extreme as:

Direction:



19. Around what value is the null distribution centered? Why does that make sense?

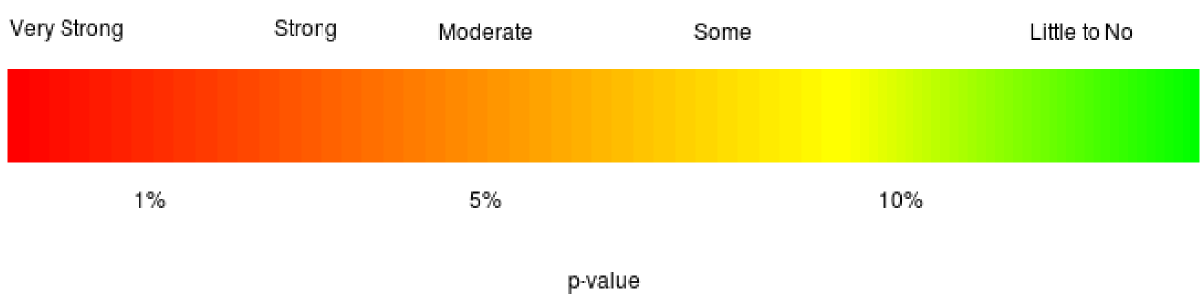
20. Where does the statistic (value from question 8) fall in the null distribution? Is it towards the center or in one of the tails?

21. Is the statistic likely to happen or unlikely to happen if the true proportion of male boxers is 0.1? Explain your answer.

22. Using the simulation, what is the proportion of samples at this summary statistic or greater, if the true proportion of male boxers is 0.1? *Hint: Look under the simulation.*

This is the p-value. The smaller the p-value the more evidence we have against the null hypothesis.

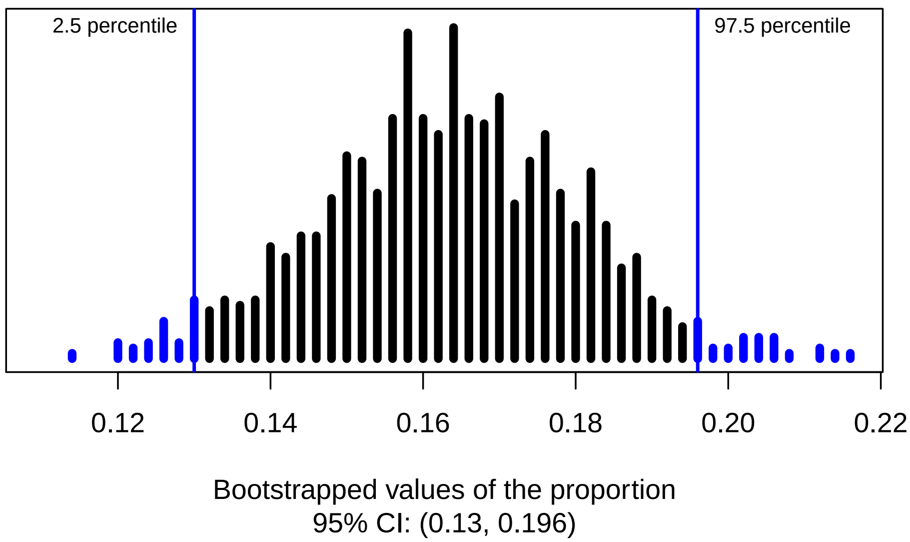
23. Using the following guidelines for the strength of evidence, how much evidence do the data provide against the null hypothesis?



A point estimate 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 a confidence interval.

We will use bootstrapping to find the 95% confidence interval.

24. What is bootstrapping?



25. Explain why the blue lines are at the 2.5th percentile and the 97.5th percentile.

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

27. What are we 95% confident is contained within this interval?

1.4.6 Communicate the results and answer the research question

When we write a conclusion we answer the research question by stating how much evidence there is for the alternative hypothesis.

28. Write a paragraph summarizing the results. Be sure to include:

- Summary statistic
- P-value
- Conclusion in context
- Confidence Interval
- Interpretation of the Confidence Interval

1.4.7 Revisit and look forward

29. Suggest a new research question that you might investigate, building on what you learned in this study.

1.5 Additional notes

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

ACTIVITY 2

Helmet Use and Head Injuries

2.1 Learning objectives

- Write out the null and alternative hypothesis for two categorical variables
- Assess the conditions to use the standard normal distributions
- Calculate the Z test statistic for a difference in proportions
- Find the p-value and assess the strength of evidence
- Create and interpret a confidence interval for the difference in proportions

2.2 Terminology review

Here are a few terms we will use in today's activity.

- Conditional proportion
- Z test
- z* multiplier
- Null Hypothesis
- Alternative Hypothesis
- Test statistic

Review Chapter 5 in your textbook for more information on these topics.

2.3 Helmet Use and Head Injuries

In "Helmet Use and Risk of Head Injuries in Alpine Skiers and Snowboarders" by Sullheim et. al., in the Journal of the American Medical Association, Vol. 295, No. 8, we can see the results from a random sample 3562 skiers and snowboarders involved in accidents.

	Head Injury	No Head Injury	Total
Wore Helmet	96	656	752
Did Not Wear Helmet	480	2330	2810
Total	576	2986	3562

Is there evidence that safety helmet use reduces the risk of head injury for skiers and snowboarders?

2.3.1 Vocabulary review

1. What is the explanatory variable?

2. What is the response variable?

3. Is this an experiment or observational study?

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		

5. What is the conditional proportion of skiers/snowboarders with a head injury that wore a helmet?

6. What is the conditional proportion of skiers/snowboarders with a head injury that did not wear a helmet?

2.3.2 Ask a research question

In this study we are looking at the relationship between two groups or two parameters (π_1 and π_2). Remember we define the parameter as the true proportion of observational units that represent the variable of interest.

7. What is the variable of interest in this study?

8. Write the two parameters of interest for this study. Let 1 = skier/snowboarder wore helmet, 2 = skier/snowboarder did not wear helmet.

   π_1 -

π_2 -

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

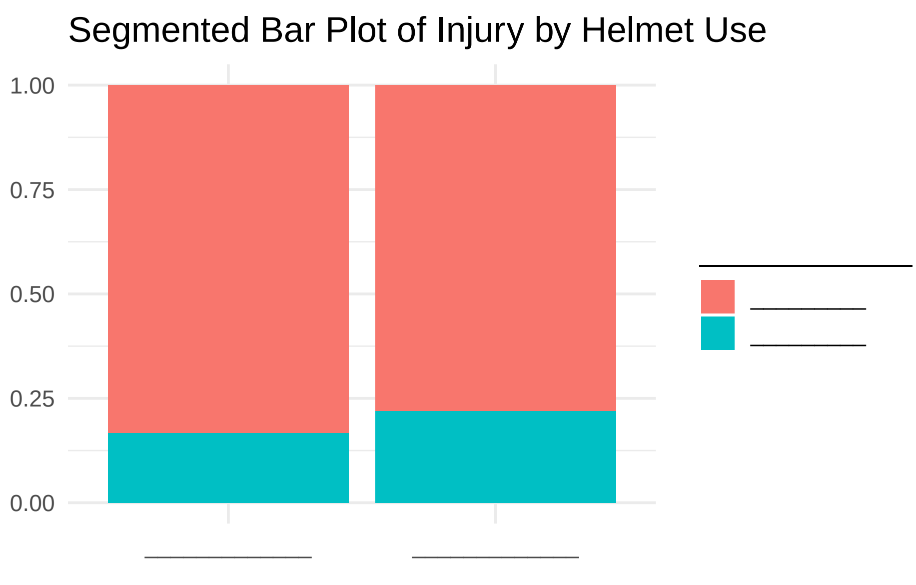
9. Write the null hypothesis out in words using your answers to question 8.

10. What is the research question?

11. Based on the research question fill in the appropriate sign for the alternative hypothesis:

$H_A: \pi_1 - \pi_2$ ______________ 0

2.3.3 Summarize and visualize the data



12. Fill in the blanks on the graph with the appropriate variables and values to plot a segmented bar plot of injury by helmet use.

13. Based on the bar plot, Does there appear to be an association between helmet use and head injury? Explain.

14. Calculate the point estimate for this study. We will use helmet use minus no helmet use as the order of subtraction.

15. What is the notation used for the value calculated in question 14?

2.3.4 Use statistical analysis methods to draw inferences from the data

To test the null hypothesis we could use simulation methods as we did with a single categorical variable. In this activity we will focus on theory-based methods. Like with a single proportion, the difference in proportions can be mathematically modeled using the normal distribution if certain conditions are met.

Conditions for the sample distribution of $\hat{p}_1 - \hat{p}_2$

- Independence: The data are independent within and between the two groups.
- Success-Failure Condition: The success-failure condition holds for each group.

16. Is the independence condition met? Explain your answer.

17. Is the success-failure condition met for each group? Explain your answer.

To calculate the test statistic we use:

$$Z = \frac{\text{point estimate} - \text{null value}}{SE}$$

where the standard error is calculated using the pooled proportion of successes.

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}_{pool}(1 - \hat{p}_{pool})(\frac{1}{n_1} + \frac{1}{n_2})}, \text{where}$$

$$\hat{p}_{pool} = \frac{\text{number of "successes"}}{\text{number of cases}} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$

18. Calculate the $SE(\hat{p}_1 - \hat{p}_2)$.

19. Calculate the test statistic.

We will use the pnorm function in R to find the p-value.

```
#> [1] 0.002118205
```

20. Report the p-value.

21. How much evidence does the p-value provide against the null hypothesis?

To find a confidence interval for the difference in proportions we will add and subtract the margin of error from the point estimate to find the two endpoints.

$$\hat{p}_1 - \hat{p}_2 \pm z^* SE(\hat{p}_1 - \hat{p}_2), \text{where}$$

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\left(\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}\right)}$$

Note that the formula changes when calculating the variability around the statistic in order to calculate a confidence interval! Here use the sample proportions for each group to calculate the standard error for the difference in proportions. The z^* multiplier is found under the normal distribution. We find the values that encompass the middle 95% of the data.

```
#> [1] 1.959964
```

22. Calculate the standard error for a difference in proportions to create a 95% confidence interval.

23. Using the multiplier of $z^* = 1.96$, calculate the 95% confidence interval for the difference in true proportion of head injuries for those that used helmets minus those who did not.

24. Interpret the confidence interval found in question 23 in context of the problem.

25. Write a paragraph summarizing the results of the study. Be sure to include:

- Summary statistic
- P-value
- Conclusion (written to answer the research question)
- Confidence interval
- Interpretation of the confidence interval
- Scope of inference

2.3.5 Types of errors

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.

		Test Conclusion	
Truth	H_0 true	good decision	Type 1 Error
	H_A true	Type 2 Error	good decision

A Type 1 Error is rejecting the null hypothesis when $H_0$is actually true. A Type 2 Error is failing to reject the null hypothesis when the alternative is actually true.

26. Using a significance level of 0.05, what decision do you make in regards to the null hypothesis?

27. What type of error could we have made?

28. Write this error in context of the problem.

2.4 Additional notes

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

ACTIVITY 3

COVID-19 and Air Pollution

3.1 Learning outcomes

- Given a research question, construct the null and alternative hypotheses in words and using appropriate statistical symbols
- Describe and perform simulation-based hypothesis for paired quantitative data
- Interpret and evaluate a p-value
- Find a confidence interval for the mean difference using bootstrapping
- Use a confidence interval to determine the conclusion of a hypothesis test

3.2 Terminology review

The following terms will be covered in this activity.

- Mean difference
- Paired data
- Independent groups
- Shifted Null Distribution

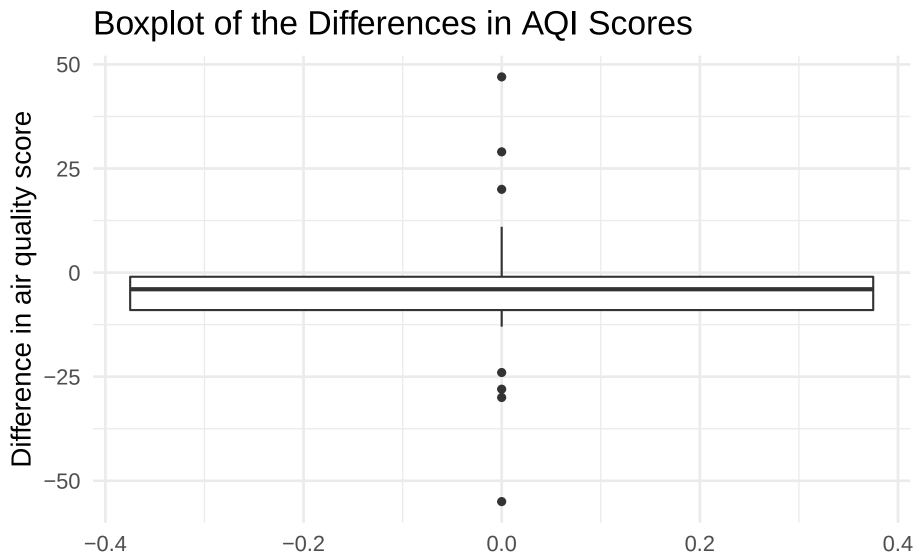
For further explanation of these topics see Section 6.2 in the textbook.

3.3 COVID-19 and air pollution

The social distancing efforts and stay-at-home directives to help combat the spread of COVID-19 have appeared to help 'flatten the curve' across the United States, albeit at a high cost to many individuals and businesses. The impact of these measures, though, goes far beyond the infection and death rates from the disease. You may have seen images comparing air quality in large international cities like Rome, Milan, Wuhan, and New Delhi such as the one pictured below which seem to indicate, perhaps unsurprisingly, that fewer people driving and factories being shut down have reduced air pollutants.

Have high population-density U.S. cities seen the same improved air quality conditions? To study this question, data was gathered from the U.S. Environmental Protection Agency (EPA) AirData website which records the ozone (O3) and fine particulate matter (PM2.5) values for cities across the U.S. These measures are used to calculate an air quality index (AQI) score for each city each day of the year. Thirty-three of the most densely populated U.S. cities were selected and the AQI score recorded for April 20, 2020 as well as the five-year median AQI score for April 20th (2015 - 2019). Note that higher AQI scores indicate worse air quality.

3.3.1 Vocabulary review



	Mean	Standard deviation	Sample Size
Current	$\bar{x}_1 = 47.394$	$s_1 = 14.107$	$n_1 = 33$
5 Year Median	$\bar{x}_2 = 51.545$	$s_2 = 17.447$	$n_2 = 33$
Differences	$\bar{x}_d =$ -4.152	$s_d = 17.096$	$n_d = 33$

1. What is the sample size?

2. Identify the variables in this study. What role do each have?

3. Why is this treated as a paired study design and not two independent samples?

4. Is this an experiment or observational study?

3.3.2 Ask a research question

5. What are the two competing possibilities to run a hypothesis test?

6. Write the null hypothesis in words.

7. What is the research question?

8. Write the alternative hypothesis in notation.

3.3.3 Summarize and visualize the data

9. Report the summary statistic for the data.

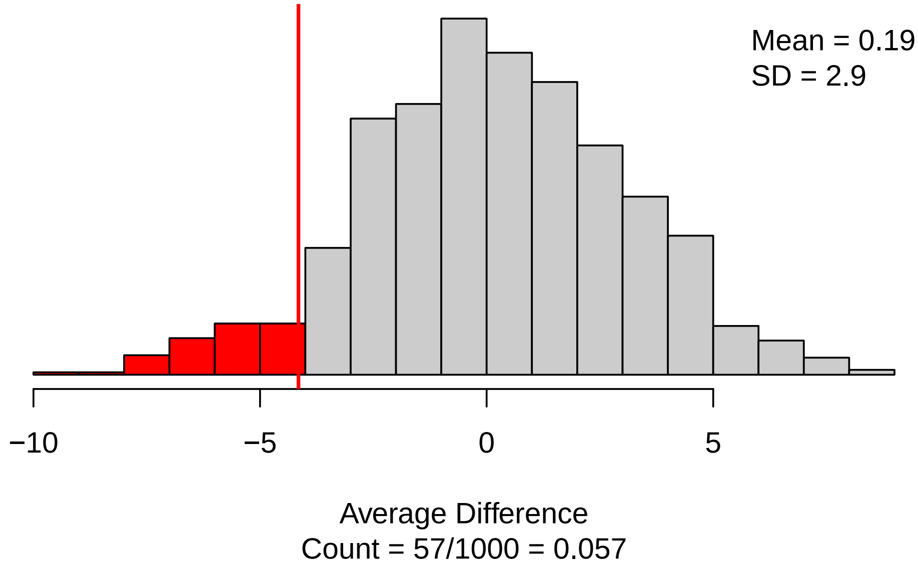
10. What notation is used for the value in question 9?

3.3.4 Use statistical inferential methods to draw inferences from the data

To simulate the null distribution we will use a bootstrapping method - sampling with replacement from the data set. Before bootstrapping we will need to shift the each data point by the difference $\mu_0 - \bar{x}$. This will ensure that the simulated null distribution will be centered at the null value.

11. Calculate the difference $\mu_0 - \bar{x}$. Will we need to shift the data up or down?

The image below gives the null distribution from one possible set of 1000 bootstrap samples:



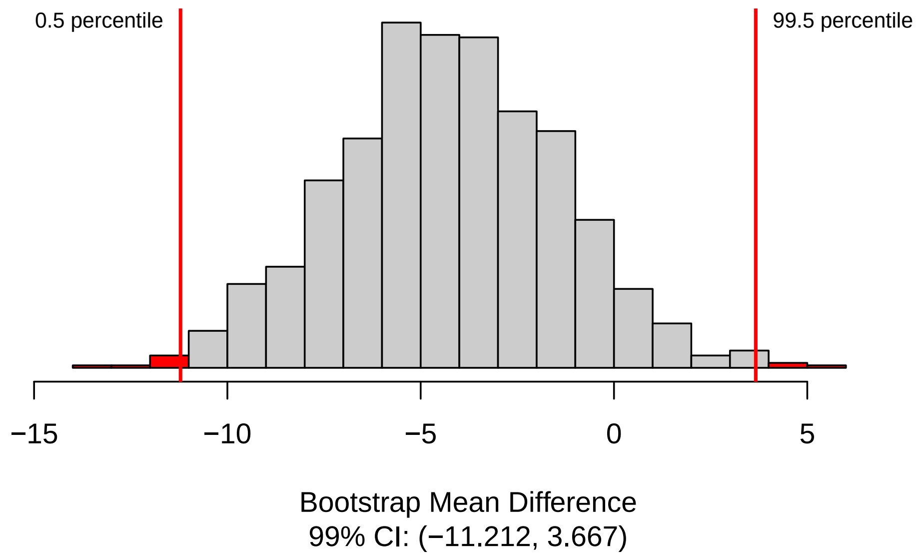
12. Explain why the null distribution is centered at zero.

13. What proportion of samples are beyond the sample mean difference in AQI Scores for current scores minus 5 year median scores?

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

15. How much evidence does this provide for improved air quality in US cities?

16. Write out the parameter of interest in context of the study.



17. Use bootstrapping to find a 99% confidence interval for the parameter of interest. Fill in the data, number of bootstrap samples, and confidence level. Report the confidence interval in interval notation.

3.3.5 Communicate the results and answer the research question.

18. Interpret the 99% confidence interval in the following figure in the context of the problem.

19. Write a paragraph summarizes the results of this study. Be sure to include:

- Summary statistic
- P-value
- Interpretation of the p-value

- Confidence Interval
- Interpretation of the confidence interval
- Conclusion in context

3.3.6 Revisit and look forward

3.4 Additional notes

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

ACTIVITY 4

Record Snowfall

4.1 Learning objectives

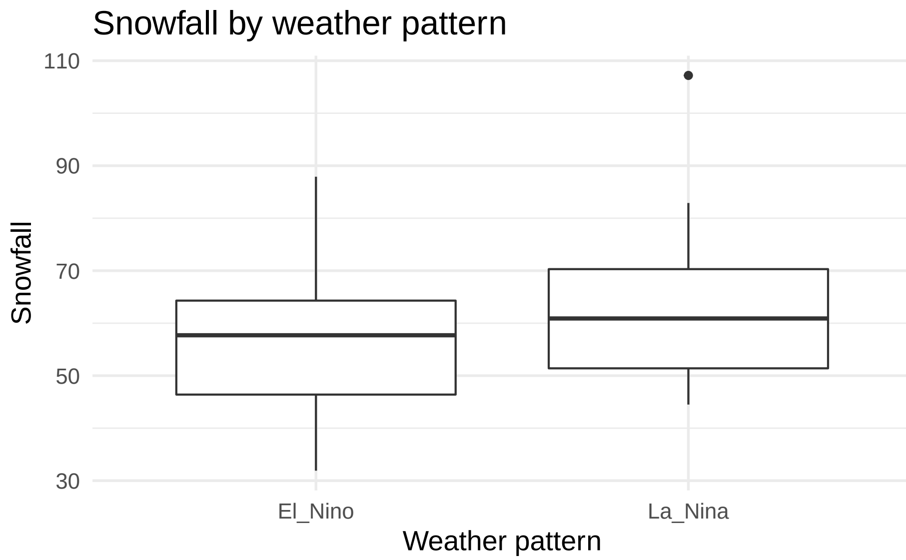
- Write out the null and alternative hypothesis for one categorical and one quantitative Variable
- Calculate and carry-out simulation based hypothesis test for a difference in means
- Interpret and evaluate a p-value
- Find a bootstap confidence interval for the difference in means
- Use a confidence interval to determine the conclusion of a hypothesis test

4.2 Record snowfall

In the winter of 2018-2019, Bozeman had a record snowfall which resulted in the collapse of two flat-roofed buildings on the MSU campus. A writer for the Washington Post predicted the heavy snowfall for 2018-2019 due to the El Nino weather pattern that occurred in that season. A meteorologist in Montana wanted to see if the weather pattern really was associated with total snowfall. She obtained historical data from 44 years on the weather pattern (El Nino or La Nina) and snowfall (in inches) at the Billings Weather Station.

```
favstats(Snowfall~WeatherPattern, data=Snow)
#>   WeatherPattern  min   Q1 median   Q3   max     mean       sd  n missing
#> 1        El_Nino 31.9 46.4   57.7 64.3  87.9 56.23043 13.00823 23       0
#> 2        La_Nina 44.5 51.4   60.9 70.3 107.2 63.13333 15.48626 21       0
```

```
ggplot(data = Snow,
       aes(x = WeatherPattern, y = Snowfall)) +
    geom_boxplot() +
    labs(title = "Snowfall by weather pattern",
         x = "Weather pattern")
```



4.2.1 Quantitative variables review

1. The two variables assessed in this study are the type of weather pattern and snowfall. Identify the role for each variable (explanatory, response).

2. Which group (El Nino or La Nina) has the highest center? Explain what measure you are using?

3. Using the side-by-side boxplots, which group has the largest spread? How did you make that choice?

4. Is this an experiment or an observational study? Explain your reasoning.

5. Is this a paired data set or independent groups? Explain your answer.

4.2.2 Ask a research question.

6. Write out the parameter of interest in context of the study. Use proper notation and be sure to define your subscripts. Use El Nino minus La Nina as the order of subtraction.

7. What are the two competing possibilities we will evaluate in this study?

8. Identify which is the null hypothesis and which is the alternative hypothesis.

4.2.3 Summarize and visualize the data

9. Calculate the summary statistic. Use El Nino minus La Nina as the order of subtraction. What is the appropriate notation for the statistic?

4.2.4 Use statistical inferential methods to draw inferences from the data

Remember that the null distribution is created based on the assumption the null hypothesis is true. In this study, we asssume there is no association between variables. This means that a snowfall value could be in either an El Nino year or a La Nina year.

To demonstrate this your instructor will use cards to represent the sample.

10. How many cards will we start with?

11. What will we write on each card?

12. Next we will mix the cards together and shuffle into two piles. How many cards will go into each pile? What should we label the piles?

13. What value is calculated from the cards and plotted on the null distribution?

14. Once we create a null distribution of 1000 simulations, at what value do you expect the distribution to be centered at? Explain your answer.

simulation

15. Load the package CatStats. Using the ….Enter the values for the ….

16. Report the p-value. How much evidence does the p-value provide against the null hypothesis?

17. Using bootstrapping find a 90% confidence interval.

bootstrapping simulation

18. Interpret the interval you calculated in Question 17.

4.2.5 Communicate the results and answer the research question

19. Write a paragraph summarizing the results of the study. Be sure to include:

- Summary statistic
- P-value
- Conclusion in context
- Confidence interval
- Interpretation of the confidence interval
- Scope of inference

4.2.6 Revisit and look rorward

20. Would the results from the theory-based test match the results we saw with the simulation? Explain why or why not.

21. If we had data on 45 La Nina years and 47 El Nino years and found a similar summary statistic, what would happen to the p-value? The width of the confidence interval? The power?

4.3 Additional notes

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

ACTIVITY 5

Hand Dexterity

5.1 Learning outcomes

- Given a research question, construct the null and alternative hypotheses in words and using appropriate statistical symbols
- Describe and perform theory-based hypothesis tests for the slope
- Interpret and evaluate a p-value
- Construct and interpret a theory-based confidence interval for slope
- Use a confidence interval to determine the conclusion of a hypothesis test

5.2 Terminology review

The following terms will be covered in this activity.

- Correlation
- Slope
- Coeffiecient of determination

For further explanation of these topics review Chapter 3 and 7 in the textbook.

5.3 Hand dexterity

Physical therapists often evaluate manual (hand) dexterity by having patients complete simple tasks, such as moving pegs on a board or threading objects through holes. Researchers want to examine the manual dexterity of children as part of a follow-up study of a test originally designed for adults to see how manual dexterity changes with age. In this test, 174 participants were given a board with 16 pegs, each in their own hole, arranged in a 4x4 grid. Participants were instructed to pick up the peg with one hand, flip it over by rotating their wrist, then reinsert it in the same hole. Using this test, researchers want to know if as people age the speed at which they can flip all 16 pegs increases.

The variables in this dataset consist of the following:

- **time:** Recorded time to flip all 16 pegs, measured in seconds.
- **speed:** The average speed to flip a peg for each participant (seconds per peg).
- **age:** Age of the participants, measured in years.
- **dominant:** Whether the participant's dominant hand was used, coded as 0 for no, 1 for yes.
- **gender:** The participant's gender, recorded as a binary variable, 0 for male, 1 for female.
- **HD:** The dominant hand of the participant, recorded as "R" for right hand, "L" for left hand.

- **handUsed:** Which hand the participant used to complete the test, recorded as "R" for right hand, "L" for left hand.

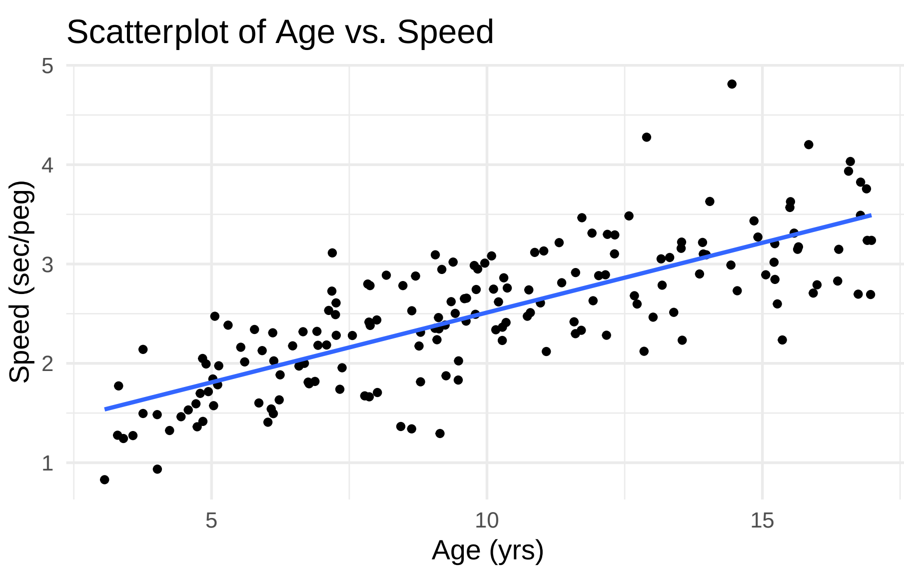
Data source: Hand Dexterity in Children: Administration and Normative Values of the Functional Dexterity Test (FDT), Gogola, G., et al., 2013

5.3.1 Vocabulary review

1. Explain why regression methods are appropriate to use to address the researchers' question. Make sure you clearly define the variables of interest in your explanation and their roles.

2. What is the scope of inference for this study? Explain your answer.

3. Create a scatterplot to examine the relationship between the speed at which a participant can flip a peg and the age of the participant. Provide this plot. Based on your plot, does it appear that there is a relationship between `age` and `speed`? Note: `age` should be on the x-axis.



4. Describe the features of the plot you created in Question 3.

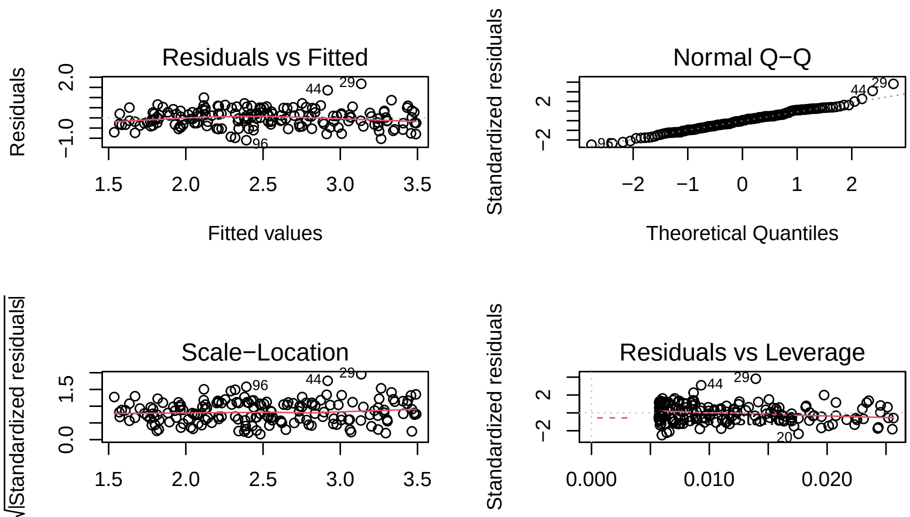
If you indicated there are potential outliers, which points are they?

5.3.2 Conditions for the least squares line

When performing inference on a least squares line, the follow conditions are generally required

- Linearity: the data should follow a linear trend
- Nearly normal residuals: residuals must be nearly normal
- Constant variability: the variability of points around the least squares line remains roughly constant
- Independent observations: individual data points must be independent

The scatterplot and the residual plot will be used to assess the conditions for approximating the data with the T-distribution



5. Are the conditions met

5.3.3 Ask a reseach question

6. Write out the null hypothesis in words.

7. Using the research question, write the alternative hypothesis in notation.

5.3.4 Summarize and visualize the data

Enter the variable names into the linear model function to get the linear model output.

```
#>               Estimate  Std. Error  t value     Pr(>|t|)
#> (Intercept) 1.1070563 0.093326769 11.86215 4.237698e-24
#> age         0.1404378 0.008796699 15.96483 8.777516e-36
```

8. Using the output above, write the equation of the regression line.

9. Interpret the slope in context of the problem.

10. Using your estimated line of best fit, predict the per peg speed for a participant who was 9.18 years old. Show all work.

11. Calculate the residual associated with the observation (9.18, 2.95), using your estimated regression line from question 8.

5.3.5 Use statistical inferential methods to draw inferences from the data

To find the value of the test statistic to test the slope we will use,

$T = \frac{\text{slope estimate}}{SE} = T = \frac{b_1}{SE(b_1)}$

We will use the linear model output above to get the estimate for slope and standard error.

12. Calculate the test statistic for slope.

13. What value does the value calculated in question 12 match in the linear model output?

14. Interpret the test statistic in context of the problem.

15. Using the linear model output, report the p-value for the test of significance.

16. Based on the p-value, how much evidence is there against the null hypothesis?

Recall that a confidence interval is calculated by adding and subtracting the margin of error to the point estimate.

point estimate $\pm t^* SE(estimate)$ $b_1 \pm t^* SE(b_1)$

The t^* multiplier comes from the t-distribution with $n - 2$ df.

```
#> [1] 1.973852
```

17. Calculate the 95% confidence interval for the true slope.

18. Calculate the coefficient of determination for a linear model that describes the relationship between `age` and `speed`. Use proper notation.

19. Interpret this value in the context of the study.

5.3.6 Communicate the results and answer the research question

20. Based on the p-value, write a conclusion in context of the problem.

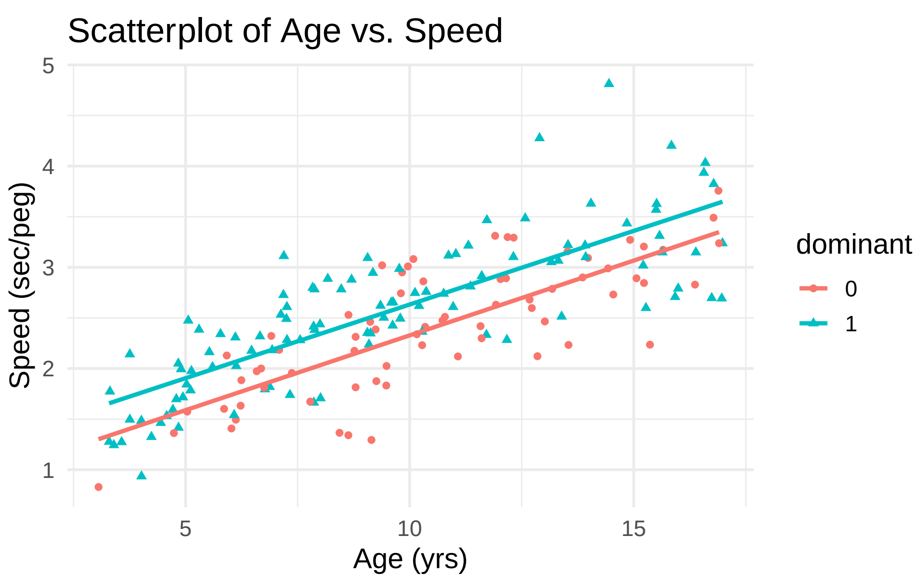
21. Interpret the 95% confidence interval in context of the problem.

22. Summarize the results of the study in a written paragraph. Be sure to include.

- Summary statistic
- Test statistic and interpretation
- P-value and interpretation
- Confidence interval and interpretation
- Conclusion in context
- Scope of inference

5.3.7 Revisit and look forward

23. Is there an effect due to gender on this linear relationship? Explain your answer using the scatterplot.



5.4 Additional notes

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