

Paired Data

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
- Interpret and evaluate a p-value
- Construct and interpret a theory-based confidence interval
- Use a confidence interval to determine the conclusion of a hypothesis test

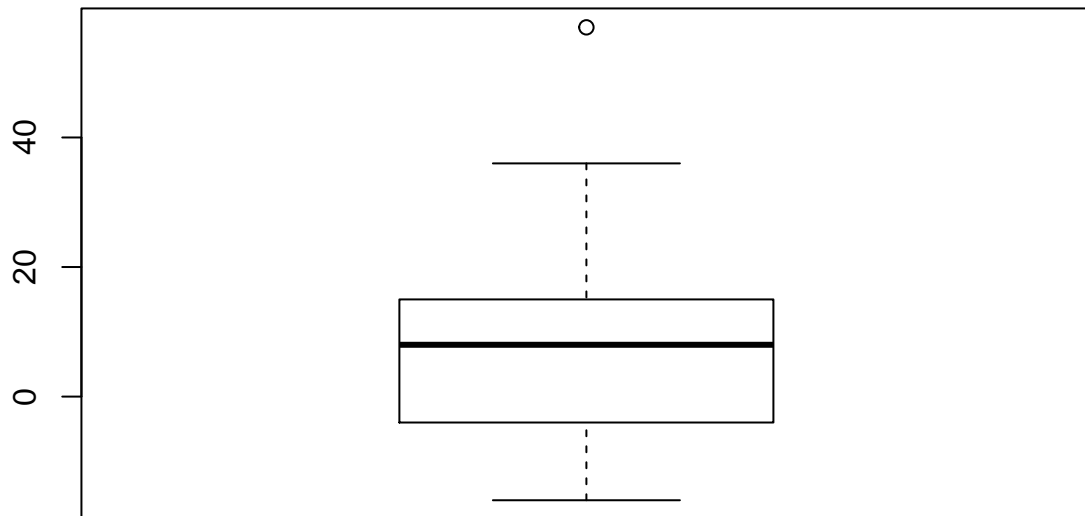
Mean Difference in Heart Rates for Jumping Jacks and Bicycle Kicks

Which exercise, jumping jacks or bicycle kicks will raise your heart rate more? Students in an introductory statistics class were asked to participate in an experiment to answer this question. Each student flipped a coin to determine which exercise to complete first. If the coin landed on heads the student would do jumping jacks for 30 seconds and then measure their heart rate. After a 2 minute break the student would do bicycle kicks for 30 seconds and then record their heart rate. If the coin landed on tails the student would complete bicycle kicks first followed by jumping jacks using the same times as above.

Review

```
heartrate <- read.csv("../data/HeartRates.csv")
boxplot(heartrate$Diff,
        main="Boxplot of the Differences in Heart Rates for Exercises")
```

Boxplot of the Differences in Heart Rates for Exercises



```
fav_stats(heartrate$Diff)
```

```
##  min Q1 median Q3 max    mean    sd  n missing
## -16 -4      8 15  57 7.604651 15.91666 43      0
```

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. What is the purpose of randomizing the order of jumping jacks and bicycle kicks before measuring heart rates?

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.

Summarize and Visualize the Data

9. Report the summary statistic for the data.

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

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?

Add simulation here

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

13. What proportion of samples are beyond the sample mean difference in heart beats for jumping jacks minus bicycle kicks?

Communicate the results and answer the research question.

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

To calculate a confidence interval to estimate the mean difference in heart rates, we will use this equation:

$$\bar{x}_d \pm t^* SE(\bar{x}_d), \text{ where } SE(\bar{x}_d) = \sqrt{\left(\frac{\bar{x}_d}{n_d}\right)}$$

To find the t^* multiplier for a 95% confidence interval we will find the value in the t-distribution that represents the endpoints for the middle 95% of the data.

```
qt(0.95+0.025, df = 42) #Multiplier for 95% confidence with 42 (n-1) df
```

```
## [1] 2.018082
```

15. Calculate the $SE(\bar{x}_d)$.

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

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

18. Based off your p-value and confidence interval, write a conclusion.

Revisit and Look Forward

Suppose we had a sample of 90 students instead of 43 resulting in the same summary statistic.

19. Would this new data provide more or less evidence against the null hypothesis? Explain your answer.

20. Would this result in a wider or narrower confidence interval?