Memory Study

Comparing Two Means: Simulation-Based Approach

LEARNING GOALS

- State the null and the alternative hypotheses for an explanatory variable with two categories and a quantitative response variable.
- Implement the 3S strategy to compare two means: Find a **statistic**, **simulate**, and compute the **strength** of evidence against observed study results happening by chance alone.
- Use the **Comparing Groups** applet to conduct a simulation of the null hypothesis.
- Find and interpret the standardized statistic and the p-value for a test of two means.
- State a complete conclusion about the alternative hypothesis (and null hypothesis) based on the p-value and/or standardized statistic and the study design, including statistical significance, estimation, generalizability, and causation.

STEP 1: Ask a research question.

1. What sorts of things do you think affect a person's memory?

This statistical study will investigate how the way the letters in a sequence are chunked or grouped affects people's ability to memorize a string of letters in 20-seconds.

Our research question: "Does the type of chunking make a difference in the number of letters from a sequence that people can remember correctly after memorizing for 20-seconds?"

STEP 2: Design a study and collect data.

In this study, each person was randomly assigned to one of two sequences A or B, where the letters in the sequences were chunked or grouped in two different ways.

Sequence A (FBI-O): FBI-OMG-CIA-USA-SAT-GPA-ACT-NBA-CPR-LOL
Sequence B (FBIO-): FBIO-MGC-IAU-SASA-TGP-AAC-TNB-AC-PRL-OL

For each person we recorded the number of letters correctly recalled after memorizing for 20-seconds, as well as whether they had caffeine that day and how much sleep they got the previous night.

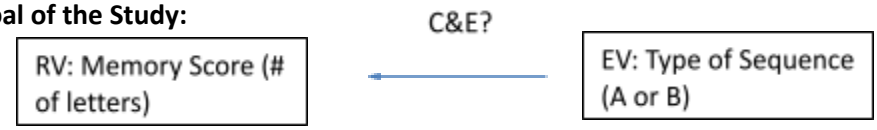
2. What is the observational (or experimental) unit in this study and how many are there?

3. Identify the explanatory variable in this study and classify it as categorical or quantitative.

4. Identify the response variable in this study and classify it as categorical or quantitative.

5. Was this a designed experiment or an observational study? Explain how you are deciding.

Goal of the Study:



6. Will each person have the exact same memory score? If not, what are some possible sources of variability in memory scores?

A **Sources of Variation Diagram** helps us think about the study in terms of the variability we expect to see in the memory scores. Fill in the diagram below with what you already know about the study including discussion of any "controls" used in the study to minimize unexplained variation.

Observed Variation in:	**Sources of explained variation**	**Sources of unexplained variation**
	•	•
Inclusion criteria *Design*		

STEP 3: Explore the data.

7. At the bottom of the data collection page, select the option to the **Load the results into the Comparing Groups applet**. Make sure the memory score is the response variable and sequence is the explanatory variable. Select Show groups.

   a. Notice that the applet creates parallel dotplots, one for each sequence group. Check the **Add boxplots** button to add boxplots to these graphs. Based on these graphs alone, which group (**Sequence A: FBI-O** or **Sequence B: FBIO-**) tends to have the higher memory scores?

   b. Notice also that the applet computes numerical summaries of the data, such as the mean and standard deviation (SD) of the memory score for each sequence.

      i. For **Sequence A: FBI-O**, record the sample size (*n*), mean, and SD.

n_{SeqA} = Mean$_{SeqA}$ = SD $_{SeqA}$ =

ii. For **Sequence B: FBIO-**, record the sample size (*n*), mean, and SD.

n_{SeqB} = Mean$_{SeqB}$ = SD $_{SeqB}$ =

c. Based on the numerical summaries, which group (**Sequence A: FBI-O** or **Sequence B: FBIO-**) had the higher mean memory score? Which sequence had more variability in memory scores?

STEP 4: Draw inferences beyond the data.

8. What are two possible explanations for Sequence A having had a higher mean memory score than Sequence B?

9. Which explanation do you think is responsible for the difference in the mean memory scores between the two sequences?

3S Strategy

Statistic

10. A natural statistic for measuring how different the observed group means are from each other is the difference in the mean memory score between the two sequences. Report the value of this statistic (Sequence A minus Sequence B).

Simulate*:*

11. To conduct this simulation:

a. How many cards do you need?

b. What will you write on each card?

c. To conduct *one repetition* of this simulation, shuffle the stack of cards well and then randomly distribute cards into two stacks: one for sequence A and one for sequence B. (Make sure the sample sizes are the same as what happened in the actual study.)

i. Calculate and report the sample means for each rerandomized group.

ii. Calculate the difference in the mean scores for each sequence (A minus B)

d. Combine this result with your classmates' to create a dotplot that shows the distribution of several possible values of the difference in sample means that could have happened due to pure chance if the type of sequence has no impact on memory score. Sketch the dotplot, being sure to label and scale the horizontal axis.

e. At about what value is the dotplot centered? Explain why this makes sense. (*Hint*: What are we assuming to be true when we conduct the simulation?)

f. Where is the observed difference in means from the original study (as reported in #10) on the dotplot? Did this value happen often, somewhat rarely, or very rarely? How are you deciding?

12. You would now like to conduct many, many more repetitions to determine what is typical and what is not typical for the difference in group means, assuming that cap type has no impact on rolling time in seconds. We think you would prefer to use a computer applet to do this rather than continue to shuffle cards for a very long time, calculating the difference of group means by hand. Go back to the **Comparing Groups** applet, check the **Show Shuffle Options** box, select the **Plot** display, and press **Shuffle Responses**.

a. Describe what the applet is doing and how this relates to your null hypothesis.

b. Record the shuffled difference in sample means for the rerandomized groups, as given in the applet output. Is this difference more extreme than the observed difference from the actual study? How are you deciding?

c. Click on **Shuffle Responses** again and record the simulated difference in sample means for the rerandomized groups. Did it change from #12b?

d. Click on **Re-Randomize** again and record the simulated difference in sample means for the rerandomized groups. Did it change from #12b and #12c?

Now to see many more possible values of the difference in sample means, assuming cap type has no impact on rolling time, do the following in the **Comparing Groups** applet:

- Change **Number of Shuffles** from 1 to 997.
- Press **Shuffle Responses** to produce a total of 1,000 shuffles and rerandomized statistics.

e. Consider the histogram of the 1,000 could-have-been values of difference in sample means, assuming that sequence no effect on the memory scores.

i. What does one dot on the dotplot represent? (*Hint*: Think about what you would have to do to put another dot on the graph.)

ii. Describe the overall shape of the null distribution displayed in this dotplot.

iii. Where does the observed difference in sample means (as reported in #10) fall in this dotplot: near the middle or out in a tail? Are there a lot of dots that are even more extreme than the observed difference, assuming the cap type has no impact on rolling time in seconds? How are you deciding?

f. To estimate a p-value, continue with the **Comparing Groups** applet. Type in the observed difference in group means (as reported in #10), choose the appropriate alternative hypothesis in the **Count Samples** box, and press **Count**. What is your approximate p-value?

g. Complete the following sentence to provide the interpretation of the p-value.

The p-value of _____ is the probability of observing _____assuming _____.

Strength of evidence

13. Based on the p-value, evaluate the strength of evidence provided by the data against the null hypothesis: not much evidence, moderate evidence, strong evidence, or very strong evidence?

STEP 5: Formulate conclusions.

14. ***Generalization***
Were the participants in this study randomly selected from a larger "population"? Describe the population to which you would feel comfortable generalizing the results of this study. (See Inclusion Criteria.)

15. ***Causation***
Were the participants in the study randomly assigned to a sequence? What type of conclusion can we draw?

16. ***Conclusion:*** *Write a conclusion which answers the research question with regard to the strength of evidence in the context of the study.*

17. ***Estimation:*** Fill in the following interpretation of what this confidence interval reveals, paying particular attention to whether the interval is entirely positive, entirely negative, or contains zero. (*Hint*: Include the appropriate numbers and then choose the appropriate "direction" (higher or lower) in your interpretation.)

I'm 95% confident that the long-run mean rolling time in seconds with the _____ cap type is _ (higher/lower) to _____(higher/lower) than the long-run mean rolling time in seconds with the _____ cap type.

STEP 6: Look back and ahead.

18. ***Looking back:*** Did anything about the design and conclusions of this study concern you? In particular, are there things that could have been done to give a better chance finding strong evidence of a true difference between the two groups? Issues you may want to critique include:

    - Any mismatch between the research question and the study design
    - How the experimental units were selected
    - How the treatments were assigned to the experimental units
    - How the measurements were recorded
    - The number of experimental units in the study
    - Whether what we observed is of practical value

19. ***Looking ahead:*** What should the researchers' next steps be to fix the limitations or build on this knowledge?