Exploration 7.1-7.2: Does rinsing with pink artificially-sweetened solution improve running speed?

Part 1: Paired Data

LEARNING GOALS

- Identify a study design as having pairing or independent groups.
- Identify a study design as paired using repeated measures or paired using matching.

Athletes, coaches, and scientists have worked for years to identify nutritional products that can enhance performance during exercise. Endurance athletes commonly consume drinks containing water, electrolytes, and carbohydrates during training and competition to replace what they lose via metabolism and sweat. More recently, scientists have shown that simply rinsing your mouth out with a carbohydrate solution while running can increase performance. Researchers Brown et al. (2021) wanted to see whether there was a placebo-effect with these types of rinses. More specifically, they wanted to see whether this effect would be greater when the solution was dyed pink compared to a clear solution as previous studies have shown that the color pink leads to greater perceived sweetness. To test this, they planned to have the participants run on a treadmill and see how far they would run in 30 minutes.

Think about designing a study to investigate this question.

1. Identify the explanatory and response variables in this study.

2. Explain why it would be impossible to design an observational study to investigate this question and if you could, why it would not allow you to decide whether there was a greater effect with the pink solution.

3. Suppose 20 runners volunteered to participate in this experiment. Suppose that you also plan to assign a single solution, either pink or clear, to each runner. How would you decide which runner used which solution?

A reasonable experimental design would be to randomly assign 10 of the 20 runners to use the pink solution and the other 10 to use the clear solution.

4. Some runners are faster than others. Explain how random assignment controls for this, so that speed is not likely to be a confounding variable in this study.

Even though random assignment tends to balance out other variables (such as natural speed) between the two groups, there's still a chance that most of the fast runners could end up in one group and most of the slow runners in the other group. More importantly, there's likely to be a good bit of variability in the runners' speeds, and that variability would make it harder to spot a difference between the two solutions even if one solution is really better than the other.

5. Suggest a different way of conducting the experiment to make sure that speed is completely balanced between the two groups.

Definitions

- For a ***paired design***, response values come in pairs, one from each group. Sometimes the pairs come from matching individuals to create groups of two; sometimes the pairs come from measuring the same individual twice, once under each condition.
- For an ***independent groups design***, each individual in a group is unrelated to all the other individuals in the study. Each individual provides only one response value.

In this study each runner can rinse with *both* solutions. That way, we can be sure that neither treatment has more of the fast or slow runners, and we can also expect that *differences* in distances for each runner will show considerably less variability than individual running distances.

6. What aspect of this experiment should be determined randomly? (*Hint*: The treatment is not determined randomly, because each runner experiences both treatments. But what other factor could still have an effect on the response unless it was randomized?)

7. What do you suggest using as the variable to be analyzed with this paired-design experiment? (*Hint*: Think of a better option than simply analyzing the set of distances using the pink solution and the set of distances using the clear solution separately the way you would for an independent groups design of the sort described in #3 and #4.)

With a paired design, we analyze the *differences* in the response between the two treatments. In this case we would calculate the difference in running distances between using the pink solution and the clear solution for each runner and then analyze the sample of differences.

The order in which the participants run using the two solutions should be determined randomly; otherwise, the order could be a confounding variable: Perhaps runners will generally be slower on their first session and faster on their second or vice versa. Randomizing the order takes away any worries about an order effect.

8. So far you have explored three designs for this study. The first (#2) was observational. The second (#3 and #4) was a randomized experiment with independent groups. The third (#5 and #6) used a paired design with pairs created by *repeated measures*. Consider a fourth design: Suppose you have 20 runners, as before, and you have the time for each runner for a recent 5K race. Explain how you could use this information to create pairs of runners and how you would assign one runner in each pair to the pink solution and the other to the clear solution. (This method is called paired design using *matching*.)

9. Of the four designs (observational with independent groups, experimental with independent groups, paired design using repeated measures, and paired design using matching), which do you think is best for this context? Explain why. (*Hint*: Pairing works best when the units in a pair are as similar to each other as possible.)

Part 2: Simulation-based approach for analyzing paired data

LEARNING GOALS

- Understand the difference between independent samples and paired samples in terms of the study design and how variability can be lower in a paired design and how this can influence the strength of evidence.
- Complete a simulation-based test of significance of a paired design by writing out the hypothesis, determining the observed statistic, computing the p-value, and writing out an appropriate conclusion.

STEP 1: Ask a research question. Now let's look at how we will analyze paired data using a simulation-based approach. Remember that our research question is to see whether this effect would be different if the solution was dyed pink compared to a clear solution. Previous studies have shown that the color pink leads to greater perceived sweetness and thus may make the participants run farther, but the researchers are open to pink possibly having the opposite effect.

STEP 2: Design a study and collect data. To test this, researchers Brown et al. (2021) recruited 10 participants (6 males and 4 females) for the study. All were experienced runners that regularly ran at least three times per week. The participants refrained from strenuous exercise and the consumption of alcohol and caffeine for 24 hours prior to being tested and food for 4 hours prior to being tested. Before their initial session, they watched a video showing the benefits of a carbohydrate mouth rinse while exercising and were told they were going to compare two commercial sports drinks (this, of course, was a lie). Two non-caloric artificially sweetened solutions (0.12 g of sucralose dissolved in 500 ml of water) were prepared for each runner. One was dyed pink, and the other was left clear. After a warm-up, each runner was instructed to run for 30 minutes on a treadmill at a self-selected pace maintaining a rating of perceived exertion of 15 (on a scale of 6 to 20) which can be described as hard. The runners rinsed their mouths out with 25ml of the solution, randomly assigned to be pink or clear, for 5 seconds before spitting out. They repeated this every 5 minutes during their runs. One week later, all the participants returned to repeat the protocol but rinsing with a solution of the other color (clear or pink) than the one they were originally assigned. The distance each participant ran (measured in meters) for the 30 minutes was recorded for each of their runs and is shown in the following table.

Participant	Pink_Distance	Clear_Distance
1	4105	3483
2	4361	3862
3	4105	4172
4	4828	4758
5	4845	4791
6	4845	4995
7	5205	5062
8	5912	5443
9	5827	5702
10	6440	6086

10. Explain why it is reasonable to say that the two distances collected for each runner should *not* be treated as *independent* data.

11. Is the pairing done here use matching or repeated measures? Explain.

12. Notice that the distances using the clear solution are ordered from smallest to largest. What do you notice about the ordering of the pick solution distances? What does that tell you about an advantage of pairing with these data?

Because the data are *paired*, we will compare the two times for each runner by calculating the difference in distances between the two solutions. Thus, we can define our parameter of interest to be

μ_d = long-run mean *difference* in running distance when rinsing with a pink solution and clear solution (pink – clear) in the population of interest.

Note that the subscript "*d*" in μ_d is used to denote that we are looking at an average of *differences*.

13. State the null and alternative hypotheses (using μ_d) to test whether the mean difference in running distance is not 0. (Note: We are doing a two-sided test here like most researchers would actually do.)

> **Key idea**
> When the parameter of interest is the long-run mean difference or population mean difference, the corresponding statistic is the sample mean difference.

STEP 3: Explore the data.

14. Find the average of the differences between the two distances. This is the statistic we will use to summarize the data.

STEP 4: Draw inferences beyond the data. Your null hypothesis should essentially state that there is no difference in the running distances between using the two solutions, on average. If that is the case, it doesn't really matter whether we swap someone's distance using the pink solution with that that person's using the clear solution. This is how we will model the null hypothesis to develop a null distribution. To randomly swap some of the values we can just use a coin flip. If the coin lands heads, you will swap the two distances. If the coin lands tails, you won't swap the distances.

15. Flip a coin for each pair of distances and switch the appropriate ones. Recalculate the differences in distances and find the new simulated mean difference. Pool this value together with those of your classmates'. Where does the actual statistic you found in #14 fit in this null distribution? Is it out in the tail?

16. As you know, it would be better to have many more simulations than what your class just did. We will do this by using an applet.
    - Go to the **Matched Pairs** applet.
    - Press **Clear** to erase the default data and then copy and paste the **RunDistance** data into the data window. Then press **Use Data**.
    - Notice that the applet graphs the individual distances in each group, along with the means and standard deviations for each group.

- Below that, the applet provides a dotplot of the differences in the distances in the sample. Note that some of these difference values are negative numbers because you are looking at *change* or *difference* in distances. The graph of these differences also shows the mean of the differences and the standard deviation of the differences.
- Write down these values in the following table:

Condition	**Sample mean, $\bar{x}$**	**Sample SD, *s***
Pink solution	$\bar{x}_{pink}$ =	s_{pink} =
Clear solution	$\bar{x}_{clear}$ =	s_{clear} =
Diff = Pink – Clear	$\bar{x}_d$ =	s_d =

17. How does the SD of the difference in distances (s_d) compare to the SD of the individual distances for each color (s_{pink} and s_{clear})? Explain what this is telling us in terms of variability in runner distances.

The **Matched Pairs** applet will perform the simulation similar to what you did with flipping a coin.

- Check the **Randomize** box and click on **Randomize**.
- Once the coin tosses have determined which distance will be in which column, the applet displays the rerandomized data (the colors show you the original column for each observation, so you should see a mix in each group now).
- The could-have-been value for the mean difference is added to the **Average Difference** graph.

18. What is the value of your simulated mean difference? Is the actual mean difference more extreme than your simulated mean difference?

19. Update the number of times to **Randomize** to 99 (for a total of 100 repetitions), uncheck **Animate**, and press **Randomize**. Consider the Average Difference graph the applet has created.

    a. How many dots are in this graph?

    b. What does each dot represent?

The table below summarizes the key aspects of the simulation:

Null hypothesis	=	Long-run average difference in distances is 0
One repetition	=	Rerandomizing (possibly swapping) distances within runners
Statistic	=	Average difference in distances in the sample

20. To better see the long-run pattern of mean difference in sample means that could have been, IF the two distances were *swappable*, update the number of times to **Randomize** to 900 and press **Randomize** (for a total of 1,000 repetitions). Describe the updated graph of average differences with the 1,000 samples or repetitions, with regard to the following features.

    a. What is the shape of the graph?

    b. About what number is this graph centered? Explain why you were expecting this.

c. This graph also reports a value for standard deviation, SD. Report this value and give a simple *interpretation* of this value answering, “What is this value measuring?”

21. You now should have generated 1,000 possible values of the mean difference in distances between using the two solutions that were simulated assuming the null hypothesis was true. How does the observed mean difference from the study (as reported in #14) compared to these simulated values? Is an average difference in distances like that observed in the actual study unlikely to happen by chance alone if distance using the pink solution and distance using the clear solution are the same, on average? How are you deciding?

To quantify the strength of evidence against the null hypothesis, you can find the p-value.

- Go back to the **Matched Pairs** applet.
- In the **Count Samples** box, make an appropriate selection from the drop-down menu (*Hint:* In what direction does your alternative hypothesis look?) and enter the appropriate number in the box (*Hint:* At least as extreme as what number?).

22. Report your approximate p-value.

3S Strategy

As you may have already noticed, the strategy we used to find the p-value is the same 3S strategy that has been used previously.

1. **Statistic:** Compute the statistic in the sample. In this case, the statistic you looked at was the observed mean difference in distances.
2. **Simulate:** Identify a chance model that reflects the null hypothesis. To simulate what could have been if the null hypothesis is true, you can toss a coin for each runner, and if it lands heads, swap the two distances recorded for that runner. If the coin lands tails, do not swap the distances. Repeat this process 1,000 times (or more), recording the mean difference in distances each time and thus obtaining a distribution of these mean differences that were simulated assuming the null hypothesis were true.
3. **Strength of evidence:** If your actual observed statistic falls in the tail of the null distribution, then you have strong evidence that there is a difference in the average distances between those that rinse with the two solutions.

Note: The distances using both solutions were *paired* on the same individuals, and so you used a simulation method that lets you use this information.

23. Alternatively, you can summarize the strength of evidence using a standardized statistic. Find the standardized statistic and confirm that the strength of evidence you receive from the p-value is approximately the same as with the standardized statistic. (Remember that the standard statistic is how many standard deviations the observed statistic is above the theoretical mean of the null distribution.)

24. We can again use the 2SD method to approximate a 95% confidence interval for the mean difference in distances between using the two solutions. The overall structure of the 2SD interval formula is the same:

estimate ± 2(SD)

where the estimate is the sample mean difference in distances and SD is the standard deviation of your null distribution when you did 1,000 repetitions in the applet (NOT the standard deviation from the data). Use these numbers to find an approximate 95% confidence interval for the mean difference in distances between using the two solutions.

STEP 5: Formulate conclusions.

25. Use the p-value obtained in #22 to state a conclusion in the context of the problem. Be sure to comment on statistical significance. Can you conclude that there is strong evidence of a *difference* in average running distance between rinsing with a pink solution and clear solution in the long run? Why or why not? Can you conclude that there is strong evidence that those rinsing with the pink solution will have a *larger* average running distance than those rinsing with the clear solution in the long run?

26. Can you draw a cause-and-effect conclusion? Explain.

27. To what population are you willing to generalize the results?

28. Provide an interpretation of your confidence interval from #24, being sure to describe the parameter in this context.

STEP 6: Look back and ahead. The researchers went to great lengths to keep the conditions similar for all runners to try to reduce variability in the data. For example, they all used the same treadmill in the same laboratory. The participants all warmed up using the same protocol, all rinsed at the same time intervals, all were tested at the same time of day. Trying to keep all of these variables constant helps the researchers convince others that the color of the solution, not any of these other variables, is causing any difference observed between the two distances. It also helps reduce variability in the data.

29. One of the last statements in the researchers' paper is, "Future research should seek to elucidate the link between mouth rinse colour, perceived carbohydrate intake and psychophysiological outcomes in exercising humans." [Note: Psychophysiological outcomes are physiological outcomes that are affected by psychological processes.] What are they saying here that should be done that they haven't already shown in their paper?

Exploring Further

Let's check out how things would have worked had we ignored the pairing and analyzed the data as if the distances between using the two solutions had come from two totally different samples that were independent of each other.

30. Go to the Multiple Means applet and analyze the data as though we have two independent samples, as you did in Chapter 6. Before you paste in the data, click on the **Unstacked** box, clear the data out that is shown, and paste in the **RunDistance** data set. Click on **Use Data** and make sure the statistic selected is the **Difference in Means.** Develop a null distribution with at least 1,000 shuffles.

    a. What is the difference in means as reported in the applet? How does this compare to the mean of the differences from the Matched Pairs applet you reported in #22?

b. What is the mean and standard deviation of your null distribution?

c. What is the approximate p-value for a two-sided test?

31. Compare the null SD obtained using the *two-independent-samples* method to that obtained using the *paired samples* method. Which null SD is larger?

32. Compare the p-value obtained using the *two-independent-samples* method to that obtained using the *paired samples* method. Which p-value is smaller and hence provides stronger evidence against the null hypothesis of no difference?

Note: Using a paired samples method will often give a smaller p-value and hence stronger evidence against the null hypothesis than the two-independent-samples method. This is what you should have found #32. This happens because runners that tend to run farther than most other runners using the pink solution also run farther using the clear solution. Similarly, for those that run the shorter distances. This makes the variability of the differences smaller than the variability of the individual data. (You should have noted this difference when you answered #17.)

Reference

Brown, Daniel R., e al. Mouth Rinsing with a Pink Non-caloric, Artificially-Sweetened Solution Improves Self-Paced Running Performance and Feelings of Pleasure in Habitually Active Individuals. *Frontiers in Nutrition* 8:217 doi:10.3389/fnut.2021.678105(2021). https://www.frontiersin.org/article/10.3389/fnut.2021.678105