Exploration 2.2: How does caffeine affect finger tapping?

Approximately 90% of North American adults consuming caffeine daily (Lovett, *New Sci*, 187, 2005). Caffeine is widely used as a stimulant—helping you stay awake late studying, waking up for an early morning class, or energizing you enough to focus well on the task at hand. Researchers wanted to investigate whether there are other ways to get the same effects, with little to no downside? To begin to answer this question, they compared the effects of a placebo, caffeine, and theobromine, which is the active chemical naturally found in chocolate and is an alkaloid with a similar molecular structure and effects on people as caffeine.

To carry out their study, the researchers trained healthy young adults to tap their fingers in such a way that the tapping rate could be measured (number of taps per minute). After learning/ practicing this type of finger tapping, participants were assigned one of the three treatments and their finger tapping rate was measured 2 hours later. The treatments were a 200mg caffeine pill, a 200mg theobromine pill, or a 200mg placebo pill.

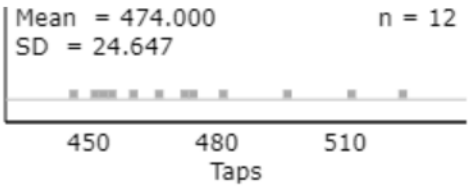
(Scott & Chen, "Comparison of the Action of l-ethly Theobromine and Caffeine in Animals and Man," *Journal of Pharmacologic Experimental Therapy,* 82, 1944).

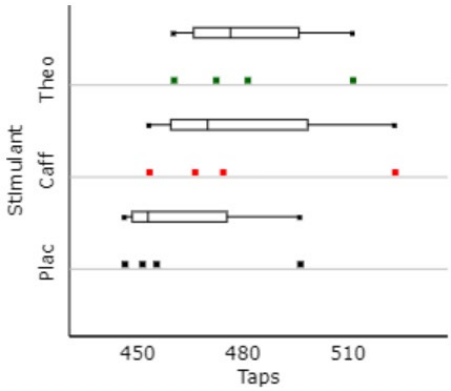
The researchers would like to know whether the type of pill affects the finger tapping rate and specifically, whether theobromine affects finger tapping differently from caffeine.

1. Identify the observational units, the explanatory variable, and the response variable. Complete the hypothesized Sources of Variation diagram.

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

2. Consider the following output from the **Comparing Groups** applet.





	n	Mean	SD
Theo	4	481.00	21.77
Caff	4	479.00	30.58
Plac	4	462.00	22.96
pooled	12	474.00	25.41

a. Calculate the observed effects for each type of pill.

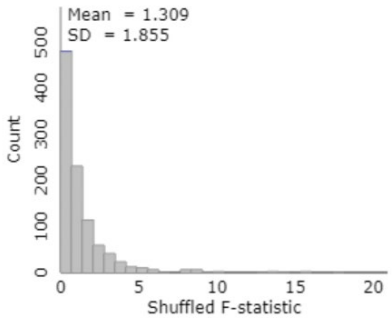
Theobromine: Caffeine: Placebo:

b. Does the type of pill appear to explain some of the variability in the tapping rate? How are you deciding? [Note: Descriptively, not inferentially]

3. Consider the following ANOVA table for the (one variable) separate means model and a simulated null distribution of the *F*-statistic.

ANOVA table:

Source	df	Sum Sq	Mean Sq	F
Treatments	2	872.00	436.00	0.68
Error	9	5810.00	645.56	
Total	11	6682.00		



How much variation in tapping rate is explained by the type of pill? Based on the *F*-statistic and the null distribution, should we find a large p-value or a small p-value? How are you deciding?

4. Suggest some possible explanations for finding an insignificant result here.

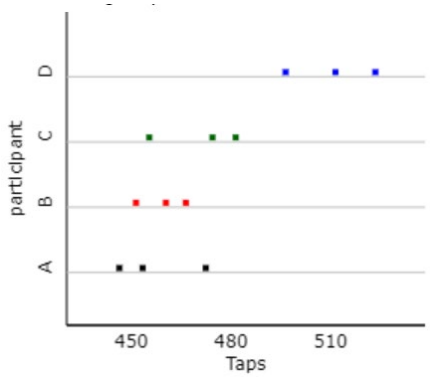
Block Design

The researchers attempted to minimize sources of unexplained variation in the finger tapping rates by waiting exactly 2 hours for each subject before taking the finger tapping measurements and measuring the finger tapping rates exactly the same way for each subject. In addition, the researchers included only young, healthy adults in their study. Because of this strict inclusion criteria, we expect the variability in finger tapping rates to be smaller than if the study had a broader group of participants, e.g., older adults, adults with health problems, etc. However, even within this group of young, healthy adults, there's likely to be considerable person-to-person variation in finger tapping rate, caused by things like genetics, motivation, history of caffeine use, etc. To account for these person-to-person differences, we could assign each participant to take each of the three pills (in random order), measuring the finger tapping rate 2 hours after each type of pill is taken, and then repeating for the next pill etc Of course, a sufficient amount of time should be waited between each pill so that the effects can "wash out" of each person's system. This type of experimental design is called a "block design."

Definition: A ***block study design*** creates ***blocks*** of experimental units that are similar to each other, randomly assigns the treatments within each block, and then analyzes the data in a way which accounts for the more sophisticated design. When there are only two groups being compared a block study design is called a *matched pairs design*. (The term *block* comes from the first block designs which were agricultural experiments in large fields where separate parts of the field were called "blocks.")

In fact, in this study design, each person was considered a block, and all of the treatments were assigned to each person (in random order). This is also called a *repeated measures design*.

5. Based on the following output, do the data confirm or refute the researchers' prediction that different people tend to have different finger tapping rates? Does one of the participants appear to have a faster average rate of tapping? A slower average rate of tapping?



	n	Mean	SD
D	3	510.00	13.53
C	3	470.00	13.45
B	3	459.00	7.55
A	3	457.00	13.45

Because the participants appear to have very different finger tapping rates, it's a good thing the researchers decided to use a block design!

Modifying the Analysis to Consider Blocking

Now that we know more about the actual way this study was designed, our analysis should reflect the study design that was used.

Key Idea: The data analysis should correspond to the study's design.

Because the study design was actually a block design, we need to re-consider how to simulate the behavior of the *F*-statistic assuming the null hypothesis to be true.

6. Think about simulating a null distribution of *F*-statistics. Describe how you could shuffle the observed finger tapping rates in a way that mirrors the actual study design, but assuming the null hypothesis is true.

The null distribution in #3 assumes all 12 responses were re-randomized across the 3 groups – ignoring the blocking.

Analysis #1: Ignore the Blocking—Incorrect

Open the **One Blocking Variable** applet and use the pull-down menu to select the *Finger tapping* data. Press the **Use Data** button.

Replicate the null distribution shown in #3 by checking the **Show Shuffle Options** box and pressing **Shuffle Responses**. Notice that the "Completely randomized" option is selected. Watch one shuffle and then change the **Number of Shuffles** to some large number like 1000. Include a screen capture of your results.

7. Record the mean and standard deviation of this simulated "completely randomized" null distribution.

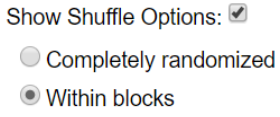
Mean: Std Dev:

8. Use the observed *F*-statistic from #3 and the simulated completely randomized null distribution to approximate the p-value.

Analysis #2: Revising the Simulation to Account for the Blocking

Now let's consider modifying the analysis that does account for the blocking as you considered in #6.

Consider a simulation that takes into account the blocked design. Select **Within blocks** as the randomization method. This color codes the finger tapping rates by participant (block). Change the **Number of Shuffles** back to one. Press **Shuffle Responses** so that the applet shuffles the responses one participant at a time. Notice that this re-randomization, separate for each person (block), now mimics how the randomization was carried out in the actual study.



9. Consider participant D (blue dots). Describe how the finger tapping rates changed for this participant after one random shuffling of the data.

10. Is it possible with this method of shuffling, for 2 (or 3) of each participant's finger tapping rates to end up in the same treatment group? Explain why or why not. (Shuffle Responses again if you need to watch the simulation again, to be sure what is happening.)

Now change the **Number of Shuffles** to carry out a total of at least 1000 re-randomizations of the finger tapping rates *within each participant block* to the treatment groups.

11. What are the mean and standard deviation of the simulated null distribution of *F*-statistics shuffling **within blocks**? How do these compare to the mean and standard deviation of the simulated null distribution of F-statistics obtained using a completely randomized approach (Analysis #1, Question #7)?

Mean using blocking: SD using blocking:

You should see less random variation in the simulated *F*-statistics when shuffling within each participant block (Analysis #2).

12. Explain why this makes sense. That is, explain why the *variation* of the null distribution of *F*-statistics will tend to be smaller when we re-shuffle within blocks as compared to shuffling in a completely randomized way.

Key idea: The null distribution is typically less variable when shuffling within blocks

13. How will the smaller standard deviation in the shuffled within blocks null distribution impact the strength of evidence provided by the observed *F*-statistic, 0.68? Explain.

Strength of evidence: Use the applet to approximate the p-value for this study from this new simulated null distribution. Include a screen capture of your results.

14. How has the strength of evidence (p-value) against the null hypothesis changed when considering blocks (Analysis #2) vs. Analysis #1 which did not consider the blocks? Why?

Simulated p-value without blocking (Question #8):

Simulated p-value using blocking:

Important note: Analysis 1 is not an appropriate analysis of these data, because it does not reflect the study design by ignoring the fact that the design was blocked. We only showed you Analysis 1 here in order to illustrate the benefits of blocking.

Analysis #3: Revising the *F*-statistic to Account for Blocking

In Analysis #2, you modified the simulation to reflect the study design. This had the consequence of reducing the variability in the null distribution, which led to a smaller p-value for the *F*-statistic.

Another way to analyze a block design is to modify the *F*-statistic for the treatments after accounting for the person-to-person variability in finger tapping rates. In particular, we want to measure the differences in finger tapping rates among the treatments *within each person*, and then average those differences across the participants (e.g., are the rates of the placebo treatment consistently lower, regardless of who's doing the tapping?).

To do this, we want to adjust the responses in a way that takes into account the person-to-person differences. This idea is a bit like how a handicap works in golf. When players of differing abilities play one another, they are given different handicaps. A good golfer might have a handicap of 3 (because he/she is expected to score 3 strokes above par; lower scores are better in golf) and a bad golfer might have a handicap of 15 (he/she is expected to score 15 strokes above par). At the end of the round, the final scores are adjusted by adding each player's handicap to their score. So, the good golfer would need to score 12 strokes better than the bad golfer in order to win. By adjusting the golfer's final scores based on handicaps, the golfers are compared on the same scale, and a true "winner" can be determined.

So first, let's estimate those person-to-person differences by computing the *effect* of each person. These are called ***block effects***.

Definition: ***Block effects*** are calculated by subtracting the overall response mean across all responses from the mean of the responses within a block (block mean – overall mean).

15. Calculate the *block effect* for each of the four participants by subtracting the overall average finger tapping rate (474) from the average finger tapping rate for each participant. Note: the block effects will be positive for participants with above average finger tapping rates and negative for participants with below average finger tapping rates.

	Block mean	–	Overall mean	= Block effect
Participant A block effect:		–	474	=
Participant B block effect:		–	474	=
Participant C block effect:		–	474	=
Participant D block effect:		–	474	=

Confirm your calculations by returning to the blocking applet and checking the **Show Block Effects** box (under the dotplots).

16. As with the treatment effects, when each block mean comes from the same number of observations (three in this case), the block effects will sum to zero. Confirm this by summing the four block effects from Question #15.

17. Now, compute "adjusted" finger tapping rates for each participant's three finger tapping rates by subtracting off the participant's (block) effects. For example, because participant A is below average by 17 taps, the adjusted finger tapping rate for the placebo pill for this participant is computed as 446 – (–17) = 446 + 17 = 463.

Finger Tapping Rate	Type of Pill	Participant	Block effect	Adjusted finger tapping rate (Finger tapping rate minus block effect)
446	Placebo	A	-17	463
453	Caffeine	A		
472	Theobromine	A		
451	Placebo	B	-15	
466	Caffeine	B		
460	Theobromine	B		
455	Placebo	C	-4	
474	Caffeine	C		
481	Theobromine	C		
496	Placebo	D	36	
523	Caffeine	D		
511	Theobromine	D		

Key idea: Adjusting for block effects can be thought of as "leveling the playing field" between the blocks, much like if you were in a race and gave a less-abled opponent a head start.

18. Compare the block-adjusted finger tapping values to the original finger tapping values. In terms of the variation, how do the two sets of original and adjusted values compare?

Return to the blocking applet, and check **"Adjust data for block effects."**

19. What happens to the finger tapping values after they have been adjusted for the block effects? In particular, what happens to the mean of each treatment group? What happens to the variation within each treatment group? Why?

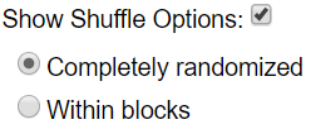
Treatment group means:

Variation within each treatment group:

Explanation:

20. What impact will this reduction in the variation within the treatment groups have on the adjusted *F*-statistic for the treatments?

Confirm that the new block-adjusted *F*-statistic is 7.88 (shown in the applet as Observed F).
Now shuffle the block-adjusted data values using the completely randomized approach. (Why do we no longer shuffle within blocks?)



Key idea: When simulating the null distribution of the *F*-statistic computed from the block-adjusted data you no longer need to re-randomize within the blocks—instead, re-randomize the block-adjusted values into the treatments across all of the experimental units (ignoring the blocks). The blocks have already been accounted for in the analysis by adjusting the data up front.

21. Compare this null distribution to the one obtained earlier shuffling within blocks (Analysis #2). How has the null distribution changed?

Mean: Standard deviation:

22. What is the simulated p-value using the "adjusted" *F*-statistic from the adjusted data?

23. In terms of the effect of the type of pill on finger tapping, what do you conclude after analyzing the block-adjusted *F*-statistic? How does this conclusion compare to the conclusion from Analysis #2, shuffling within blocks? Why does this make sense?

24. We can use an ANOVA table to keep track of the sources of variation. With two variables (treatments and blocks), this is called a ***two-variable ANOVA.*** In the **applet,** check the box to **Show ANOVA table**. Now the blocking variable is also included as an explained source of variation.

Source	DF	SS	MS	F
Treatments				
Blocks				
Error				
Total				

a. Explain how each degrees of freedom (DF) value is calculated.

b. Verify that the sum of squares for the error is what is left over after subtracting the *SStreatment* and *SSblock* values from the *SSTotal*.

c. In the applet, if you uncheck the **Adjust for block effects** box, this shows the ANOVA table corresponding to Analysis #1. Compared to Analysis #1, how has the *SStreatment* changed when we accounted for blocks (Analysis #2)? How has the *SSerror* changed compared to Analysis #1? Why?

d. What percentage of the total variation in the finger tapping rates is due to the participant and what percentage is due to the type of pill? Which variable, participant or type of pill, explains a larger proportion of variation in finger tapping rates? What does this say about the importance of this blocking variable in this study?

Theory-based approaches for finding p-values require certain validity conditions to be met. Here are the validity conditions for ANOVA with a blocking variable.

Validity conditions for ANOVA with a blocking variable

The *F*-distribution is a good approximation to the null distribution of the *F*-statistic as long as:

1. Either the sample size is at least 20 for each the treatment groups being compared, without strong skewness or outliers in the block-adjusted responses; or if the sample sizes are less than 20 for each treatment group, then the distributions of the block-adjusted response variable in each treatment group is approximately symmetric (examine dotplots for strong skewness or outliers OR examine the residuals together as one distribution).
2. The standard deviations of the block-adjusted response variable in each treatment group being compared are approximately equal to each other (the largest standard deviation is not more than twice the value of the smallest standard deviation).
3. The experimental units are randomly assigned to the treatments within each block which means the block-adjusted responses should be approximately independent between and within the treatment groups.

25. Comment on how you the analysis you just conducted meets the validity conditions by evaluating the three conditions listed in the previous box.

26. Another way to evaluate validity conditions is to see how similar the p-values are and how similar the null distributions are between the theory-based and simulated analyses. Comment on the similarity and explain how this provides evidence that the validity conditions are met.

27. What is the p-value reported for the treatment effects for the theory-based analysis? What null and alternative hypotheses are tested by this p-value?

Because the actual study used blocking and we modified the analysis to account for the blocking variable, we need to modify the null and alternative hypothesis being tested as shown here.

> Null hypothesis: The long-run average finger tapping rate in each of the three experimental conditions is the same, *after accounting for differences in finger tapping rate among different people.*
>
> Alt hypothesis: At least one of the three experimental conditions have a different long-run average finger tapping rate, *after accounting for differences in finger tapping rate among different people.*

28. Based on these modified null and alternative hypotheses, state a conclusion about the study.

29. Summarize the conclusions you would draw from this study with regards to generalizability and causation.