

1. **Polio**<sup>1</sup>. Polio is now a very rare disease, thanks to medical research, which developed a vaccine, and thanks to statistical work, which proved the vaccine to be effective. The testing of the vaccine involved two different studies. One was a true experiment, with the treatment and placebo assigned by a coin flip, but the other study was observational.

The observational study used the children who didn't get parental permission for the vaccine as a control group and gave the vaccine to everyone who got permission. In what ways do you think the treatment and control groups are likely to be different? How might that make it more challenging to identify the causal factor?

---

<sup>1</sup>This is a revised version of question 4.E.5 from Cobb, , 1998.

2. **The Free-Choice Paradigm.** The Free-Choice Paradigm<sup>2</sup> is a term for a type of experiment designed to test what is termed the “cognitive dissonance” of a subject. Cognitive dissonance says that individuals strive to decrease the discomfort generated by conflicting cognitions (i.e. opinions or thoughts) by modifying them.

A traditional free-choice paradigm experiment has a subject do the following:

*Step A.* Rate items individually according to their desirability (e.g. oranges, clementines, apples, bananas, etc.), say on a score between 1-10.

*Step B.* Choose between pairs of items that they previously rated similarly (“do you prefer oranges or clementines?”, “do you prefer apples or bananas?”, etc.)

*Step C.* Rate items individually from the first step (again).

Results from such experiments over the past decades show that items ranked highly in the 1st step get increased ratings when they are repeated in step C, and those ranking lowly in the 1st step get even lower ratings in the repetition of step C. To be concrete, consider the following response: for each subject, take those items rated with a value above 5 in step A, calculate a difference in the rankings between step C and step A and take an average over all such items for the subject. Then this response is found to be significantly greater than zero based on appropriate statistical tests.

The cognitive dissonance theory explains this phenomena by the theory that forcing individuals to specify preferences between particular items in step B (“do you prefer oranges or clementines?”) forces them to recognize inconsistencies in their preferences and to alter them, hence the changes between step A and step C (i.e. the process of going through step B caused the change in preferences seen in step C).

- a. What is the nature of the treatments?
- b. What are the units?
- c. How were units assigned to treatments?
- d. What is the nature of the response?
- e. Do any of the answers to the previous questions parts make it difficult to assess the psychological theory that they’re trying to test? If so, explain why.

To help you answer this last question, consider the follow-up study proposed by a critic of this paper. The critic reruns the above experiment, only changing the order of the steps: step A, step C, step B. They find a similar value of the response (after appropriate statistical tests) as when running the original experiment of step A, step B, step C.

---

<sup>2</sup>Salti et al, “Cognitive dissonance resolution is related to episodic memory,” PLoS One, 2014.

3. **Estimation under Imbalance.** Consider the setting where we seek to estimate the Average Treatment Effect (ATE) using as our estimator the observed difference in the mean responses in the two groups,  $\bar{Y}_1 - \bar{Y}_0$ . Let  $D_i$  track whether unit  $i$  receives the treatment (1) or the control (0) and let the set of indices in the treatment and control groups be  $g_1$  and  $g_0$  respectively. Also let  $n_1$  and  $n_0$  be the number of units in each group.

So far, we've analyzed the case where  $n_1 = n_0$ , a case where the groups are *balanced*. What happens to the estimator when the groups are *unbalanced*?

Determine whether or not the estimator is biased when  $n_1$  is twice the size of  $n_0$ . Units are assigned to groups by randomly shuffling the list of their unique identifiers and assigning the first  $n_1$  to the treatment and the remaining to the control. If the estimator is biased, suggest an alternative bias-corrected estimator.

---

**Sampling Distributions.** Consider the following schedule of potential outcomes corresponding to six students in the anchoring experiment.

```
library(tidyverse)
anchor_mini_sched <- tibble(Y_0 = c(15, 15, 19, 2, 10, 8),
                           Y_1 = c(31, 22, 45, 20, 20, 15))
anchor_mini_sched
```

```
# A tibble: 6 × 2
  Y_0  Y_1
  <dbl> <dbl>
1    15    31
2    15    22
3    19    45
4     2    20
5    10    20
6     8    15
```

4. If the random assignment in the experiment involved assigning indices to the students, randomly permuting a vector of three 1s and three 0s, then assigning the units to either treatment (1) or control (0) based on whether the index in the vector was a 1 or a 0, what is the sampling distribution of  $\widehat{ATE}$ ? Display the distribution as a table with two columns (one for the value of the RV, the other for the probability) and as a plot.
5. Using this sampling distribution, calculate the expected value and variance of  $\widehat{ATE}$ .
6. What is the smallest number of permutations / partitions that you need to consider in order to know the sampling distribution?

7. **Testing Anchoring.** Conduct a hypothesis test that the first question on the Anchoring Experiment had no effect on the guess for the percent of UN Nations that are in Africa. Use  $\alpha = .05$ , consider only the randomness induced by the random assignment to  $X = 11$  or  $X = 73$ , use the sharp null hypothesis, use the difference in group means as your statistic, and use a two-tailed test. Approximate the sampling distribution of the test statistic by first creating the schedule of outcomes implied by the null hypothesis, then use the `rand_stats()` function in the code sample from lecture to calculate 1000 test statistics under the null.

Provide the code that you used, a plot of the sampling distribution (a.k.a. the null distribution), the p-value, and a clearly written interpretation about what this analysis says about our original research question.

Use the full anchoring data set collected on the first day of class:

```
library(tidyverse)
anchoring <- read.csv("https://stat158.berkeley.edu/spring-2026/data/anchoring/
anchoring.csv")
```