

1. **Cancer and Group Therapy II.** Researchers randomly assigned metastatic breast cancer patients to either a control group or a group that received weekly 90-minute sessions of group therapy and self-hypnosis. The group therapy involved discussion and support for coping with the disease. The goal of the experiment was to see whether the latter treatment improved the patients' quality of life at the time, but a followup study on these patients collected data on the number of months of survival after the beginning of the study<sup>1</sup>.

You can access the data from this study using the following code.

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

`GROUP` is the original group assigned to the subject. `SURVIVAL` is the survival time in months from the beginning of the study. At the time the survival data was collected (10 years later), some subjects were still alive. They are flagged in the `CENSOR` column. Data (in this case survival time) that can only be known up to some bound is called *censored*.

- a. The parameter of central interest to researchers is the *ATE*. What is the point estimate of this parameter using this data?
  - b. Form a 95% confidence interval associated the point estimate using the *t* distribution. As before use the pooled estimate of the standard deviation<sup>2</sup>.
  - c. Form a 95% CI using RBI and the constant-effect assumption.
  - d. Methods b. and c. rely upon different assumptions. Describe what they are and which seem more realistic in this particular setting.
2. **Confident in your confidence intervals?** In lecture we formed a randomization-based confidence interval for the difference in mean response between the  $X = 11$  group and the  $X = 73$  group in the anchoring experiment. Your goal for this exercise is to use simulation to check the coverage of your interval and evaluate the consequences of using the constant effect assumption when it is violated.

This simulation requires knowing the value of the true parameter, *ATE*, so you'll be playing this question as an all knowing god with all potential outcomes at your fingertips. You can find two full schedules of potential outcomes at the following links.

```
godlikeA <- read.csv("https://stat158.berkeley.edu/spring-2026/data/anchoring/
godlike_schedule_A.csv")
godlikeB <- read.csv("https://stat158.berkeley.edu/spring-2026/data/anchoring/
godlike_schedule_B.csv")
```

Schedule A represents a world with constant effects: the individual treatment effect for each unit is the same. Schedule B represents a world with variable effects: each unit has a different individual treatment effect.

<sup>1</sup>Data from Q 4.31 in *Statistical Sleuth*.

<sup>2</sup>In R, you can do this either "by hand" using the point estimate plus and minus a value of `qt()` times your estimate of the SE. You can also return to the last time you saw this dataset, in a testing framework. The object created by `t.test()` has bundled with it the associated confidence interval.

- a. For each god-like schedule A and B, calculate the true  $ATE$  parameters.
- b. For god-like schedule A, use simulation to create many original experiments and then for each construct a randomization-based CI with a constant effect assumption and evaluate whether it contains the true parameters. Select a  $1 - \alpha$  confidence level to use that is between 60% and 90%. For A and B separately, report the proportion of confidence intervals that contain the parameter.
- c. Repeat b. but use god-like schedule B.
- d. Did your randomization-based CIs with the constant effect assumption from the previous two questions succeed in capturing the parameter with probability  $1 - \alpha$ ? If not, provide some intuition for why not.

This is a challenging simulation to set up correctly. If you get stuck please consult the hints<sup>345</sup>.

---

<sup>3</sup>The idea is to recreate the random process that leads to the create of the random confidence interval many times and check the proportion of those intervals that contain the parameter. In randomization-based inference, the only randomness comes from the assignments of units to treatment and control at the start of the original experiment, so that's where each one of your simulations begins.

<sup>4</sup>For each of settings A and B above, you will want to write a function. One approach is to write a function that:

1. Randomly assigns each unit to one of the two groups.
2. Sets the potential outcome that is *not* observed to be NA. At this stage, you have the analog of an original experimental data set.
3. Calculates  $\widehat{ATE}$ .
4. Uses  $\widehat{ATE}$  to impute the missing potential outcomes.
5. Using this filled-in schedule, form a randomization-based confidence interval.
6. Check whether the interval contains the true parameter and return that logical value.

Then you can replicate this function many times and count the proportion of TRUEs to estimate the coverage probability. See <https://andrewpbray.github.io/designrbi> for functions to facilitate steps 1-5.

<sup>5</sup>This code may take awhile to run - you're simulating many randomization-based CIs - so start with small numbers of replicates while you're still testing your code. You can also take advantage of your computers ability to run multiple R sessions at the same time using the `future_replicate()` function inside the `future.apply` package.

3. **Growing Herbs II:** Botanists studying optimal greenhouse conditions for a fast-growing herb ran a fully-crossed three-way factorial experiment. They manipulated light intensity (low, medium, or high), watering frequency (low, medium, or high), and soil type (sandy or loam). Each of the  $3 \times 3 \times 2 = 18$  treatment combinations was replicated twice, yielding  $n = 36$  plants. The response is shoot dry weight (grams) measured after four weeks.

Their results from running a full three-way ANOVA model are shown in the table below.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
light	2	190.54	95.27	135.084	1.45e-11	***
water	2	131.53	65.76	93.246	3.17e-10	***
soil	1	22.88	22.88	32.442	2.11e-05	***
light:water	4	33.41	8.35	11.842	6.85e-05	***
light:soil	2	4.98	2.49	3.533	0.0508	.
water:soil	2	1.10	0.55	0.781	0.4727	
light:water:soil	4	6.34	1.59	2.248	0.1043	
Residuals	18	12.70	0.71			
---						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

- How many distinct null hypotheses are being tested in this ANOVA table?
- Consider a decision rule that rejects an individual null hypothesis when the p-value is less than or equal to  $\alpha = .05$ . If we assume that the tests are independent, what is the probability of making at least one Type I error?
- If you wanted to control the strong family-wise error rate at  $\gamma = .05$  using a Bonferroni correction, what would be the new significance threshold for each individual test?
- How many of the tests in the ANOVA table would be considered statistically significant at the  $\alpha = .05$  level? How many would be considered statistically significant at the Bonferroni-corrected threshold you calculated in part c.?
- Consider the proof we went through in class showing that a decision rule that uses the Bonferroni correction keeps the family-wise error rate less than or equal to  $\gamma$ . Under what scenario would this bound be tight? That is, when would the family-wise error rate actually equal  $\gamma$  when using the Bonferroni correction? Does this seem like a realistic scenario in practice? What does that tell you about the actual family-wise error rate when using the Bonferroni correction in practice?

4. **Babies Walking: Blocks.** Consider the Babies Walking study from lecture where babies were randomly assigned to one of four programs and the time it took for them to walk was recorded.
- Why would it be difficult to use *parent ID* as a blocking factor in this study?
  - List two factors that *could* be used as blocking factors in this study. For each factor, why you would expect it to explain variation in walking time.
  - Imagine that one of your blocking factors has five levels, allowing for a complete block design. The data frame below provides the id of each baby in the original study along with their corresponding block. Fill in the program column with a random assignment of the four programs to the babies within each block.

id	block	program
1	A	NA
2	A	NA
3	A	NA
4	A	NA
5	B	NA
6	B	NA
7	B	NA
8	B	NA
9	C	NA
10	C	NA
11	C	NA
12	C	NA
13	D	NA
14	D	NA
15	D	NA
16	D	NA
17	E	NA
18	E	NA
19	E	NA
20	E	NA

5. **Blocking and Variance: Model-based.** In lecture we worked through the exact standard error of  $\widehat{ATE}$  from the Randomization-Based framework. Consider an analogous parametric framing using the following model:

$$Y_i \stackrel{\text{iid}}{\sim} N(\mu_{j(i)}, \sigma_{j(i)}^2)$$

where there are  $J = 2$  groups of size  $n_1$  and  $n_2$ , and  $j(i)$  is the treatment that unit  $i$  belongs to.

- a. What is the variance of  $\widehat{ATE}$  under this model?
  - b. What guidance does this give us in making designs that maximize precision and power? Specifically comment on four quantities inspected in the lecture: the total number of units, the variances of the outcomes in each group, the balance of the design, and covariance between the outcomes in the two groups. Is there any difference in the guidance provided by this model-based framing compared to the randomization-based framing?
6. **Blocking and Variance: Unequal Treatment Allocation.** In the GCB design from lecture, each block assigned exactly half its units to treatment and half to control. Suppose instead that block  $A$  (of size  $n_A$ ) assigns  $n_{A1}$  units to treatment and  $n_{A0} = n_A - n_{A1}$  to control, and similarly block  $B$  assigns  $n_{B1}$  to treatment and  $n_{B0} = n_B - n_{B1}$  to control, where  $n_{A1}/n_A \neq n_{B1}/n_B$ . For these calculations, use the (simpler) model-based framing from the previous question
- a. Is the estimator  $\widehat{ATE} = \frac{n_A}{n} \widehat{ATE}_A + \frac{n_B}{n} \widehat{ATE}_B$  still unbiased for the overall ATE? Justify your answer.
  - b. Write out the full expression for  $SE(\widehat{ATE})$ , substituting in the CR[1] formula for each block-level SE (treating each block as its own completely randomized experiment with potentially unequal group sizes).
  - c. Suppose  $\sigma_1^2 > \sigma_0^2$  within block  $A$ . Based on the CR[1] SE formula, what allocation ratio  $n_{A1}/n_{A0}$  minimizes  $Var(\widehat{ATE}_A)$ ? What does this suggest about how you should design the experiment when you have prior knowledge about within-block variance?