Homework 3

Lee Kennedy-Shaffer

Course Week 8

Show your work and express all answers in lowest terms. You are welcome to use results and definitions from the posted document on the Moodle, but be sure to explain why a result or property can be used if it is not obvious (e.g., if certain conditions must be met, how do we know they are?).

Note: If you use R to get results, please also write the answers in context or make clear which number is which result.

Question 1. Power and Sample Size as a Function of Inputs

Consider a RCT targeting a difference in means of a continuous outcome.

A. Assume that the variance of the outcome in both arms is 1, and that we want 80% power to detect an effect of 0.5 units with a significance level of 0.05. What is the total number of participants (in both arms combined) that we need for this study?

For the remaining parts, use the parameters given above unless otherwise specified. All plots can be done in Excel, R, or another software of your choice. Some sample code in R is given at the end of this assignment.

B. Plot the total sample size required vs. the power, as power ranges from 50% to 99%.

C. Plot the total sample size required vs. the minimum detectable effect, as the effect ranges from 0.1 to 2.0.

D. Plot the total sample size required vs. the standard deviation of the outcome, as the SD ranges from 0.1 to 10.0.

E. Plot the total sample size required vs. the significance level, as the significance level ranges from 0.01 to 0.10.

Question 2: Empirical Power and Simulation for RCTs

Kalla and Broockman (2016) studied the effect of campaign contributions on access to Congressional officials. In that study, the outcomes were binary (whether the constituent met with an official at a certain level or above). Suppose instead we measure a continuous outcome, like a favorability rating the official's office gives to the constituent's organization or issue. We are still limited to 192 total representatives, so we will conduct a 1:1 randomized study with 96 representatives in each treatment arm (we ignore matching and the 1:2 randomization from the real study for this case). Suppose the outcome (favorability rating on a 0–1 scale) has a variance of 0.01 under either control or the intervention. We are interested in a minimum detectable effect size of 0.05 and a significance level of $\alpha = 0.05$.

A. How much power will our study have for this MDE?

B. Simulate this experiment 1000 times assuming the outcome is normally distributed and has a mean of 0.5 under the control condition. Find the empirical expectation of the estimator (average of the estimates across the simulations), empirical variance of the estimator (variance of the estimates across the simulations) empirical power (percentage of the simulations that find a significant effect), and empirical 95% confidence interval coverage (percentage of the simulations where the true treatment effect is within the 95% CI).

*Sample code for the simulation is available at the end of this assignment if you'd like: you'll need to put in the appropriate values everywhere you see **XX**. You don't need to change any other parts of that code to get the simulation output for this part.*

C. Graph a histogram of the distribution of estimates from your simulation. What do you notice?

D. We assumed a variance of 0.01 for the outcome in our design. If the outcome variance is higher than expected, what should happen to our empirical expectation, variance, power, and coverage?

E. Re-do the simulation with an outcome variance of 0.02. How do the empirical operating characteristics change?

F. We also assumed a normal distribution in the design. If instead the distribution is skewed, we can see how that affects our operating characteristics. Change the distribution in the simulation from a normal distribution with control mean 0.5 and variance 0.01 to a log-normal distribution with control mean 0.5 and variance 0.01 (you can do this by re-copying the `One.Sim` function, commenting out the `rnorm()` lines and uncommenting the `rlnorm()` lines, and using the correct means and variances). Both arms should have outcome variances of 0.01 again. Note that the `rlnorm` lines include functions to find the parameters that give the same mean and variance. Re-do the simulation. How do the empirical operating characteristics change?

G. From the simulations in E and F, do you think it is more important to correctly estimate the outcome variance or the distribution of the outcome in the design phase? Explain why to someone who has taken probability/statistics but not study design.

Question 3: Drop-Out, Missing Data, and Bias in RCTs

In this problem, we're going to see how study participants dropping out of a trial can cause bias, especially if the treatment assignment is unmasked. We will see bias both under the null hypothesis and under the alternative hypothesis when there is differential drop-out.

Suppose we are running an RCT of a new migraine medicine. The endpoint of interest is Y, the number of migraines the trial participants have per week, which ranges from 0 to 4 for everyone in the trial. Our estimand will be $\theta = \mu_1 - \mu_0 = E[Y_1] - E[Y_0]$, the difference in population means if everyone took the new intervention vs. everyone taking the placebo.

In the placebo group, the distribution of outcomes follows a Poisson distribution with parameter $\lambda = 2$: $Y_0 \sim Pois(2)$. To start, imagine that the new medicine has no effect at all, so the treatment group has the same distribution of outcomes as the placebo group. Then $\theta = 0$.

We run a trial by randomly sampling 50 people from the population into each arm (100 people total) and recording their outcomes after treatment. The estimator is the average outcome among the people in the intervention arm minus the average outcome among the people in the control arm.

A. Show that $\hat{\theta}$ is unbiased for θ and find $Var(\hat{\theta})$ and $SD(\hat{\theta})$.

B. Suppose that the trial is open-label (no masking of treatment assignment). The individuals in the placebo group who still have 4 or more migraines per week drop out of the trial so they can try a different treatment and do not report their outcomes. If the estimator just averages the remaining outcomes in the control arm and subtracts them from the average outcome in the intervention arm, what will the bias be? Does this estimator, on average, make the treatment look more or less effective than it really is? **Hint:** find $E[Y|Y < 4]$ when $Y \sim Pois(2)$.

C. Instead of what happens in part B, now suppose that the new medicine causes side effects. Suppose that 10 random individuals from the intervention arm get bad side effects and drop out of the trial. The probability of someone getting side effects is independent of their migraine frequency. If the estimator just averages the remaining outcomes in the intervention arm and subtracts the average outcome in the control arm, what will the bias be? What will the variance be?

For the remaining parts, suppose that the new medicine is effective. The distribution of outcomes among people in the intervention arm will be Poisson distributed with parameter $\lambda = 1.5$: $Y_1 \sim Pois(1.5)$. The distribution in the placebo arm is the same as it was previously.

D. What is the true value of the estimand?

E. If there is no drop-out, what will the bias and variance of the estimator be?

F. If the placebo arm members with 4 or more migraines per week drop out (as in B), what will the bias be? Does the drug look more or less effective than it really is?

G. If instead a random 10 individuals from the intervention arm drop out (as in C), what will the bias and variance be?

Question 4: Difference in Maximum Estimand

We have focused so far on estimands that compare the mean of an outcome (or, equivalently, proportion of a binary outcome) between the two arms. In some trials, however, interest might lie with a different quantity. For example, if we're comparing the run times of two computer algorithms, we might want to know the difference between the maximum run times for each algorithm. In this question, we're going to see how to find the operating characteristics of an estimator for this type of estimand.

The design is that we run each algorithm 10 times (randomizing time, machine, inputs, etc. to get exchangeability) and use the estimator $\hat{\theta} = \max Y_{1i} - \max Y_{2j}$, where $i, j \in \{1, 2, \ldots, 10\}$. To ease notation, you can let $X_1 = \max Y_{1i}$ be the max from algorithm 1 and $X_2 = \max Y_{2j}$ be the max from algorithm 2.

Assume that under both algorithms, each observation will be i.i.d. following a continuous uniform distribution with minimum value m (the same for both algorithms) and maximum value n_1 or n_2 for algorithm 1 or 2, respectively. That is: $Y_{1i} \overset{i.i.d.}{\sim} Unif(m, n_1)$ and $Y_{2j} \overset{i.i.d.}{\sim} Unif(m, n_2)$. You can further assume that all times for algorithm 1 are independent of all times from algorithm 2. The estimand we're interested in is the difference in maximums: $\theta = n_1 - n_2$.

A. Find the cumulative distribution function for $X_1 = \max Y_{1i}$. Be sure to note the support of X_1. *Hint:* We need to find $F_1(x) = P[X_1 \leq x]$, which is equal to the probability that all 10 observations from algorithm 1 are less than or equal to x. Use the independence of these observations.

B. Find the probability density function for X_1.

C. Use the PDF to find the expectation and variance of X_1. You can solve the integrals by hand (*hint:* replace x by $(x - m) + m$ in the term that's multiplied by the PDF before integrating) or using an online calculator. But do simplify as much as possible and leave constants as fractions rather than rounded decimals.

D. State $E[X_2]$ and $Var(X_2)$ by analogy to the results in C (you don't have to redo all of the work). Then use those to find $E[\hat{\theta}]$ and $Var(\hat{\theta})$.

E. Explain what happens to $E[\hat{\theta}]$ as the sample size gets very large (again, no need to redo the work, just look at what would change in your expectation as the sample size gets larger). *Hint:* $\hat{\theta}$ is what we call an **asymptotically unbiased** estimator of $n_1 - n_2$.

F. We can **de-bias** this estimator by multiplying it by a factor that depends only on the sample size. If we do this, what is the expectation and variance of the debiased estimator?

OPTIONAL: Do you think $\hat{\theta}$ is approximately normally distributed? Why or why not? You can simulate to test it: try it both under the null hypothesis and if there is a true difference.

Question 5: Ethics and Feasibility Reading Responses

Read one of the following four articles:

- Jones CP. Invited commentary: 'Race,' racism, and the practice of epidemiology. *Am J Epidemiol* 2001;154(4): 299–304.
  Available at https://doi.org/10.1093/aje/154.4.299.
- Reverby SM. Ethical failures and history lessons: The U.S. Public Health Service research studies in Tuskegee and Guatemala. *Publ Health Rev* 2012;34(1): 13.
  Available at https://publichealthreviews.biomedcentral.com/articles/10.1007/BF03391665.
- London AJ, Seymour CW. The ethics of clinical research: Managing persistent uncertainty. *J Am Med Assoc* 2023;329(11): 884–885.
  Available at http://doi.org/10.1001/jama.2023.1675.
- Wilson A, Kasina F, Nduta I, Ayumbah Akallah J. When economists shut off your water. *Arica Is a Country* (blog) Nov. 2023.
  Available at https://africasacountry.com/2023/11/when-economists-shut-off-your-water.

In 2 paragraphs, write a reflection on the implications of your chosen article for study design, specifically considering the interaction of ethics, feasibility, and statistical goals of study designs. What are the implications of the article for future study designs and how do those implications affect our statistical goals? Be sure to include specific references (including quotations and page numbers where applicable) to points discussed in the article.

Final Project Group

For the final project, you will work with a group of students in the course to design (but not actually run) a study on any topic of your choice. Select a group of 3–4 total students in the class for your project and submit that in/with this homework. If you would like me to place you with a group (or have open spots in your group and would be open to adding additional students), please note that instead/in addition. The full assignment will be posted soon.

Sample R Code for Question 1

```
## Create a table with a column for the parameter you want to vary:
DF1 <- tibble(Power=seq(0.5,0.99,by=0.01))

## Create another column in that table with the sample size needed
## for each value of that parameter (the XXs will need to be filled in)
DF1$Sample.Size <- sapply(X=DF1$Power, FUN=function(value)
    2*ceiling(power.t.test(delta=XX, sd=XX, sig.level=XX, power=value,
    type="two.sample", alternative="two.sided")$n))

## Plot the results:
ggplot(data=DF1) + geom_line(mapping=aes(x=Power, y=Sample.Size)) +
  theme_bw() +
  labs(title="Required Sample Size vs. Power", y="Total Required Sample Size")
```

Sample R Code for Question 2

```
# Fill in the appropriate parameters:
SS.PerArm <- XX
Mean.Treatment <- XX
Mean.Control <- XX
Variance.Treatment <- XX
Variance.Control <- XX
Num.Sims <- XX

set.seed(XX)

## This function will simulate the trial once using a normal distribution
## and return the estimate, variance, p-value, and lower and upper CI bounds.
One.Sim.Normal <- function(ssarm, mu.1, mu.0, sigma2.1, sigma2.0) {
  Data_Trt <- rnorm(n=ssarm, mean=mu.1, sd=sqrt(sigma2.1))
  Data_Ctrl <- rnorm(n=ssarm, mean=mu.0, sd=sqrt(sigma2.0))
  test <- t.test(x=Data_Trt, y=Data_Ctrl, paired=FALSE,
                 alternative="two.sided", conf.level=0.95)
  return(c(Estimate=mean(Data_Trt)-mean(Data_Ctrl),
           Variance=var(Data_Trt)/SS.PerArm+var(Data_Ctrl)/SS.PerArm,
           PValue=test$p.value,
           CI.Lower=test$conf.int[1],
           CI.Upper=test$conf.int[2]))
}


## This function will simulate the trial once using a log-normal distribution
## and return the estimate, variance, p-value, and lower and upper CI bounds.
One.Sim.LN <- function(ssarm, mu.1, mu.0, sigma2.1, sigma2.0) {
  Data_Trt <- rlnorm(n=ssarm, meanlog=log(mu.1)-log(sigma2.1/mu.1^2+1)/2,
                     sdlog=sqrt(log(sigma2.1/mu.1^2+1)))
  Data_Ctrl <- rlnorm(n=ssarm, meanlog=log(mu.0)-log(sigma2.0/mu.0^2+1)/2,
                      sdlog=sqrt(log(sigma2.0/mu.0^2+1)))
  test <- t.test(x=Data_Trt, y=Data_Ctrl, paired=FALSE,
                 alternative="two.sided", conf.level=0.95)
  return(c(Estimate=mean(Data_Trt)-mean(Data_Ctrl),
           Variance=var(Data_Trt)/SS.PerArm+var(Data_Ctrl)/SS.PerArm,
           PValue=test$p.value,
           CI.Lower=test$conf.int[1],
           CI.Upper=test$conf.int[2]))
}


## This function will run the simulation Num.Sims times and compile all
## of the returned values into a tibble. Then you can use summarize to
## get the empricial operating characteristics from that.
Results1 <- as_tibble(t(replicate(n=Num.Sims,
                      One.Sim.Normal(ssarm=SS.PerArm,
                              mu.1=Mean.Treatment, mu.0=Mean.Control,
                              sigma2.1=Variance.Treatment,
                              sigma2.0=Variance.Control)))))
```