Ch. 5 Optimal Experiment Design Notes

Dylan Baker

October 2024

1 Introduction

A few key definitions:

- "The significance level, also known as the probability of Type I error, is the probability of falsely rejecting the null hypothesis (i.e., a 'false positive'). "
- "The minimum detectable effect size is the magnitude of the treatment effect that the analyst desires to detect."
- "The statistical power of a hypothesis test is the probability of detecting an effect if there is an effect to be detected; or stated differently, it is the probability that we can reject the null hypothesis."

2 Variance of the ATE Estimator

In the model

$$Y_{iT} = \alpha_i + X_i\beta + \bar{\tau}D + \tau_i D + \varepsilon_i$$

where $\bar{\tau}$ is the ATE.

The variance of the ATE estimator is given by:

$$\text{var}(\hat{\tau}) = \frac{\sigma^2}{N} = \frac{\text{var}(\varepsilon)}{N * \text{var}(D)} \tag{1}$$

This expression speaks to ways that we may achieve more power. "The first, N, is well-known and commonly discussed as advice on how to achieve more power: 'obtain a larger sample size if you want more precise estimates' is what every student who wants to conduct an experiment learns. The second choice, $\text{var}(D)$, is less often discussed. What equation (1) shows is that inducing more variation in your treatment variable leads to more precise estimates, and the rate of increased precision is identical to how precision changes in N."

3 Sample Mean Qualities

Assume "that $Y_{i0} \mid X_i \sim N\left(\mu_0, \sigma_0^2\right)$ and Y_{i1} if $D = 1$ where $Y_{i1} \mid X_i \sim N\left(\mu_1, \sigma_1^2\right)$."

Then, the sample means must satisfy:

$$\frac{\overline{Y_1} - \overline{Y_0}}{\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}} = t_{\alpha/2} \Rightarrow \bar{Y}_1 - \bar{Y}_0 = t_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}$$

and:

$$\frac{\overline{Y_1} - \overline{Y_0} - MDE}{\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}} = -t_\beta \Rightarrow \bar{Y}_1 - \bar{Y}_0 = MDE - t_\beta \sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}$$

Then the minimum detectable effect size is given by:

$$\text{MDE} = \left(t_{\alpha/2} + t_\beta\right) \sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}$$

"These parameters determine the smallest value of $|\bar{\tau}| > 0$ for which the experiment will correctly reject the null hypothesis with probability $1 - \beta$ at significance level α."

"If $\sigma_1^2 = \sigma_0^2 = \sigma^2$ and we define $N = n_0 + n_1$," then we can re-write this as:

$$\text{MDE} = \left(t_{\alpha/2} + t_\beta\right) \sqrt{\frac{1}{P(1-P)} \frac{\sigma^2}{N}}$$

"Where $\frac{1}{P(1-P)} \frac{\sigma^2}{N}$ is the exact sample variance of the treatment effect estimator $\text{var}(\hat{\tau})$.

smallest sample sizes that solve the equality in equation (5.6) satisfy $n_0 = n_1 = n$ are given by:

$$n_0^* = n_1^* = n^* = 2\left(t_{\alpha/2} + t_\beta\right)^2 \left(\frac{\sigma}{MDE}\right)^2$$

If the variances of the outcomes are not equal, this solution becomes

$$N^* = \left(\frac{t_{\alpha/2} + t_\beta}{MDE}\right)^2 \left(\frac{\sigma_0^2}{\pi_0{}^*} + \frac{\sigma_1^2}{\pi_1{}^*}\right)$$

With $\pi_0^* = \frac{\sigma_0}{\sigma_0 + \sigma_1}, \pi_1{}^* = \frac{\sigma_1}{\sigma_0 + \sigma_1}$ and $N = n_0 + n_1, \pi_0 + \pi_1 = 1, \pi_0 = \frac{n_0}{n_0 + n_1}$"

Note that the decision about n_0 and n_1 is also determined by the variance of the outcomes under each condition, where a larger variance would warrant a larger sample size.

"Our optimal design equations reveal several features that the experimenter must consider. First, the optimal sample size increases proportionally with the variance of outcomes... A second insight that the design equations yield is that the optimal sample size increases nonlinearly with the significance level and the power of the test... A third insight from our optimality rules is that the optimal sample size decreases proportionally with the square of the MDE."

4 Multi-Valued Explanatory Variable

Recall

$$\text{var}(\hat{\tau}) = \frac{\text{var}(\varepsilon)}{N * \text{var}(D)}$$

then if we believe that the treatment effect is linear, then we maximize var(D) by putting half the sample in $D = 0$ and half in $D = \bar{D} > 0$.