Power Analysis

Tiffany Cheng, Matt Kinkley, Prachi Varma

June 12, 2022

In this experiment, we are examining the effect of stereotype threat on the test performance of females vs. males. We identified our experiment as potentially having heterogeneous treatment effects, so we will be using a regression model with two main effects (gender and treatment) and an interaction effect. In particular, we believe that our treatment (asking participants for their gender prior to administering a math test) will affect females and males differently. In addition, we will be assessing the significance of the interaction effect using robust standard errors.

Through reading previous research on stereotype threat, we discovered that the treatment effect was largest when the math test only contained difficult questions, so we will be only including difficult math questions on our exam. From previous studies, we found that the difference in exam scores between males and females in treatment was about 20 percentage points, while there was little to no difference in exam scores between males and females in control. Since these results were from a study where the participants were told that the exam produced differences between females and males, the treatment effect may be higher. For this reason, we are also testing lower treatment effects since we are not including that statement.

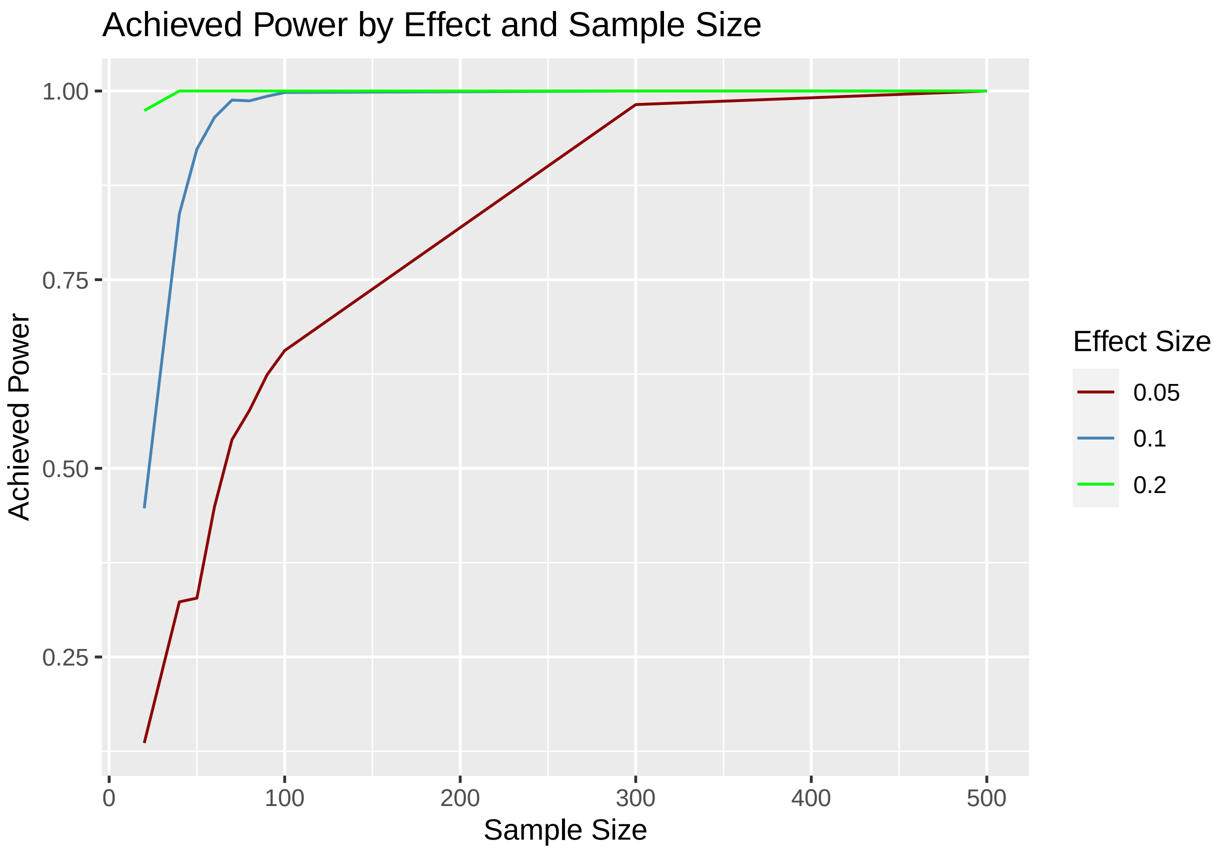
```r
samples <- c(20, 40, 50, 60, 70, 80, 90, 100, 300, 500)

find_power <- function(effect_size) {

  power <- NA

  for (j in 1:length(samples)) {

    p_values <- NA

    for (k in 1:1000) {

      d <- data.table(id = 1:samples[j])

      d[id <= samples[j] / 2, treatment := "treatment"]
      d[id > samples[j] / 2, treatment := "control"]
      d[id %% 2 == 0, gender := "male"]
      d[id %% 2 == 1, gender := "female"]

      d[treatment == "control" | gender == "male", score := rnorm(.N, mean = 0.4, sd = 0.05)]
      d[gender == "female" & treatment == "treatment", score := rnorm(.N, mean = 0.4 - effect_size, sd =

      d[ , gender := as.factor(gender)]
      d[ , gender := relevel(gender, ref = "male")]

      d[score < 0, score := 0]
      d[score > 1, score := 1]
```

```
      mod <- d[ , lm(score ~ as.factor(treatment) + as.factor(gender) + as.factor(treatment) * as.facto

      coef_test <- coeftest(mod, vcov. = vcovHC(mod))

      inter_pvalue <- coef_test["as.factor(treatment)treatment:as.factor(gender)female", "Pr(>|t|)"]

      p_values[k] <- inter_pvalue

    }

    power[j] <- mean(p_values < 0.05)

  }

  return(power)

}
```

```
power_0.05 <- find_power(0.05)
power_0.1 <- find_power(0.1)
power_0.2 <- find_power(0.2)
```

```
power_chart <- data.frame(
  sample_size = rep(samples, times = 3),
  achieved_power = c(power_0.05, power_0.1, power_0.2),
  effect_size = as.factor(rep(c("0.05", "0.1", "0.2"), each = 10))
)
```

```
ggplot(data = power_chart, aes(x = sample_size, y = achieved_power, color = effect_size)) +
  geom_line() +
  labs(title = "Achieved Power by Effect and Sample Size", x = "Sample Size", y = "Achieved Power", colo
  scale_color_manual(labels = c("0.05", "0.1", "0.2"), values = c("darkred", "steelblue", "green"))
```



To achieve above 80% power, we would need 200 subjects for an effect size of 0.05, 40 subjects for an effect size of 0.1, and less than 20 subjects for an effect size of 0.2. If we assume our treatment effect for our experiment is 0.1, then we would need at least 40 subjects. If we anticipate our effect size to be very small, then we would need to obtain around 200 subjects.