# Power Analysis - Implicit Gender Bias in Online-Hiring Experiment

Chris McClure-St. Sunayna Jain, J. Claira Kauffmann, Edward Junprung

2023-02-18

```r
#This is the code for Scenario 3 - blocking.
#Scenario 2 and 1 used a condensed version of this code, pulling the p-values for
#the treatment coefficients.

analyze_power_blocking <- function(control_response_rate_low, treatment_response_rate_low,
                          control_response_rate_high, treatment_response_rate_high,
                          n, num_loops, treat_proportion=0.5){

  p_vals <- rep(NA, num_loops)

  for(i in 1:num_loops){
    # set up data
    treat_size <- round(n * treat_proportion, 0)
    ctrl_size <- n - treat_size
    treat_x <- rep(1, treat_size)
    ctrl_x <- rep(0, ctrl_size)
    group_treat_low<-(rep(c('low_wage'),ceiling(treat_size/2)))
    group_treat_high<-(rep(c('high_wage'),floor(treat_size/2)))
    group_ctrl_low<-(rep(c('low_wage'),ceiling(ctrl_size/2)))
    group_ctrl_high<-(rep(c('high_wage'),floor(ctrl_size/2)))
    treat_y_low <- rbinom(ceiling(treat_size/2), 1, treatment_response_rate_low)
    treat_y_high <- rbinom(floor(treat_size/2), 1, treatment_response_rate_high)
    ctrl_y_low <- rbinom(ceiling(ctrl_size/2), 1, control_response_rate_low)
    ctrl_y_high <- rbinom(floor(ctrl_size/2), 1, control_response_rate_high)

    #create a table
    d <- data.table(condition = c(treat_x, ctrl_x),
                    outcome = c(treat_y_low,treat_y_high,
                    ctrl_y_low,ctrl_y_high), group=c(group_treat_low, group_treat_high,
                                                     group_ctrl_low,group_ctrl_high))
    # make the linear model
    mod <- lm(outcome ~ condition + as.factor(group), data = d)
    s<- summary(mod)

    # calculate the f-test p-value and add to the vector
    p_vals[i] <- pf(s$fstatistic[1], s$fstatistic[2], s$fstatistic[3], lower.tail = FALSE)
  }
  # return proportion of p values <= .05
  mean(p_vals <= 0.05, na.rm = TRUE)


}
```

**Scenario 1:** We plan to apply to jobs posted online, and record the number of positive responses for female candidates (treatment) versus male candidates (control). 'Positive response' is defined as a follow-up email asking for an interview. For our baseline response rate, we referenced CareerPlug's third annual Recruiting Metrics Report, which found an average applicant-to-interview conversion rate between 15-20%. [1]

For our estimated treatment effect (resumes from females), we saw studies that found no implicit gender bias in hiring accross the spectrum of industries. However, certain industries such as tech are male domminated, and we believe that this implicit bias could bleed in certain fields. [2] So, for this senario we estimated a -8% ATE when women apply to tech related jobs, which we simulated by pulling from a binomial distributioon (1 for a positive response; 0 for no response).

lm(outcome ~ condition); ATE -8%

**Scenario 2:**

Aas mentioned above, one study found little to no evidence of implicit gender bias, however it focused on the international market, whereas we are only concerned with the US job market. [3] Still, this gives us reason to believe that our ATE in scenario 1 may be too big. This concern is heightened by a more conservative statistic on applcation-success-rate of 2 to 3.4%. [4]. So, for scenario 2, we analyzed the power of our model with a more conservative ATE.

lm(outcome ~ condition); ATE -4%

**Scenario 3:**

There is a possibility that gender bias could vary depending on the industry. In one study, they found that "female applicants were favoured in female-dominated occupations. . . ." [5] This¸ could lead to our treatment effect being cancelled out if we applied to a broad spectrum of occupations. For this power analysis, then, we blocked on low-wage and high-wage occupations (low-wage jobs such as service industry, clerical work, etc., are often seen as stereotypical female occupations):

lm(outcome ~ condition + as.factor(group), data = d); ATE: -5% high-age, +4% low-wage

Here we had a gender bias favoring men in high-wage jobs and female in low-wage jobs. Without blocking these effects would cancel each other out.

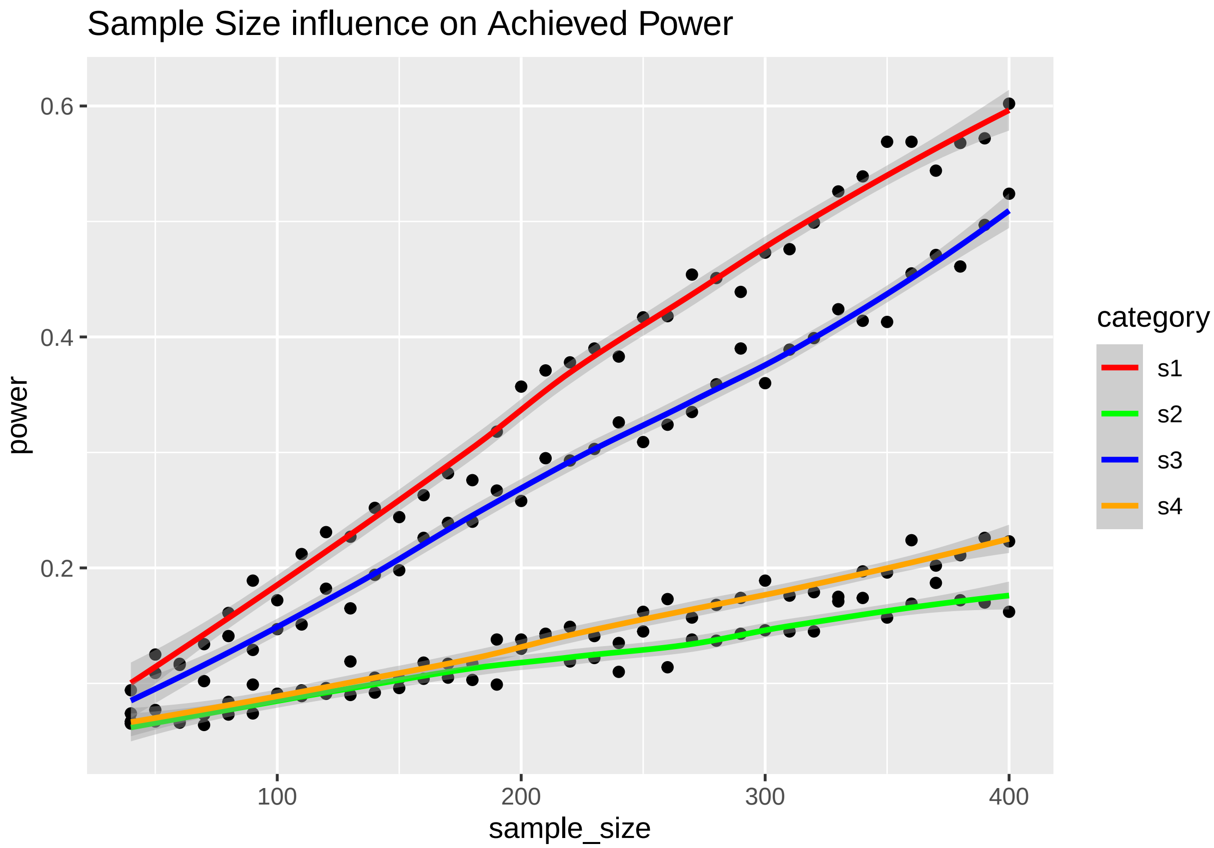
**Scenario 4:**

We heard annecdotal accounts of implicit bias against candidates with gender ambigous names. This is when the person responsible for hiring is unclear if a candidate is male or female based on a name (and not suggesting the candidate is non-binary). This potential phenomenon intrigued us as we could find no study that explored it. So, for our final scenario, we did a multi-treatment analysis by including gender-ambiguous-name resumes.

lm(outcome ~ condition + group, data = d); ATE: -4% ambiguous, -8% female

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

[1] https://www.careerplug.com/wp-content/uploads/2022/03/2022-Recruiting-Metrics-Report.pdf
[2] https://www.science.org/content/article/science-still-seen-male-profession-according-international-study-gender-bias
[3] https://academic.oup.com/esr/article/38/3/337/6412759
[4] https://www.prosperityforamerica.org/resume-statistics
[5] https://academic.oup.com/esr/article/38/3/337/6412759


Sample Size influence on Achieved Power
0.6
0.4
0.2
power
100
200
300
400
sample_size
category
s1
s2
s3
s4