

Exercise 1

2024-01-17

Exercise 1

Group 1: Angel Oluwole-Rotimi, Boyan Li, Freddy Chen, Jessica Song, Tomy Pelletier

Google study : Find the optimal number of interviews for hiring

Experimental Design:

$O_1, O_2, O_3, O_4, O_5, O_6, \dots$ Final Each observation O_i represents the score given to a candidate after each interview.

The experimental design in this scenario is a sequential observational study. This approach, which differs from a traditional experimental design involving control and treatment groups, is more aligned with data-driven decision-making based on observational data analysis.

Step 1: Carlisle collected these scores for dozens of candidates across all their interview stages.

Step 2: Carlisle then analyzed this data to find a point where additional interviews did not significantly change the candidate's average score. This involves looking for a convergence point in the scores across multiple interviews.

One possible method: For each candidate, calculate the cumulative average score up to each interview. For instance, after the first interview, the average score is just the score from the first interview. After the second interview, it's the average of the first and second interview scores, and so on. The key analysis involves looking at how the average score changes (or converges) as more interviews are conducted. The hypothesis is that after a certain number of interviews, the average score will stabilize, indicating that subsequent interviews do not significantly change the overall assessment of the candidate. Graphical Analysis: Plotting the average scores against the number of interviews for multiple candidates can visually reveal the point of convergence. The optimal number of interviews is identified at the point where additional interviews do not meaningfully change the candidate's average score. This is where the curve of the average score versus the number of interviews begins to flatten.

Step 3: To ensure that the findings are not candidate-specific, this analysis should be replicated across many candidates. This helps in determining whether the observed convergence point is consistent across different candidates.

Threats to Validity

1. Selection Definition: Systematic differences in respondent characteristics across conditions could cause the observed effect. Analyzing dozens of past interview processes without consideration to random assignment could mean systematic differences in characteristics of the different groups. We're not sure exactly that there are groups, because everyone received the same original interview style. This seemed to specifically be about finding the lowest ("optimal") interview count at which the interviewee score converged. Possible Solution: Divide the candidate pool into groups based on relevant characteristics (such as job type, experience level, education, etc.) and analyze the interview scores within these groups. This helps control for differences in candidate characteristics that might affect their scores.

2. Instrumentation Definition: The nature of a measure may change over time or over conditions in a way that could be confused with a treatment effect. There may have been a change in conditions (e.g. style) over time in the dozens of studies that were included in the analysis, even if they were the same number of interviews. Possible Solution: Standardize the scoring metrics used in different interviews to ensure consistency. This could involve aligning scoring scales or criteria across all interviews.

Managerial Challenge :

```
performance_data =read.csv("/Users/apple/Desktop/GitHub/MMA-Talent-Analytics/E1/performance_data.csv")
attach(performance_data)
summary(performance_data)
```

```
##      day      worker1      w1_intervention      worker2
## Min.   : 1.00   Min.   : -2.90994   Length:100   Min.   : -3.06956
## 1st Qu.: 25.75   1st Qu.: -0.76691   Class :character   1st Qu.: -0.75374
## Median : 50.50   Median : 0.01285   Mode  :character   Median : 0.01418
## Mean   : 50.50   Mean    : -0.06214           Mean    : 0.03260
## 3rd Qu.: 75.25   3rd Qu.: 0.66558           3rd Qu.: 0.84170
## Max.   :100.00   Max.    : 3.25155           Max.    : 3.42847
## w2_intervention      worker3      w3_intervention
## Length:100          Min.   : -2.469278   Length:100
## Class :character    1st Qu.: -0.887326   Class :character
## Mode  :character    Median : 0.002095   Mode  :character
##                      Mean    : -0.119809
##                      3rd Qu.: 0.664347
##                      Max.    : 2.219876
```

```
max(worker1)
```

```
## [1] 3.25155
```

```
# Assuming your data frame is named df and the column of interest is w2_intervention

# Count the number of "None"
num_none <- sum(performance_data$w2_intervention == "None")
# Count the number of "A"
num_a <- sum(performance_data$w2_intervention == "A")

# Print the results
print(paste("Number of 'None':", num_none))
```

```
## [1] "Number of 'None': 87"
```

```
print(paste("Number of 'A':", num_a))
```

```
## [1] "Number of 'A': 13"
```

```
12/13
```

```
## [1] 0.9230769
```

#On a total of 13 interventions, 12 of them resulted in the performance going down the next day for treatment A.

```
# Count the number of "None"
num_none <- sum(performance_data$w3_intervention == "None")

# Count the number of "A"
num_b <- sum(performance_data$w3_intervention == "B")

# Print the results
print(paste("Number of 'None':", num_none))
```

```
## [1] "Number of 'None': 87"
```

```
print(paste("Number of 'B':", num_b))
```

```
## [1] "Number of 'B': 13"
```

Key Takeaways Findings

1. On a total of 13 interventions, 12 of them resulted in the performance going down the next day for treatment A. (12/13=92%)
2. The only successful intervention is on day 49-50, and it stands out because one intervention was also conducted on day 48. So we can explore the idea that 2 interventions in a row can increase performance or maybe it is not related.
3. For group B, intervention made on a day where results were negative resulted in a performance improvement the day following it.
4. The only scenario where the expert is right saying that A helps employees and B harms it is if we consider the mean for both groups. Group A has a better average performance score than group B. This ties back to the concept of human beings being better at identifying correlation and mixing it with a causal effect.