BUAN 6392.0W1 Assignment 1

Group 10

Kushal Ayyagari Satyanarayana
Melissa Cunningham
Shubham Hota
Rohith Thotakura
Faraz Mohammed Mumtaz

1. The HMRC wants to collect taxes from delinquent taxpayers and employs the BIT (behavioral insights team) to assist. We are given some information on a hypothetical taxpayer, John Smith. One approach for changing the letter would be to encourage him to pay his taxes using facts and comparisons against other nearby taxpayers. In this case, 98% of citizens had paid their taxes, and 95% of citizens in John's hometown had paid. Future, his neighbors and parents both paid on time. The letter would point out these facts and remind John of his delinquent status.
   This strategy would encourage the taxpayer to join the status quo. If his peers are complying with paying taxes, why isn't he? It is a form of social pressure to join the crowd and be a normal, compliant citizen. This strategy focuses more on the positive aspect of behavior. Change your behavior to be regarded positively by your peer group, and do the right thing by paying your taxes.

2. The HMRC should use an A/B test to evaluate the success of the proposed letter. The control group would receive the current letter format, and the test group would receive the new letter format. It would be essential to assign these letters randomly. With unknown outcomes, it would make sense to keep the treatment group sample size as small as possible in case of a worse outcome.
   Once the test is performed, they can compare the results of the two groups using a t-test and determine if there are significant changes between the two groups. The important hypothesis does the new letter significantly increases tax payment collections.

3.

(i) What is the impact of stay-at-home orders due to COVID-19 on economic activity in a state?

a. What is the outcome variable of interest, and how would you characterize the "treatment"?
b.Indicate plausible sources of selection or omitted-variable bias that lead you to believe that direct comparisons of outcomes from observational data would *not* yield a good measure of the causal effect.

The impact of stay-at-home orders due to COVID-19 on economic activity in a state can vary greatly depending on various factors, including the specific industries and businesses affected by the orders, the resources and support available to affected workers and businesses, and the overall state of the economy before the pandemic. In general, stay-at-home orders can result in decreased economic activity, as they often limit the ability of individuals and businesses to operate as usual and result in job losses, reduced consumer spending, and other economic challenges. However, the magnitude of these impacts will depend on the specific circumstances in each state and the effectiveness of government support and stimulus measures.

The outcome variable of interest would be the level of economic activity in a state before and after implementing stay-at-home orders due to COVID-19. The treatment would be implementing the stay-at-home order to determine its effect on the state's economy.

Several sources of selection or omitted-variable bias could affect the validity of a causal inference from observational data:

Internal problem in treatment selection: Some states may have implemented stay-at-home orders earlier or later than others based on factors that are also associated with economic activity, such as the level of COVID-19 infections.

Nonrandom treatment assignment: Some states may have more strict enforcement of stay-at-home orders than others, which could affect the level of economic activity.

Confounding variables: Other factors that vary across states, such as the local economy, state policies, and consumer behavior, could affect the implementation of stay-at-home orders and the level of economic activity.

Self-selection bias: Some individuals and businesses may have chosen to stay at home or close down even before implementing stay-at-home orders, which could affect economic activity.

These biases suggest that direct comparisons of outcomes from observational data would not yield a good measure of the causal effect of stay-at-home orders on economic activity. Alternative approaches, such as natural experiments or randomized controlled trials, may better address these biases to estimate the causal effect.

(ii)Does sponsored search advertising on Google affect clicks for a website?

a. What is the outcome variable of interest, and how would you characterize the "treatment"?

b.Indicate plausible sources of selection or omitted-variable bias that lead you to believe that direct comparisons of outcomes from observational data would *not* yield a good measure of the causal effect.

The outcome variable of interest would be the number of clicks on a website.
The treatment would be the presence or absence of sponsored search advertising on Google. In other words, the variable of interest would be the number of clicks between the control group(websites without sponsored search advertising) and the treatment group (websites with sponsored search advertising). The treatment can be considered as a binary variable, taking on the value of 1 for websites with sponsored search advertising and 0 for websites without.

The outcome variable of interest will be the presence or absence of sponsored ads on google. The variable of interest will be the difference in the number of clicks between the control group(websites with no sponsored search ads) and the treatment groups(websites with sponsored ads).
Selection-omitted variable bias-> sponsored search ads might be targeted to certain
websites that choose to invest in sponsored search ads on google may have different characteristics such as large budget, attractive products, and different audiences.
The content of the website also matters popularity of ads, and the website might change due to various factors

4. See below:
a. The null hypothesis is that the proportion of callbacks for resumes sent out with white-sounding names (p1) is equal to the proportion of callbacks for resumes sent out with African American names (p2).

b. **The difference in means is**
$p_1 - p_2 = 0.12 - 0.04 = 0.08$
c. **The test statistic is 16.33, and the p-value is almost 0.**

```
# Define the sample proportions
p1 <- 0.12
p2 <- 0.04

# Define the sample sizes
n1 <- 6000
n2 <- 6000

# Calculate the standard error
SE <- sqrt((p1 * (1 - p1) / n1) + (p2 * (1 - p2) / n2))
p=(p1+p2)/(n1+n2)

#SE1<-sqrt(p*(1-p)*(1/n1+1/n2))

# Calculate the t-statistic
t <- (p1 - p2) / SE



# Print the results
cat("t-statistic:", t, "\n")
t <- 16.32
df <- n1 + n2 - 2
p_value <- 2*pt(t, df, lower.tail = FALSE)
p_value
```

```r
t <- 16.32
df <- n1 + n2 - 2
p_value <- 2*pt(t, df, lower.tail = FALSE)
p_value
```

```
> SE <- sqrt((p1 * (1 - p1) / n1) + (p2 * (1 - p2) / n2))
> p=(p1+p2)/(n1+n2)
> # Calculate the t-statistic
> t <- (p1 - p2) / SE
> # Print the results
> cat("t-statistic:", t, "\n")
t-statistic: 16.32993
> t <- 16.32
> df <- n1 + n2 - 2
> p_value <- 2*pt(t, df, lower.tail = FALSE)
> p_value
[1] 3.086338e-59
>
```

d. Since the p-value is < 0.05, we can reject the null hypothesis that the callback proportion is the same for white names vs. African American names. We accept the alternate hypothesis that the difference between callbacks between people with African - American names and white names is significant.

e. This test tells us that there was a significant difference in callback proportions for the two groups in this sample. It implies there could be discrimination in the labor market, and more research would need to be done to conclude definitive discrimination. All the submitted resumes were qualified candidates. There could be other sources of selection or omitted-variable bias that could affect the outcome.

5. For this question, a 3% improvement in the existing conversion rate is 0.2% * 1.03 = 0.206%.

   The number of impressions required per group to detect a 3% improvement in conversation rate at 5% significance and 80% power is 70,546 per group.

   R-code/output:

```r
pwr.2p.test(h=ES.h(p1 = 0.2, p2 = 0.206), sig.level = 0.05, power = 0.8)
```

```
> pwr.2p.test(h=ES.h(p1 = 0.2, p2 = 0.206), sig.level = 0.05, power = 0.8)

     Difference of proportion power calculation for binomial distribution (arcs
ine transformation)

              h = 0.01491711
              n = 70545.16
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: same sample sizes
```