

Problem Set 5

Field Experiments

Adam Yang

1. Online advertising natural experiment.

These are simulated data (closely, although not entirely) based on a real example, adopted from Randall Lewis' dissertation at MIT.

Problem Setup

Imagine Yahoo! sells homepage ads to advertisers that are quasi-randomly assigned by whether the user loads the Yahoo! homepage (www.yahoo.com) on an even or odd second of the day. More specifically, the setup is as follows. On any given week, Monday through Sunday, two ad campaigns are running on Yahoo!'s homepage. If a user goes to www.yahoo.com during an even second that week (e.g., Monday at 12:30:58pm), the ads for the advertiser are shown. But if the user goes to www.yahoo.com during an odd second during that week (e.g., Monday at 12:30:59), the ads for other products are shown. (If a user logs onto Yahoo! once on an even second and once on an odd second, they are shown the first of the campaigns the first time and the second of the campaigns the second time. Assignment is not persistent within users.)

This natural experiment allows us to use the users who log onto Yahoo! during odd seconds/the ad impressions from odd seconds as a randomized control group for users who log onto Yahoo! during even seconds/the ad impressions from even seconds. (We will assume throughout the problem there is no effect of viewing advertiser 2's ads, from odd seconds, on purchases for advertiser 1, the product advertised on even seconds.)

Imagine you are an advertiser who has purchased advertising from Yahoo! that is subject to this randomization on two occasions. Here is a link to (fake) data on 500,000 randomly selected users who visited Yahoo!'s homepage during each of your two advertising campaigns, one you conducted for product A in March and one you conducted for product B in August (~250,000 users for each of the two experiments). Each row in the dataset corresponds to a user exposed to one of these campaigns.

```
library(data.table)
library(stargazer)
library(dplyr)
library(foreign)
library(lmtest)
library(sandwich)
library(multiwayvcov)
```

```
d1 <- fread('./data/ps5_no1.csv')
d1
```

```
##      product_b total_ad_exposures_week1 treatment_ad_exposures_week1
##      1:         1                     4                        3
##      2:         1                     1                        1
##      3:         1                     3                        1
##      4:         0                     5                        0
##      5:         0                     1                        1
##      ---
## 499996:         1                     2                        1
## 499997:         1                     4                        0
```

```

## 499998:      0      5      2
## 499999:      1      5      1
## 500000:      0      3      1
##      week0 week1 week2 week3 week4 week5 week6 week7 week8 week9 week10
##      1:  5.5  6.2  0.0  0.0  0.0  0.0  0  0.0  9.7  4.1  0.0
##      2:  6.2  0.0  8.6  2.4  0.0  7.4  0  0.0  0.0  5.7  0.0
##      3:  0.0  5.3  0.0  8.1  7.8  3.3  0  0.0  9.4  0.0  0.0
##      4:  0.0  4.1  0.0  8.8  5.8  5.9  0  0.0  0.0  9.6  0.0
##      5:  7.6  3.6  4.6  5.5  7.2  7.1  0  0.0  0.0  0.0  0.0
##      ---
## 499996:  0.0  2.8  8.1  2.8  4.1  8.1  0  0.0  0.0  0.0  2.3
## 499997:  0.0  0.0  4.0  4.6  0.0  0.0  0  5.0  0.0  0.0  0.0
## 499998:  4.3  0.0  0.0  0.0  4.2  6.5  0  4.8  9.4  6.4  9.5
## 499999:  0.0  0.0  0.0  5.4  6.1  0.0  12  7.7  0.0  0.0  0.0
## 500000:  4.7  0.0  0.0  2.2  0.0  0.0  0  4.3  0.0  4.9  0.0

```

The variables in the dataset are described below:

- **product_b**: an indicator for whether the data is from your campaign for product A (in which case it is set to 0), sold beginning on March 1, or for product B, sold beginning on August 1 (in which case it is set to 1). That is, there are two experiments in this dataset, and this variable tells you which experiment the data belong to.
- **treatment_ad_exposures_week1**: number of ad exposures for the product being advertised during the campaign. (One can also think of this variable as “number of times each user visited Yahoo! homepage on an even second during the week of the campaign.”)
- **total_ad_exposures_week1**: number of ad exposures on the Yahoo! homepage each user had during the ad campaign, which is the sum of exposures to the “treatment ads” for the product being advertised (delivered on even seconds) and exposures to the “control ads” for unrelated products (delivered on odd seconds). (One can also think of this variable as “total number of times each user visited the Yahoo! homepage during the week of the campaign.”)
- **week0**: For the treatment product, the revenues from each user in the week prior to the launch of the advertising campaign.
- **week1**: For the treatment product, the revenues from each user in the week during the advertising campaign. The ad campaign ends on the last day of week 1.
- **week2-week10**: Revenue from each user for the treatment product sold in the weeks subsequent to the campaign. The ad campaign was not active during this time.

Simplifying assumptions you should make when answering this problem:

- The effect of treatment ad exposures on purchases is linear. That is, the first exposure has the same effect as the second exposure.
- There is no effect of being exposed to the odd-second ads on purchases for the product being advertised on the even second.
- Every Yahoo! user visits the Yahoo! home page at most six times a week.
- You can assume that treatment ad exposures do not cause changes in future ad exposures. That is, assume that getting a treatment ad at 9:00am doesn’t cause you to be more (or less) likely to visit the Yahoo home pages on an even second that afternoon, or on subsequent days.

Questions to Answer

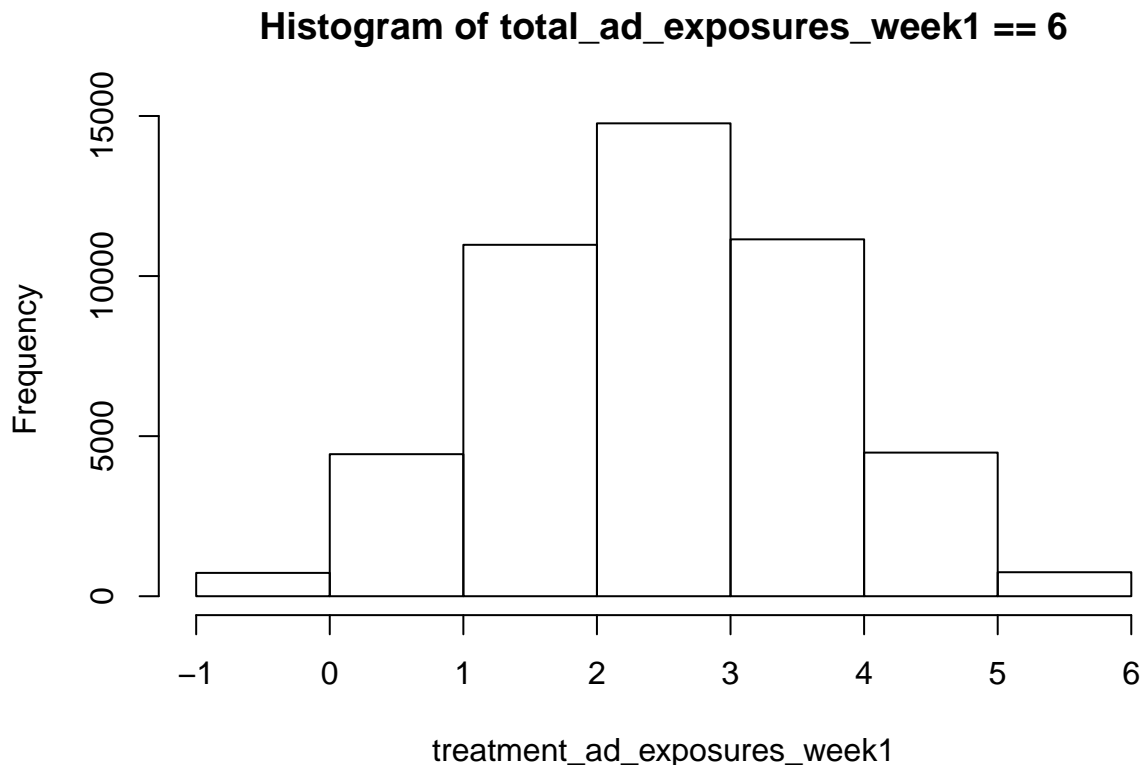
- Run a crosstab of `total_ad_exposures_week1` and `treatment_ad_exposures_week1` to sanity check that the distribution of impressions looks as it should. Does it seem reasonable? Why does it look like this? (No computation required here, just a brief verbal response.)

```
crosstab <- xtabs(~total_ad_exposures_week1+treatment_ad_exposures_week1, data = d1)
crosstab
```

```
##               treatment_ad_exposures_week1
## total_ad_exposures_week1    0    1    2    3    4    5    6
##      0 61182      0    0    0    0    0    0
##      1 36754 37215    0    0    0    0    0
##      2 21143 42036 20965    0    0    0    0
##      3 10683 32073 32314 10726    0    0    0
##      4  5044 20003 30432 20223  5115    0    0
##      5  2045 10563 20970 20793 10293 2131    0
##      6   729  4437 10977 14771 11147  4486   750
```

Answer: As shown in the crosstab above, when the `total_ad_exposures_week1` is 0, all the `treatment_ad_exposures_week1` are also 0, which means that when there are no ad exposures in week1, there will also be no treatment, which makes sense. This is also true for the other values of `total_ad_exposures_week1`. For example, when the `total_ad_exposures_week1` is 1, the `treatment_ad_exposures_week1` can only be 0 or 1. The upper right hand triangle of the crosstab all consists of 0 which shows the consistency of this rule. Also, we can see that for each row of the data, there seems to be a normal distribution. For example, when `total_ad_exposures_week1` is 6, the most likely number of `treatment_ad_exposures_week1` is 3, followed by 2 and 4. The approximate normal distribution can be seen in the histogram below. This also makes sense because assuming that treatment is randomly assigned, and we only have 2 states of treatment or control, then we would expect the number of treatments to have a normal distribution around half of the total ad exposures.

```
hist(d1[d1$total_ad_exposures_week1 == 6]$treatment_ad_exposures_week1,
     main = "Histogram of total_ad_exposures_week1 == 6",
     xlab = "treatment_ad_exposures_week1", breaks = c(-1,0,1,2,3,4,5,6))
```



- b. Your colleague proposes the code printed below to analyze this experiment: `lm(week1 ~ treatment_ad_exposures_week1, data)` You are suspicious. Run a placebo test with the

prior week's purchases as the outcome and report the results. Did the placebo test “succeed” or “fail”? Why do you say so?

```
dA <- d1[d1$product_b == 0]
dB <- d1[d1$product_b == 1]
# Placebo Test with both product A and product B
placeboTest <- lm(week0 ~ treatment_ad_exposures_week1, data = d1)
summary(placeboTest)

##
## Call:
## lm(formula = week0 ~ treatment_ad_exposures_week1, data = d1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.248 -2.196 -1.670  2.430  8.330
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.669685   0.006027   277.0  <2e-16 ***
## treatment_ad_exposures_week1 0.263099   0.003155    83.4  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.796 on 499998 degrees of freedom
## Multiple R-squared:  0.01372,    Adjusted R-squared:  0.01372
## F-statistic: 6955 on 1 and 499998 DF,  p-value: < 2.2e-16
```

Answer: In the model above, I ran a placebo test by regressing the week0 revenues with the treatment_ad_exposures_week1. The results showed that the treatment group in week one had **0.263 (0.0032)** higher revenue in week zero than the control group. The results are shown to be highly statistically significant. This means that the placebo test has failed because in week zero, no treatment has been applied at all so we would expect there to be no statistically significant difference between the two groups if we had purely unbiased random assignment. The fact that the users in the week one treatment group contributed more revenue than the week one control group before any treatment was applied, suggests that there is some unaccounted bias between the two groups. This bias seems to be causing the treatment group to contribute more revenue regardless of treatment.

```
# Placebo Test for only product A in March
placeboTest <- lm(week0 ~ treatment_ad_exposures_week1, data = dA)
summary(placeboTest)

##
## Call:
## lm(formula = week0 ~ treatment_ad_exposures_week1, data = dA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.084 -1.821 -1.568  2.179  7.832
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.568239   0.006766  231.77  <2e-16 ***
## treatment_ad_exposures_week1 0.252637   0.004045   62.46  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.584 on 300036 degrees of freedom
## Multiple R-squared:  0.01284,    Adjusted R-squared:  0.01283
## F-statistic: 3902 on 1 and 300036 DF,  p-value: < 2.2e-16

# Placebo Test for only product B in August
placeboTest <- lm(week0 ~ treatment_ad_exposures_week1, data = dB)
summary(placeboTest)

##
## Call:
## lm(formula = week0 ~ treatment_ad_exposures_week1, data = dB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.212 -2.357 -2.144  2.843  8.070
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.930027   0.011837  163.05  <2e-16 ***
## treatment_ad_exposures_week1 0.213653   0.005333   40.06  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.078 on 199960 degrees of freedom
## Multiple R-squared:  0.007963,    Adjusted R-squared:  0.007958
## F-statistic: 1605 on 1 and 199960 DF,  p-value: < 2.2e-16
```

Answer: The placebo test also fails when you look at product A and B separately.

- c. The placebo test suggests that there is something wrong with our experiment or our data analysis. We suggest looking for a problem with the data analysis. Do you see something that might be spoiling the randomness of the treatment variable? How can you improve your analysis to get rid of this problem? Why does the placebo test turn out the way it does? What one thing needs to be done to analyze the data correctly? Please provide a brief explanation of why, not just what needs to be done. (*Note: This question, and verifying that you answered it correctly in part d below, may require some thinking. If we find many people can't figure it out, we will post another hint in a few days.*)

Answer: I think that one difference between the users is how often they visit the Yahoo! homepage. It may be that the users who go online more often will also be more likely to purchase the products off the internet. The people who go online more often will also be more likely to have been exposed to the treatment more often. Therefore, this will cause a bias where we overestimate the treatment effect if we do not take the user's internet usage into consideration when we do the regression. From our data set, it seems like we can use the `total_ad_exposures_week1` variable to represent how often each user uses the internet. This is assuming that most of the users who use the internet more often will be visiting the Yahoo! homepage more often in week one. We can improve our analysis to get rid of the bias by adding `total_ad_exposures_week1` as a covariate in the regression. By doing so, we can get the treatment effect while holding the amount of times the users visit the Yahoo! homepage as equal.

- d. Implement the procedure you propose from part (c), run the placebo test for the Week 0 data again, and report the results. (This placebo test should pass; if it does not, re-evaluate your strategy before wasting time proceeding.)

```
# Placebo Test for both product A and B
placeboTest <- lm(week0 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = d1)
summary(placeboTest)

##
```

```
## Call:
## lm(formula = week0 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1,
##     data = d1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.817 -2.079 -1.589  2.455  7.823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.345375   0.007295  184.436 <2e-16 ***
## treatment_ad_exposures_week1 -0.002245   0.004629  -0.485  0.628
## total_ad_exposures_week1      0.245348   0.003149  77.922 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.779 on 499997 degrees of freedom
## Multiple R-squared:  0.02555,    Adjusted R-squared:  0.02555
## F-statistic: 6556 on 2 and 499997 DF,  p-value: < 2.2e-16
```

Answer: In the model above, I ran a placebo test again by regressing the week0 revenues with the treatment_ad_exposures_week1, but this time, I introduced total_ad_exposures_week1 as a covariate. The results showed that the treatment group in week one had approximately **0.0022 (0.0046)** lower revenue in week zero than the control group. The results are shown to be not statistically significant anymore. This means that the placebo test has passed because in week zero, no treatment has been applied at all so we would expect there to be no statistically significant difference between the two groups if we had purely unbiased random assignment. Looking at our total_ad_exposures_week1 covariate, we see that for one extra time they visit the Yahoo! homepage, the users will on average contribute **0.245 (0.0031)** more to the revenue by buying the products. This is shown to be highly statistically significant which suggests that our reasoning in part (c) makes sense. The users who tend to visit the Yahoo! homepage more often, also tend to be more likely to purchase the products online.

```
# Placebo Test for only product A
placeboTest <- lm(week0 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dA)
summary(placeboTest)
```

```
##
## Call:
## lm(formula = week0 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1,
##     data = dA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.710 -2.006 -1.324  2.194  7.745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.323594   0.007979  165.890 <2e-16 ***
## treatment_ad_exposures_week1 -0.005548   0.006053  -0.917  0.359
## total_ad_exposures_week1      0.231078   0.004048  57.082 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.57 on 300035 degrees of freedom
## Multiple R-squared:  0.02344,    Adjusted R-squared:  0.02344
```

```
## F-statistic: 3601 on 2 and 300035 DF, p-value: < 2.2e-16
# Placebo Test for only product B
placeboTest <- lm(week0 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dB)
summary(placeboTest)

##
## Call:
## lm(formula = week0 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1,
##     data = dB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.875 -2.408 -1.942  2.857  7.724
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4777918   0.0157086   94.075  <2e-16 ***
## treatment_ad_exposures_week1 0.0007448   0.0072181    0.103   0.918
## total_ad_exposures_week1     0.2320818   0.0053323   43.524  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.063 on 199959 degrees of freedom
## Multiple R-squared:  0.01727, Adjusted R-squared:  0.01726
## F-statistic: 1757 on 2 and 199959 DF, p-value: < 2.2e-16
```

Answer: The placebo test also fails when you look at product A and product B separately.

- e. Now estimate the causal effect of each ad exposure on purchases during the week of the campaign itself using the same technique that passed the placebo test in part (d).

Note to Grader: I am assuming this question is asking us to run the regression on week1 revenues only, using the same technique as part (d). I am assuming we are not expected to separate product A data from product B, and we are also not supposed to do regressions on week2-week10 revenues. (I did the regressions on week2-week10 revenues and the treatment effect loses its statistical significance).

```
# Regression for both product A and B
modelAB <- lm(week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = d1)
summary(modelAB)

##
## Call:
## lm(formula = week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1,
##     data = d1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.003 -2.104 -1.542  2.447  8.110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.317960   0.007263  181.47  <2e-16 ***
## treatment_ad_exposures_week1 0.056340   0.004609   12.22  <2e-16 ***
## total_ad_exposures_week1     0.224478   0.003135   71.61  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.767 on 499997 degrees of freedom
## Multiple R-squared:  0.02782,    Adjusted R-squared:  0.02781
## F-statistic: 7153 on 2 and 499997 DF,  p-value: < 2.2e-16
```

Answer: From the regression shown above, for each time the user gets treated in the week (0-6 times), they end up contributing, on average, **0.056 (0.0046)** more revenue. The result is highly statistically significant.

Regression for only product A

```
modelA <- lm(week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dA)
summary(modelA)
```

```
##
## Call:
## lm(formula = week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1,
##     data = dA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.933 -1.978 -1.296   2.190   7.186
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.295811   0.007939   163.22 <2e-16 ***
## treatment_ad_exposures_week1 0.068154   0.006023    11.31 <2e-16 ***
## total_ad_exposures_week1     0.204690   0.004028    50.82 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.557 on 300035 degrees of freedom
## Multiple R-squared:  0.02623,    Adjusted R-squared:  0.02622
## F-statistic: 4041 on 2 and 300035 DF,  p-value: < 2.2e-16
```

Regression for only product B

```
modelB <- lm(week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dB)
summary(modelB)
```

```
##
## Call:
## lm(formula = week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1,
##     data = dB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.021 -2.407 -1.882   2.856   8.068
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.444893   0.015651    92.32 < 2e-16 ***
## treatment_ad_exposures_week1 0.044232   0.007192     6.15 7.75e-10 ***
## total_ad_exposures_week1     0.218411   0.005313    41.11 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.052 on 199959 degrees of freedom
## Multiple R-squared:  0.01879,    Adjusted R-squared:  0.01878
```



```
## F-statistic: 1915 on 2 and 199959 DF, p-value: < 2.2e-16
```

Answer: When you look at only product A, for each time the user gets treated in the week (0-6 times), they end up contributing, on average, **0.068 (0.0060)** more revenue. When you look at only product B, for each time the user gets treated in the week (0-6 times), they end up contributing, on average, **0.044 (0.0072)** more revenue. The results are both highly statistically significant.

- f. The colleague who proposed the specification in part (b) challenges your results – they make the campaign look less successful. Write a paragraph that a layperson would understand about why your estimation strategy is superior and his/hers is biased.

Answer: His/her estimation strategy is biased because he/her did not take into consideration that people who use the internet more often, are more likely to purchase the product online. This can be seen in the `total_ad_exposures_week1` coefficient of **0.2244 (0.0031)** shown in part (e), which is highly statistically significant. By not including this covariate, my colleague would be overestimating the treatment effect of the ad campaign. My method takes each user's internet usage into consideration and I resulted in a smaller treatment effect which is also statistically significant but more accurate. My colleague's method failed the placebo test shown in part (b) while my method passed the placebo test shown in part (d).

- g. Estimate the causal effect of each treatment ad exposure on purchases during and after the campaign, up until week 10 (so, total purchases during weeks 1 through 10).

```
# Regression for both product A and B
d1$total_purchases <- d1$week1+d1$week2+d1$week3+d1$week4+d1$week5+d1$week6+d1$week7+d1$week8+d1$week9+
modelAB <- lm(total_purchases ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = d1)
summary(modelAB)
```

```
##
## Call:
## lm(formula = total_purchases ~ treatment_ad_exposures_week1 +
##     total_ad_exposures_week1, data = d1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30.597  -7.372  -0.731   6.654  59.782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15081    0.02771  618.949  <2e-16 ***
## treatment_ad_exposures_week1  0.01274    0.01758   0.724    0.469
## total_ad_exposures_week1     2.22834    0.01196  186.307  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.56 on 499997 degrees of freedom
## Multiple R-squared:  0.1321, Adjusted R-squared:  0.1321
## F-statistic: 3.804e+04 on 2 and 499997 DF, p-value: < 2.2e-16
```

Answer: The causal effect of the treatment ad exposure for both product A and product B on purchases during and after the campaign, comes with a coefficient of **0.013 (0.018)**, which is not statistically significant.

```
# Regression for only product A
dA$total_purchases <- dA$week1+dA$week2+dA$week3+dA$week4+dA$week5+dA$week6+dA$week7+dA$week8+dA$week9+
modelA <- lm(total_purchases ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dA)
summary(modelA)
```

```
##
## Call:
```

```
## lm(formula = total_purchases ~ treatment_ad_exposures_week1 +
##     total_ad_exposures_week1, data = dA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.640  -7.004  -0.722   6.269  54.923
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      16.92152    0.03097  546.458  <2e-16 ***
## treatment_ad_exposures_week1  0.04551    0.02349   1.937   0.0527 .
## total_ad_exposures_week1      2.08190    0.01571  132.510  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.973 on 300035 degrees of freedom
## Multiple R-squared:  0.1193, Adjusted R-squared:  0.1193
## F-statistic: 2.032e+04 on 2 and 300035 DF,  p-value: < 2.2e-16
```

Regression for only product B

```
dB$total_purchases <- dB$week1+dB$week2+dB$week3+dB$week4+dB$week5+dB$week6+dB$week7+dB$week8+dB$week9+
modelB <- lm(total_purchases ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dB)
summary(modelB)
```

```
##
## Call:
## lm(formula = total_purchases ~ treatment_ad_exposures_week1 +
##     total_ad_exposures_week1, data = dB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.137  -7.936  -0.655   7.245  59.128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      18.41939    0.05805  317.322  <2e-16 ***
## treatment_ad_exposures_week1 -0.02282    0.02667  -0.856   0.392
## total_ad_exposures_week1      2.11957    0.01970  107.572  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.32 on 199959 degrees of freedom
## Multiple R-squared:  0.09574, Adjusted R-squared:  0.09573
## F-statistic: 1.059e+04 on 2 and 199959 DF,  p-value: < 2.2e-16
```

Answer: The causal effect of the treatment ad exposure for only product A on purchases during and after the campaign, comes with a coefficient of **0.046 (0.023)**, which is not statistically significant. The causal effect of the treatment ad exposure for only product B on purchases during and after the campaign, comes with a coefficient of **-0.023 (0.027)**, which is not statistically significant.

- h. Estimate the causal effect of each treatment ad exposure on purchases only after the campaign. That is, look at total purchases only during week 2 through week 10, inclusive.

Regression for both product A and B

```
d1$total_after_purchases <- d1$week2+d1$week3+d1$week4+d1$week5+d1$week6+d1$week7+d1$week8+d1$week9+d1$
modelAB <- lm(total_after_purchases ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = d
```

```
summary(modelAB)
```

```
##
## Call:
## lm(formula = total_after_purchases ~ treatment_ad_exposures_week1 +
##     total_ad_exposures_week1, data = d1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.856  -7.097  -0.697   6.382  54.079
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      15.83285     0.02654  596.493 < 2e-16 ***
## treatment_ad_exposures_week1 -0.04360     0.01684  -2.588  0.00964 **
## total_ad_exposures_week1      2.00387     0.01146  174.901 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.11 on 499997 degrees of freedom
## Multiple R-squared:  0.1154, Adjusted R-squared:  0.1154
## F-statistic: 3.261e+04 on 2 and 499997 DF,  p-value: < 2.2e-16
```

Answer: The causal effect of the treatment ad exposure for both product A and product B on only purchases after the campaign, comes with a coefficient of **-0.044 (0.017)**, which is statistically significant.

Regression for only product A

```
dA$total_after_purchases <- dA$week2+dA$week3+dA$week4+dA$week5+dA$week6+dA$week7+dA$week8+dA$week9+dA$week10
modelA <- lm(total_after_purchases ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dA)
summary(modelA)
```

```
##
## Call:
## lm(formula = total_after_purchases ~ treatment_ad_exposures_week1 +
##     total_ad_exposures_week1, data = dA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.889  -6.735  -0.689   6.043  53.111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      15.62571     0.02970  526.183 <2e-16 ***
## treatment_ad_exposures_week1 -0.02265     0.02253  -1.005   0.315
## total_ad_exposures_week1      1.87721     0.01507  124.590 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.564 on 300035 degrees of freedom
## Multiple R-squared:  0.1037, Adjusted R-squared:  0.1037
## F-statistic: 1.736e+04 on 2 and 300035 DF,  p-value: < 2.2e-16
```

Regression for only product B

```
dB$total_after_purchases <- dB$week2+dB$week3+dB$week4+dB$week5+dB$week6+dB$week7+dB$week8+dB$week9+dB$week10
modelB <- lm(total_after_purchases ~ treatment_ad_exposures_week1 + total_ad_exposures_week1, data = dB)
```

```
summary(modelB)
```

```
##
## Call:
## lm(formula = total_after_purchases ~ treatment_ad_exposures_week1 +
##     total_ad_exposures_week1, data = dB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.381  -7.644  -0.678   6.923  53.521
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      16.97449     0.05556 305.508 < 2e-16 ***
## treatment_ad_exposures_week1 -0.06705     0.02553  -2.626  0.00863 **
## total_ad_exposures_week1      1.90116     0.01886 100.802 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.83 on 199959 degrees of freedom
## Multiple R-squared:  0.08318,    Adjusted R-squared:  0.08317
## F-statistic: 9070 on 2 and 199959 DF,  p-value: < 2.2e-16
```

Answer: The causal effect of the treatment ad exposure for only product A on purchases during and after the campaign, comes with a coefficient of **-0.023 (0.023)**, which is not statistically significant. The causal effect of the treatment ad exposure for only product B on purchases during and after the campaign, comes with a coefficient of **-0.067 (0.026)**, which is statistically significant.

i. Tell a story that could plausibly explain the result from part (h).

Answer: From part (h), we see that when we look at the combined revenue of both product A and product B from week2 to week10, there is a decrease of **0.044 (0.017)** in revenue as the exposure to the ad campaign increases by 1. I think this is because many of the people in the treatment group purchased the products in week1 when they were exposed to the ad campaign. In the weeks after the ad campaign, they already owned the product, so they are less likely to purchase it again. From part (g), we see that there seems to be no statistically significant treatment effect if we include the revenues from week1 to week10. This means that the ad campaign gives us a short term boost in revenue, but it does not give us a long term boost in revenue.

If we only look at the results for product A in part (h), we see that there is a decrease of **0.023 (0.023)** in revenue as the exposure to the ad campaign increases by 1, but it is not statistically significant. For product B, however, we see that there is a decrease of **0.067 (0.026)** in revenue as the exposure to the ad campaign increases by 1, which is statistically significant. In this case, we can guess that the revenue dropoff effect is stronger in product B than product A. Perhaps product A is something that the users are more likely purchase more than once.

j. Test the hypothesis that the ads for product B are more effective, in terms of producing additional revenue in week 1 only, than are the ads for product A. (*Hint: The easiest way to do this is to throw all of the observations into one big regression and specify that regression in such a way that it tests this hypothesis.*) (*Hint 2: There are a couple defensible ways to answer this question that lead to different answers. Don't stress if you think you have an approach you can defend.*)

```
# Regression for both product A and B
```

```
modelAB <- lm(week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1 + product_b + treatment_a
summary(modelAB)
```

```
##
## Call:
```

```
## lm(formula = week1 ~ treatment_ad_exposures_week1 + total_ad_exposures_week1 +
##      product_b + treatment_ad_exposures_week1 * product_b, data = d1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.032 -2.194 -1.500  2.439  8.071
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   1.289270   0.008010 160.960
## treatment_ad_exposures_week1    0.061251   0.005637  10.867
## total_ad_exposures_week1        0.210868   0.003230  65.277
## product_b                      0.170320   0.013186  12.917
## treatment_ad_exposures_week1:product_b -0.010100   0.006490  -1.556
##                                Pr(>|t|)
## (Intercept)                   <2e-16 ***
## treatment_ad_exposures_week1    <2e-16 ***
## total_ad_exposures_week1        <2e-16 ***
## product_b                      <2e-16 ***
## treatment_ad_exposures_week1:product_b    0.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.766 on 499995 degrees of freedom
## Multiple R-squared:  0.02848,    Adjusted R-squared:  0.02847
## F-statistic: 3664 on 4 and 499995 DF,  p-value: < 2.2e-16
```

Answer: I introduced the `product_b` variable into the regression along with the interaction regressor of `treatment_ad_exposures_week1*product_b`. From this, we see that the coefficient for product A is **0.0612 (0.0056)** which is statistically significant (Note that the value is very similar to the one we found in part (e)). The coefficient for product B would be **0.0612 (0.0056) - 0.0101 (0.0065) = 0.0511** which is smaller than product A (Note that the value is larger than the one we found in part (e), but the overall story is unchanged. The ads for product B seems to be less effective). From the regression shown above, it seems like product B is less effective than product A. However, the difference in the coefficient between product A and product B is not statistically significant which means we cannot quite reject the hypothesis that the effectiveness for the two ads are equal.

- k. You notice that the ads for product A included celebrity endorsements. How confident would you be in concluding that celebrity endorsements increase the effectiveness of advertising at stimulating immediate purchases?

Answer: I would not be very confident concluding this at all. First of all, the difference between the coefficients as shown in part (j) is not statistically significant, so we cannot reject the hypothesis that the effectiveness for the two ads are equal. Secondly, even if we could say that the ad campaign for product A is more effective than the ad campaign for product B, it does not necessarily have to be because of the celebrity endorsements. Maybe the ad campaign for product A is created to be much more attractive than product B with more vibrant colors. Also, maybe the time in which the two ads were shown would play a part in their effectiveness since the ad for product A was shown in March and the ad for product B was shown in August. Finally, the product itself can play a role in the effectiveness of the ad campaigns. If product A is a much more desirable product than product B, then it would naturally sell better. I would not be confident on concluding that celebrity endorsements increase the effectiveness of advertising because we are not doing an apples to apples comparison.

2. Vietnam Draft Lottery

A famous paper by Angrist exploits the randomized lottery for the Vietnam draft to estimate the effect of education on wages. (*Don't worry about reading this article, it is just provided to satisfy your curiosity; you can answer the question below without referring to it. In fact, it may be easier for you not to, since he has some complications to deal with that the simple data we're giving you do not.*)

Problem Setup

Angrist's idea is this: During the Vietnam era, draft numbers were determined randomly by birth date – the army would literally randomly draw birthdays out of a hat, and those whose birthdays came up sooner were higher up on the list to be drafted first. For example, all young American men born on May 2 of a given year might have draft number 1 and be the first to be called up for service, followed by November 13 who would get draft number 2 and be second, etc. The higher-ranked (closer to 1) your draft number, the likelier it was you would be drafted.

We have generated a fake version of this data for your use in this project. You can find real information (here)[<https://www.sss.gov/About/History-And-Records/lotter1>]. While we're defining having a high draft number as falling at 80, in reality in 1970 any number lower than 195 would have been a “high” draft number, in 1971 anything lower than 125 would have been “high”.

High draft rank induced many Americans to go to college, because being a college student was an excuse to avoid the draft – so those with higher-ranked draft numbers attempted to enroll in college for fear of being drafted, whereas those with lower-ranked draft numbers felt less pressure to enroll in college just to avoid the draft (some still attended college regardless, of course). Draft numbers therefore cause a natural experiment in education, as we now have two randomly assigned groups, with one group having higher mean levels of education, those with higher draft numbers, than another, those with lower draft numbers. (In the language of econometricians, we say the draft number is “an instrument for education,” or that draft number is an “instrumental variable.”)

Some simplifying assumptions:

- Suppose that these data are a true random sample of IRS records and that these records measure every living American's income without error.
- Assume that the true effect of education on income is linear in the number of years of education obtained.
- Assume all the data points are from Americans born in a single year and we do not need to worry about cohort effects of any kind.

Questions to Answer

- a. Suppose that you had not run an experiment. Estimate the “effect” of each year of education on income as an observational researcher might, by just running a regression of years of education on income (in R-ish, `income ~ years_education`). What does this naive regression suggest?

```
model <- lm(income~years_education, data = d2)
summary(model)

##
## Call:
## lm(formula = income ~ years_education, data = d2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -91655 -17459   -837   16346 141587
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -23354.64   1252.74  -18.64  <2e-16 ***
## years_education  5750.48    83.34   69.00  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26590 on 19565 degrees of freedom
## Multiple R-squared:  0.1957, Adjusted R-squared:  0.1957
## F-statistic: 4761 on 1 and 19565 DF, p-value: < 2.2e-16
```

Answer: In the naive regression shown above, we see that for each year of education, a person would approximately have an income that is **\$5750.48 (83.34)** larger. This result is highly statistically significant. This naive regression suggests that each year of education would result in an increased income of almost \$6000. However, in the next questions we will talk about how this number might be overestimated because of possible bias.

- b. Continue to suppose that we did not run the experiment, but that we saw the result that you noted in part (a). Tell a concrete story about why you don't believe that observational result tells you anything causal.

Answer: The observational result does not tell us anything causal because the treatment of pursuing higher education was not randomly assigned but rather self selected. In this case, it is possible that the smarter and/or more ambitious people are actively pursuing higher education and that might be what is causing the higher income rather than the level of education itself. Another possibility is that people from wealthy families tend to go to college and are actually able to afford college whereas poorer families might not. These wealthy families might also have much better connections that can result in a high paying job which the poorer families do not have access too. This can also cause bias that would overestimate our results.

- c. Now, let's get to using the natural experiment. We will define "having a high-ranked draft number" as having a draft number of 80 or below (1-80; numbers 81-365, for the remaining 285 days of the year, can be considered "low-ranked"). Create a variable in your dataset indicating whether each person has a high-ranked draft number or not. Using regression, estimate the effect of having a high-ranked draft number, the dummy variable you've just created, on years of education obtained. Report the estimate and a correctly computed standard error. (*Hint: Pay special attention to calculating the correct standard errors here. They should match how the draft is conducted.)

```
# Create a dummy variable indicating whether a person is a high-ranked draft
d2$high_rank <- as.numeric(d2$draft_number <= 80)
```

```
# Using regression, estimate the effect of having a high-ranked draft number on years of education obtained
model <- lm(years_education~high_rank, data = d2)
```

```
# Use coeftest to find the cluster-robust standard error and p value
coeftestvalues <- coeftest(model, cluster.vcov(model, ~ draft_number))
coeftestvalues
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.434305  0.017703 815.345 < 2.2e-16 ***
## high_rank    2.125756  0.038188  55.666 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: As shown above, people with high-ranked draft numbers seem to have approximately **2.126** more years of education than the people with low-ranked draft numbers. The clustered standard error is **0.0382** which makes the result highly statistically significant.

- d. Using linear regression, estimate the effect of having a high-ranked draft number on income. Report the estimate and the correct standard error.

```
# Using regression, estimate the effect of having a high-ranked draft number on years of education obtained
model <- lm(income~high_rank, data = d2)

# Use coeftest to find the cluster-robust standard error and p value
coeftestvalues <- coeftest(model, cluster.vcov(model, ~ draft_number))
coeftestvalues
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60761.89      244.36 248.656 < 2.2e-16 ***
## high_rank    6637.55      511.90 12.966 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: As shown above, people with high-ranked draft numbers seem to have approximately **\$6637.60** higher income than the people with low-ranked draft numbers. The clustered standard error is **\$511.90** which makes the result highly statistically significant.

- e. Divide the estimate from part (d) by the estimate in part (c) to estimate the effect of education on income. This is an instrumental-variables estimate, in which we are looking at the “clean” variation in both education and income that is due to the draft status, and computing the slope of the income-education line as “clean change in Y” divided by “clean change in X”. What do the results suggest?

```
# Divide the estimate from part (d) by the estimate in part (c) to estimate the effect of education on income
model1 <- lm(years_education~high_rank, data = d2)
model2 <- lm(income~high_rank, data = d2)
clean.change.in.x <- summary(model1)$coef[2]
clean.change.in.y <- summary(model2)$coef[2]
clean.change.in.y/clean.change.in.x

## [1] 3122.444
```

Answer: The result suggests that for each year of education, we can expect **\$3122.44** higher income.

- f. Natural experiments rely crucially on the “exclusion restriction” assumption that the instrument (here, having a high draft rank) cannot affect the outcome (here, income) in any other way except through its effect on the “endogenous variable” (here, education). Give one reason this assumption may be violated – that is, why having a high draft rank could affect individuals’ income other than because it nudges them to attend school for longer.

Answer: A possible way that the exclusion restriction can be violated is if going to war will lead to a lower income. If having a high draft rank means that you are very likely to be drafted and then go to war, and that going to war leads to a lower income, the exclusion restriction will be violated. Some reasons that this may be true is that going to war prevented these people from picking up trade experiences and skills (assuming they would not have gone to college) which means they are less desirable for higher paying jobs. Also, it is possible that they have some negative psychological effects from going to war such as PTSD that causes them to not be able to obtain and keep a high paying job.

- g. Conduct a test for the presence of differential attrition by treatment condition. That is, conduct a formal test of the hypothesis that the “high-ranked draft number” treatment has no effect on whether

we observe a person's income. (Note, that an earning of \$0 *actually* means they didn't earn any money.)

Note to Grader: Since there are no NA values in the data set, I am assuming that people who attrited from the experiment are not included within the data set. In this case, I think the question is asking about whether there are, on average, fewer people per birthdate in the high draft rank group.

```
# First convert data frame into a data table
d2 <- as.data.table(d2)
# Then create a aggregate table to count the number of rows with each draft_number
d2g <- d2[, .(N), by = .(draft_number, high_rank)]
# Create a linear regression
model <- lm(N ~ high_rank, data = d2g)
# Show results with regular standard errors
summary(model)
```

```
##
## Call:
## lm(formula = N ~ high_rank, data = d2g)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.986  -4.986   0.014   5.014  24.300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  54.9860     0.4346 126.532 < 2e-16 ***
## high_rank    -6.2860     0.9282  -6.772 5.13e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.336 on 363 degrees of freedom
## Multiple R-squared:  0.1122, Adjusted R-squared:  0.1097
## F-statistic: 45.86 on 1 and 363 DF, p-value: 5.128e-11
```

```
# Use coeftest to find the cluster-robust standard error and p value
print("For Cluster-Robust Standard Error")
```

```
## [1] "For Cluster-Robust Standard Error"
```

```
coeftestvalues <- coeftest(model, cluster.vcov(model, ~ draft_number))
coeftestvalues
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  54.98596     0.43149 127.4326 < 2.2e-16 ***
## high_rank    -6.28596     0.94482  -6.6531 1.059e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: According to the model shown above, being of high draft rank leads to an average of **6.29 (0.9)** fewer people being included in the study as compared to the low draft rank. This result is highly statistically significant. This may mean that people in the high draft rank are attriting out of this experiment because we don't have access to their income or information.

h. Tell a concrete story about what could be leading to the result in part (g).

Answer: There are many possible reasons why we are getting fewer participants in the high draft rank group. The most obvious one is that it is possible that the high draft rank group is more likely to go to war and if they are killed in combat, they are attrited from the study. Another possible reason is if veterens become homeless it may be hard to find them and get their income reports.

- i. Tell a concrete story about how this differential attrition might bias our estimates.

Answer: This differential attrition might bias our estimates because there might be some intrinsic difference between someone that survives war and someone who doesn't. Maybe the people who did not survive the war are the best and brightest elite soldiers who are sent on more dangerous missions. Being the best and brightest might've equated to a higher income if they ended up surviving. Or the opposite can be true where the smarter soldiers are given desk jobs or act as translators which means they are more likely to survive the war. These soldiers might have useful skills that would result in a higher income after the war and the soldiers who died might've had a lower income if they survived the war. Also the homeless veterens who we are unable to get income reports from would also bias our data because we are not including their income into our study.

3. Dinner Plates

Suppose that researchers are concerned with the health consequences of what people eat and how much they weigh. Consider an experiment designed to measure the effect of a proposal to help people diet. Subjects are invited to a dinner and are randomly given regular-sized or slightly larger than regular sized plates. Hidden cameras record how much people eat, and the researchers find that those given larger plates eat substantially more food than those assigned small plates.

A statistical test shows that the apparent treatment effect is far greater than one would expect by chance. The authors conclude that a minor adjustment, reducing plate size, will help people lose weight.

- How convincing is the evidence regarding the effect of plate size of what people eat and how much they weigh?
- What design and measurment improvements do you suggest?

Answer: I don't think the evidence regarding the effect of plate size of what people eat and how much they weigh is very convincing. One reason is that people may just be polite becaues they were invited to dinner and therefore would try their best to finish everything on their plate. Also there might be some interference between the people at dinner. Maybe seeing everyone else finish their food is causing some subjects to also try and finish their food. Furthermore, the study done only tells us how much the plate size may affect how much people eat, not their weight loss. It is possible that someone given a smaller plate will end up eating less, but also will end up eating more snacks later on because they are still hungry. In that case, having a smaller plate would affect how much they eat for one meal, but it will not necessarily mean they will lose weight.

To improve the experiment, instead of inviting the subjects to dinner, you can conduct the experiment at an event where free food is provided to the attendees and give the attendees different size plates. In this way, the subjects might not feel the need to be polite and finish all their food. Also at the facility of this event, you might have small tables that only fit one or two people, which might minimize the effect of interence. It would also be ideal to run the experiment over time so that we can truly see the effect of weight loss. We would need a way to measure the weight of each subject. If we invite the subjects to a free health screening before we provide the free lunch, it may give us an opportunity of getting their weight. We can do this repeatedly over time to see if the weight of each subject is changing. However, we have to keep in mind that this can introduce bias because people might become aware of their weight and choose to eat less for that meal. Assuming that this bias would only affect people who are heavier, we can introduce the subjects' starting weight as a covariate to mitigate this bias a bit. We also don't want the subjects to be aware that plate size is what we're varying so it might be ideal to separate the two groups. A sneaky way to get the pateint's weights without their knowing is to use chairs that have the ability to measure their weight. That

way, all they know is that they're eating free food and do not know that their weight is being measured. We will likely want multiple readings on the same people over time and if we do it by inviting them to an event with free food, there might be some problems of attrition. The people who do not show up to all the events with free food might be different than the people who do. To mitigate this, we might want to keep track of people who are required to attend these events. Maybe doing the experiment at a school on children eating cafeteria food is a possibility.

These changes would only tell us the effect of plate size in regards to eating food that is given to the subjects. In terms of dieting, most people would make their own food. It might be interesting to do a different experiment where you invite people over one at a time to cook their own food and randomly give them different sized plates. You can then see what kind of food they cook and how much they cook in relation to their assigned plate size.

4. Think about Treatment Effects

Throughout this course we have focused on the average treatment effect. Think back to *why* we are concerned about the average treatment effect. What is the relationship between an ATE, and some individuals' potential outcomes? Make the strongest case you can for why this is *good* measure.

Answer: According to Gerber and Green, “the average treatment effect (ATE) is the average difference in the potential outcomes $Y_i(1)$ and $Y_i(0)$ for the entire collection of subjects. When the outcomes for the entire treatment and control groups are observed, the treatment group's outcomes are a random sample of $Y_i(1)$ for the pool of subjects, and the control group's outcomes are a random sample of $Y_i(0)$ values. The average outcome in each experimental group provides an unbiased estimate of $\mu_{Y(1)}$ and $\mu_{Y(0)}$ ”. In other words, though we cannot possibly know both potential outcomes of every single subject, if we can provide random assignment to the treatment and control groups and make sure we hold everything equal except for the variable that we applied intervention to, then the average outcome of each group is a good estimate of the average potential outcome of each group. The average treatment effect then tells us the difference between these average potential outcome of the treatment group and the average potential outcome of the control group. This is good because it helps us generalize the treatment effect of our applied variance through interference and gets us closer to assuming causality.