

Problem Set #4

Experiments and Causality

Romulo Manzano (12Nov2017)

```
In [1]: library("foreign")
#loadpackages
library("data.table")
library("foreign")
library("lmtest")
library("sandwich")
library("multiwayvcov")
library("AER")
```

Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

as.Date, as.Date.numeric

Loading required package: car

Loading required package: survival

FE exercise 5.2

a. Make up a hypothetical schedule of potential outcomes for three Compliers and three Never-Takers where the ATE is positive but the CACE is negative.

Answer

Below table consistst of three compliers and three never-takers

```
In [2]: table1 <- data.frame(Observation = 1:6,
                             y_d_0 = c(4,6,1,5,6,5),
                             y_d_1 = c(9,2,6,4,7,1),
                             d_z_0 = c(0, 0,0,0,0,0),
                             d_z_1 = c(0,1,0,1,0,1) )
table1$Complier <- table1$d_z_1
table1
```

Observation	y_d_0	y_d_1	d_z_0	d_z_1	Complier
1	4	9	0	0	0
2	6	2	0	1	1
3	1	6	0	0	0
4	5	4	0	1	1
5	6	7	0	0	0
6	5	1	0	1	1

ATE

```
In [3]: mean(table1$y_d_1 - table1$y_d_0)
0.333333333333333
```

CACE

```
In [4]: mean(table1[table1$d_z_1 == 1,]$y_d_1 - table1[table1$d_z_1 == 1,]$y_d_0)
-3
```

b. Suppose that an experiment were conducted on your pool of subjects. In what ways would the estimated CACE be informative or misleading?

Answer CACE will be informative if the researcher is trying to understand the average treatment effect for the population on which the treatment was prescribed and actually administered (assume one-sided non compliance). For example this can be very useful when assessing the effect of a new drug/medicine on individuals.

At the same time, CACE can also be misleading as the estimation cannot be extrapolated into the effect of the treatment assignment overall. For example, if the researcher's goal is to measure the effect of running a new program where the population may or may not comply, CACE will not be the appropriate measure.

c. **In addition, please also answer this question:** Which population is more relevant to study for future decision making: the set of Compliers, or the set of Compliers plus Never-Takers? Why?

Answer Arguably both, as explained on the prior question, both assessing the effect of the treatment actually being administered and the effect of treatment assignment can be relevant for future decision making (if the results are interpreted correctly). More specifically, a government conducting an experiment to study the effect of a new policy may see value in assessing the overall effect of treatment assignment (compliers plus never-takers) despite knowing some subjects might not comply.

FE exercise 5.6

Suppose that a researcher hires a group of canvassers to contact a set of 1,000 voters randomly assigned to a treatment group. When the canvassing effort concludes, the canvassers report that they successfully contacted 500 voters in the treatment group, but the truth is that they only contacted 250. When voter turnout rates are tabulated for the treatment and control groups, it turns out that 400 of the 1,000 subjects in the treatment group voted, as compared to 700 of the 2,000 subjects in the control group (none of whom were contacted).

a. If you believed that 500 subjects were actually contacted, what would your estimate of the CACE be?

Answer

Intent to Treat

In [5]:
$$ITT = (400/1000) - (700/2000)$$

ITT

0.05

ITT_D

Proportion of subjects treated in the event that they were assigned to treatment (we assume one side non compliance so we don't need to worry about subtracting the proportion who would have been treated if in control group). This quantity is given in the question, which is 0.5 (500 out of 1000)

CACE

In [6]:
$$ITT/0.5$$

0.1

CACE would be 10% if we believed 500 subjects were actually contacted

b. Suppose you learned that only 250 subjects were actually treated. What would your estimate of the CACE be?

CACE

In [7]:
$$ITT/0.25$$

0.2

CACE would be 20% under this new proportion

c. Do the canvassers' exaggerated reports make their efforts seem more or less effective? Define effectiveness either in terms of the ITT or CACE. Why does the definition matter?

Answer

The exaggerated canvassers' efforts seem **less** effective from a CACE perspective, but **more** effective under the ITT.

From a CACE perspective, the exaggerated report suggests that the effect of actually contacting the subjects was only 10% (as opposed to 20%) suggesting that the effort to persuade voters once engaged were not as material as they actually were.

From an ITT perspective, exaggerated reports cause the efforts seem more effective than they actually were, as only a small proportion (25%) of the treatment group was actually actively encouraged to vote.

FE exercise 5.10

Guan and Green report the results of a canvassing experiment conducted in Beijing on the eve of a local election. Students on the campus of Peking University were randomly assigned to treatment or control groups. Canvassers attempted to contact students in their dorm rooms and encourage them to vote. No contact with the control group was attempted. Of the 2,688 students assigned to the treatment group, 2,380 were contacted. A total of 2,152 students in the treatment group voted; of the 1,334 students assigned to the control group, 892 voted. One aspect of this experiment threatens to violate the exclusion restriction. At every dorm room they visited, even those where no one answered, canvassers left a leaflet encouraging students to vote.

```
In [8]: d3 <- read.dta("./data/Guan_Green_CPS_2006.dta")
        head(d3)
```

turnout	contact	dormid	treat2
0	0	1010101	0
0	0	1010101	0
0	0	1010101	0
0	0	1010102	0
0	0	1010102	0
0	1	1010103	1

a. Using the data set from the book's website, estimate the ITT. First, estimate the ITT using the difference in two-group means. Then, estimate the ITT using a linear regression on the appropriate subset of data. *Heads up: There are two NAs in the data frame. Just na.omit to remove these rows.*

Answer

ITT using two group means

```
In [9]: mean(na.omit(d3[d3$treat2==1,]$turnout))-mean(na.omit(d3[d3$treat2==0,]$turnout))
0.131929570928821
```

ITT using linear regression

```
In [10]: result <- lm(turnout ~ treat2, data = na.omit(d3))
summary(result)
```

Call:

```
lm(formula = turnout ~ treat2, data = na.omit(d3))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.8006	0.1994	0.1994	0.1994	0.3313

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.66867	0.01162	57.521	<2e-16 ***
treat2	0.13193	0.01422	9.278	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4246 on 4020 degrees of freedom
 Multiple R-squared: 0.02096, Adjusted R-squared: 0.02072
 F-statistic: 86.08 on 1 and 4020 DF, p-value: < 2.2e-16

b. Use randomization inference to test the sharp null hypothesis that the ITT is zero for all observations, taking into account the fact that random assignment was clustered by dorm room. Interpret your results.

Answer

First we define a couple of functions to do randomized inference

```
In [11]: atePs4 <- function(dataset)
{ return (mean(dataset[dataset$treat2==1,]$turnout)
          - mean(dataset[dataset$treat2==0,]$turnout))
}
```

```
In [12]: randomizePs4 <- function(size) sample(c(0,1), size, replace = TRUE)
```

```
In [13]: randomSimulationATEPS4 = function(dataset) {
  trialX <- cbind(dataset)
  #randomizing for a given room
  dorms <- cbind(data.frame(dormid = (unique((trialX$dormid)))))
  dorms$treat2 <- randomizePs4(nrow(dorms))
  trialX <- trialX[ , !(names(trialX) %in% c("treat2"))]
  trialX <- merge(x = trialX, y = dorms, by = "dormid", all = TRUE)
  effect <- atePs4(trialX)
  return(effect)
}
```

Running 10k simulations

```
In [14]: simulationsITT <- c()
for (i in seq(1,10000))
{
  simulationsITT[i] <- randomSimulationATEPS4(na.omit(d3))
}
```

One tailed p-value

```
In [15]: mean(simulationsITT > 0.131929570928821)
```

```
0
```

This means that we can successfully reject the null hypothesis at a 95% confidence level. In other words, the observed average effect for the realized randomization is well beyond two standard deviations of the distribution generated via randomized inference, and it is highly unlikely the observed effect is the result of chance.

c. Assume that the leaflet had no effect on turnout. Estimate the CACE. Do this in two ways: First, estimate the CACE using means. Second, use some form of linear model to estimate this as well. If you use a 2SLS, then report the standard errors and draw inference about whether the leaflet had any causal effect among compliers.

Answer

Means method

Next, we determine the proportion of the treatment group that was actually contacted

```
In [16]: d3Treated <- cbind(d3[d3$treat2==1,])
         mean(na.omit(d3Treated$contact))
```

```
0.885873605947955
```

CACE

```
In [17]: 0.131929570928821/0.885873605947955
```

```
0.148925952915874
```

Linear regression approach

```
In [18]: d3Leaflet <- cbind(d3)
         #d3Leaflet$leaflet <- (!d3Leaflet$contact & d3Leaflet$treat2)
```

```
In [19]: result2 <- ivreg(turnout ~ contact | treat2, data = na.omit(d3Leaflet))
         coeftest(result2)
```

t test of coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.668666    0.011652  57.3864 < 2.2e-16 ***
contact      0.148940    0.016091   9.2563 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The coefficient associated with 'contact' confirms the effects calculated by hand in the earlier part of this question. The CACE is a 14.89% increase in turnout rate

In order to confirm the treatment effect on the leaflets only, we could run a simple regression on the treatment assignment and contact

```
In [20]: #result3 <- lm(turnout ~ treat2, data = na.omit(d3Leaflet[d3Leaflet$contact==0,]))
result3 <- lm(turnout ~ treat2 + contact, data = na.omit(d3Leaflet))
coeftest(result3)
```

t test of coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.668666   0.011624  57.5256 < 2.2e-16 ***
treat2       0.103321   0.026874   3.8446 0.0001226 ***
contact      0.032297   0.025745   1.2545 0.2097342
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this we can infer that the treatment effect (turnout rate increase) associated with the leaflets was ~10%

FE exercise 5.11

Nickerson describes a voter mobilization experiment in which subjects were randomly assigned to one of three conditions: a baseline group (no contact was attempted); a treatment group (canvassers attempted to deliver an encouragement to vote); and a placebo group (canvassers attempted to deliver an encouragement to recycle). Based on the results in the table below answer the following questions

Treatment Assignment	Treated ?	N	Turnout
Baseline	No	2572	31.22%
Treatment	Yes	486	39.09%
Treatment	No	2086	32.74%
Placebo	Yes	470	29.79%
Placebo	No	2109	32.15%

First Use the information to make a table that has a full recovery of this data. That is, make a `data.frame` or a `data.table` that will have as many rows as there are observations in this data, and that would fully reproduce the table above. (Yes, this might seem a little trivial, but this is the sort of "data thinking" that we think is important.)

```
In [21]: createRandomData <- function(treat,nrecords,treated,turnoutrate)
{
  x <- as.integer(nrecords*turnoutrate)
  t <- c(rep(1,x),rep(0,as.integer(nrecords-x)))
  return (data.frame(treatment = rep(treat,nrecords),
                     treated = rep(treated,nrecords ),
                     #turnout = rbinom(nrecords, 1, turnoutrate))
                     turnout = t
                ))
}
```

```
In [22]: library('data.table')
dataTest <- list(createRandomData('Baseline',2572,'N',0.3122),
                 createRandomData('Treatment',486,'Y',0.3909),
                 createRandomData('Treatment',2086,'N',0.3274),
                 createRandomData('Placebo',470,'Y',0.2979),
                 createRandomData('Placebo',2109,'N',0.3215))
testDataFrame <- rbindlist(dataTest)
#mean(dataTest$turnout)
```

a. We are rewriting part (a) as follows: "Estimate the proportion of Compliers by using the data on the Treatment group. Then compute a second estimate of the proportion of Compliers by using the data on the Placebo group. Are these sample proportions statistically significantly different from each other? Explain why you would not expect them to be different, given the experimental design." (Hint: ITT_D means "the average effect of the treatment on the dosage of the treatment." I.E., it's the contact rate α in the async).

Answer

Compliers on the tratment group

```
In [23]: nrow(testDataFrame[testDataFrame$treatment == "Treatment" & testDataFrame$treated==1])
2086
```

b. Do the data suggest that Never Takers in the treatment and placebo groups have the same rate of turnout? Is this comparison informative?

Answer

Data do suggest Never Takers in the treatment and placebo groups have the same turnout rate. The comparison is not informative though as we can infer that the treatment an placebo groups exhibit similar behavior when untreated, suggesting the groups are not inherintly different

c. Estimate the CACE of receiving the placebo. Is this estimate consistent with the substantive assumption that the placebo has no effect on turnout?

Answer

```
In [24]: ITTPlacebo = mean(testDataFrame[testDataFrame$treatment == "Placebo",]$turnout) -
ITTPlacebo
0.00535760481988451
```

```
In [25]: ITTdPlacebo = (nrow(testDataFrame[testDataFrame$treatment == "Placebo" & testDataF
/nrow(testDataFrame[testDataFrame$treatment == "Placebo",]))
ITTdPlacebo
0.182241178751454
```

```
In [26]: CACEPlacebo = ITTPlacebo/ITTdPlacebo
CACEPlacebo
0.0293984315542173
```


The data suggest that placebo does have a small effect in turnout.

d. Estimate the CACE of receiving the treatment using two different methods. First, use the conventional method of dividing the ITT by the ITT_{D}.

Answer

Conventional method

```
In [27]: ITTTreat = mean(testDataFrame[testDataFrame$treatment == "Treatment",]$turnout) -
         ITTPlacebo
```

0.0268273716951788

```
In [28]: ITTdTreat = (nrow(testDataFrame[testDataFrame$treatment == "Treatment" & testDataFrame$D == 1]) -
         nrow(testDataFrame[testDataFrame$treatment == "Treatment" & testDataFrame$D == 0]))
```

0.18895800933126

```
In [29]: CACETreat = ITTTreat/ITTdTreat
         CACETreat
```

0.141975308641975

Two-stage linear regression

```
In [30]: testDataFrame$isTreated = testDataFrame$treated == "Y"
         testDataFrame$isTreatment = testDataFrame$treatment == "Treatment"
```

```
In [31]: resultTreat <- ivreg(turnout ~ isTreated | isTreatment, data = testDataFrame[te
         coefTest(resultTreat)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3118196	0.0092393	33.7492	< 2e-16 ***
isTreatedTRUE	0.1419753	0.0691496	2.0532	0.04011 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Both the conventional method and the 2SLR concur on a CACE of 14.19% increase in turnout

e. Then, second, compare the turnout rates among the Compliers in both the treatment and placebo groups. Interpret the results.

Answer

When doing a naive ATE on placebo vs. treatment we can observe a treatment effect of just ~2% on turnout rate.

```
In [32]: resultTreatoverPlacebo <- ivreg(turnout ~ isTreatment , data = testDataFrame[testDataFrame$treatment == "Treatment",])
         coeftest(resultTreatoverPlacebo)
```

t test of coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.3171772   0.0092434  34.3139   <2e-16 ***
isTreatmentTRUE 0.0214698   0.0130810   1.6413    0.1008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
In [33]: TreatOverPlacebo = mean(testDataFrame[testDataFrame$treatment == "Treatment",]$turnout)
         TreatOverPlacebo
```

0.0214697668752943

More Practice

Determine the direction of bias in estimating the ATE for each of the following situations when we randomize at the individual level. Do we over-estimate, or underestimate? Briefly but clearly explain your reasoning.

a. In the advertising example of Lewis and Reiley (2014), assume some treatment-group members are friends with control-group members.

Answer

The ATE would potentially be underestimated because members of the control group could effectively be indirectly be influenced by the treatment, that is, members of the control group would have learned about the advertisements their friends were exposed too which could have increased their likelihood to make a purchase.

That would have translated into higher expenditure measures for the control group, closing the gap between treatment and control.

b. Consider the police displacement example from the bulleted list in the introduction to FE 8, where we are estimating the effects of enforcement on crime.

Answer

We would overestimate ATE because of crime being displaced from treated locations to adjacent areas. A byproduct of that displacement is the crime rates going down in the treatment areas while the opposite (crime rates increase) happens on adjacent untreated areas. For example, if the crime rate in both the treated and control areas BEFORE the experiment were 500 crimes per 100,000 general population each, and AFTER the treatment the rates turn out to be 350 per 100,000 (treated) and 650 per 100,000 (control) respectively, we could incorrectly conclude that the policing has a positive effect even though the generalized crime rate (treatment+control) remained unchanged.

c. Suppose employees work harder when you experimentally give them compensation that is more generous than they expected, that people feel resentful (and therefore work less hard) when they learn that their compensation is less than others, and that some treatment-group members talk to control group members.

Answer

Given that some treatment group members talk to control group members, there is clear interference effects in the control group, that is, control group may learn that their compensation is less than others (members in the treatment group) and therefore work less hard. Hence, the treatment effect has a direct positive impact in the productivity of the treatment group and a negative one in that of the control group. As a result the ATE will be overstated as the gap it measures accounts for more than just the increase in productivity of the treatment group members.

d. When Olken (2007) randomly audits local Indonesian governments for evidence of corruption, suppose control-group governments learn that treatment-group governments are being randomly audited and assume they are likely to get audited too.

Answer

Under this case, the ATE would also overestimate the true effect of the audits. Essentially there is interference with the control group as they would actively take steps towards hiding evidence that otherwise would have been caught and then appear to be less corrupt as a whole. This in essence has a direct impact in the level corruption assessed for the control groups.

FE exercise 8.2

National surveys indicate that college roommates tend to have correlated weight. The more one roommate weights at the end of the freshman year, the more the other freshman roommate weights. On the other hand, researchers studying housing arrangements in which roommates are randomly paired together find no correlation between two roommates' weights at the end of their freshman year. *Explain how these two facts can be reconciled.*

Answer

The fact that the national survey data captures a correlation between two roommates weights doesn't shed any light into the underlying details on housing arrangements. If roommates have total freedom to choose who they live with at the beginning of the year, they might find themselves choosing people with similar lifestyles, similar backgrounds or even choose to live with existing acquaintances, which could very well have similar eating habits, and be more likely to develop a social lifes within the same circles.

Conversely, as part of the randomized experiment, housing arrangements are made without regard to any pre-existing criteria. As a result, the chances of two roommates having similar eating habits at the begining of the experiment are totally random (hence less likely). Furthermore, such random assignment would also lead for roommates to have just as different cultural backgrounds and interests and could result in a situation where each roommate develops his/her own social life totally independently and hence have less influence over each others actions (including eating habits!).

FE exercise 8.10

A doctoral student conducted an experiment in which she randomly varied whether she ran or walked 40 minutes each morning. In the middle of the afternoon over a period of 26 days she measured the following outcome variables: (1) her weight; (2) her score in Tetris; (3) her mood on a 0-5 scale; (4) her energy; and (5) whether she got a question right on the math GRE.

```
In [34]: d2 <- read.dta("./data/Hough_WorkingPaper_2010.dta")
         head(d2)
```

day	run	weight	tetris	mood	energy	appetite	gre
1	1	21	11092	3	3	0	1
2	1	21	14745	3	1	2	0
3	0	20	11558	3	3	0	1
4	0	21	11747	3	1	1	1
5	0	21	14319	2	3	3	1
6	1	19	7126	3	2	0	1

a. Suppose you were seeking to estimate the average effect of running on her Tetris score. Explain the assumptions needed to identify this causal effect based on this within-subjects design. Are these assumptions plausible in this case? What special concerns arise due to the fact that the subject was conducting the study, undergoing the treatments, and measuring her own outcomes?

Answer

In order to conduct a within subject experiment and identify causal effects we'd need to take into account the following two assumptions:

- **Non anticipation:** potential outcomes are unaffected by treatments that are administered in the future.
- **Non persistence:** potential outcomes are unaffected by prior interventions.

While the nonanticipation assumption is indeed plausible, the non persistence is clearly not as her Tetris score certainly affected by past interventions.

Additional concerns arise from the fact the subject was conducting her own outcomes. All sorts of biases could have influenced the measurement of the data points across the board

b. Estimate the effect of running on Tetris score. Use regression to test the sharp null hypothesis that running has no immediate or lagged effect on Tetris scores. (**Note** the book calls for randomization inference, but this is a bit of a tough coding problem. **HINT:** For the second part of part (b), run one regression of today's score on both today's treatment assignment and yesterday's treatment assignment. Then, calculate the p-value that both effects are zero.)

Answer

Measuring ATE using a simple (average) approach

```
In [35]: mean(na.omit(d2[d2$run == 1,$tetris])) - mean(na.omit(d2[d2$run == 0,$tetris]))
```

13613.1

Regressing on treatment assignment

```
In [36]: #adding prior day run
assignment = cbind(d2$run)
#increasing it by one
assignment <- append(assignment, NaN, after = 0)
d2$priorTreatment <- assignment[1:length(assignment)-1]
d2
```

day	run	weight	tetris	mood	energy	appetite	gre	priorTreatment
1	1	21	11092	3	3	0	1	NaN
2	1	21	14745	3	1	2	0	1
3	0	20	11558	3	3	0	1	1
4	0	21	11747	3	1	1	1	0
5	0	21	14319	2	3	3	1	0
6	1	19	7126	3	2	0	1	0
7	0	20	16067	3	4	0	0	1
8	0	20	3939	3	2	0	1	0
9	1	21	28230	4	2	0	0	0
10	0	21	17396	4	4	1	1	1
11	1	20	36152	1	4	0	0	0
12	0	20	16567	4	4	1	1	1
13	0	20	NA	NA	NA	NA	NA	0
14	1	18	11853	4	2	0	1	0
15	1	18	20433	4	2	2	1	1
16	1	18	20701	3	4	0	0	1
17	0	20	NA	NA	NA	NA	1	1
18	1	19	17509	3	3	1	1	0
19	0	21	9779	3	3	1	0	1
20	0	22	18598	3	3	1	1	0
21	1	20	36665	2	3	0	1	0
22	0	21	8094	4	3	1	1	1
23	1	19	48769	2	5	0	0	0
24	1	20	22601	4	4	1	1	1
25	1	19	37950	4	4	0	1	1
26	1	20	56047	4	4	0	1	1

On treatment assignment:

```
In [37]: treatRegression <- lm(tetris ~ run, data = d2)
         coeftest(treatRegression)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	12806.4	3708.5	3.4532	0.002264	**
run	13613.1	4855.6	2.8036	0.010351	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

P-value suggests that the sharp null hypothesis of no immediate treatment effect can be rejected at a 99% confidence level.

```
In [38]: priorTreatRegression <- lm(priorTreatment ~ run, data = d2)
         coeftest(priorTreatRegression)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.58333	0.14924	3.9086	0.0007057	***
run	-0.12179	0.20696	-0.5885	0.5619394	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

P-value suggests that the sharp null hypothesis of no effect of prior treatment (lagged effect) can not be rejected at a relevant statistical significance level. This seems to support the argument that the non persistence assumption holds.

c. One way to lend credibility to with-subjects results is to verify the no-anticipation assumption. Use the variable `run` to predict the `tetris` score *on the preceeding day*. Presume that the randomization is fixed. Why is this a test of the no-anticipation assumption? Does a test for no-anticipation confirm this assumption?

Answer

```
In [39]: #adding prior day run
tetrisMeasuresSchedule = cbind(d2$tetris)
#increasing it by one
tetrisMeasuresSchedule <- append(tetrisMeasuresSchedule, NaN, after = 0)
d2$priorDayTetrisScore <- tetrisMeasuresSchedule[1:length(tetrisMeasuresSchedule)-
d2
```

day	run	weight	tetris	mood	energy	appetite	gre	priorTreatment	priorDayTetrisScore
1	1	21	11092	3	3	0	1	NaN	NaN
2	1	21	14745	3	1	2	0	1	11092
3	0	20	11558	3	3	0	1	1	14745
4	0	21	11747	3	1	1	1	0	11558
5	0	21	14319	2	3	3	1	0	11747
6	1	19	7126	3	2	0	1	0	14319
7	0	20	16067	3	4	0	0	1	7126
8	0	20	3939	3	2	0	1	0	16067
9	1	21	28230	4	2	0	0	0	3939
10	0	21	17396	4	4	1	1	1	28230
11	1	20	36152	1	4	0	0	0	17396
12	0	20	16567	4	4	1	1	1	36152
13	0	20	NA	NA	NA	NA	NA	0	16567
14	1	18	11853	4	2	0	1	0	NA
15	1	18	20433	4	2	2	1	1	11853
16	1	18	20701	3	4	0	0	1	20433
17	0	20	NA	NA	NA	NA	1	1	20701
18	1	19	17509	3	3	1	1	0	NA
19	0	21	9779	3	3	1	0	1	17509
20	0	22	18598	3	3	1	1	0	9779
21	1	20	36665	2	3	0	1	0	18598
22	0	21	8094	4	3	1	1	1	36665
23	1	19	48769	2	5	0	0	0	8094
24	1	20	22601	4	4	1	1	1	48769
25	1	19	37950	4	4	0	1	1	22601
26	1	20	56047	4	4	0	1	1	37950

```
In [40]: runRegression <- lm(priorDayTetrisScore ~ run, data = d2)
summary(runRegression)
```

Call:

```
lm(formula = priorDayTetrisScore ~ run, data = d2)
```

Residuals:

Min	1Q	Median	3Q	Max
-15610	-7521	-2337	2424	29220

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18903.8	3335.8	5.667	1.26e-05 ***
run	645.6	4823.5	0.134	0.895

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11560 on 21 degrees of freedom

(3 observations deleted due to missingness)

Multiple R-squared: 0.0008524, Adjusted R-squared: -0.04673

F-statistic: 0.01792 on 1 and 21 DF, p-value: 0.8948

This is a reasonable approach to test the non anticipation assumption. Both the coefficient (negligible) and it's p-value (not significant) suggest there is no correlation whatsoever between the future treatment and the observed effect.

However this test does not necessarily confirm the assumption, as we're only regressing on the immediately preceeding score, rather than trying out shifting the treatment applied in subsequent times (higher lag) in the future.

d. If Tetris responds to exercise, one might suppose that energy levels and GRE scores would as well. Are these hypotheses borne out by the data?

Answer

Energy


```
In [41]: energyRegression <- lm(energy ~ run, data = d2)
summary(energyRegression)
```

Call:

```
lm(formula = energy ~ run, data = d2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.0714	-1.0179	0.0000	0.9286	1.9286

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.00000	0.33662	8.912	9.4e-09 ***
run	0.07143	0.44074	0.162	0.873

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.064 on 22 degrees of freedom

(2 observations deleted due to missingness)

Multiple R-squared: 0.001192, Adjusted R-squared: -0.04421

F-statistic: 0.02627 on 1 and 22 DF, p-value: 0.8727

The hypothesis that energy levels respond to exercise is not supported by the data. The non persistence assumption might play a bigger role here. One can argue that prior sessions of running/walking have an persistent effect on the energy levels of the subject, that is, the longer she's greater she has been subject to exersice in the past, the more conditioned her body becomes resulsting in long lasting effects on energy levels. This will produce biased estimates of the effect as there is not enough time for the effects of the prior treatments to 'wash out'

GRE

```
In [42]: greRegression <- lm(gre ~ run, data = d2)
summary(greRegression)
```

Call:

```
lm(formula = gre ~ run, data = d2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.8182	-0.6429	0.1818	0.3571	0.3571

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.8182	0.1385	5.909	5.05e-06 ***
run	-0.1753	0.1850	-0.948	0.353

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4592 on 23 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.03757, Adjusted R-squared: -0.004275

F-statistic: 0.8978 on 1 and 23 DF, p-value: 0.3532

The hypothesis that GRE responds to exercise is not supported by the data at a statistically significant level. Once again, the non persistence assumption may play a big role here. As time goes by the subjects ability to correctly answer a math question on the GRE could have improved (simply by getting better with practice!) without regards to treatment assignment.

e. **Additional Mandatory Question:** Note that the observations in this regression are not necessarily all independent of each other. In particular, think about what happens when we lag a variable. Given this non-independence, would you expect randomization inference to give you a better answer than the regression answer you just obtained in (b)? Why? Which number(s) do you expect to be different in regression than in randomization inference? What is the direction of the bias? (This is a conceptual question, so you do not need to conduct the randomization inference to answer it. However, you are certainly welcome to try that exercise if you are curious.)

Answer

The linear regression approach is overstating the treatment effect in tetris scores for multiple reasons. First tetris score can be thought of as both the post-treatment variable but also a pre-treatment variable. That is, the prior tetris score of the subject is instrumental to understand the effect of exercising in the post-treatment score.

Also, as we are dealing with a within-subject experiment, a simple regression fails to account for two problems: (1) the fact that the tetris score improves overtime - hence the difference in score should be measured, and (2) that the distribution of the treatment is skewed overtime, that is in the first couple of weeks, where tetris scores were lower, the subject exercised ~42% of the time, whereas on the last two weeks (when scores were higher) the subject exercised ~70% of the time. By virtue of that instatiation of the assignment the linear regression might attribute the effects of playing tetris over time to the treatment attribute.

Under a randomization inference approach, we would expect the p-value to be higher under a randomization inference approach as it take into account all potential realizations where the treatment assignment is as unbalanced as the one observed in the experiment.