Juanjo Carin

W241 (Field Experiments) – Problem Set #4 – MIDS Spring 2015

March 30, 2015

Contents

W241 – Problem Set #4

1. FE exercise 5.2.

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

Supposed that the subjects are ordered so that the first three subjects are the Compliers, and the last three subjects are the Never-Takers, we have:

$$ATE = E[Y_i(d=1) - Y_i(d=0)] = \frac{1}{6}\sum_{i=1}^{6}\{Y_i(d=1) - Y_i(d=0)\}$$

$$CACE = E[Y_i(d=1) - Y_i(d=0) \mid d_i(z=1) = 1] = \frac{1}{3}\sum_{i=1}^{3}\{Y_i(d=1) - Y_i(d=0)\}$$

(We would also have $ITT = E[Y_i(z=1, d(1)) - Y_i(z=0, d(0))] = \frac{1}{6}\sum_{i=1}^{6}\{Y_i(z=1, d(1)) - Y_i(z=0, d(0))\} = \frac{1}{6}\sum_{i=1}^{3}\{Y_i(d=1) - Y_i(d=0)\} = \frac{CACE}{2}$, and $ITT_D = E[d_i(z=1) - d_i(z=1)] = \frac{3}{6} = \frac{1}{2}$.)

If the $ATE > 0$ and the $CACE < 0$, the latter equation implies that:

$$\sum_{i=1}^{3} Y_i(d=1) < \sum_{i=1}^{3} Y_i(d=0)$$

So we could write $\sum_{i=1}^{3} Y_i(d=0)$ as $\sum_{i=1}^{3} Y_i(d=1) + \epsilon$, where $\epsilon > 0$.

The first equation implies that:

$$-\epsilon + \sum_{i=4}^{6} Y_i(d=1) - \sum_{i=4}^{6} Y_i(d=0) > 0$$

So we could write $\sum_{i=4}^{6} Y_i(d=1)$ as $\sum_{i=4}^{6} Y_i(d=0) + \epsilon + \delta$, where $\delta > 0$.

Any combination of the 12 possible values of Y_i that satisfies both conditions would work, leading to:

$$ATE = \frac{1}{6}\left(\sum_{i=1}^{3}\{Y_i(d=1) - Y_i(d=0)\} + \sum_{i=4}^{6}\{Y_i(d=1) - Y_i(d=0)\}\right) = \frac{(-\epsilon) + (\epsilon + \delta)}{6} = \frac{\delta}{6} > 0$$

$$CACE = \frac{1}{3}\sum_{i=1}^{3}\{Y_i(d=1) - Y_i(d=0)\} = -\frac{\epsilon}{3} < 0$$

To take a simple case, let's assume that the effect is constant among Compliers, and also among Never-Takers:

$$\epsilon' = \epsilon/3$$

$$\delta' = \delta/3$$

$$\forall i \in \{1,2,3\} \quad Y_i(d=1) - Y_d(i=0) = -\epsilon'$$

$$\forall i \in \{4,5,6\} \quad Y_i(d=1) - Y_d(i=0) = \epsilon' + \delta'$$

In that case:

$$ATE = \frac{3(-\epsilon') + 3(\delta' + \epsilon')}{6} = \frac{\delta'}{2} = \frac{\delta}{6}$$

$$CACE = \frac{3(-\epsilon')}{3} = -\epsilon' = -\frac{\epsilon}{3}$$

(And, of course, $ITT = -\epsilon/6$.)

Based on that, a hypothetical schedule of potential outcomes could be:

Observation	$Y_i(d=0)$	$Y_i(d=1)$	$d_i(z=0)$	$d_i(z=1)$	Type
1	α	$\alpha - \epsilon'$	0	1	Complier
2	β	$\beta - \epsilon'$	0	1	Complier
3	γ	$\gamma - \epsilon'$	0	1	Complier
4	λ	$\lambda + \epsilon' + \delta'$	0	0	Never-Taker
5	μ	$\mu + \epsilon' + \delta'$	0	0	Never-Taker
6	ν	$\nu + \epsilon' + \delta'$	0	0	Never-Taker

for any values of α, β, γ, λ, μ, and ν, as long as $\epsilon', \delta' > 0$.

I.e., if the effect for each Complier is negative, and the effect for each Never-Taker is positive, and larger than the absolute value of the effect for the Compliers. (That was just one possibility; other possible schedule would be the same as the previous, but with no effect for two of the Compliers and two of the Never-Takers, and so forth.)

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

Depending on which subjects are assigned to treatment, the estimated CACE would be negative (if at least one of the Compliers is assigned to treatment) or null (if only Never-Takers are assigned to treatment). (We should keep in mind that the value of $-\epsilon/3$ that we obtained is the average; the estimate would depend on the realization).

The CACE would be misleading if we assume that the ATE has the same sign—i.e., that it is negative (or zero, in 1 of the $\binom{6}{3} = 20$ possible allocations, supposed that the treatment and control groups have the same size), when it is actually the opposite. In other words, that the effect would be negative for everyone, supposed that we are able to treat them.

When wouldn't the CACE be so misleading? To explore that possibility we should analyze the definition of "Never-Takers": the word "never" in that term can be misleading as well—it just means that, for a particular realization of an experiment, the treatment was not applied to those subjects, even when they were assigned to it. But that can change: the experimental can avoid or reduce the one-sided noncompliance... as well as who is a Complier or a Never-Taker[1]. If Never-Takers were always the same (i.e., if they would really **never** be treated), then the CACE would be informative, because we would always treat those for whom the effect is negative, and none of those for whom the effect is positive (and larger in absolute value).

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?

In terms of analyzing the results (which drive decision making), the set of Compliers can be more relevant, because it allows us to estimate the CACE, i.e., the treatment effect (among Compliers), not just the intent-to-treat effect. If we are not able to determine the size of this set[2] (and hence the compliance rate), we can only estimate the ITT, which would be OK if we're just interested in the overall effectiveness of a program (in that case the whole set is enough).

But I would add that in terms of designing the experiment, the set of Compliers plus Never-Takers is more relevant, mainly because we usually don't know in advance who the Compliers are (so we cannot work with that subset if we don't know who's part of it). If we were able to know in advance what is the set of Compliers (not only its size but also who is part of it; that would imply that the Never-Takers would actually **never** be treated, no matter how we try to apply the treatment), we could include only Compliers in our sample—and exclude all Never-Takers—, but I think that's rarely the case: we know that (sometimes only the size of the group) **after** the experiment has been conducted.

---

[1] Going back to the example of canvassing, some Never-Takers would have been at home (and some Compliers would have not) if the canvassers had gone at a different time. Supposed that we had a situation like the one we're discussing in this exercise, the CACE would be positive in some realizations of the experiment, where a Never-Taker becomes a Complier, and vice versa.

[2] E.g., if we're not able to accurately measure who has been effectively treated, or up to what extent.

2. FE exercise 5.6.

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

Since we don't know the voter turnout rate among the Compliers, first we need to calculate the ITT and the compliance rate (a.k.a. the intent-to-treat effect of z_i on d_i):

$$ITT = E[Y_i(z=1)] - E[Y_i(z=0)] = \frac{400}{1000} - \frac{700}{2000} = 40\% - 35\% = 5\%$$

$$ITT_D = \alpha = E[d_i(z=1)] - E[d_i(z=0)] = \frac{500}{1000} - \frac{0}{2000} = 50\%$$

$$CACE = \frac{ITT}{ITT_D} = \frac{0.05}{0.5} = 0.05 \times 2 = \boxed{10\%}$$

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

The ITT would remain, but the compliance rate would change in this case:

$$ITT_D = \alpha = E[d_i(z=1)] - E[d_i(z=0)] = \frac{250}{1000} - 0 = 25\%$$

$$CACE = \frac{ITT}{ITT_D} = \frac{0.05}{0.25} = 0.05 \times 4 = \boxed{20\%}$$

c. Do the canvassers' exaggerated reports make their efforts seem more or less effective? When formulating your answer, you may define effectiveness in terms of either the ITT or the CACE.

Less effective in terms of the CACE. The ITT, which is related to the overall effectiveness (among Compliers and non-Compliers) would remain constant, as commented above.

3. FE exercise 5.10.

a. Estimate the ITT.

First we read the dataset and explore it (noticing that the numbers given in the book—at least in the 1st edition, 2012—are not always the same than we get from the dataset):

```
# Read dataset
library(foreign)
Peking <- read.dta("Guan_Green_CPS_2006.dta.dta")

# Explore dataset
library(pastecs)
format(stat.desc(subset(Peking, select = -dormid )), digits = 0,
       scientific = FALSE)
```

```
##                turnout contact treat2
## nbr.val          4022    4024   4024
## nbr.null          978    1641   1334
## nbr.na              2       0      0
## min                 0       0      0
## max                 1       1      1
## range               1       1      1
## sum              3044    2383   2690
## median              1       1      1
## mean                1       1      1
## SE.mean             0       0      0
## CI.mean.0.95        0       0      0
## var                 0       0      0
## std.dev             0       0      0
## coef.var            1       1      1
```

```
# Omit NA values
    # 2 observations without information about turnout
Peking <- na.omit(Peking)

# Check there is only one-sided noncompliance
    # Not everyone who was assigned was treated
    # but everyone who was not assigned was not treated
Peking_table <- table(Peking[, c("treat2", "contact", "turnout")])
ftable(addmargins(Peking_table, FUN = list(Total = sum), quiet = TRUE))
```

```
##                 turnout    0    1 Total
## treat2 contact
## 0      0                 442  892  1334
##        1                   0    0     0
##        Total             442  892  1334
## 1      0                  70  237   307
##        1                 466 1915  2381
##        Total             536 2152  2688
## Total  0                 512 1129  1641
##        1                 466 1915  2381
##        Total             978 3044  4022
```

```
# Turnout rates in each group
round(100*prop.table(Peking_table, c(2, 1)), 2)
```

```
## , , turnout = 0
##
##        contact
## treat2    0      1
##      0 33.13
##      1 22.80 19.57
##
## , , turnout = 1
##
##        contact
## treat2    0      1
##      0 66.87
##      1 77.20 80.43
```

Omitting the observations with NA values, the dataset contains 4022 observations. Of the 2688 students assigned to the treatment group, 2381 were contacted. A total of 2152 students in the treatment group voted; of the 1334 students assigned to the control group, 892 voted.

```
# Function to estimate the ITT (without SEs)
est.ITT <- function(outcome, assign)
    mean(outcome[assign ==1 ]) - mean(outcome[assign == 0])

(Peking_ITT_estimate <- est.ITT(Peking$turnout, Peking$treat2))
```

```
## [1] 0.1319296
```

As shown above, the ITT is 0.132 = 13.2%.

b. Use randomization inference or regression 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.

I'll use both methods. First I use randomization inference, using some of the functions of the "*ri*" package (as demonstrated in this script from the book:

```
library(ri)
perms <- genperms(Peking$treat2, maxiter=10e3, clustvar = Peking$dormid)
```

```
## Too many permutations to use exact method.
## Defaulting to approximate method.
## Increase maxiter to at least 5.83377797524832e+275 to perform exact estimation.
```

```
numiter <- ncol(perms)
probs <- genprobexact(Peking$treat2, clustvar = Peking$dormid)
Peking_ITT_RI <- estate(Peking$turnout, Peking$treat2, prob = probs)
Ys <- genouts(Peking$turnout, Peking$treat2, ate = 0)
distout <- gendist(Ys, perms, prob=probs)
sum(distout >= Peking_ITT_RI) # one-tailed comparison
```

```
## [1] 0
```

```r
sum(abs(distout) >= abs(Peking_ITT_RI)) # two-tailed comparison
```

```
## [1] 0
```

```r
dispdist(distout, Peking_ITT_RI, display.plot = FALSE)
```

```
## $two.tailed.p.value
## [1] 0
##
## $two.tailed.p.value.abs
## [1] 0
##
## $greater.p.value
## [1] 0
##
## $lesser.p.value
## [1] 1
##
## $quantile
##        2.5%       97.5%
## -0.04285999  0.04318501
##
## $sd
## [1] 0.02186358
##
## $exp.val
## [1] 8.647986e-05
```

The "*two.tailed.p.value.abs*" parameter shown above is 0, which indicates that none of the 10000 randomizations that were simulated assuming the sharp null hypothesis yielded an absolute estimated ITT greater than or equal to the absolute estimated ITT (which is equal to what we had previously got: $0.132 = 13.2\%$). That's equivalent to a two-sided p-value smaller than 10^{-4}.

Next we use regression:

```r
Peking_ITT_reg <- lm(turnout ~ treat2, data = Peking)

# Function to calculate CSEs
cl <- function(fm, cluster){
    require(sandwich, quietly = FALSE)
    require(lmtest, quietly = FALSE)
    M <- length(unique(cluster))
    N <- length(cluster)
    K <- fm$rank
    dfc <- (M/(M-1))*((N-1)/(N-K))
    uj <- apply(estfun(fm),2, function(x) tapply(x, cluster, sum));
    vcovCL <- dfc*sandwich(fm, meat=crossprod(uj)/N)
    coeftest(fm, vcovCL)
}

cl(Peking_ITT_reg, Peking$dormid)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
```

```
## (Intercept) 0.668666    0.020241 33.0349 < 2.2e-16 ***
## treat2      0.131930    0.023271  5.6692 1.536e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Estimate_Peking_ITT <- cl(Peking_ITT_reg, Peking$dormid)[2, 1]
```

The result is in line with the previous one: the *p*-value of the estimate—again, 0.132 (0.023)—is as small as 1.5e-08 (so, on average, we should run about $1/1.5e-08 = 6.5e+07$ randomizations to get an estimated ITT under the sharp null hypothesis as great as the one we obtained).

So both results confirm that the intent-to-treat effect is highly statistically significant—it would be extremely unlikely that we obtained such a result just by chance—so we reject the sharp null hypothesis.

I would add that a 13.2 percentage-point increase in turnout, caused by canvassing, even when not everybody assigned to be treated was actually treated. also has a high practical significance.

c. Assume that the leaflet had no effect on turnout. Estimate the CACE.

```
Peking_ITTd_reg <- lm(contact ~ treat2, data = Peking)
cl(Peking_ITTd_reg, Peking$dormid)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 6.8456e-14 8.1713e-17 837.763 < 2.2e-16 ***
## treat2      8.8579e-01 1.2169e-02  72.789 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Estimate_Peking_ITTd <- cl(Peking_ITTd_reg, Peking$dormid)[2, 1])
```

```
## [1] 0.8857887
```

```
library(AER)
Peking_CACE_reg <- ivreg(turnout ~ contact, ~ treat2, data = Peking)
cl(Peking_CACE_reg, Peking$dormid)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.668666    0.020241 33.0349 < 2.2e-16 ***
## contact     0.148940    0.026311  5.6607 1.613e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Estimate_Peking_CACE <- cl(Peking_CACE_reg, Peking$dormid)[2, 1]
```

The estimated CACE is 0.149 (0.026).
(As expected, the ITT is about the 88.6% of the CACE.)

Another method to calculate the CACE is based on the following values:

```
(Turnout_Treatment <- mean(Peking$turnout[Peking$treat2 == 1]))
```

```
## [1] 0.8005952
```

```
(Turnout_Treatment_Compliers <- mean(Peking$turnout[Peking$treat2 == 1 & Peking$contact == 1]))
```

```
## [1] 0.8042839
```

```
(Turnout_Treatment_NeverTakers <- mean(Peking$turnout[Peking$treat2 == 1 & Peking$contact == 0]))
```

```
## [1] 0.771987
```

```
(ComplianceRate  <- mean(Peking$contact[Peking$treat2 == 1]))
```

```
## [1] 0.8857887
```

```
(Turnout_Control <- mean(Peking$turnout[Peking$treat2 == 0]))
```

```
## [1] 0.6686657
```

```
# Due to random assignment, we assume that the compliance rate would have been
    # the same in the control group... so we can estimate the turnout rate among
    # Never-Takers in that group
Turnout_Control_NeverTakers  <-  Turnout_Treatment_NeverTakers
# The turnout rate among Compliers in the control group can be easily
    # calculated: the overall turnout rate in that group must be the one
    # estimated above
(Turnout_Control_Compliers <- (Turnout_Control - (1 - ComplianceRate) *
                                        Turnout_Control_NeverTakers) / ComplianceRate)
```

```
## [1] 0.6553437
```

```
# Now that we know the turnout rate among Compliers in both groups, the CACE
    # is just the difference between both quantities
```

So this situation—where the leaflet is assumed to have no effect on turnout— can be summarized as follows:

	Compliers	Never-Takers	Total
Proportion	88.58%	11.42%	100%
Turnout rate in Treatment	80.43%	77.2%	80.06%
Turnout rate in Control	65.53%	77.2%	66.87%
Effect	14.89%	0%	13.19%
	CACE		ITT

d. Assume that the leaflet raised the probability of voting by one percentage point among both Compliers and Never-Takers. Write down a model of the expected turnout rates in the treatment and control groups, incorporating the average effect of the leaflet.

Incorporating the average effect of the leaflet does not change the overall turnout rates in both groups, and hence the ITT (I mean, it does change the turnout rate in the treatment group, but that change is already included in the ITT value of 88.6% that we already estimated—so we expect the effect of the canvasser's speech to be lower), but only the expected turnout rates among the Compliers and the Never-Takers in the control group.

Now the difference between each turnout rate (i.e., the ITT) is an effect of the leaflet (among all the participants assigned to treatment) plus the canvasser's speech (among those who answered the door: the Compliers).

The ratio of Compliers in the treatment group (i.e., the compliance rate) does not change, either, but we can no longer estimate the CACE as the division of those two parameters, because the **exclusion restriction** is violated: the assignment z_i, and not only the treatment d_i, has an effect in the outcome Y_i.

In general, the ITT can be decomposed as follows:

$ITT = E[Y_i(z=1)] - E[Y_i(z=0)] = E[Y_i(z=1, d(1))] - E[Y_i(z=0, d(0))]$
$= \{E[Y_i(z=1, d=1) \mid d_i(1) = 1]ITT_D + E[Y_i(z=1, d=0) \mid d_i(1) = 0](1 - ITT_D)\}$
$- \{E[Y_i(z=0, d=0) \mid d_i(1) = 1]ITT_D + E[Y_i(z=0, d=0) \mid d_i(1) = 0](1 - ITT_D)\}$
$= \{E[Y_i(z=1, d=1) \mid d_i(1) = 1] - E[Y_i(z=0, d=0) \mid d_i(1) = 1]\} ITT_D$
$+ \{E[Y_i(z=1, d=0) \mid d_i(1) = 0] - E[Y_i(z=0, d=0) \mid d_i(1) = 0]\} (1 - ITT_D)$
$= CACE \cdot ITT_D + \{E[Y_i(z=1, d=0) \mid d_i(1) = 0] - E[Y_i(z=0, d=0) \mid d_i(1) = 0]\} (1 - ITT_D)$

Under the exclusion restriction assumption, the second term of the last expression is zero. When that assumption is violated (as we are supposing now), the ITT (i.e., the difference between the expected turnout rate in the entire treatment and control groups) is a result of the average effect of the leaflet among those not home and the effects of conversations and the leaflets among those who are home:

$$ITT = ITT_D \cdot CACE + (1 - ITT_D) \cdot E[Y_i(z=1, d=0) - Y_i(z=0, d=0)]$$

So, as we mentioned, the effect among those who are home is smaller than before (when it was assumed to be the only cause):

$$CACE = \frac{ITT}{ITT_D} - \frac{1 - ITT_D}{ITT_D} E[Y_i(z=1, d=0) - Y_i(z=0, d=0)]$$

For this particular example (where the effect of the leaflet is $E[Y_i(z=1, d=0) - Y_i(z=0, d=0)] = 1\%$), the model would be:

$$0.1319 = 0.8858 \cdot CACE + (1 - 0.8858) \cdot 0.01$$

e. Given this assumption, estimate the CACE of canvassing.

We just need to solve the previous equation:

```
(New_Estimate_Peking_CACE <- Estimate_Peking_ITT / Estimate_Peking_ITTd -
    (1 - Estimate_Peking_ITTd) / Estimate_Peking_ITTd * 0.01)
```

```
## [1] 0.1476509
```

So the CACE is <mark>0.1477,</mark> slightly smaller than what we had previously obtained (0.1489, which was overestimated under this new assumption), as we expected. The difference is small because the effect of assignment is small and the compliance rate is high.

This new estimate of the CACE includes the effect of both the leaflet and the canvasser's speech (because the turnout rate of those who hear the latter are also affected by the former).

Another explanation about how to estimate the CACE is as follows:

Since the CACE is $E[(Y_i(d=1) - Y_i(d=0)) \mid d_i(1) = 1]$, we'd need to know $E[Y_i(d=0) \mid d_i(1) = 1]$. We are not able to observe the turnout rate of Compliers in the treatment group had they not been assigned to treatment, but—thanks to the random assignment—we know that (at least on average) the compliance rate would be the same in the control group, and hence how many people in the control group would have answered the door had they been assigned to treatment.

Therefore, we know that the overall turnout rate in the control group could be decomposed as follows:

$$E[Y_i(z=0)] = E[Y_i(z=0) \mid d_i(1)=1]ITT_D + E[Y_i(z=0) \mid d_i(1)=0](1 - ITT_D)$$

$$\Rightarrow E[Y_i(z=0) \mid d_i(1)=1] = \frac{E[Y_i(z=0)] - E[Y_i(z=0) \mid d_i(1)=0](1 - ITT_D)}{ITT_D}$$

We know the values of every term in the right side of the last expression (including $E[Y_i(z=0) \mid d_i(1)=0] = E[Y_i(z=1) \mid d_i(1)=0] - 1\%$), so we can build a table like the one in part (c), for this new situation:

```
(Turnout_Control_NeverTakers  <-  Turnout_Treatment_NeverTakers - 0.01)
```

```
## [1] 0.761987
```

```
(Turnout_Control_Compliers <- (Turnout_Control - (1 - ComplianceRate) *
                                    Turnout_Control_NeverTakers) / ComplianceRate)
```

```
## [1] 0.6566331
```

	Compliers	Never-Takers	Total
Proportion	88.58%	11.42%	100%
Turnout rate in Treatment	80.43%	77.2%	80.06%
Turnout rate in Control	65.66%	76.2%	66.87%
Effect	14.77%	1%	13.19%
	CACE		ITT

f. Suppose, instead, that the leaflet had no effect on Compliers (who heard the canvasser's speech and ignore the leaflet) but raised turnout among Never-Takers by 3 percentage points. Given this assumption, estimate the CACE of canvassing.

We just have to solve the following equation again:

$$CACE = \frac{ITT}{ITT_D} - \frac{1 - ITT_D}{ITT_D} E[Y_i(z=1, d=0) - Y_i(z=0, d=0)]$$

Taking into account that now $E[Y_i(z = 0) \mid d_i(1) = 0] = E[Y_i(z = 1) \mid d_i(1) = 0] - 1\%$.

```
(Newest_Estimate_Peking_CACE <- Estimate_Peking_ITT / Estimate_Peking_ITTd -
    (1 - Estimate_Peking_ITTd) / Estimate_Peking_ITTd * 0.03)
```

[1] 0.1450721

Now the CACE is 0.1451. It is even lower to what we had previously obtained (0.1477), because the effect of the assignment to treatment (i.e., the effect of the leaflet) is bigger.

The fact that the leaflet has no effect on Compliers implies that now the CACE only includes—is a result of—the effect of the canvasser's speech (the treatment).

Same as before, another way to obtain this value would be building a new table:

```
(Turnout_Control_NeverTakers  <-  Turnout_Treatment_NeverTakers - 0.03)
```

[1] 0.741987

```
(Turnout_Control_Compliers <- (Turnout_Control - (1 - ComplianceRate) *
                                    Turnout_Control_NeverTakers) / ComplianceRate)
```

[1] 0.6592118

	Compliers	Never-Takers	Total
Proportion	88.58%	11.42%	100%
Turnout rate in Treatment	80.43%	77.2%	80.06%
Turnout rate in Control	65.92%	74.2%	66.87%
Effect	14.51%	3%	13.19%
	CACE		ITT

4. FE exercise 5.11.

a. 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.

We could estimate those proportions directly from the table in page 171 of the book:

```
(ComplianceRate_Treatment <- 486 / (486 + 2086))
```

```
## [1] 0.188958
```

```
(ComplianceRate_Placebo <- 470 / (470 + 2109))
```

```
## [1] 0.1822412
```

The proportions are 18.9% and 18.22%, respectively.

But since we have to find out whether they are statistically significantly different from each other, we will first recreate the data in the table:

```
# Create the dataframe based on the table in page 171
Nickerson <- data.frame(Assignment = c(rep("Baseline", 2572),
                                       rep("Treatment", 486 + 2086),
                                       rep("Placebo", 470 + 2109)),
                        Treated = c(rep(0, 2572), rep(1, 486), rep(0, 2086),
                                    rep(1, 470), rep(0, 2109)),
                        Turnout = c(rep(1, round(31.22/100*2572)),
                                    rep(0, 2572 - round(31.22/100*2572)),
                                    rep(1, round(39.09/100*486)),
                                    rep(0, 486 - round(39.09/100*486)),
                                    rep(1, round(32.74/100*2086)),
                                    rep(0, 2086 - round(32.74/100*2086)),
                                    rep(1, round(29.79/100*470)),
                                    rep(0, 470 - round(29.79/100*470)),
                                    rep(1, round(32.15/100*2109)),
                                    rep(0, 2109 - round(32.15/100*2109))))
# Not required, but randomly permute the order of subjects
Nickerson <- Nickerson[sample(1:dim(Nickerson)[1], replace = FALSE), ]

Nickerson$Treated <- factor(Nickerson$Treated, levels = c(0:1),
                            labels = c("No", "Yes"))
Nickerson$Turnout <- factor(Nickerson$Turnout, levels = c(0:1),
                            labels = c("No", "Yes"))

# Check that we've correctly replicated the table in page 171
Nickerson_table <- table(Nickerson)
ftable(addmargins(Nickerson_table, FUN = list(Total = sum), quiet = TRUE))
```

```
##                    Turnout   No  Yes Total
## Assignment Treated
```

```
## Baseline    No              1769  803  2572
##             Yes                0    0    0
##             Total           1769  803  2572
## Placebo     No              1431  678  2109
##             Yes              330  140   470
##             Total           1761  818  2579
## Treatment   No              1403  683  2086
##             Yes              296  190   486
##             Total           1699  873  2572
## Total       No              4603 2164  6767
##             Yes              626  330   956
##             Total           5229 2494  7723
```

```
round(100*prop.table(Nickerson_table, c(2, 1)), 2)
```

```
## , , Turnout = No
##
##            Treated
## Assignment    No   Yes
##    Baseline  68.78
##    Placebo   67.85 70.21
##    Treatment 67.26 60.91
##
## , , Turnout = Yes
##
##            Treated
## Assignment    No   Yes
##    Baseline  31.22
##    Placebo   32.15 29.79
##    Treatment 32.74 39.09
```

```
Nickerson$Turnout <- as.numeric(Nickerson$Turnout == "Yes")
Nickerson$Treated <- as.numeric(Nickerson$Treated == "Yes")

Nickerson$Assigned_Treatment <- ifelse(Nickerson$Assignment == "Treatment", 1, 0)
Nickerson$Assigned_Placebo <- ifelse(Nickerson$Assignment == "Placebo", 1, 0)
```

To find out the compliance rate among each group, we just need to run a regression of the treatment on the assignment (to any of the groups which receive a treatment):

```
# Nickerson_ITTd_Treatment <- lm(Treated ~ Assigned_Treatment,
#                                Nickerson[Nickerson$Assignment != "Placebo", ])
# Nickerson_ITTd_Placebo <- lm(Treated ~ Assigned_Placebo,
#                                Nickerson[Nickerson$Assignment != "Treatment", ])
Nickerson_ITTd <- lm(Treated ~ Assigned_Treatment + Assigned_Placebo, Nickerson)
summary(Nickerson_ITTd)
```

```
##
## Call:
## lm(formula = Treated ~ Assigned_Treatment + Assigned_Placebo,
##     data = Nickerson)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1890 -0.1890 -0.1822  0.0000  0.8178
##
```

```
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.669e-15  6.262e-03    0.00        1
## Assigned_Treatment 1.890e-01 8.855e-03   21.34   <2e-16 ***
## Assigned_Placebo   1.822e-01 8.849e-03   20.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3176 on 7720 degrees of freedom
## Multiple R-squared:  0.07061,    Adjusted R-squared:  0.07037
## F-statistic: 293.3 on 2 and 7720 DF,  p-value: < 2.2e-16
```

As expected, we obtain the same results... which now we can confirm they're statistically significant.

That was equivalent to an ANOVA analysis of the 3 possible assignments:

```
summary.lm(aov(Treated ~ Assignment, Nickerson))
```

```
##
## Call:
## aov(formula = Treated ~ Assignment, data = Nickerson)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1890 -0.1890 -0.1822  0.0000  0.8178
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.669e-14  6.262e-03    0.00        1
## AssignmentPlacebo  1.822e-01  8.849e-03   20.59   <2e-16 ***
## AssignmentTreatment 1.890e-01 8.855e-03   21.34   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3176 on 7720 degrees of freedom
## Multiple R-squared:  0.07061,    Adjusted R-squared:  0.07037
## F-statistic: 293.3 on 2 and 7720 DF,  p-value: < 2.2e-16
```

But are they statistically significantly different from each other? To answer that question, we might simply run a difference-in-means analysis:

```
t.test(Nickerson$Treated[Nickerson$Assigned_Treatment == 1],
       Nickerson$Treated[Nickerson$Assigned_Placebo == 1])
```

```
##
## 	Welch Two Sample t-test
##
## data:  Nickerson$Treated[Nickerson$Assigned_Treatment == 1] and Nickerson$Treated[Nickerson$Assigned_Placebo ==
## t = 0.6199, df = 5147.566, p-value = 0.5354
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01452611  0.02795977
## sample estimates:
## mean of x mean of y
## 0.1889580 0.1822412
```

Which, again, is equivalent to an ANOVA analysis, this time excluding the baseline:

```
summary.lm(aov(Treated ~ Assignment,
                    Nickerson[Nickerson$Assignment != "Baseline", ]))
```

```
##
## Call:
## aov(formula = Treated ~ Assignment, data = Nickerson[Nickerson$Assignment !=
##     "Baseline", ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1890 -0.1890 -0.1822 -0.1822  0.8178
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.182241   0.007657   23.80   <2e-16 ***
## AssignmentTreatment 0.006717   0.010836    0.62    0.535
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3888 on 5149 degrees of freedom
## Multiple R-squared:  7.462e-05,  Adjusted R-squared:  -0.0001196
## F-statistic: 0.3843 on 1 and 5149 DF,  p-value: 0.5354
```

The *p*-value (of the *F*-statitisc) is 0.54, which means that **the difference between both proportions is not statitiscally significant** (i.e., we cannot reject the null hypothesis that both assignments—treatment or placebo—lead to the same proportion of Compliers).

That confirms what we expected: due to the random assignment, each of the 3 groups (including the Baseline, although we cannot confirm this in its case) should include the same proportion of Compliers (as well as the proportion of any other characteristic), at least on average.

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

The question we're asked is whether the difference between the corresponding turnout rates (32.74% and 32.15%) is statistically significant:

```
# Never-Takers in Placebo vs. Never-Takers in Treatment
t.test(Nickerson$Turnout[Nickerson$Assigned_Treatment == 1 &
                             Nickerson$Treated == 0],
       Nickerson$Turnout[Nickerson$Assigned_Placebo == 1 &
                             Nickerson$Treated == 0])
```

```
##
##  Welch Two Sample t-test
##
## data:  Nickerson$Turnout[Nickerson$Assigned_Treatment == 1 & Nickerson$Treated ==  and Nickerson$Turnout[Nickers
## t = 0.4109, df = 4191.963, p-value = 0.6812
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02240809  0.03429115
## sample estimates:
## mean of x mean of y
## 0.3274209 0.3214794
```

```
# Another way to do that
summary.lm(aov(Turnout ~ Assignment,
                    Nickerson[Nickerson$Assignment != "Baseline" &
                              Nickerson$Treated == 0, ]))
```

```
##
## Call:
## aov(formula = Turnout ~ Assignment, data = Nickerson[Nickerson$Assignment !=
##     "Baseline" & Nickerson$Treated == 0, ])
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -0.3274 -0.3274 -0.3215  0.6726  0.6785
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.321479   0.010197  31.528   <2e-16 ***
## AssignmentTreatment 0.005942   0.014460   0.411    0.681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4683 on 4193 degrees of freedom
## Multiple R-squared:  4.027e-05,  Adjusted R-squared:  -0.0001982
## F-statistic: 0.1688 on 1 and 4193 DF,  p-value: 0.6812
```

Yes, the difference in the Never-Takers' turnout between the treatment and placebo groups is not statistically significant ($p = 0.68$).

That allows us to directly compare Compliers in the treatment group to Compliers in the placebo group—we don't have to assume that their proportion is similar, as we would do with the control group, because we know exactly what that proportion is.

In addition, comparison to the turnout rates of the reminding subjects (those in the control group, and Compliers in both the treatment and placebo group)) would allow us to verify that neither the placebo nor the assignment itself has an effect on the turnout.

```
# Baseline vs. Never-Takers in Placebo & Never-Takers in Treatment
summary.lm(aov(Turnout ~ Assignment, Nickerson[Nickerson$Treated == 0, ]))
```

```
##
## Call:
## aov(formula = Turnout ~ Assignment, data = Nickerson[Nickerson$Treated ==
##     0, ])
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -0.3274 -0.3215 -0.3122  0.6726  0.6878
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.312208   0.009198  33.945   <2e-16 ***
## AssignmentPlacebo   0.009271   0.013703   0.677    0.499
## AssignmentTreatment 0.015213   0.013744   1.107    0.268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4665 on 6764 degrees of freedom
```

```
## Multiple R-squared:  0.000187,   Adjusted R-squared:  -0.0001086
## F-statistic: 0.6327 on 2 and 6764 DF,  p-value: 0.5312
```

```r
# Baseline vs. Compliers in Placebo & Compliers in Treatment
summary.lm(aov(Turnout ~ Assignment,
                    Nickerson[Nickerson$Assignment == "Baseline" |
                                        Nickerson$Treated == 1, ]))
```

```
##
## Call:
## aov(formula = Turnout ~ Assignment, data = Nickerson[Nickerson$Assignment ==
##     "Baseline" | Nickerson$Treated == 1, ])
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -0.3910 -0.3122 -0.3122  0.6878  0.7021
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.312208   0.009194  33.959  < 2e-16 ***
## AssignmentPlacebo   -0.014336   0.023389  -0.613 0.539962
## AssignmentTreatment  0.078738   0.023062   3.414 0.000647 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4663 on 3525 degrees of freedom
## Multiple R-squared:  0.003677,   Adjusted R-squared:  0.003111
## F-statistic: 6.504 on 2 and 3525 DF,  p-value: 0.001515
```

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

```r
Nickerson_CACE_placebo <-
    ivreg(Turnout ~ Treated, ~ Assigned_Placebo,
            data = Nickerson[Nickerson$Assignment != "Treatment", ])

# A function to calculate RSEs
RSEs <- function(model){
    require(sandwich, quietly = TRUE)
    require(lmtest, quietly = TRUE)
    newSE <- vcovHC(model)
    coeftest(model, newSE)
    }

RSEs(Nickerson_CACE_placebo)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3122084  0.0091408 34.1555   <2e-16 ***
## Treated     0.0272650  0.0710617  0.3837   0.7012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The CACE of receiving the placebo is <mark>0.027</mark> (0.071). It is not statistically significant at all, which is consistent with the assumption that the placebo has no effect on 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. Second, compare turnout rates among Compliers in both the treatment and placebo groups. Interpret the results.**

The first method can be accomplished by simply dividing the ITT by the ITT_D:

```
(Nickerson_ITT_Treatment <-
    mean(Nickerson$Turnout[Nickerson$Assignment == "Treatment"]) -
    mean(Nickerson$Turnout[Nickerson$Assignment == "Baseline"]))
```

```
## [1] 0.02721617
```

```
(Nickerson_ITTd_Treatment <-
    dim(Nickerson[Nickerson$Assignment == "Treatment" &
                     Nickerson$Treated == 1, ])[1] /
    dim(Nickerson[Nickerson$Assignment == "Treatment", ])[1])
```

```
## [1] 0.188958
```

```
(Nickerson_CACE_Treatment <- Nickerson_ITT_Treatment / Nickerson_ITTd_Treatment)
```

```
## [1] 0.1440329
```

Or by using regression:

```
Nickerson_CACE_Treatment_reg <-
    ivreg(Turnout ~ Treated, ~ Assigned_Treatment,
            data = Nickerson[Nickerson$Assignment != "Placebo", ])
RSEs(Nickerson_CACE_Treatment_reg)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3122084  0.0091408 34.1555  < 2e-16 ***
## Treated     0.1440329  0.0691939  2.0816  0.03743 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Or by a combination of both:

```
Nickerson_ITT_Treatment_reg <-
    lm(Turnout ~ Assigned_Treatment,
            data = Nickerson[Nickerson$Assignment != "Placebo", ])
Nickerson_ITTd_Treatment_reg <-
    lm(Treated ~ Assigned_Treatment,
            data = Nickerson[Nickerson$Assignment != "Placebo", ])
RSEs(Nickerson_ITT_Treatment_reg)[2, 1] /
    RSEs(Nickerson_ITTd_Treatment_reg)[2, 1]
```

```
## [1] 0.1440329
```

The estimate we get is 0.144 (0.069).

When we compare turnout rates among Compliers in both the treatment and placebo groups we obtain:

```
Nickerson_CACE_Treatment_Placebo_reg <-
    lm(Turnout ~ Assigned_Treatment, Nickerson[Nickerson$Treated == 1, ])
RSEs(Nickerson_CACE_Treatment_Placebo_reg)
```

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         0.297872   0.021140 14.0906 < 2.2e-16 ***
## Assigned_Treatment 0.093074   0.030641  3.0376  0.002449 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I.e., an estimated CACE of 0.093 (0.031).

Both estimates are statistically significant. The reason why they are not equal to each other is that, due to sampling variability (and possibly heterogeneity of the treatment):

1. the proportion of Compliers based on subjects' responses to the treatment or placebo is not exactly the same (though the difference is not statistically significant)
2. the turnout rate among Never-Takers is also not exactly the same among both groups (though, again, the difference is not statiscally significant)

Had those values been exactly equal, the estimate of the CACE would have been the same using both methods.

5. Bias in estimating the ATE.

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

If treatment-group members talk to their friends in the control group about the adverts they have been exposed to, they may influence their purchase behavior, pressumably making some of them to buy (because the effect of advertising on sales was positive). That would lead to more sales in the control group than there would be in the absence of interference, and hence to a smaller difference between the treatment and control groups. So we would **underestimate** the ATE.

E.g., let's say each participant purchases $40 in the absence of treatment. Assuming the ATE is $5 (treated participants purchase $45 on average) when the non-interference assumption is met, breaking that assumption might imply say a $$42 average in the control group, so the estimated ATE would be $45 − $42 = $3.

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.

If the treatment in one location displaces the (source of the) outome to another location, assigned to control, the outcome in that latter location would be greater than it would normally be in the absence of treatment. Since the treatment causes the outcome (crime rate) to be lower, we would **overestimate** the ATE.

E.g., let's say crime rate is 10% in the absence of treatment. Assuming the ATE is −2% (crime rate turns out to be 8% in treated locations) when the non-interference assumption is met, breaking that assumption might imply say a 1.5% increase in the control group, so the estimated ATE would be 8% − 11.5% = −3.5% (greater than −2%, in absolute value).

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.

In this case we would **overestimate** the ATE: the difference between the productivity in each group comes not only from the fact that employees in the treatment work harder (and hence the minuend is greater), but also from the fact that some workers in the control group are affected by the treatment and work less hard, causing the subtrahend to be smaller.

E.g., let's say each employee produces $100 in the absence of treatment. Assuming the ATE is $20 (treated employees produce $120 on average) when the non-interference assumption is met, breaking that assumption might imply say a $10 decrease in the control-group average productivity, so the estimated ATE would be $120 − $90 = $30.

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.

This would be an example of deterrence: because governments in the control group would assume they are likely to get audited, they would also try to reduce corruption (as governments in the treatment control do), so we would **underestimate** the ATE.

E.g., let's say the corruption rate is 20% in the absence of treatment. Assuming the ATE is -10% (corruption rate is reduced to 10%) when the non-interference assumption is met, breaking that assumption might imply say a -3% decrease in the control-group corruption rate, so the estimated ATE would be $10\% - 17\% = -7\%$ (smaller than -10%, in absolute value).

---

6. FE exercise 8.2.

When people are given the freedom to choose their roommate, they tend to choose someone with similar interests and habits. E.g., it is more likely that two individuals that play a lot of sports, or that play a lot of video games and are more sedentary, or that are vegan, or that are heavy-drinkers, become roommates, than people with very different lifestyles.

Random assignment of roommates increases the variability within couples of roommates. Randomness will cause that the two members of some couples of roommates will be similar (in terms of all the aspects of lifestyle that lead to gaining weight), but they will be very different in some other couples (and hence the correlation between the two members' weight will be negative), so on average the correlation will be zero.

---

7. FE exercise 8.6.

a. Estimate $E[Y_{01} - Y_{00}]$ for the random assignment that places the treatment at location A.

In general:

$$\hat{E}[Y_{01} - Y_{00}] = \frac{\sum_{i=1}^{N} \frac{d_{01,i} Y_{01,i}}{p_{01,i}}}{\sum_{i=1}^{N} \frac{d_{01,i}}{p_{01,i}}} - \frac{\sum_{i=1}^{N} \frac{d_{00,i} Y_{00,i}}{p_{00,i}}}{\sum_{i=1}^{N} \frac{d_{00,i}}{p_{00,i}}}$$

So for this example, when location A is assigned to treatment (and hence location B is assigned to spillover):

$$\hat{E}[Y_{01} - Y_{00}] = \frac{\sum_{i=1}^{1} \frac{Y_{01,i}}{p_{01,i}}}{\sum_{i=1}^{1} \frac{1}{p_{01,i}}} - \frac{\sum_{i=3}^{5} \frac{Y_{00,i}}{p_{00,i}}}{\sum_{i=3}^{5} \frac{1}{p_{00,i}}} = \frac{\frac{0}{0.2}}{\frac{1}{0.2}} - \frac{\frac{0}{0.4} + \frac{6}{0.6} + \frac{6}{0.8}}{\frac{1}{0.4} + \frac{1}{0.6} + \frac{1}{0.8}} = \frac{0}{5} - \frac{0+10+7.5}{2.5+1.6667+1.25} = 0 - \frac{17.5}{5.4167} = -3.2308$$

```
table8_2 <- data.frame(Y00 = c(0, 6, 0, 6, 6), Y10 = c(2, 2, 4, 6, NA),
                       Y01 = c(0, 10, 4, 6, 3),
                       P00 = c(0.6, 0.4, 0.4, 0.6, 0.8),
                       P10 = c(0.2, 0.4, 0.4, 0.2, 0), P11 = rep(0.2, 5))
table8_2$exposure <- c(01, 10, 00, 00, 00)
table8_2[, c("Y", "weight")] <- t(apply(table8_2, 1, function(x) {
    if (x[7] == 00) {c(x[1], 1/x[4])}
    else if (x[7]== 10) {c(x[2], 1/x[5])}
    else if (x[7]== 01) {c(x[3], 1/x[6])}
    }))
E_Y01 <- weighted.mean(table8_2$Y[table8_2$exposure == 01],
                       table8_2$weight[table8_2$exposure == 01])
E_Y00 <- weighted.mean(table8_2$Y[table8_2$exposure == 00],
                       table8_2$weight[table8_2$exposure == 00])
E_Y01 - E_Y00
```

```
## [1] -3.230769
```

b. Estimate $E[Y_{10} - Y_{00}]$ for the random assignment that places the treatment at location A, restricting the sample to the set of villages that have a non-zero probability of expressing both of these potential outcomes.

In general:

$$\hat{E}[Y_{10} - Y_{00}] = \frac{\sum_{i=1}^{N} \frac{d_{10,i} Y_{10,i}}{p_{10,i}}}{\sum_{i=1}^{N} \frac{d_{10,i}}{p_{10,i}}} - \frac{\sum_{i=1}^{N} \frac{d_{00,i} Y_{00,i}}{p_{00,i}}}{\sum_{i=1}^{N} \frac{d_{00,i}}{p_{00,i}}}$$

So for this example, when location A is assigned to treatment (and hence location B is assigned to spillover, which would never be the case of location E):

$$\hat{E}[Y_{10} - Y_{00}] = \frac{\sum_{i=2}^{2} \frac{Y_{10,i}}{p_{10,i}}}{\sum_{i=2}^{2} \frac{1}{p_{10,i}}} - \frac{\sum_{i=3}^{4} \frac{Y_{00,i}}{p_{00,i}}}{\sum_{i=3}^{4} \frac{1}{p_{00,i}}} = \frac{\frac{2}{0.4}}{\frac{1}{0.4}} - \frac{\frac{0}{0.4} + \frac{6}{0.6}}{\frac{1}{0.4} + \frac{1}{0.6}} = \frac{5}{2.5} - \frac{0+10}{2.5+1.6667} = 2 - \frac{10}{4.1167} = 2 - 2.4 = \boxed{-0.4}$$

```r
table8_2_b <- table8_2[table8_2[, 4] != 0 & table8_2[, 5] != 0, ]

E_Y10 <- weighted.mean(table8_2_b$Y[table8_2_b$exposure == 10],
                       table8_2_b$weight[table8_2_b$exposure == 10])
E_Y00 <- weighted.mean(table8_2_b$Y[table8_2_b$exposure == 00],
                       table8_2_b$weight[table8_2_b$exposure == 00])
E_Y10 - E_Y00
```

```
## [1] -0.4
```

c. In order to make a more direct comparison between these two treatment effects, estimate $E[Y_{01} - Y_{00}]$, restricting the sample to the same set of villages as in part (b).

$$\hat{E}[Y_{01} - Y_{00}] = \frac{\sum_{i=1}^{1} \frac{Y_{01,i}}{p_{01,i}}}{\sum_{i=1}^{1} \frac{1}{p_{01,i}}} - \frac{\sum_{i=3}^{4} \frac{Y_{00,i}}{p_{00,i}}}{\sum_{i=3}^{4} \frac{1}{p_{00,i}}} = \frac{\frac{0}{0.2}}{\frac{1}{0.2}} - \frac{\frac{0}{0.4} + \frac{6}{0.6}}{\frac{1}{0.4} + \frac{1}{0.6}} = \frac{0}{5} - \frac{0+10}{2.5+1.6667} = 0 - \frac{10}{4.1167} = \boxed{-2.4}$$

```r
table8_2_c <- table8_2_b

E_Y01 <- weighted.mean(table8_2_c$Y[table8_2_c$exposure == 01],
                       table8_2_c$weight[table8_2_c$exposure == 01])
E_Y00 <- weighted.mean(table8_2_c$Y[table8_2_c$exposure == 00],
                       table8_2_c$weight[table8_2_c$exposure == 00])
E_Y01 - E_Y00
```

```
## [1] -2.4
```

8. FE exercise 8.9.

a. For the subset of 11 hotspot locations that lie outside the range of possible spillovers, calculate $E[Y_{01} - Y_{00}]$, the ATE of immediate police surveillance.

```r
# Read the dataset
police <- read.csv("GerberGreenBook_Chapter8_Table_8_4_8_5.csv")

# Discard hotspot locations that lie within the range of possible spillovers
police_outside <- police[police$hotwitin500 == 0, ]

# The weights are the inverse of the probabilities
police_outside$weight <- ifelse(police_outside$exposure == 01,
                                1/police_outside$prob01,
                                1/police_outside$prob00)

E_Y01 <- weighted.mean(police_outside$y[police_outside$exposure == 01],
                       police_outside$weight[police_outside$exposure == 01])
E_Y00 <- weighted.mean(police_outside$y[police_outside$exposure == 00],
                       police_outside$weight[police_outside$exposure == 00])
E_Y01 - E_Y00
```

```
## [1] 3.333333
```

$E[Y_{01} - Y_{00}] = 3.33,$ i.e., that's how much crime rate is increased.

b. For the remaining 19 hotspot locations that lie within the range of possible spillovers, calculate $E[Y_{01} - Y_{00}]$, $E[Y_{10} - Y_{00}]$, and $E[Y_{11} - Y_{00}]$.

```r
# Discard hotspot locations that lie outside the range of possible spillovers
police_within <- police[police$hotwitin500 != 0, ]

# The weights are the inverse of the probabilities
for (i in 1:dim(police_within)[1]) {
    if (police_within$exposure[i] == 00)
        {police_within$weight[i] <- 1/police_within$prob00[i]}
    else if (police_within$exposure[i] == 01)
        {police_within$weight[i] <- 1/police_within$prob01[i]}
    else if (police_within$exposure[i] == 10)
        {police_within$weight[i] <- 1/police_within$prob10[i]}
    else if (police_within$exposure[i] == 11)
        {police_within$weight[i] <- 1/police_within$prob11[i]}
    }

E_Y00 <- weighted.mean(police_within$y[police_within$exposure == 00],
                       police_within$weight[police_within$exposure == 00])
E_Y01 <- weighted.mean(police_within$y[police_within$exposure == 01],
                       police_within$weight[police_within$exposure == 01])
E_Y10 <- weighted.mean(police_within$y[police_within$exposure == 10],
                       police_within$weight[police_within$exposure == 10])
E_Y11 <- weighted.mean(police_within$y[police_within$exposure == 11],
                       police_within$weight[police_within$exposure == 11])
E_Y01 - E_Y00
```

```
## [1] -16.03258
```

```
E_Y10 - E_Y00
```

```
## [1] -0.03702306
```

```
E_Y11 - E_Y00
```

```
## [1] -9.605768
```

$E[Y_{01} - Y_{00}] = -16.03.$
$E[Y_{10} - Y_{00}] = -0.04.$
$E[Y_{11} - Y_{00}] = -9.61.$

> **c. Use the data at http://isps.research.yale.edu/FEDAI to estimate the average effect of spillover on nonexperimental units. Note that your estimator must make use of the probability that each unit lies within 500 meters of a treated expiermental unit; exclude from your analysis any units that have zero probability of experiencing spillovers.**

```
# Read the dataset
police2 <- read.csv("GerberGreenBook_Chapter8_Exercise_9c.csv.csv")

# Discard hotspot locations that have zero probability of experiencing spillovers
police2 <- police2[police2$prob10 != 0, ]

police2$weight <- ifelse(police2$exposure == 10, 1/police2$prob10,
                                1/police2$prob00)

E_Y10 <- weighted.mean(police2$y[police2$exposure == 10],
                        police2$weight[police2$exposure == 10])
E_Y00 <- weighted.mean(police2$y[police2$exposure == 00],
                        police2$weight[police2$exposure == 00])
E_Y10 - E_Y00
```

```
## [1] 4.602226
```

$E[Y_{10} - Y_{00}] = 4.6.$

9. FE exercise 8.10.

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

First of all, if assignment is not random, receiving treatment may be systematically related to potential outcomes. But we are told that the subject randomly varied whether she ran or walked 40 minutes each morning.

The other main assumptions to identify the causal effect are somehow equivalent to the non-interference assumption in a between-subjects design:

1. **No anticipation**: potential outcomes are unaffected by treatments that are administered in the future $(Y_{001} = Y_{000})$.
2. **No persistence**: potential outcomes in one period are unaffected by treatments administered in prior periods $(Y_{100}) = Y_{000})$.

With regards to the no persistence assumption, is one day a "washout period" long enough? Based on the particular variable of interest (Tetris score) it seems plausible, but we are not able to know that with certainty without comparing to similar studies.

The no-anticipation assumption is more likely to be broken in this case, especially because the subject is aware of the study and actually conducting it. Whatever the random process she uses to randomly vary the treatment is (be it a flip coin or any other mechanism), though every realization is independent from the others—so, for example, the chance of landing a head up is 0.5, no matter what the results were in prior days), the subject may think that—by regression to the mean—if she was not assigned to run many days in a row, she will probably be treated the next day, and vice versa. If she thinks there is a causal relation between running and her Tetris score, she may not try very hard to play well, thinking something like "*Tomorrow I'll probably do better.*"

Apart from those assumptions, the main concerns come from the fact that the subject is conducting the study. She may be subconsciously trying to confirm her prior theories, e.g., playing better when she has run and worse when she has not (not to mention that she might be tempted to cheat, recording different results or changing the treatment schedule).

b. Estimate the effect of running on Tetris score. Use randomization inference to test the sharp null hypothesis that running has no immediate or lagged effect on Tetris scores.

```
# Read the dataset
Hough <- read.dta("Hough_WorkingPaper_2010.dta.dta")
Hough$run_pre <- c(NA, Hough$run[1:dim(Hough)[1]-1])

model_run <- lm(tetris ~ run_pre + run, Hough)
summary(model_run)


##
## Call:
## lm(formula = tetris ~ run_pre + run, data = Hough)
```

```
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -19693.2  -7830.5   -46.2  5359.5  27539.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11793       4754   2.481  0.02211 *
## run_pre         1689       4948   0.341  0.73643
## run            15026       4986   3.014  0.00686 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11740 on 20 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3125, Adjusted R-squared:  0.2437
## F-statistic: 4.545 on 2 and 20 DF,  p-value: 0.0236
```

As shown above, the *p*-value of the *F*-statistic is 0.0236.

The *F*-statistic in a regression tests the overall significance of the regression model, i.e., the null hypothesis that **all** of the regression coefficients (which correspond to today's and yesterday's treatment assignment here) are equal to zero. So its *p*-value is the probability that the null hypothesis for the full model is true.

In our case, the *p*-value is below 0.05 (and hence statistically significant), so at least one of the variables (today's treatment assignment—Run—, based on the *p*-value of each coefficient) makes this model fit the data better than the simplest model, with no variables (which would be the mean of the variable Tetris).

> **b'. Do you think the randomization inference answers would differ? Why or why not? (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.)**

I've actually conducted randomization inference, though I'm not 100% sure I have done it the right way. (Anyway, the answer to this part is at the end of it, in **bold** letters.)

```r
randomize <- function() sample(Hough$run)
numiter <- 100e3

F <- summary(model_run)$fstatistic[1]

dist_F_SNH <- function(){
    run_ri <- randomize()
    run_pre_ri <- c(NA, run_ri[1:dim(Hough)[1]-1])
    # Yesterday's treatment assignment is not random but depends on today's
        # (or vice versa)
    summary(lm(Hough$tetris ~ run_ri + run_pre_ri))$fstatistic[1]
    }

F_ri <- replicate(numiter, dist_F_SNH())
mean(F_ri > F)
```

```
## [1] 0.01682
```

```
round(1 - pf(summary(model_run)$fstatistic[1],
             summary(model_run)$fstatistic[2],
             summary(model_run)$fstatistic[3]), 4)
```

```
##  value
## 0.0236
```

```
plot(density(F_ri), col = "red",
     main = "Distribution of F-statistics under SNH")
# It really looks like an F-distribution!!
abline(v = F, col = "blue")
```
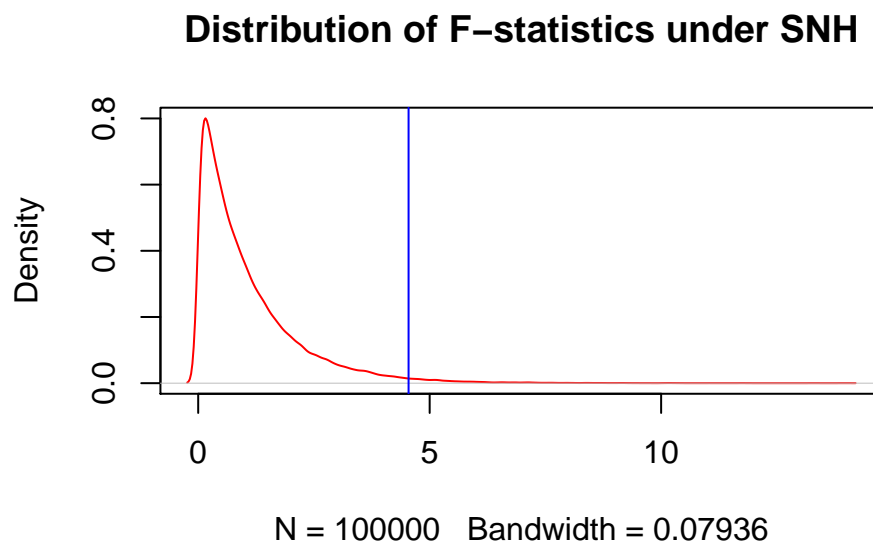

Figure 1: Distribution of the F-statistics under the SNH

Under the null hypothesis that neither today's treatment assignment nor yesterday's treatment assignment help predict the Tetris score, only 1.68% of the F-statistics are greater than the one we initially obtained (4.545). So the p-value we obtain using randomization inference is smaller than the one obtained using regression (2.36%).

Anyway, we don't really need to conduct randomization inference to know that: randomization inference is more accurate than regression (or a difference-in-means) when the assumptions needed for regression are not met.

At the end, the regression we've conducted is an ANOVA test: the null hypothesis of the F-statistic is that there are no significant differences in the Tetris scores whether the subject ran that day (or the previous day) or not.

```
summary.lm(aov(tetris ~ run_pre + run, Hough))
```

```
##
## Call:
## aov(formula = tetris ~ run_pre + run, data = Hough)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19693.2  -7830.5    -46.2   5359.5  27539.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11793       4754   2.481  0.02211 *
## run_pre         1689       4948   0.341  0.73643
## run            15026       4986   3.014  0.00686 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11740 on 20 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3125, Adjusted R-squared:  0.2437
## F-statistic: 4.545 on 2 and 20 DF,  p-value: 0.0236
```

So I'll just check the assumptions of ANOVA. In this example the assumptions are met only up to some extent, that's why the results of regression do not differ too much from those of randomization inference.

The sample distribution of Tetris scores within groups (of the variable Run, to make things simpler) are slightly right-skewed due to a few large scores:
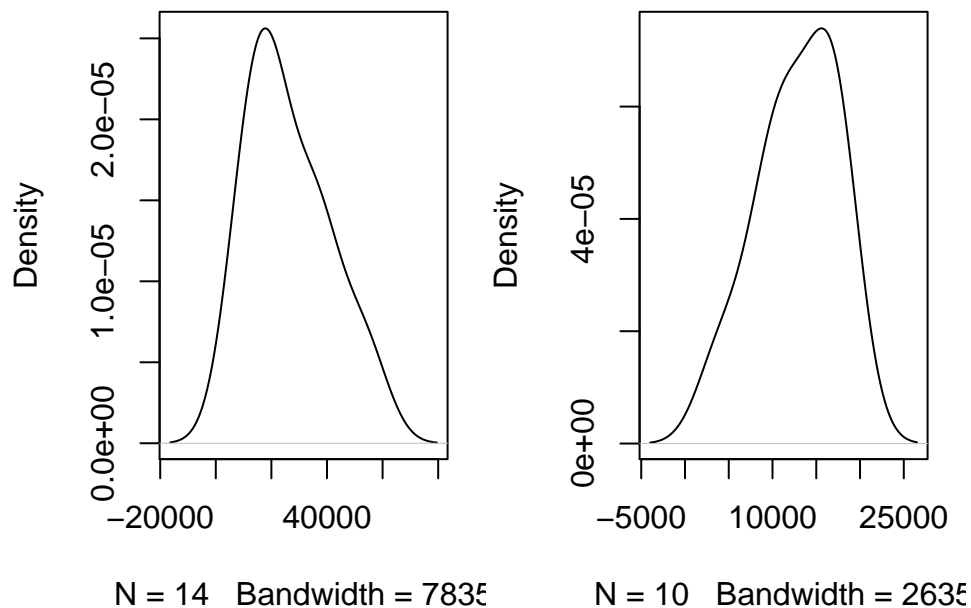

Figure 2: Sample distribution of Tetris scores in treatment and control days

But group sizes are almost equal, so the F-statistic will be quite robust to violations of normality.

The variance of the residuals at each level of Run is not constant (homoscedasticity assumption is broken):

```
leveneTest(model_run$residuals,
            as.factor(Hough$run[!is.na(Hough$tetris) & !is.na(Hough$run_pre)]))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  1   5.936 0.02382 *
##       21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
var(Hough$tetris[Hough$run==1])
```

```
## [1] 217798209
```

```
var(na.omit(Hough$tetris[Hough$run==0]))
```

```
## [1] 21594680
```

And here lies the reason why **the F-statistic from the regression is conservative (its p-value is larger)** than the one from randomization inference: the group with larger sample size ($Run = 1$, 14 observations) has a variance about 10 times the variance of the group with smaller sample size ($Run = 0$, 10 observations). Had the situation been the opposite (if the group with larger sample size had smaller variance), the F-statistic would have been liberal (its p-value would have been lower, making it more likely to detect an effect when it does not really exist).

c. One way to lend credibility to within-subjects results is to verify the no-anticipation assumption. Use the variable Run to predict the Tetris score *on the preceding day*. Presumably, the true effect is zero. Does randomization inference confirm this prediction?

```
Hough$tetris_pre <- c(NA, Hough$tetris[1:dim(Hough)[1]-1])
```

```
model_no_anticipation <- lm(tetris_pre ~ run, Hough)
summary(model_no_anticipation)
```

```
##
## Call:
## lm(formula = tetris_pre ~ run, data = Hough)
##
## 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
```

```
    # Equivalent to:
    # Hough$run_post <- c(Hough$run[2:dim(Hough)[1]], NA)
    # model_no_anticipation2 <- lm(tetris ~ run_post, Hough)
    # summary(model_no_anticipation2)
```

Yes, *regression* confirms that the no-anticipation assumption is met: the coefficient of Run, when we regress the Tetris score *on the preceding day* on it, is not statistically significant at all, so it does not predict thats score.

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?

```
model_energy <- lm(energy ~ run + run_pre, Hough)
summary(model_energy)
```

```
##
## Call:
## lm(formula = energy ~ run + run_pre, data = Hough)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4153 -0.7049  0.2131  0.5847  2.2131
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.6230     0.4313   6.081 6.07e-06 ***
## run           0.1639     0.4524   0.362    0.721
## run_pre       0.6284     0.4490   1.400    0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.065 on 20 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.09044,    Adjusted R-squared:  -0.000516
## F-statistic: 0.9943 on 2 and 20 DF,  p-value: 0.3875
```

```
model_GRE <- lm(gre ~ run + run_pre, Hough)
summary(model_GRE)
```

```
##
## Call:
## lm(formula = gre ~ run + run_pre, data = Hough)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7918 -0.5763  0.2082  0.3511  0.4237
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.86441    0.18949   4.562  0.00017 ***
## run         -0.21550    0.19648  -1.097  0.28516
## run_pre     -0.07264    0.19648  -0.370  0.71531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.4722 on 21 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.05557,    Adjusted R-squared:  -0.03438
## F-statistic: 0.6178 on 2 and 21 DF,  p-value: 0.5486
```

No, these hypotheses are not borne out by the data: the F-statistic is not statistically significant at all, in any of the two regressions, so neither Energy levels nor GRE scores respond to neither today's treatment assignment nor yesterday's.