

Field Experiments: Design, Analysis and Interpretation

Solutions for Chapter 8 Exercises

Alan S. Gerber and Donald P. Green*

Question 1

Important concepts:

- a) Interpret the expression $Y_i(\mathbf{d}) = Y_i(d)$ and explain how it conveys the non-interference assumption.

Answer:

The expression $Y_i(d)$ refers to the potential outcome that would be expressed based on the input d , which refers to the treatment that subject i receives. By contrast, $Y_i(\mathbf{d})$ refers to the potential outcome that subject i would express based on the assignments that all of the subjects receive. The equality means that the only input that matters is the treatment that subject i receives.

- b) Why are experiments that involve possible spatial spillover effects (such as the example described in section 8.4) said to involve “implicit” clustered assignment?

Answer:

Because certain units are so closely spaced that if a subject receives spillovers from one of the units, it must receive spillovers from all of the units. In that sense, spillovers are assigned as clusters.

- c) In what ways might a within-subjects design violate the non-interference assumption?

Answer:

In a within-subjects design, the unit of observation is the time period. Non-interference presupposes that each unit’s potential outcomes are solely a function of the treatments administered in that period. Possible violations include the following: subjects in one period are affected by the treatments that they may have received in a previous period; subjects in one period may be affected because they anticipate treatments that will be administered in a subsequent period.

- d) What are the attractive properties of a waitlist (or stepped-wedge) design?

Answer:

If the assumptions for unbiased inference are met, this within-subjects design may provide precise estimates of causal effects even when the number of subjects is limited. In terms of implementation, it may be easier to gain the cooperation of subjects or groups administering the treatment that might otherwise object to the use of a control group because everyone in the study eventually receives the treatment.

*Solutions prepared by Peter M. Aronow and revised by Alexander Coppock

Question 2

National surveys indicate that college roommates tend to have correlated weights. The more one roommate weighs at the end of the freshman year, the more the other freshman roommate weighs. 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:

A correlation describes the relationship between roommates' weights. A correlation might arise even if subjects have no effect on one another; for example, if subjects from similar regional or ethnic backgrounds tend to room together, one might observe a correlation. Random pairing of roommates means that, in expectation, there would be no correlation between their weights unless they either affected one another's weights or were both affected by similar environmental factors.

Question 3

Sometimes researchers are reluctant to randomly assign individual students in elementary classrooms because they are concerned that treatments administered to some students are likely to spill over to untreated students in the same classroom. In an attempt to get around possible violations of the non-interference assumption, they assign classrooms as clusters to treatment and control, and administer the treatment to all students in a classroom.

- a) State the non-interference assumption as it applies to the proposed clustered design.

Answer:

The non-interference assumption depends on the estimand. If the aim is to estimate the causal effect of the intervention on individual students, the non-interference assumption is the same as usual, namely, that each student's potential outcomes are affected only by the treatment administered to that subject. If one is concerned about transmission of treatments between students in the same classroom, that concern would still apply to a clustered design, since potential outcomes may be affected by the treatments that other subjects in the same classroom receive. On the other hand, if one is interested in classroom-level treatment effects (i.e., the difference between a 100% treated classroom and a 0% treated classroom), this design sidesteps concerns about within-classroom interference because they are built into the definition of the estimand. In the latter case, the relevant non-interference assumption holds that classroom outcomes are unaffected by the treatment status of other classrooms (e.g., other classrooms in the same school or grade).

- b) What causal estimand does the clustered design identify? Does this causal estimand include or exclude spillovers within classrooms?

Answer:

The causal estimand identified by the clustered design is the average effect of a classroom being 100% treated versus 0% treated. This includes within-classroom spillovers at the individual level, but assumes that across-classroom spillovers do not occur.

Question 4

Recall from Chapter 3 (exercise 9) the field experiment conducted by Camerer in which he selected pairs of similar horses running in the same race and randomly placed large wagers on one of them

to see if his bets affected the amount of money that other bettors placed on both horses.

- a) Define the potential outcomes in Camerer's study. What non-interference assumption is invoked?

Answer:

Two definitions for potential outcomes are possible.

The outcome variable in this study was defined as the change in total bets that occur between the pre-intervention period and the post-intervention period. Presumably, the original study assumed that potential outcomes $Y_i(D_i)$ reflect only the treatment assignment of horse i , not the treatment or non-treatment of the other horse in the experimental pair.

Nevertheless, one could define four potential outcomes for each horse: $Y_i(D_i = 1, D_j = 1)$, $Y_i(D_i = 1, D_j = 0)$, $Y_i(D_i = 0, D_j = 1)$, and $Y_i(D_i = 0, D_j = 0)$, which correspond to the change in bets for horse i depending on whether i , j , both, or neither are treated with experimental bets. Because of the matched pair randomization, horses only ever revealed the second or third of these potential outcomes.

- b) What is the causal parameter that this study identifies?

Answer:

Depending on which non-interference assumption is made, the same difference-in-means estimate refers either to the average difference between $Y_i(D_i = 1)$ and $Y_i(D_i = 0)$ or the average difference between $Y_i(D_i = 1, D_j = 0)$ and $Y_i(D_i = 0, D_j = 1)$. Stated plainly, the first is the average effect of treatment on the change in bets. The second is the average effect of being the treated horse in a pair in which one horse receives the experimental bets.

The treatment effect for a given horse is assumed to be $Y_i(1) - Y_i(0)$, where $Y_i(0)$ is the untreated potential outcome that would be revealed regardless of whether neither horse were treated or only the other paired horse were treated. In other words, the bets that flow to horse i have nothing to do with whether the other paired horse receives the experimental wager.

Question 5

In their study of spillover effects, Sinclair, McConnell, and Green sent mailings to randomly selected households encouraging them to vote in an upcoming special election.¹ The mailings used a form of "social pressure," disclosing whether the targeted individual had voted in previous elections. Because this type of mail had proven to increase turnout by approximately 4-5 percentage points in previous experiments, Sinclair, McConnell, and Green used it to study whether treatment effects are transmitted across households. Employing a multi-level design, they randomly assigned all, half, or none of the members of each nine-digit zip code to receive mail. For purposes of this example, we focus only on households with one registered voter. The outcome variable is voter turnout as measured by the registrar of voters. The results are as follows. Among registered voters in untreated zip codes, 1,021 of 6,217 cast ballots. Among untreated voters in zip codes where half of the households received mail, 526 of 3,316 registered voters cast ballots. Among treated voters in zip codes where half of the households received mail, 620 of 2,949 voted. Finally, among treated voters in zip codes where every household received mail, turnout was 1,316 of 6,377.

- a) Using potential outcomes, define the treatment effect of receiving mail addressed to subject i .

Answer:

The definition of personally receiving mail could be defined in three ways (given our focus on

¹Sinclair, McConnell, and Green 2010.

one-voter households). It could be (a) the effect of mail on those whose zip code neighbors receive no mail, (b) the effect of mail on those for whom half of the neighboring households in the zip code receive mail, or (c) the effect of mail on those whose zip code neighbors all receive mail. Given the design of this study, only (b) can be estimated empirically because no one receives mail in an untreated zip code, and everyone receives mail in a 100% treated zip code.

- b) Define the “spillover” treatment effect of being in a zip code where varying fractions of households are treated.

Answer:

Holding constant one’s own treatment status, one may define three potential outcomes depending on whether none, half, or all of the neighboring households are treated. When defining the ATE of spillover, one may compare half to none, full to half, or full to none.

- c) Propose an estimator for estimating the firsthand and secondhand treatment effects. Show that the estimator is unbiased, explaining the assumptions required to reach this conclusion.

Answer:

The firsthand effects can be estimated only for those in half-treated zip codes by comparing average outcomes among treated and untreated subjects. One can assess the spillover effect among subjects who receive no mail themselves but reside in either half-treated or untreated zip codes. Similarly, one can assess the spillover effect among subjects who received mail themselves and reside in either 100% or 50% treated zip codes. The three assumptions are random assignment (satisfied by design because direct treatments and rates of treatment among neighbors are randomly assigned), non-interference (satisfied if we believe that potential outcomes are solely a function of firsthand treatment and treatment of others in the same zip code; treatment of those outside the zip code is ignored), and excludability (satisfied if we believe that potential outcomes are affected only by firsthand and second hand receipt of mail and not by other factors that might be correlated with treatment assignment).

- d) Based on these data, what do you infer about the magnitude of the mailing’s direct and indirect effects?

```
In [1]: clear
        qui set obs 18859
        qui egen z_ind = fill(0,0)
        qui replace z_ind = 1 in 9534/18859

        // ssc install egenmore (uncomment to install the package)
        qui egen z_zip = repeat(), values("none")
        qui replace z_zip = "half" in 6218/12482
        qui replace z_zip = "all" in 12483/18859

        qui egen Y = fill(1,1)
        qui replace Y=0 in 1022/6217
        qui replace Y = 0 in 6744/9533
        qui replace Y=0 in 10154/12482
        qui replace Y=0 in 13799/18859

In [2]: qui mean Y if z_ind==1 & z_zip=="half"
```

```

    scalar ate_treat_half = _b[Y]
    qui mean Y if z_ind==0 & z_zip=="half"
    scalar ate_untreat_half = _b[Y]
    qui mean Y if z_ind==0 & z_zip=="none"
    scalar ate_untreat_none = _b[Y]
    qui mean Y if z_ind==1 & z_zip=="all"
    scalar ate_treat_all = _b[Y]

In [3]: // ate.fristhand.half
        disp ate_treat_half - ate_untreat_half

.05161591

In [4]: // ate.secondhandf.untreated
        disp ate_untreat_half - ate_untreat_none

-.00560227

In [5]: // ate.secondhandf.treated
        disp ate_treat_all - ate_treat_half

-.00387413

```

Here, the firsthand effects can be estimated only for those in half-treated zip codes: $620/2949 - 526/3316 = 0.052$, or 5.2 percentage points. One can assess spillover effect by way of two different comparisons. In order to assess the effects of spillover among subjects who receive no mail themselves, compare voting rates for those living in 50% treated zip code to those living in the 0% treated zip code: $526/3316 - 1021/6217 = -0.006$, or negative 0.6 percentage points. In order to assess the effects of spillover among subjects who received mail themselves, compare voting rates for those living in 100% treated zip codes to those living in 50% treated zip codes: $1316/6377 - 620/2949 = -0.004$, or negative 0.4 percentage points. Although the estimated firsthand effect is strongly positive, both of the estimated spillover effects are close to zero.

Question 6

Using the potential outcomes from the clinic example in Table 8.2, calculate the following estimates.

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

Answer:

The treated and untreated averages are $\frac{0}{0.2} = 0$ and $\frac{\frac{0}{0.4} + \frac{6}{0.6} + \frac{6}{0.8}}{\frac{1}{0.4} + \frac{1}{0.6} + \frac{1}{0.8}} = 3.23$, respectively. The estimated ATE is -3.23 .

- 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.

Answer:

The treated and untreated averages are $\frac{\frac{2}{0.4}}{\frac{1}{0.4}} = 2$ and $\frac{\frac{0}{0.4} + \frac{6}{0.6}}{\frac{0.4}{0.4} + \frac{0.6}{0.6}} = 2.4$, respectively. The estimated ATE is -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).

Answer:

The treated and untreated averages are $\frac{\frac{0}{1}}{\frac{1}{0.2}} = 0$ and $\frac{\frac{0}{0.4} + \frac{6}{0.6}}{\frac{0.4}{0.4} + \frac{0.6}{0.6}} = 2.4$, respectively. The estimated ATE is -2.4 .

Question 7

Lab experiments sometimes pair subjects together and have them play against one another in games where each subject is rewarded financially according to the game's outcome. One such game involves making monetary contributions to a public good (e.g., preserving the environment); the game can be arranged such that each player gains financially if both of them make a contribution, but each player is better off still if they contribute nothing while their partner in the game makes a contribution. The treatment is whether the pair of players is allowed to communicate prior to deciding whether to contribute. Suppose that a lab experimenter recruits four subjects and assigns them randomly as pairs to play this game. The outcome is whether each player makes a contribution: Y_i is 1 if the player contributes and 0 otherwise. Each player has three potential outcomes: Y_{0i} is the outcome if players are prevented from communicating, Y_{1i} is the outcome if a player communicates with another player who is "persuasive," and Y_{2i} is the outcome if a player communicates with another player who is "unpersuasive." The table below shows the schedule of potential outcomes for four players, two of whom are persuasive and two of whom are unpersuasive.

Table 1: Question 7 Table

Subject	Type	Y_{0i}	Y_{1i}	Y_{2i}
1	Persuasive	0	1	0
2	Persuasive	1	1	0
3	Unpersuasive	0	0	0
4	Unpersuasive	1	1	1

- a) Calculate the average treatment effect of $Y_{1i} - Y_{0i}$. Calculate the average treatment effect of $Y_{2i} - Y_{0i}$.

Answer:

The ATE of talking with a persuasive person is $(3/4) - (1/2) = (1/4)$. The ATE of talking with an unpersuasive person is $(1/4) - (1/2) = -(1/4)$.

- b) How many random pairings are possible with four subjects?

Answer:

There are $4!/(2!(4-2)!) = 6$ pairings.

- c) Suppose that the experimenter ignores the distinction between Y_{1i} and Y_{2i} and considers only two treatment conditions: the control condition prevents communication between pairs of players, and the treatment condition allows communication. Call the observed outcomes in the

communication condition Y_{1i}^* . Across all possible random pairings of subjects, what is the average difference-in-means estimate when the average Y_{1i}^* is compared to the average Y_{0i} ? Does this number correspond to either of the two estimands defined in part (a)? Does it correspond to the average of these two estimands?

Answer:

The average difference-in-means estimate is $\frac{0-0.5-1+0+0.5+0.5}{6} = -\frac{1}{12}$. This does not correspond to any of the estimands defined in part a), nor does it correspond to the average of these estimands.

- d) What is the probability that a persuasive subject is treated by communicating with an unpersuasive subject? What is the probability that an unpersuasive subject is treated by communicating with an unpersuasive subject?

Answer:

Subject 1 has a 1/6 chance of being assigned to communicate with a persuasive subject (subject 2) and has a 1/3 chance of being assigned to communicate with an unpersuasive subject (subjects 3 or 4). The same probabilities apply to Subject 2. Subject 3 has a 1/6 chance of communicating with an unpersuasive subject (subject 4). The same probabilities apply to Subject 4.

- e) Briefly summarize why a violation of the non-interference assumption leads to biased difference-in-means estimates in this example.

Answer:

One's potential outcomes change depending on how the randomization happened to come out. Bias occurs because the probability of encountering a persuasive or unpersuasive partner is related to potential outcomes.

- f) Would bias be eliminated if the experimenter replicated this study (with four subjects) each day and averaged the results over a series of 100 daily studies?

Answer:

It depends. Replicating small experiments with the same combination of persuasive and unpersuasive subjects simply reproduces the bias described above, because each experiment is subject to the same bias. On the other hand, if one imagines replicating this study with a random draw of the four subject types (see part G below), no bias results because selecting one subject for treatment does not prevent a subject of the same type from being assigned to control.

- g) Would bias be eliminated if the experimenter assembled 400 subjects at the same time (imagine 100 subjects for each of the four potential outcomes profiles in the table) and assigned them to pairs? Hint: Answer the question based on the intuition suggested by part (d).

Answer:

Bias becomes negligible as the size of a given experiment increases, because in a large experiment the probability of encountering a persuasive partner is nearly the same for both persuasive and unpersuasive subjects.

Question 8

Concerns about interference between units sometimes arise in survey experiments. For example, surveys sometimes administer a series of “vignettes” involving people with different attributes. A respondent might be told about a low-income person who is randomly described as white or black; after hearing the description, the respondent is asked to rate whether this person deserves public assistance. The respondent is then presented with a vignette about a second person, again randomly

described as white or black, and asked about whether this person deserves public assistance. This design creates four experimental groups: (a) two vignettes about blacks, (b) two vignettes about whites, (c) a black vignette followed by a white vignette, and (d) a white vignette followed by a black vignette. Each respondent provides two ratings.

- a) Propose a model of potential outcomes that reflects the ways that subjects might respond to the treatments and the sequences in which they are presented.

Answer:

One could model this scenario as a two-period within-subjects experiment. Call the black vignette $D_i = 1$ and the white vignette $D_i = 0$. Assume that respondents are affected only by the treatments they have received in the past or present; future treatments are irrelevant. For each respondent, the relevant potential outcomes are denoted $Y_{t-1,t}$, where the $t - 1$ subscript refers to which (if any) treatment is administered in the preceding time period, t refers to the current time period. This notation allows us to use Y_{01} in period i , for example, to refer to that period's potential outcome if a respondent were given the white vignette in the preceding period and the black vignette in the current period. If no question is asked in the preceding period (because the outcome is the response to the first question), denote the outcome as Y_{*0} or Y_{*1} .

- b) Using your model of potential outcomes, define the ATE or ATEs that a researcher might seek to estimate.

Answer:

One estimand is the average difference between Y_{*1} and Y_{*0} , which is the effect of the vignette on responses to the first question. Another is the average difference between Y_{01} and Y_{00} , which is the effect of race on responses to the second question for those who are asked about whites in the first vignette. Similarly, one might consider the average difference between Y_{11} and Y_{10} , which is the effect of race on responses to the second question for those who are asked about blacks in the first vignette.

- c) Suggest an identification strategy for estimating this (these) causal estimand(s) using the data from this experiment.

Answer:

Each of the estimands mentioned in part (b) can be estimated by using the between-subjects part of the design. To estimate $E[Y_{*1} - Y_{*0}]$, compare average responses to the first question among those randomly assigned black or white vignettes. To estimate $E[Y_{01} - Y_{00}]$, compare average responses to the second question among those randomly assigned black or white vignettes who earlier received a white vignette. To estimate $E[Y_{11} - Y_{10}]$, compare average responses to the second question among those randomly assigned black or white vignettes who earlier received a black vignette.

- d) Suppose a researcher analyzing this experiment estimates the average “race effect” by comparing the average evaluation of the white recipient to the average evaluation of the black recipient. Is this a sound approach?

Answer:

This approach amounts to pooling the respondents' answers to both questions; if there are N respondents, this analysis analyzes $2N$ observations. If the three race effects defined above differ, this approach will estimate a weighted average of the three estimands, which may be uninterpretable as a causal effect.

Question 9

Use data from the hotspots experiment in Table 8.4 (these data are also available at <http://isps.research.yale.edu/FEDAI>) and the probabilities that each unit is exposed to immediate or spillover treatments (Table 8.5) to answer the following questions:

- 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.

```
In [1]: qui import delimited "./data/chapter08/GerberGreenBook_Chapter8_Table_8_4_8_5.csv",

In [2]: qui mean y01 if prox500==0
         scalar mean_y01 = _b[y01]
         qui mean y00 if prox500==0
         scalar mean_y00 = _b[y00]
         scalar true_ate = mean_y01-mean_y00

In [3]: disp true_ate

-5

In [4]: qui mean y if prox500==0 & assignment ==1
         scalar mean_y = _b[y]
         qui mean y00 if prox500==0 & assignment==0
         scalar mean_y00_0 = _b[y00]
         scalar ate_hat = mean_y - mean_y00_0

In [5]: disp ate_hat

3.3333333
```

The true ATE for the observations that lie outside the range of possible spillovers is -5. The estimated ATE using the observed random assignment is 3.33.

- 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 $E[Y_{11} - Y_{00}]$.

```
In [6]: qui mean y01 if prox500==1
         scalar mean_y01 = _b[y01]
         qui mean y00 if prox500==1
         scalar mean_y00 = _b[y00]
         qui mean y10 if prox500==1
         scalar mean_y10 = _b[y10]
         qui mean y11 if prox500==1
         scalar mean_y11 = _b[y11]

In [7]: scalar true_ate_01 = mean_y01 - mean_y00
         scalar true_ate_10 = mean_y10 - mean_y00
         scalar true_ate_11 = mean_y11 - mean_y00
```

```

In [8]: disp true_ate_01

-5

In [9]: disp true_ate_10

5

In [10]: disp true_ate_11

-7

In [11]: qui gen q =.
          qui replace q=prob10 if exposure==10
          qui replace q=prob11 if exposure==11
          qui replace q=prob01 if exposure==01
          qui replace q=prob00 if exposure==00
          qui gen weights = 1/q

In [12]: // fit.01
          qui regress y i.exposure if ///
          prox500>0 & (exposure==0 | exposure==1) [aweight=weights]
          estimates store m1, title(Model 1)

In [13]: // fit.10
          qui regress y i.exposure if ///
          prox500>0 & (exposure==0 | exposure==10) [pweight=weights]
          estimates store m2, title(Model 2)

In [14]: // fit.11
          qui regress y i.exposure if ///
          prox500>0 & (exposure==0 | exposure==11) [pweight=weights]
          estimates store m3, title(Model 3)

In [15]: estout m1 m2 m3, ///
          cells(b(star fmt(3)) se(par fmt(3))) ///
          legend label varlabels(_cons Constant) ///
          stats(N r2)

```

Among observations that lie outside the range of possible spillovers, the ATE of direct treatment is -5, the ate of indirect treatment is 5, and the ATE of direct and indirect treatment together is -7.

By comparing weighted averages, with weights equal to the inverse of the probability that an observation is assigned to its observed treatment condition, we obtain estimates for the three ATEs: -16.0, -0.04, -9.6, respectively.

- 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

Table 2: Question 9c: Treatment Effect Estimates

	Crime Rate		
	(1)	(2)	(3)
exposure01	-16.033 (8.065)		
exposure10		-0.037 (9.074)	
exposure11			-9.606 (7.725)
Constant	62.606 (5.222)	62.606 (4.976)	62.606 (4.918)
N	12	14	11
R ²	0.283	0.00000	0.147

unit lies within 500 meters of a treated experimental unit; exclude from your analysis any units that have zero probability of experiencing spillovers.

```
In [16]: clear
         qui import delimited "./data/chapter08/GerberGreenBook_Chapter8_Exercise_9c.csv", c
```

```
In [17]: qui gen q=.
         qui replace q=prob10 if exposure==10
         qui replace q=prob00 if exposure==0
         qui gen weights = 1/q
```

```
In [18]: regress y i.exposure if (prob10>0) & (prob10<1) [pweight=weights]
```

(sum of wgt is 1.5089e+02)

```
Linear regression               Number of obs   =           71
                               F(1, 69)        =          56.95
                               Prob > F         =          0.0000
                               R-squared         =          0.5496
                               Root MSE      =          2.0976
```

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
10.exposure		4.602226	.6098582	7.55	0.000	3.385592	5.81886
_cons		4.285784	.5233643	8.19	0.000	3.241701	5.329867

The estimate of the spillover effects of treatment on the non-experimental units is 4.6.

Question 10

A doctoral student conducted an experiment in which she randomly varied whether she ran or walked 40 minutes each morning.² In the middle of each afternoon over a period of 26 days, she measured the following outcome variables: (1) her weight (minus a constant, for privacy's sake), (2) her score in a game of Tetris, (3) her mood on a 0-5 scale, with 5 being the most pleasant, (4) her energy level on a 0-5 scale, with 5 being the most energetic, and (5) whether she answered correctly a randomly selected problem from the math section of the GRE. Outcomes are missing for days 13 and 17. The data are listed below.

- 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 instance? What special concerns arise due to the fact that the subject was conducting the study, undergoing the treatments, and measuring her own outcomes?

Answer:

Suppose the effect were defined as $Y_i(1)$ and $Y_i(0)$, where the subscript refers to day. This formulation assumes potential outcomes respond only to the treatments administered that day, with no carryover from days past and no anticipation of treatments to come. The no-anticipation assumption seems reasonable; more questionable is the assumption that Tetris scores respond only to today's treatment, not yesterday's. The potential outcomes above presuppose that the cognitive or physiologic effects of running disappear after a night's sleep. In order to relax this assumption, one could expand the schedule of potential outcomes to include pairs (or longer sequences) of treatments on successive days. There is a risk of an excludability violation when measuring one's own outcomes; what if the subject tries harder when playing tetris in the wake of a running treatment?

- 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.

```
In [1]: set more off
        clear
        macro drop _all
        qui import delim ./data/chapter08/Hough_WorkingPaper_2010, clear
        set seed 1234567

In [2]: rename tetris Y
        rename run Z

        // set time variable
        qui tsset day
        qui gen Zlag = 1.Z
        qui gen Ylag = 1.Y
```

²Hough 2010.

```

In [3]: /*-----fit1-----*/
        qui reg Y Z
        global fit1 = _b[Z]

In [4]: /*-----fit2-----*/
        qui reg Y Z Zlag
        global fit2=e(F)

In [5]: /*-----fit3-----*/
        qui reg Ylag Z
        global fit3=_b[Z]

In [6]: /*-----fit4-----*/
        qui reg energy Z
        global fit4=_b[Z]

In [7]: /*-----fit5-----*/
        qui reg gre Z
        global fit5=_b[Z]

In [8]: cap program drop coeff_Z
        program define coeff_Z, rclass
            args Y
            tempvar rannum Zri
            gen `rannum' = uniform()
            egen `Zri' = cut(`rannum'), group(2)
            qui reg `Y' `Zri'
            return scalar coef_Z = _b[`Zri']
        end

In [9]: cap program drop F_stat
        program define F_stat, rclass
            args Y day
            tsset `day'
            tempvar rannum Zri
            gen `rannum' = uniform()
            egen `Zri' = cut(`rannum'), group(2)
            qui reg `Y' `Zri' l.`Zri'
            return scalar fstat = e(F)
        end

In [10]: /*----- dist1 -----*/
        preserve
        simulate coef_Z=r(coef_Z), reps(10000) nodots: coeff_Z Y
        qui count if coef_Z >= $fit1

        // one-tailed p-value: does running increase Tetris scores
        // mean(dist1 >= fit1)

```

```

disp r(N)/_N
restore

.004

In [11]: /*----- dist2 -----*/
preserve
simulate fstat=r(fstat), reps(10000) nodots: F_stat Y day
qui count if fstat >= $fit2

// one-tailed p-value: does running increase Tetris scores
// mean(dist2 >= fit2) */
disp r(N)/_N
restore

.0146

In [12]: /*----- dist3 -----*/
preserve
simulate coef_Z=r(coef_Z), reps(10000) nodots: coeff_Z Ylag
qui count if abs(coef_Z) >= abs($fit3)

// two-tailed p-value: placebo fit
// mean(dist3 >= fit3)
disp r(N)/_N
restore

.8879

In [13]: /*----- dist4 -----*/
preserve
simulate coef_Z=r(coef_Z), reps(10000) nodots: coeff_Z energy
qui count if coef_Z >= $fit4

// one-tailed p-value: does running improve energy
// mean(dist4 >= fit4)
disp r(N)/_N
restore

.5033

In [14]: /*----- dist5 -----*/
preserve
simulate coef_Z=r(coef_Z), reps(10000) nodots: coeff_Z gre
qui count if coef_Z >= $fit5

// one-tailed p-value: does running improve GRE
// mean(dist5 >= fit5)

```

```
disp r(N)/_N
restore
```

.8137

Focusing solely on the immediate effect of running that day's tetris scores, we see that on non-running days the average tetris score is 12,806, as compared to 26,419 on running days, for a difference of 13,613. Randomization inference indicates that this observed difference has a one-tailed p -value of 0.004. Using the F-statistic to assess the joint significance of immediate and one-period lagged effects, we obtain a p -value of 0.0146, again allowing us to reject the null hypothesis of no effect.

- 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?

Answer:

See code block above for calculations. As expected, the means are similar (18,903 for control and 19,549 for treatment), and the two-tailed p -value is 0.8879.

- 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:

No, energy has a difference-in-means of just 0.07 and a p -value of 0.5033; and GRE's effect goes (insignificantly, $p=0.8137$) in the wrong direction, with the treatment diminishing the probability of a right answer by 0.175.

Question 11

Return to the stepped-wedge advertising example in section 8.6 and the schedule of assigned treatments in Table 8.8.

- a) Estimate $E[Y_{01} - Y_{00}]$ by restricting your attention to weeks 2 and 3. How does this estimate compare to the estimate of $E[Y_{11} - Y_{00}]$ presented in the text, which is also identified using observations from weeks 2 and 3?

```
In [1]: clear
        qui set obs 16
        //(uncomment to install the package)
        //ssc install egenmore
        qui egen week = repeat(), values("2")
        qui replace week = "3" in 9/1
        qui egen prob00 = fill(0.5,0.5)
        qui replace prob00=0.25 in 9/1
        qui egen prob01 = fill(0.25,0.25)
        qui egen prob11 = fill(0.25,0.25)
        qui replace prob11 = 0.5 in 9/1
```

```
In [5]: input int y str2 z
        9 "11"
```

```

5 "00"
2 "01"
3 "00"
3 "00"
8 "11"
3 "00"
1 "01"
4 "11"
7 "01"
10 "11"
10 "01"
3 "00"
10 "11"
4 "00"
3 "11" end

```

```

In [6]: gen prob01z = prob01 if z=="01"
        replace prob01z = prob01 if prob01z==.
        //difference in the way that R and STATA weighted mean functions
        gen y01z = y if z=="01"
        egen mean01 = wtmean(y01z), weight(1/prob01z)
        gen prob00z = prob00 if z=="00"
        replace prob00z = prob00 if prob01z==.
        gen y00z = y if z=="00"
        egen mean00 = wtmean(y00z), weight(1/prob00z)
        gen ate01_00 = mean01 - mean00
        disp ate01_00

```

1.5

The effect of immediate treatment appears to be 1.5, which is weaker than the effect mentioned in the text (4.13), possibly suggesting that the effect of immediate treatment is weaker than the effect of immediate and lagged treatment.

- b) Estimate $E[Y_{11} - Y_{00}]$ without imposing the assumption that treatment effects disappear after two weeks by restricting your attention to week 2.

```

In [7]: gen prob01z2 = prob01 if z=="01" & week=="2"
        gen y11z2 = y if z=="11" & week=="2"
        replace prob01z2 = .25 if prob01z2==.
        egen mean11 = wtmean(y11z2), weight(1/prob01z2)
        gen prob00z2 = prob01 if z=="00" & week=="2"
        replace prob00z2 = prob01 if prob00z2==.
        gen y00z2 = y if z=="00" & week=="2"
        drop mean00
        egen mean00 = wtmean(y00z2), weight(1/prob00z2)

```



```
gen ate11_00 = mean11 - mean00  
disp ate11_00
```

5

Without imposing this assumption and focusing only on week two, the estimated ATE of immediate and lagged treatment is 5.