Field Experiments: Design, Analysis and Interpretation
Solutions for Chapter 6 Exercises

Alan S. Gerber and Donald P. Green*

Question 1

The following three quantities are similar in appearance but refer to different things. Describe the differences.

- $E[Y_i(d(1))|D_i = 1]$
  Answer:
  This expression refers to the expected potential outcome of Y_i given the treatment received by the assigned treatment group $D_i(1)$ for the subgroup of subjects who actually receive the treatment $(D_i = 1)$.

- $E[Y_i(d(1))|d_i(1) = 1]$
  Answer:
  This expression refers to the expected potential outcome of Y_i given the treatment received by the assigned treatment group $D_i(1)$ for the subgroup of subjects who receive the treatment if assigned to it $(D_i(1) = 1)$. In the case of one-sided non-compliance, this subgroup is the Compliers. For two-sided non-compliance this is composed of Always-Takers and Compliers.

- $E[Y_i(d(1))|d_i(1) = d_i(0) = 1]$
  Answer:
  This expression refers to the expected potential outcome of Y_i given the treatment received by the assigned treatment group $D_i(1)$ for the subgroup of subjects known as Always-Takers, who always receive the treatment regardless of whether they are assigned to the treatment group $(D_i(1) = D_i(0) = 1)$.

Question 2

Question 3

Assuming that the excludability and non-interference assumptions hold, are the following statements true or false? Explain your reasoning.

---

a) Among Compliers, the ITT equals the ATE.
   Answer:
   True. For Compliers, treatment assigned equals treatment received, and so ITT = ATE.

b) Among Defiers, the ITT equals the ATE.
   Answer:
   False: For Compliers, treatment assigned is the opposite of treatment received, and so ITT = -ATE.

c) Among Always-Takers and Never-Takers, the ITT and ATE are zero.
   Answer:
   False. For Always-takers and Never-takers, the ITT is zero because they respond the same to both experimental assignments. The ATE among these subgroups may not be nonzero; the ATE is not revealed empirically.

Question 4

Question 5

Suppose that a sample contains 30% Always-Takers, 40% Never-Takers, 15% Compliers, and 15% Defiers. What is the ITT_D?
Answer:
Recall from equation (6.19): $ITT_D = \pi_C + \pi_{AT} - (\pi_D + \pi_{AT}) = \pi_C - \pi_D$, which in this case implies that the $ITT_D = 0$.

Question 6

Question 7

In experiments with one-sided noncompliance, the ATE among subjects who receive the treatment (sometimes called the average treatment-on-the-treated effect, or ATT) is the same as the CACE, because only Compliers receive the treatment. Explain why the ATT is not the same as the CACE in the context of two-sided noncompliance.
Answer:

Under two-sided noncompliance, both Compliers and Always-takers receive treatment when assigned to the treatment group, and Always-takers receive treatment when assigned to the control group. Therefore, as we move from one-sided to two-sided noncompliance, "the treated" no longer refers to Compliers, and the ATT no longer equals the CACE.

Question 8

Question 9

In their study of the effects of conscription on criminal activity in Argentina, Galiani, Rossi, and Schargrodsky use official records of draft lottery numbers, military service, and prosecutions for a cohort of men born between 1958 and 1962.[1] Draft eligibility is scored 1 if an individual had a draft lottery number that caused him to be drafted, and 0 otherwise. Draft lottery numbers were selected randomly by drawing balls from an urn. Military service is scored 1 if the individual actually served in the armed services, and 0 otherwise. Subsequent criminal activity is scored 1 if the individual had a judicial record of prosecution for a serious offense. For a sample of 5,000 observations, the authors report an $\widehat{ITT_D}$ of 0.6587 (SE = 0.0012), an $\widehat{ITT}$ of 0.0018 (SE = 0.0006), and a $\widehat{CACE}$ of 0.0026 (SE = 0.0008). The authors note that the $\widehat{CACE}$ implies a 3.75% increase in the probability of criminal prosecution with military service.

a) Interpret the $\widehat{ITT_D}$, $\widehat{ITT}$, $\widehat{CACE}$, and their standard errors.
   Answer:
   The $\widehat{ITT_D}$ refers to the difference in rates of military service between the treatment and control groups. Evidently, the treatment group was 65.87 percentage points more likely to serve in the military than the control group. The $\widehat{ITT}$ refers to the difference in prosecution rates between the assigned treatment and control groups (irrespective of whether a subject actually served). The estimate of 0.0018 implies that the treatment group was 0.18 percentage points more likely to be prosecuted than the assigned control group. The $\widehat{CACE}$ is the estimated ATE among Compliers, those who serve in the military if and only if they have a draft-eligible number. This estimate is 0.0026, which implies that Compliers become 0.26 percentage points more likely to be prosecuted as a result of serving in the military. The standard errors are a measure of statistical uncertainty, and a rule of thumb is that a 95% confidence interval may be formed by adding and subtracting +/- 2SEs. In this case, the 95% interval for $\widehat{ITT_D}$ is 65.87 +/- 0.0024; for $\widehat{ITT}$ is 0.0018 +/- 0.0012; for $\widehat{CACE}$, it is 0.0026 +/- 0.0016. The margin of uncertainty for the $\widehat{ITT}$ and $\widehat{CACE}$ is fairly wide, but the intervals are on the positive side of zero, suggesting that military service (if the exclusion restriction holds) has a criminogenic effect.

b) The authors note that 4.21% of subjects who were not draft eligible nevertheless served in the armed forces. Based on this information and the results shown above, calculate the proportion of Never-Takers, Always-Takers, and Compliers under the assumption of monotonicity.

---

[1]Galiani, Rossi, and Schargrodsky 2010.

Answer:

Monotonicity means that the proportion of Defiers is zero. The 4.21% who served without being drafted implies that Always-takers are 4.21% of the subject pool. From the $\widehat{ITT_D}$ of 0.6587 we infer the Compliers are 65.87% of the subject pool. That leaves 1 - 4.21% - 65.87% = 29.9% who are Never-takers.

c) Discuss the plausibility of the monotonicity, non-interference, and excludability assumptions in this application. If an assumption strikes you as implausible, indicate whether you think the $\widehat{CACE}$ is biased upward or downward.

Answer:

Let's analyze each assumption. Monotonicity implies no Defiers. Defiers are those who serve in the military if and only if they are not drafted. Given that one ordinarily think of people who join the military on their own volution as being willing to go if drafted, it is difficult to imagine that many people fit this description, so this assumption seems plausible. Random assignment implies that treatment assignment is independent of the potential outcomes. Although some lotteries are implemented incompetently or corruptly, we are given no reason to suspect that here. Non-interference means that potential outcomes reflect only the treatment or control status of the subject in question and do not depend on the status of other observations. In this case the potential outcome is whether a subject will be prosecuted. It seems possible that one's criminal career could be shaped by whether one's friends are or aren't drafted, but it is not clear how this violation of non-interference would bias the results, since if my friends are drafted it might make me more likely to engage in criminal conduct regardless of whether I am assigned to treatment or control. Excludability means that potential outcomes respond solely to receipt of the treatment (military service) and not the random assignment of the treatment or any indirect byproduct of random assignment (e.g., draft dodging). If citizens that are drafted are more easily monitored (e.g., their finger prints are recorded) then there might be an upward bias in the measurement of the crime committed by those assigned to treatment simply because it is easier to solve a crime committed by them.

Question 10

Question 11

A large-scale experiment conducted between 2002 and 2005 assessed the effects of Head Start, a preschool enrichment program designed to improve school readiness.[2] The assigned treatment encouraged a nationally representative sample of eligible (low-income) parents to enroll their four-year-olds in Head Start. Of the 1,253 children assigned to the Head Start treatment, 79.8% actually enrolled in Head Start; 855 of the children assigned to the control group (13.9%) nevertheless enrolled in Head Start. One of the outcomes of interest is pre-academic skills, as manifest at the end of the yearlong intervention. The principal investigators report that scores averaged 365.0

---

[2]Puma et al. 2010. We focus here on one part of the study, the sample of four-year-old subjects.

among students assigned to the treatment group and 360.5 among students assigned to the control group, with a two-tailed p-value of .041. Two years later, students completed first grade. Their first grade scores on a test of academic skills averaged 447.7 in the treatment group and 449.0 in the control group, with a two-tailed p-value of 0.380.

a) Estimate the CACE for this experiment, using pre-academic skills scores as the outcome.
   Answer:
   The estimated CACE is: $\widehat{CACE} = \frac{365-360.5}{0.798-0.139} = 6.82$

b) Estimate the CACE for this experiment, using academic skills in first grade as an outcome.
   Answer:
   The estimated CACE is: $\widehat{CACE} = \frac{447.7-449.0}{0.798-0.139} = -1.97$

c) Estimate the average downstream effect of pre-academic skills on first grade academic skills. Hint: Divide the estimated ITT (from a regression of first grade academic skills on assigned treatment) by the estimated ITT_D (from a regression of pre-academic skills on assigned treatment). Interpret your results. Are the assumptions required to identify this downstream effect plausible in this application? If not, would you expect the apparent downstream effect to be overestimated or underestimated?
   Answer:
   The estimated downstream CACE is: $\widehat{CACE} = \frac{447.7-449.0}{365-360.5} = -0.29$
   The results suggest, surprisingly, that an improvement in pre-academic skills among Compliers (those whose pre-academic skills change if they are exposed to the treatment) led to a deterioration of academic skills in first grade. For every one-point gain in pre-academic skills, there was a 0.29 drop in first grade skills. Ordinarily, one would expect a positive relationship (building early skills help build skills later on). One possible explanation for this anomalous result is sampling variability. Another is a violation of the exclusion restriction. Suppose, for the sake of argument, that Head Start teachers were coaching students to help them perform better on tests of pre-academic skills. (One could define this sort of teaching-to-the-test as the effect of Head Start, in which case there would be no excludability violation.) Suppose that coaching boosts pre-academic skills scores but lowers first grade scores because the same tricks that are used on the pre-academic skills test lower grades on the first grade test. The excluded factor of coaching boosts the denominator and lowers the numerator, and so the net bias is difficult to predict.