

Field Experiments: Design, Analysis and Interpretation

Solutions for Chapter 9 Exercises

Alan S. Gerber and Donald P. Green*

Question 1

Important concepts:

- a) Define CATE. Is a Complier average causal effect (CACE) an example of a CATE?

Answer:

CATE stands for conditional average treatment effect, or the ATE among a subgroup. Typically, the subgroup in question is defined by some observable covariate(s), such as the CATE for women over 40 years of age. One could, however, define a CATE for a latent group such as Compliers (those who take the treatment if and only if assigned to the treatment group). Therefore, a CACE is a CATE.

- b) What is an interaction effect?

Answer:

An interaction refers to systematic variation in treatment effects. A treatment-by-covariate interaction refers to variation in ATEs that is a function of covariates. A treatment-by-treatment interaction refers to variation in the average effect of one randomized intervention that occurs as a function of other assigned treatments.

- c) Describe the multiple comparisons problem and the Bonferroni correction.

Answer:

The multiple comparisons problem refers to the distortion in p -values that occurs when researchers conduct a series of hypothesis tests. When several hypothesis tests are conducted, the chances that at least one of them appears significant may be substantially greater than 0.05, the nominal size of each test. The Bonferroni correction reestablishes the proper size of each test when several hypothesis tests are conducted. If k tests are conducted at the 0.05 level, the Bonferroni-corrected target significance level is $0.05/k$.

Question 2

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

Question 3

One way to reduce variance in $Y_i(0)$ is to block on a prognostic covariate. When blocking is used, the joint distribution of $Y_i(0)$ and $Y_i(1)$ is simulated within blocks using the bounding procedure described in section 9.2. Using the schedule of potential outcomes below, show how the maximum and minimum values of the covariance of $Y_i(0)$ and $Y_i(1)$ compare to the maximum and minimum values of the covariance of $Y_i(0)$ and $Y_i(1)$ for the dataset as a whole (i.e., had blocking not been used).

Table 1: Question 3 Table

Block	Subject	Yi(0)	Yi(1)
A	A-1	0	2
A	A-2	1	5
A	A-3	1	3
A	A-4	2	1
B	B-1	2	3
B	B-2	3	3
B	B-3	4	9
B	B-4	4	7

```

block <- c(rep("A", 4), rep("B", 4))
Y0 <- c(0,1,1,2,2,3,4,4)
Y1 <- c(2,5,3,1,3,3,9,7)

# function for calculating population covariances
cov.pop <- function(x,y){sum((x-mean(x))*(y-mean(y)))/(length(x))}

# Ignoring blocks
Y1.lowtohigh <- sort(Y1)
Y1.hightolow <- sort(Y1, decreasing=TRUE)

cov.min <- cov.pop(Y0, Y1.hightolow)
cov.max <- cov.pop(Y0, Y1.lowtohigh)
cov.min

## [1] -3.141

cov.max

## [1] 3.234

# Including blocks
Y1.lowtohigh.block <- c(sort(Y1[block=="A"]), sort(Y1[block=="B"]))
Y1.hightolow.block <- c(sort(Y1[block=="A"],decreasing=TRUE),
                        sort(Y1[block=="B"],decreasing=TRUE))
cov.min.block <- cov.pop(Y0, Y1.hightolow.block)

```

```

cov.max.block <- cov.pop(Y0, Y1.lowtohigh.block)
cov.min.block

## [1] -0.01562

cov.max.block

## [1] 2.984

```

The lowest and highest covariances under simple random assignment are -3.14 and 3.23. In order to find the lowest and highest covariances under blocked assignment, sort the potential outcomes within blocks before calculating the covariances for all observations. Under blocked random assignment, the lowest covariance is -0.02, and the highest covariance is 2.98. Taking advantage of the blocks reduces the range of possible covariances.

Question 4

Question 5

The table below shows hypothetical potential outcomes for an experiment in which low-income subjects in a developing country are randomly assigned to receive (i) loans to aid their small businesses; (ii) business training to improve their accounting, hiring, and inventory-management skills; (iii) both; or (iv) neither. The outcome measure is business income during the subsequent year. The table also includes a pre-treatment covariate, an indicator scored 1 if the subject was judged to be proficient in these basic business skills.

Table 2: Question 5 Table

Subject	$Y_i(\text{loan})$	$Y_i(\text{training})$	$Y_i(\text{both})$	$Y_i(\text{Neither})$	Prior business skills
1	2	2	3	2	0
2	2	3	2	1	0
3	5	6	6	4	1
4	3	1	5	1	1
5	4	4	5	0	0
6	10	8	11	10	1
7	1	3	3	1	0
8	5	5	5	5	1
Average	4	4	5	3	0.5

a) What is the ATE of the loan if all subjects were also to receive training?

Answer:

The relevant comparison is the average potential outcomes under “both” to the average potential outcome under only “training.” The ATE is 5-4=1.

- b) What is the ATE of the loan if no subjects receive training?

Answer:

The relevant comparison is the average potential outcomes under “loan” to the average potential outcome under only “neither.” The ATE is 4-3=1.

- c) What is the ATE of the training if all subjects also receive a loan?

Answer:

The relevant comparison is the average potential outcomes under “both” to the average potential outcome under only “loan.” The ATE is 5-4=1.

- d) What is the ATE of the training if no subjects receive a loan?

Answer:

The relevant comparison is the average potential outcomes under “training” to the average potential outcome under only “neither.” The ATE is 4-3=1.

- e) Suppose subjects were randomly assigned to one of the four experimental treatments in equal proportions. Use the table above to fill in the expected values of the four regression coefficients for the model and interpret the results:

$$Y_i = \alpha_0 + \alpha_1 Loan_i + \alpha_2 Training_i + \alpha_3 (Loan_i * Training_i) + e_i \quad (1)$$

The four coefficients are $\alpha_0 = 3$, the average outcome under “neither”; $\alpha_1 = 1$, the ATE of loan when there is no training; $\alpha_2 = 1$, the ATE of training when there is no loan; and $\alpha_3 = 0$ the change in the effect of training that occurs when our focus switches from those who receive no loan to those who receive a loan. Note that this interaction term can also be interpreted as the change in the ATE of loans that we observe when we move from the untrained subgroup to the trained subgroup.

$$Y_i = 3 + 1 * Loan_i + 1 * Training_i + 0 * (Loan_i * Training_i) + e_i \quad (2)$$

- f) Suppose a researcher were to implement a block randomized experiment, such that two subjects with business skills are assigned to receive loans, and two subjects without business skills are assigned to receive loans, and the rest are assigned to control. No subjects are assigned to receive training. The researcher estimates the model

$$Y_i = \gamma_0 + \gamma_1 Loan_i + \gamma_2 Skills_i + \gamma_3 (Loan_i * Skills_i) + e_i \quad (3)$$

Over all 36 possible random assignments, the average estimated regression is as follows:

$$Y_i = 1.00 + 1.25 Loan_i + 4.00 Skills_i - 0.50 (Loan_i * Skills_i) \quad (4)$$

Interpret the results and contrast them with the results from part (e). (Hint: the block randomized design does not affect the interpretation. Focus on the distinction between treatment-by-treatment and treatment-by-covariate interactions.)

Answer:

The key thing to bear in mind when interpreting these results is that the interaction between

loans and skills is a treatment-by-covariate interaction because skills are not randomly assigned. The results seem to suggest that loans are more effective amongst those without skills ($CATE = 1.25$) than among those with skills ($CATE = 1.25 - 0.5 = 0.75$). These CATEs may describe the ATEs in these two skill groups, but the change in CATEs does not necessarily imply that a random increase in skill would diminish the effects of loans.

Question 6

Question 7

In their 2004 study of racial discrimination in employment markets, Bertrand and Mullainathan sent resumes with varying characteristics to firms advertising job openings. Some firms were sent resumes with putative African American names, while other firms received resumes with putatively Caucasian names. The researchers also varied other attributes of the resume, such as whether the resume was judged to be of high or low quality (based on labor market experience, career profile, gaps in employment, and skills listed).¹ The table below shows the rate at which applicants were called back by employers, by the city in which the experiment took place and by the randomly assigned attributes of their applications.

Table 3: Question 7 Table

		Boston				Chicago			
		Low-quality resume		High-quality resume		Low-quality resume		High-quality resume	
		Black	White	Black	White	Black	White	Black	White
% Received Call		7.01	10.15	8.5	13.12	5.52	7.16	5.28	8.94
(N)		(542)	(542)	(541)	(541)	(670)	(670)	(682)	(682)

- a) For each city, interpret the apparent treatment effects of race and resume quality on the probability of receiving a follow-up call.

Answer:

For Boston, the effect of (white) race is $10.15 - 7.01 = 3.14$ when resume quality is low and $13.12 - 8.50 = 4.62$ when resume quality is high. For Chicago, the effect of (white) race is $7.16 - 5.52 = 1.64$ when resume quality is low and $8.94 - 5.28 = 3.66$ when resume quality is high. Note that another, equally valid way to interpret the table is to assess the effect of resume quality for each race, but the substantive focus of this study is on race effects.

- b) Propose a regression model that assesses the effects of the treatments, interaction between them, and interactions between the treatments and the covariate, city.

Answer:

¹Bertrand and Mullainathan 2004, p. 994.

This model is similar to the interactive regression specifications described above, but it contains treatment-by-treatment interactions (race x resume) and treatment-by-covariate interactions (race x city, resume x city) and a higher order interaction (race x resume x city) that allows for the possibility that the race x resume interaction differs by city. Here, City is scored 1 if Chicago. Race = 1 if white. Resume =1 if high quality. Notice that the “saturated” regression model contains eight parameters, one for each cell of the table.

$$Y_i = \gamma_0 + \gamma_1 Race_i + \gamma_2 Resume_i + \gamma_3 City_i + \gamma_4(Race_i * Resume_i) + \gamma_5(Race_i * City_i) + \gamma_6(Resume_i * City_i) + \gamma_7(Race_i * Resume_i * City_i) + e_i$$

- c) Estimate the parameters in your regression model. Interpret the results (This can be done by hand based on the percentages given in the table.)

Answer:

Because there as many parameters as experimental groups, the estimated coefficients reproduce the percentages given in the table:

$$Y_i = 7.01 + 3.14Race_i + 1.49Resume_i - 1.49City_i + 1.48(Race_i * Resume_i) - 1.50(Race_i * City_i) - 1.73(Resume_i * City_i) + 0.54(Race_i * Resume_i * City_i) + e_i$$

Additional response (Boston = 1; Black = 1; Low quality = 1)

$$Y_i = 8.94 - 3.66Race_i - 1.78Resume_i + 4.18City_i + 2.02(Race_i * Resume_i) - 0.96(Race_i * City_i) - 1.19(Resume_i * City_i) - 0.54(Race_i * Resume_i * City_i) + e_i$$

Additional response (Boston = 1; Black = 1; High quality = 1)

$$Y_i = 7.16 - 1.64Race_i + 1.78Resume_i + 2.99City_i - 2.02(Race_i * Resume_i) - 1.50(Race_i * City_i) + 1.19(Resume_i * City_i) + 0.54(Race_i * Resume_i * City_i) + e_i$$

```
Y <- c(rep(1, 38), rep(0, 542-38), rep(1, 55), rep(0, 542-55),
      rep(1, 46), rep(0, 541-46), rep(1, 71), rep(0, 541-71),
      rep(1, 37), rep(0, 670-37), rep(1, 48), rep(0, 670-48),
      rep(1, 36), rep(0, 682-36), rep(1, 61), rep(0, 682-61))

boston <- c(rep(1, 542+542+541+541), rep(0, 670+670+682+682))
chicago <- 1-boston
lowquality <- c(rep(1, 542+542), rep(0, 541+541), rep(1, 670+670), rep(0, 682+682))
highquality <- 1-lowquality
black<- c(rep(1, 542), rep(0,542), rep(1, 541), rep(0,541),
         rep(1, 670), rep(0,670), rep(1, 682), rep(0,682))
white <- 1-black

# All the models are
# Y ~ race + quality + city + race*quality + race*city + quality*city + race*quality*city
# In principle, there are 8 possibilities... here are 4.
```

```

fit_1 <- lm(Y ~ white + highquality + chicago + white*highquality +
            white*chicago + highquality*chicago + white*highquality*chicago)
fit_2 <- lm(Y ~ black + highquality + chicago + black*highquality +
            black*chicago + highquality*chicago + black*highquality*chicago)
fit_3 <- lm(Y ~ white + highquality + boston + white*highquality +
            white*boston + highquality*boston + white*highquality*boston)
fit_4 <- lm(Y ~ black + lowquality + chicago + black*lowquality +
            black*chicago + lowquality*chicago + black*lowquality*chicago)
stargazer(fit_1, fit_2, fit_3, fit_4, style = "apsr", notes = c("Coefficients depend on study"))

```

Question 8

Question 9

An example of a two-factor design that encounters one-sided noncompliance may be found in Fieldhouse et al.'s study of voter mobilization in the United Kingdom.² In this study, the first factor is whether each voter was mailed a letter encouraging him or her to vote in the upcoming election. The second factor is whether each voter was called with an encouragement to vote. Noncompliance occurs in the case of phone calls, as some targeted voters cannot be reached when called. The experimental design consists of four groups: a control group, a mail-only group, a phone-only group, and a group targeted for both mail and phone. The following table shows the results by assigned experimental group.

- a) Show that, under certain assumptions, this experimental design allows one to identify the following parameters: (i) the ATE of mail, (ii) the Complier average causal effect (CACE) of phone calls, (iii) the CATE of mail among those who comply with the phone call treatment, (iv) the CATE of mail among those who do not comply with the phone call treatment, and (v) the CACE of phone calls among those who receive mail.

Answer:

- (i) **ATE of mail.** The ATE of mail is identified using the core assumptions of chapter 2 (random assignment, non-interference, and excludability). Excludability in this holds that the only way that random assignment of mail affects outcomes is through the mail treatment itself.
- (ii) **Complier average causal effect (CACE) of phone calls.** In order to identify the CACE of phone calls, we must invoke the assumptions of Chapter 5, since this is a case of one-sided non-compliance. Again, the exclusion restriction holds that the only way that the

²Fieldhouse et al. 2010.

Table 4: Question 7C Table

	Y			
	(1)	(2)	(3)	(4)
white	0.031* (0.016)		0.016 (0.015)	
black		-0.031* (0.016)		-0.046*** (0.016)
highquality	0.015 (0.016)	0.030* (0.016)	-0.002 (0.015)	
lowquality				-0.030* (0.016)
chicago	-0.015 (0.016)	-0.030* (0.016)		-0.042*** (0.016)
boston			0.015 (0.016)	
white:highquality	0.015 (0.023)		0.020 (0.021)	
white:chicago	-0.015 (0.022)			
black:highquality		-0.015 (0.023)		
black:lowquality				0.015 (0.023)
black:chicago		0.015 (0.022)		0.010 (0.022)
highquality:chicago	-0.017 (0.022)	-0.012 (0.022)		
white:highquality:chicago	0.005 (0.031)			
black:highquality:chicago		-0.005 (0.031)		
white:boston			0.015 (0.022)	
highquality:boston			0.017 (0.022)	
white:highquality:boston			-0.005 (0.031)	
lowquality:chicago				0.012 (0.022)
black:lowquality:chicago				0.005 (0.031)
Constant	0.070*** (0.012)	0.101*** (0.012)	0.055*** (0.010)	0.131*** (0.012)
N	4,870	4,870	4,870	4,870
R ²	0.008	0.008	0.008	0.008
Adjusted R ²	0.006	0.006	0.006	0.006
Residual Std. Error (df = 4862)	0.271	0.271	0.271	0.271
F Statistic (df = 7; 4862)	5.359***	5.359***	5.359***	5.359***

*p < .1; **p < .05; ***p < .01

Coefficients depend on student's parameterization

Table 5: Question 9 Table

	Control	Mail Only	Phone Only	Mail and Phone
N	5179	4367	3466	2287
Number Contacted by Phone	0	0	2003	1363
Among those Assigned to this Experimental Group, Percent who Voted	0.397	0.403	0.397	0.418
Among those Contacted by Phone, Percent who Voted	NA	NA	0.465	0.468

assignment of phone calls affects outcomes is through actual phone contacts. The CACE here is ATE among those who receive phone calls if assigned to the treatment group.

- (iii) **CATE of mail among those who comply with the phone call treatment.** The CATE of mail is identified in the same way as an ATE, except that it is restricted to those who actually receive phone calls
 - (iv) **CATE of mail among those who do not comply with the phone call treatment.** Same as above, but among those who are not treated when called.
 - (v) **CACE of phone calls among those who receive mail.** The CACE of phone calls among those who receive mail is identified among the same group of compliers as the CACE above, since mail is assigned and received randomly (because we assume full compliance with the mail treatment). However, the ATE of the calls among Compliers may differ from the ATE among Compliers who also receive mail, due to a treatment-by-treatment interaction.
- b) Using the identification strategies you laid out in part (a), estimate each of the five parameters using the results in the table.
- (i) **ATE of mail.** The estimated ATE of mail is $40.3 - 39.7 = 0.6$ percentage points.
 - (ii) **Complier average causal effect (CACE) of phone calls.** The estimated CACE of phone calls is the ITT divided by the share of compliers: $(39.7 - 39.7) / (2003/3466) = 0$.
 - (iii) **CATE of mail among those who comply with the phone call treatment.** The CATE of mail among those who receive a call is $46.8 - 46.5 = 0.3$ percentage points.
 - (iv) **CATE of mail among those who do not comply with the phone call treatment.** In order to figure out the CATE of mail among those who did not comply when called, we must first back out the voting rates given the numbers presented above. For example, the overall voting rate in the treatment group of 41.8 is a weighted average of the voting rates among the contacted and uncontacted. Thus, $41.8 = 46.8(1363/2287) + X(923/2287)$. Solving for X gives 34.4, and repeating the same calculation for the control group gives 30.4. Therefore, the estimated effect of mail for this subgroup is 4.0 percentage points.
 - (v) **CACE of phone calls among those who receive mail.** The CACE of phone calls among those who receive mail is the ITT divided by the contact rate: $(41.8 - 40.3) / (1363/2287) = 2.5$ percentage points
- c) In Chapters 5 and 6, we discussed the use of instrumental variables regression to estimate CACEs when experiments involve noncompliance. Here, we can apply instrumental variables regression

to a factorial experiment in which one factor encounters noncompliance. With the replication dataset at <http://isps.research.yale.edu/FEDAI>, use instrumental variables regression to estimate the parameters of the Vote equation in the following three-equation regression model:

$$\begin{aligned} \text{PhoneContact}_i &= \alpha_0 + \alpha_1 \text{Mail}_i + \alpha_2 \text{PhoneAssign}_i + \alpha_3 (\text{PhoneAssign}_i * \text{Mail}_i) + e_i \\ \text{PhoneContact}_i * \text{Mail}_i &= \gamma_0 + \gamma_1 \text{Mail}_i + \gamma_2 \text{PhoneAssign}_i + \gamma_3 (\text{PhoneAssign}_i * \text{Mail}_i) + \epsilon_i \\ \text{Vote}_i &= \beta_0 + \beta_1 \text{Mail}_i + \beta_2 \text{PhoneContact}_i + \beta_3 (\text{PhoneContact}_i * \text{Mail}_i) + u_i \end{aligned}$$

Interpret the regression estimates in light of the five parameters you estimated in part (b). Which causal parameters does instrumental variables regression estimate or fail to estimate?

```
fieldhouse <- within(fieldhouse,{
  mail <- m
  phone <- p
  phone_contact <- c
  vote <- y
})

fit.iv <- ivreg(vote ~ mail + phone_contact + phone_contact * mail
               | mail + phone + mail * phone,
               data = fieldhouse)

summary(fit.iv)

##
## Call:
## ivreg(formula = vote ~ mail + phone_contact + phone_contact *
##       mail | mail + phone + mail * phone, data = fieldhouse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4282 -0.4030 -0.3970  0.5970  0.6032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3969878  0.0068099  58.296  <2e-16 ***
## mail           0.0060348  0.0100684   0.599    0.549
## phone_contact  -0.0001781  0.0186142  -0.010    0.992
## mail:phone_contact 0.0253338  0.0282311   0.897    0.370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4901 on 15296 degrees of freedom
## Multiple R-Squared:  0.001046, Adjusted R-squared:  0.0008504
## Wald test: 1.126 on 3 and 15296 DF, p-value: 0.3369
```

The intercept is the voting rate in the control group. The coefficient for “phone contact” is the estimated CACE for phones when no mail is assigned. The effect for “mail” is the ATE for mail when no phone calls are assigned. The coefficient for “mail:phone contact” is the extent to which the apparent CACE of phone calls increases when we move from the no-mail to the mail group. These estimates reproduce the estimates generated by hand above. Notice that IV regression does not report the effect of mail for non-compliers.