Field Experiments: Design, Analysis and Interpretation
Solutions for Chapter 4 Exercises

Alan S. Gerber and Donald P. Green[*]

Question 1

Important concepts: [10pts]

a) Define "covariate." Explain why covariates are (at least in principle) measured prior to the random allocation of subjects to treatment and control.
Answer:
A covariate is a variable that is (1) unaffected by the treatment and (2) used to predict outcomes. In order to increase the credibility of the claim that a given covariate is unaffected by the treatment, researchers typically restrict the set of covariates to those variables that are measured (or are measurable) prior to the random allocation of treatments.

b) Define "disturbance term."
Answer:
The disturbance term comprises all sources of variation in potential outcomes other than the average treatment effect. For example, in equation (4.7), the disturbance term is $u_i = Y_i(0) - \mu_{Y(0)} + [(Y_i(1) - \mu_{Y(1)}) - (Y_i(0) - \mu_{Y(0)})]D_i$. The disturbance term comprises the idiosyncratic variation in untreated responses $Y_i(0) - \mu_{Y(0)}$, plus the idiosyncratic variation in treatment effects $[(Y_i(1) - \mu_{Y(1)}) - (Y_i(0) - \mu_{Y(0)})]D_i$.

c) In equation (4.2), we demonstrated that rescaling the outcome by subtracting a pre-test leads to unbiased estimates of the ATE. Suppose that instead of subtracting the pre-test X_i, we subtracted a rescaled pretest cX_i, where c is some positive constant. Show that this procedure produces unbiased estimates of the ATE.
Answer:
The proof is similar to equation (4.2) and again makes use of the fact that the expected value of X_i is the same in the treatment and control groups when treatments are allocated randomly:

$$\begin{aligned}
E[\widehat{ATE}] &= E[Y_i - cX_i|D_i = 1] - E[Y_i - cX_i|D_i = 0)] \\
&= E[Y_i|D_i = 1] - E[cX_i|D_i = 1] - E[Y_i|D_i = 0] + E[cX_i|D_i = 0] \\
&= E[Y_i|D_i = 1] - cE[X_i|D_i = 1] - E[Y_i|D_i = 0] + cE[X_i|D_i = 0] \\
&= E[Y_i(1)] - E[Y_i(0)]
\end{aligned}$$

d) Show that the parameter b in equation (4.7) is identical to the ATE.
Answer:

Recall from Equation (4.7) that:

$$\begin{aligned}
Y_i &= Y_i(0)(1 - D_i) + Y_i(1)D_i \\
&= Y_i(0) + (Y_i(1) - Y_i(0))D_i \\
&= \mu_{Y(0)} + [\mu_{Y(1)} - \mu_{Y(0)}]D_i + Y_i(0) - \mu_{Y(0)} + [(Y_i(1) - \mu_{Y(1)}) - (Y_i(0) - \mu_{Y(0)})]D_i \\
&= a + bD_i + u_i
\end{aligned}$$

This equation implies that $b = \mu_{Y(1)} - \mu_{Y(0)}$, which is the ATE because the expected value of $Y_i(1)$ is $\mu_{Y(1)}$, and the expected value of $Y_i(0)$ is $\mu_{Y(0)}$.

Question 2

Question 3

The table below illustrates the problems that may arise when researchers exercise discretion over what results to report to readers. Suppose the true ATE associated with a given treatment were 1.0. The table reports the estimated ATE from nine experiments, each of which involves approximately 200 subjects. Each study produces two estimates, one based on a difference-in-means and another using regression to control for covariates. In principle, both estimators generate unbiased estimates, and covariate adjustment has a slight edge in terms of precision. Suppose the researchers conducting each study use the following decision rule: "Estimate the ATE using both estimators and report whichever estimate is larger." Under this reporting policy, are the reported estimates unbiased? Why or why not? [6 pts]

Answer:

This procedure leads to biased estimates. Although each estimator is unbiased, the greater of two unbiased estimates is not unbiased. One can think of this procedure as "Report the no-covariates estimate unless the with-covariates estimate is larger, in which case report the with-covariates estimate." On its own, the no-covariates estimate is unbiased, but it tends to be corrected when it generates a lower-than-average estimate. In this example, the average estimate generated by this reporting procedure is $12/9 = 1.33$, which is greater than the true ATE of 1.0.

Table 1: Question 3 table

Study	No covariates	With covariates	Greater of two estimates
1	5	4	5
2	3	3	3
3	2	2	2
4	6	5	6
5	1	1	1
6	0	0	0
7	-3	-1	-1
8	-5	-4	-4
9	0	-1	0
Average	1	1	1.33
Standard Deviation	3.54	2.83	3.08

Question 4

Question 5

Randomizations are said to be "restricted" when the set of all possible random allocations is narrowed to exclude allocations that have inadequate covariate balance. Suppose, for example, that the assignment of treatments (D_i) in Table 4.1 was conducted subject to the restriction that a regression of D_i on X_i (the pretest) does not allow the researcher to reject the sharp null hypothesis of no effect of X_i on D_i at the 0.05 significance level) produces a p-value on that is greater than 0.05. In other words, had the researcher found that the assigned D_i were significantly predicted by X_i, the random allocation would have been conducted again, until the D_i met this criterion. [10pts]

a) Conduct a series of random assignments in order to calculate the weighting variable w_i; for units in the treatment group, this weight is defined as the inverse of the probability of being assigned to treatment, and for units in the control group, this weight is defined as the inverse of the probability of being assigned to control. See Table 4.2 for an example. Does w_i appear to vary within the treatment group or within the control group?

```
In [1]: clear
        clear matrix
        clear mata
        set matsize 11000
        set maxvar 32767
```

```stata
        set seed 1234567
        set more off
```

In [2]:
```stata
// loop to simulte random assignment and save to a matrix
        cap matrix drop z
        matrix z=J(40, 10000, .)

        qui forvalues i = 1/10000 {
        //create and save 50 permutations of treatment
                import delim
        "./data/chapter04/GerberGreenBook_Chapter4_Exercises_4-5.csv"
                tempvar teststat Z
                gen `Z' = .
                gen `teststat' = -1
                while `teststat' < 0.05{
                        tempvar rannum Zri t
                    gen `rannum'=uniform()
                        egen `Zri' = cut(`rannum'), group(2)
                        qui reg `Zri' x
                        gen `t' = _b[x]/_se[x]
                        replace `teststat' = 2*ttail(e(df_r),abs(`t'))
                }
                replace `Z' = `Zri'
                forvalues j = 1/40 {
                matrix z[`j', `i'] = `Z'[`j']
                }
                drop _all
        }
```

In [3]:
```stata
qui import delim
        "./data/chapter04/GerberGreenBook_Chapter4_Exercises_4-5.csv", clear
        rename d D
        rename y Y
        matrix rowm = z * J(colsof(z), 1, 1/colsof(z))
        matrix colnames rowm=probs
        svmat double rowm, names(col)
        qui gen weights = (1/probs)*D +(1/(1-probs))*(1-D)
        tabstat weights, by(D) stat(v)
```

```
Summary for variables: weights
     by categories of: D (D)

      D |  variance
---------+----------
      0 |  .0004185
      1 |   .000625
---------+----------
```

```
   Total |   .0005084
--------------------
```

The variance of the weights is 4×10^{-4} in the treatment condition and 6×10^{-4} in the control condition. Indeed, units do have different probabilities of assignments as a result of the restriction scheme, but the differences are small.

b) Use randomization inference to test the sharp null hypothesis that D_i has no effect on Y_i by regressing Y_i on D_i and comparing the estimate to the sampling distribution under the null hypothesis. Make sure that your sampling distribution includes only random allocations that satisfy the restriction mentioned above. Be sure to weight units by inverse probability weights as produced by the random allocation procedure. Estimate the ATE, calculate the p-value, and interpret the results.

```
In [4]: qui reg Y D [pw=weights]
        global ate_restricted_RA = _b[D]

        di "ATE (Restricted Assignment)= " $ate_restricted_RA

ATE (Restricted Assignment)= 10.712532

In [5]: svmat z

        cap matrix drop y_dis
        matrix y_dis=J(10000, 1, .)

        forvalues i = 1/10000 {
                tempvar weight`i'
                gen `weight`i'' = (1/probs)*z`i' +(1/(1-probs))*(1-z`i')
                qui reg Y z`i' [pw=`weight`i'']
                matrix y_dis[`i', 1] = _b[z`i']

        }

In [6]: preserve
         svmat y_dis
         // p value
         count if abs(y_dis1) > abs($ate_restricted_RA)
         di r(N)/_N
         restore

   56

.0056
```

The IPW estimate of the ATE is 10.71, which is close to the unweighted estimate above. Using a two-tailed test in order to evaluate the null hypothesis that the treatment has no effect for

5

any subject, we find a p-value of 0.0056, which leads us to reject the null hypothesis in favor of the alternative hypothesis that the treatment has some effect.

c) Use randomization inference to test the sharp null hypothesis that D_i has no effect on Y_i by regressing Y_i on D_i and X_i and comparing the estimate to the sampling distribution under the null hypothesis. Estimate the ATE, calculate the p-value, and interpret the results.

```
In [7]: qui reg Y x D [pw=weights]
        global ate_cov_restricted_RA = _b[D]
        di "ATE Controlling Covariance (Restricted Assignment)= "
                          $ate_cov_restricted_RA

ATE Controlling Covariance (Restricted Assignment)= 5.3186105

In [8]: cap matrix drop cov_dis
        matrix cov_dis=J(10000, 1, .)
        forvalues i = 1/10000{
                        tempvar weight`i'
                        gen `weight`i'' = (1/probs)*z`i' +(1/(1-probs))*(1-z`i')
                        qui reg Y x z`i' [pw=`weight`i'']
                        matrix cov_dis[`i', 1] = _b[z`i']

        }

In [9]: preserve
        svmat cov_dis

        // p.value controlling covariance
        count if abs(cov_dis1) > abs($ate_cov_restricted_RA)
        di r(N)/_N
        restore

   28

.0028
```

The IPW estimate of the ATE is 5.32, which is close to the unweighted estimate above. We again use a two-tailed test in order to evaluate the null hypothesis that the treatment has no effect for any subject. We find a p-value of 0.0028, which leads us to reject the null hypothesis in favor of the alternative hypothesis that the treatment has some effect.

d) Compare the sampling distributions under the null hypothesis in parts (a) and (b) to the sampling distributions obtained in exercises 4(d) and 4(e), which assumed that the randomization was unrestricted.

```
In [10]: /*----------------------se_complete_RA

         // calculate ate_complete_RA
         qui reg Y D
```

6

```stata
        global ate_complete_RA = _b[D]
        di "ATE under Complete Assignment = " $ate_complete_RA
```

ATE under Complete Assignment = 10.7

```stata
In [11]: // RI under the null ate=ate_complete_RA
        cap drop Y0_sim Y1_sim Y_sim
        qui gen Y0_sim = Y
        qui gen Y1_sim = Y
        qui gen Y_sim = .
        qui replace Y0_sim = Y - $ate_complete_RA if D==1
        qui replace Y1_sim = Y + $ate_complete_RA if D==0
```

```stata
In [12]: capture program drop ate_complete_RA_ri
        program define ate_complete_RA_ri, rclass
                replace Y_sim = Y0_sim*(1-D) + Y1_sim*(D)
                regress Y_sim D
            return scalar Ys_complete_RA=_b[D]
        end

        tsrtest D r(Ys_complete_RA) using distout_complete_RA.dta, ///
        overwrite: ate_complete_RA_ri
```

Two-sample randomization test for
theta=r(Ys_complete_RA) of ate_complete_RA_ri by D

Combinations: 137846528820 = (40 choose 20)
Assuming null=0
Observed theta: 10.7

Minimum time needed for exact test (h:m:s):   392479:42:00
Reverting to Monte Carlo simulation.
Mode: simulation (10000 repetitions)

progress: |...|

 p=0.48995 [one-tailed test of Ho: theta(D==0)<=theta(D==1)]
 p=0.50995 [one-tailed test of Ho: theta(D==0)>=theta(D==1)]
 p=0.48995 [two-tailed test of Ho: theta(D==0)==theta(D==1)]

Saving log file to distout_complete_RA.dta...done.

```stata
In [13]: // calculate se_complete_RA
        preserve
        use "distout_complete_RA.dta", clear
        qui drop if _n==1
        tabstat theta, stat(sd)
        restore
```

```
    variable |         sd
-------------+----------
       theta |   4.673591
----------------------
```

In [14]: `/*--------------------se_cov_complete_RA`

```
         // calculate ate_cov_complete_RA
         qui reg Y D x
         global ate_cov_complete_RA = _b[D]
         di "ATE Controlling Covariance under Complete Assignment= " $ate_cov_complete_RA
```

```
ATE Controlling Covariance under Complete Assignment= 5.3155362
```

In [15]: `// RI under the null ate= ate_cov_complete_RA`
```
         cap drop Y0_sim Y1_sim Y_sim
         qui gen Y0_sim = Y
         qui gen Y1_sim = Y
         qui gen Y_sim = .
         qui replace Y0_sim = Y - $ate_cov_complete_RA if D==1
         qui replace Y1_sim = Y + $ate_cov_complete_RA if D==0
```

In [16]: `capture program drop ate_cov_complete_RA_ri`
```
         program define ate_cov_complete_RA_ri, rclass
                 replace Y_sim = Y0_sim*(1-D) + Y1_sim*(D)
                 regress Y_sim D x
             return scalar Ys_cov_complete_RA=_b[D]
         end

         tsrtest D r(Ys_cov_complete_RA) using distout_cov_complete_RA.dta, ///
         overwrite: ate_cov_complete_RA_ri
```

```
Two-sample randomization test for
theta=r(Ys_cov_complete_RA) of ate_cov_complete_RA_ri by D

Combinations:   137846528820 = (40 choose 20)
Assuming null=0
Observed theta: 5.316

Minimum time needed for exact test (h:m:s):   408561:47:43
Reverting to Monte Carlo simulation.
Mode: simulation (10000 repetitions)

progress: |...|

 p=0.50495 [one-tailed test of Ho:   theta(D==0)<=theta(D==1)]
```

```
 p=0.49495 [one-tailed test of Ho:  theta(D==0)>=theta(D==1)]
 p=0.50495 [two-tailed test of Ho:  theta(D==0)==theta(D==1)]

Saving log file to distout_cov_complete_RA.dta...done.
```

In [17]: `// calculate se_cov_complete_RA`
```
        preserve
        use "distout_cov_complete_RA.dta", clear
        qui drop if _n==1
        tabstat theta, stat(sd)
        restore
```

```
    variable |        sd
-------------+----------
       theta |  1.577595
-----------------------
```

In [18]: `/*--------------------se_restricted_RA`

```
        // calculate ate_restricted_RA
        di "ATE under Restricted Assignment= " $ate_restricted_RA
```

```
ATE under Restricted Assignment= 10.712532
```

In [19]: `// RI under the null ate= ate_restricted_RA`
```
        cap drop Y0_sim Y1_sim Y_sim
        qui gen Y0_sim = Y
        qui gen Y1_sim = Y
        qui gen Y_sim = .
        qui replace Y0_sim = Y - $ate_restricted_RA if D==1
        qui replace Y1_sim = Y + $ate_restricted_RA if D==0
```

In [20]: `// clean space`
```
        drop __00*
```

In [21]: `cap matrix drop distout_restricted_RA`
```
        matrix distout_restricted_RA=J(10000, 1, .)

        qui forvalues i = 1/10000 {
                tempvar weight`i'
                gen `weight`i'' = (1/probs)*z`i' +(1/(1-probs))*(1-z`i')
                replace Y_sim = Y0_sim*(1-z`i') + Y1_sim*(z`i')
                qui reg Y_sim z`i' [pw=`weight`i'']
                matrix distout_restricted_RA[`i', 1] = _b[z`i']

        }
```

```
In [22]: /*se_restricted_RA*/
         preserve
         svmat distout_restricted_RA
         tabstat distout_restricted_RA, stat(sd)
         restore

    variable |        sd
-------------+----------
distout_re~1 |  4.139767
------------------------

In [23]: /*--------------------se_cov_restricted_RA

         // calculate ate_restricted_RA
         di "ATE Controlling Covariance(Restricted Assignment)=
          " $ate_cov_restricted_RA

ATE Controlling Covariance(Restricted Assignment)= 5.3186105

In [24]: // RI under the null ate= ate_cov_restricted_RA

         cap drop Y0_sim Y1_sim Y_sim
         qui gen Y0_sim = Y
         qui gen Y1_sim = Y
         qui gen Y_sim = .
         qui replace Y0_sim = Y - $ate_cov_restricted_RA if D==1
         qui replace Y1_sim = Y + $ate_cov_restricted_RA if D==0

In [25]: // clean space
         drop __00*

In [26]: cap matrix drop distout_cov_restricted_RA
         matrix distout_cov_restricted_RA=J(10000, 1, .)

         qui forvalues i = 1/10000 {
                 replace Y_sim = Y0_sim*(1-z`i') + Y1_sim*(z`i')
                 tempvar weight`i'
                 gen `weight`i'' = (1/probs)*z`i' +(1/(1-probs))*(1-z`i')
                 qui reg Y_sim x z`i' [pw=`weight`i'']
                 matrix distout_cov_restricted_RA[`i', 1] = _b[z`i']

         }

In [27]: /*se_restricted_RA*/
         preserve
         svmat distout_cov_restricted_RA
         tabstat distout_cov_restricted_RA, stat(sd)
         restore
```

```
    variable |         sd
-------------+----------
distout_co~1 |  1.592966
-----------------------
```

Table 2: Summary of Estimated Standard Errors

	Without Covariates	With Covariates
Complete Random Assignment	4.674	1.578
Restricted Random Assignment	4.140	1.593

Without covariates and assuming complete randomization, we obtain a standard error of 4.674. Under restricted randomization, the standard error declines to 4.140. Including a covariate and assuming complete randomization, we obtain a standard error of 1.578. Under restricted randomization, the standard error remains essentially unchanged at 1.593. Restricted randomization is akin to blocking, in that it rules out random allocations that result in imbalance; however, its advantages in terms of precision are limited when the researcher controls for a strongly prognostic covariate, which achieves most of the precision gains associated with blocking.

Question 6

Question 7

Researchers may be concerned about using block randomization when they are unsure whether the variable used to form the blocks actually predicts the outcome. Consider the case in which blocks are formed randomly – in other words, the variable used to form the blocks has no prognostic value whatsoever. Below is a schedule of potential outcomes for four observations. [10pts]

Table 3: Question 7 Table

Subject	Y(0)	Y(1)
A	1	2
B	0	3
C	2	2
D	5	5

Table 4: Question 7a table

Treated Units	$Y(1)$	$Y(0)$	$\widehat{ATE}$
A and B	2.5	3.5	-1
A and C	2	2.5	-0.5
A and D	3.5	1	2.5
B and C	2.5	3	-0.5
B and D	4	1.5	2.5
C and D	3.5	0.5	3

a) Suppose you were to use complete random assignment such that $m = 2$ units are assigned to treatment. What is the sampling variance of the difference-in-means estimator across all six possible random assignments?

The average estimated ATE is 1.0, which is the true ATE. The variance of the estimated ATEs over all 6 possible randomizations is 2.833.

b) Suppose you were to form blocks by randomly pairing the observations. Within each pair, you randomly allocate one subject to treatment and the other to control so that $m = 2$ units are assigned to treatment.There are three possible blocking schemes; for each blocking scheme, there are four possible random assignments. What is the sampling variance of the difference-in-means estimator across all twelve possible random assignments?

Table 5: Question 7b table

	Treated Units	$Y(1)$	$Y(0)$	$\widehat{ATE}$
AB and CD blocked	A,C	2	2.5	-0.5
	A,D	3.5	1	2.5
	B,D	4	1.5	2.5
	B,C	2.5	3	-0.5
AC and BD blocked	A,B	2.5	3.5	-1
	A,D	3.5	1	2.5
	C,B	2.5	3	-0.5
	C,D	3.5	0.5	3
AD and BC blocked	A,B	2.5	3.5	-1
	A,C	2	2.5	-0.5
	D,B	4	1.5	2.5
	D,C	3.5	0.5	3

Across the 12 possible random assignments, the variance of the estimated ATE is again 2.833. Notice that every estimate in the previous table appears in this table twice.

c) From this example, what do you infer about the risks of blocking on a non-prognostic covariate?
Answer:
There is no risk of increasing variance with a useless blocking variable; at worst, the variable

will be random noise, in which case the sampling variance will be the same as a design without blocking.

Question 8

Question 9

Gerber and Green conducted a mobilization experiment in which calls from a large commercial phone bank urged voters in Iowa and Michigan to vote in the November 2002 election.[1] The randomization was conducted within four blocks: uncompetitive congressional districts in Iowa, competitive congressional districts in Iowa, uncompetitive congressional districts in Michigan, and competitive congressional districts in Michigan. Table 4.3 presents results only for one-voter households in order to sidestep the complications of cluster assignment. [10pts]

a) Within each of the four blocks, what was the apparent effect of being called by a phone bank on voter turnout?
   Answer:
   From the "Estimated ATE" Row: Block 1: .0096, Block 2: -.0078, Block 3: -.0136, Block 4: .0083. Substantively, these results suggest that calls encouraging voter turnout had effects ranging from -1.4 percentage points to +1.0 percentage point.

b) When all of the subjects in this experiment are combined (see the rightmost column of the table), turnout seems substantially higher in the treatment group than the control group. Explain why this comparison gives a biased estimate of the ATE.
   Answer:
   This estimator is biased because individuals in each stratum had different propensities to enter into treatment. The uncompetitive Michigan block has the lowest rate of treatment and also has the lowest rate of voting in the control group. Overall, blocks with higher rates of treatment tend to have higher rates of voting in the control group, which accounts for the upward bias.

c) Using the weighted estimator described in Chapter 3, show the calculations used to generate an unbiased estimate of the overall ATE.

```
In [1]: clear
        input ests shareoftotalN
        .00964 .049487
        -.007829 .1520981
        -.01362 .626616
        .008271 .171799
```

[1]Gerber and Green 2005.

```
          end

          qui gen overall_ate = ests*shareoftotalN
          total(overall_ate)

          ests  shareof~N

Total estimation                        Number of obs    =          4

          ----------------------------------------------------------------
                     |       Total    Std. Err.     [95% Conf. Interval]
          -----------+-----------------------------------------------------
          overall_ate |   -.0078273    .0090322      -.0365719     .0209173
          ----------------------------------------------------------------
```

d) When analyzing block randomized experiments, researchers frequently use regression to estimate the ATE by regressing the outcome on the treatment and indicator variables for each of the blocks (omitting one block if the regression includes an intercept.) This regression estimator places extra weight on blocks that allocate approximately half of the subjects to the treatment condition (i.e., $P_j = 0.5$) because these blocks tend to estimate the within-block ATE with less sampling variability. Compare the four OLS weights to the weights W_j used in part (c).
Answer:
The weights used in part (c) are based on the share of the subject pool that is in each block. This weighting scheme places a great deal of weight on the relatively large Michigan block. By contrast, the OLS weights are a blend of the number of subjects in each block and each block's balance between treatment and control allocations. Because the blocks do not differ very much in terms of their allocation rates, the OLS weights tend to be similar across blocks.

e) Regression provides an easy way to calculate the weighted estimate of the ATE in part (c) above. For each treatment subject i, compute the proportion of subjects in the same block who were assigned to the treatment group. For control subjects, compute the proportion of subjects in the same block who were assigned to the control group. Call this variable q_i. Regress outcomes on treatment, weighting each observation by $1/q_i$, and show that this type of weighted regression produces the same estimate as weighting the estimated ATEs for each block.

```
In [2]: clear
        qui import delim ./data/chapter04/Gerber_Green_AAAPSS_2005, clear
        qui bysort strata: egen blockpr = mean(treat2)
        qui gen q = blockpr*treat2 + (1-blockpr)*(1-treat2)
        regress vote02 treat2 [aw=1/q]

(sum of wgt is   1.8814e+06)

      Source |       SS           df       MS      Number of obs   =    940,715
-------------+----------------------------------   F(1, 940713)    =      57.98
       Model |   14.4110635        1   14.4110635   Prob > F        =     0.0000
```

```
   Residual |  233826.144     940,713   .248562679     R-squared       =    0.0001
------------+-------------------------------     Adj R-squared   =    0.0001
      Total |  233840.555     940,714   .248577734     Root MSE        =    .49856


--------------------------------------------------------------------------
     vote02 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+-------------------------------------------------------------
     treat2 |   -.007828   .0010281    -7.61   0.000    -.0098429    -.005813
      _cons |   .4661975   .0007269   641.31   0.000     .4647727    .4676223
--------------------------------------------------------------------------
```

The coefficient on the treatment indicator is -0.0078, which is the same as was found in part c.

Question 10