

# 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:

---

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

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

A researcher working with Israeli elementary school students sought to improve students' ability to solve logic puzzles.<sup>1</sup> Students in the treatment and control group initially took a computer-administered test, and the number of correctly solved puzzles was recorded. A few days later, students assigned to the control group were then given 30 minutes to improve their puzzle-solving skills by playing on a computer. During the same allotment of time, students in the treatment group listened to an instructor describe some rules of thumb to keep in mind when solving logic puzzles. All subjects then took a computer-administered post-test, and the number of correctly solved puzzles was recorded. The table below shows the results for each subject. [10pts]

Table 1: Question 2 Table

Subject	D	Pre-test	Post-test	Improvement
1	1	10	10	0
2	1	9	11	2
3	1	5	6	1
4	1	3	6	3
5	1	3	6	3
6	1	6	7	1
7	1	6	7	1
8	1	5	6	1
9	1	6	7	1
10	0	9	9	0
11	0	6	7	1
12	0	11	10	-1
13	0	4	5	1
14	0	3	3	0
15	0	10	10	0
16	0	7	8	1
17	0	7	7	0
18	0	8	10	2

- a) As a randomization check, use randomization inference to test the null hypothesis that the pre-test scores are unaffected by treatment assignment.

<sup>1</sup>Dan Gendelman conducted this study in 2004 and shared it with us via personal communication.

```

In [1]: qui import delim ./data/chapter04/Gendelman_2004, clear
        set seed 1234567
        rename treatment D
        rename posttest Y
        rename pretest X

In [2]: capture program drop Fstat
        program define Fstat, rclass
            regress D X
            return scalar Fs=e(F)
        end

        // calculate 48620 (18 choose 9) permutations
        tsrtest D r(Fs) using 4_2_Fstat.dta, overwrite: Fstat

```

Two-sample randomization test for  $\theta = r(F_s)$  of Fstat by D

Combinations: 48620 = (18 choose 9)  
 Assuming null=0  
 Observed  $\theta$ : 1.274

Minimum time needed for exact test (h:m:s): 0:03:27  
 Mode: exact

progress: |...|

p=0.31345 [one-tailed test of  $H_0: \theta(D=0) \leq \theta(D=1)$ ]  
 p=0.76851 [one-tailed test of  $H_0: \theta(D=0) \geq \theta(D=1)$ ]  
 p=0.31345 [two-tailed test of  $H_0: \theta(D=0) = \theta(D=1)$ ]

Saving log file to 4\_2\_Fstat.dta...done.

```

In [3]: // ate
        di r(obsvStat)

```

1.2743363

```

In [4]: // p.valueb
        di r(uppertail)

```

.31345125

We calculated the F-statistic of a regression of treatment assignment on the pretest score for all possible randomizations, and found that the observed F-statistic was larger than 31.35% of the simulated statistics, implying a  $p$ -value of 0.313. As expected, we fail to reject the null

hypothesis that the treatment assignment is unrelated to the pretreatment covariate, pretest.

- b) Use difference-in-means estimation to estimate the effect of the treatment on the post-test score. Form a 95% confidence interval.

```
In [5]: // calculate ate
        qui reg Y D
        global tau = _b[D]
        di "ATE = " $tau
```

ATE = -.33333333

```
In [6]: // RI under the null ate=ate
        qui gen Y0_sim = Y
        qui gen Y1_sim = Y
        qui gen Y_sim = .
        qui replace Y0_sim = Y - $tau if D==1
        qui replace Y1_sim = Y + $tau if D==0

        capture program drop ate_ci
        program define ate_ci, rclass
            replace Y_sim = Y0_sim*(1-D) + Y1_sim*(D)
            regress Y_sim D
            return scalar Ys=_b[D]
        end

        tsrtest D r(Ys) using ate_ci.dta, overwrite: ate_ci
```

Two-sample randomization test for  $\theta = r(Y_s)$  of ate\_ci by D

Combinations: 48620 = (18 choose 9)

Assuming null=0

Observed theta: -.3333

Minimum time needed for exact test (h:m:s): 0:03:26

Mode: exact

progress: |...|

p=0.50428 [one-tailed test of  $H_0: \theta(D=0) \leq \theta(D=1)$ ]

p=0.50000 [one-tailed test of  $H_0: \theta(D=0) \geq \theta(D=1)$ ]

p=0.76450 [two-tailed test of  $H_0: \theta(D=0) = \theta(D=1)$ ]

Saving log file to ate\_ci.dta...done.

```

In [7]: preserve
        use "ate_ci.dta", clear
        qui drop if _n==1

        sort theta
        // 95% confidence interval (CI)
        di "(" round(theta[floor(_N*0.025)], 0.001) ", "
        round(theta[floor(_N*0.975)], 0.001) ")"

        restore

(-2.259, 1.593)

```

We obtained a difference-in-means estimate of the ATE of  $-0.3333333$  and a 95% confidence interval of  $[-2.26, 1.59]$ . This confidence interval is wide enough to include much larger and much smaller treatment effects – even crossing zero.

- c) Use difference-in-differences estimation to estimate the effect of the treatment on the post-test score. Form a 95% confidence interval, and compare it to the interval in part (b).

```

In [8]: rename improvement Y_improve

        // calculate ate.improve
        qui reg Y_improve D
        global tau_im = _b[D]
        di "ATE.improve = " $tau_im

ATE.improve = 1

In [9]: // RI under the null ate=ate.improve
        qui replace Y0_sim = Y_improve
        qui replace Y1_sim = Y_improve
        qui replace Y_sim = .

        qui replace Y0_sim = Y_improve - $tau_im if D==1
        qui replace Y1_sim = Y_improve + $tau_im if D==0

In [10]: capture program drop ate_im_ci
         program define ate_im_ci, rclass
             replace Y_sim = Y0_sim*(1-D) + Y1_sim*(D)
             regress Y_sim D
             return scalar Ys_im=_b[D]
         end

         tsrtest D r(Ys_im) using ate_im_ci.dta, overwrite: ate_im_ci

```

Two-sample randomization test for  $\theta = r(Ys\_im)$  of `ate_im_ci` by `D`

```

Combinations: 48620 = (18 choose 9)
Assuming null=0
Observed theta: 1

Minimum time needed for exact test (h:m:s): 0:03:29
Mode: exact

progress: |...|

p=0.58739 [one-tailed test of Ho: theta(D==0)<=theta(D==1)]
p=0.59726 [one-tailed test of Ho: theta(D==0)>=theta(D==1)]
p=0.58739 [two-tailed test of Ho: theta(D==0)==theta(D==1)]

Saving log file to ate_im_ci.dta...done.

In [11]: use "ate_im_ci.dta", clear
         drop if _n==1

         sort theta

         // 95% confidence interval (CI)

         // 95% CI is different from R result due to rounding
         // the permutation test is exactly the same

         di "95%CI = " "(" round(theta[floor(_N*0.025)], 0.001) " , "
         round(theta[floor(_N*0.975)], 0.001) ")"

95%CI = (.111 , 1.889)

```

By subtracting a pre-test, we have sharpened our estimates. The difference-in-differences estimate of the ATE is 1 and the 95% confidence interval is [0.11, 1.89]. No longer does the 95% confidence interval cross zero, meaning we can be confident at the 95% level that the estimated ATE is larger than zero. This contrasts with part b) where the background variability in test scores made the estimation of a small treatment effect more difficult.

### 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 2: 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

Table 4.1 contains a column of treatment assignments,  $D_i$ , that reflects a complete random assignment of 20 schools to treatment and 20 schools to control. [14pts]

- a) Use equation (2.2) to generate observed outcomes based on these assigned treatments. Regress  $Y_i$  on  $D_i$  and interpret the slope and intercept. Is the estimated slope the same as the estimated ATE based on a difference-in-means?

```
In [1]: qui import delim
        "./data/chapter04/GerberGreenBook_Chapter4_Exercises_4-5.csv",clear
        set seed 1234567
```

```
In [2]: rename d D
        regress y D
        qui summ y if D==1, meanonly
        scalar treat_mean = r(mean)
        qui summ y if D==0, meanonly
        scalar control_mean = r(mean)
```

Source		SS	df	MS	Number of obs	=	40
-----+-----							
Model		1144.9	1	1144.9	F(1, 38)	=	5.16
Residual		8427.5	38	221.776316	Prob > F	=	0.0288
-----+-----							
Total		9572.4	39	245.446154	R-squared	=	0.1196
					Adj R-squared	=	0.0964
					Root MSE	=	14.892

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
D	10.7	4.709313	2.27	0.029	1.166494	20.23351
_cons	26.85	3.329987	8.06	0.000	20.10879	33.59121

```
In [3]: disp "ATE =" treat_mean - control_mean
```

```
global tau = treat_mean - control_mean
```

```
ATE =10.7
```

The estimate obtained with OLS regression (10.7) is identical to the estimate obtained with difference-in-means (10.7).

- b) Regress treated and untreated outcomes on  $X_i$  to see whether the condition in equation (4.6) appears to hold. What do you infer about the advisability of rescaling the dependent variable so that the outcome is a change (i.e.,  $Y_i - X_i$ )?

```
In [4]: qui regress y x if D==1
        local co_treat = _b[x]
        qui regress y x if D==0
        local co_control= _b[x]
        disp "Sum of Coefficients = "`co_treat' + `co_control'
```

```
Sum of Coefficients = 1.8460607
```

```
In [5]: qui gen ydiff = y - x
        regress ydiff D
```

Source		SS	df	MS	Number of obs	=	40
-----+-----							
Model		235.225	1	235.225	F(1, 38)	=	8.93
Residual		1000.55	38	26.3302632	Prob > F	=	0.0049
-----+-----							
Total		1235.775	39	31.6865385	R-squared	=	0.1903
					Adj R-squared	=	0.1690
					Root MSE	=	5.1313



ydifff	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D	4.85	1.62266	2.99	0.005	1.565096	8.134904
_cons	-1	1.147394	-0.87	0.389	-3.322778	1.322778

Substituting regression estimates for the true ratio of covariances to variances satisfies the inequality, suggesting that the use of this covariate will improve precision.

$$\frac{Cov(\widehat{Y_i(0)}, X_i)}{Var(X_i)} + \frac{Cov(\widehat{Y_i(1)}, X_i)}{Var(X_i)} = 0.8995221 + 0.9465386 = 1.8460607$$

We also see that the standard errors have shrunk substantially – the standard error for a regression of  $Y$  on  $D$  is 4.7093133, whereas the standard error for the regression of the change in  $Y$  on  $D$  is 1.6226603

- c) Regress  $Y_i$  on  $D_i$  and  $X_i$ . Interpret the regression coefficients, contrasting these results with those obtained from a regression of  $Y_i$  on  $D_i$  alone.

```
In [6]: regress y D x
        global tau_c = _b[D]
```

Source	SS	df	MS	Number of obs	=	40
Model	8627.78597	2	4313.89298	F(2, 37)	=	168.97
Residual	944.614033	37	25.530109	Prob > F	=	0.0000
Total	9572.4	39	245.446154	R-squared	=	0.9013
				Adj R-squared	=	0.8960
				Root MSE	=	5.0527

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D	5.315536	1.628474	3.26	0.002	2.015935	8.615138
x	.9204212	.0537624	17.12	0.000	.8114883	1.029354
_cons	1.216271	1.875729	0.65	0.521	-2.584317	5.016858

The estimated ATE (5.32) is now roughly half the size as the original difference-in-means. (This estimate also happens to be much closer to the true ATE of 4.0.) Comparing the estimated standard errors from both regressions suggests that the inclusion of a covariate has greatly improved precision.

- d) With the estimates obtained in part (a), use randomization inference (as described in Chapter 3) to evaluate the sharp null hypothesis of no effect for any school. To obtain the sampling distribution under the sharp null hypothesis, simulate 100,000 random assignments, and for each simulated sample, estimate the ATE using a regression of  $Y_i$  on  $d_i$ . Interpret the results.

```
In [7]: capture program drop ate
        program define ate, rclass
            regress y D
            return scalar Ys=_b[D]
        end

        tsrtest D r(Ys) using ate.dta, overwrite: ate
```

Two-sample randomization test for  $\theta = r(Y_s)$  of ate by D

Combinations: 137846528820 = (40 choose 20)

Assuming null=0

Observed theta: 10.7

Minimum time needed for exact test (h:m:s): 161969:40:17

Reverting to Monte Carlo simulation.

Mode: simulation (10000 repetitions)

progress: |...|

p=0.01490 [one-tailed test of  $H_0: \theta(D=0) \leq \theta(D=1)$ ]

p=0.98600 [one-tailed test of  $H_0: \theta(D=0) \geq \theta(D=1)$ ]

p=0.03110 [two-tailed test of  $H_0: \theta(D=0) = \theta(D=1)$ ]

Saving log file to ate.dta...done.

```
In [8]: //ate
        di r(obsvStat)
```

10.7

```
In [9]: // p.value
        di r(twotail)
```

.03109689

```
In [10]: preserve
          //historgam
```

```

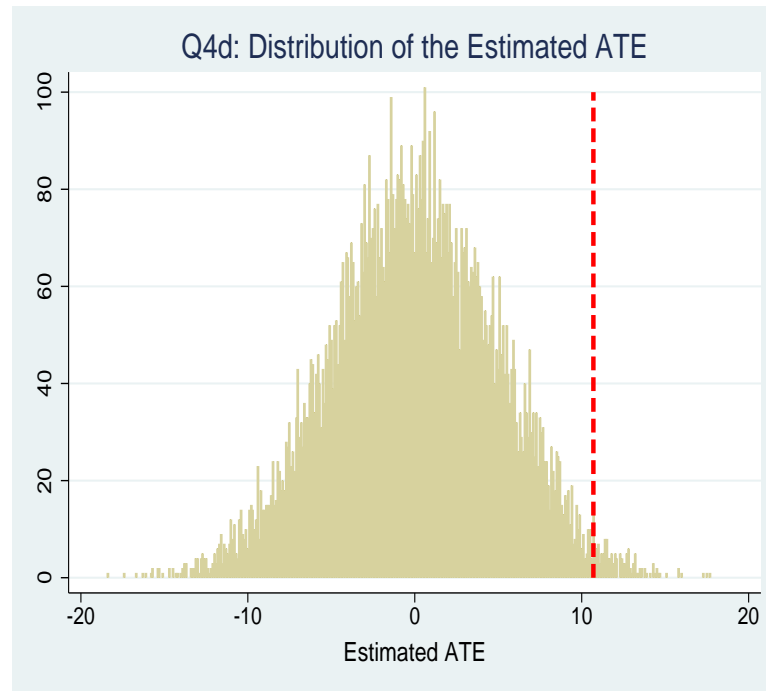
use "ate.dta", clear
global tau = theta[1]
drop if _n==1

graph twoway (histogram theta,frequency bin(1000)) ///
(scatteri 0 $tau 100 $tau,
c(1) lc(red) lw(thick) lp(dash) m(i)), legend(off) ///
b1title("Estimated ATE")
title("Q4d: Distribution of the Estimated ATE") ///
xtitle("")

graph export ./exercise_4_4_d_graph.pdf

restore

```



We use a two-tailed test in order to evaluate the null hypothesis that the treatment has no effect for any subject. We find a two-tailed  $p$ -value of 0.031, which leads us to reject the null hypothesis in favor of the alternative hypothesis that the treatment has some positive effect.

- e) Using the estimator in part (c), use randomization inference to evaluate the sharp null hypothesis of no effect for any school. To obtain the sampling distribution under the sharp null hypothesis, simulate 100,000 random assignments, and for each simulated sample, estimate the ATE using a regression of  $Y_i$  on  $D_i$  and  $X_i$ . Interpret the results.

```

In [11]: capture program drop ate_cov
          program define ate_cov, rclass
            regress y D x
            return scalar Ys=_b[D]

```

```

end

tsrtest D r(Ys) using ate_cov.dta, overwrite: ate_cov

Two-sample randomization test for theta=r(Ys) of ate_cov by D

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

Minimum time needed for exact test (h:m:s): 164458:34:01
Reverting to Monte Carlo simulation.
Mode: simulation (10000 repetitions)

progress: |...|

p=0.00150 [one-tailed test of Ho: theta(D==0)<=theta(D==1)]
p=0.99840 [one-tailed test of Ho: theta(D==0)>=theta(D==1)]
p=0.00290 [two-tailed test of Ho: theta(D==0)==theta(D==1)]

Saving log file to ate_cov.dta...done.

In [12]: //ate_cov
         di r(obsvStat)

5.3155362

In [13]: // p.value_cov
         di r(twotail)

.00289971

```

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 two-tailed p-value of 0.003, which leads us to reject the null hypothesis in favor of the alternative hypothesis that the treatment has some effect.

- f) Use the estimated ATE in part (a) to construct a full schedule of potential outcomes for all schools, assuming that every school has the same treatment effect. Using this simulated schedule of potential outcomes, construct a 95% confidence interval for the sample average treatment effect in the following way. First, assign each subject to treatment or control, and estimate the ATE by a regression of  $Y_i$  on  $D_i$ . Repeat this procedure until you have 100,000 estimates of the ATE. Order the estimates from smallest to largest. The 2,501st estimate marks the 2.5th percentile, and the 97,500th estimate marks the 97.5th percentile. Interpret the results.

```

In [14]: 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 - $tau if D==1
        qui replace Y1_sim = y + $tau if D==0

In [15]: capture program drop ate_null
        program define ate_null, rclass
            replace Y_sim = Y0_sim*(1-D) + Y1_sim*(D)
            regress Y_sim D
            return scalar Ys=_b[D]
        end

        ritest D Ys = r(Ys), reps(10000)
        sav(ate_null.dta, replace) nodots: ate_null

```

Source		SS	df	MS	Number of obs	=	40
-----+-----					F(1, 38)	=	5.16
Model		1144.9	1	1144.9	Prob > F	=	0.0288
Residual		8427.5	38	221.776316	R-squared	=	0.1196
-----+-----					Adj R-squared	=	0.0964
Total		9572.4	39	245.446154	Root MSE	=	14.892

Y_sim		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----							
D		10.7	4.709313	2.27	0.029	1.166494	20.23351
_cons		26.85	3.329987	8.06	0.000	20.10879	33.59121

```

        command: ate_null
        Ys: r(Ys)
        res. var(s): D
        Resampling: Permuting D
        Clust. var(s): __00000J
        Clusters: 40
        Strata var(s): none
        Strata: 1

```

T		T(obs)	c	n	p=c/n	SE(p)	[95% Conf. Interval]	
-----+-----								
Ys		10.7	4996	10000	0.4996	0.0050	.4897515	.5094487

Note: Confidence interval is with respect to p=c/n.

Note:  $c = \#\{ |T| \geq |T(\text{obs})| \}$

```
In [16]: preserve
         // CI and Hist
         use "ate_null.dta", clear

         sort Ys
         global ylower= round(Ys[251], .01)

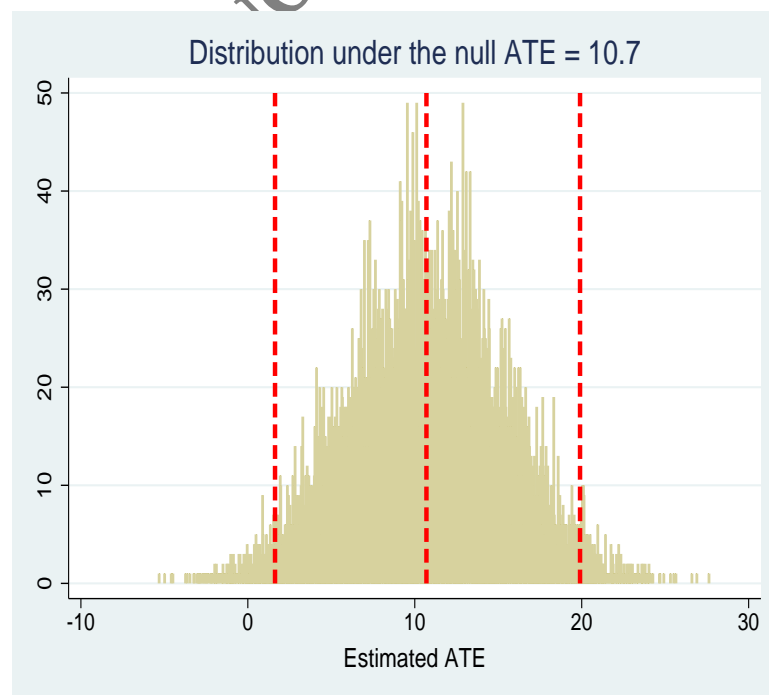
         global yupper= round(Ys[9750], .01)
         di "95%CI: ($ylower, $yupper)"

         graph twoway (histogram Ys,frequency bin(1000))
         (scatteri 0 $tau 50 $tau, c(1) lc(red) lw(thick) lp(dash) m(i))
         (scatteri 0 $ylower 50 $ylower, c(1) lc(red) lw(thick) lp(dash) m(i))
         (scatteri 0 $yupper 50 $yupper, c(1) lc(red) lw(thick) lp(dash) m(i)),
         legend(off)
         b1title("Estimated ATE")
         title("Distribution under the null ATE = 10.7") ///
         xtitle("")

         graph export ./exercise_4_4_f_graph.pdf

         restore
```

95%CI: (1.53, 19.83)



The confidence interval stretches from [1.53, 19.83] implying that the ATE is positive but its location is subject to a great deal of statistical uncertainty. Our best guess is 10.7, but the interval ranges from a small positive value to a truly massive effect.

- g) Use the estimated ATE in part (c) to construct a full schedule of potential outcomes for all schools, assuming that every school has the same treatment effect. Using this simulated schedule of potential outcomes, simulate the 95% confidence interval for the sample average treatment effect estimated by a regression of  $Y_i$  on  $D_i$  and  $X_i$ . Interpret the results. Is this confidence interval narrower than one you generated in question (f)?

```
In [17]: 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 - $tau_c if D==1
        qui replace Y1_sim = y + $tau_c if D==0

In [18]: capture program drop ate_null_cov
        program define ate_null_cov, rclass
            replace Y_sim = Y0_sim*(1-D) + Y1_sim*(D)
            regress Y_sim D x
            return scalar Ys_cov=_b[D]
        end

        ritest D Ys_cov = r(Ys_cov), reps(10000)
        sav(ate_null_cov.dta, replace) nodots: ///
        ate_null_cov
```

Source	SS	df	MS	Number of obs	=	40
Model	8627.78597	2	4313.89298	F(2, 37)	=	168.97
Residual	944.614033	37	25.530109	Prob > F	=	0.0000
Total	9572.4	39	245.446154	R-squared	=	0.9013
				Adj R-squared	=	0.8960
				Root MSE	=	5.0527

Y_sim	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D	5.315536	1.628474	3.26	0.002	2.015935	8.615138
x	.9204212	.0537624	17.12	0.000	.8114883	1.029354
_cons	1.216271	1.875729	0.65	0.521	-2.584317	5.016858

```
command: ate_null_cov
Ys_cov: r(Ys_cov)
res. var(s): D
Resampling: Permuting D
```

```

Clust. var(s):  __00000M
      Clusters:  40
Strata var(s):  none
      Strata:    1

```

```

-----
T          |      T(obs)      c      n      p=c/n      SE(p) [95% Conf. Interval]
-----+-----
      Ys_cov |      5.315536      4961      10000      0.4961      0.0050      .4862528      .5059495
-----

```

Note: Confidence interval is with respect to  $p=c/n$ .

Note:  $c = \#\{|T| \geq |T(\text{obs})|\}$

```

In [19]: preserve
         // CI and Hist
         use "ate_null_cov.dta", clear

         sort Ys_cov
         global yclower= round(Ys_cov[251], .001)
         global ycupper= round(Ys_cov[9750], .001)

         di "95%CI: ($yclower, $ycupper)"

         graph twoway (histogram Ys_cov,frequency bin(1000))
         (scatteri 0 $tau_c 40 $tau_c, c(1) lc(red) lw(thick) lp(dash) m(i))
         (scatteri 0 $yclower 40 $yclower, c(1) lc(red) lw(thick) lp(dash) m(i))
         (scatteri 0 $ycupper 40 $ycupper, c(1) lc(red) lw(thick) lp(dash) m(i)),
         legend(off)
         b1title("Estimated ATE controlling covariance")
         title("Distribution under the null ATE = 5.3")
         xtitle("")

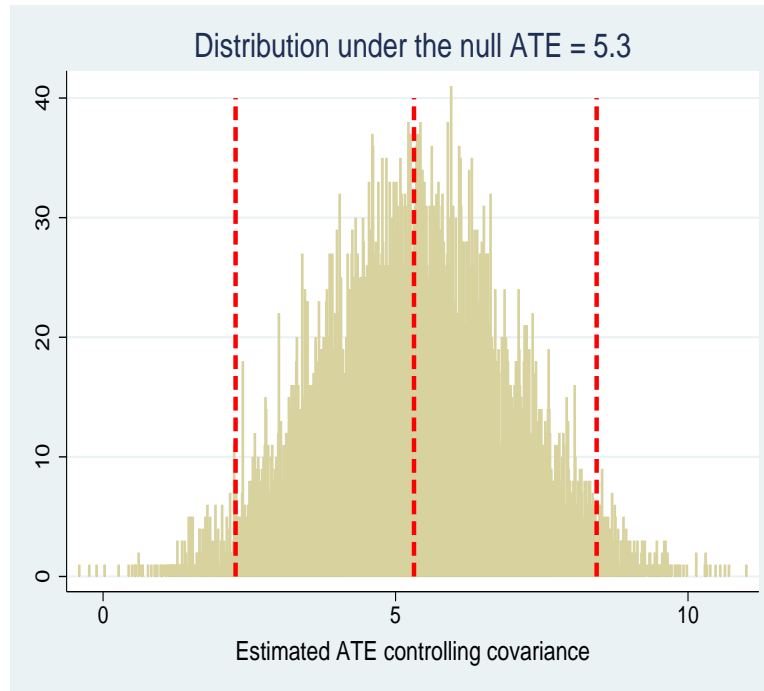
         graph export ./exercise_4_4_g_graph.pdf

         restore

```

95%CI: (2.213, 8.460)





The confidence interval now stretches from [2.22, 8.46]. Interestingly, this interval no longer contains the estimate obtained without controls for covariates. Our best guess is now 5.3, and our 95% interval is now roughly one-third as wide as before.

## 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
        set seed 1234567
        set more off
```

```

In [2]:
// loop to simulate 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]: 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  $4 \times 10^{-4}$  in the treatment condition and  $6 \times 10^{-4}$  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 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
        global ate_complete_RA = _b[D]
        di "ATE under Complete Assignment = " $ate_complete_RA

ATE under Complete Assignment = 10.7

In [11]: // RI under the null ate=ate_complete_RA
        cap drop YO_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

```

```

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:   $\theta(D==0) \leq \theta(D==1)$ ]
p=0.50995 [one-tailed test of Ho:   $\theta(D==0) \geq \theta(D==1)$ ]
p=0.48995 [two-tailed test of Ho:   $\theta(D==0) == \theta(D==1)$ ]

```

Saving log file to distout\_complete\_RA.dta...done.

```

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 YO_sim Y1_sim Y_sim
qui gen YO_sim = Y
qui gen Y1_sim = Y
qui gen Y_sim = .
qui replace YO_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 = YO_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 YO_sim Y1_sim Y_sim
qui gen YO_sim = Y
qui gen Y1_sim = Y
qui gen Y_sim = .
qui replace YO_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 = YO_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 3: 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

One way to practice your experimental design skills is to undertake a mock randomization of an existing non-experimental dataset. In this exercise, the existing dataset is treated as though it were a baseline data collection effort that an experimental researcher gathered in preparation for a random intervention. The actual data in question come from a panel study of Russian villagers. Villagers from randomly selected rural areas of Russia were interviewed in 1995 and re-interviewed in 1996 and 1997. Our attention focuses on the 462 respondents who were interviewed in all three waves and provided answers to questions about their income, church membership, and evaluation of national conditions (i.e., how well are things going in Russia?). Imagine that an experimental intervention occurred after the 1996 survey and that national evaluations in the 1997 survey were the experimental outcome of interest. The dataset provided at [isps.research.yale.edu/FEDAI](https://isps.research.yale.edu/FEDAI) contains the following pre-treatment covariates that may be used for blocking: sex, church membership, social class, and evaluations of national conditions in 1995 and 1996. As you design your experiment, imagine that “post-intervention” evaluations of national conditions in 1997 were unknown. [10pts]

- a) One way to develop a sense of which variables are likely to predict post-intervention evaluations of national conditions in 1997 is to regress evaluations of national conditions in 1996 on sex, church membership, social class, and evaluations in 1995. Which of these variables seem to most strongly predict evaluations of national conditions in 1996? What is the  $R^2$  from this regression?

```
In [1]: clear
        qui import delim ./data/chapter04/OBrien_Patsiorkovski_ICPSR_1999,clear
        qui recode sexresp6 ( 2 = 1) (1 = 0), gen(female)
        qui recode memberc6 ( 1 = 1) (else = 0), gen(church_member)

In [2]: // fit1
        qui reg index96 index95 i.female i.church_member i.group6
        disp %8.4f e(r2)

0.3937
```

```
In [3]: // fit2
        qui reg index96 i.female i.church_member i.group6
        disp %8.5f e(r2)
```

0.02828

The regression treats “index95” as a continuous variable and all others as categorical. the R-squared is 0.394, which implies that the regressors predict about 40% of the variance in “index96”. The strongest predictor is 95, the lagged dependent variable. Had we omitted this variable from the model, the R-squared would have fallen to 0.028.

- b) Suppose you were to design a block random assignment in order to predict evaluations in 1997. Use the R package `blockTools` (for example code, see [isps.research.yale.edu/FEDAI](https://isps.research.yale.edu/FEDAI)) to perform a block random assignment, blocking on sex, church membership, social class, and evaluations in 1996. Decide for yourself how many subjects to include in each block. Compare the treatment and control groups to verify that blocking produced groups that have the same profile of sex, church membership, social class, and evaluations in 1996.

```
In [4]: //ssc install randtreat
        //(install the package for random assignment)

        qui randtreat, generate(treat) ///
        strata(female church_member index96 group6) misfits(wglobal)
```

```
In [5]: regress treat female church_member i.group6 index96
```

Source	SS	df	MS	Number of obs	=	462
Model	.042679933	6	.007113322	F(6, 455)	=	0.03
Residual	115.45732	455	.253752352	Prob > F	=	0.9999
Total	115.5	461	.250542299	R-squared	=	0.0004
				Adj R-squared	=	-0.0128
				Root MSE	=	.50374

	treat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
female		-.0075749	.0552815	-0.14	0.891	-.1162136 .1010638
church_member		.0134376	.0817447	0.16	0.870	-.1472064 .1740816
group6						
2		-.0484915	.1590765	-0.30	0.761	-.3611072 .2641243
3		-.0451665	.1563562	-0.29	0.773	-.3524364 .2621034
4		-.0456885	.2234276	-0.20	0.838	-.4847664 .3933895
index96		-.001982	.0129813	-0.15	0.879	-.0274928 .0235288
_cons		.5608169	.1703183	3.29	0.001	.2261088 .895525

Using the package `randtreat`, we created blocks of size 2 based on gender, church membership, evaluations in 1996, and social class. The package also conducts complete random assignment – with some work, this assignment can be extracted. Regressing this treatment assignment on the set of pretreatment covariates reveals that the groups are well balanced.

- c) Suppose you wanted to assess how well your blocking design performed in terms of increasing the precision with which treatment effects are estimated. Of course, there was no actual treatment in this case, but imagine that shortly after the survey in 1996, a treatment were administered to a randomly selected treatment group. (Here is an instance in which the sharp null hypothesis of no effect is known to be true!) The outcome from this imaginary experiment is evaluations of national conditions in 1997. Compare the sampling distribution of the estimated treatment effect (which should be centered on zero) under balanced complete random assignment to the sampling distribution of the estimated treatment effect under block random assignment.

Answer:

See below

- d) Calculate the sampling distribution of the estimated treatment effect under balanced complete random assignment using regression to control for the variables that would have otherwise been used to form blocks. Compare the resulting distribution to the sampling distribution of the estimated treatment effect under block random assignment. Does blocking produce an appreciable gain in precision over what is achieved by covariate adjustment?

```
In [6]: set seed 1234567
        // complete assignment
        clear*

        capture program drop com_ra
        program define com_ra, rclass
        import delim ./data/chapter04/OBrien_Patsiorkovski_ICPSR_1999,clear
        capture drop complete
        randtreat, generate(complete)
        regress index97 complete
        return scalar coef = _b[complete]
        exit
        end

        qui simulate coef=r(coef), reps(10000): com_ra
        qui sum
        scalar complete_est = r(mean)
        scalar complete_sd = r(sd)

        di "Average Estimate (complete)" %8.2f complete_est
        di "Standard Error (complete)" %8.2f complete_sd
```

```
Average Estimate (complete)    0.00
```

```
Standard Error (complete)      0.17
```

```

In [7]: // adjusted
        capture program drop adj_ra
        program define adj_ra, rclass
        import delim ./data/chapter04/OBrien_Patsiorkovski_ICPSR_1999,clear
        capture drop complete
        randtreat, generate(complete)
        regress index97 complete sexresp6 memberc6 i.group6 index96
        return scalar coef = _b[complete]
        exit
        end

        qui simulate coef=r(coef), reps(10000): adj_ra
        qui sum
        scalar adjusted_est = r(mean)
        scalar adjusted_sd = r(sd)

        di "Average Estimate (adjusted)"%8.2f adjusted_est
        di "Standard Error (adjusted)"%8.2f adjusted_sd

Average Estimate (adjusted)    -0.00

Standard Error (adjusted)      0.13

```

```

In [8]: // blocked
        capture program drop block_ra
        program define block_ra, rclass
        import delim ./data/chapter04/OBrien_Patsiorkovski_ICPSR_1999,clear
        capture drop block
        randtreat, generate(block) ///
        strata(sexresp6 memberc6 group6 index96) misfits(wglobal)
        regress index97 block
        return scalar coef = _b[block]
        exit
        end

        qui simulate coef=r(coef), reps(10000): block_ra
        qui sum
        scalar block_est = r(mean)
        scalar block_sd = r(sd)

        di "Average Estimate (blocked)"%8.2f block_est
        di "Standard Error (blocked)"%8.2f block_sd

Average Estimate (blocked)      0.00

```

Standard Error (blocked)      0.12

```
In [9]: // pool results together
        disp %8.2f complete_est %8.2f adjusted_est %8.2f block_est
        disp %8.2f complete_sd %8.2f adjusted_sd %8.2f block_sd
```

Table 4: Comparison of 3 estimators

	complete	adjusted	blocked
Average Estimate	0.00	-0.00	0.00
Standard Error	0.17	0.13	0.12

The table above shows a comparison of three estimators of the ATE: difference-in-means under complete random assignment, OLS with covariate adjustment under complete random assignment, and difference-in-means under blocked random assignment. All three estimators are centered on the true ATE of zero. The least precise method is complete random assignment with the difference-in-means estimator, which produces a standard error of 0.17. The most precise approach is blocking, which produces a standard error of 0.12. Slightly inferior to blocking is covariate adjustment, which produces a standard error of 0.13. Blocking's slight superiority stems from the fact that, under blocking, there is no incidental correlation between the covariates and random assignments and therefore no "collinearity penalty."

## 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 5: Question 7 Table

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

- 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

Table 6: 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

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 7: 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

Sometimes researchers randomly assign subjects from lists that are later discovered to have duplicate entries. Suppose, for example, that a fund-raising experiment randomly assigns 500 of 1,000

names to a treatment that consists of an invitation to contribute to a charitable cause. However, it is later discovered that 600 names appear once and 200 names appear twice. Before the invitations are mailed, duplicate invitations are discarded, so that no one receives more than one invitation. [10pts]

- a) What is the probability of assignment to the treatment group among those whose names appeared once in the original list? What is the probability of assignment to the treatment group among those whose names appeared twice in the original list?

Answer:

The probability of being assigned to treatment if your name appears once is 0.5. The probability of being assigned to treatment if your name is a duplicate is  $0.5 + (0.5)(0.5) = 0.75$ , where the first term is the probability you were assigned to treatment the first time your name came up and the second term is the probability you were assigned to control the first time multiplied by the probability you were assigned to treatment the second time.

- b) Of the 800 unique names in the original list, how many would you expect to be assigned to treatment and control?

Answer:

Of the 600 unique names that appear once, 300 are, in expectation, allocated to treatment. Of the 200 unique names that appear twice, 150 are, in expectation, allocated to treatment. Thus, we expect a total of 450 unique names in the treatment group.

- c) What estimation procedure should one use in order to obtain unbiased estimates of the ATE?

Answer:

One should analyze the experiment as though it were randomized in two blocks: the names that appear once and the names that appear twice. Use an estimator like equation (4.11).

## 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.<sup>2</sup> 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

---

<sup>2</sup>Gerber and Green 2005.





```
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
Model	14.4110635	1	14.4110635	F(1, 940713)	=	57.98
Residual	233826.144	940,713	.248562679	Prob > F	=	0.0000
Total	233840.555	940,714	.248577734	R-squared	=	0.0001
				Adj R-squared	=	0.0001
				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

The 2003 Kansas City voter mobilization experiment described in Chapter 3 is a cluster randomized design in which 28 precincts comprising 9,712 voters were randomly assigned to treatment and control.<sup>3</sup> The study contains a wealth of covariates: the registrar recorded whether each voter participated in elections dating back to 1996. The dataset may be obtained at [isps.research.yale.edu/FEDAI](https://isps.research.yale.edu/FEDAI). [10pts]

- a) Test the balance of the treatment and control groups by looking at whether past turnout predicts treatment assignment. Regress treatment assignment on the entire set of past votes, and calculate the F-statistic. Use randomization inference to test the null hypothesis that none of the past turnout variables predict treatment assignment. Remember that to simulate the distribution of the F-statistic, you must generate 1,000 random cluster assignments and calculate the F-statistic for each simulated assignment. Judging from the p-value of this test, what does the F-statistic seem to suggest about whether subjects in the treatment and control groups have comparable background characteristics?

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

<sup>3</sup>Arceneaux 2005.

```

set maxvar 32767
set seed 1234567

qui import delim ./data/chapter04/Arceneaux_AAAPSSsubset_2005, clear
set seed 1234567

In [2]: rename treatment Z
        rename vote03 Y
        rename unit clust

        global cov v_p2003 v_m2003 v_g2002 v_p2002 v_m2002 ///
        v_s2001 v_g2000 v_p2000 v_m2000 v_s1999 ///
        v_m1999 v_g1998 v_m1998 v_s1998 v_m1997 ///
        v_s1997 v_g1996 v_p1996 v_m1996 v_s1996

In [3]: rittest Z Fstat=e(F), right clu(clust) reps(10000) nodots: reg Z $cov

        Fstat:  e(F)
        res. var(s):  Z
        Resampling:  Permuting Z
        Clust. var(s):  clust
        Clusters:  28
        Strata var(s):  none
        Strata:  1

-----
T          |      T(obs)      c      n      p=c/n      SE(p) [95% Conf. Interval]
-----+-----
          Fstat |    1.649235    9392   10000   0.9392   0.0024   .9343371   .9438051
-----+-----

Note: Confidence interval is with respect to p=c/n.
Note: c = #{T >= T(obs)}

In [4]: // one-tail p-value
        di %8.3f el(r(p),1,1)

0.939

```

Using randomization inference, we recover a  $p$ -value of 0.939; we therefore cannot reject the null hypothesis of random assignment.

- b) Regress turnout in 2003 (after the treatment was administered) on the experimental assignment and the full set of covariates. Interpret the estimated ATE. Use randomization inference to test the sharp null hypothesis that experimental assignment had no effect on any subject's decision to vote.

```
In [5]: // calculate the probabibility under cluster assignment
preserve
collapse Z, by(clust)
qui sum Z
global p = r(mean)
restore

// probability of being assigned to treatment
gen probs = $p
sum probs
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
probs	9,712	.5	0	.5	.5

```
In [6]: //ate
qui reg Y Z $cov
di "ATE= " _b[Z]
```

ATE= .05596196

```
In [7]: riteest Z ate_sim=_b[Z], ///
right clu(clust) reps(10000) nodots: ///
reg Y Z $cov
```

```
res. var(s):  Z
Resampling:  Permuting Z
Clust. var(s): clust
Clusters:    28
Strata var(s): none
Strata:      1
```

T		T(obs)	c	n	p=c/n	SE(p)	[95% Conf. Interval]
-----+-----							
ate_sim		.055962	13	10000	0.0013	0.0004	.0006924 .002222

Note: Confidence interval is with respect to  $p=c/n$ .

Note:  $c = \#\{T \geq T(\text{obs})\}$

```
In [8]: // ate
di %8.5f el(r(b),1,1)
```

0.05596

```
In [9]: // one-tail p-value
        di %8.4f el(r(p),1,1)

0.0013
```

The estimate of the treatment effect is 0.056, implying that treatment increased turnout by 5.6 percentage points. This finding is statistically significant. Under the sharp null, estimates as large or larger only occur 0.13% of the time.

- c) When analyzing cluster randomized experiments with clusters of varying size, one concern is that difference-in-means estimation is prone to bias. This concern also applies to regression. In order to sidestep this problem, researchers may choose to use the difference-in-totals estimator in equation (3.24) to estimate the ATE. Estimate the ATE using this estimator.

```
In [10]: qui gen weights = Z/probs + (1 - Z)/(1 - probs)
        qui tabstat Y [aw=weights], by(Z) stat(sum) save

        scalar ateHT = (el(r(Stat2),1,1) - el(r(Stat1),1,1))/_N
        di "ate_HT ="%8.5f ateHT

ate_HT = 0.05395
```

The difference-in-totals estimate of the treatment effect is that treatment increased turnout by 5.4 percentage points.

- d) Use randomization inference to test the sharp null hypothesis that treatment assignment had no effect, using the difference-in-totals estimator.

```
In [11]: cap program drop di_in_total
        program define di_in_total, rclass
        tabstat Y [aw=weights], by(Z) stat(sum) save
        return scalar ateHT_sim = (el(r(Stat2),1,1) - el(r(Stat1),1,1))/_N
        end

        rittest Z ateHT_sim=r(ateHT_sim), ///
        right clu(clust) reps(10000) nodots: di_in_total
```

```
Resampling: Permuting Z
Clust. var(s): clust
Clusters: 28
Strata var(s): none
Strata: 1
```

```
-----+-----
T          |      T(obs)      c      n      p=c/n      SE(p) [95% Conf. Interval]
-----+-----
```

ateHT_sim	.0539539	1954	10000	0.1954	0.0040	.1876681	.2033097
-----------	----------	------	-------	--------	--------	----------	----------

-----  
Note: Confidence interval is with respect to  $p=c/n$ .

Note:  $c = \#\{T \geq T(\text{obs})\}$

```
In [12]: // one-tail p-value
```

```
di %8.3f el(r(p),1,1)
```

```
0.195
```

Estimates generated under the sharp null equaled or exceeded the observed difference-in-totals 19.5% of the time, meaning we cannot reject the null.

- e) The difference-in-totals estimator can generate imprecise estimates, but its precision can be improved by incorporating information about covariates. Create a new outcome variable that is the difference between a subject's turnout (1 = vote, 0 = abstain) and the average rate of turnout in all past elections. Now, using this "differenced" outcome variable, estimate the ATE using the difference-in-totals estimator, and test the sharp null hypothesis of no effect.

```
In [13]: mkmat $cov , mat(cov_m)
```

```
matrix rowm = cov_m * J(colsof(cov_m), 1, 1/colsof(cov_m))
```

```
matrix colnames rowm=row_mean
```

```
svmat double rowm, names(col)
```

```
qui gen Y_diff = Y - row_mean
```

```
In [14]: cap program drop di_in_total2
```

```
program define di_in_total2, rclass
```

```
tabstat Y_diff [aw=weights], by(Z) stat(sum) save
```

```
return scalar ateHT_sim2 = (el(r(Stat2),1,1) - el(r(Stat1),1,1))/_N
```

```
end
```

```
ritest Z ateHT_sim=r(ateHT_sim2), ///
```

```
right clu(clust) reps(10000) nodots: di_in_total2
```

Resampling: Permuting Z

Clust. var(s): clust

Clusters: 28

Strata var(s): none

Strata: 1

T		T(obs)	c	n	p=c/n	SE(p)	[95% Conf. Interval]
-----+							
ateHT_sim		.0487438	110	10000	0.0110	0.0010	.0090491 .013243

-----  
Note: Confidence interval is with respect to  $p=c/n$ .  
Note:  $c = \#\{T \geq T(\text{obs})\}$

```
In [15]: // ateHT2  
         di %8.5f el(r(b),1,1)
```

0.04874

```
In [16]: // one-tail p-value  
         di %8.3f el(r(p),1,1)
```

0.011

Using the differenced outcome variable tightened our estimates – the  $p$ -value under the sharp null is now 0.011, meaning we can reject the sharp null of no effect for any unit.

DO NOT DISTRIBUTE