Green // Experimental Research // 2016

# Final Exam

1. Identification of Terms. Define and state the significance of the following terms or phrases. Be concise, and focus on key issues.
2. Treatment-by-covariate interaction

A treatment-by-covariate interaction investigates systematic variance in treatment effects as a function of covariates: in other words, the change in conditional average treatment effects between subgroups partitioned on the basis of this covariate. While treatment-by-covariate interactions are often of great interest to researchers, there are some concerns in estimating and interpreting these effects. In particular, these raise the issue of multiple comparisons, (which increase the probability of type I errors unless corrected for). In addition, researchers frequently want to apply a causal interpretation to interaction effects, but this is not appropriate when covariates are not randomly assigned.

1. Trimming bounds vs. extreme value bounds

Trimming bounds indicate bounds on the ATE *for always-reporters* under the additional assumption of monotonicity in reporting (that you might have if-treated reporters or if-untreated reporters, but not both). Trimming bounds are created by estimating the proportion of if-treated reporters **π** in the treatment group (or the proportion of if-untreated reporters in the control group, if reporting is higher in control) and then removing the highest or lowest **π** outcomes from treatment and calculating the difference-in-means under each scenario.

Extreme value bounds indicate bounds on the ATE for all subjects. Extreme value bounds indicate the largest and smallest estimates one would obtain if one were to substitute the largest or smallest possible outcomes in place of missing data in each experimental group.

1. Within-subjects design

In contrast to a between-subjects design, in which different subjects are randomly allocated to treatment and control groups, a within-subjects design tracks the same subject over time, and assigns certain periods to treatment or control. The fact that the same individual is followed as he/she/it travels from a treated to an untreated state (or vice versa) poses a potential threat to the non-interference assumption.

The non-interference assumption states that a subject’s potential outcomes are unaffected by the treatment assignment of any other subject, or Yi(d) = Yi(d). (pp. 253, FEDAI)

Non-interference may be jeopardized if there is “anticipation” (i.e., my potential outcomes in the present period are affected by my future treatment status) or by “carry-over” (i.e., my current treatment status affects my potential outcomes in future periods.)

In the context of a within-subjects design, random assignment’s role is fairly limited, as it merely ensures that the timing of the treatment is independent of potential outcomes. The main threat to inference is carry-over or anticipation. To some extent, these concerns may be allayed by designs that blind subjects (mitigating anticipation) or allow for a wash-out period (mitigating spillover).

1. Mediation

Mediation refers to the process by which causal effects are transmitted from a treatment to an outcome. For example, the relationship between consumption of limes and the occurrence of scurvy is mediated by the amount of vitamin C in the bloodstream. Mediation poses fundamental challenges to causal inference. Let M(d) be a potential outcome of the mediator when d=0 (control) or d=1 (treatment). When we speak of the effect of the (unmediated) direct effect of a treatment on an outcome (i.e., the effect of D, holding M(d) constant), we are in effect trying to estimate empirically a quantity such as:

E[Y(M(1),1) – Y(M(1),0)] or E[Y(M(0),1) – Y(M(0),0)],

but the second and third potential outcomes in the line above are fundamentally unobservable. We cannot observe, for example, how Y would respond if a subject were untreated and yet M took on the value that follows from the administration of the treatment. The same problem also applies to average mediation effects (i.e., the effect of varying M in response to treatment while holding D constant):

E[Y(M(1),1) – Y(M(0),1)] or E[Y(M(1),0) – Y(M(0),0)].

Again, the second and third terms are complex counterfactuals that are inherently unobservable. An important special case occurs when M(1) = M(0) for all observations: in that case, it follows from the above equation that the average causal mediation effect must be zero. Empirically, this result implies that we can rule out mediators when the sharp null hypothesis of no effect of the treatment on the mediator seems to hold.

An analysis NOT justified by randomized experimental design is a regression of Y on M and the treatment. Such a regression is prone to post-treatment bias because unobserved causes of M may be related to unobserved causes of Y. Not only is such a regression prone to bias, it also apportions the “total” effect of the treatment on the outcome into “direct” and “indirect” components under the special case where the effects of the treatment and the mediator are constant across subjects.

1. Meta-analysis

Meta-analysis refers to statistical procedures designed to summarize the results of research literatures. Meta-analysis is sometimes described as a “systematic” method for constructing a literature review because it summarizes research findings based on a replicable formula. Specifically, when meta-analysis is used to pool several studies, each study’s experimental result is weighted according to a formula that follows from an underlying statistical model. In chapter 11, the model involves random sampling from a large population, and the formula (fixed effects meta analysis) weights each study to the inverse of its precision, or squared standard error. This procedure is superior to a count of studies that show significant or insignificant results because the latter potentially accords too much weight to small studies that produce statistically insignificant results and too little weight to large studies that convincingly demonstrate an effect when other, smaller studies fail to do so. Another advantage of meta-analysis over this head-count method is that meta-analysis generates a point estimate and confidence interval, which is more informative than a summary statement about statistical significance.

1. Short answer
2. Briefly summarize the implications of clustered random assignment for experimental design and analysis. What complications arise when clusters (e.g., media markets) contain different numbers of subjects?

**Implications for analysis:**

If cluster sizes are equal, then difference-in-means is an unbiased estimator of the ATE. If cluster sizes are unequal, then the difference-in-means estimator is biased because the denominator is now a random variable, and the ratio of an expectation is not equal to the expectation of a ratio. If the number of clusters is relatively small, use difference-in-totals instead (Eq, 3.24, pp. 83). Better yet, in the design phase, block on cluster size before random assignment.

We cannot use standard methods to estimate uncertainty: since the effective N is smaller (due to randomization at the cluster level), the sampling variability typically increases in the variability of the cluster means (pp. 82).

The true SE of the estimated ATE, assuming fixed cluster means, is described by Equation 3.22 (pp. 82). This quantity cannot be identified because we cannot observe the covariance term, so we estimate standard errors using Equation 3.23 (pp. 83).

We can form confidence intervals by creating a schedule of potential outcomes under the assumption of constant treatment effects (τi = ATEhat); we also apply a degrees of freedom adjustment that expands the width of the interval by the square root of [(k-1)/(k‐2)] where k is the number of clusters (pp. 83, including footnote 20).

**Implications for design (based on Eq 3.22 and extensions to the discussion in pp. 57-59):**We can decrease the SE (i.e. shrink the sampling distribution) by:

(1) increasing the number of clusters;

(2) placing greater number of clusters into the treatment group where cluster-level means of potential outcomes associated with that group have a higher  variance, but if unknown use a balanced design;

(3) examine treatments that minimize (or have negative) covariance between  average treated potential outcomes and average untreated potential  outcomes at the cluster level;

(4) minimize the cluster level mean treated potential outcome variance and the cluster level mean untreated potential outcome variance.

(5) block on cluster size to avoid the issue of bias in difference-in-means

1. Explain (preferably using a bit of algebra) why rejecting the null hypothesis that implies rejection of the null hypothesis of homogeneous treatment effects (i.e., the hypothesis that )).

See pp. 293, FEDAI.

Var(Yi(1))

=  Var(Yi(0) + τi)

=  Var(Yi(0)) + Var(τi) + 2 Cov( Yi(0), τi ) and the equality Var(Yi(1)) = Var(Yi(0)) holds when  Var(τi) = -2 Cov ( Yi(0), τi )

Under the null hypothesis that τi is constant across subjects, both sides of Equation (9.3) are zero, since the covariance between a variance and a constant is zero. Thus rejecting the hypothesis Var(Yi(1)) = Var(Yi(0)) means rejecting the null hypothesis that Var(τi) = 0.

1. Modeling and data analysis

The table below shows the results of an experiment in which 630,640 subjects were randomly sent a “social pressure” mailing immediately prior to an election in June of 2012. Social pressure mailings showed voters whether they and their neighbors voted in the last election. The remaining 33,380 subjects were sent nothing. Turnout in that election is indicated by the variable votedS. This variable equals 1 when a subject voted; 0 otherwise. Later that year, a presidential election occurred, and subjects voted or abstained (see the variable votedG).

Suppose you sought to estimate the “downstream” effect of votedS on votedG.

1. Briefly explain why the identification of “downstream” effects is akin to the identification of the CACE in the presence of two-sided noncompliance.
2. Show algebraically how one can identify the average causal effect among those who vote in the June election if and only if they are encouraged by the mailer. Indicate what assumptions you invoke in the course of your identification proof.
3. Explain and critically evaluate the excludability assumption in this particular study.
4. With a hand calculator (or a calculator on your cell phone), use the results below to estimate this average causal effect. (Don’t worry about estimating standard errors.)

Subjects assigned to the control group

| votedS

votedG | 0 1 | Total

-----------+----------------------+----------

abstained | 7,990 1,275 | 9,265

| 69.96 5.81 | 27.76

-----------+----------------------+----------

voted | 3,431 20,684 | 24,115

| 30.04 94.19 | 72.24

-----------+----------------------+----------

Total | 11,421 21,959 | 33,380

Subjects assigned to the treatment (mail) group

| votedS

votedG | 0 1 | Total

-----------+----------------------+----------

abstained | 147,147 24,721 | 171,868

| 70.46 5.86 | 27.25

-----------+----------------------+----------

voted | 61,691 397,081 | 458,772

| 29.54 94.14 | 72.75

-----------+----------------------+----------

Total | 208,838 421,802 | 630,640

The downstream effect of **votedS** on **votedG** may be identified by instrumenting for **votedS** using random assignment to the mailer.

Let Z = random assignment to mailer D = votedS = voted in spring election Y = votedG = voted in general election





Critically evaluate the exclusion restriction assumption: The assumption would be violated if the mailer affected voting in the general election via some causal pathway other than voting in the spring. This is likely, for instance, if the mailer were so memorable that people were influenced by it when deciding whether to vote in November.

1. Interpreting results

Researchers in Uganda recently conducted a study in which popular movies were shown free of charge in 56 rural trading centers during a four-week film festival. In 24 randomly selected trading centers, messages encouraging empathy for women who suffer medical complications arising from (illegal) abortions were aired during commercial breaks, whereas in the remaining 32 trading centers, messages about other topics were aired. Outcomes are measured at the level of the individual villager. (For purposes of answering the questions below, you can ignore the complications associated with clustered assignment in this study.) Outcome measures were collected six weeks later via an ostensibly unrelated survey that covered a variety of topics, including abortion. For our purposes, the outcome is scored 1 if the respondent expresses willingness to help a woman ostracized on account of abortion and 0 otherwise. (In the tables below, these outcomes are labeled “help” and “no help,” respectively.) Compliance was measured at the end of the end-line survey; subjects were asked whether they attended any of the films and, if not, whether they knew others who had attended. Compliers are those who attended; partial compliers are those who know others who attended; and never-takers are those who neither attended nor knew others who attended.

* 1. Define the relevant potential outcomes: exposure to the abortion message, knowing others who were exposed to the abortion message, and not being exposed directly or indirectly to this message.

Potential outcomes might be defined as follows (notation may of course vary):

Y(d=1) = feelings towards women who have experienced abortion given exposure to the abortion message directly

Y(d= .5) = feelings towards women who have experienced abortion given exposure to the abortion message indirectly via others who have seen the message

Y(d = 0) = feelings towards women who have experienced abortion given no exposure to the abortion message either directly or indirectly

Note that we might also imagine a fourth potential outcome of Y(d=1.5), reflecting feelings towards abortion given exposure to the abortion message both directly and indirectly via others who have seen the message. However, given the design approach in this study, it is not possible to distinguish between this potential outcome and Y(d=1).

Also note that due to one-sided noncompliance present in this study, some subjects do not have the potential to reveal all potential outcomes (eg. Never-takers will always reveal Y(d=0), regardless of treatment assignment).

* 1. Define average treatment effects of interest in this study. Explain whether these estimands are identified given the design and whatever assumptions you see fit to invoke. Be sure to make your assumptions explicit.

We might be interested in identifying both the direct effects of the treatment (Y(d=1) – Y(d=0)) and indirect effects (Y(d=.5) – Y(d=0)). We CAN identify both of these quantities for two nonrandom subsets of subjects: compliers and partial compliers, respectively. Because this is a placebo design, partial and full compliers are identified in both groups and we can thus compare outcomes directly between units in treatment and control. Note that this procedure recovers the CACE for non-random groups of the population – compliers are the types to attend a free film festival and partial compliers are the types to not attend themselves but know people who would.

We might also be interested in the intent-to-treat effect across the population in order to understand the effectiveness of the program on the study population, which is a weighted average of direct, indirect, and null treatment effects taking into account the proportions of different complier types and never takers.

Assumptions: The usual assumptions apply regarding randomization/probabilities of assignment. Noninterference in this case implies that the treatment status of one trading center does not affect individuals in another trading center. There’s an additional assumption in condensing potential outcomes that there is no distinction between Y(1.5) and Y(1) – ie no added effect of knowing people who attend screenings if an individual himself observes the screenings. Excludability requires that the only effect on potential outcomes occurs through watching the ads, and that there were no differences between placebo and treatment film festivals aside from the ads.

* 1. The table below shows the results for each of the three compliance strata. Estimate the causal estimands you identified in part (b), and interpret the results.

E[Y(1)-Y(0)|C] = 83.52 – 72 = 11.52 percentage points more likely to express sympathy for women who need an abortion

E[Y(0.5) – Y(0)|PC] = 73.6 – 68.18 = 5 percentage points more likely to express sympathy for women who need an abortion

E[Y(1)-Y(0)|NT] = 74.62 – 74.66 = -.04 (Not an estimand of interest but a good check to verify that never-takers in the placebo group are equivalent to those in treatment/that the exclusion restriction holds)

Group 1: Never-takers

| Z

Y | Control ABO Treat | Total

-----------+----------------------+----------

No Help | 112 84 | 196

| 25.34 25.38 | 25.36

-----------+----------------------+----------

Help | 330 247 | 577

| 74.66 74.62 | 74.64

-----------+----------------------+----------

Total | 442 331 | 773

| 100.00 100.00 | 100.00

-----------------------------------------------------------------------

Group 2: Partial Compliers

| Z

Y | Control ABO Treat | Total

-----------+----------------------+----------

No Help | 238 141 | 379

| 31.82 26.40 | 29.56

-----------+----------------------+----------

Help | 510 393 | 903

| 68.18 73.60 | 70.44

-----------+----------------------+----------

Total | 748 534 | 1,282

| 100.00 100.00 | 100.00

-----------------------------------------------------------------------

Group 3: Compliers

| Z

Y | Control ABO Treat | Total

-----------+----------------------+----------

No Help | 56 29 | 85

| 28.00 16.48 | 22.61

-----------+----------------------+----------

Help | 144 147 | 291

| 72.00 83.52 | 77.39

-----------+----------------------+----------

Total | 200 176 | 376

| 100.00 100.00 | 100.00