Investigator-Designed Randomized Experiments

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Investigator-Designed Randomized Experiments

1. Introduction
2. The Potential Outcomes Framework Revisited
3. Designing a 2-Group Randomized Experiment
4. An Example of a 2-Group Experiment
5. Analyzing Data from Randomized Experiments

Introduction

The School Voucher Experiment

- In February 1997, the privately-funded School Choice Scholarships Foundation (SCSF) announced that it would provide 1300 public elementary school children from low-income families with vouchers worth up to $1400 toward tuition at private elementary schools.
- There were more than 10,000 applications in a 3 month period.
- In May, 1997, the SCSF held a lottery to determine which children received scholarship vouchers.
- Besides enhancing the perception of fairness, the randomized lottery also provided a rare opportunity to produce a completely randomized experiment to investigate the causal effect of a *scholarship offer*.

The Potential Outcomes Framework Revisited

Potential Outcomes

- Consider the ith child in the experiment.
- Prior to the randomization, each child has two potential outcomes.
- $Y_i(1)$ is the outcome if offered the scholarship/voucher.
- $Y_i(0)$ is the outcome if not offered the scholarship/voucher.
- Note that, while prior to the randomization either of these outcomes is possible, after the lottery/randomization, it is only possible to observe one of the two outcomes for each individual.

The Potential Outcomes Framework Revisited

Individual Treatment Effects

- If we did somehow have access to $Y_i(1)$ and $Y_i(0)$ for each individual then the Individual Treatment Effect (ITE) for the ith individual could be calculated as

$$ITE_i = Y_i(1) - Y_i(0) \tag{1}$$

The Potential Outcomes Framework Revisited

Average Treatment Effect

- Of particular interest would be the Average Treatment Effect (ATE) across all the children in the population.
- Using expected value notation, we write

$$ATE = E\left(Y_i(1) - Y_i(0)\right) \tag{2}$$

- Unfortunately, we cannot estimate this directly from observed values of both potential outcomes, because each child has only one of the two potential outcomes.
- However, under certain assumptions formalized by Rubin and others, one can *estimate* the ATE with an unbiased estimator in a truly randomized experiment.

The Potential Outcomes Framework Revisited

Average Treatment Effect

- In a properly designed 2-group randomized design, the estimated ATE turns out to be simply the difference between the experimental and control group means.
- Specifically,

$$\widehat{ATE} = \overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 0} \tag{3}$$

where $\overline{Y}_{\bullet 1}$ is the sample mean for those receiving a scholarship offer, and $\overline{Y}_{\bullet 0}$ is the sample mean for those not receiving an offer.

The Potential Outcomes Framework Revisited

The Stable Unit Treatment Value Assumption (SUTVA)

- This key assumption states that the value of Y for unit u exposed to treatment t will be the same no matter what mechanism is used to assign treatment t to unit u, and no matter what treatments the other units receive.

The Potential Outcomes Framework Revisited

Violating the SUTVA

- Morgan and Winship (2007, p. 37–38) give an example of how the SUTVA can be violated via what they call a "treatment effect dilution."
- In this situation, the more units (i.e., subjects) assigned to a treatment, the less effective the treatment.
- In their table on the next slide, we see a set of treatment patterns for a highly stylized $n = 3$ experiment.
- Next to each of the first 3 treatment assignment patterns is the potential outcome pair for each unit receiving the treatment.
- Note that the treatment effect is $+2$ for each unit receiving the treatment.
- In the second row grouping of three treament assignment patterns, note that the individual treatment effects are all reduced to $+1$. Because in these groupings, 2 units are assigned to the treatment condition, it appears that assigning more units to the treatment has reduced the effect.

The Potential Outcomes Framework Revisited

Violating the SUTVA

Table 2.2: A Hypothetical Example in Which SUTVA is Violated

Treatment assignment patterns	Potential outcomes	
$\begin{bmatrix} d_1 = 1 \\ d_2 = 0 \\ d_3 = 0 \end{bmatrix}$ or $\begin{bmatrix} d_1 = 0 \\ d_2 = 1 \\ d_3 = 0 \end{bmatrix}$ or $\begin{bmatrix} d_1 = 0 \\ d_2 = 0 \\ d_3 = 1 \end{bmatrix}$	$y_1^1 = 3$ $y_2^1 = 3$ $y_3^1 = 3$	$y_1^0 = 1$ $y_2^0 = 1$ $y_3^0 = 1$
$\begin{bmatrix} d_1 = 1 \\ d_2 = 1 \\ d_3 = 0 \end{bmatrix}$ or $\begin{bmatrix} d_1 = 0 \\ d_2 = 1 \\ d_3 = 1 \end{bmatrix}$ or $\begin{bmatrix} d_1 = 1 \\ d_2 = 0 \\ d_3 = 1 \end{bmatrix}$	$y_1^1 = 2$ $y_2^1 = 2$ $y_3^1 = 2$	$y_1^0 = 1$ $y_2^0 = 1$ $y_3^0 = 1$

The Potential Outcomes Framework Revisited

Violating the SUTVA

Example (Effect of Catholic Schooling)

Suppose that a study attempted to assess the impact on learning of attending a Catholic parochial school vs. a public school. If the study became large, then the influx of a large number of public school students into the Catholic schools may disrupt "what is special" about the Catholic shools, and thereby cause a violation of the SUTVA.

The Potential Outcomes Framework Revisited

Violating the SUTVA

Example (Effect of Retraining)

Suppose a study sought to estimate the effects of labor-retraining programs on income. It might be that when a small-scale program is put in place in an area where there is a large market for a kind of laborer, the effect will be quite positive, while if a large-scale program is introduced into a smaller area, then the effect might be reduced.

Designing a 2-Group Randomized Experiment

- Murnane and Willett discuss the following steps in constructing a 2-group randomized experiment:
  1. Randomly sample subjects from a well-defined population.
  2. Randomly assign subjects to experimental conditions.
  3. A well-defined manipulation is implemented faithfully in the Treatment group, but not th e control group. All other conditions remain constant.
  4. A value on the dependent variable is measured identically for all participants.
  5. An estimate of the *ATE* is constructed as the mean difference between Treatment and Control conditions.

Designing a 2-Group Randomized Experiment

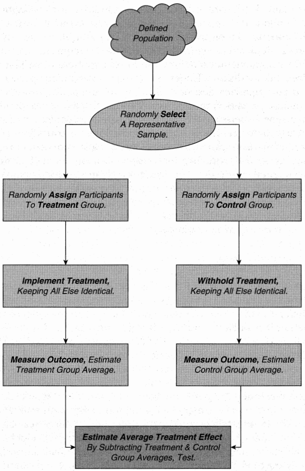

Figure 4.1 Conducting a two-group randomized experiment.

An Example of a 2-Group Experiment

The NYSP Study

- 11,105 children had applications submitted. *This is the target population.*
- 2260 children were chosen as subjects, and of these, 1300 (Treatment group) received vouchers worth up to $1400, and 960 were assigned to the Control group.

An Example of a 2-Group Experiment

The NYSP Study

Can you think of a way that the defined target population in this study might differ from the broader population of children from low-income families?

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

- In this very simple example, we can perform some very simple analyses.
- We can use this simple special case to demonstrate some important general principles that will serve us well in more complex designs.

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

- In this simple two-group experiment, we have the option of either performing a 2-sample t test (and associated confidence interval) or expressing the analysis in terms of an equivalent linear regression model.
- Both analyses are demonstrated by Murnane and Willett (pp. 48–60) on a subsample of African American children, of whom 291 were assigned to the Treatment group and 230 to the Control group.
- We'll replicate their analysis in R.

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

Table 4.1 Alternative analyses of the impact of voucher receipt (*VOUCHER*) on the third-grade academic achievement (*POST_ACH*) for a subsample of 521 African-American children randomly assigned to either a "voucher" treatment or a "no voucher" control group (n = 521)

Strategy #1: Two-Group t-Test

	Number of Observations	Sample Mean	Sample Standard Deviation	Standard Error
VOUCHER = 1	291	26.029	19.754	1.158
VOUCHER = 0	230	21.130	18.172	1.198
Difference		4.899		1.683
t-statistic		2.911		
df		519		
p-value		0.004		

Strategy #2: Linear Regression Analysis of POST_ACH on VOUCHER

Predictor	Parameter	Parameter Estimate	Standard Error	*t*-Statistic	*p*-value
INTERCEPT	β_0	21.130	1.258	16.80	0.000
VOUCHER	β_1	4.899	1.683	2.911	0.004
R^2 Statistic		0.016			
Residual Variance		19.072			

Strategy #3: Linear Regression Analysis of POST_ACH on VOUCHER, with PRE_ACH as Covariate

Predictor	Parameter	Parameter Estimate	Standard Error	*t*-Statistic	*p*-value
INTERCEPT	β_0	7.719	1.163	6.64	0.000
VOUCHER	β_1	4.098	1.269	3.23	0.001
PRE_ACH	γ	0.687	0.035	19.90	0.000
R^2 Statistic		0.442			
Residual Variance		14.373			

achievement tests prior to entering the NYSP experiment and at the end of their third year of participation. Of these, 291 were participants in the "voucher receipt" group and 230 in the "no voucher" group. Following the procedure adopted by Howell et al. (2002), we have averaged each child's national percentile scores on the reading and mathematics tests to obtain variables measuring composite academic achievement on entry into the study (which we refer to subsequently as covariate *PRE_ACH*) and after the third year of the experiment (which we refer to subsequently as outcome *POST_ACH*).

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

- Start by loading in the data:

```
> data <- read.csv("ch04.csv")
> attach(data)
```

- Next, we fit a simple linear regression model with the dichotomous *VOUCHER* variable as the predictor.
- We find that the coefficient attached to *VOUCHER* has an estimated value of 4.899 with an estimated standard error of 1.683.

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

```
> fit.1 <- lm(post_ach ~ voucher)
> summary(fit.1)

Call:
lm(formula = post_ach ~ voucher)

Residuals:
   Min     1Q Median     3Q    Max
-25.53 -15.03  -4.63  10.47  63.37

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   21.130      1.258  16.802  < 2e-16 ***
voucher        4.899      1.683   2.911  0.00375 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.07 on 519 degrees of freedom
Multiple R-squared:  0.01607,       Adjusted R-squared:  0.01417
F-statistic: 8.475 on 1 and 519 DF,  p-value: 0.003755
```

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

Give a brief verbal description of the meaning and interpretation of the values 4.899 and 1.683.

Analyzing Data from Randomized Experiments

NYSP African-American Children Subpopulation

- Murnane and Willet also examine a model in which a preachievement variable is added as a covariate.

```
> fit.2 <- lm(post_ach ~ voucher + pre_ach)
> summary(fit.2)

Call:
lm(formula = post_ach ~ voucher + pre_ach)

Residuals:
    Min      1Q  Median      3Q     Max
-47.337  -9.533  -2.124   7.973  59.781

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.71888    1.16298   6.637 8.08e-11 ***
voucher      4.09761    1.26873   3.230  0.00132 **
pre_ach      0.68731    0.03454  19.897  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.37 on 518 degrees of freedom
Multiple R-squared:  0.4423,	Adjusted R-squared:  0.4401
F-statistic: 205.4 on 2 and 518 DF,  p-value: < 2.2e-16
```