Chapter 6

Per Johansson and Donald B. Rubin

Yau Mathematical Sciences Center, Tsinghua University

June 30, 2021

Introduction

During the same period in which Fisher was developing the FEP approach, Neyman (1923, 1990) was focusing on methods for the estimation of, and inference for,average treatment effects, also using the distribution induced by randomization of the units in the experiment from a larger population of units.

At a general level, he was interested in the long-run operating characteristics of statistical procedures under both repeated sampling from the population and randomized assignment of treatments to the units in the sample.

Specifically, he attempted to find point estimators that were unbiased, and also interval estimators that had the specified nominal coverage in large samples.

As noted before, his focus on average effects was different from the focus of Fisher; the average effect across a population may be equal to zero even when some, or even

Introduction

Neyman's basic questions were the following.

- What would the average outcome be if all units were exposed to the active treatment, $\overline{Y}(1)$ in our notation?
- How did that compare to the average outcome if all units were exposed to the control treatment, $\overline{Y}(0)$ in our notation?
- Most importantly, what is the difference between these averages, the average treatment effect $\tau_{\text{FS}} = \overline{Y}(1) - \overline{Y}(0) = \sum_{i=1}^{N}(Y_i(1) - Y_i(0))/N$?

Neyman's approach was to develop an estimator of τ_{FS} and derive its mean and variance under repeated sampling.

His approach is similar to Fisher's, in that both consider the distribution of statistics (functions of the observed $\mathbf{W}$ and $\mathbf{Y}^{\text{obs}}$) under the randomization distribution, with all potential outcomes regarded as fixed. The similarity ends there.

Introduction

In Neyman's analysis, we do not start with an assumption that allows us to fill in all values of the missing potential outcomes

Neyman's primary concern was whether an estimator was unbiased for τ_{FS}.

A secondary goal was to construct an interval estimator for τ_{FS}, which he hoped to base on an unbiased estimator for the sampling variance of the average treatment effect estimator.

Confidence intervals, as they were called later by Neyman (1934), are stochastic intervals that are constructed in such a way that they include the true value of the estimand (here τ_{FS}) with probability, over repeated draws, at least equal to some fixed value, the confidence coefficient.

The Duflo-Hanna-Ryan Teacher Incentive Experiment

Use data from a randomized experiment conducted in rural India by Duflo, Hanna, and Ryan (2012). It was designed to study the effect of financial incentives on teacher performance, both measured directly by teacher absences, as well as by educational output measures, such as average class test scores.

A sample of 113 single-teacher schools was selected, and in a randomly selected subset of 57 schools, the salary structure was changed so that teachers were given a salary that was tied to their attendance over a month long period, whereas in the remaining 56 schools, the salary structure was unchanged.

In all schools, the teachers were given cameras with time stamps and asked to have students take pictures of the class with the teacher (beginning, and at the end of every school day). In addition, there were random unannounced visits to the schools by program officials to see whether the school was open or not.

The Duflo-Hanna-Ryan Teacher Incentive Experiment

6 schools with missing data are dropped, why $N = 107$ ($N_t = 53$ and $N_c = 54$) with recorded values on five key variables: four outcomes and one covariate.

Outcomes

- `open`: the proportion of times the school was open during a random visit.
- `pctpostwritten`: he percentage of students who completed a writing test
- `written`: writing test score averaged over all the students in each school who took the test. (In each class at least some students took the writing test at the end of the study.)
- `written_all`: average writing test score with zeros imputed for the students who did not take the test

One covariate, `pctprewritten`: the percentage of students who took the written test prior to the study.

The Duflo-Hanna-Ryan Teacher Incentive Experiment

Table 6.1: Summary Statistics for Duflo-Hanna-Ryan Teacher-Incentive Observed Data

	Variable	Control ($N_c = 54$) avg	(s.d.)	Treated ($N_t = 53$) avg	(s.d.)	min	max
pretreatment	pctprewritten	0.19	0.19	0.16	0.17	0.00	0.67
posttreatment	open	0.58	0.19	0.80	0.13	0.00	1.00
	pctpostwritten	0.47	0.19	0.52	0.23	0.05	0.92
	written	0.92	0.45	1.09	0.42	0.07	2.22
	written_all	0.46	0.32	0.60	0.39	0.04	1.43

Unbiased Estimation of the Average Treatment Effect

As before, for each unit in the population of N units, there exist two potential outcomes, $Y_i(0)$ and $Y_i(1)$, corresponding to the outcome under control and treatment respectively.

The only random component is $\mathbf{W}$, with ith element W_i, which by definition has a known distribution in a completely randomized experiment (CRE).

Neyman was interested in the population average treatment effect: $\tau_{\text{FS}} = \overline{Y}(1) - \overline{Y}(0)$

A natural estimator for the average treatment effect is:

$$\hat{\tau}^{\text{dif}} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{\text{obs}} - \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{\text{obs}}$$

$$= \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}},$$

Unbiased Estimation of the Average Treatment Effect

where

$$\overline{Y}_c^{\text{obs}} = \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{\text{obs}} \qquad \text{and} \quad \overline{Y}_t^{\text{obs}} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{\text{obs}}.$$

Theorem

The estimator $\hat{\tau}^{\text{dif}}$ is unbiased for τ_{S}.

PROOF OF THEOREM 1: Using the fact that $Y_i^{\text{obs}} = Y_i(1)$ if $W_i = 1$, and $Y_i^{\text{obs}} = Y_i(0)$ if $W_i = 0$, we can write the estimator $\hat{\tau}^{\text{dif}}$ as:

$$\hat{\tau}^{\text{dif}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{W_i \cdot Y_i(1)}{N_t/N} - \frac{(1-W_i) \cdot Y_i(0)}{N_c/N} \right).$$

Unbiased Estimation of the Average Treatment Effect

Because the potential outcomes as fixed, the only component in this statistic that is random is the treatment assignment, W_i.

Given the set up of a CRE, by Section 3.5,
$\Pr_W(W_i = 1|\mathbf{Y}(0), \mathbf{Y}(1)) = \mathbb{E}_W[W_i|\mathbf{Y}(0), \mathbf{Y}(1)] = N_t/N$. Thus, $\hat{\tau}^{\text{dif}}$ is unbiased for the average treatment effect τ_{FS}:

$$\mathbb{E}_W\left[\hat{\tau}^{\text{dif}}|\mathbf{Y}(0), \mathbf{Y}(1)\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbb{E}_W[W_i]\cdot Y_i(1)}{N_t/N} - \frac{\mathbb{E}_W[1-W_i])\cdot Y_i(0)}{N_c/N}\right)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i(1) - Y_i(0)\Big) = \tau_{\text{FS}}.$$

Unbiased Estimation of the Average Treatment Effect

Note that the estimator is unbiased, irrespective of the share of treated and control units in the randomized experiment. This does not imply, however, that this share is irrelevant for inference; it can greatly affect the precision of the estimator.

For the teacher-incentive experiment, taking the proportion of days that the school was open (`open`) as the outcome of interest, this estimator for the average effect is

$$\hat{\tau}^{\text{dif}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} = 0.80 - 0.58 = 0.22,$$

.

The Sampling Variance of the Neyman Estimator

Neyman was also interested in constructing interval estimates for the average treatment effect, which he later (Neyman, 1935a) termed confidence intervals.
Three steps, (1) derive the sampling variance of the estimator, (2) develop estimators for this sampling variance and (3) use central limit argument over the estimators distribution, and use (2) to create a large sample confidence interval of τ_{S}.
Here we focus on the (1). This derivation is relatively cumbersome because the assignments for different units are not independent in a CRE.

We start with two units and then expand to the general case.

The Sampling Variance of the Neyman Estimator With Two Units

The estimand, in this case is

$$\tau_{\text{FS}} = \frac{1}{2} \cdot \Big[(Y_1(1) - Y_1(0)) + (Y_2(1) - Y_2(0)) \Big]. \tag{1}$$

In a CRE, both units cannot receive the same treatment; it follows that $W_1 = 1 - W_2$.

The estimator is therefore:

$$\hat{\tau}^{\text{dif}} = W_1 \cdot \Big(Y_1^{\text{obs}} - Y_2^{\text{obs}} \Big) + (1 - W_1) \cdot \Big(Y_2^{\text{obs}} - Y_1^{\text{obs}} \Big).$$

If $W_1 = 1$, our estimate will be $\hat{\tau}^{\text{dif}} = Y_1^{\text{obs}} - Y_2^{\text{obs}} = Y_1(1) - Y_2(0)$. If on the other hand, $W_1 = 0$, the estimate will be $\hat{\tau}^{\text{dif}} = Y_2^{\text{obs}} - Y_1^{\text{obs}} = Y_2(1) - Y_1(0)$, so that we can also write:

$$\hat{\tau}^{\text{dif}} = W_1 \cdot \Big(Y_1(1) - Y_2(0) \Big) + (1 - W_1) \cdot \Big(Y_2(1) - Y_1(0) \Big).$$

The Sampling Variance of the Neyman Estimator With Two Units

Let $D = 2 \cdot W_1 - 1$, so that $D \in \{-1, 1\}$, $W_1 = (1 + D)/2$ and $W_2 = 1 - W_1 = (1 - D)/2$.

Because $\mathbb{E}_W[W_1] = 1/2$ is $\mathbb{E}_W[D] = 0$ and the variance is $\mathbb{V}_W(D) = \mathbb{E}_W[D^2] = 1$.

In terms of D and the potential outcomes, we can write the estimator $\hat{\tau}^{\text{dif}}$ as:

$$\hat{\tau}^{\text{dif}} = \frac{D+1}{2} \cdot \Big(Y_1(1) - Y_2(0)\Big) + \frac{1-D}{2} \cdot \Big(Y_2(1) - Y_1(0)\Big),$$

which can be rewritten as:

$$\hat{\tau}^{\text{dif}} = \frac{1}{2} \cdot \Big[\Big(Y_1(1) - Y_1(0)\Big) + \Big(Y_2(1) - Y_2(0)\Big)\Big] + \frac{D}{2} \cdot \Big[\Big(Y_1(1) + Y_1(0)\Big) - \Big(Y_2(1) + Y_2(0)\Big)\Big]$$

$$= \tau_{\text{FS}} + \frac{D}{2} \cdot \Big[\Big(Y_1(1) + Y_1(0)\Big) - \Big(Y_2(1) + Y_2(0)\Big)\Big].$$

The Sampling Variance of the Neyman Estimator With Two Units

Because $\mathbb{E}_W[D] = 0$, we can see immediately that $\hat{\tau}^{\text{dif}}$ is unbiased for τ_{FS}, .

However, the representation in terms of D also makes the calculation of its sampling variance straightforward:

$$\mathbb{V}_W(\hat{\tau}^{\text{dif}}) = \mathbb{V}_W\left(\tau_{\text{FS}} + \frac{D}{2} \cdot \left[\Big(Y_1(1) + Y_1(0)\Big) - \Big(Y_2(1) + Y_2(0)\Big)\right]\right)$$

$$= \frac{1}{4} \cdot \mathbb{V}_W(D) \cdot \left[\Big(Y_1(1) + Y_1(0)\Big) - \Big(Y_2(1) + Y_2(0)\Big)\right]^2,$$

The Sampling Variance of the Neyman Estimator With Two Units

Given that $\mathbb{V}_W(D) = 1$, it follows that the sampling variance of $\hat{\tau}^{\text{dif}}$ is equal to:

$$\mathbb{V}_W(\hat{\tau}^{\text{dif}}) = \frac{1}{4} \cdot \Big[\Big(Y_1(1) + Y_1(0)\Big) - \Big(Y_2(1) + Y_2(0)\Big)\Big]^2. \tag{2}$$

This representation of the sampling variance shows that this will be an awkward object to estimate, because it depends on all four potential outcomes, including products of the different potential outcomes for the same unit that are never jointly observed.

The Sampling Variance of of the Neyman Estimator with N Units

To calculate the sampling variance of $\hat{\tau}^{\text{dif}}$ we need the expectations of the second and cross moments of the treatment indicators W_i for $i = 1, \ldots, N$.

Because $W_i \in \{0, 1\}$, $W_i^2 = W_i$, and thus

$$\mathbb{E}_W\left[W_i^2\right] = \mathbb{E}_W\left[W_i\right] = \frac{N_t}{N}, \qquad \text{and} \quad \mathbb{V}_W(W_i) = \frac{N_t}{N} \cdot \left(1 - \frac{N_t}{N}\right).$$

With the number of treated units fixed at N_t, the two events—unit i being treated and unit i' being treated—are not independent.

Therefore

$$\mathbb{E}_W[W_i \cdot W_{i'}] = \Pr_W(W_i = 1) \cdot \Pr_W(W_{i'} = 1 | W_i = 1) = \frac{N_t}{N} \cdot \frac{N_t - 1}{N - 1}, \qquad \text{for} \quad i \neq j,$$

because conditional on $W_i = 1$ there are $N_t - 1$ treated units remaining, out of a total of $N - 1$ units remaining.

The Sampling Variance of of the Neyman Estimator with N Units

Theorem

The sampling variance of $\hat{\tau}^{\text{dif}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$ *is:*

$$\mathbb{V}_W\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}, \tag{3}$$

where S_c^2 *and* S_t^2 *are the variances of* $Y_i(0)$ *and* $Y_i(1)$ *in the sample, defined as:*

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(0) - \overline{Y}(0)\Big)^2, \qquad \text{and} \quad S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(1) - \overline{Y}(1)\Big)^2,$$

and S_{tc}^2 *is the sample variance of the unit-level treatment effects, defined as:*

$$S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(1) - Y_i(0) - (\overline{Y}(1) - \overline{Y}(0))\Big)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(1) - Y_i(0) - \tau_{\text{FS}}\Big)^2.$$

The Sampling Variance of of the Neyman Estimator with N Units

The first two elements follow by using standard results from the analysis of simple random samples: given a CRE, the N_t and N_c units provide a simple random sample of the N values of $Y_i(1)$ of $Y_i(0)$, respectively.

We estimate $\overline{Y}(1)$, by the average outcome for the N_t treated units, $\overline{Y}_t^{\text{obs}}$. This estimator is unbiased for $\overline{Y}(1)$.

The population variance of $Y_i(1)$ is $S_t^2 = \sum_i (Y_i(1) - \overline{Y}(1))^2/(N-1)$. The sampling variance for an average from a sample of size N_t is $S_t^2/N_t = \sum_i (Y_i(1) - \overline{Y}(1))^2/(N_t(N-1))$.

Similarly, the average outcome for the N_c units assigned to control, $\overline{Y}_c^{\text{obs}}$, is unbiased for $\overline{Y}(0)$, and its sampling variance is S_c^2/N_c.

The Sampling Variance of of the Neyman Estimator with N Units

S_{tc}^2/N, is the sample variance of the unit-level treatment effects, $Y_i(1) - Y_i(0)$.

If the treatment effect is constant in the population, $S_{tc}^2 = 0$, if not $S_{tc}^2 > 0$.

Because it is subtracted from the sum of the first two elements, the positive value for S_{tc}^2 reduces the sampling variance of this estimator for the average treatment effect.

Note that

$$S_{tc}^2 = S_c^2 + S_t^2 - 2 \cdot \rho_{tc} \cdot S_c \cdot S_t,$$

where

$$\rho_{tc} = \frac{1}{(N-1) \cdot S_c \cdot S_t} \sum_{i=1}^{N} \Big(Y_i(1) - \overline{Y}(1)\Big) \cdot \Big(Y_i(0)) - \overline{Y}(0)\Big). \tag{4}$$

The Sampling Variance of of the Neyman Estimator with N Units

By definition, ρ_{tc} is a correlation coefficient, and so lies in the interval $[-1, 1]$.

Substituting this representation of S_{tc}^2 into equation (3), the alternative expression for the sampling variance of $\hat{\tau}^{\text{dif}}$ (alternative to (3)) is:

$$\mathbb{V}_W\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right) = \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 + \frac{2}{N} \cdot \rho_{tc} \cdot S_c \cdot S_t. \tag{5}$$

The sampling variance of our estimator is smallest when the potential outcomes are perfectly negatively correlated ($\rho_{tc} = -1$), so that

$$S_{tc}^2 = S_c^2 + S_t^2 + 2 \cdot S_c \cdot S_t,$$

and

$$\mathbb{V}_W\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \,\middle|\, \rho_{tc} = -1\right) = \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 - \frac{2}{N} \cdot S_c \cdot S_t,$$

The Sampling Variance of of the Neyman Estimator with N Units

$$\mathbb{V}_W\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \middle| \rho_{tc} = -1\right) = \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 - \frac{2}{N} \cdot S_c \cdot S_t,$$

$$= \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{(S_c - S_t)^2}{N}. \tag{6}$$

When the treatment effect is constant and additive, $Y_i(1) - Y_i(0) = \tau$ for all $i = 1, \ldots, N$,

$$\mathbb{V}^{\text{const}} = \mathbb{V}_W\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \middle| \rho_{tc} = 1, S_c^2 = S_t^2\right) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}. \tag{7}$$

The fact that the sampling variance of $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$ is largest when the treatment effect is constant (*i.e.*, not varying) across units may appear somewhat counterintuitive. Howebr the two-unit case provides the intuition for why this is the case.

The Sampling Variance of the Neyman Estimator with N Units

Consider two numerical examples.
(1) $Y_1(0) = Y_1(1) = 10$, and $Y_2(0) = Y_2(1) = -10$, corresponding to a zero treatment effect for both units.

In this example $\rho_{tc} = 1$ as can be seen from that the numerator in (4) equals

$$\frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(1) - \overline{Y}(1)\Big)\cdot\Big(Y_i(0) - \overline{Y}(0)\Big)$$

$$= \Big((Y_1(1) - 0)\cdot(Y_1(0) - 0) + (Y_2(1) - 0)\cdot(Y_2(0) - 0)\Big) = 200,$$

and the two components of the denominator in (4) equal

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(0) - \overline{Y}(0)\Big)^2 = \Big((10-0)^2 + (-10-0)^2\Big) = 200,$$

The Sampling Variance of the Neyman Estimator with N Units

and

$$S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(1) - \overline{Y}(1)\Big)^2 = \Big((10-0)^2 + (-10-0)^2\Big) = 200,$$

(2) suppose that $Y_1(0) = Y_2(1) = -10$, and $Y_1(1) = Y_2(0) = 10$. A similar calculation shows that the $\rho_{tc} = -1$

In both examples, the *average* treatment effect is zero, but in (1), it constant. In (2), the treatment effect for unit 1 is 20, and for unit 2 the it is equal to -20.

The Sampling Variance of the Neyman Estimator with N Units

In (1) the two possible values of the estimator are $Y_1^{\text{obs}} - Y_2^{\text{obs}} = 20$ (if $W_1 = 1$ and $W_2 = 0$) and $Y_2^{\text{obs}} - Y_1^{\text{obs}} = -20$ (if $W_1 = 0$ and $W_2 = 1$).

In (2) the two values of the estimator are both equal to 0.

Hence the sampling variance of the estimator in (1), with $\rho_{tc} = +1$, is positive (in fact, equal to 20^2), whereas in(2), with $\rho_{tc} = -1$, the sampling variance is 0.

Estimating the Sampling Variance

The next step is to develop an estimator for the sampling variance. To do this, we consider separately each of the three elements of the sampling variance.

The numerator of the first term, the sample variance of the potential control outcome vector, $\mathbf{Y}(0)$, is equal to S_c^2.

From standard results on simple random samples, an unbiased estimator for S_c^2 is

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i(0) - \overline{Y}_c^{\text{obs}} \right)^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}} \right)^2.$$

Analogously, we can estimate S_t^2, the population variance of $Y_i(1)$, by

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i(1) - \overline{Y}_t^{\text{obs}} \right)^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}} \right)^2.$$

Estimating the Sampling Variance

The third term, S_{tc}^2 is generally impossible to estimate empirically because we never observe both $Y_i(1)$ and $Y_i(0)$ for the same unit.

As noted previously, if the treatment effects are constant and additive ($Y_i(1) - Y_i(0) = \tau_{\mathrm{S}}$ for all units), then this component of the sampling variance is equal to zero and the third term vanishes.
Thus we have proved:

Theorem

If the treatment effect $Y_i(1) - Y_i(0)$ is constant, then an unbiased estimator for the sampling variance is

$$\hat{\mathbb{V}}^{\text{neyman}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}. \tag{8}$$

Estimating the Sampling Variance

This estimator for the sampling variance is widely used, even when the assumption of an additive treatment effect may be known to be inaccurate.

Two main reasons, fist by implicitly setting the third element of the estimated sampling variance equal to zero, the expected value of $\hat{\mathbb{V}}^{\text{neyman}}$ is at least as large as the true sampling variance of $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$.

Hence, in large samples, confidence intervals generated using this estimator of the sampling variance will have coverage at least as large, but not necessarily equal to, their nominal coverage (i.e. conservative inference).

The second reason is that it is always unbiased for the sampling variance of $\hat{\tau}^{\text{dif}}$ as an estimator of the infinite super population average treatment effect.

Estimating the Sampling Variance

Here we consider two alternative estimators for the sampling variance of $\hat{\tau}^{\text{dif}}$.

The first explicitly allows for treatment effect heterogeneity.

Under treatment effect heterogeneity, the estimator for the sampling variance in equation (8), $\hat{\mathbb{V}}^{\text{neyman}}$, provides an upwardly biased estimate: the third term, which vanishes if the treatment effect is constant, is now negative.

The question arises whether we can improve upon the Neyman variance estimator without risking under coverage in large samples.

Estimating the Sampling Variance

Note that the implication of constant treatment effects is $S_c^2 = S_t^2$

If $S_c^2 \neq S_t^2$ this would in large samples lead to a difference in the corresponding estimates s_c^2 and s_t^2.

The sampling variance given in equation (5) is

$$\mathbb{V}_W\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right) = S_c^2 \cdot \frac{N_t}{N \cdot N_c} + S_c^2 \cdot \frac{N_c}{N \cdot N_t} + \rho_{tc} \cdot S_c \cdot S_t \cdot \frac{2}{N}.$$

Thus, if ρ_{tc} was known we could estimate the sampling variance

$$\hat{\mathbb{V}}^{\rho_{tc}} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + \rho_{tc} \cdot s_c \cdot s_t \cdot \frac{2}{N}. \tag{9}$$

Estimating the Sampling Variance

This variance is maximized if $\rho_{01} = 1$. Thus an alternative conservative estimator that exploits this bound is:

$$\hat{\mathbb{V}}^{\rho_{tc}=1} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_1^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N}$$

$$= \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} - \frac{(s_t - s_c)^2}{N}. \tag{10}$$

If s_c^2 and s_t^2 are unequal, then $\hat{\mathbb{V}}^{\rho_{tc}=1}$ will be smaller than $\hat{\mathbb{V}}^{\text{neyman}}$.

Using $\hat{\mathbb{V}}^{\rho_{tc}=1}$ to construct confidence intervals will result in tighter confidence intervals than using $\hat{\mathbb{V}}^{\text{neyman}}$, without compromising their large sample validity.

Estimating the Sampling Variance

The intervals based on $\hat{\mathbb{V}}^{\rho_{tc}=1}$ will still be conservative, in large samples, because $\hat{\mathbb{V}}^{\rho_{tc}=1}$ is still upwardly biased when the true correlation is smaller than one, although less so than $\hat{\mathbb{V}}^{\text{neyman}}$.

Note, however, that with no information beyond the fact that $S_c^2 \neq S_t^2$, all choices for ρ_{tc} smaller than unity raise the possibility that we will underestimate the sampling variance and construct invalid confidence intervals.

The second estimator builds on the assumption that $Y_i(1) - Y_i(0) = \tau$ for all i. Under the constant treatment assumption, the population variances of the two potential outcomes, S_c^2 and S_t^2, must be equal.

Estimating the Sampling Variance

We can therefore define $S^2 \equiv S_c^2 = S_t^2$ and pool the outcomes for the the treated and control units to estimate this common variance:

$$s^2 = \frac{1}{N-2} \cdot \left(s_c^2 \cdot (N_c - 1) + s_t^2 \cdot (N_t - 1) \right)$$

$$= \frac{1}{N-2} \cdot \left(\sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}} \right)^2 + \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}} \right)^2 \right). \tag{11}$$

The larger sample size for this estimator, leads to a more precise estimator if the treatment effect is constant, namely

$$\hat{\mathbb{V}}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t} \right). \tag{12}$$

When the treatment effects are constant this estimator is preferable to either $\hat{\mathbb{V}}^{\text{neyman}}$ or $\hat{\mathbb{V}}^{\rho_{tc}=1}$, but if not, it need not be valid. Both $\hat{\mathbb{V}}^{\text{neyman}}$ and $\hat{\mathbb{V}}^{\rho_{tc}=1}$ are valid generally, and therefore may be preferred.

The Duflo-Hanna-Ryan Teacher Incentive Experiment

We found that the incentives-based salary rather than the conventional salary structure increased the probability that the school was open by 0.22. What is the variance of this estimator?

The estimates of S_c^2, S_t^2, and the combined variance S^2 are

$$s_c^2 = 0.19^2, \qquad s_t^2 = 0.13^2, \qquad \text{and} \quad s^2 = 0.16^2.$$

Note that the two sample variances s_c^2 and s_t^2 are quite different.

This means that

$$\hat{\mathbb{V}}^{\text{neyman}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} = 0.0311^2$$

and

The Duflo-Hanna-Ryan Teacher Incentive Experiment

$$\hat{\mathbb{V}}^{\rho_{tc}=1} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N} = 0.0305^2.$$

By construction this estimator is smaller than $\hat{\mathbb{V}}^{\text{neyman}}$.

However, even though the variances s_c^2 and s_t^2 differ by more than a factor of two, the difference in the estimated sampling variances $\hat{\mathbb{V}}^{\rho_{tc}=1}$ and $\hat{\mathbb{V}}^{\text{neyman}}$ is very small in this example, less than 1%.

In general, the standard variance $\hat{\mathbb{V}}^{\text{neyman}}$ is unlikely to be substantially larger than $\hat{\mathbb{V}}^{\rho_{tc}=1}$, as suggested by this example.

The Duflo-Hanna-Ryan Teacher Incentive Experiment

The third and final estimate of the sampling variance, which relies on a constant treatment effect for its validity, is

$$\hat{\mathbb{V}}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t} \right) = 0.0312^2,$$

slightly larger than the other estimates, but essentially the same for practical purposes.

The Duflo-Hanna-Ryan Teacher Incentive Experiment

Table 6.2: Estimates for Effect of Teacher Incentives on Proportion of Days that School is Open

Estimated Means	$\overline{Y}_c^{\text{obs}}$	0.58
	$\overline{Y}_t^{\text{obs}}$	0.80
	$\hat{\tau}$	0.22
Estimated Variance Components	s_c^2	0.19^2
	s_t^2	0.13^2
	s^2	0.16^2
	N_c	54
	N_t	53
Sampling Variance Estimates	$\hat{\mathbb{V}}^{\text{neyman}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$	0.03^2
	$\hat{\mathbb{V}}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t}\right)$	0.03^2
	$\hat{\mathbb{V}}^{p_{tc}=1} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N}$	0.03^2

Confidence Intervals and Testing

This section discuss a number of ways to construct a confidence interval (CI) and to conduct tests for hypotheses concerning the average treatment effect.

By a CI with confidence coefficient $1-\alpha$, we mean a pair of functions $C_L(\mathbf{Y}^{\text{obs}}, \mathbf{W})$ and $C_U(\mathbf{Y}^{\text{obs}}, \mathbf{W})$, defining an interval $[C_L(\mathbf{Y}^{\text{obs}}, \mathbf{W}), C_U(\mathbf{Y}^{\text{obs}}, \mathbf{W})]$, such that

$$\Pr_W(C_L(\mathbf{Y}^{\text{obs}}, \mathbf{W}) \leq \tau \leq C_U(\mathbf{Y}^{\text{obs}}, \mathbf{W})) \geq 1-\alpha.$$

The only reason the lower and upper bounds in this interval are random is through their dependence on $\mathbf{W}$.

Note that, in this expression, the probability of including the true value τ may exceed $1-\alpha$, in which case the interval is valid but conservative.

The CI is based on a normal approximation to the randomization distribution of $\hat{\tau}^{\text{dif}}$.

This approximation is somewhat intellectually inconsistent with our stress on finite sample properties of the estimator for τ and its sampling variance, but it is driven by the common lack of empirical *a priori* information about the joint distribution of the potential outcomes.

Let $\hat{\mathbb{V}}$ be an estimate of the sampling variance of $\hat{\tau}^{\text{dif}}$ over its randomization distribution (in practice we recommend using $\hat{\mathbb{V}}^{\text{neyman}}$).

Confidence Intervals

Normality is often a good approximation to the randomization distribution of standard estimates, even in fairly small samples. To further improve on this approximation, we could approximate the distribution of $\hat{\mathbb{V}}$ by a chi-squared distribution, and then use that to approximate the distribution of $\hat{\tau}^{\text{dif}}/\sqrt{\hat{\mathbb{V}}}$ by a t-distribution.

A central CI with nominal confidence level $(1-\alpha)\times 100\%$, is

$$\text{CI}^{1-\alpha}(\tau_{\text{FS}}) = \left(\hat{\tau}^{\text{dif}} + z_{\alpha/2}\cdot\sqrt{\hat{\mathbb{V}}}, \hat{\tau}^{\text{dif}} + z_{1-\alpha/2}\cdot\sqrt{\hat{\mathbb{V}}}\right),$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the standard normal distribution, respectively.

Thus, with a wish to construct a 90% CI, the nominal central 90% CI is

$$\text{CI}^{0.90}(\tau_{\text{FS}}) = \left(\hat{\tau}^{\text{dif}} - 1.645\cdot\sqrt{\hat{\mathbb{V}}}, \hat{\tau}^{\text{dif}} + 1.645\cdot\sqrt{\hat{\mathbb{V}}}\right).$$

Confidence Intervals

Note that the validity of the CIs is under the same assumptions that make the corresponding estimator for the sampling variance an unbiased or upwardly biased estimator of the true sampling variance.

Based on the three estimators $\hat{\mathbb{V}}^{\text{neyman}}$, $\hat{\mathbb{V}}^{\text{const}}$ and $\hat{\mathbb{V}}^{\rho_{tc}=1}$, the 90% confidence intervals are

$$\text{CI}^{0.90}_{\text{neyman}}(\tau_{\text{FS}}) = (0.2154 - 1.645 \cdot 0.0311, 0.2154 + 1.645 \cdot 0.0311) = (0.1642, 0.2667),$$

$$\text{CI}^{0.90}_{\text{const}}(\tau_{\text{FS}}) = (0.1640, 0.2668),$$

and

$$\text{CI}^{0.90}_{\rho_{tc}=1}(\tau_{\text{FS}}) = (0.1652, 0.2657),$$

Testing

Suppose we wish to test:

$$H_0^{\text{neyman}}: \quad \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) = 0$$

$$H_a^{\text{neyman}}: \quad \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) \neq 0$$

A natural test statistic to use for Neyman's "average null" is the ratio of the point estimate to the estimated standard error.

We have $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} = 0.2154$ and $\hat{\mathbb{V}}^{\text{neyman}}$=0.0311.

Testing

The resulting t-statistic is therefore

$$t = \frac{\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}}{\sqrt{\hat{\mathbb{V}}^{\text{neyman}}}} = -\frac{0.2154}{0.0311} = 6.9.$$

The associated p-value for a two-sided test, based on the normal approximation to the distribution of the t-statistic, is $2 \cdot (1 - \Phi(6.9)) < 0.001$.

At conventional significance levels, we clearly reject the (Neyman) null hypothesis that the average treatment effect is zero.

Testing

There are two important differences between the Neyman and Fisher approaches.

1. they assess different null hypotheses.
2. the Neyman test relies on a large sample normal approximation for its validity

With regard to (1), The Neyman null assesses whether the average treatment effect is zero and with the alternative being different from zero, while in the Fisher approach the null hypothesis is

$$H_0^{\text{fisher}}: \qquad Y_i(1) - Y_i(0) = 0 \qquad \text{for all } i = 1, \ldots, N,$$

and the (implicit) alternative hypothesis is

$$H_a^{\text{fisher}}: \qquad Y_i(1) - Y_i(0) \neq 0 \qquad \text{for some } i = 1, \ldots, N.$$

Testing

Depending on the implementation of the FEP approach, this difference in null hypotheses may be unimportant.

To illustrate thsi point consider as an example, a population where for all units $Y_i(0) = 2$. For 1/3 of the units the treatment effect is 2. For 2/3 of the units the treatment effect is -1.

In this case the Neyman null hypothesis of a zero average effect is true. The Fisher null hypothesis of no effect whatsoever is not true.

Testing

Whether we can detect this violation depends on the choice of statistic. The FEP approach, with the statistic equal to the average difference in outcomes by treatment status, has no power against this alternative.

However, the FEP approach, with a different statistic, based on the average difference in outcomes after transforming the outcomes by taking logarithms, does have power in this setting.

In this artificial example, the expected difference in logarithms by treatment status is -0.23. The FEP based on the difference in average logarithms will detect this difference in large samples.

Testing

The second difference between the two procedures is in the approximate nature of the Neyman test, compared to the exact results for the FEP approach.

We use two approximations in the Neyman approach. First, we use the *estimated* variance (e.g., $\hat{\mathbb{V}}^{\text{neyman}}$) instead of the *actual* variance ($\mathbb{V}_W(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}})$).

Second, we use a normal *approximation* for the repeated sampling distribution of the difference in averages $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$. Both approximations are justified in large samples.

If the sample is reasonably large, and if there are few or no outliers, as in the application in this chapter, these approximations will likely be accurate.

Inference for Population Average

Suppose that the population of N subjects taking part in the CRE is itself a simple random sample from a larger population, which, for simplicity, we assume is infinite.

Viewing our N units as a sample of the target super population, rather than viewing them as the population itself, induces a distribution on the two potential outcomes for each unit.

The pair of potential outcome values for an observed unit i is simply one draw from the distribution in the population and is, therefore, itself stochastic.

Inference for Population Average

To be clear about this super population perspective, we used the subscript FS to denote the finite sample average treatment effect and SP to denote the super population average treatment effect:

$$\tau_{\mathrm{FS}} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0)\right) \qquad \text{and} \quad \tau_{\mathrm{SP}} = \mathbb{E}_{\mathrm{SP}}\left[Y_i(1) - Y_i(0)\right].$$

the subscript SP indicates that the expectation is taken over the distribution generated by random sampling from the super population, and not solely over the randomization distribution.

Inference for Population Average

$\tau_{\text{SP}} = \mathbb{E}_{\text{SP}}[Y_i(1) - Y_i(0)]$ is denoted as the average treatment effect in the super population.

Because of the random sampling, τ_{SP} is also equal to the expected value of the finite sample average treatment effect,

$$\mathbb{E}_{\text{SP}}\left[\tau_{\text{FS}}\right] = \mathbb{E}_{\text{SP}}\left[\overline{Y}(1) - \overline{Y}(0)\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\text{SP}}\left[Y_i(1) - Y_i(0)\right] = \tau_{\text{SP}}. \tag{13}$$

Let σ_c^2 and σ_t^2 denote the population variances of the two potential outcomes, or the super population expectations of S_c^2 and S_t^2:

$$\sigma_c^2 = \mathbb{V}_{\text{SP}}(Y_i(0)) = \mathbb{E}_{\text{SP}}\left[(Y_i(0) - \mathbb{E}_{\text{SP}}[Y_i(0))^2\right],$$

and

$$\sigma_t^2 = \mathbb{V}_{\text{SP}}(Y_i(1)) = \mathbb{E}_{\text{SP}}\left[(Y_i(1) - \mathbb{E}_{\text{SP}}[Y_i(1))^2\right].$$

Inference for Population Average

The variance of the unit-level treatment effect in the super-population is similarly defined, $\sigma^2_{tc} = \mathbb{V}_{\text{SP}}(Y_i(1) - Y_i(0)) = \mathbb{E}_{\text{SP}}[(Y_i(1) - Y_i(0) - \tau_{\text{SP}})^2]$.

As seen from this definition the variance of τ_{FS} across repeated random samples is equal to

$$\mathbb{V}_{\text{SP}}(\tau_{\text{FS}}) = \mathbb{V}_{\text{SP}}\left(\overline{Y}(1) - \overline{Y}(0)\right) = \sigma^2_{tc}/N. \tag{14}$$

Now consider the sampling variance of $\hat{\tau}^{\text{dif}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$, given this sampling from the super-population.

The expectation and variance operators without subscripts denote expectations and variances taken over both the randomization distribution and the random sampling from the super-population.

Inference for Population Average

We have

$$\mathbb{V}\left(\hat{\tau}^{\text{dif}}\right) = \mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \mathbb{E}\left[\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \mathbb{E}_{\text{SP}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right],$$

where the second equality holds because $\mathbb{E}\left[\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right] = \mathbb{E}_{\text{SP}}[\overline{Y}(1) - \overline{Y}(0)] = \tau_{\text{SP}}$.

Adding and subtracting $\overline{Y}(1) - \overline{Y}(0)$ within the expectation, this sampling variance, over both randomization and random sampling, is equal to:

Inference for Population Average

$$\mathbb{V}\left(\hat{\tau}^{\text{dif}}\right)$$

$$= \mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right) + \left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{SP}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right)^2\right]$$

$$+\mathbb{E}_{\text{SP}}\left[\left(\left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{SP}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right]$$

$$+2\cdot\mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right)\cdot\left(\left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{SP}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)\right]$$

Inference for Population Average

The third term is equal to zero because the expectation of the first factor, $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - (\overline{Y}(1) - \overline{Y}(0))$, conditional on the N–vectors $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$, is zero.

Hence the sampling variance reduces to:

$$\mathbb{V}\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right) = \mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \overline{Y}(1) - \overline{Y}(0)\right)^2\right]$$

$$+\mathbb{E}_{\text{SP}}\left[\left(\overline{Y}(1) - \overline{Y}(0) - \mathbb{E}_{\text{SP}}\left[Y(1) - Y(0)\right]\right)^2\right]. \qquad (15)$$

Because $\mathbb{E}_W\left[\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \middle| \mathbf{Y}(0), \mathbf{Y}(1)\right] = \tau_{\text{FS}} = \overline{Y}(1) - \overline{Y}(0)$ the first term is equal to the expectation of the conditional variance of $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$ (conditional on the N-vector of potential outcomes $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$).

Inference for Population Average

This means that conditional variance is equal to

$$\mathbb{E}_W\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \overline{Y}(1) - \overline{Y}(0)\right)^2 \middle| \mathbf{Y}(0), \mathbf{Y}(1)\right] = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}, \tag{16}$$

as in equation (3).

The expectation of (16) over the distribution of $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ generated by sampling from the superpopulation is then

$$\begin{aligned}
&\mathbb{E}\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \overline{Y}(1) - \overline{Y}(0)\right)^2\right] \\
&= \mathbb{E}_{\text{SP}}\left[\mathbb{E}_W\left[\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} - \overline{Y}(1) - \overline{Y}(0)\right)^2 \middle| \mathbf{Y}(0), \mathbf{Y}(1)\right]\right] \\
&= \mathbb{E}_{\text{SP}}\left[\frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}\right] = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t} - \frac{\sigma_{tc}^2}{N}.
\end{aligned}$$

Inference for Population Average

As the expectation of the second term on the right side of equation (15), is equal to σ^2_{tc}/N this means that the variance of $\hat{\tau}^{\text{dif}}$ over sampling from the superpopulation equals:

$$\mathbb{V}_{\text{SP}} = \mathbb{V}_{\text{SP}}\left(\hat{\tau}^{\text{dif}}\right) = \frac{\sigma^2_c}{N_c} + \frac{\sigma^2_t}{N_t}, \tag{17}$$

which we can estimate without bias by substituting s^2_c and s^2_t for σ^2_c and σ^2_t, respectively:

$$\hat{\mathbb{V}}^{\text{SP}} = \frac{s^2_c}{N_c} + \frac{s^2_t}{N_t}.$$

The estimator $\hat{\mathbb{V}}^{\text{SP}}$ is identical to the previously introduced conservative estimator of the sampling variance for the finite population average treatment effect estimator, $\hat{\mathbb{V}}^{\text{neyman}}$

Inference for Population Average

Under simple random sampling from the super-population, the expected value of the estimator $\hat{\mathbb{V}}^{\text{neyman}}$ equals $\mathbb{V}_{\text{SP}}$.

Neither $\hat{\mathbb{V}}^{\text{const}}$ nor $\hat{\mathbb{V}}^{\rho_{tc}=1}$ in equation (10)have this attractive quality.

Thus, despite the fact that $\hat{\mathbb{V}}^{\text{const}}$ may be a better estimator of the sampling variance in the finite population when the treatment effect is constant, and $\hat{\mathbb{V}}^{\rho_{tc}=1}$ may be a better estimator of $\mathbb{V}_{\text{FS}}$, $\hat{\mathbb{V}}^{\text{neyman}}$ is used almost uniformly in practice in our experience, although the logic for it appears to be rarely explicitly discussed.

Neyman's Approach With Covariates

One can easily extend Neyman's approach for estimating average treatment effects to settings with discrete covariates.

In this case, one would partition the sample into subsamples defined by the values of the covariate and then conduct the analysis separately within these subsamples.

The resulting within-subsample estimators would be unbiased for the within-subsample average treatment effect.

Taking an average of these estimates, weighted by subsample sizes, gives an unbiased estimate of the overall average treatment effect (more on this in Chapter 9 on stratified random experiments).

It is impossible, however, in general to derive estimators that are exactly unbiased

Neyman's Approach With Covariates

It is impossible, however, in general to derive estimators that are exactly unbiased under the randomization distribution, conditional on the covariates, when there are covariate values for which we have only treated or only control units, which is likely to happen with great frequency in settings with covariates that take on many values.

In such settings, building a model for the potential outcomes, and using this model to create an estimator of the average treatment effect, is a more appealing option. We turn to this topic in the next two chapters.

Results for the Duflo-Hanna Teacher Incentive Data

We analyze four outcomes in turn, plus one "pseudo-outcome", (`pctprewritten`). In general, it can be useful to carry out such analyses as a check on the success of the randomization.

1. `pctprewritten`, the point estimate is -0.03 and the 95% confidence interval is (-.10,0.04)).
2. `open`, the proportion of days that the school was open during the days it was subject to a random check. 0.22 and $[0.15, 0.28]$.
3. `pctpostwritten`, the percentage of students in the school who took the written test. 0.05 and $[-0.03, 0.13]$
4. `written`, the average score on the writing test. 0.17 and $[0.00, 0.34]$.
5. `written_all`, the average test score, assigning zeros to students not taking the test. 0.14 and $[0.00, 0.28]$

Results for the Duflo-Hanna Teacher Incentive Data

Table 6.3: Estimates of, and Confidence Intervals for, Average Treatment Effects for Duflo-Hanna-Ryan Teacher-Incentive Data

		$\widehat{\text{ate}}$	$\widehat{\text{(s.e.)}}$	95% c.i.
pretreatment	pctprewritten	-0.03	(0.04)	[-0.10,0.04]
posttreatment	open	0.22	(0.03)	[0.15,0.28]
	pctpostwritten	0.05	(0.04)	[-0.03,0.13]
	written	0.17	(0.08)	[0.00,0.34]
	written_all	0.14	(0.07)	[0.00,0.28]

Results for the Duflo-Hanna Teacher Incentive Data

In the final analysis, we look at estimates separately for two subsamples, defined by whether the proportion of students taking the initial writing test was zero or positive, to illustrate the application of the methods developed in this chapter to subpopulations defined by covariates.

Again, these analyses are for illustrative purposes only, and we do not take account of the fact that we do multiple tests.

The first subpopulation (`pctprewritten`$= 0$) comprises 40 schools (37%) and the second (`pctprewritten`> 0) 67 schools (63%). We analyze separately the effect of assignment to attendance-based teacher incentives on all four outcomes.

Results for the Duflo-Hanna Teacher Incentive Data

Table 6.4: Estimates of, and Confidence Intervals for, Average Treatment Effects for Duflo-Hanna-Ryan Teacher-Incentive Data

	pctprewritten = 0 ($N = 40$)			pctprewritten > 0 ($N = 67$)			Difference		
variable	$\hat{\tau}$	$\widehat{\text{(s.e.)}}$	95% c.i.	$\hat{\tau}$	$\widehat{\text{(s.e.)}}$	95% c.i.	est	$\widehat{\text{(s.e.)}}$	95% c.i.
open	0.23	(0.05)	[0.14,0.32]	0.21	(0.04)	[0.13,0.29]	0.02	(0.06)	[-0.10,0.14]
pctpostwritten	-.004	0.06	[-0.16,0.07]	0.11	(0.05)	[0.01,0.21]	-0.15	(0.08)	[-0.31,0.00]
written	0.20	(0.10)	[0.00,0.40]	0.18	(0.10)	[-0.03,0.38]	0.03	(0.15)	[-0.26,0.31]
written_all	0.04	(0.07)	[-0.10,0.19]	0.22	(0.09)	[0.04,0.40]	-0.18	(0.12)	[-0.41,0.05]