# Unbiasedness of Difference-in-Means Estimator under Random Assignment*

Thomas Leavitt

This guide proves that the Difference-in-Means estimator is unbiased for the Sample Average Treatment Effect (under both complete and simple random assignment) and, when the study sample is itself a random draw from a superpopulation, that it is also unbiased for the Population Average Treatment Effect. It is self-contained, but it shares its framework with the companion guides "Notation and Setup: The Finite Population Potential Outcomes Framework" (the full development of the notation used here), "Variance and Conservative Estimation of the Difference-in-Means Estimator (SATE)," and "Variance and Conservative Estimation of the Difference-in-Means Estimator (PATE)."

## 1 Estimation of Sample Average Treatment Effect (SATE)

### 1.1 Setup

Let the index $i \in \{1, \dots, n\}$ run over $n$ units in a finite sample, $\mathcal{S}_n$, where $n \geq 4$. Of these $n$ units, $n_T \geq 2$ are assigned to the treatment condition and $n_C \geq 2$ are assigned to the control condition, where $n_T + n_C = n$. Although not necessary for the unbiasedness of the estimator or for the derivation of the Difference-in-Means estimator's variance, these assumptions on the sizes of $n$, $n_T$ and $n_C$ ensure that the conservative (Neyman) estimator of the Difference-in-Means estimator's variance is well defined. That estimator forms a separate sample variance within each treatment arm, dividing by $n_T - 1$ for the treated units and by $n_C - 1$ for the control units, and so requires

*This is a live document that is subject to updating at any time.

at least two units in each arm. We therefore assume that $n_T \geq 2$ and $n_C \geq 2$, so that $n \geq 4$. Let the binary indicator variable $Z_i \in \{0, 1\}$ denote whether unit $i$ is assigned to treatment ($Z_i = 1$) or control ($Z_i = 0$), and collect these indicators in the random assignment vector $\boldsymbol{Z} = \left[Z_1, \ldots, Z_n\right]^\top$. The set of all logically possible assignment vectors is $\{0, 1\}^n$. The assignment mechanism, however, places positive probability only on those vectors with exactly $n_T$ treated units; we collect these in the set $\Omega = \left\{\boldsymbol{z} \in \{0, 1\}^n : \sum_{i=1}^n z_i = n_T\right\}$, so that $\Omega$ is the support of $\boldsymbol{Z}$. Under complete random assignment, $\boldsymbol{Z}$ is distributed uniformly on $\Omega$, and the number of elements in $\Omega$, denoted $|\Omega|$,[1] is $\binom{n}{n_T}$. By contrast, under $n$ independent Bernoulli assignments, there would be $2^n$ possible assignment vectors. However, even if $n_T$ is not fixed by design (as in complete random assignment), we can fix $n_T$ by conditioning on its observed value. The randomization distribution conditional on the realized $n_T$ yields the same randomization distribution one would obtain if $n_T$ had been fixed ex ante by design. Hence, this general setup and the proof to follow pertains to both simple and complete random assignment even though the argument by which one can regard $n_T$ as fixed is slightly different under simple and complete assignment mechanisms.

Adopting the terminology of Freedman (2009) and later Gerber and Green (2012), define a potential outcomes schedule as a vector-valued function $\boldsymbol{y} : \{0, 1\}^n \to \mathbb{R}^n$ that maps each possible assignment vector $\boldsymbol{z} \in \{0, 1\}^n$ to an $n$-dimensional vector of real-valued outcomes. More intuitively, a potential outcomes schedule is a listing of how each of the $n$ study participants would respond to every assignment $\boldsymbol{z}$ that the experiment could in principle produce. The $i$th entry of $\boldsymbol{y}(\boldsymbol{z})$ is the outcome that unit $i$ would exhibit under assignment $\boldsymbol{z}$. The potential outcomes are *fixed* features of the units; they do not vary with the random assignment. Randomness enters only later, through which assignment $\boldsymbol{z}$ the mechanism happens to select.

Since the assignment space $\{0, 1\}^n$ contains $2^n$ assignment vectors, the schedule specifies, in principle, $2^n$ outcome vectors. However, under the Stable Unit Treatment Value Assumption (SUTVA)[2] (Cox, 1958; Rubin, 1980, 1986), unit $i$'s outcome depends only on its own assignment $z_i$ and not on the assignments of the other units. Accordingly, let $y_{Ti}$ denote the common outcome value of unit $i$ across all assignments $\boldsymbol{z}$ with $z_i = 1$, and let $y_{Ci}$ denote the common outcome value of unit $i$ across all assignments $\boldsymbol{z}$ with $z_i = 0$. SUTVA thus collapses the entire schedule down to just two fixed numbers per unit, $y_{Ti}$ and $y_{Ci}$. The individual causal effect for unit $i$ on the additive scale is $\tau_i = y_{Ti} - y_{Ci}$. The vectors $\boldsymbol{y_T}$ and $\boldsymbol{y_C}$ collect the treatment and control potential outcomes, respectively, for all $n$ units, and $\boldsymbol{\tau}$ collects the $n$ individual, additive effects. These are all *fixed*, finite population quantities. The observed outcome for unit $i \in \{1, \ldots, n\}$ is $Y_i = Z_i y_{Ti} + (1 - Z_i)\, y_{Ci}$, which equals $y_{Ti}$ when $Z_i = 1$ and $y_{Ci}$ when $Z_i = 0$. Because $y_{Ti}$ and $y_{Ci}$ are fixed, the *only* source

[1] For an arbitrary set $W$, let $|W|$ denote the cardinality of (i.e., the number of elements in) the set $W$.

[2] SUTVA implies that (1) units in the experiment respond to only the treatment condition to which each unit is individually assigned and (2) the treatment condition is actually the same treatment for all units assigned to treatment and the control condition is the same for all units assigned to control.

of randomness in the observed outcome $Y_i$ is the assignment indicator $Z_i$.

It will be convenient to write the finite population means of the treatment and control potential outcomes as $\bar{y}_T \coloneqq n^{-1} \sum_{i=1}^{n} y_{Ti}$ and $\bar{y}_C \coloneqq n^{-1} \sum_{i=1}^{n} y_{Ci}$, both of which are fixed quantities. The target of interest is the Sample Average Treatment Effect (SATE),

$$\tau_{\text{SATE}} \coloneqq n^{-1} \sum_{i=1}^{n} \tau_i = n^{-1} \sum_{i=1}^{n} \left(y_{Ti} - y_{Ci}\right) = \bar{y}_T - \bar{y}_C,$$

which is likewise a fixed (though unknown) quantity. Define the Difference-in-Means estimator of $\tau_{\text{SATE}}$ as

$$\hat{\tau} \coloneqq \frac{\sum\limits_{i=1}^{n} Z_i Y_i}{\sum\limits_{i=1}^{n} Z_i} - \frac{\sum\limits_{i=1}^{n} \left(1 - Z_i\right) Y_i}{\sum\limits_{i=1}^{n} \left(1 - Z_i\right)}. \tag{1}$$

Unlike the estimand $\tau_{\text{SATE}}$, the estimator $\hat{\tau}$ is a *random variable*, since it is a function of the random assignment vector $\boldsymbol{Z}$ and inherits all of its randomness from $\boldsymbol{Z}$. Under complete random assignment the denominators are fixed, $\sum_{i=1}^{n} Z_i = n_T$ and $\sum_{i=1}^{n} \left(1 - Z_i\right) = n_C$, so $\hat{\tau}$ can equivalently be written

$$\hat{\tau} = \frac{1}{n_T} \sum_{i=1}^{n} Z_i Y_i - \frac{1}{n_C} \sum_{i=1}^{n} \left(1 - Z_i\right) Y_i,$$

which is the form used in the companion guides on the estimator's variance. For the expectation of $\hat{\tau}$ in Equation (1), I write $\mathrm{E}_{\Omega}\left[\cdot\right]$ to emphasize that the expectation is taken over only the randomness of the assignment process, i.e., over the distribution of $\boldsymbol{Z}$ on $\Omega$.

### 1.2 Estimation under Complete Random Assignment

**Lemma 1.** *Under complete, uniform random assignment in which $n_T$ out of $n$ total units are assigned to treatment,* $\mathrm{E}_{\Omega}\left[Z_i\right] = \dfrac{n_T}{n}$ *for all $i \in \{1, \ldots, n\}$ units.*

*Proof.* We will complete this proof in two steps: We will show that (1) the proportion of assignments in which unit $i$ is in the treatment condition is $\dfrac{n_T}{n}$ and (2) under uniform, random assignment, the probability that $Z_i = 1$ is equal to this proportion $\dfrac{n_T}{n}$.

**Step 1:** First note that the number of ways to choose a subset of $n_T$ treated units from a fixed population of $n$ units is as follows:

$$\binom{n}{n_T} = \frac{n!}{(n-n_T)!n_T!} = \frac{n!}{n_C!n_T!}, \tag{2}$$

where $n_C = n - n_T$ is the number of units assigned to the control condition.

Given that an arbitrary unit $i$ is in the treatment condition and only $n_T$ total units can be in the treatment condition, there are $\binom{n-1}{n_T-1}$ ways in which $n_T - 1$ other units could be in the treatment condition. Hence, the number of assignments in which unit $i$ is treated and $n_T - 1$ other units are treated is:

$$\binom{n-1}{n_T-1} = \frac{(n-1)!}{\big((n-1)-(n_T-1)\big)!\,(n_T-1)!} \tag{3}$$

To get the proportion of assignments in which unit $i$ is treated, we need to divide (3) by (2):

$$\frac{\binom{n-1}{n_T-1}}{\binom{n}{n_T}} = \frac{\left(\dfrac{(n-1)!}{\big((n-1)-(n_T-1)\big)!\,(n_T-1)!}\right)}{\left(\dfrac{n!}{n_C!n_T!}\right)} \tag{4}$$

Now notice that:

$$\begin{aligned}(n-1)-(n_T-1) &= n-1-n_T+1\\ &= n-n_T\\ &= n_C\end{aligned}$$

We can therefore substitute $n_C$ for $(n-1)-(n_T-1)$ in (4), which gives us:

$$\frac{\left(\dfrac{(n-1)!}{n_C!\,(n_T-1)!}\right)}{\left(\dfrac{n!}{n_C!n_T!}\right)} \tag{5}$$

Now we can simply manipulate (5) and cancel terms until we are left with $\frac{n_T}{n}$:

$$
\begin{aligned}
&= \left(\frac{(n-1)!}{n_C!\,(n_T-1)!}\right)\left(\frac{n_C!n_T!}{n!}\right) \\
&= \left(\frac{(n-1)\,(n-2)\ldots 1}{n_C\,(n_C-1)\ldots 1\,(n_T-1)\ldots 1}\right)\left(\frac{n_C\,(n_C-1)\ldots 1 n_T\,(n_T-1)\ldots 1}{n\,(n-1)\ldots 1}\right) \\
&= \frac{\color{blue}{(n-1)\,(n-2)\ldots 2}\color{red}{n_C\,(n_C-1)\ldots 2}n_T\color{green}{(n_T-1)\ldots 2}}{\color{red}{n_C\,(n_C-1)\ldots 2}\color{green}{(n_T-1)\ldots 2}n\color{blue}{(n-1)\ldots 2}}
\end{aligned}
$$

All of the respective matching colors in the numerator and denominator cancel, which leaves us with $\frac{n_T}{n}$. Therefore, exactly the fraction $\frac{n_T}{n}$ of all possible assignments are those in which unit $i$ is in the treatment condition.

**Step 2:** The total probability of all assignments in which $i$ is treated is simply the sum of the probabilities of those assignments in which unit $i$ is in the treatment condition. Under uniform random assignment, the probability of each assignment permutation is $\frac{1}{|\Omega|}$. Thus, the probability that unit $i$ is treated is as follows:

$$
\begin{aligned}
\left(\frac{1}{|\Omega|}\right)\left(\frac{n_T}{n}\right)|\Omega| &= \left(\frac{1}{|\Omega|}\right)\frac{n_T\,|\Omega|}{n} \\
&= \frac{n_T\,|\Omega|}{|\Omega|\,n} \\
&= \frac{n_T}{n}
\end{aligned}
$$

Since $\Pr(Z_i = 1) = \frac{n_T}{n}$ for all $i \in \{1, \ldots, n\}$ units, it follows that the expected value of $Z_i \in \{0, 1\}$ is $\mathrm{E}_\Omega[Z_i] = 1\left(\frac{n_T}{n}\right) + 0\left(1 - \frac{n_T}{n}\right) = \frac{n_T}{n}$. $\square$

**Proposition 1.** *Under complete, uniform random assignment,* $\mathbb{E}_\Omega[\hat{\tau}] = \tau_{SATE}$.

*Proof.* First, the linearity of expectations implies that

$$\begin{aligned}
\mathbb{E}_\Omega\left[\hat{\tau}\right] &= \mathrm{E}_\Omega\left[\frac{\sum\limits_{i=1}^{n} Z_i Y_i}{\sum\limits_{i=1}^{n} Z_i} - \frac{\sum\limits_{i=1}^{n}\left(1 - Z_i\right) Y_i}{\sum\limits_{i=1}^{n}\left(1 - Z_i\right)}\right] \\
&= \mathrm{E}_\Omega\left[\frac{\sum\limits_{i=1}^{n} Z_i Y_i}{\sum\limits_{i=1}^{n} Z_i}\right] - \mathrm{E}_\Omega\left[\frac{\sum\limits_{i=1}^{n}\left(1 - Z_i\right) Y_i}{\sum\limits_{i=1}^{n}\left(1 - Z_i\right)}\right]
\end{aligned}$$

and, since the number of treated and control units are fixed at $n_T$ and $n_C$ under complete random assignment,

$$\mathbb{E}_\Omega\left[\hat{\tau}\right] = \frac{1}{n_T}\,\mathrm{E}_\Omega\left[\sum_{i=1}^{n} Z_i Y_i\right] - \frac{1}{n_C}\,\mathrm{E}_\Omega\left[\sum_{i=1}^{n}\left(1 - Z_i\right) Y_i\right]$$

Since, under SUTVA, the observed outcomes for treated units is equal to those units' treatment potential outcomes, we can substitute $Z_i y_{Ti}$ for $Z_i Y_i$. Analogously, we can substitute $\left(1 - Z_i\right) y_{Ci}$ for $\left(1 - Z_i\right) Y_i$. That is, with $Y_i = Z_i y_{Ti} + (1 - Z_i) y_{Ci}$, it follows that

$$\begin{aligned}
Z_i Y_i = Z_i\left(Z_i y_{Ti} + (1 - Z_i) y_{Ci}\right) &= Z_i y_{Ti} \text{ and} \\
(1 - Z_i) Y_i = (1 - Z_i)\left(Z_i y_{Ti} + (1 - Z_i) y_{Ci}\right) &= (1 - Z_i) y_{Ci},
\end{aligned}$$

which leaves us with

$$\begin{aligned}
\mathbb{E}_\Omega\left[\hat{\tau}\right] &= \frac{1}{n_T}\,\mathrm{E}_\Omega\left[\sum_{i=1}^{n} Z_i y_{Ti}\right] - \frac{1}{n_C}\,\mathrm{E}_\Omega\left[\sum_{i=1}^{n}\left(1 - Z_i\right) y_{Ci}\right] \\
&= \left(\frac{1}{n_T}\right)\mathbb{E}_\Omega\left[Z_1 y_{T1} + \cdots + Z_n y_{Tn}\right] - \left(\frac{1}{n_C}\right)\mathbb{E}_\Omega\left[\left(1 - Z_1\right) y_{C1} + \cdots + \left(1 - Z_n\right) y_{Cn}\right] \\
&= \left(\frac{1}{n_T}\right)\left(\mathbb{E}_\Omega\left[Z_1 y_{T1}\right] + \cdots + \mathbb{E}_\Omega\left[Z_n y_{Tn}\right]\right) - \left(\frac{1}{n_C}\right)\left(\mathbb{E}_\Omega\left[\left(1 - Z_1\right) y_{C1}\right] + \cdots + \mathbb{E}_\Omega\left[\left(1 - Z_n\right) y_{Cn}\right]\right) \\
&= \left(\frac{1}{n_T}\right)\left(y_{T1}\mathbb{E}_\Omega\left[Z_1\right] + \cdots + y_{Tn}\mathbb{E}_\Omega\left[Z_n\right]\right) - \left(\frac{1}{n_C}\right)\left(y_{C1}\mathbb{E}_\Omega\left[\left(1 - Z_1\right)\right] + \cdots + y_{Cn}\mathbb{E}_\Omega\left[\left(1 - Z_n\right)\right]\right)
\end{aligned}$$

By Lemma 1, $\mathbb{E}_\Omega\left[Z_i\right] = \left(\frac{n_T}{n}\right)$ for all $i \in \{1, \ldots, n\}$, which implies that $\mathbb{E}_\Omega\left[(1 - Z_i)\right] = 1 - \left(\frac{n_T}{n}\right) = \left(\frac{n_C}{n}\right)$ for all $i \in \{1, \ldots, n\}$. Hence, we can substitute $\left(\frac{n_T}{n}\right)$ for $\mathbb{E}_\Omega\left[Z_i\right]$ and $\left(\frac{n_C}{n}\right)$ for $\mathbb{E}_\Omega\left[1 - Z_i\right]$, respectively, which then yields

$$
\begin{aligned}
&= \left(\frac{1}{n_T}\right)\left(\frac{n_T}{n}\right)(y_{T1} + \cdots + y_{Tn}) - \left(\frac{1}{n_C}\right)\left(\frac{n_C}{n}\right)(y_{C1} + \cdots + y_{Cn}) \\
&= \left(\frac{1}{n}\right)(y_{T1} + \cdots + y_{Tn}) - \left(\frac{1}{n}\right)(y_{C1} + \cdots + y_{Cn}) \\
&= \frac{(y_{T1} + \cdots + y_{Tn})}{n} - \frac{(y_{C1} + \cdots + y_{Cn})}{n} \\
&= \bar{y}_T - \bar{y}_C \\
&= \tau_{\text{SATE}}.
\end{aligned}
$$

□

### 1.3 Estimation under Simple Random Assignment

Under simple random assignment, let the number of experimental units, $n$, remain a fixed quantity, but let the numbers of treatment and control units now be *random variables*, defined as $\tilde{n}_T \coloneqq \sum_{i=1}^{n} Z_i$ and $\tilde{n}_C \coloneqq \sum_{i=1}^{n} (1 - Z_i) = n - \tilde{n}_T$.

A note on notation is needed here. Elsewhere we distinguish a random quantity from its realized value by writing the random quantity with a capital letter and its realization with the corresponding lowercase letter, as with the assignment indicator $Z_i$ and its realization $z_i$. We do *not* follow that convention for the arm counts. The natural uppercase symbol, $N$, is already reserved, both by long-standing convention in the sampling literature and in the companion guide on the Population Average Treatment Effect, for the size of the (super)population, so writing the random treatment count as $N_T$ would invite confusion with a population size. To avoid this clash, in this one case we mark randomness with a *tilde* rather than with capitalization. The random treatment and control counts are $\tilde{n}_T$ and $\tilde{n}_C$, and the un-tilded $n_T$ and $n_C$ denote their realized values, that is, the values on which we condition.

We take the support of $\tilde{n}_T$ to be $\{1, \ldots, n-1\}$, so that neither $\tilde{n}_T$ nor $\tilde{n}_C$ can take on the value 0 (which would leave one of the two groups empty and the estimator $\hat{\tau}$ undefined). In this setting, the set of possible assignments is $\Omega = \left\{\boldsymbol{z} \in \{0,1\}^n : 0 < \sum_{i=1}^{n} z_i < n\right\}$, which contains $2^n - 2$ elements. In the proof to follow, note that whenever we take an expectation conditional on a realized number of treated units $n_T$ (that is, on the event $\{\tilde{n}_T = n_T\}$), the expectation is over $\Omega = \left\{\boldsymbol{z} : \sum_{i=1}^{n} z_i = n_T\right\}$. When not conditioning on a value of $n_T$, the expectation is over $\Omega = \left\{\boldsymbol{z} : 0 < \sum_{i=1}^{n} z_i < n\right\}$. For simplicity, I do not change the notation $\Omega$ for these two sets of assignments under complete and simple random assignment.

In the proof that follows, we will draw upon the Law of Iterated Expectations, which states in general that, for two random variables $X$ and $Y$, $\mathrm{E}\left[X\right] = \mathrm{E}_Y\Big[\mathrm{E}_X\left[X \mid Y = y\right]\Big]$, where $\mathrm{E}_X$ refers to the expectation over $X$ and $\mathrm{E}_Y$ refers to the expectation over $Y$.

**Proposition 2.** *Under simple, uniform random assignment,* $\mathbb{E}_\Omega\left[\hat{\tau}\right] = \tau_{SATE}$.

*Proof.* By the law of iterated expectations, the expected value of the Difference-in-Means estimator, $\hat{\tau}$, can be decomposed by conditioning on the realized number of treated units, $\tilde{n}_T$:

$$\mathrm{E}_\Omega\left[\hat{\tau}\right] = \mathrm{E}_\Omega\Big[\hat{\tau} \mid \tilde{n}_T = 1\Big] \Pr\left(\tilde{n}_T = 1\right) + \cdots + \mathrm{E}_\Omega\Big[\hat{\tau} \mid \tilde{n}_T = n-1\Big] \Pr\left(\tilde{n}_T = n-1\right). \tag{6}$$

Conditional on the event $\{\tilde{n}_T = n_T\}$, simple random assignment reduces to complete random assignment with $n_T$ treated units, so Proposition 1 applies. By that proposition, the expected value of the estimator conditional on any realized number of treated units $n_T \in \{1, \ldots, n-1\}$ is equal to $\tau_{\mathrm{SATE}}$. Hence, it follows that Equation (6) can be rewritten as

$$\mathrm{E}_\Omega\left[\hat{\tau}\right] = \tau_{\mathrm{SATE}} \Pr\left(\tilde{n}_T = 1\right) + \cdots + \tau_{\mathrm{SATE}} \Pr\left(\tilde{n}_T = n-1\right),$$

which we can rewrite as

$$\mathrm{E}_\Omega\left[\hat{\tau}\right] = \tau_{\mathrm{SATE}}\Big[\Pr\left(\tilde{n}_T = 1\right) + \cdots + \Pr\left(\tilde{n}_T = n-1\right)\Big].$$

Finally, note that, since $\tilde{n}_T$ takes values in $\{1, \ldots, n-1\}$ with probability one, the second and third axioms of probability imply that $\Big[\Pr\left(\tilde{n}_T = 1\right) + \cdots + \Pr\left(\tilde{n}_T = n-1\right)\Big] = 1$. Hence, it follows that

$$\begin{aligned} \mathrm{E}_\Omega\left[\hat{\tau}\right] &= \tau_{\mathrm{SATE}}\Big[1\Big] \\ \mathrm{E}_\Omega\left[\hat{\tau}\right] &= \tau_{\mathrm{SATE}}, \end{aligned}$$

which proves the proposition. □

## 2 Estimation of the Population Average Treatment Effect (PATE)

### 2.1 Setup

So far we have treated the $n$ units of $\mathcal{S}_n$ as the entire population of interest and shown that $\hat{\tau}$ is unbiased for the SATE. Suppose now that these $n$ units are themselves a random sample from a larger superpopulation, and ask whether $\hat{\tau}$ is unbiased for the average treatment effect in that superpopulation. (The full superpopulation framework, and the variance of $\hat{\tau}$ about the PATE, are developed in the companion guides; here we introduce only what unbiasedness requires.)

Consider a superpopulation $\mathcal{P}_N$ of size $N \geq n$, indexed by $i \in \{1, \ldots, N\}$, in which every unit has fixed potential outcomes $y_{Ti}$ and $y_{Ci}$ and a fixed individual effect $\tau_i = y_{Ti} - y_{Ci}$. Let $n$ of these $N$ units be drawn into the experimental sample $\mathcal{S}_n$ by simple random sampling, recorded by the indicator $R_i \in \{0, 1\}$, where $R_i = 1$ if unit $i$ is sampled and $R_i = 0$ otherwise; the sampling vector $\boldsymbol{R} = \left[R_1, \ldots, R_N\right]^\top$ is distributed uniformly on $\Pi = \left\{\boldsymbol{r} \in \{0, 1\}^N : \sum_{i=1}^N r_i = n\right\}$. Among the $n$ sampled units, $n_T$ are assigned to treatment and $n_C$ to control by complete random assignment, exactly as before. The estimand is the Population Average Treatment Effect,

$$\tau_{\text{PATE}} \coloneqq \frac{1}{N} \sum_{i=1}^{N} \tau_i,$$

a fixed (though unknown) feature of the superpopulation, and the Difference-in-Means estimator becomes

$$\hat{\tau} \coloneqq \frac{1}{n_T} \sum_{i=1}^{N} R_i Z_i Y_i - \frac{1}{n_C} \sum_{i=1}^{N} R_i \left(1 - Z_i\right) Y_i, \tag{7}$$

in which $R_i Z_i = 1$ only for units that are both sampled and treated, $R_i \left(1 - Z_i\right) = 1$ only for units that are both sampled and control, and unsampled units ($R_i = 0$) contribute nothing. There are now *two* sources of randomness, the sampling indicators $\{R_i\}_{i=1}^N$ and the assignment indicators $\{Z_i\}_{i=1}^n$. We write $\mathrm{E}\left[\cdot\right]$ for the expectation over both, $\mathrm{E}_\Pi\left[\cdot\right]$ for the expectation over the sampling process, and $\mathrm{E}_\Omega\left[\cdot\right]$ for the expectation over the assignment process conditional on the realized sample.

**Proposition 3.** *Under simple random sampling of $n$ units from $\mathcal{P}_N$ and complete random assignment among the sampled units,* $\mathrm{E}\left[\hat{\tau}\right] = \tau_{PATE}$.

*Proof.* We take the expectation over both stages of randomness by iterating, first over the assignment (conditional on the realized sample) and then over the sample,

$$\mathrm{E}\left[\hat{\tau}\right] = \mathrm{E}_\Pi\left[\mathrm{E}_\Omega\left[\hat{\tau}\right]\right].$$

Conditional on any realized sample, the $n$ sampled units undergo complete random assignment, so Proposition 1 applies *within* the sample and gives $\mathrm{E}_\Omega\left[\hat{\tau}\right] = \tau_{\text{SATE}}$, where $\tau_{\text{SATE}} = \frac{1}{n}\sum_{i=1}^N R_i \tau_i$ is the average effect among the sampled units. Substituting,

$$\mathrm{E}\left[\hat{\tau}\right] = \mathrm{E}_\Pi\left[\tau_{\text{SATE}}\right] = \mathrm{E}_\Pi\left[\frac{1}{n}\sum_{i=1}^{N} R_i \tau_i\right].$$

Because the individual effects $\tau_i$ are fixed and only the sampling indicators $R_i$ are random, the linearity of expectations gives

$$\mathrm{E}\left[\hat{\tau}\right] = \frac{1}{n}\sum_{i=1}^{N} \tau_i \ \mathrm{E}_\Pi\left[R_i\right].$$

It remains to find $\mathrm{E}_\Pi\left[R_i\right]$. Under simple random sampling of $n$ units from $N$, the event $\{R_i = 1\}$ plays exactly the role that $\{Z_i = 1\}$ played under complete random assignment of $n_T$ units from $n$. The proportion of samples that include unit $i$ is $\frac{n}{N}$, and under uniform sampling the probability of inclusion equals that proportion. The argument of Lemma 1, with $N$ and $n$ in place of $n$ and $n_T$, therefore gives $\mathrm{E}_\Pi\left[R_i\right] = \frac{n}{N}$ for every $i \in \{1, \ldots, N\}$. Substituting,

$$\mathrm{E}\left[\hat{\tau}\right] = \frac{1}{n}\sum_{i=1}^{N} \tau_i \left(\frac{n}{N}\right) = \frac{1}{N}\sum_{i=1}^{N} \tau_i = \tau_{\text{PATE}},$$

which proves the proposition. $\square$

This argument reuses the earlier results rather than repeating them. The inner expectation is handled by the complete random assignment proposition, and the outer expectation by the same counting argument as Lemma 1, now applied to sampling $n$ of $N$ units instead of assigning $n_T$ of $n$. The result also makes precise the sense in which the Difference-in-Means estimator answers two questions at once. With respect to the assignment alone, it is unbiased for the SATE, the average effect among the units actually studied. Once we also average over which units are sampled, it is unbiased for the PATE, the average effect in the superpopulation.

## References


Cox, D. R. (1958). *Planning of Experiments.* New York: John Wiley & Sons.

Freedman, D. A. (2009). *Statistical Models: Theory and Practice* (Revised ed.). Cambridge: Cambridge University Press.

Gerber, A. S. and D. P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York: W. W. Norton.

Rubin, D. B. (1980). Comment on "randomization analysis of experimental data: The Fisher randomization test" by D. Basu. *Journal of the American Statistical Association 75*(371), 591–593.

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association 81*(396), 961–962.