Variance of the Difference-in-Means Estimator*

Thomas Leavitt

1 Variance of Difference-in-Means estimator for the 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 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 estimator of the Difference-in-Means estimator's variance is well defined. Let the binary indicator variable $Z_i \in \{0, 1\}$ denote whether unit i is assigned to treatment ($Z_i = 1$) or control ($Z_i = 0$). The set $\Omega = \left\{\boldsymbol{z} : \sum_{i=1}^{n} z_i = n_T\right\}$ contains the possible values of $\boldsymbol{Z} = \left[Z_1, \dots, Z_n\right]^{\top}$. Under complete random assignment, the number of elements in the set Ω 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.

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

Adopting the terminology of Freedman (2009) and later Gerber and Green (2012), define a potential outcomes schedule as a vector-valued function, $\boldsymbol{y} : \Omega \mapsto \mathbb{R}^n$, which maps the set of assignments, Ω, to an n-dimensional vector of real numbers, $\mathbb{R}^n$. More intuitively, a potential outcomes schedule is a listing of how each study participant would have responded to any $\boldsymbol{z} \in \Omega$ that a random assignment process could produce. The vectors of potential outcomes are the elements in the image of the potential outcomes schedule, $\boldsymbol{y} : \Omega \mapsto \mathbb{R}^n$, and the individual potential outcomes for unit $i \in \{1, \ldots, n\}$ are the ith entries of each of the n-dimensional vectors of potential outcomes.

With $|\Omega|$ possible assignments, where $|\Omega| = \binom{n}{n_T}$ under complete random assignment, there are in principle $|\Omega|$ vectors of potential outcomes.[1] However, under the Stable Unit Treatment Value Assumption (SUTVA)[2] (Cox, 1958; Rubin, 1980, 1986), let y_{Ti} denote the common outcome value of unit i for all $\boldsymbol{z} \in \Omega$ with $z_i = 1$. Likewise, let y_{Ci} denote the common outcome value of unit i for all $\boldsymbol{z} \in \Omega$ with $z_i = 0$. The individual causal effect for unit i on the additive scale is $\tau_i = y_{Ti} - y_{Ci}$. The vectors $\boldsymbol{y_C}$ and $\boldsymbol{y_T}$ denote the collection of control and treatment potential outcomes, respectively, for all n units, and $\boldsymbol{\tau}$ denotes the collection of individual, additive effects for all n units. The observed outcome for unit $i \in \{1, \ldots, n\}$ is $Y_i = Z_i y_{Ti} + (1 - Z_i)\, y_{Ci}$, which is either y_{Ti} or y_{Ci} depending on whether the randomly selected $\boldsymbol{z} \in \Omega$ is with $z_i = 1$ or $z_i = 0$.

The target of interest is the Sample Average Treatment Effect (SATE), $\tau_{\text{SATE}} := n^{-1} \sum\limits_{i=1}^{n} \tau_i$. Define the Difference-in-Means estimator of τ_{SATE} under complete random assignment as

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

For the expectation and variance of this estimator in Equation (1) with respect to the SATE, I write $\text{E}_{\Omega}[\cdot]$ and $\text{Var}_{\Omega}[\cdot]$ to indicate that the expectation and variance pertain to only randomness of the assignment process.

1.2 Derivation of variance of Difference-in-Means estimator for the SATE

Proposition 1. *The variance of $\hat{\tau}$ for the τ_{SATE} under complete random assignment is*

$$\text{Var}_{\Omega}[\hat{\tau}] = \frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}, \tag{2}$$

[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.

where

$$S_n^2(\boldsymbol{y_T}) = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right)^2 \tag{3}$$

$$S_n^2(\boldsymbol{y_C}) = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right)^2 \tag{4}$$

$$S_n^2(\boldsymbol{\tau}) = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(\tau_i - \frac{1}{n}\sum_{i=1}^{n} \tau_i\right)^2 . \tag{5}$$

Proof. Building on Imbens and Rubin (2015, Chapter 6, Appendix A), we will break the proof into several steps:

Step 1: We will show that we can write the Difference-in-Means estimator in (1) in terms of *not* only Z_i, but also the centered treatment variable $A_i = Z_i - \mathrm{E}_\Omega[Z_i]$. Doing so will greatly simplify the subsequent algebra for our derivation of the variance.

Step 2: The potential outcomes, $\{y_{Ti}, y_{Ci}\}_{i=1}^n$, are fixed and observed outcomes inherit their randomness only from the treatment assignment variable. Thus, to derive the variance of the Difference-in-Means estimator (written now in terms of A_i), we will need to derive $\mathrm{E}_\Omega[A_i]$, $\mathrm{Var}_\Omega[A_i]$ and $\mathrm{Cov}\left[A_i, A_j\right]$ for $i \neq j$.

Step 3: We will rely on the expressions for $S_n^2(\boldsymbol{y_T})$ and $S_n^2(\boldsymbol{y_C})$ above, as well as on the following algebraic equivalence that we will need to prove:

$$S_n^2(\boldsymbol{\tau}) = \left(S_n^2(\boldsymbol{y_T}) - S_n^2(\boldsymbol{y_C}) - \frac{2}{n(n-1)}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right)\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right)\right). \tag{6}$$

Step 4: Once we have completed the three previous steps, we will rely only on the linearity of expectations, rules of variance and (often very messy) algebra.

Step 1

First, note that we can re-write the estimator as

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

Because $n = n_C + n_T$, note that $1 = \frac{n_C + n_T}{n}$, so instead of $(1 - Z_i)$ we can write $\left(\frac{n_C + n_T}{n} - Z_i\right)$:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{n}{n_T}Z_i y_{Ti} - \frac{n}{n_C}\left(\frac{n_C + n_T}{n} - Z_i\right) y_{Ci},$$

which after a little bit of algebra yields

$$\frac{1}{n}\sum_{i=1}^{n}\frac{n}{n_T}\left(\underbrace{Z_i - \frac{n_T}{n}}_{=A_i} + \frac{n_T}{n}\right) y_{Ti} - \frac{n}{n_C}\left[\frac{n_C}{n} - \left(\underbrace{Z_i - \frac{n_T}{n}}_{=A_i}\right)\right] y_{Ci}, \tag{7}$$

where A_i is the centered treatment variable, $A_i = Z_i - \mathrm{E}_\Omega\left[Z_i\right]$, because, under complete random assignment, $\mathrm{E}_\Omega\left[Z_i\right] = \frac{n_T}{n}$.

Therefore, we can re-write (7) as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{n}{n_T}\left(A_i + \frac{n_T}{n}\right) y_{Ti} - \frac{n}{n_C}\left(\frac{n_C}{n} - A_i\right) y_{Ci},$$

which, with a bit more algebra, is equivalent to

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} y_{Ti} - y_{Ci}}_{\tau_{\text{SATE}}} + \frac{1}{n}\sum_{i=1}^{n} A_i\left(\frac{n}{n_T}y_{Ti} + \frac{n}{n_C}y_{Ci}\right) = \tau_{\text{SATE}} + \frac{1}{n}\sum_{i=1}^{n} A_i\left(\frac{n}{n_T}y_{Ti} + \frac{n}{n_C}y_{Ci}\right)$$

Step 2

Thus far, we have shown that the Difference-in-Means estimator in (1) is equivalent to

$$\tau_{\text{SATE}} + \frac{1}{n}\sum_{i=1}^{n} A_i\left(\frac{n}{n_T}y_{Ti} + \frac{n}{n_C}y_{Ci}\right), \tag{8}$$

in which τ_{SATE}, n_C, n_T, n, and $\{y_{Ci}, y_{Ti}\}_{i=1}^{n}$ are all fixed quantities. The only random quantities in (8) are $\{A_i\}_{i=1}^{n}$.

Therefore, to derive the variance of (8), we now need to derive $\mathrm{E}_\Omega\left[A_i\right]$, $\mathrm{Var}_\Omega\left[A_i\right]$ and $\mathrm{E}_\Omega\left[A_i A_j\right]$ for $i \neq j$.

To do so, let's write the sample space of A_i as follows:

$$A_i = Z_i - \frac{n_T}{n} = \begin{cases} \frac{n_C}{n} & \text{if } Z_i = 1 \\ -\frac{n_T}{n} & \text{if } Z_i = 0 \end{cases} \tag{9}$$

and recall that, under complete random assignment, $\Pr(Z_i = 1) = \frac{n_T}{n}$ and $\Pr(Z_i = 0) = 1 - \frac{n_T}{n} = \frac{n}{n} - \frac{n_T}{n} = \frac{n - n_T}{n} = \frac{n_C}{n}$.

Derivation of $\mathrm{E}_\Omega[A_i]$**:** It is straightforward to see that $\mathrm{E}_\Omega[A_i] = \mathrm{E}_\Omega\left[Z_i - \frac{n_T}{n}\right] = \mathrm{E}_\Omega[Z_i] - \frac{n_T}{n} = \frac{n_T}{n} - \frac{n_T}{n} = 0$ or, equivalently,

$$\begin{aligned} \mathrm{E}_\Omega[A_i] &= \frac{n_C}{n}\Pr(Z_i = 1) - \frac{n_T}{n}\Pr(Z_i = 0) \\ &= \frac{n_C}{n}\frac{n_T}{n} - \frac{n_T}{n}\left(1 - \frac{n_T}{n}\right) \\ &= \frac{n_C}{n}\frac{n_T}{n} - \frac{n_T}{n}\frac{n_C}{n} \\ &= 0. \end{aligned}$$

Derivation of $\mathrm{Var}_\Omega[A_i]$**:** Note that $\mathrm{Var}_\Omega[A_i] = \mathrm{E}_\Omega\left[A_i^2\right] - \mathrm{E}_\Omega[A_i]^2 = \mathrm{E}_\Omega\left[A_i^2\right] - 0$. Thus,

$$\begin{aligned} \mathrm{Var}_\Omega[A_i] &= \mathrm{E}_\Omega\left[A_i^2\right] \\ &= \left(\frac{n_C}{n}\right)^2 \Pr(Z_i = 1) + \left(-\frac{n_T}{n}\right)^2 \Pr(Z_i = 0) \\ &= \frac{n_C^2}{n^2}\Pr(Z_i = 1) + \frac{n_T^2}{n^2}\Pr(Z_i = 0) \\ &= \frac{n_C^2}{n^2}\frac{n_T}{n} + \frac{n_T^2}{n^2}\frac{n_C}{n} \\ &= \frac{n_C^2 n_T}{n^3} + \frac{n_T^2 n_C}{n^3} \end{aligned}$$

$$
\begin{aligned}
&= \frac{n_C^2 n_T + n_T^2 n_C}{n^3} \\
&= \frac{n_C n_T \left(n_C + n_T\right)}{n^3} \\
&= \frac{n_C n_T \left(n\right)}{n^3} \\
&= \frac{n_C n_T}{n^2}.
\end{aligned}
$$

Derivation of $\text{Cov}\left[A_i, A_j\right]$ **for** $i \neq j$**:** Note that $\{A_i\}_{i=1}^{n}$ all have the same expected value — i.e., $\text{E}_\Omega\left[A_i\right] = \frac{n_T}{n}$ for all i — but are *not* independent. We therefore will need to derive $\text{Cov}\left[A_i, A_j\right]$ for $i \neq j$. To do so, first note that

$$
\begin{aligned}
\text{Cov}\left[A_i, A_j\right] &= \text{E}_\Omega\left[\left(A_i - \text{E}_\Omega\left[A_i\right]\right)\left(A_j - \text{E}_\Omega\left[A_j\right]\right)\right] \\
&= \text{E}_\Omega\left[\left(A_i - 0\right)\left(A_j - 0\right)\right] \\
&= \text{E}_\Omega\left[A_i A_j\right],
\end{aligned}
$$

which is easier to work with.

The possible values that $A_i A_j$ could take on are

$$
A_i A_j = \begin{cases}
\left(\frac{n_C}{n}\right)\left(\frac{n_C}{n}\right) = \frac{n_C^2}{n^2} & \text{if } Z_i = 1 \text{ and } Z_j = 1 \\
\left(-\frac{n_T}{n}\right)\left(\frac{n_C}{n}\right) = \frac{-n_T n_C}{n^2} & \text{if } Z_i = 0 \text{ and } Z_j = 1 \text{ or } Z_i = 1 \text{ and } Z_j = 0 \\
\left(-\frac{n_T}{n}\right)\left(-\frac{n_T}{n}\right) = \frac{n_T^2}{n^2} & \text{if } Z_i = 0 \text{ and } Z_j = 0.
\end{cases}
$$

Having derived the sample space of $A_i A_j$, we now need to derive the probabilities that correspond to each of the events in the sample space of $A_i A_j$, namely, $\Pr\left(Z_i = 1, Z_j = 1\right)$, $\Pr\left(Z_i = 0, Z_j = 1\right)$, $\Pr\left(Z_i = 1, Z_j = 0\right)$ and $\Pr\left(Z_i = 0, Z_j = 0\right)$.

Remember that Z_i and Z_j are *not* independent, which implies that, e.g., $\Pr\left(Z_i = 1, Z_j = 1\right) \neq \Pr\left(Z_i = 1\right)\Pr\left(Z_j = 1\right)$. Instead, we will appeal to the definition of joint probability in which

$\Pr\left(Z_i = z, Z_j = z'\right) = \Pr\left(Z_i = z\right)\Pr\left(Z_j = z' \mid Z_i = z\right)$:

$$\Pr\left(Z_i = z, Z_j = z'\right) = \begin{cases} \left(\frac{n_T}{n}\right)\left(\frac{n_T - 1}{n-1}\right) & \text{if } Z_i = 1 \text{ and } Z_j = 1 \\ \\ \left(\frac{n_T}{n}\right)\left(\frac{n_C}{n-1}\right) & \text{if } Z_i = 1 \text{ and } Z_j = 0 \\ \\ \left(\frac{n_C}{n}\right)\left(\frac{n_T}{n-1}\right) & \text{if } Z_i = 0 \text{ and } Z_j = 1 \\ \\ \left(\frac{n_C}{n}\right)\left(\frac{n_C}{n-1}\right) & \text{if } Z_i = 0 \text{ and } Z_j = 0. \end{cases}$$

We can therefore write the probability distribution function (PDF) of A_iA_j as

$$\Pr\left(A_iA_j\right) = \begin{cases} \left(\frac{n_T}{n}\right)\left(\frac{n_T - 1}{n-1}\right) = \frac{n_T\left(n_T - 1\right)}{n\left(n-1\right)} & \text{if } A_iA_j = \frac{n_C^2}{n^2} \\ \\ \left(\frac{n_T}{n}\right)\left(\frac{n_C}{n-1}\right) + \left(\frac{n_C}{n}\right)\left(\frac{n_T}{n-1}\right) = \frac{2\left(n_Tn_C\right)}{n\left(n-1\right)} & \text{if } A_iA_j = \frac{-n_Tn_C}{n^2} \\ \\ \left(\frac{n_C}{n}\right)\left(\frac{n_C}{n-1}\right) = \frac{n_C\left(n_C - 1\right)}{n\left(n-1\right)} & \text{if } A_iA_j = \frac{n_T^2}{n^2}, \end{cases}$$

which implies that

$$\mathrm{E}_\Omega\left[A_iA_j\right] = \frac{n_C^2}{n^2}\left(\frac{n_T\left(n_T-1\right)}{n\left(n-1\right)}\right) + \frac{-n_Tn_C}{n^2}\left(\frac{2\left(n_Tn_C\right)}{n\left(n-1\right)}\right) + \frac{n_T^2}{n^2}\left(\frac{n_C\left(n_C-1\right)}{n\left(n-1\right)}\right)$$

or equivalently, after some algebra,

$$\frac{-n_Tn_C}{n^2\left(n-1\right)}.$$

Step 3

In this step, we will prove the algebraic equivalence between the expression for $S_n^2(\boldsymbol{\tau})$ in Equation (5) above and

$$S_n^2(\boldsymbol{\tau}) = \left(S_n^2(\boldsymbol{y_T}) - S_n^2(\boldsymbol{y_C}) - \frac{2}{n\left(n-1\right)}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right)\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right)\right), \tag{10}$$

which will be valuable for our derivation of $\text{Var}_\Omega\left[\hat{\tau}\right]$ in **Step 4** to follow.

Recall from Equation (5) that

$$S_n^2(\boldsymbol{\tau}) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\tau_i - \frac{1}{n}\sum_{i=1}^{n}\tau_i\right)^2.$$

Hence,

$$\begin{aligned}
S_n^2(\boldsymbol{\tau}) &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\tau_i - \frac{1}{n}\sum_{i=1}^{n}\tau_i\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\underbrace{(y_{Ti} - y_{Ci})}_{=\tau_i} - \frac{1}{n}\sum_{i=1}^{n}\underbrace{(y_{Ti} - y_{Ci})}_{=\tau_i}\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n}(y_{Ti} - y_{Ci}) - \frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)^2,
\end{aligned}$$

which, after expanding and then simplifying, is equivalent to

$$\begin{aligned}
S_n^2(\boldsymbol{\tau}) = &\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)^2}_{=S_n^2(\boldsymbol{y_T})} + \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)^2}_{=S_n^2(\boldsymbol{y_C})} \\
&- \frac{2}{n-1}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right) \\
= &\, S_n^2(\boldsymbol{y_T}) + S_n^2(\boldsymbol{y_C}) - \frac{2}{n-1}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right).
\end{aligned}$$

Step 4

Now, to complete the derivation of its variance, note that the variance of the Difference-in-Means estimator based on the expression of the Difference-in-Means in (8) is:

$$\text{Var}_\Omega\left[\hat{\tau}\right] = \text{Var}_\Omega\left[\tau_{\text{SATE}} + \frac{1}{n}\sum_{i=1}^{n}A_i\left(\frac{n}{n_T}y_{Ti} + \frac{n}{n_C}y_{Ci}\right)\right]$$

$$= \frac{1}{n^2} \operatorname{Var}_\Omega \left[\sum_{i=1}^{n} A_i \underbrace{\left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right)}_{\text{constant}} \right],$$

which, since $\mathrm{E}_\Omega [A_i] = 0$, is equivalent to

$$\operatorname{Var}_\Omega [\hat{\tau}] = \frac{1}{n^2} \left(\mathrm{E}_\Omega \left[\left(\sum_{i=1}^{n} A_i \underbrace{\left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right)}_{\text{constant}} \right)^2 \right] - \mathrm{E}_\Omega \left[\sum_{i=1}^{n} A_i \underbrace{\left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right)}_{\text{constant}} \right]^2 \right)$$

$$= \frac{1}{n^2} \mathrm{E}_\Omega \left[\left(\sum_{i=1}^{n} A_i \left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right) \right)^2 \right].$$

Expanding the expression immediately above yields

$$\operatorname{Var}_\Omega [\hat{\tau}] = \frac{1}{n^2} \mathrm{E}_\Omega \left[\left(\sum_{i=1}^{n} A_i \left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right) \right)^2 \right]$$

$$= \frac{1}{n^2} \mathrm{E}_\Omega \left[\left(\sum_{i=1}^{n} A_i \left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right) \right) \left(\sum_{j=1}^{n} A_j \left(\frac{n}{n_T} y_{Tj} + \frac{n}{n_C} y_{Cj} \right) \right) \right]$$

$$= \frac{1}{n^2} \mathrm{E}_\Omega \left[\sum_{i=1}^{n} \sum_{j=1}^{n} A_i A_j \left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right) \left(\frac{n}{n_T} y_{Tj} + \frac{n}{n_C} y_{Cj} \right) \right].$$

In **Step 2** above, we showed that

$$\begin{cases} \mathrm{E}_\Omega \left[A_i A_j \right] = \dfrac{-n_T n_C}{n^2 (n-1)} & \text{when } i \neq j \\ \\ \mathrm{E}_\Omega \left[A_i A_j \right] = \mathrm{E}_\Omega \left[A_i^2 \right] = \operatorname{Var}_\Omega [A_i] = \dfrac{n_C n_T}{n^2} & \text{when } i = j, \end{cases}$$

so it follows that the Difference-in-Means estimator's variance is

$$\operatorname{Var}_\Omega [\hat{\tau}] = \frac{1}{n^2} \sum_{j=1}^{n} \left(\frac{n}{n_T} y_{Ti} + \frac{n}{n_C} y_{Ci} \right)^2 \underbrace{\mathrm{E}_\Omega \left[A_i^2 \right]}_{= \frac{n_C n_T}{n^2}}$$

$$+\sum_{i=1}^{n}\sum_{j\neq i}\left(\frac{n}{n_T}y_{Ti}+\frac{n}{n_C}y_{Ci}\right)\left(\frac{n}{n_T}y_{Tj}+\frac{n}{n_C}y_{Cj}\right)\underbrace{\mathrm{E}_\Omega\left[A_iA_j\right]}_{=\frac{-n_Tn_C}{n^2(n-1)}}.$$

Substituting the expressions we derived for $\mathrm{E}_\Omega\left[A_i^2\right]$ and $\mathrm{E}_\Omega\left[A_iA_j\right]$ and a lot of algebra yields

$$\begin{aligned}\mathrm{Var}_\Omega\left[\hat{\tau}\right] &= \frac{n_C}{nn_T(n-1)}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)^2+\frac{n_T}{nn_C(n-1)}\sum_{i=1}^{n}\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)^2\\ &+\frac{2}{n(n-1)}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right).\end{aligned}\tag{11}$$

Now recall that

$$S_n^2(\boldsymbol{y_T})=\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)^2 \quad\text{and}$$

$$S_n^2(\boldsymbol{y_C})=\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)^2,$$

which yields

$$\mathrm{Var}_\Omega\left[\hat{\tau}\right]=\frac{n_C}{nn_T}S_n^2(\boldsymbol{y_T})+\frac{n_T}{nn_C}S_n^2(\boldsymbol{y_C})+\frac{2}{n(n-1)}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right).$$

Finally, at long last, it follows that

$$\begin{aligned}\mathrm{Var}_\Omega\left[\hat{\tau}\right] &= \frac{n_C}{nn_T}S_n^2(\boldsymbol{y_T})+\frac{n_T}{nn_C}S_n^2(\boldsymbol{y_C})+\frac{2}{n(n-1)}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)\\ &= \frac{S_n^2(\boldsymbol{y_T})}{n_T}+\frac{S_n^2(\boldsymbol{y_C})}{n_C}-\frac{S_n^2(\boldsymbol{y_T})}{n}-\frac{S_n^2(\boldsymbol{y_C})}{n}+\frac{2}{n(n-1)}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)\\ &= \frac{S_n^2(\boldsymbol{y_T})}{n_T}+\frac{S_n^2(\boldsymbol{y_C})}{n_C}-\frac{1}{n}\underbrace{\left(S_n^2(\boldsymbol{y_T})-S_n^2(\boldsymbol{y_C})-\frac{2}{(n-1)}\sum_{i=1}^{n}\left(y_{Ti}-\frac{1}{n}\sum_{i=1}^{n}y_{Ti}\right)\left(y_{Ci}-\frac{1}{n}\sum_{i=1}^{n}y_{Ci}\right)\right)}_{=S_n^2(\boldsymbol{\tau})}.\end{aligned}$$

As we showed in **Step 3**, the third term in the expression immediately above is equal to S_τ^2, which leaves us with

$$\frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}. \tag{12}$$

□

The expression we just derived for $\text{Var}_\Omega[\hat{\tau}]$ is the *true* variance of the Difference-in-Means estimator. It is mathematically equivalent to Equation 3.4 in Gerber and Green (2012, 57), which differs from the expression in Equation (12) above because the variances and covariance of treated and control potential outcomes in Gerber and Green (2012, Equation 3.4, 57) use a denominator of n as opposed to $n-1$ as in Equations (3), (4) and (5) above. The corrolary below establishes this equivalence.

Corollary 1. *An equivalent expression for the finite sample variance of the Difference-in-Means estimator under complete random assignment is*

$$\text{Var}_\Omega[\hat{\tau}] = \frac{1}{n-1}\left(\frac{n_C\sigma_n^2(\boldsymbol{y_T})}{n_T} + \frac{n_T\sigma_n^2(\boldsymbol{y_C})}{n_C} + 2\sigma_n(\boldsymbol{y_C}, \boldsymbol{y_T})\right), \tag{13}$$

where

$$\sigma_n^2(\boldsymbol{y_T}) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right)^2 \tag{14}$$

$$\sigma_n^2(\boldsymbol{y_C}) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right)^2 \tag{15}$$

$$\sigma_n(\boldsymbol{y_C}, \boldsymbol{y_T}) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right)\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right). \tag{16}$$

Proof. Recall the expression for $\text{Var}_\Omega[\hat{\tau}]$ in Equation (11):

$$\begin{aligned}\text{Var}_\Omega[\hat{\tau}] = \frac{n_C}{nn_T(n-1)}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right)^2 + \frac{n_T}{nn_C(n-1)}\sum_{i=1}^{n}\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right)^2 \\ + \frac{2}{n(n-1)}\sum_{i=1}^{n}\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{n} y_{Ti}\right)\left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{n} y_{Ci}\right).\end{aligned}$$

Then, appealing to the definitions $\sigma_n^2(\boldsymbol{y_T})$, $\sigma_n^2(\boldsymbol{y_C})$ and $\sigma_n(\boldsymbol{y_C}, \boldsymbol{y_T})$ above, yields

$$\mathrm{Var}_\Omega\left[\hat{\tau}\right] = \frac{n_C}{n_T\left(n-1\right)}\sigma_n^2(\boldsymbol{y_T}) + \frac{n_T}{n_C\left(n-1\right)}\sigma_n^2(\boldsymbol{y_C}) + \frac{2}{\left(n-1\right)}\sigma_n(\boldsymbol{y_C}, \boldsymbol{y_T})$$

$$= \frac{1}{n-1}\left(\frac{n_C\sigma_n^2(\boldsymbol{y_T})}{n_T} + \frac{n_T\sigma_n^2(\boldsymbol{y_C})}{n_C} + 2\sigma_n(\boldsymbol{y_C}, \boldsymbol{y_T})\right),$$

which completes the proof. □

2 Variance of Difference-in-Means estimator for the Population Average Treatment Effect (PATE)

2.1 Setup

Thus far, we have considered estimation in only a finite sample. Now consider a superpopulation, $\mathcal{P}_N$, of size $N \geq n \geq 4$ and let the index $i \in \{1, \dots, N\}$ run over the N units in $\mathcal{P}_N$. Let n units from $\mathcal{P}_N$ be randomly selected into an experimental sample, $\mathcal{S}_n$, while the remaining $N - n$ units in $\mathcal{P}_N$ are unsampled. Of these n sampled units, $n_T \geq 2$ are randomly assigned to treatment and $n - n_T = n_C \geq 2$ are randomly assigned to control.

The binary indicator variable $R_i \in \{0, 1\}$ denotes whether individual unit i is included ($R_i = 1$) or excluded ($R_i = 0$) in the random sample from the superpopulation $\mathcal{P}_N$. Let the set $\Pi = \left\{\boldsymbol{r} : \sum_{i=1}^N r_i = n\right\}$ contain all possible values of $\boldsymbol{R} = \left[R_1, \dots, R_N\right]^\top$.

The target of interest is the Population Average Treatment Effect (PATE), $\tau_{\text{PATE}} \coloneqq N^{-1}\sum\limits_{i=1}^N \tau_i$. Define the Difference-in-Means estimator of τ_{PATE} as

$$\hat{\tau} \coloneqq \frac{1}{n_T}\sum_{i=1}^N R_i Z_i Y_i - \frac{1}{n_C}\sum_{i=1}^N R_i\left(1 - Z_i\right) Y_i. \tag{17}$$

Otherwise, the setup for inference of the τ_{PATE} is identical to the setup in Section 1.1 except that now, under simple random sampling of n units from a population of size N and complete random assignment with n_T treated units out of n sampled units, there are $\binom{N}{n}\binom{n}{n_T}$ possible assignments, which reflect the set of $\binom{n}{n_T}$ random assignments for any single realized sample and the $\binom{N}{n}$ different possible samples that could be realized. Hence, the potential outcomes schedule is now defined as a mapping from the set of $\binom{N}{n}\binom{n}{n_T}$ possible assignments to an N-dimensional vector of real numbers, $\mathbb{R}^N$, to reflect the potential outcomes of all N units in $\mathcal{P}_N$. The SUTVA assumption for this potential outcomes schedule is analogous to the SUTVA assumption in Section 1.1 above.

In addition, note that, by the law of total variance, the variance of the Difference-in-Means estimator of the PATE is

$$\operatorname{Var}\left[\hat{\tau}\right] = \operatorname{E}_{\Pi}\left[\operatorname{Var}_{\Omega}\left[\hat{\tau}\right]\right] + \operatorname{Var}_{\Pi}\left[\operatorname{E}_{\Omega}\left[\hat{\tau}\right]\right]. \tag{18}$$

As Equation (18) makes clear, the Difference-in-Means estimator of the PATE in (17) has two sources of randomness: $\{R_i\}_{i=1}^N$ and $\{Z_i\}_{i=1}^N$, which reflect a random sampling process and a random assignment process. By contrast, the Difference-in-Means estimator of the SATE in (1) has only one source of randomness, $\{Z_i\}_{i=1}^N$. For the overall expectation and variance of the Difference-in-Means estimator of the PATE, I simply write $\operatorname{E}[\cdot]$ and $\operatorname{Var}[\cdot]$. However, I write either $\operatorname{E}_{\Omega}[\cdot]$ and $\operatorname{Var}_{\Omega}[\cdot]$ or $\operatorname{E}_{\Pi}[\cdot]$ and $\operatorname{Var}_{\Pi}[\cdot]$ when the expectation and variance are taken over randomness of either the assignment process or sampling process.

2.2 Derivation of variance of Difference-in-Means estimator for the PATE

Proposition 2. *The variance of $\hat{\tau}$ for τ_{PATE} under simple random sampling from the units in $\mathcal{P}_N$ and complete random assignment among the units in $\mathcal{S}_n$ is*

$$\operatorname{Var}\left[\hat{\tau}\right] = \frac{N}{N-1}\left(\frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{\sigma_N^2(\boldsymbol{\tau})}{N}\right), \tag{19}$$

where

$$\sigma_N^2(\boldsymbol{y_T}) = \left(\frac{1}{N}\right)\sum_{i=1}^N\left(y_{Ti} - \frac{1}{N}\sum_{i=1}^N y_{Ti}\right)^2$$

$$\sigma_N^2(\boldsymbol{y_C}) = \left(\frac{1}{N}\right)\sum_{i=1}^N\left(y_{Ci} - \frac{1}{N}\sum_{i=1}^N y_{Ci}\right)^2$$

$$\sigma_N^2(\boldsymbol{\tau}) = \frac{1}{N}\sum_{i=1}^N\left(y_{Ti} - y_{Ci} - \frac{1}{N}\sum_{i=1}^N \tau_i\right)^2.$$

Proof. As mentioned above, the law of total variance implies that the variance of the Difference-in-Means estimator of the PATE is

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

Since $\hat{\tau}$ is unbiased for τ_{SATE} over Ω, we know that $\text{E}_{\Omega}\left[\hat{\tau}\right] = \tau_{\text{SATE}}$. From Proposition 1, we know that

$$\text{Var}_{\Omega}\left[\hat{\tau}\right] = \frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}.$$

Thus, it follows that

$$\begin{aligned}\text{Var}\left[\hat{\tau}\right] &= \text{E}_{\Pi}\left[\text{Var}_{\Omega}\left[\hat{\tau}\right]\right] + \text{Var}_{\Pi}\left[\text{E}_{\Omega}\left[\hat{\tau}\right]\right] \\ &= \text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}\right] + \text{Var}_{\Pi}\left[\tau_{\text{SATE}}\right].\end{aligned}$$

Therefore, we need to derive $\text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}\right]$ and $\text{Var}_{\Pi}\left[\tau_{\text{SATE}}\right]$.

Elementary theory from survey sampling (Cochran, 1977; Kish, 1965; Lohr, 2010) implies that

$$\text{Var}_{\Pi}\left[\frac{1}{n}\sum_{i=1}^{N} R_i\tau_i\right] = \frac{N-n}{(N-1)}\frac{\sigma_N^2(\boldsymbol{\tau})}{n}.$$

Now we only need to derive $\text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}\right]$:

$$\begin{aligned}\text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{y_T})}{n_T} + \frac{S_n^2(\boldsymbol{y_C})}{n_C} - \frac{S_n^2(\boldsymbol{\tau})}{n}\right] &= \text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{y_T})}{n_T}\right] + \text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{y_C})}{n_C}\right] - \text{E}_{\Pi}\left[\frac{S_n^2(\boldsymbol{\tau})}{n}\right] \\ &= \frac{1}{n_T}\text{E}_{\Pi}\left[S_n^2(\boldsymbol{y_T})\right] + \frac{1}{n_C}\text{E}_{\Pi}\left[S_n^2(\boldsymbol{y_C})\right] - \frac{1}{n}\text{E}_{\Pi}\left[S_n^2(\boldsymbol{\tau})\right],\end{aligned}$$

since, under simple random sampling with a fixed n and complete random assignment, n_T, n_C and n are all fixed constants.

Recalling the definitions of $S_n^2(\boldsymbol{y_T})$, $S_n^2(\boldsymbol{y_C})$ and $S_n^2(\boldsymbol{\tau})$ in (3), (4) and (5) in the finite sample context, we can re-write each as

$$S_N^2(\boldsymbol{y_T}) = \left(\frac{1}{n-1}\right)\sum_{i=1}^{N} R_i\left(y_{Ti} - \frac{1}{n}\sum_{i=1}^{N} R_i y_{Ti}\right)^2$$

$$S_N^2(\boldsymbol{y_C}) = \left(\frac{1}{n-1}\right) \sum_{i=1}^{N} R_i \left(y_{Ci} - \frac{1}{n}\sum_{i=1}^{N} R_i y_{Ci}\right)^2$$

$$S_N^2(\boldsymbol{\tau}) = \left(\frac{1}{n-1}\right) \sum_{i=1}^{N} R_i \left(\tau_i - \frac{1}{n}\sum_{i=1}^{N} R_i \tau_i\right)^2,$$

which make explicit that the randomness of $S_N^2(\boldsymbol{y_T})$, $S_N^2(\boldsymbol{y_C})$ and $S_N^2(\boldsymbol{\tau})$ (all of which would be fixed in a finite sample setting) stems from the N sample inclusion variables $\{R_i\}_{i=1}^N$.

Adapting Cochran (1977, Theorem 2.4), it follows that

$$\mathrm{E}\left[S_N^2(\boldsymbol{y_T})\right] = \frac{N}{N-1}\sigma_N^2(\boldsymbol{y_T})$$

$$\mathrm{E}\left[S_N^2(\boldsymbol{y_C})\right] = \frac{N}{N-1}\sigma_N^2(\boldsymbol{y_C})$$

$$\mathrm{E}\left[S_N^2(\boldsymbol{\tau})\right] = \frac{N}{N-1}\sigma_N^2(\boldsymbol{\tau}),$$

which implies that

$$\mathrm{E}_\Pi\left[\frac{S_N^2(\boldsymbol{y_T})}{n_T} + \frac{S_N^2(\boldsymbol{y_C})}{n_C} - \frac{S_N^2(\boldsymbol{\tau})}{n}\right] = \frac{N}{N-1}\left(\frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{\sigma_N^2(\boldsymbol{\tau})}{n}\right).$$

With these expressions for $\mathrm{E}_\Pi\left[\frac{S_N^2(\boldsymbol{y_T})}{n_T} + \frac{S_N^2(\boldsymbol{y_C})}{n_C} - \frac{S_N^2(\boldsymbol{\tau})}{n}\right]$ and $\mathrm{Var}_\Pi\left[\tau_{\text{SATE}}\right]$, it follows that

$$\begin{aligned}
\mathrm{Var}\left[\hat{\tau}\right] &= \mathrm{E}_\Pi\left[\mathrm{Var}_\Omega\left[\hat{\tau}\right]\right] + \mathrm{Var}_\Pi\left[\mathrm{E}_\Omega\left[\hat{\tau}\right]\right] \\
&= \frac{N}{N-1}\left(\frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{\sigma_N^2(\boldsymbol{\tau})}{n}\right) + \left(\frac{N-n}{N-1}\right)\frac{\sigma_N^2(\boldsymbol{\tau})}{n} \\
&= \frac{1}{N-1}\left(\frac{N\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{N\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{N\sigma_N^2(\boldsymbol{\tau})}{n} + \frac{(N-n)\,\sigma_N^2(\boldsymbol{\tau})}{n}\right) \\
&= \frac{1}{N-1}\left(\frac{N\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{N\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{N\sigma_N^2(\boldsymbol{\tau})}{n} + \frac{\sigma_N^2(\boldsymbol{\tau})N - \sigma_N^2(\boldsymbol{\tau})n}{n}\right)
\end{aligned}$$

$$= \frac{1}{N-1}\left(\frac{N\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{N\sigma_N^2(\boldsymbol{y_C})}{n_C} - \sigma_N^2(\boldsymbol{\tau})\right)$$

$$= \frac{1}{N-1}\left(\frac{N\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{N\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{N}{N}\sigma_N^2(\boldsymbol{\tau})\right)$$

$$= \frac{N}{N-1}\left(\frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{\sigma_N^2(\boldsymbol{\tau})}{N}\right).$$

□

As we can see from the expression in Equation (19), as $N \to \infty$,

$$\frac{N}{N-1}\left(\frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C} - \frac{\sigma_N^2(\boldsymbol{\tau})}{N}\right) \to \frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C}.$$

Hence, if the superpopulation is infinite, then

$$\text{Var}\left[\hat{\tau}\right] = \frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C}$$

and, so long as N is very large,

$$\text{Var}\left[\hat{\tau}\right] \approx \frac{\sigma_N^2(\boldsymbol{y_T})}{n_T} + \frac{\sigma_N^2(\boldsymbol{y_C})}{n_C}.$$

References

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Hoboken, NJ: John Wiley & Sons. 14, 15

Cox, D. R. (1958). *Planning of Experiments.* New York, NY: Wiley. 2

Freedman, D. A. (2009). *Statistical Models: Theory and Practice.* New York, NY: Cambridge University Press. 2

Gerber, A. S. and D. P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York, NY: W.W. Norton. 2, 11

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY: Cambridge University Press. 3

Kish, L. (1965). *Survey Sampling.* New York, NY: Wiley & Sons. 14

Lohr, S. L. (2010). *Sampling: Design and Analysis* (Second ed.). Boston, MA: Brooks/Cole. 14

Rubin, D. B. (1980). Comment on 'Randomization analysis of experimental data in the Fisher randomization test' by Basu, D. *Journal of the American Statistical Association 75*(371), 591–593. 2

Rubin, D. B. (1986). Which ifs have causal answers? (Comment on 'Statistics and causal inference' by Paul W. Holland). *Journal of the American Statistical Association 81*(396), 961–962. 2