Online Supplement

Thomas Leavitt[*] and Viviana Rivera-Burgos[†]

[*]Corresponding author: thomas.leavitt@baruch.cuny.edu, Marxe School of Public and International Affairs, Baruch College, City University of New York (CUNY)

[†]Political Science, Baruch College, City University of New York (CUNY)

1 Variance expressions and their estimators

As Neyman (1923) shows (see also Imbens and Rubin, 2015, pp. 87-92), the variance of the Difference-in-Means is

$$\text{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{y}\left(\boldsymbol{Z}\right)\right)\right] = \frac{S^2(1))}{n_1} + \frac{S^2(0))}{n_0} - \frac{S^2(1,0))}{N}, \tag{1}$$

where

$$\begin{aligned} S^2(1) &= \frac{1}{(N-1)} \sum_{i=1}^{N} \left(y_i(1) - \bar{y}(1)\right)^2 \\ S^2(0) &= \frac{1}{(N-1)} \sum_{i=1}^{N} \left(y_i(0) - \bar{y}(0)\right)^2 \\ S^2(1,0) &= \frac{1}{(N-1)} \sum_{i=1}^{N} \left(\tau_i - \tau\right)^2 . \end{aligned}$$

This variance can be conservatively estimated by dropping the third term in (1) and plugging in unbiased estimators for $S^2(1)$ and $S^2(0)$. This conservative variance estimator is

$$\widehat{\text{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{y}\left(\boldsymbol{Z}\right)\right)\right] = \frac{\hat{S}^2(1)}{n_1} + \frac{\hat{S}^2(0)}{n_0}, \tag{2}$$

where

$$\begin{aligned} \hat{S}^2(1) &= \frac{1}{(n_1-1)} \sum_{i=1}^{N} Z_i \left(y_i(1) - \frac{1}{n_1} \sum_{i=1}^{N} Z_i y_i(1)\right)^2 \quad \text{and} \\ \hat{S}^2(0) &= \frac{1}{(n_0-1)} \sum_{i=1}^{N} (1-Z_i) \left(y_i(0) - \frac{1}{n_0} \sum_{i=1}^{N} (1-Z_i) y_i(0)\right)^2 . \end{aligned}$$

This variance estimator is conservative in the sense that its expected value is always greater than or equal to the variance in (1).

The variance of the scaled Difference-in-Means is always at least as great as the variance in (1). In general, for an arbitrary constant, $a \in \mathbb{R}$, and random variable, X, $\text{Var}\left[aX\right] = a^2 \text{Var}\left[X\right]$. Since the inverse proportion of Always-Openers is always greater than or equal to 1, it follows that the variance of the scaled Difference-in-Means is always greater than or equal to the variance of the Difference-in-Means. More formally, the variance of the scaled Difference-in-Means is

$$\left(\frac{1}{\pi_{\text{AO}}^2}\right) \text{Var}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{y}\left(\boldsymbol{Z}\right)\right)\right], \tag{3}$$

and its conservative estimator is

$$\left(\frac{1}{\pi_{\text{AO}}^2}\right) \widehat{\text{Var}}\left[\hat{\tau}\left(\boldsymbol{Z}, \boldsymbol{y}\left(\boldsymbol{Z}\right)\right)\right]. \tag{4}$$

When researchers measure opening without error, the Difference-in-Means conditional on measured opening is equivalent to conditioning on a baseline covariate since, as Assumption 4 implies, measured opening is fixed across assignments. Hence, the Difference-in-Means conditional on measured opening is equivalent to the variance of the post-stratified Difference-in-Means in Miratrix et al. (2013). To express this variance, first let $W_{\text{AO}}(z)$ denote the number of Always-Openers randomized into either treatment or control:

$$W_{\text{AO}}(1) = \sum_{i=1}^{N} Z_i \tilde{m}_i \tag{5}$$

$$W_{\text{AO}}(0) = \sum_{i=1}^{N} (1 - Z_i) \tilde{m}_i. \tag{6}$$

Since the Difference-in-Means conditional on Always-Openers is undefined whenever $W_{\text{AO}}(1) < 1$ or $W_{\text{AO}}(0) < 1$, we follow Miratrix et al. (2013) by conditioning on the set of assignments in which the estimator is defined. We leave this conditioning implicit and write the variance of the Difference-in-Means conditional on opening as

$$\frac{1}{N_{\text{AO}}} \left(\beta_{\text{AO},1} S_{\text{AO}}^2(1) + \beta_{\text{AO},0} S_{\text{AO}}^2(0) + 2\gamma_{\text{AO}}(1,0)\right), \tag{7}$$

where

$$\beta_{\text{AO},z} \coloneqq \mathrm{E}\left[W_{\text{AO}}(1-z)/W_{\text{AO}}(z)\right]$$

$$S^2_{\text{AO}}(z) \coloneqq \frac{1}{N_{\text{AO}}-1}\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\}\left(y_i(z) - \bar{y}_{\text{AO}}(z)\right)^2$$

$$\gamma_{\text{AO}}(1,0) \coloneqq \frac{1}{N_{\text{AO}}-1}\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\}\left(y_i(1) - \bar{y}_{\text{AO}}(1)\right)\left(y_i(0) - \bar{y}_{\text{AO}}(0)\right)$$

$$\bar{y}_{\text{AO}}(z) \coloneqq \frac{1}{N_{\text{AO}}}\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\} y_i(z).$$

Both $\beta_{\text{AO},0}$ and $\beta_{\text{AO},1}$ can be estimated via numerical simulation of the randomization process. Each $S^2_{\text{AO}}(z)$ can be unbiasedly estimated by its sample analogue, $\hat{S}^2_{\text{AO}}(z)$, as in (9) and (10) below. However, $\gamma_{\text{AO}}(1,0)$ cannot be directly estimated. Instead, to be conservative, we use the expression for the covariance when potential outcomes are perfectly positively correlated:

$$\frac{\gamma_{\text{AO}}(1,0)}{\sqrt{S^2_{\text{AO}}(1)S^2_{\text{AO}}(0)}} = 1,$$

which implies

$$\gamma_{\text{AO}}(1,0) = \sqrt{S^2_{\text{AO}}(1)S^2_{\text{AO}}(0)}.$$

The AM-GM inequality then implies that

$$\sqrt{S^2_{\text{AO}}(1)S^2_{\text{AO}}(0)} \leq \frac{S^2_{\text{AO}}(1) + S^2_{\text{AO}}(0)}{2}, \tag{8}$$

which, after substituting the right hand side of (8) for $\gamma_{\text{AO}}(1,0)$, implies that

$$2\gamma_{\text{AO}}(1,0) \leq S^2_{\text{AO}}(1) + S^2_{\text{AO}}(0).$$

Therefore, the upper-bound of $\text{Var}\left[\hat{\tau}^{\text{Open}}\left(\boldsymbol{Z}, \tilde{\boldsymbol{m}}, \boldsymbol{y}(\boldsymbol{Z})\right)\right]$ is given by

$$\text{Var}\left[\hat{\tau}^{\text{Open}}\left(\boldsymbol{Z}, \tilde{\boldsymbol{m}}, \boldsymbol{y}(\boldsymbol{Z})\right)\right] \leq \frac{1}{N_{\text{AO}}}\left(\beta_{\text{AO},1} S^2_{\text{AO}}(1) + \beta_{\text{AO},0} S^2_{\text{AO}}(0) + S^2_{\text{AO}}(1) + S^2_{\text{AO}}(1)\right).$$

All terms in this quantity can be directly estimated via

$$\hat{S}^2_{\text{AO}}(1) = \left(\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = k\right\} Z_i - 1\right)^{-1} \sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = k\right\} Z_i \left(y_i(1) - \hat{\bar{y}}_{\text{AO}}(1)\right)^2 \text{ or} \tag{9}$$

$$\hat{S}^2_{\text{AO}}(0) = \left(\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\} (1 - Z_i) - 1\right)^{-1} \sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\} (1 - Z_i) \left(y_i(0) - \hat{\bar{y}}_{\text{AO}}(0)\right)^2 \tag{10}$$

where

$$\hat{\bar{y}}_{\text{AO}}(1) = \left(\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\} Z_i\right)^{-1} \sum_{i=1}^{n} \mathbb{1}\left\{\tilde{m}_i = 1\right\} Z_i y_i(1) \text{ and}$$

$$\hat{\bar{y}}_{\text{AO}}(0) = \left(\sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\} (1 - Z_i)\right)^{-1} \sum_{i=1}^{N} \mathbb{1}\left\{\tilde{m}_i = 1\right\} (1 - Z_i) y_i(0).$$

2 Worked examples

We now provide simple worked analyses of our proposed solutions to facilitate their implementation. The code for both of these examples is available at Leavitt and Rivera-Burgos (2024). To fix ideas, imagine a simple example of only 10 state legislators. Each legislator is emailed a constituency service request from an email address with a name and subject line independent of treatment assignment. Then within the body of the emails, the researcher randomized a signal of race, either Black (treatment) or white (control). In every email, both treated and control, the text also includes a signal of co-partisanship with whichever legislator to whom the email is sent.

Imagine that the observed experimental data are as follows.

Legislator	Treatment	Open	Reply
1	1	1	1
2	1	0	0
3	1	0	0
4	0	1	1
5	0	0	0
6	1	0	0
7	0	1	0
8	0	0	0
9	0	0	0
10	1	1	0

Table 1: Imaginary audit experiment among legislators under no racial discrimination in opening

Consistent with Assumption 2, no legislators reply to an email without first opening it. In addition, under the assumption of no racial discrimination in opening, all legislators with `Open` = 1 are Always-Openers. In this example, there are no replies to emails marked as unopened, so we would fail to reject the null hypothesis of no measurement error. We therefore conduct our analysis among only these legislators, i.e., as if our data were as follows.

Legislator	Treatment	Open	Reply
1	1	1	1
4	0	1	1
7	0	1	0
10	1	1	0

Table 2: Imaginary audit experiment among only Always-Open legislators

The post-stratified Difference-in-Means simply calculates the Difference-in-Means among only the units in Table 2 ignoring the third column. Hence, in this simple example, $\hat{\tau}^{\text{Open}}(\boldsymbol{Z}, \tilde{\boldsymbol{m}}, \boldsymbol{y}(\boldsymbol{Z})) = (0+1)/2 - (1+0)/2 = 0$. A conservative variance estimate (see Miratrix et al., 2013) is approximately 0.58.

We can also imagine another example as follows.

Legislator	Treatment	Open	Reply
1	1	1	1
2	1	0	1
3	1	1	0
4	0	1	1
5	0	0	0
6	1	0	1
7	0	1	0
8	0	0	0
9	0	0	1
10	1	1	0

Table 3: Imaginary audit experiment among legislators under no racial discrimination in opening

In this example, because there are three legislators who replied to the email but are measured as not opening the email, we would reject the null hypothesis of no measurement error. In addition, because there are 5 other units measured as opening, it follows that the proportion of Always-Openers is bounded between 0.8 and 1. In the baseline case in which we assume all units are Always-Reporters, the point estimate is 0.2 and the variance estimate is 0.12. As we decrease the proportion of Always-Openers from 1 to 0.9 and then to its lower-bound of 0.8, the point estimate increases to 0.22 and 0.25, respectively. The variance estimate also increases to 0.15 and 0.19, which would yield a wider CI in each case.

The aim of these two simple worked examples is to show that certain experimental designs permit exceedingly simple data analyses. For the purposes of the argument in this paper, the essential ingredient of such an experimental design is plausibly satisfying the no racial discrimination in opening assumption (and, although not absolutely necessary, the measurement of opening). Insofar as researchers design their audit experiments in thoughtful ways, then no complicated post-hoc statistical adjustments are required to infer relevant causal targets, like the ATE among Always-Openers.

References

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York, NY: Cambridge University Press. 1

Leavitt, T. and V. Rivera-Burgos (2024). Replication Data for: Audit experiments of racial discrimination and the importance of symmetry in exposure to cues. 4

Miratrix, L. W., J. S. Sekhon, and B. Yu (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(2), 369–396. 2, 5

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych 10*, 1–51. 1