Neymanian Repeated Sampling Inference in Completely Randomized Experiments

Peng Ding, Email: pengdingpku@berkeley.edu

1. Finite population quantities

Consider a completely randomized experiment with n units, where n_1 of them receive the treatment and n_0 of them receive the control. For unit $i = 1, \ldots, n$, we have potential outcomes $Y_i(1)$ and $Y_i(0)$, and individual effect $\tau_i = Y_i(1) - Y_i(0)$. The potential outcomes have finite population means

$$\bar{Y}(1) = n^{-1} \sum_{i=1}^{n} Y_i(1), \quad \bar{Y}(0) = n^{-1} \sum_{i=1}^{n} Y_i(0),$$

variances

$$S^2(1) = (n-1)^{-1} \sum_{i=1}^{n} \{Y_i(1) - \bar{Y}(1)\}^2, \quad S^2(0) = (n-1)^{-1} \sum_{i=1}^{n} \{Y_i(0) - \bar{Y}(0)\}^2,$$

and covariance

$$S(1,0) = (n-1)^{-1} \sum_{i=1}^{n} \{Y_i(1) - \bar{Y}(1)\}\{Y_i(0) - \bar{Y}(0)\}.$$

The individual effects have mean

$$\tau = n^{-1} \sum_{i=1}^{n} \tau_i = \bar{Y}(1) - \bar{Y}(0).$$

and variance

$$S^2(\tau) = (n-1)^{-1} \sum_{i=1}^{n} (\tau_i - \tau)^2.$$

We have the following relationship between the variances and covariance.

Lemma 1. $2S(1,0) = S^2(1) + S^2(0) - S^2(\tau)$.

Proof of Lemma 1. The conclusion follows from the simple fact that $2ab = a^2 + b^2 - (a-b)^2$ with $a = Y_i(1) - \bar{Y}(1)$ and $b = Y_i(0) - \bar{Y}(0)$. $\square$

These fixed quantities are functions of the Science Table. We are interested in estimating the average causal effect τ based on the data $(Z_i, Y_i)_{i=1}^n$ from a completely randomized experiment.

2. Neyman (1923)'s theorem

Based on the observed outcomes, we can calculate the sample means

$$\hat{\bar{Y}}(1) = n_1^{-1} \sum_{i=1}^n Z_i Y_i, \quad \hat{\bar{Y}}(0) = n_0^{-1} \sum_{i=1}^n (1 - Z_i) Y_i,$$

the sample variances

$$\hat{S}^2(1) = (n_1 - 1)^{-1} \sum_{i=1}^n Z_i \{Y_i - \hat{\bar{Y}}(1)\}^2, \quad \hat{S}^2(0) = (n_0 - 1)^{-1} \sum_{i=1}^n (1 - Z_i) \{Y_i - \hat{\bar{Y}}(0)\}^2.$$

But there are no sample versions of $S(1,0)$ and $S^2(\tau)$ because the potential outcomes $Y_i(1)$ and $Y_i(0)$ are never jointly observed. Neyman (1923) proved the following theorem.

Theorem 1. In a completely randomized experiment,

(1) the difference-in-means $\hat{\tau} = \hat{\bar{Y}}(1) - \hat{\bar{Y}}(0)$ is unbiased for τ, i.e., $E(\hat{\tau}) = \tau$;

(2) it has variance

$$\begin{aligned} \text{var}(\hat{\tau}) &= \frac{S^2(1)}{n_1} + \frac{S^2(0)}{n_0} - \frac{S^2(\tau)}{n} & (1) \\ &= \frac{n_0}{n_1 n} S^2(1) + \frac{n_1}{n_0 n} S^2(0) + \frac{2}{n} S(1,0), & (2) \end{aligned}$$

(3) the variance estimator

$$\hat{V} = \frac{\hat{S}^2(1)}{n_1} + \frac{\hat{S}^2(0)}{n_0}$$

is conservative for estimating $\text{var}(\hat{\tau})$:

$$E(\hat{V}) - \text{var}(\hat{\tau}) = \frac{S^2(\tau)}{n} \geq 0$$

with the equality holding if and only if $\tau_i = 0$ for all units.

A conservative 95% confidence interval based on Normal approximation is

$$\hat{\tau} \pm 1.96\sqrt{\hat{V}},$$

which is the same as the confidence interval for the standard two-sample problem asymptotically.

Some remarks:

- The point estimator is a standard one but it has a non-trivial variance under Neyman (1923)'s model. The variance is non-trivial because it not only depends on the finite population variances of the potential outcomes but also depends on the finite population variance of the individual effects, or, equivalently, the finite population covariance of the potential outcomes. Unfortunately, $S^2(\tau)$ and $S(1,0)$ are not identifiable from the data because $Y_i(1)$ and $Y_i(0)$ are never jointly observed.

- The variance estimator and confidence interval are conservative statistically in the sense of being overestimating or overcovering the true quantities. This may be not a good idea in some applications, for example, studies on side effects of drugs.

- It is a little puzzling that the more heterogeneous the individual effects are the smaller the variability of $\hat{\tau}$ is. Why?

- The Normal approximation is based on the finite population central limit theorem (Li and Ding 2017):
$$\frac{\hat{\tau} - \tau}{\sqrt{\mathrm{var}(\hat{\tau})}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1),$$
and $\hat{V}/\mathrm{var}(\hat{\tau}) \rightarrow C_{\mathrm{NF}}^2 \geq 1$ in probability.

3. Proof

First, the unbiasedness of $\hat{\tau}$ follows from the representation

$$\begin{aligned}\hat{\tau} &= n_1^{-1}\sum_{i=1}^n Z_iY_i - n_0^{-1}\sum_{i=1}^n (1-Z_i)Y_i \\ &= n_1^{-1}\sum_{i=1}^n Z_iY_i(1) - n_0^{-1}\sum_{i=1}^n (1-Z_i)Y_i(0)\end{aligned}$$

and the linearity of the expectation:

$$\begin{aligned}E(\hat{\tau}) &= E\left\{n_1^{-1}\sum_{i=1}^n Z_iY_i(1) - n_0^{-1}\sum_{i=1}^n (1-Z_i)Y_i(0)\right\} \\ &= n_1^{-1}\sum_{i=1}^n E(Z_i)Y_i(1) - n_0^{-1}\sum_{i=1}^n E(1-Z_i)Y_i(0) \\ &= n_1^{-1}\sum_{i=1}^n \frac{n_1}{n}Y_i(1) - n_0^{-1}\sum_{i=1}^n \frac{n_0}{n}Y_i(0) \\ &= n^{-1}\sum_{i=1}^n Y_i(1) - n^{-1}\sum_{i=1}^n Y_i(0) \\ &= \tau.\end{aligned}$$

Second, we can further write $\hat{\tau}$ as

$$\hat{\tau} = \sum_{i=1}^n Z_i\left\{\frac{Y_i(1)}{n_1} + \frac{Y_i(0)}{n_0}\right\} - n_0^{-1}\sum_{i=1}^n Y_i(0).$$

The variance of $\hat{\tau}$ follows from the lemma of simple random sampling:

$$\begin{aligned}\operatorname{var}(\hat{\tau}) &= \frac{n_1n_0}{n(n-1)}\sum_{i=1}^n Z_i\left\{\frac{Y_i(1)}{n_1} + \frac{Y_i(0)}{n_0} - \frac{\bar{Y}(1)}{n_1} + \frac{\bar{Y}(0)}{n_0}\right\}^2 \\ &= \frac{n_1n_0}{n(n-1)}\left[\frac{1}{n_1^2}\sum_{i=1}^n\{Y_i(1)-\bar{Y}(1)\}^2 + \frac{1}{n_0^2}\sum_{i=1}^n\{Y_i(0)-\bar{Y}(0)\}^2 + \frac{2}{n_1n_0}\sum_{i=1}^n\{Y_i(1)-\bar{Y}(1)\}\{Y_i(0)-\bar{Y}(0)\}\right] \\ &= \frac{n_0}{n_1n}S^2(1) + \frac{n_1}{n_0n}S^2(0) + \frac{2}{n}S(1,0).\end{aligned}$$

From Lemma 1, we can also write the variance as

$$
\begin{aligned}
\mathrm{var}(\hat{\tau}) &= \frac{n_0}{n_1 n} S^2(1) + \frac{n_1}{n_0 n} S^2(0) + \frac{1}{n}\{S^2(1) + S^2(0) - S^2(\tau)\} \\
&= \frac{S^2(1)}{n_1} + \frac{S^2(0)}{n_0} - \frac{S^2(\tau)}{n}.
\end{aligned}
$$

Third, because the treatment group is a simple random sample of size n_1 from the n units, the sample variance of $Y_i(1)$'s is unbiased for its population variance, i.e.,

$$
E\{\hat{S}^2(1)\} = S^2(1).
$$

Similarly, $E\{\hat{S}^2(0)\} = S^2(0)$. Therefore, $\hat{V}$ is unbiased for the first two terms in (1).

4. Neyman–Fisher controversy

By duality, we can test the following weak null hypothesis

$$
H_{0\mathrm{N}} : \bar{Y}(1) = \bar{Y}(0)
$$

based on

$$
\frac{\hat{\tau}}{\sqrt{\hat{V}}} \xrightarrow{\mathrm{d}} C_{\mathrm{NF}} \times \mathrm{N}(0,1),
$$

with $C_{\mathrm{NF}} \leq 1$. Pretending that $C_{\mathrm{NF}} = 1$, we will have a conservative test for $H_{0\mathrm{N}}$.

Whether $H_{0\mathrm{F}}$ or $H_{0\mathrm{N}}$ makes more sense? This is a famous controversy between Neyman and Fisher in a meeting at the Royal Statistical Society, UK (Neyman 1935; Sabbaghi and Rubin 2014; Ding 2016).

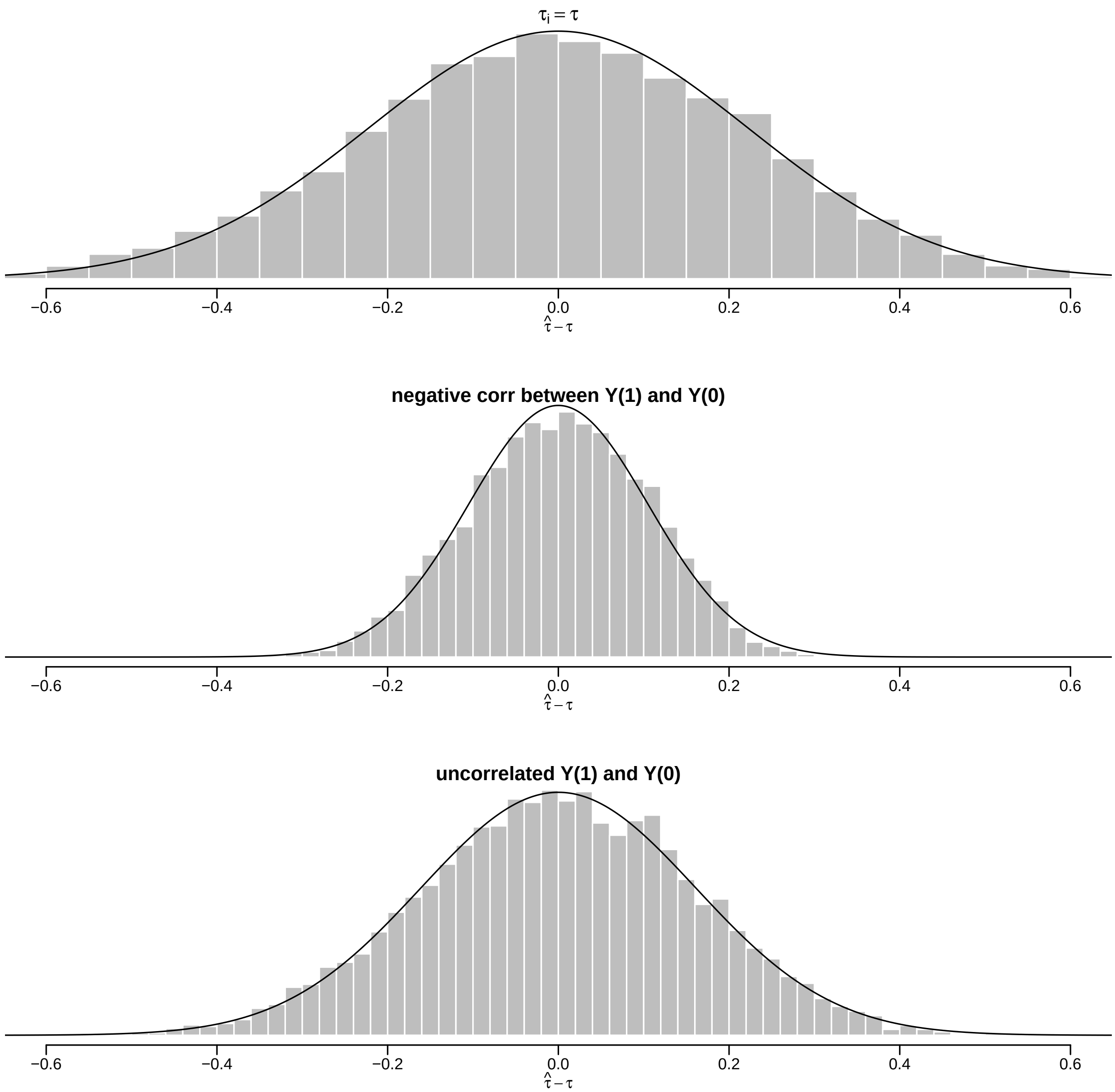
Recently, Ding and Dasgupta (2017) advocated using $\hat{\tau}/\sqrt{\hat{V}}$, i.e., the studentized statistic t, in the FRT so that the p-value is finite sample exact for $H_{0\mathrm{F}}$ and asymptotically valid for $H_{0\mathrm{N}}$.

5. Examples

Associations between potential outcomes are different but the marginal distributions are the same.

Different variances and coverage rates, but almost identical estimated variances.

	constant	negative	independent
var	0.036	0.007	0.020
estimated var	0.036	0.036	0.036
coverege rate	0.947	1.000	0.989



References

Ding, P. (2016). A paradox from randomization-based causal inference (with discussion). *Statistical Science*, 32:331–345.

Ding, P. and Dasgupta, T. (2017). A randomization-based perspective on analysis of variance: a test statistic robust to treatment effect heterogeneity. *Biometrika*, 105:45–56.

Li, X. and Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112:1759–1769.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles (with discussion). section 9 (translated). reprinted ed. *Statistical Science*, 5:465–472.

Neyman, J. (1935). Statistical problems in agricultural experimentation (with discussion). *Supplement to the Journal of the Royal Statistical Society*, 2:107–180.

Sabbaghi, A. and Rubin, D. B. (2014). Comments on the Neyman–Fisher controversy and its consequences. *Statistical Science*, 29:267–284.