Solutions to Exercises of Chapter 5

October 8, 2025
Ting-Chih Hung
nuit0815@gmail.com

Problem 5.1 (Covariate balance in the CRE)

Consider a completely randomized experiment (CRE) with n units, among which n_1 are assigned to treatment ($Z_i = 1$) and n_0 to control ($Z_i = 0$). Let $X_i \in \{1, \dots, K\}$ be the stratum membership of unit i, where K is the number of strata. Let $n_{[k]} \equiv \sum_{i=1}^n \mathbf{1}(X_i = k)$ be the number of units in stratum k, $\pi_{[k]} \equiv n_{[k]}/n$ be the proportion of units in stratum k, and $n_{[k]z} \equiv \sum_{i=1}^n \mathbf{1}(X_i = k, Z_i = z)$ be the number of units in stratum k assigned to treatment z.

We are asked to show that

$$\mathrm{E}\left(\frac{n_{[k]1}}{n_1} - \frac{n_{[k]0}}{n_0}\right) = 0. \tag{5.2}$$

This can directly be shown by the following calculation:

$$\begin{aligned}
&\mathrm{E}\left(\frac{n_{[k]1}}{n_1} - \frac{n_{[k]0}}{n_0}\right) \\
&= \mathrm{E}\left(\frac{n_{[k]1}}{n_1}\right) - \mathrm{E}\left(\frac{n_{[k]0}}{n_0}\right) && \text{(Linearity of } \mathrm{E}(\cdot)) \\
&= \frac{1}{n_1}\mathrm{E}\left(\sum_{i=1}^n Z_i \mathbf{1}(X_i = k)\right) - \frac{1}{n_0}\mathrm{E}\left(\sum_{i=1}^n (1 - Z_i)\mathbf{1}(X_i = k)\right) && \text{(Definition of } n_{[k]z}) \\
&= \frac{1}{n_1}\sum_{i=1}^n \mathbf{1}(X_i = k)\,\mathrm{E}(Z_i) - \frac{1}{n_0}\sum_{i=1}^n \mathbf{1}(X_i = k)\,\mathrm{E}(1 - Z_i) && \text{(Linearity of } \mathrm{E}(\cdot)) \\
&= \frac{1}{n_1}\sum_{i=1}^n \mathbf{1}(X_i = k)\frac{n_1}{n} - \frac{1}{n_0}\sum_{i=1}^n \mathbf{1}(X_i = k)\frac{n_0}{n} && (\mathrm{E}(Z_i) = n_1/n \text{ in the CRE}) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i = k) - \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i = k) && \text{(Simplification)} \\
&= 0.
\end{aligned}$$

Problem 5.2 (Consequence of the constant propensity score)

Following the notation in Problem 5.1, let $e_{[k]} \equiv n_{[k]1}/n_{[k]}$ be the propensity score of stratum k. Define

$$\begin{aligned}
\hat{\tau} &\equiv \frac{1}{n_1}\sum_{i=1}^n Z_i Y_i - \frac{1}{n_0}\sum_{i=1}^n (1 - Z_i) Y_i, \\
\hat{\tau}_{\mathrm{S}} &\equiv \sum_{k=1}^K \pi_{[k]} \hat{\tau}_{[k]},
\end{aligned}$$

where

$$\hat{\tau}_{[k]} \equiv \frac{1}{n_{[k]1}} \sum_{i:X_i=k} Z_i Y_i - \frac{1}{n_{[k]0}} \sum_{i:X_i=k} (1-Z_i) Y_i.$$

We are asked to show that when the propensity score is constant across strata,

$$\hat{\tau} = \hat{\tau}_{\mathrm{S}}. \tag{5.3}$$

Notice that if the propensity score is constant across strata, then $e_{[k]} = n_{[k]1}/n_{[k]} = n_1/n$ for all $k = 1, \ldots, K$. Therefore, we have

$$\begin{aligned}
&\hat{\tau}_{\mathrm{S}} \\
&\equiv \sum_{k=1}^{K} \pi_{[k]} \hat{\tau}_{[k]} && \text{(Definition of } \hat{\tau}_{\mathrm{S}}) \\
&= \sum_{k=1}^{K} \frac{n_{[k]}}{n} \left(\frac{1}{n_{[k]1}} \sum_{i:X_i=k} Z_i Y_i - \frac{1}{n_{[k]0}} \sum_{i:X_i=k} (1-Z_i) Y_i \right) && \text{(Definition of } \pi_{[k]} \text{ and } \hat{\tau}_{[k]}) \\
&= \frac{1}{n} \sum_{k=1}^{K} \left(\frac{n_{[k]}}{n_{[k]1}} \sum_{i:X_i=k} Z_i Y_i - \frac{n_{[k]}}{n_{[k]0}} \sum_{i:X_i=k} (1-Z_i) Y_i \right) && \text{(Factor out } 1/n) \\
&= \frac{1}{n} \sum_{k=1}^{K} \left(\frac{n}{n_1} \sum_{i:X_i=k} Z_i Y_i - \frac{n}{n_0} \sum_{i:X_i=k} (1-Z_i) Y_i \right) && \text{(Use } e_{[k]} = n_1/n) \\
&= \frac{1}{n_1} \sum_{k=1}^{K} \sum_{i:X_i=k} Z_i Y_i - \frac{1}{n_0} \sum_{k=1}^{K} \sum_{i:X_i=k} (1-Z_i) Y_i && \text{(Factor out } n/n_z) \\
&= \frac{1}{n_1} \sum_{i=1}^{n} Z_i Y_i - \frac{1}{n_0} \sum_{i=1}^{n} (1-Z_i) Y_i && \text{(Combine sums)} \\
&\equiv \hat{\tau}, && \text{(Definition of } \hat{\tau})
\end{aligned}$$

which completes the proof.

Problem 5.3 (Consquence of constant individual causal effects)

Problem 5.4 (Compare the CRE and SRE)

We are asked to compare the sampling variance of the difference-in-means estimator under the CRE and the stratified randomized experiment (SRE).

Suppose that we have a dataset from a SRE with K strata. Assume that $e_{[k]} = n_{[k]1}/n_{[k]}$ is constant across strata.

We are asked to show that when $n_{[k]}$ is large enough for all $k = 1, \ldots, K$, the difference between $\mathrm{Var}_{\mathrm{CRE}}(\hat{\tau})$ and $\mathrm{Var}_{\mathrm{SRE}}(\hat{\tau}_{\mathrm{S}})$ is always non-negative:

$$\begin{aligned}
&\mathrm{Var}_{\mathrm{CRE}}(\hat{\tau}) - \mathrm{Var}_{\mathrm{SRE}}(\hat{\tau}_{\mathrm{S}}) \\
&= \sum_{k=1}^{K} \left[\frac{\pi_{[k]}}{n_1} \left(\bar{Y}_{[k]}(1) - \bar{Y}(1) \right)^2 + \frac{\pi_{[k]}}{n_0} \left(\bar{Y}_{[k]}(0) - \bar{Y}(0) \right)^2 - \frac{\pi_{[k]}}{n} \left(\tau_{[k]} - \tau \right)^2 \right] \geq 0.
\end{aligned}$$

For the simplicity of notation, let $a_{[k]} \equiv \bar{Y}_{[k]}(1) - \bar{Y}(1)$ and $b_{[k]} \equiv \bar{Y}_{[k]}(0) - \bar{Y}(0)$. Then, $a_{[k]} - b_{[k]} = \tau_{[k]} - \tau$. Therefore, we have

$$
\begin{aligned}
&\operatorname{Var}_{\mathrm{CRE}}(\hat{\tau}) - \operatorname{Var}_{\mathrm{SRE}}(\hat{\tau}_{\mathrm{S}}) \\
&= \sum_{k=1}^{K} \left[\frac{\pi_{[k]}}{n_1} a_{[k]}^2 + \frac{\pi_{[k]}}{n_0} b_{[k]}^2 - \frac{\pi_{[k]}}{n} (a_{[k]} - b_{[k]})^2 \right] && \text{(Substitution)} \\
&= \sum_{k=1}^{K} \pi_{[k]} \left[\frac{1}{n_1} a_{[k]}^2 + \frac{1}{n_0} b_{[k]}^2 - \frac{1}{n} (a_{[k]}^2 - 2a_{[k]} b_{[k]} + b_{[k]}^2) \right] && \text{(Expand the square)} \\
&= \sum_{k=1}^{K} \pi_{[k]} \left[\left(\frac{1}{n_1} - \frac{1}{n} \right) a_{[k]}^2 + \left(\frac{1}{n_0} - \frac{1}{n} \right) b_{[k]}^2 + \frac{2}{n} a_{[k]} b_{[k]} \right] && \text{(Rearrangement)} \\
&= \sum_{k=1}^{K} \pi_{[k]} \left(\frac{n_0}{n n_1} a_{[k]}^2 + \frac{n_1}{n n_0} b_{[k]}^2 + \frac{2}{n} a_{[k]} b_{[k]} \right) && \text{(Simplification)} \\
&= \sum_{k=1}^{K} \frac{\pi_{[k]}}{n} \left(\frac{n_0}{n_1} a_{[k]}^2 + \frac{n_1}{n_0} b_{[k]}^2 + 2a_{[k]} b_{[k]} \right) && \text{(Factor out } 1/n\text{)} \\
&= \sum_{k=1}^{K} \frac{\pi_{[k]}}{n} \left(\sqrt{\frac{n_0}{n_1}} a_{[k]} + \sqrt{\frac{n_1}{n_0}} b_{[k]} \right)^2 && \text{(Rewrite the first two terms)} \\
&\geq 0. && \text{(Square is non-negative)}
\end{aligned}
$$

This completes the proof.

Problem 5.5 (From the CRE to the SRE)

In a CRE with n units, among which n_1 are assigned to treatment ($Z_i = 1$) and n_0 to control ($Z_i = 0$). Suppose that we have a categorical covariate $X_i \in \{1, \dots, K\}$, where K is the number of strata. Now, the quantities $n_{[k]1}$ and $n_{[k]0}$ will be random variables because the number of units in stratum k assigned to treatment z will vary across randomizations. Let $\mathbf{Z} \equiv (Z_1, \dots, Z_n)$ be the treatment assignment vector, and $\mathbf{n} \equiv \{n_{[k]1}, n_{[k]0}\}_{k=1}^{K}$ be the collection of the numbers of units in each stratum assigned to treatment and control.

We are asked to show that

$$
P_{\mathrm{CRE}}(\mathbf{Z} = \mathbf{z} \mid \mathbf{n}) = \frac{1}{\prod_{k=1}^{K} \binom{n_{[k]}}{n_{[k]1}}}.
$$

That is, conditional on $\mathbf{n}$, the treatment assignment $\mathbf{Z}$ is equivalent to that from a SRE.

Note that for any realization of $\mathbf{n}$, the number of possible treatment assignments $\mathbf{z}$ that satisfy $\mathbf{n}$ is exactly

$$
\prod_{k=1}^{K} \binom{n_{[k]}}{n_{[k]1}},
$$

and each of these treatment assignments has the same probability $1/\binom{n}{n_1}$ under the CRE. Therefore, we have

$$
P_{\mathrm{CRE}}(\mathbf{Z} = \mathbf{z} \mid \mathbf{n})
$$

$$
\begin{aligned}
&= \frac{P_{\mathrm{CRE}}(\mathbf{Z} = \mathbf{z}, \mathbf{n})}{P_{\mathrm{CRE}}(\mathbf{n})} && \text{(Definition of conditional probability)} \\
&= \frac{P_{\mathrm{CRE}}(\mathbf{Z} = \mathbf{z})}{P_{\mathrm{CRE}}(\mathbf{n})} && (\mathbf{n} \text{ is determined by } \mathbf{Z}) \\
&= \frac{\frac{1}{\binom{n}{n_1}}}{\frac{\prod_{k=1}^{K} \binom{n_{[k]}}{n_{[k]1}}}{\binom{n}{n_1}}} && \left(P_{\mathrm{CRE}}(\mathbf{Z} = \mathbf{z}) = 1/\tbinom{n}{n_1} \text{ and } P_{\mathrm{CRE}}(\mathbf{n}) = \textstyle\prod_{k=1}^{K} \tbinom{n_{[k]}}{n_{[k]1}}/\tbinom{n}{n_1}\right) \\
&= \frac{1}{\prod_{k=1}^{K} \binom{n_{[k]}}{n_{[k]1}}},
\end{aligned}
$$

which completes the proof.

Problem 5.6 (More FRTs for Section 5.2.2)

Acronyms

CRE	completely randomized experiment. 1–3
SRE	stratified randomized experiment. 2, 3