Chapter 10 Pairwise Randomized Experiments

Donald B. Rubin

Yau Mathematical Sciences Center, Tsinghua University

August 19, 2021

Introduction

This chapter considers a special case of the SRE.

Each stratum contains exactly two units, with one randomly selected to be assigned to the treatment group, and the other one assigned to the control group.

The design is known as a *pairwise randomized experiment* (PRE) or *paired comparison*.

There are two features of this design that warrant special attention.

1. The Neyman sampling variance estimator cannot be used as that estimator requires the presence of at least two units assigned to each treatment in each stratum.
2. Each stratum has the same proportion of treated units, which allows us to analyze the within-stratum estimates symmetrically; the natural estimator for the average treatment effect weights each stratum equally.

Introduction

As in the case of SRE, the motivation in a PRE is that of *a priori* removing the assignment vectors that are expected to lead to less informative inferences.

This argument relies on the within-pair variation in potential outcomes being small relative to the between-pair variation. Often the assignment to pairs is based on covariates.

Units are matched to other units based on their similarity in these covariates.

Suppose, for example, that the treatment is an expensive surgical procedure for a relatively common condition. It may not be financially feasible to apply the treatment to many individuals.

Introduction

To increase the precision of an experiment, it may, in such cases, be sensible to randomly draw J individuals from the target population of individuals who have the condition for which the surgery may be beneficial.

Then, for each of these J individuals, find a matching individual in the same population, as similar as possible to the original unit in terms of the characteristics that may be correlated with potential outcomes and efficacy of the treatment.

If the population is relatively large, it may be possible to get very close matches with respect to a large number of characteristics.

Given these J matched pairs, one can then conduct a PRE by randomly selecting one member of each pair to be assigned to the active treatment.

Introduction

In this lecture we will discuss the analyses of such PRE using the FEP, Neyman's repeated sampling, as well as regression and model-based inference.

The data to illustrate the concepts discussed is from a randomized experiment designed by Samuel Ball, Gary Bogatz, Donald Rubin, and Albert Beaton (1973) to evaluate "The Electric Company," an educational television program aimed at improving reading skills for young children, somewhat similar to Sesame Street.

The Children's Television Workshop Experiment

The experiment was conducted in two locations, Youngstown, Ohio, and Fresno, California where the Electric Company was not broadcast on local stations. In each location a number of schools was selected.

Here we focus on the data from Youngstown where two first grade classes from each of eight schools participated in the experiment.

The data for the 16 classes for the Youngstown location from this experiment are displayed in Table 10.1

Table 10.1: Data from Youngstown Children's Television Workshop Experiment

Pair G_i	Treatment W_i	Pre-test Score X_i	Post-test Score Y_i^{obs}	Normalized Rank Post-test Score R_i
1	0	12.9	54.6	-7.5
1	1	12.0	60.6	2.5
2	0	15.1	56.5	-4.5
2	1	12.3	55.5	5.5
3	0	16.8	75.2	0.5
3	1	17.2	84.8	4.5
4	0	15.8	75.6	1.5
4	1	18.9	101.9	7.5
5	0	13.9	55.3	-6.5
5	1	15.3	70.6	-1.5
6	0	14.5	59.3	-3.5
6	1	16.6	78.4	2.5
7	0	17.0	87.0	5.5
7	1	16.0	84.2	3.5
8	0	15.8	73.7	-0.5
8	1	20.1	108.6	7.5

Pairwise Randomized Experiments

The number of units, N, is even, the number of strata is $J = N/2$, with one treated unit and one control unit in each stratum so that each stratum is a pair.

Let G_i be the variable indicating the pair, with $P_i \in \{1, \ldots, N/2\}$, where P_i can be thought of as a function of covariates.

Within each pair there are $\begin{pmatrix} N(j) \\ N_t(j) \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2$ possible assignments, so that

$$p(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{j=1}^{N/2} \begin{pmatrix} N(j) \\ N_{tj} \end{pmatrix}^{-1} = \prod_{j=1}^{N/2} \frac{1}{2} = 2^{-N/2}, \qquad \text{for } \mathbf{W} \in \mathbb{W}^+,$$

where

$$\mathbb{W}^+ = \left\{ \mathbf{W} \,\middle|\, \sum_{i:G_i=j} W_i = 1 \text{ for } j = 1, \ldots, N/2 \right\}.$$

Pairwise Randomized Experiments

For convenience label the two units within a pair as units "A" and "B". Then, for all pairs $j = 1, \ldots, N/2$, let $(Y_{j,A}(0), Y_{j,A}(1))$ and $(Y_{j,B}(0), Y_{j,B}(1))$ be the potential outcomes for units A and B respectively in pair j, and let $W_{j,A}$ and $W_{j,B}$ be the treatment indicators for these units.

As one unit in each pair is randomly assigned 'treatment' while the other is a 'control' $W_{j,A} = 1 - W_{j,B}$, with $\Pr(W_{j,A} = 1 | \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}) = 1/2$.

Define also

$$Y_{j,A}^{\text{obs}} = \begin{cases} Y_{j,A}(0) & \text{if } W_{j,A} = 0, \\ Y_{j,A}(1) & \text{if } W_{j,A} = 1, \end{cases} \quad \text{and} \quad Y_{j,B}^{\text{obs}} = \begin{cases} Y_{j,B}(0) & \text{if } W_{j,A} = 1, \\ Y_{j,B}(1) & \text{if } W_{j,A} = 0, \end{cases}$$

Pairwise Randomized Experiments

The average treatment effect within pair j is $\tau_{\text{pair}}(j)$,

$$\tau_{\text{pair}}(j) = \frac{1}{2}\sum_{i:G_i=j}\Big(Y_i(1) - Y_i(0)\Big) = \frac{1}{2}\Big((Y_{j,A}(1) - Y_{j,A}(0)) + (Y_{j,B}(1) - Y_{j,B}(0))\Big).$$

The finite sample average treatment effect is

$$\tau_{\text{S}} = \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i(1) - Y_i(0)\Big) = \frac{2}{N}\sum_{j=1}^{N/2}\tau_{\text{pair}}(j).$$

Also define the pair of observed variables, one treated and one control from each pair:

$$Y_{j,c}^{\text{obs}} = \begin{cases} Y_{j,A}^{\text{obs}} & \text{if } W_{i,A} = 0, \\ Y_{j,B}^{\text{obs}} & \text{if } W_{i,A} = 1, \end{cases} \quad \text{and} \quad Y_{j,t}^{\text{obs}} = \begin{cases} Y_{j,B}^{\text{obs}} & \text{if } W_{i,A} = 0, \\ Y_{j,A}^{\text{obs}} & \text{if } W_{i,A} = 1. \end{cases}$$

Table 10.2: Potential Outcomes and Covariates from Children's Television Workshop Experiment

	Unit A					Unit B				
Pair	$Y_{i,A}(0)$	$Y_{i,A}(1)$	$W_{i,A}$	$Y_{i,A}^{\text{obs}}$	$X_{i,A}$	$Y_{i,B}(0)$	$Y_{i,B}(1)$	$W_{i,B}$	$Y_{i,B}^{\text{obs}}$	$X_{i,B}$
1	54.6	?	0	54.6	12.9	?	60.6	1	60.6	12.0
2	56.5	?	0	56.5	15.1	?	55.5	1	55.5	13.9
3	75.2	?	0	75.2	16.8	?	84.8	1	84.8	17.2
4	76.6	?	0	75.6	15.8	?	101.9	1	101.9	18.9
5	55.3	?	0	55.3	13.9	?	70.6	1	70.6	15.3
6	59.3	?	0	59.3	14.5	?	78.4	1	78.4	16.6
7	87.0	?	0	87.0	17.0	?	84.2	1	84.2	16.0
8	73.7	?	0	73.7	15.8	?	108.6	1	108.6	20.1

Fisher's Exact P-values in PRE:s

Let us focus in this discussion on the usual Fisher null hypothesis

$$H_0: \quad Y_i(0) = Y_i(1), \quad \text{for all } i = 1, \dots, N.$$

With the assignment mechanism fully known, for any fixed statistic, one can derive the randomization distribution and thus calculate the corresponding p-value.

An obvious statistic is:

$$T^{\text{avg}} = \left| \frac{1}{J} \sum_{j=1}^{J} \left(Y_{j,t}^{\text{obs}} - Y_{j,c}^{\text{obs}} \right) \right|$$

$$= \left| \frac{1}{J} \sum_{j=1}^{J} \left(W_{i,A} \cdot \left(Y_{j,A}^{\text{obs}} - Y_{j,B}^{\text{obs}} \right) + (1 - W_{i,A}) \cdot \left(Y_{j,B}^{\text{obs}} - Y_{j,A}^{\text{obs}} \right) \right) \right|.$$

Fisher's Exact P-values in PRE:s

Because each pair has a single treated and a single control unit, this also equals the 'traditional' difference in means estimator: $T^{\text{avg}} = \left| \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \right|$.

The p-value for this statistic will be different than when this statistic is used under a CRE because here the randomization distribution is based on the assignment mechanism corresponding to a PRE, leading to fewer elements in $\mathbb{W}^+$ than under complete randomization.

Alternative statistics include the average of within-pair differences in logarithms or other transformations of the basic outcomes, such as ranks.

Fisher's Exact P-values in PRE:s

To calculate the rank statistic, let R_i be the rank of Y_i^{obs} among the N values $Y_1^{\text{obs}}, \ldots, Y_N^{\text{obs}}$, normalized to have mean zero, and let $R_{j,A}$ and $R_{j,B}$ be the rank of the A and B units in pair j, among all N units.

The ranks for the sixteen classes are displayed in the last column in Table 10.1.

The rank statistic is

$$T^{\text{rank}} = \left|\overline{R}_t - \overline{R}_c\right| = \left|\frac{1}{J}\sum_{j=1}^{J}\left(W_{j,A}\cdot(R_{j,A} - R_{j,B}) + (1 - W_{j,A})\cdot(R_{j,B} - R_{j,A})\right)\right|.$$

Using ranks in PRE:s has the same advantages as using ranks in CRE:s, namely reducing the sensitivity to outliers.

Fisher's Exact P-values in PRE:s

Another statistic that is specific to PRE:s is based on the average within-pair rank of the observed outcomes.

$$T^{\text{rank,pair}} = \left| \frac{2}{N} \sum_{j=1}^{N/2} \left(\mathbf{1}_{Y_{j,1}^{\text{obs}} > Y_{j,0}^{\text{obs}}} - \mathbf{1}_{Y_{j,1}^{\text{obs}} < Y_{j,0}^{\text{obs}}} \right) \right|,$$

where $\mathbf{1}_{Y_{j,1}^{\text{obs}} > Y_{j,0}^{\text{obs}}}$ and $\mathbf{1}_{Y_{j,1}^{\text{obs}} < Y_{j,0}^{\text{obs}}}$ are indicators for whether the observed outcome for the 'treated' ('control') is larger than the observed outcome for the 'control' ('treated').

When there is substantial variation in the level of the outcomes between the pairs, $T^{\text{rank,pair}}$ has more power than the statistic T^{rank} against alternatives under which the treatment effect is constant.

Fisher's Exact P-values in PRE:s

Although the p-value is only valid for a single statistic, for illustrative purposes, we do the analysis of testing H_0 for three statistics.

$$T^{\text{avg}} = 13.4, \;\; p = 0.031, \;\; T^{\text{rank}} = 3.8, \;\; p = 0.031 \;\; \texttt{and} \;\; T^{\text{rank,pair}} = 0.5, \;\; p = 0.145.$$

The reason that the p-value for the within-pair rank statistic is larger than for the other statistics is that for the two pairs where the outcome for the treated unit is less than the outcome for the control unit in the pair, the difference in outcomes is small.

These small differences do not affect the average difference much, but they do matter for the within-pair rank statistic.

The other two p-values suggest that the television program did affect reading ability at conventional significance levels.

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Consider first the analysis of the average treatment effect in a single pair. The obvious estimator for the average treatment effect in pair j, $\tau_{\text{pair}}(j)$, is,

$$\hat{\tau}^{\text{pair}}(j) = Y_{j,t}^{\text{obs}} - Y_{j,c}^{\text{obs}} = \sum_{i:G_i=j} (2 \cdot W_i - 1) \cdot Y_i^{\text{obs}}.$$

The values of $\hat{\tau}^{\text{pair}}(j)$ for the eight pairs are displayed in the next slide.

Table 10.3: Observed Outcome Data from Children's Television Workshop Experiment by Pair

Pair	Outcome for Control Unit	Outcome for Treated Unit	Difference ($\hat{\tau}_{\text{pair}}(j)$)
1	54.6	60.6	6.0
2	56.5	55.5	-1.0
3	75.2	84.8	9.6
4	75.6	101.9	26.3
5	55.3	70.6	15.3
6	59.3	78.4	19.1
7	87.0	84.2	-2.8
8	73.7	108.6	34.9
mean	67.2	80.6	13.4
sample std	12.2	18.6	13.1

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Next, let us consider inference, first for the within-pair average treatment effect $\tau_{\text{pair}}(j)$.

For each pair we have a CRE with two units of which one unit is assigned to active treatment.

From the results in Chapter 6 it follows that $\hat{\tau}^{\text{pair}}(j)$ is unbiased for $\tau_{\text{pair}}(j)$ and that its sampling variance, based on the randomization distribution, is equal to

$$\mathbb{V}_W(\hat{\tau}^{\text{pair}}(j)) = \frac{S_c(j)^2}{N_c(j)} + \frac{S_t^2(j)}{N_t(j)} - \frac{S_{ct}(j)^2}{N(j)}.$$

With $N(j) = 2$ and $N_c(j) = N_t(j) = 1$, this expression simplifies to

$$\mathbb{V}_W(\hat{\tau}^{\text{pair}}(j)) = S_c(j)^2 + S_t^2(j) - \frac{S_{ct}(j)^2}{2}.$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

The within-pair variances can be written as

$$S_c^2(j) = \sum_{i:G_i=j} \left(Y_i(0) - \overline{Y}_j(0) \right)^2 = \frac{1}{2} \cdot \left(Y_{j,A}(0) - Y_{j,B}(0) \right)^2 ,$$

$$S_t^2(j) = \sum_{i:P_i=j} \left(Y_i(1) - \overline{Y}_j(1) \right)^2 = \frac{1}{2} \cdot \left(Y_{j,A}(1) - Y_{j,B}(1) \right)^2 ,$$

and

$$S_{ct}^2(j) = \frac{1}{2} \cdot \left((Y_{j,A}(1) - Y_{j,A}(0)) - (Y_{j,B}(1) - Y_{j,B}(0)) \right)^2 ,$$

where

$$\overline{Y}_j(0) = \frac{1}{2} \cdot \left(Y_{j,A}(0) + Y_{j,B}(0) \right) \qquad \text{and} \quad \overline{Y}_j(1) = \frac{1}{2} \cdot \left(Y_{j,A}(1) + Y_{j,B}(1) \right).$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

If the primary interest is in the finite sample average treatment effect, τ_{S}, that is, the within-pair average treatment effect averaged over the $N/2$ pairs,

$$\tau_{\mathrm{S}} = \frac{1}{N/2} \sum_{j=1}^{N/2} \tau_{\mathrm{pair}}(j),$$

the natural estimator is

$$\hat{\tau}^{\mathrm{dif}} = \frac{1}{N/2} \sum_{j=1}^{N/2} \hat{\tau}^{\mathrm{pair}}(j) = \left(\overline{Y}_t^{\mathrm{obs}} - \overline{Y}_c^{\mathrm{obs}} \right). \tag{1}$$

By unbiasedness of the within-pair estimators, $\hat{\tau}^{\mathrm{dif}}$ is unbiased for τ_{S} and

$$\mathbb{V}_W(\hat{\tau}^{\mathrm{dif}}) = \frac{1}{(N/2)^2} \sum_{j=1}^{N/2} \left(S_c^2(j) + S_t^2(j) - \frac{S_{ct}^2(j)}{2} \right).$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

So far the discussion is exactly analogous to the discussion for SRE.

In a CRE (and similarly, within a stratum in the SRE) the standard estimator for the sampling variance for the observed difference in treatment and control averages is

$$\hat{\mathbb{V}}^{\text{neyman}}\left(\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right) = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t},$$

with

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i(0) - \overline{Y}_c^{\text{obs}}\right)^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right)^2,$$

and analogously

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}}\right)^2.$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Because $N_c = N_t = 1$, these estimators, s_c^2 and s_t^2, cannot be used, thus disabling the standard estimator for the sampling variance.

One solution to this problem is to assume that the treatment effect is constant and additive, not only within pairs, but also across pairs.

Then it follows that the within-pair sampling variance is

$$\mathbb{V}_W(\hat{\tau}^{\text{pair}}(j)) = 2 \cdot S^2(j), \qquad \text{where} \;\; S^2(j) = S_c^2(j) = S_t^2(j).$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Moreover, if the treatment effect is constant across pairs, $\tau_{\text{pair}}(j) = \tau_{\text{S}}$ for all j, then $S(j)^2 = S^2$ for all j, and

$$\mathbb{V}_W\left(\hat{\tau}^{\text{dif}}\right) = \frac{1}{(N/2)^2}\sum_{j=1}^{N/2}\left(S_c^2(j) + S_t^2(j) - \frac{S_{ct}^2(j)}{2}\right) = \frac{4}{N}\cdot S^2,$$

which can be estimated as :

$$\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right) = \frac{4}{N\cdot(N-2)}\cdot\sum_{j=1}^{N/2}\left(\hat{\tau}^{\text{pair}}(j) - \hat{\tau}^{\text{dif}}\right)^2.$$

If there is heterogeneity in the treatment effects, then this estimator is upwardly biased and the corresponding confidence intervals will be conservative in the usual sense.

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Theorem

Suppose we have J pairs of units, and randomly assign one unit from each pair to the active treatment and the other unit to the control treatment. Then, (i) $\hat{\tau}^{\mathrm{dif}}$ is unbiased for τ_{S}, (ii) the sampling variance of $\hat{\tau}^{\mathrm{dif}}$ is

$$\mathbb{V}_W\left(\hat{\tau}^{\mathrm{dif}}\right) = \frac{1}{N^2}\sum_{j=1}^{N/2}\Big(Y_{j,A}(0)+Y_{j,A}(1)-\Big(Y_{j,B}(0)+Y_{j,B}(1)\Big)\Big)^2,$$

and (iii) the estimator for the sampling variance

$$\hat{\mathbb{V}}^{\mathrm{pair}}\left(\hat{\tau}^{\mathrm{dif}}\right) = \frac{4}{N\cdot(N-2)}\cdot\sum_{j=1}^{N/2}\left(\hat{\tau}^{\mathrm{pair}}(j)-\hat{\tau}^{\mathrm{dif}}\right)^2,$$

satisfies

$$\mathbb{E}\left[\hat{\mathbb{V}}^{\mathrm{pair}}\left(\hat{\tau}^{\mathrm{dif}}\right)\right] = \mathbb{V}_W(\hat{\tau}^{\mathrm{dif}}) + \frac{4}{N\cdot(N-2)}\cdot\sum_{j=1}^{N/2}\left(\tau_{\mathrm{pair}}(j)-\tau\right)^2,$$

with the expected value equal to $\mathbb{V}_W(\hat{\tau}^{\mathrm{dif}})$ if the treatment effect is constant across and within pairs.

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

The average of the within-pair differences $\hat{\tau}^{\text{pair}}(j)$, displayed in Table 10.3 is

$$\hat{\tau}^{\text{dif}} = \frac{1}{8} \cdot \sum_{j=1}^{8} \hat{\tau}^{\text{pair}}(j) = 13.4,$$

and its estimated sampling variance is

$$\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right) = \frac{1}{8 \cdot (8-1)} \cdot \sum_{j=1}^{8} \left(\hat{\tau}^{\text{pair}}(j) - \hat{\tau}^{\text{dif}}\right)^2 = 4.6^2.$$

The standard, normal distribution-based 95% confidence interval is

$$\mathrm{CI}^{0.95}(\tau_{\mathrm{S}}) = \left(\hat{\tau} - 1.96 \times \sqrt{\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right)}, \hat{\tau} + 1.96 \times \sqrt{\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right)}\right) = (4.3, 22.5). \quad (2)$$

$$\mathrm{CI}^{0.95}_{\mathrm{t}(7)}(\tau_{\mathrm{S}}) = \left(\hat{\tau} - 2.365 \times \sqrt{\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right)}, \hat{\tau} + 2.365 \times \sqrt{\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right)}\right) = (2.5, 24.3). \quad (3)$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Because we only have eight pairs of classes, one may wish to use a confidence interval based on the t-distribution with degrees of freedom equal to $N/2 - 1 = 7$,

$$\mathrm{CI}^{0.95}_{\mathrm{t}(7)}(\tau_{\mathrm{S}}) = \left(\hat{\tau} - 2.365 \times \sqrt{\hat{\mathbb{V}}^{\mathrm{pair}}\left(\hat{\tau}^{\mathrm{dif}}\right)}, \hat{\tau} + 2.365 \times \sqrt{\hat{\mathbb{V}}^{\mathrm{pair}}\left(\hat{\tau}^{\mathrm{dif}}\right)}\right) = (2.5, 24.3). \tag{4}$$

The Analysis of PRE:s from Neyman's Repeated Sampling Perspective

Suppose we had done a CRE, and had the same assignment vector. In that case we would have the same point estimate, namely $\hat{\tau}^{\mathrm{dif}} = \overline{Y}_t^{\mathrm{obs}} - \overline{Y}_c^{\mathrm{obs}} = 13.4$.

However, we would have a different estimate of the sampling variance. Using the standard Neyman estimated sampling variance we get

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{\mathrm{obs}} - \overline{Y}_c^{\mathrm{obs}} \right)^2 = 18.5^2, \qquad \text{and} \quad s_t^2 = 12.2^2,$$

and

$$\hat{\mathbb{V}}^{\mathrm{neyman}} = \frac{s_c^2}{8} + \frac{s_t^2}{8} = 7.8^2.$$

As this is substantially larger than $\hat{\mathbb{V}}^{\mathrm{pair}} = 4.6^2$ in this application, the assignment to pairs is effective, in the sense that it is based on factors that make the within-pair units substantially more similar than randomly selected units, probably leading to substantially more precise estimates.

Regression-based Analysis of PRE

In the discussions of regression-based analyses in CRE and SRE, the basic outcome in the analysis was Y_i^{obs}, the observed outcome for unit i.

Here, instead, we use as the primary outcome in the regression analysis the within-pair difference in observed outcomes of the treated and the control unit in the pair,

$$\hat{\tau}^{\text{pair}}(j) = Y_{j,t}^{\text{obs}} - Y_{j,c}^{\text{obs}},$$

with the pair serving as the unit of analysis.

We take a population perspective, where the pairs of units are drawn randomly from a large population, and one member of each pair is randomly assigned to the treatment group, and the other to the control group.

The population average treatment effect is $\tau_{\text{SP}} = \mathbb{E}_{\text{SP}}[\tau(j)]$, with the expectation taken over the random sampling of the pairs.

Regression-based Analysis of PRE

The standard estimator for the average treatment effect in a pairwise randomized experiment is the simple average of the within-pair differences,

$$\hat{\tau}^{\text{dif}} = \frac{2}{N} \sum_{j=1}^{N/2} \hat{\tau}^{\text{pair}}(j).$$

This estimator can also be interpreted as a regression estimator, where the regression function is specified simply as a constant:

$$\hat{\tau}^{\text{pair}}(j) = \tau_{\text{SP}} + \varepsilon_j.$$

The more interesting question is how to include additional covariates, beyond the implicit use of the pair indicators, into the regression function.

Regression-based Analysis of PRE

The goal when including additional covariates is to improve the precision of the estimator in cases where the covariates are strongly correlated with the treatment-control differences in potential outcomes.

Before discussing particular specifications, we first define $X_{j,A}$ and $X_{j,B}$ to be the covariate values for unit A and B respectively within pair j. Then define the within-pair observed difference in covariates between the treated and control units:

$$\Delta_{X,j} = \left(W_{j,A} \cdot (X_{j,A} - X_{j,B}) + (1 - W_{j,A}) \cdot (X_{j,B} - X_{j,A})\right),$$

and the average covariate value within the pair:

$$\overline{X}_j = \left(X_{j,A} + X_{j,B}\right)/2.$$

Regression-based Analysis of PRE

There are two leading approaches to including the covariates in the regression analysis.

1. include them in the form of the within-pair difference $\Delta_{X,j}$.

   This makes sense if X_i is associated with both potential outcomes $Y_i(0)$ and $Y_i(1)$ to approximately equal degrees.

2. include the average value of the covariates $\overline{X}_j$.

   This is a natural specification if one thinks the treatment effect, rather than the level of the potential outcomes, is linear in X_i.

The most general version of the regression function we consider includes the covariates both as within-pair differences and pair averages, where the latter is in deviations from the overall covariate mean $\overline{X}$:

$$\hat{\tau}^{\text{pair}}(j) = \tau + \beta \cdot \Delta_{X,j} + \gamma \cdot (\overline{X}_j - \overline{X}) + \varepsilon_j,$$

Regression-based Analysis of PRE

Let $(\tau^*, \beta^*, \gamma^*)$ be the population values:

$$(\tau^*, \beta^*, \gamma^*) = \arg\min_{\tau,\beta,\gamma} \mathbb{E}\left[\left(\hat{\tau}^{\text{pair}}(j) - \tau - \beta \cdot \Delta_{X,j} - \gamma \cdot (\overline{X}_j - \mu_X)\right)^2\right],$$

where $\mu_X = \mathbb{E}_{\text{SP}}(X)$ is the super population mean of X_i.

Here we use again the convention that the expectation operator without subscript is over both the randomization distribution and over the distribution induced by the random sampling from the super population.

Also let $(\hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}}, \hat{\gamma}^{\text{ols}})$ be the least squares estimators,

$$(\hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}}, \hat{\gamma}^{\text{ols}}) = \arg\min_{\tau,\beta,\gamma} \sum_{i=1}^{N} \left(\hat{\tau}^{\text{pair}}(j) - \tau - \beta \cdot \Delta_{X,j} - \gamma \cdot (\overline{X}_j - \overline{X})\right)^2.$$

Regression-based Analysis of PRE

Theorem

Suppose we conduct a pairwise randomized experiment in a sample of pairs drawn at random from the super population. Then, (i)

$$\tau^* = \tau_{\mathrm{SP}},$$

and (ii),

$$\sqrt{N} \cdot \left(\hat{\tau}^{\mathrm{ols}} - \tau_{\mathrm{SP}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}_{\mathrm{SP}}\left[\left(\hat{\tau}^{\mathrm{pair}}(j) - \tau^* - \beta^* \cdot \Delta_{X,j} - \gamma^* \cdot (\overline{X}_j - \mu_X)\right)^2\right]\right).$$

Regression-based Analysis of PRE

Now let us estimate the average treatment effect using 4 different specifications.

(1) $\hat{\tau}^{\text{ols}} = \frac{2}{N} \sum_{j=1}^{N/2} \hat{\tau}^{\text{pair}}(j) = \hat{\tau}^{\text{dif}}$

(2) $\hat{\tau}^{\text{pair}}(j) = \tau + \beta \cdot \Delta_{X,j} + \varepsilon_j$.

(3) $\hat{\tau}_{\text{pair}}(j) = \tau + \gamma \cdot \overline{X}_j + \varepsilon_j$.

(4) $\hat{\tau}^{\text{pair}}(j) = \tau + \beta \cdot \Delta_{X,j} + \gamma \cdot (\overline{X}_j - \overline{X}) + \varepsilon_j$

Note that in (1) we do not directly include the treatment indicator, because the unit of the least squares analysis here is the pair, not the individual unit, thus (1) is equal to the estimator in Equation (1).

Regression-based Analysis of PRE

Applying (1) leads to

$$\hat{\tau}^{\text{ols}} = 13.4 \ (\widehat{\text{s.e.}} \ 4.3),$$

(2) leads to

$$\hat{\tau}_{\text{pair}}(j) = \underset{(1.5)}{9.0} + \underset{(0.6)}{5.4} \times \Delta_{X,j},$$

(3) leads to

$$\hat{\tau}^{\text{pair}}(j) = \underset{(3.5)}{13.4} + \underset{(1.7)}{3.9} \times \bar{X}_j.$$

(4) leads to

$$\hat{\tau}^{\text{pair}}(j) = \underset{(1.5)}{8.5} + \underset{(0.8)}{5.9} \times \Delta_{X,j} \underset{(0.7)}{-1.0} \times \bar{X}_j,$$

Model-based Analysis of PRE

The analysis of PRE is little different from that for the case of SRE

In both cases the analysis is based on stratum membership, G_i.

The starting point is, again, a model for the joint distribution of the potential outcomes given the covariates, including the pair indicators, in terms of an unknown parameter θ:

$$f(\mathbf{Y}(0), \mathbf{Y}(1) | \mathbf{X}, \mathbf{G}, \theta),$$

in combination with a prior distribution on θ,$p(\theta)$.

In combination with the known assignment mechanism, this allow us to obtain the joint distribution $\mathbf{Y}^{\mathrm{mis}}$ given the observed data $(\mathbf{X}, \mathbf{G}, \mathbf{Y}^{\mathrm{obs}}, \mathbf{W})$, and thus allow us to obtain the posterior distribution of the estimand of interest, e.g., the average effect of the treatment.

Model-based Analysis of PRE

First we assume that, conditional on $(\mathbf{X}, \mathbf{G}, \mathbf{W})$ and θ, the potential outcomes are independent by the usual appeal to De Finetti's theorem:

$$f(\mathbf{Y}(0), \mathbf{Y}(1) | \mathbf{X}, \mathbf{G}, \mathbf{W}, \theta) = \prod_{i=1}^{N} f(Y_i(0), Y_i(1) | X_i, G_i, \theta).$$

The specific model we consider has a hierarchical structure, with pair-specific mean parameters μ_j, for $j = 1, \ldots, J$, that is,

$$\left(\begin{array}{c} Y_i(0) \\ Y_i(1) \end{array} \right) \Bigg| \; G_i = j, X_i = x, \mu_1, \ldots, \mu_{N/2}, \gamma, \beta, \sigma_c^2, \sigma_t^2$$

$$\sim \mathcal{N}\left(\left(\begin{array}{c} \mu_j + x \cdot \beta \\ \mu_j + \gamma + x \cdot \beta \end{array} \right), \left(\begin{array}{cc} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{array} \right) \right)$$

Model-based Analysis of PRE

Note that given this model, γ corresponds to the super population average treatment effect, τ_{SP}.

However, in this discussion we focus on inference for τ_{S}, by multiple imputing the missing potential outcomes. For that reason, the interpretation of the parameters in the statistical model is incidental.

Next, we specify a model for the pair specific means μ_j:

$$\left(\begin{array}{c} \mu_1 \\ \vdots \\ \mu_{N/2} \end{array} \right) \Bigg| \mathbf{G}, \mathbf{X}, \mathbf{W}, \gamma, \beta, \sigma_c^2, \sigma_t^2, \mu \sim \mathcal{N}\left(\left(\begin{array}{c} \mu \\ \vdots \\ \mu \end{array} \right), \left(\begin{array}{ccc} \sigma_\mu^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_\mu^2 \end{array} \right) \right).$$

Model-based Analysis of PRE

Even in simple cases, there are no analytic expressions for the posterior distributions for estimands of interest in such hierarchical models.

However, as we discussed in Chapter 8, this is of no intrinsic importance.

For the analysis of Childrens Television Workshop data we specify independent prior distributions for μ, σ_μ^2, σ_c^2, σ_t^2, γ, and β.

For μ, γ and β, we use normal prior distributions centered at zero, with variance 100^2.

For the three variance parameters $(\sigma_\mu^2, \sigma_c^2, \sigma_t^2)$, we use, again, inverse gamma distributions, here with parameters 1 and 1.

Model-based Analysis of PRE

The posterior mean and variance for the average treatment effect are

$$\mathbb{E}[\tau_{\mathrm{S}}|\mathbf{Y}^{\mathrm{obs}}, \mathbf{W}, \mathbf{X}, \mathbf{G}] = 8.4, \qquad \mathbb{V}(\tau_{\mathrm{S}}|\mathbf{Y}^{\mathrm{obs}}, \mathbf{W}, \mathbf{X}, \mathbf{G}) = 1.7^2$$

These estimates are quite similar to those for the regression model with the covariate equal to difference in pretreatment variables, where we estimated the average effect to be 9.0 with a standard error of 1.5.

Table 10.4 we report posterior means and standard deviations for all parameters.

Table 10.4: Posterior Moments and Quantiles

Parameter	Mean	Standard Deviation	Quantiles 0.025	0.975
γ	8.6	(1.6)	5.1	11.7
β	5.9	(0.6)	4.8	7.0
$\ln(\sigma_c)$	1.1	0.5	-0.3	1.9
$\ln(\sigma_t)$	0.5	0.7	-0.8	1.7
μ	-9.2	(2.2)	-13.6	-4.7
$\ln(\sigma_\mu)$	1.5	0.4	0.4	2.2