Predicting Treatment Effects in Randomized Clinical Trial

Randomized clinical trial (RCT) is a type of scientific experiment which is commonly used in testing the new medicines, healthcare service or health technologies. Blindness is usually used in RCTs. One way to apply blindness is that the new medicine and placebo were blind by attaching different tags (A or B). When assigning the medicines to patients, sometimes the doctors, patients and statisticians are blind to the medicines they used. None of them knows about which kind of medicines that patients are assigned to. Usually, the statistician will be asked to predict the difference between A or B. Suppose we observe the simple sample mean for patients assigned to treatment A, is 80 mmg, for diastolic blood pressure (DBP) and simple sample mean for treatment B, is 90 mmg. How should the statistician predict the effect of treatment?

The traditional approach is to use the difference between the simple sample means of A and B to estimate the treatment effect. In this approach, the treatments are considered as the fixed effects. In practice, one way to carry out the clinical trial is to have a list of patients and then randomly choose several sample groups, and for each sample group, we randomly assign a certain treatment. In this situation, we develop a predictor of the difference between A and B for the realized treatment assignment. Interestingly, the predictor of the realized treatment effect does not equal the simple difference in sample means. The predictor has smaller mean squared error (MSE) than the estimator using sample means has. The approach considers treatments as random effects when there are only two treatments. We use the historical definition of random effects applied to a finite set of treatments. These results are considered to be of broad interest.

The basic issue under the derivation is whether we consider the treatments are random or fixed. Suppose we have three treatments a, b, d, if we put three different shapes of tags in the drugs. Actually, we could get six permutations of these three treatments. When we apply the three treatments to the patients, we use one of the realizations of the six permutations.
The treatments are randomly assigned to different shaped tags. According to the traditional concept, we will think the treatments are considered random if the treatments are sampled from a large population, and what we are concerned is about the underlying population. In this case, we are not sampling treatments from a large treatment population, usually we will think of the treatments as fixed effects. But even we have all the treatments, we only use one of the realization of six the permutations of treatments. Is there any reason we could think the treatments are random effects in the model?

Notice that there are two processes involved in randomizing patients to different treatments. First is choosing simple random samples from the list of patients. The second process is to allocate the sample groups to each of the treatments. We use the permutation and sampling to mimic the two processes.

First, according to the potential observable outcome concepts, suppose we have \( A \) treatments, then for each patient, they have \( A \) potential observable outcomes, and we have totally \( N \) patients on the list. Assuming there are no measurement errors, we could use \( y_{sa} \) to indicate the potential observable outcomes for patient \( s \) under the treatment \( a \). \( s \) is one of \( (1, 2, \ldots, N) \) and \( a \) is one of \( (1, 2, \ldots, A) \).

Then totally the potential observable population is \( N \times A \) matrix. We could express the population in \( (y_{sa}) \).
Second, suppose we take $A$ random samples each with sample size $n$, and then randomly assign $A$ treatments to $A$ samples. In order to express this process, we could permute the matrix by columns and by rows. And then we could get the new matrix as

$$
\begin{pmatrix}
    y_{11} & y_{12} & \cdots & y_{1,(A-1)} & y_{1A} \\
    y_{21} & y_{22} & \cdots & y_{2,(A-1)} & y_{2A} \\
    \vdots & \vdots & & \vdots & \vdots \\
    y_{(N-1)1} & y_{(N-1)2} & \cdots & y_{(N-1),(A-1)} & y_{(N-1)A} \\
    y_{N1} & y_{N2} & \cdots & y_{N,(A-1)} & y_{NA}
\end{pmatrix}
$$

$Y_{ij}$ is the value of the $i^{th}$ patient given the $j^{th}$ treatment. $i$ indicates the patient in the $i^{th}$ position after permutation of all the patients. And $j$ indicates the treatment in the $j^{th}$ position after permutation of all the treatments.

Thirdly, we choose the first $n$ patients from the first column, and the second $n$ from the second column, et al. as $A$ samples so that we observe the values in these samples.

$$
\begin{pmatrix}
    Y_{11} & Y_{12} & \cdots & Y_{1,(A-1)} & Y_{1A} \\
    Y_{21} & Y_{22} & \cdots & Y_{2,(A-1)} & Y_{2A} \\
    \vdots & \vdots & & \vdots & \vdots \\
    Y_{(N-1)1} & Y_{(N-1)2} & \cdots & Y_{(N-1),(A-1)} & Y_{(N-1)A} \\
    Y_{N1} & Y_{N2} & \cdots & Y_{N,(A-1)} & Y_{NA}
\end{pmatrix}
$$

Fourthly, we rearrange the random variables and get a set corresponding to the weighted sample totals and a set corresponding to the weighted remaining totals for each treatment.
We define \( Y_I^* = \frac{n}{N} \begin{bmatrix} \bar{Y}_{j=1I} \\ \bar{Y}_{j=2I} \\ \vdots \\ \bar{Y}_{j=AI} \end{bmatrix} \) and \( Y_{II}^* = \begin{bmatrix} \bar{Y}_{j=1I} \\ \bar{Y}_{j=2I} \\ \vdots \\ \bar{Y}_{j=AI} \end{bmatrix} \). Based on the BLUP approach, we derive the predictor as

\[
\hat{P}_{ij}^* = \frac{\sigma_I^2}{\sigma_A^2 + \frac{\sigma_S^2}{n}} (\bar{Y}_{ij} - \bar{Y}_{j=1I}) \quad \text{and} \quad \text{MSE} \left( \hat{P}_{ij}^* \right) = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma_S^2}{n}} \left( \frac{2\sigma_S^2}{n} \right)
\]

where \( \sigma_S^2 \) is variation between units and \( \sigma_A^2 \) is variation between treatments.

To obtain the empirical predictor we could plug in the method of moment estimates of \( \sigma_S^2 \) and \( \sigma_A^2 \) to the predictors. Method of moment estimators will be used to get the estimators of the variance components. According to the ANOVA
table for the one factor experimental design, the sum of squares of errors is expressed as

$$SSE = \sum_{j=1}^{A} \sum_{i=(j-1)n+1}^{i} (Y_{ij} - \bar{Y}_{j})^2$$

and the mean squared errors is denoted as

$$MSE = \frac{\sum_{j=1}^{A} \sum_{i=(j-1)n+1}^{i} (Y_{ij} - \bar{Y}_{j})^2}{A(n-1)}.$$  The sum of squares of treatments is defined as

$$SST = \sum_{j=1}^{A} n (\bar{Y}_{j} - \bar{Y})^2$$

and the mean squared of treatments is denoted as

$$MST = \frac{\sum_{j=1}^{A} n (\bar{Y}_{j} - \bar{Y})^2}{A-1}.$$  Hence, the method of moment estimators for variance components is as followings:

$$\hat{\sigma}_A^2 = MSE \quad \text{and} \quad \hat{\sigma}_S^2 = \frac{(MST - MSE)}{n}, \quad \text{if} \quad MST > MSE.$$

$$\hat{\sigma}_S^2 = \frac{SST + SSE}{nA-1} \quad \text{and} \quad \hat{\sigma}_A^2 = 0, \quad \text{if} \quad MST \leq MSE.$$

Plug in the shrinkage constant $k$, the empirical predictor is written as

$$\hat{P}_{E_{ij}} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_S^2} \left( \bar{Y}_{ij} - \bar{Y}_{j} \right).$$

Compare the predictor, empirical predictor and sample means.
Table 1: MSE for different predictors or estimators

<table>
<thead>
<tr>
<th>Different predictors (Estimators)</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean</td>
<td>( \hat{P}_{sij} = \bar{Y}<em>j - \bar{Y}</em>{j+} )</td>
</tr>
<tr>
<td>RP predictor</td>
<td>( \hat{P}_{ij} = \frac{\sigma_s^2}{\sigma_A^2 + \sigma_s^2} )</td>
</tr>
<tr>
<td>Empirical RP predictor</td>
<td>( \hat{P}_{Eij} = \frac{\sigma_s^2}{\sigma_A^2 + \sigma_s^2} )</td>
</tr>
</tbody>
</table>

According to Table 1, if the variance components are known, the \( MSE(\hat{P}_{ij}) \) is less than \( MSE(\hat{P}_{sij}) \). If unknown, the simulation study is conducted to evaluate the empirical predictors and the sample mean.

Simulation results for MSE for Empirical predictor and the difference between sample means. In order to execute the simulation, we use \( A = 2 \) treatments \( N = 120 \) as an example. And overall mean is defined as \( \mu_0 = 4 \), and \( \sigma_s^2 = 4 \), we choose target power as 0.98, 0.95, 0.90, 0.86, 0.80, 0.76, 0.70 and calculate the \( \sigma_s^2 \) based on the relation between power and non-central parameters. In each case, we resample 5000 times and calculate MSE for the empirical predictor and MSE for the estimator using the difference between sample means.