Predictors of treatment means for a One Factor Completely Randomized Design

A Dissertation Prospectus Presented

April 26, 2007

by

Bo Xu

Approved as to style and content by:

_______________________________
Edward J. Stanek III, Chair

_______________________________
John P. Buonaccorsi, Member

_______________________________
Recai Yucel, Member

Graduate Program Director
Shlomo Barnoon
# Table of Contents

1. Introduction ........................................... 3
   1.1 Causality and concept of potential observable outcomes .......... 5
   1.2 Completely randomized design and Randomization .................. 6
   1.3 Random Effects ....................................... 7
   1.4 Sampling Strategy and Random Permutation ......................... 8
   1.5 BLUP for random effects .................................. 10
2. Literature Review ...................................... 11
3. Proposal .................................................. 25
4. Example ................................................... 28
5. Preliminary Results ..................................... 30
Reference .................................................... 36
1. Introduction

Prediction of random effects in one way random effects model is very important in biology, agriculture and pharmaceutical areas. This proposed research will study predictors of realized treatments in the context of this problem. Specifically, we study predictors of treatment means in a one factor completely randomized design using a random permutation model prediction framework. The context we consider is for a finite population with units labeled by \( s = 1, \ldots, N \) and finite levels (treatments) of the factor labeled by \( a = 1, \ldots, A \). The experiment consists of randomly assigning a simple random sample of \( n \) units to each of the \( A \) levels where \( nA \leq N \). The goal is to predict the mean of a treatment, or linear combinations of treatment means. Usually, when treatments are a random sample from a large population of possible treatments, random effects models are used.

Data from a one factor completely randomized design can be represented via two conceptual models, each related to the manner in which data accrue. One approach to the conduct of a study is to consecutively select samples based on simple random sample (SRS) without replacement, and associated the first sample with treatment 1, the second with treatment 2, etc. The other is to consecutively select samples based on SRS without replacement, and then randomly assign a treatment to sample 1, another to sample 2, etc. The difference of the two ways in which the study may be conducted lead to models that corresponds to considering treatments to be fixed or random. In the first approach, the treatments are fixed, and in the second approach, the treatments should be considered as random effects. In our research, we use the latter approach and develop inference based on a random permutation model to capture the mechanism of such random sampling strategy.
The random permutation model is a design-based model that is directly linked to a finite population. This model does not require assumptions about parametric distributions of response. Random variables correspond to permutations of units in the finite population, where each permutation of the population occurs with equal probabilities. Parameters in the population are represented as a non-stochastic vector. When the concept of potential observable outcomes for each treatment is used, the potential observable population is extended into a population of size $N \times A$. The randomness of units is thus attributable to random sampling. Since we permute the treatments before assigning them, another set of random variables is used to indicate the permutation of treatments.

Predictors of treatment means in the random permutation model will be developed in a similar way as model-based predictors obtained via super-population approaches for survey sampling. This research will compare estimators from a one factor fixed effect model with predictors from a one factor random effects model, and study the advantages and disadvantages of the different approaches.

There are several issues and concepts that are needed to place this research in perspective. We begin with a brief presentation of these concepts. Following this discussion, we present a more detailed review of the literature. We conclude with an outline of the proposed research.

In the introduction part, we begin with a brief introduction to the concepts of causality and potential observable outcomes. Next, we discuss randomization in a completely randomized design, and the role of randomization. We follow this with a short discussion of the term ‘random effect’, and introduce the sampling framework in the context of a random permutation model. We conclude with discussion of prediction of realized random effects.
1.1 Causality and Concept of Potential Observable Outcomes

Causality is the relationship between cause and effect (Kempthorne1952). Causal effects are defined as comparisons of potential outcomes under different treatments on a common set of units (Rubin, 2005). A unit is defined as the person, place, or thing upon which a treatment will operate, at a particular time. A treatment is defined to be an intervention, the effects of which the investigator wishes to assess relative to no intervention. A potential outcome is the value of a unit’s measurement of interest after application of the treatment (Rubin 2005).

One of the assumptions required for causal inference is the stable unit treatment value assumption (SUTVA) which comprises two sub-assumptions (Little and Rubin 2000). First, it assumes that there is no interference between units, implying that none of the potential observable outcomes for a unit is affected by the treatment assignment that any other unit receives. Second, the SUTVA assumption implies that there are no hidden versions of treatments; no matter how unit \( s \) received treatment \( a \), the outcome that would be observed is \( y_{sa} \). The main idea is that the causal effect, defined as the difference in response with vs. without the treatment, for a particular unit does not depend on assignments of other units. Under SUTVA, each unit has one non-stochastic potential value under each treatment. Hence, potential observable outcomes will form an \( N \times A \) array.

We use this concept of potential observable outcomes in formulating one factor experimental design. We assume that each unit could be potentially observed under each of the treatments, and represent the response for unit \( s \) given treatment \( a \) by the non-stochastic value \( y_{sa} \). The average for treatment \( a \) is defined by \( \mu_a = \frac{1}{N} \sum_{s=1}^{N} y_{sa} \). The average for unit \( s \)
is defined by $\mu_s = \frac{1}{A} \sum_{a=1}^{A} y_{sa}$. The overall mean is $\mu = \frac{1}{NA} \sum_{a=1}^{A} \sum_{s=1}^{N} y_{sa}$. If $a = 1$ represents control, and $a = 2$ represents a treatment, the average causal effect of the treatment is given by $\mu_2 - \mu_1$.

1.2 Completely randomized designs and Randomization

The completely randomized design is the simplest form of an experimental design. A key feature of this design is that treatments are randomly assigned to the units. However, the design may also be described as units being randomly assigned to treatments. When formulating models for such designs, the treatments are usually considered as fixed effects even though the description may state that treatments are randomly assigned to the units. The completely randomized design can use any number of treatments and permit different treatments to be assigned to different numbers of units.

Kempthorne (1952) described the randomization process for completely randomized design through a set of random variables (design random variables). The study was represented mathematically and a linear model was derived based on specification of a parameterization of the population. This model is called randomization model.

Randomization models are based on the random assignment of units to treatments, and have been used in inferences for treatment contrasts, requiring less restrictive assumptions than for the ordinary normal theory models (Kempthorne, 1952 and Scheffe 1959). This model captures the mechanism of the random assignment of units to treatments and the prior distribution of the potential population plays no role in the inference.

The use of randomization insures that the choice of samples is not affected by any biased or preconceived notions on the part of the experimenter, thus making the results of the experiments...
more readily acceptable to others, as well as protecting the experimenter against self-deception.

1.3 Random effects

The term *random effects* in the context of analysis of variance is usually used to denote the levels of a factor in an ANOVA design which are sampled from a population of possible levels (Kunter, Nachtsheim, et al. 1974). Usually, the random effects refer to the deviation of the treatment means from the overall mean so that the random effects have zero mean. Traditionally, the judgment as to whether the effects are random or fixed depends in part on how the observations were obtained. If the levels of the factor are a random sample from the population or come from a probability distribution, it should be considered as random effects (Eisenhart 1947, Searle and Casella et al. 1992). Some authors (Searle and Casella et al. 1992) consider the target for inference to be important when deciding whether a factor is fixed or random: when inferences will be made about a population of effects from which those in the data are considered to be a random sample, the effects are considered as random; and when inferences are going to be confined to the effects in the experiment, the effects are considered as fixed.

Historically, the population that random effects come from is usually considered to be of infinite size, either because this is the case or because, although finite, the population is large enough to be taken as infinite (Searle and Casella et al. 1992, Henderson (ref to animal breeding book) Bennett and Franklin (1954, p404) and Kempthorne 1955 ). However, the definition of random effects does not require an infinite size population of such effects (Searle and Casella et al. 1992). We can consider the conceptual population of effects to be of three sizes: infinite, finite but so large as to be deemed infinite, and finite.

The settings we discuss assume that there are a finite number of treatments in one factor.
experimental design. As a result, we define our population based on potential observable outcomes concept with finite units and finite treatments. Completely randomization design is conducted by the random assignment of the treatments to samples of units. In our framework, we permute the treatments before we assign them which lead to consider the treatments are random. Next we are going to present a way to use random variables to capture the mechanism of permutation which forms a basis of sampling and forms the random permutation model in our settings.

### 1.4 Sampling Strategy and Random Permutation

Let an experiment consist of randomly selecting $A$ consecutive simple random samples of $n_a$ units without replacement, and then randomly assigning $A$ treatments to the samples $(nA A N \leq N)$. We permute the treatments prior to assigning them to sample groups.

One way to represent the experimental design is to think of all possible permutations of the units and all possible permutations of treatments. Let us index the positions in a permutation of units by $i = 1, \ldots, N$. We will refer to the unit in position $i$ as PSU $i$. Meanwhile, permute the treatments and index the positions in the permutation of treatments by $j = 1, \ldots, A$. We represent the experiment by sets of random variables which have stochastic properties corresponding to the joint permutation of units and treatments.

As we have mentioned before, the potential observable population in the setting is an $N \times A$ array with the non-stochastic values $y = \left((y_{si})\right)$ where $s = 1, \ldots, N$ which denote the label of a unit, and $a = 1, \ldots, A$ denote the label of a treatment. Random sampling without replacement of units is introduced by defining a set of random variables of $U_{is}$, $i = 1, \ldots, N$ and $s = 1, \ldots, N$. The permutation of units can be defined in terms of $U_{is}$. Explicitly, the random variable $U_{is}$ takes on a value of one if unit $s$ is assigned to position $i$ in a permutation, or
zero otherwise. The matrix of indicator random variables for units is given by \( U_{N \times N} = \left( U_{\alpha} \right) \).

Similarly, let \( j = 1, \ldots, A \) index the position of a treatment in a permutation of treatments. The permutation of the treatments can be defined in terms of the indicator random variables \( V_{ja} \), where \( V_{ja} \) takes on a value of one if the treatment \( a \) is assigned to position \( j \) in a permutation, or zero otherwise. The matrix of indicator random variables for treatments is \( V_{A \times A} = \left( V_{ja} \right) \).

We use \( U \) and \( V \) to represent a joint permutation of units and treatments. Explicitly, let

\[
U = \left( \begin{array}{cccc}
U_1 & U_2 & \cdots & U_N
\end{array} \right) = \left( \begin{array}{cccc}
U_{\alpha1} & U_{\alpha2} & \cdots & U_{\alpha N}
\end{array} \right) \text{ is the } i^{th} \text{ row vector of } U \text{ and } V = \left( \begin{array}{cccc}
V_1 & V_2 & \cdots & V_A
\end{array} \right) = \left( \begin{array}{cccc}
V_{j1} & V_{j2} & \cdots & V_{jA}
\end{array} \right) \text{ is the } j^{th} \text{ row vector of } V .
\]

Then

\[
Y_{ij} = U'_{i} y V_{j} = \sum_{a=1}^{A} \sum_{a=1}^{A} U_{ia} V_{ja} y_{sa}.
\]

As a result, the population after the joint random permutation is defined as \( Y = \left( \begin{array}{c}
Y_{ij}
\end{array} \right) \) = \( U y V' \). Using the property of the permutation matrix (Stanek, Argentina2006-lec1a.doc), we get probability statements for permutation matrix \( U \) and \( V \). The joint distribution of \( Y \) could be derived.

Next, we choose \( i = 1, \ldots, n \) and \( j = 1 \) which is \( \left( Y_{11} \ Y_{21} \ \cdots \ Y_{ni} \right)' \) as the first sample group. We choose \( i = n+1, \ldots, 2n \) and \( j = 2 \) which is \( \left( Y_{(n+1)1} \ Y_{(n+2)1} \ \cdots \ Y_{(2n)1} \right)' \) as the second sample group, et al. Hence the last sample group should be

\[
\left( Y_{1(A-1)n+1,1} \ Y_{1(A-1)n+2,1} \ \cdots \ Y_{An,1} \right)' \text{ which indicates } i = \left( (A-1)n + 1 \right), \ldots, An \text{ and } j = A.
\]

Finally, we have \( A \) sample groups and each has \( n \) units and been associated with one of the \( A \) treatments. This procedure fulfills our approach to sample units by SRS without replacement and random assign treatments to each sample of units. Our goal is to predict the mean
of a realized treatment based on a sample of \( Y \) corresponding to the realized random variables in a one factor completely randomized design.

1.5 BLUP for random effects

When predicting the mean of a realized treatment based on a sample of \( Y \), we use BLUP approach. The acronym BLUP stands for “best linear unbiased prediction”. It is a common used technique for predicting the random effects in both the random effects model and the mixed model. In our framework, the target is a linear combination of random variables in the population, for example, 

\[
P_j = \frac{1}{N} \sum_{i=1}^{N} Y_{ji}
\]

where \( P_j \) indicates the mean of the \( N \) random variables corresponding to the treatment assigned to the \( j^{th} \) position in the permutation of treatments.

BLUP in the context of survey sampling is defined as (1) the predictor is a linear combination of the random variables of the sample; (2) unbiased in the sense that unconditionally expected values of the difference between the predictors and targets equal to zero (unbiased refers to average bias over all the possible samples equal to zero); (3) minimize the expected mean squared error (MSE) of the predictors. This prediction approach is parallel to the prediction-based approach used in survey sampling with super population models (Scott and Smith 1969; Pfeffermann and Nathan 1981). In the super-population prediction approach, statistical inference is based on a model for super population. The finite population is defined as a realization of super population random variables. However, the actual sample design plays no role in the inference. It usually is considered as the model-based approach. On the contrary, the approach we develop is based on random variables that arise directly from the experimental design, while explicitly incorporates permutation of the treatments. For this reason, it is a design based prediction approach (Stanek, 2004).
2. Literature Review

Neyman (1923) built up the model in agricultural experiments to compare a number of crop varieties. In this important paper, Neyman contributed a lot to understanding the experiment design and the analysis method. In designing the experiments, first the field is divided to \( m \) plots. In order to compare \( v \) variety, Neyman introduced a double-indexed array to indicate the unknown potential yields. \( U_{ij} \) indicates the true yield of the \( i^{th} \) variety for \( j^{th} \) plots where \( i = 1, \ldots, v \) and \( j = 1, \ldots, m \). This can be considered as there are \( v \) urns and in each urn, there are \( m \) balls.

So the number of \( a_i = \frac{\sum_{j=1}^{m} U_{ij}}{m} \) is the best estimated of the yield from the \( i^{th} \) variety on the field and \( \sigma_i^2 = \frac{\sum_{j=1}^{m} (U_{ij} - a_i)^2}{m} \) indicates the population variance of the \( i^{th} \) variety.

Actually when conducting the experiment, one plot can only be measured under one variety. This introduces the property that if one ball is taken from one of urns, then balls having the same plot (label) will disappear from all the other urns. Suppose the experimenter carried out \( k \) trials under sampling without replacement in certain variety, Neyman proved that the trials conducted in this way are not independent. Any pair of the random outcomes from the \( k \) trials has the covariance \( E(x_{ij} - a_i)(x_{ik} - a_i) = -\frac{\sigma_i^2}{m-1} \). He proved that the arithmetic mean of \( k \) experiments \( \overline{x}_i = \frac{\sum_{j=1}^{k} x_{ij}}{k} \) is the best estimate of \( a_i \). When running into the problem of determining the difference between the yields of two varieties, it is easy to see that \( E(\overline{x}_i - \overline{x}_j) = a_i - a_j \). It means that the expected value of the difference of the partial
averages of yields from two different varieties is equal to the difference of their expectations.

Contributions of Neyman’s method could be understood as following:

First, even he did not illustrate very clearly, he built his method based on the finite population and the sampling framework of simple sampling without replacement. In the whole setting, he supposed there are \( m \) plots and sample without replacement for the \( k \) plots to each of the variety. No assumption is about the underlying distribution of the population and the possible outcomes are not independent. Explicit or implicit, each unit of the finite population is not random, but when simple sampling is conducted the possible outcomes are random. Since we could enumerate all the possible samples, we could get the expected values and variance of the sample means.

Secondly, Neyman introduced the concept of the “potential yield” which is a very good start to understand the causal inference. In this setting, each plot could have a true yield under certain variety. Here the population like balls in urns is no longer \( m \) balls, but \( m \) times \( v \) balls. When defining variations and covariance between different varieties, all the possible yields are used.

Last, Neyman said “for the time being, we will conclude that since it is impossible to calculate directly an estimate of \( r \) (correlation between treatments), it is necessary to take \( r=1 \)”, this corresponds to the frequently term “unit-treatment additivity” which has been used by Kempthorne, 1952, Cox, 1958, and Holland, 1986.

Another point which is worth notifying is that Neyman in this paper assumed the measurement could be made with high accuracy, and the observed yield is essentially equal to the true yield.

Unfortunately, even with great ideas, Neyman does not express it very clearly with his
Later on, Randomization models was discussed by Neyman (1935) for the randomized blocks, by Welch (1937) and Pitman (1937) for the Latin square under a certain null hypothesis, and by Kempthorne (1952, 1955) and Wilk (1955) for many other designs.

Further step notably on this direction goes by Kempthorne in 1952. He presented the randomization model mathematically. He defined the term of experimental units (EU) as the piece of experimental material, to which a treatment is assigned and applied. He also gave the formal definition of the completely randomized design: “Suppose we have $N = tr$ homogeneous EUs with $t$ treatments. Let the $tr$ EUs be partitioned randomly with equal probability into $t$ sets of $r$ EUs. Let the $t$ treatments be assigned to the $t$ sets such that the $i^{th}$ treatment is applied to each of the $r$ EUs in the $i^{th}$ set. This procedure defines the completely randomized equal replication design for $t$ treatments.” He emphasized that “one has a randomized design, if and only if, one has randomized the assignment of the treatments to the EUs.” Even he did not formally use the potential observed outcomes for each units, he defined “the treatment $i$ is applied to EU $k$, the true response is a number of $T_{ik}$. We suppose that if we could, in fact, impose every treatment on every EU, we could observe the totality of numbers $\{T_{ik}\}$.” In this context, he continued the idea of “potential observable yield” from Neyman. Kempthorne built up “design random variables” to express the completely randomized design process mathematically. Let $\delta_{ij}^k = 1$ if EU $k$ is associated with label $ij$, 0 otherwise. Therefore, the probability statements are $P(\delta_{ij}^k = 1) = \frac{1}{N}$, $P(\delta_{ij}^k = 1, \delta_{ij'}^{k'} = 1) = \frac{1}{N(N-1)}$, when $k \neq k', ij \neq i'j'$, there are $(rt)^2$ random variables, which are simple Bernoulli (0,1) random variables and that they are identically but not independently distributed. If denoting the expectation and variance and
covariance under the randomization model as \( E_R, \) \( Var_R \) and \( cov_R \), respectively, we could obtain

\[
E_R(\delta^k_y) = \frac{1}{N}, \quad Var_R(\delta^k_y) = \frac{1}{N} \left(1 - \frac{1}{N}\right),
\]

\[
\text{cov}_R(\delta^k_y, \delta^{k'}_y) = -\frac{1}{N^2} \quad \text{if} \quad k = k', \quad ij \neq i'j',
\]

\[
\text{cov}_R(\delta^k_y, \delta^k_y) = -\frac{1}{N^2} \quad \text{if} \quad k \neq k', \quad ij = i'j',
\]

\[
\text{cov}(\delta^k_y, \delta^{k'}_y) = \frac{1}{N^2(N - 1)} \quad \text{if} \quad k \neq k', \quad ij \neq i'j'.
\]

In his context, there are \( N \times t \) fixed \( T_{ik} \), which forms a finite population with potential (conceptual) responses. Using the design random variables, \( \delta^k_y \), he linked the \( y_y \) and the \( T_{ik} \) as

\[ y_y = \sum_{k=1}^{N} \delta^k_y T_{ik}. \]

Here, \( y_y \) is a random variable, and since we know the probability statement of \( \delta^k_y \), after some simple derivation we could easily get the first and second moments for \( y_y \).

Another point that Kempthorne contributed to the randomization model is that what he called “the additivity in the Broad Sense”. At the beginning, Kempthorne assumed \( T_{ik} = T_i + U_k \), this indicates that the response of treatment \( i \) applied to \( EU \) \( k \) is made up additively from a contribution due to the \( i^{th} \) treatment \( T_i \) and a contribution due to the \( k^{th} \) \( EU \) \( U_k \). This concept agrees with the idea Neyman argued that \( r = 1 \) (the correlation between different treatment is equal to 1). Kempthorne called this additivity in strict sense. In the additivity in the Broad Sense, he attached the technical error \( M_{ik} \) (Wilk and Kempthorne 1956) to \( T_{ik} \). He partitioned \( M_{ik} \) as \( E_{ik} \) which is experimental error, and \( O_{ik} \) which reflects observational error. Hence, the conceptual response of applying treatment \( i \), to \( EU \) \( k \), is \( Y_{ik} = T_i + U_k + E_{ik} + O_{ik} \) hence

\[ y_y = \sum_{k=1}^{N} \delta^k_y Y_{ik}. \]

With carefully and appropriately defining the parameters \( y_y \) could be written as
\[ y_{ij} = \mu + \tau_i + \sum_{k=1}^{N} \delta_{ij}^k u_k + \sum_{k=1}^{N} \delta_{ij}^k E_{ik} + \sum_{k=1}^{N} \delta_{ij}^k O_{ik} \]
\[ = \mu + \tau_i + \omega_{ij} + v_{ij} + \eta_{ij} \]

The last three items in the right hand side of the equation are random. In this framework, \( y_{ij} \) is not independent with \( y_{ij'} \). But when we run into contract problem, Kempthorne found that the expected value and variance of \( \sum_i c_i \bar{y}_i \) are the same as those based on the assumption of treating \( \omega_{ij} \) as independently, identically distributed (i.i.d.) random variables with mean zero and variance \( \sigma_u^2 \).

Kempthorne provided the ANOVA tables which included the expected values of the means of sum squares. One step further, he showed how the randomization process together with the analysis of variance leads to a simple procedure for testing hypotheses about treatment effects. The null hypothesis people concerned is “the treatments have no differential effects” which means we would observe the same response on an EU regardless of which treatment has been used. In this sense, the experimental plan we have used is a random one of the possible plans \( s = \frac{N!}{(r!)^2} \). If chosen a criterion \( C \), we are able to evaluate \( C \) for all possible plans which give us a set of numbers \( C \). Then we declare that the significance level against the null hypothesis of no differential treatment effects, with the chosen criterion, is

\[ SL = \frac{1}{S} \left[ \text{number of } C_r \left( r = 1, 2, \ldots, s \right) \geq C_0 \right] \]

This is the exact test do not based on the assumption of the population distribution and independence properties. But the tackling point is that this randomization test includes cumbersome calculation even after some simplification.

After comparing the first two moments of the distribution of \( Z = \frac{SS(T)}{SS(T) + SS(E)} \),
Kempthorne draw the conclusion that distribution of $Z$ under normal theory and under randomization theory are in good agreement. And further simulation study also approved that the ordinary $F$-test for testing the hypothesis of no differences among the treatment effects as a good approximation to the randomization test. These results can go back to Fisher (1935), Pitman (1937), and Welch (1937).

Scheffe (1959, Chapter 9) summarized the randomization models which take into account a randomization employed in assigning the treatment combination to the experiment units. He extended to formulate the randomized blocks and Latin square designs. He judged it is worth while to generalize the randomization model for the reason that we could gain a better “understanding of the nature of the error distribution generated by the physical act of randomization” and this should be part of out knowledge of the basic theory of the analysis of variance. Also, he continued the idea of “additivity” which means that no interactions exist between units and treatments. When it comes to the randomization test, Scheffe emphasized that randomization test exists whenever the joint distribution of the observations under the hypothesis has a certain kind of symmetry (The distribution is invariant under a group of permutations.). He pointed out the $F$-test derived for the corresponding hypothesis under a fixed-effects normal-theory model including assumptions of normality, independence, and equality of variance, or a slight modification of this test, can often be regarded as a good approximation to a permutation test in randomized block design and Latin square design.

Tukey (1949) contributed a test for Non-additivity which provides a means to test whether the additivity assumption is true or false. Ghosh and Sharma (1963) stated Tukey’s Test for Non-Additivitiy very clearly.
In a two-way classification, if the $x_{ij}$ is the observation in the $i^{th}$ row and $j^{th}$ column where $i = 1, 2, \cdots, p ; j = 1, 2, \cdots, q$, and the $x_{ij}$ are independent normal variables with mean $\mu_{ij}$ and variance $\sigma^2$. The model could be written as $\mu_{ij} = \mu + \alpha_i + \beta_j + c_{ij}$ where $\sum_i (\alpha_i) = \sum_j (\beta_j) = \sum_i (c_y) = \sum_j (c_y) = 0$. Under additivity, we have $\mu_{ij} = \mu + \alpha_i + \beta_j$.

Tukey developed a test criterion for non-additivity. Let

$$S_1 = \frac{\sum_{i,j} (x_{ij} - \bar{x})(x_{ij} - \bar{x})}{\sum_i (x_i - \bar{x})^2 \sum_j (x_{ij} - x)^2}$$

The statistic is defined by Tukey is

$$T = \frac{S_1 (pq - p - q)}{ErrorS.S - S_1}$$

And the hypothesis of additivity is rejected for large values of $T$.

The distribution of the statistic under the hypothesis that additivity holds was show to be an $F$ distribution with d.f.1 and $pq - p - q$. Scheffé (1959) has considered the restricted class of alternatives in which $c_{ij} = c \alpha_i \beta_j$. Both of the tests assume the error term is independently, identical and normally distributed.

The random-effects models for the analysis of variance are also called variance components models. It could be traced back to late nineteenth century in astronomy area long time before statistician efforts to handle it. Airy (1861) has formulated a variance components model of the one-way layout for night observations. And he treated the night effects as random effects.

Fisher (1925, Sec.40) implicitly introduced variance components models, and they are of course behind his idea of the intra-class correlation coefficient. Eisenhart (1947) puts forward the terminology of “Model I” (used in fixed effect models) and “Model II” (used as random effects...
Scheffe (1959) sketched mathematical models for the random effects as below:

$$\Omega: y_{ij} = \mu + a_i + e_{ij}$$

The $I + IJ$ random variables $\{a_i\}, \{e_{ij}\}$ are completely independently,

The $\{a_i\}$ are $N\left(0, \sigma^2_a\right)$; The $\{e_{ij}\}$ are $N\left(0, \sigma^2_e\right)$.

The assumptions emphasized here are additivity, normality, equal variances, and independence.

When coming to random effects model, Cornfield and Tukey (1956) pointed out pigeonhole model which provides a very good connection to finite treatments and elements. They considered a finite or infinite number of elements. These elements are classified to $R$ rows and $C$ columns. Each of the $RC$ pigeonholes contains $N$ elements. In Tukey’s opinion, if a sample of $r$ rows be drawn from the $R$ potential rows, and a sample of $c$ columns be drawn from the $C$ potential columns, then the column and row factors could be treated as random effects $(r < R, c < C)$. As a result, when calculating the variance components, the sampling fraction should go into the expected values. The only assumption Tukey made here is that all the samplings of the rows, of columns, and with pigeonholes are at random and independent of one another. The assumption depends on the nature of the sampling and randomization involved in obtaining the data.

Kempthorne in his paper “Fixed, Mixed, and Random Models” attempted to classify the randomization model to the fixed models, mixed models and random models. According to his idea, if $a$ levels of the factors are sampling from $A$ total levels where $A \gg a$, this factor should be considered as random effects. Further if $a = A$, it is considered as fixed effects. Then the
sampling fractions will go into the expectation of mean squares. Cornfield in 1953 and Bennett
and Franklin in 1954 also discussed how to think about the effect of sampling from finite
population of units and treatments into experimental designs. Searle (1970) developed a general
rule to change from infinite to finite population for the expected value of mean square error.

However, none of them discussed how the prediction of the random effects under the
randomization models with sampling from finite populations of units and treatments.

When it comes to the prediction problem, Henderson C.R. (1950) put forth a model as

$$y_i = \sum_{i=1}^{p} b_i x_{ia} + \sum_{i=1}^{q} u_i z_{ia} + e_{ia}$$

Where $b_i$ are unknown fixed parameters, $x_{ia}$ and $z_{ia}$ are observable parameters, $u_i$
are a random sample from a multivariate normal distribution with means zero and covariance
matrix $\sigma^2$, and $e_{ia}$ are normally and independently distributed with means zero and variance $\sigma^2_{e_{ia}}$. Henderson thought that maximum likelihood solution lead to a satisfactory estimation procedure
of $\mu$ assuming the variance components are known comparing with solutions by least square
method.

Best linear unbiased prediction (BLUP), in general, is a method of estimating random
effects. BLUP estimates of the realized values of the random variables are linear in the sense that
they are linear function of the data $y$; unbiased in the sense that the average value of the estimates
is equal to the average value of the quantity being estimated; best in the sense that they have
minimum mean squared error within the class of linear unbiased estimators; and the predictors
used to distinguish them from estimators of the fixed effects (Robinson 1991).

When it comes to the difference between prediction and estimation, it has become common
practice to “estimate” fixed effects and to “predict” random effects. But Henderson in 1950 still
used estimation both for fixed quantity and random variables. Henderson in 1984 expressed his
doubts about the appropriateness of the terminology:

“Which is the more logical concept, prediction of a random variable or estimation of the
realized value of a random variable? If we haven animal already born, it seems reasonable to
describe the evaluation of its breading value as an estimation problem. On the other hand, if we
are interested in evaluating the potential breeding value of a mating between two potential parents,
this would be a problem in prediction. If we are interested in future records, the problem is clearly
one of prediction.”

Robinson preferred understanding prediction as estimation of a realized value. He indicated
to use “estimation” of both fixed and random effects.

When trying to find BLUP, Henderson (1950,1959, 1973) described the BLUP estimates as
being “joint maximum likelihood estimates”. He brought forth mixed model equations (MME)
and mixed model solutions. He actually assumed that the random effects and error items are
normally distributed and to maximize the joint density function of $y$ and $u$ with respect to
$\beta$ and $u$. Dempfle (1977) and Lindley and Smith (1972) presented the Bayesian Derivation of
BLUP. They assumed that the prior density is a normal distribution; hence posterior density is
proportional to the joint density in Henderson’s derivation. Goldberger’s Derivation in 1962 gave
a clear definition of best linear unbiased predictors: suppose the best linear unbiased prediction
of $y_\ast$, if $y$ is the vector of the observations, we seek the linear predictor $p = c'y$ such that
$E( p - y_\ast)( p - y_\ast)'$ is a minimum subject to $E( p - y_\ast) = 0$. Here “best” is subject to
minimizing the mean squared errors. Goldberger’s derivation only based on the first two moments.
The predictors he got are the same as Henderson’s predictors form the normality assumption.
One way random effects model could be regarded as a special case of the mixed model

where $X \beta$ reduce to $\mu$. Without effort, we could get the predictors from Henderson’s Mixed model solution: 

$$BLUP(u + a_i) = \text{GLSE}(u) + \frac{n_i \sigma_a^2}{\sigma_e^2 + n_i \sigma_a^2} (\bar{y}_i - \text{GLSE}(u))$$

assuming $\sigma_a^2$ and $\sigma_e^2$ are known. Box and Tiao (1968) presented a detailed derivation of BLUP for balanced random effects model: 

$$BLUP(u + a_i) = \bar{y}_i - \frac{\sigma_e^2}{\sigma_e^2 + n\sigma_a^2} (\bar{y}_i - \bar{y})$$

which is agreement with the result from Henderson method turning to the balanced cases.

The solutions of the predictors employed difference criteria in the sense of “best”:

Henderson: (1950,1959) Kempthorne (1959): minimized the maximum likelihood function ;
Bayes method, minimized Bayes risk; Goldberger, minimized the mean squared errors only using first two moments.

None of the above approaches account for the impact of a finite population on the predictors.

Predictors of linear combinations of responses of elements of a finite population in a two stage sampling setting were developed by Scott and Smith (1969) using a super population model, according to which a finite population is viewed as the realization of a set of random variables.

Predictors are constructed using the joint distribution assumed for the super population. The first and second moment assumptions of the super-population match the mixed model assumptions.

This approach to statistical inference is based on the model assumed for the super population. Scott and Smith (1969) presented a Bayesian derivation (assuming that the super population is normally distributed), and a distribution-free derivation based on minimizing the expected mean squared error of a linear predictor.

Predictors in finite population are also considered in the setting of random permutation
Random permutation models are used to provide a probabilistic “link” to relate a sample to its parent population (Stanek 2003). Permutation models were discussed more comprehensively in the earlier work of Cassel, Sarndal and Wretman (1977) and Rao and Bellhouse (1978). This framework does not require assumptions about parametric distributions. The permuted population consists of random variables corresponding to permutations of units in the finite population, where permutations occur with equal probabilities. The population is represented as a non-stochastic vector. The randomness is thus attributable to random sampling. Under this framework, the structural relationship between the population and the permuted one can be explicitly represented using matrices. (Li, wenjun, 2003). The random permutation model Cassel et. 1977 mentioned is a random permutation super-population models which has been discussed Mukhopadhyay (1984) and Rao (1984), Padmawar and Mukhopadhyay (1985). Stanek and Julio (2004) introduced the design-based approach to inference under simple random sampling of a finite population which encompasses a simple random permutation super-population model. A unified approach to estimation and prediction under simple random sampling in such design-based random permutation model was discussed by (Stanek 2004) . Additionally, this method has been applied to predicting random effects from finite population clustered samples with response error (Stanek 2004) and applied to rate estimation and standardization (Li, Wenjun 2003) .

One step further in the prediction approach is to predict random effects when the variance components are unknown. It seems to be the one step toward an ultimate goal of finding good predictors for whatever the general mixed linear model and the random effects model. Some estimators of the shrinkage constants have been developed, for example, analysis of variance (ANOVA) estimator (Searle 1971, Ch.9), James-Stein positive-part estimator (Efron and Morris
1973) and so on. Peixoto and Harville in 1986 made their effort to compare alternative predictors under the Balanced One-Way Random Model. They used ANOVA estimators, James-Stein positive-part estimator, the maximum likelihood (ML), restricted maximum likelihood (REML), and modified – REML estimators. They found when the intra-class correlation is small, the truncated restricted maximum likelihood predictor would appear to be better than others via simulation study and provide a smaller expected MSE. But the overall performance is difficult to evaluate since the behaviors of different predictors will change with the true intra-class correlation, the replicates and levels of the factor and the variance of errors $\sigma^2$.

What I am going to do in my thesis is to apply the prediction approach in the random permutation models to one factor completely randomized design considering the permutation of finite treatments. This approach is distribution free and design-based.

The preliminary result shows that shrinkage constants go into the formulae of predictors for the treatment mean. When the variance is unknown, if we simply plug in the method of moment estimators in stead of the variance components, the MSE is no longer smaller than that from simple sample mean. It is a critical point that how to estimate the shrinkage problem.

Further needs to say about the expanded model and Rao Bellhouse Theorem.

Usually, under the assumption of simple random sampling without replacement, the typical random permutation probability model assigns equal probability to all permutations of the finite population units. We index each unit’s position in the permutation by $i = 1, \ldots, N$. The value in position $i$ for a randomly selected permutation is defined by the realization of the random variable $Y_i = \sum_{j=1}^{N} U_{ij} Y_j$. This representation of random variables does not allow units to be identified and hence does not permit inference about unit parameters. The expanded model has
been proposed by Stanek, Singer, Lencina in 2004. They expanded $Y_i = \sum_{j=1}^{N_i} U_{ij} y_j$ to individual variables $\left( U_{i1} y_1 \ U_{i2} y_2 \ \ldots \ U_{iN} y_N \right)$. An advantage of the expanded model is the ability to identify a parameter associated with a labeled unit. The question pops up: “Do we need to use expanded model to develop our predictors? In other words, are the predictors from the typical model sufficient and optimal?” Rao Bellhouse Theorem provides a way to solve this question.

Rao and Bellhouse (1978) state in Theorem 1.1 conditions under which an estimator is optimal in a space. They define the expectations with respect to a probability design, a model and response error using the subscripts $p, m, r$. The theorem is as follows:

Let $C$ denote a class of $pm$ – unbiased estimators of $\mu$. Let $C_0$ denote a class of $pm$ – unbiased estimators, $e_d$ of zero. An estimator of $\overline{e}_d$ in $C$ is “optimal” for $\mu$ if and only if for every estimator $e_d$ of zero belonging to $C_0$ it is true that $E(\overline{e}_d - \mu) e_d \equiv 0$.

This theorem provides a way to testify that an estimator under certain condition is optimal or not. It could be used to see during the data reduction whether the information is lost, in other words, whether the data summary is sufficient or not. In the random permutation model, $Y_i = \sum_{j=1}^{N_i} U_{ij} y_j$ could be seen as a reduction of $\left( U_{i1} y_1 \ U_{i2} y_2 \ \ldots \ U_{iN} y_N \right)$. Stanek 2006 applied this theorem to show that a predictor of a linear combination of PSUs can be optimally developed from the usual random variables, as opposed to the expanded random variables.

In my thesis, the expanded model also will be used. The Rao Bellhouse theorem could help me to compare the predictors from typical random permutation model and expanded one. What’s more, in order to simplify the calculation, the observed values in each treatment will be collapsed to the sample mean. We can also apply the theorem to see whether there is any information lost.
during the collapsed process.

3. Proposal

This research will investigate using random effect models for treatments in experimental design settings and the implication of inference from such models relative to inference from using fixed effect models for treatments. The results will be developed in a design-based random permutation model framework. The objectives of this research can be grouped into five areas. First, optimal predictors of linear combinations of realized treatment means will be developed and compared with similar linear combination of treatment means from fixed effect models. Next, we will develop an expanded model for a random treatment effect setting, and develop predictors in this context. Third, we will evaluate the sufficiency of model frameworks and compare expanded with collapsed random effect models. Fourth, measurement error will be added to the modeling framework. Finally, as time permits, we will evaluate more general models.

3.1 Develop the predictors of realized treatment parameters

The first objective of this research is development of the predictors of a linear combination of realized treatment parameters in a 1 factor completely randomized design using a random permutation model prediction framework. The potential population that we consider is the $N \times A$ array in which each element is non-stochastic. Parameters can be parameterized as

$$y_{sa} = \mu + \beta_s + \alpha_a + \epsilon_{sa}$$

where $\beta_s = \mu_s - \mu$ as the effect of unit $s$, $\alpha_a = \mu_a - \mu$ as the effect of treatment $a$, $\epsilon_{sa}$ as the interaction between unit $s$ and treatment $a$. We will use $U$ and $V$ to represent a joint permutation of units and treatments, and derive the joint random permutation
model for the population as \( Y = (Y_j) = UyV' \) where \( Y_j \) indicates the response of PSU \( i \) in the treatment group \( j \).

The BLUP approach in a finite population will be used in derivation of the predictors. The theoretical mean squared error can be obtained from the random permutation model. Evaluate the predictors relative to other competitors via comparison of the expected mean squared error (MSE) will be made theoretically and via simulation. Finally, we plan to develop and evaluate empirical predictors assuming unknown variance components. We will estimate shrinkage constants and use them to obtain empirical predictors. Comparison and evaluation of the empirical predictors will be made via simulation and via theoretical approximation, where possible.

### 3.2 Evaluate expanded model

An expanded random permutation model has been proposed by Stanek and Singer, et al. (2004). The main feature is characterizing the problem with an expanded set of random variables induced by the design-based approach. An advantage of the expanded model is the ability to identify a parameter associated with a labeled unit, or treatment.

We propose to develop an expanded model in a one way experimental study. Since a random variable can be defined as \( Y_{ij} = U'yV_j = \sum_{s=1}^{N} \sum_{a=1}^{d} U_{is} V_{ja} y_{ia} \), we propose not to expand all the individual random variables but to expand random variable \( V_j \) in \( Y_{ij} \) so as to express it as a vector of the random variables as \( V_j \left( \sum_{s=1}^{N} U_{is} y_{ia} \right) V_j \left( \sum_{s=1}^{N} U_{is} y_{ia} \right) \ldots V_{jd} \left( \sum_{s=1}^{N} U_{is} y_{ia} \right) \).

Hence the number of random variables will be expanded from \( NA \) to \( NA^2 \). We plan to use similar methods to develop the joint probability of these random variables as in Part 3.1 and derive
the predictors. We will evaluate MSE in the expanded model and compare the results with the
typical random permutation model in Part 3.1. Hence we will use these results to help understand
clearly whether there is improvement by using the expanded random variables.

3.3 Apply Rao-Bellhouse theorem

We plan to apply a theorem due to Rao Bellhouse (1976) to investigate the sufficiency of
the space spanned by random variables for a predictor. The predictors we consider are linear
combinations of realized treatment means. In this proposal we are going to use three models: the
expanded random permutation model with $NA^2$ random variables, the typical random
permutation model with $NA$ random variables, and the typical random permutation model
collapsed to the sample and remainder means (consisting of $2A$ random variables). We propose
to use the Rao Bellhouse Theorem to compare these three frameworks in order to see whether a
lower dimensional framework is sufficient to develop optimal predictors.

3.4 Allow measurement errors

The population parameters in 3.1-3.3 are defined by $y_{sa} = \mu + \beta_s + \alpha_a + \epsilon_{sa}$. No
measurement error is considered. Since the measurement error for the observation is an important
problem and may not be avoided, we propose to investigate this problem. We will add
measurement error to the unit, and represent the response by $Y_{sak} = \mu + \beta_s + \alpha_a + \epsilon_{sa} + W_{sak}$
where $W_{sak}$ represents response error with zero expected value and response error variance $\sigma_{sa}^2$. A
similar strategy of permuting treatments and units is applied by using $U$ and $V$. Follow the same
plan in Part 3.1 we attempt to find the predictors in the framework allowing measurement error.
3.5 Investigate general models

If time permits, we will investigate more general models. One more general model that we will consider is a model that includes an interaction item $e_{n_1}$. Another situation that we will consider occurs when samples are of unequal size for treatments. For each setting, we will consider all of the steps in 3.1-3.4.

4. Example

We illustrate the ideas and preliminary results proposed for this research with an example. The example is small and artificial, but conveys the main ideas of the research. Table 1 describes an example with a population size of $N = 5$. Two treatments will be assigned at random. We assume that there is no interaction between units and treatments and STUVA holds. Our goal is to find predictors for the means of treatment 1 and treatment 2. $R$ indicates the sampling mechanism based on a simple random sampling without replacement. Suppose the sample size $n$ is equal to 2 for each treatment. One of the realized values of $R$ is listed in Table 1. It indicates units 2 and 5 are assigned to treatment 1 and units 1 and 4 are assigned to treatment 2. Then, we will observe the values without parenthesis.
Table 1: Artificial example of potential observable responses

<table>
<thead>
<tr>
<th>Subject (s)</th>
<th>$Y_{s1}$</th>
<th>$Y_{s2}$</th>
<th>$Y_{s1} - Y_{s2}$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>15</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>30</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>20</td>
<td>20</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>35</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>65</td>
<td>45</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>49</td>
<td>29</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

This realized sample is summarized in Table 2. $\sigma^2_s$ is defined as the variance of units within each treatment. $\sigma^2_A$ is defined as the variance of the treatments. $f$ is the fraction factor which is defined by $\frac{n}{N}$. The sample mean for treatment 1 is 42.5; the sample mean for treatment 2 is 37.5.

Under our framework, the predictor for a realized treatment mean is given by:

$$\hat{P}_j = \overline{Y}_j + \frac{\sigma^2_s}{\sigma^2_A + \frac{\sigma^2_S}{n}} (\overline{Y}_j - \overline{Y}_)$$

Based on this formula, our predictor for treatment 1 is 41.4, and the predictor for treatment 2 is 38.6. Compared the sample means and the predictors, we find the predictors we developed seems more close to the true means.

Table 2: Summary of Sample 1

<table>
<thead>
<tr>
<th>Unit (s)</th>
<th>Response $y_{is}$</th>
<th>True Means</th>
<th>$\hat{P}_j$</th>
<th>Sample Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRT 1</td>
<td>2</td>
<td>50</td>
<td>49</td>
<td>53.2</td>
</tr>
<tr>
<td>TRT 1</td>
<td>5</td>
<td>65</td>
<td>49</td>
<td>53.2</td>
</tr>
<tr>
<td>TRT 2</td>
<td>1</td>
<td>15</td>
<td>29</td>
<td>29.3</td>
</tr>
<tr>
<td>TRT 2</td>
<td>4</td>
<td>35</td>
<td>29</td>
<td>29.3</td>
</tr>
</tbody>
</table>

$$\sigma^2_s = 142.5 \quad \sigma^2_A = 200 \quad f = \frac{n}{N} = 0.4$$
In order to answer whether this result is inevitable or not, we list all the possible samples and for each sample, we calculate sample means, the predictors and the sum of squares (SSEs) and MSEs for sample means and our predictors respectively. Average MSEs from all the samples, we obtain the expected MSEs of sample means and our predictors. The comparison between two expected MSEs can help evaluate the behaviors of the predictors we develop.

Table 3 lists all the possible samples. The mean squared errors for sample means and our predictors are obtained as 42.75 and 33.39 respectively. This indicates our predictors provide a smaller expected MSE than the sample mean.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>trt1</td>
<td>1</td>
<td>35</td>
<td>49</td>
<td>42.5</td>
<td>40.53</td>
<td>42.25</td>
<td>71.74</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>trt1</td>
<td>2</td>
<td>50</td>
<td>49</td>
<td>42.5</td>
<td>40.53</td>
<td>42.25</td>
<td>71.74</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>trt2</td>
<td>3</td>
<td>20</td>
<td>29</td>
<td>27.5</td>
<td>29.47</td>
<td>2.25</td>
<td>0.22</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>trt2</td>
<td>4</td>
<td>35</td>
<td>29</td>
<td>27.5</td>
<td>29.47</td>
<td>2.25</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>1</td>
<td>trt2</td>
<td>5</td>
<td>45</td>
<td>29</td>
<td>40</td>
<td>40.66</td>
<td>121.00</td>
<td>135.88</td>
</tr>
<tr>
<td>240</td>
<td>1</td>
<td>trt2</td>
<td>4</td>
<td>35</td>
<td>29</td>
<td>40</td>
<td>40.66</td>
<td>121.00</td>
<td>135.88</td>
</tr>
<tr>
<td>240</td>
<td>2</td>
<td>trt1</td>
<td>3</td>
<td>40</td>
<td>49</td>
<td>45</td>
<td>44.34</td>
<td>16.00</td>
<td>21.68</td>
</tr>
<tr>
<td>240</td>
<td>2</td>
<td>trt1</td>
<td>2</td>
<td>50</td>
<td>49</td>
<td>45</td>
<td>44.34</td>
<td>16.00</td>
<td>21.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average</td>
<td>42.75 33.39</td>
</tr>
</tbody>
</table>

5. Preliminary Results

5.1 Population

We use the potentially observable population concept of Rubin (2005) in formulating the
problem. We assume that each unit could be potentially observed under each of the treatments, and represent the response for unit \( s \) given treatment \( a \) by the non-stochastic value \( y_{sa} \). The mean for treatment \( a \) is defined by \( \mu_a = \frac{1}{N} \sum_{s=1}^{N} y_{sa} \). The mean for unit \( s \) is defined by \( \mu_s = \frac{1}{A} \sum_{a=1}^{A} y_{sa} \). The overall mean over all the potential observable values is

\[
\mu = \frac{1}{NA} \sum_{a=1}^{A} \sum_{s=1}^{N} y_{sa}.
\]

### 5.2 Parameterization

We define additional parameters in terms of these basic values. First, define \( \beta_s = \mu_s - \mu \) as the effect of unit \( s \), \( \alpha_a = \mu_a - \mu \) as the effect of treatment \( a \), \( \varepsilon_{sa} \) as the interaction between unit \( s \) and effect of treatment \( a \). As a result,

\[
y_{sa} = \mu + \beta_s + \alpha_a + \varepsilon_{sa} \quad (1).
\]

We represent the potentially observable responses for treatment \( a \) by \( y_a \) as \( y_a = (y_{1a}, y_{2a}, \ldots, y_{Na})' \), and the \( N \times A \) matrix of potentially observable responses as \( y = (y_1, y_2, \ldots, y_A) \). The vector of treatment parameters is given by

\[
\mu = \frac{1}{N} 1_N y'
\]

where \( 1_N \) is \( N \times 1 \) vector with all elements equal to one. We can express \( y \) in terms of \( \mu \), \( \beta = (\beta_1, \beta_2, \ldots, \beta_N)' \), \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_A)' \) and \( \varepsilon = (\varepsilon_{sa}) \):

\[
y = 1_N 1_A' \mu + \beta 1_A' + 1_N \alpha' + \varepsilon \quad (2)
\]

We define the co-variance matrix of the response for treatments as \( \frac{N-1}{N} \Sigma \), where

\[
\Sigma = (\Sigma_{aa'})_{a,A} = \frac{1}{N-1} y'y_N \text{ and } \Sigma_{aa'} = \frac{1}{N-1} \sum_{s=1}^{N} (y_{sa} - \mu_a)(y_{sa'} - \mu_{a'}) \text{ for all } a, a'
\]
\(a = 1, \ldots, A\) and \(a^* = 1, \ldots, A\). We define \(P_K = I_K - \frac{1}{K} J_K\) where \(I_K\) is \(K \times K\) identity matrix and \(J_K = I_K I_K'\). Further, we define \(\bar{\sigma}^2 = \frac{1}{A} \text{tr}(\Sigma) = \frac{\sum_{a=1}^{d} \sigma_{aa}}{A}\). Similarly, we define the variance of the response for unit \(s\) as \(\frac{A-1}{A} \sigma_{ss}\) where \(\sigma_{ss} = \frac{1}{A-1} \sum_{a=1}^{d} (y_{sa} - \mu_s)^2\), and the covariance for unit \(s\) and unit \(s^*\) as \(\frac{A-1}{A} \sigma_{ss^*}\), where

\[
\sigma_{ss^*} = \frac{1}{A-1} \sum_{a=1}^{d} (y_{sa} - \mu_s)(y_{s^*a} - \mu_{s^*}),
\]
and summarizing the covariance as \(\frac{A-1}{A} \Sigma_s\) where

\[
\Sigma_s = \left(\left(\sigma_{ss^*}\right)\right) = \frac{1}{A-1} yP_s y'.
\]

### 5.3 Derive the Random Permutation Model

We use \(U\) and \(V\) to represent a joint permutation of units and treatments. Explicitly, let

\[
U = (U_1 \ U_2 \ \cdots \ U_N)' = (\left(\left(U_i\right)\right))' \quad \text{and} \quad V = (V_1 \ V_2 \ \cdots \ V_N)' = (\left(\left(V_j\right)\right))'.
\]

Using the property of the permutation matrix (Stanek, Argentina2006-lec1a.doc), for any permutation matrix, we get

\[
E(U) = \frac{1}{N} J_N
\]

\[
\text{Var}_U\left[\text{vec}(U)\right] = \frac{1}{N-1} (P_N \otimes P_N)
\]

\[
\text{Var}(U_i) = \frac{1}{N} P_N
\]

\[
\text{Cov}(U_i, U_{i^*}) = -\frac{1}{N(N-1)} P_N \quad \text{when} \ i \neq i^*
\]

\[
E(U_i U_{i^*}) = \frac{1}{N} I_N \quad \text{when} \ i = i^* \quad (3)
\]

\[
E(U_i' U_{i^*}) = \frac{1}{N(N-1)} (J_N - I_N) \quad \text{when} \ i \neq i^* \quad (4)
\]

We define \(P_N = I_N - \frac{1}{N} J_N\) where \(I_N\) is \(N \times N\) identity matrix and \(J_N = I_N I_N'\).
Similar property occurs for $\mathbf{V}$.

We define the joint random permutation model for the population as

$$\mathbf{Y} = \left(\mathbf{Y}_i\right) = \mathbf{UyV}$$

where $Y_{ij}$ indicates the response of PSU $i$ in the treatment group.

Using (1),

$$\mathbf{Y} = \mathbf{U1_N1'_A V'\mu + U\boldsymbol{\beta}1'_\mathbf{V} + U1_N\mathbf{\tau V' + E}}$$

(5)

where $\mathbf{E} = \mathbf{UEV'}$. Note that $\mathbf{U1_N} = 1_N$, $1'_\mathbf{A V} = 1'_\mathbf{A}$. As a result,

$$\mathbf{Y} = 1_N1'_\mathbf{A V'}$$

(6)

We also define $\mathbf{B = U\beta}$ and $\mathbf{T = V\tau}$. Then

$$\mathbf{Y} = 1_N1'_\mathbf{A V'} + \mathbf{B1'_\mathbf{A} + 1_N T' + E}$$

(7)

We refer to this as the random one factor permutation model.

5.4 Expectation and Variance:

We determine the expected value and the variance of the random one factor permutation model next. We use $E_{uv}(\mathbf{Y}) = E_{uv}\left(E_{uv}(\mathbf{Y})\right)$ to determine the expected value. Using the definition of $\mu$, $E_{uv}(\mathbf{Y}) = \mu1_N1'_\mathbf{A}$.

We use a column expansion of $\mathbf{Y}$ to develop an expression for the variance.

$$Var_{uv}\left[vec(\mathbf{Y})\right] = Var_{uv}\left(E_{uv}\left[vec(\mathbf{Y})\right]\right) + E_{uv}\left(Var_{uv}\left[vec(\mathbf{Y})\right]\right)$$

$$Var_{uv}\left[vec(\mathbf{Y})\right] = \frac{A\sigma^2}{A-1}(\mathbf{J}_A - 1_N) \otimes \mathbf{P}_N + \frac{A\sigma^2}{A-1} \mathbf{P}_A \otimes \mathbf{P}_N + \sigma^2 \mathbf{P}_A \otimes \mathbf{J}_N$$

5.5 Simplifying the Model Assuming no Interaction:

In order to simplify the problem, we first assume $\epsilon_{sa} = 0$ (No interaction between effect of units and effect of treatments. This implies the treatment effect will not change with respect to
different unit and the unit effect will not change with respect to different treatments). With this assumption we simplify the model (1) as \( y_{au} = \mu + \beta_a + \alpha_a \).

As a result, \( \sigma_{au} = \frac{1}{N-1} \sum_{u=1}^{N} (y_{au} - \mu_a)(y_{au} - \mu_a) = \frac{1}{N-1} \sum_{u=1}^{N} \beta_a^2 = \sigma_S^2. \) \( \Sigma \) can be simplified as \( \Sigma = \sigma_S^2 J_A \) and \( \sigma^2 = \sigma_S^2. \) Then

\[
\text{Var}_{uv} \left[ \text{vec}(Y) \right] = \sigma_S^2 J_A \otimes P_N + \sigma_A^2 P_A \otimes J_N.
\]

### 5.6 Re-arranging Terms into the Sample and Remainder

We re-arrange and collapse the random variables into a set corresponding to the weighted sample totals, \( Y'_I \), and a set corresponding to the weighted remaining totals for each treatment,

\[
Y'_I \text{ such that } \begin{pmatrix} Y'_I \\ Y''_I \end{pmatrix} = \begin{pmatrix} K_I \\ K''_I \end{pmatrix} Y^* \text{ where } Y^* = \text{vec}(Y)
\]

and \( \begin{pmatrix} K_I \\ K''_I \end{pmatrix} = \frac{1}{N} \begin{pmatrix} \otimes_{j=1}^A \left( \delta'_j \otimes 1_n' \right) & 0 \end{pmatrix}_{1 \times (N-a_A)} \) and \( \begin{pmatrix} Y'_I \\ Y''_I \end{pmatrix} = \begin{pmatrix} fY_I \\ (1-f)Y''_I \end{pmatrix} \).

We use these expressions to form the expected value and variance of \( \begin{pmatrix} Y'_I \\ Y''_I \end{pmatrix} \).

\[
E_{uv} \left[ \begin{pmatrix} Y'_I \\ Y''_I \end{pmatrix} \right] = \begin{pmatrix} f \otimes_{j=1}^A \mu \\ (1-f) \otimes_{j=1}^A \mu \end{pmatrix}
\]

\[
\text{var}_{uv} \left[ \begin{pmatrix} Y'_I \\ Y''_I \end{pmatrix} \right] = \begin{pmatrix} V_I & V_{I,II} \\ V'_{I,II} & V''_I \end{pmatrix}, \text{ where}
\]

\[
V_I = \frac{f}{N} \sigma_S^2 (I_A - fJ_A) + f^2 \sigma_A^2 P_A
\]

\[
V_{I,II} = \frac{f}{N} \left[ fJ_A - I_A \right] + f(1-f) \sigma_A^2 P_A
\]
\[ V_{II} = f \frac{\sigma^2_S}{N} [I_A - fJ_A] + (1 - f)^2 \sigma^2_A P_A. \]

5.7 Find the BLUP

Refer to Argentina2006-lec1a.doc which gives the general method to find the BLUP.

For our special sample:

Target: \( P_j = \frac{1}{N} \sum_{i=1}^{N} Y_{ij} \)

Collapsing: \( \begin{pmatrix} Y^*_i \\ Y^*_{II} \end{pmatrix} = \begin{pmatrix} \bar{Y}_i \\ (1 - f) \bar{Y}_{II} \end{pmatrix} \)

We know \( E_{II} \begin{pmatrix} Y^*_i \\ Y^*_{II} \end{pmatrix} = \begin{pmatrix} f \bar{Y}_i \\ (1 - f) \bar{Y}_{II} \end{pmatrix}, \) \( \mu \) and \( \begin{pmatrix} V_I \\ V_{II} \end{pmatrix}, \) \( \begin{pmatrix} V'_{II} \\ V_{II} \end{pmatrix} \)

Now \( (g'_I, g'_{II}) = (e'_I, e'_j), \) \( X_i = f^1 A \) (8) and \( X_{II} = (1 - f) A \) (9).

\[ \hat{p}_j = g'_I Y^*_i + g'_{II} \left[ X_{II} \alpha + V_{II}^{-1} \left( Y^*_I - X_i \alpha \right) \right] \] where \( \alpha = \left( X_i' V_{II}^{-1} X_i \right)^{-1} X_i' V_{II}^{-1} Y^*_I. \)

After simplification, we obtain predictors as

\[ \hat{p}_j = \bar{y}_I + \frac{\sigma^2_A}{\sigma^2_A + \frac{\sigma^2_S}{n}} \left( \bar{y}_{II} - \bar{y}_i \right). \]

MSE of the Sample mean

\[ MSE(\bar{y}_{II}) = (1 - f) \frac{\sigma^2_S}{n} \]

MSE of the random permutation predictor

\[ MSE(\hat{p}_j) = MSE(\bar{y}_{II}) - \frac{A - 1}{A} \frac{\sigma^2_S}{n} \frac{\sigma^2_A}{\sigma_A^2 + \sigma_S^2} \]

\[ = MSE(\bar{y}_{II}) - \frac{A - 1}{A} \frac{A - 1}{A(\phi + 1) - 1} \frac{\sigma^2_S}{n} \]
Reference


