The Effect of Blocking on the Experimental Precision of Tomato Yield

Adam Taiye Mudi and Aliyu Usman
Department of Mathematics, Statistics and Computer Science, Kaduna Polytechnic, Kaduna

Abstract: Past experiments on the effect of fertilizer levels on tomato yield have frequently used Randomized Complete Blocked Designs (RCBD), even though the experimental areas seem relatively uniform with regard to external variables. Consequently, Completely Randomized Designs (CRD) may have had better precision with which to estimate treatment effects compared with RCBD. The objective of this study was to estimate the Relative Efficiency (ERE) of a RCBD compared with a CRD from the data sets obtained from Kwara State Agricultural Development Project, Nigeria for three farm trials with the aim of estimating the effect with less effort. The result of the analysis showed that ERE (RCBD to CRD) are (1.064, 0.9164 and 0.9068) respectively for the three trials. Hence blocking is inefficient for future similar experiment under the same conditions.

Keywords: CRD: completely randomized design, ERE: estimated relative efficiency, MSB,: mean square for blocking effect, MSE : error mean square, RCBD: randomized complete block design, SSB,: sum of squares for blocking effect, SSE,: error sum of squares

INTRODUCTION

When designing field experiments, blocking is often used to increase precision on treatment effects by reducing experimental error variance (Montgomery, 2009). Blocking stratifies experimental units into homogenous subgroups and in field trials, plots are normally blocked according to their proximity to each other. Blocking will increase treatment precision only if plots are blocked according to one or more varying external factors (irrigation gradient, soil nutrient status, etc.). If however, an experimental area is homogenous, blocking may actually decrease the precision of estimating treatment effects. This results from a Larger Means Square error (MS_{e}) term in the ANOVA since error degrees of freedom are reduced without a comparable reduction in error Sum of Squares (SS_{e}). In this situation, a Completely Randomized Design (CRD) would estimate treatment effects more precisely than a Randomized Complete Block Design (RCBD) (Anbari and Lucas, 1994). Block designs arise in comparative experimentation as fundamental devices for improving precision when working with heterogeneous experimental units. The blocks are simply a partition of the units into (say) b sets exhibiting homogeneity within sets. Given the blocks, of sizes k_1, k_2, k_3, ..., k_b, a block design is an assignment of v treatments to the \sum_{i=1}^{b} k_i experimental units. Optimality theory for block designs attempts to determine which of the many possible assignments is in some sense the best. The randomized block, Latin square and other complete block types of experiments may be inefficient for large number of treatments, because of their occasional failure to adequately minimize the effect of soil heterogeneity. Generally, the greater the heterogeneity within blocks, the poorer the precision of treatment effect estimates (Bisgaard and Steinberg, 2007). But in particular, it is of interest to investigate this precision with respect to some specific crops - tomatoes in this study.

Randomized Complete Block Design (RCBD): In a Completely Randomized Design (CRD), treatments are assigned to the experimental units in a completely random manner. The random error component arises because of all the variables which affect the dependent variable except the one controlled variable, the treatment (Federer, 1967). Naturally, the experimenter wants to reduce the errors which account for differences among observations within each treatment. One of the ways in which this could be achieved is through blocking. This is done by identifying supplemental variables that are used to group experimental subjects that are homogeneous with respect to that variable. These create differences among the blocks and make observations within a block similar. The simplest design that would accomplish this is known as a Randomized Complete Block Design (RCBD). Each block is divided into k sub-blocks of equal size. Within each block the k treatments are assigned at random to the sub-blocks. The design is complete in the sense that each block contains all the k treatments. The following layout shows a RCBD with k treatments and b blocks (Gomez and Gomez, 1984). There is one observation per treatment in each block and the treatments are run in a random order within each block. Hence, the statistical model is:
where, \( y_{ij} \) = The \( j \)th observation on treatment \( i \)
\( \mu \) = The overall mean
\( \tau_i \) = The \( i \)th treatment effect
\( \beta_j \) = The \( j \)th block effect
\( \varv_{ij} \) = A random error term associated with the \( ij \)th observation

Relative efficiency of the randomized complete block design: As with CRD, the expected mean squares for treatment and error are functions of the sample sizes \((b, \text{ the number of blocks})\), the true treatment effects \((\alpha_1, \ldots, \alpha_t)\) as well as the variance of the random error terms \(\sigma^2\) (Ganju and Lucas, 1999). By assigning all treatments to units within blocks error variance is smaller for RCBD than CRD (which combines block variations and random error into error term). The relative efficiency answers the question of how many times as many replicates would be needed for CRD to have as precise of estimates of treatment means as RCBD does? (Kuehl, 1994). Moreover, this gives the notion of Relative Efficiency (RE) in the context of estimation of treatment comparisons. For two designs, \(D_1\) and \(D_2\) say, the RE of \(D_1\) to \(D_2\) is defined as follows:

\[
RE(D_1, D_2) = \frac{\text{Efficiency}_{D_1}}{\text{Efficiency}_{D_2}}
\]

or

\[
RE(D_1, D_2) = \frac{\text{Var}_{D_2}}{\text{Var}_{D_1}}
\]

\[
RE(D_1, D_2) = \frac{\text{MSE}(D_1)}{\text{MSE}(D_2)}
\]

where; \(\text{Var}_{D_i}\) refers to \(\sum c_i \varv_i^2\) for design \(D_i\) \((i = 1, 2)\). In our case \(D_1\) is RCBD with \(t\) treatments and \(b\) blocks and \(D_2\) is a CRD with \(r\) replications for each of the \(t\) treatments. The RE defined depends on the true variances for the two designs which, of course, are unknown. The best we can do then is to obtain the estimated RE, which we shall denote by ERE (Hinkelmann and Kempthorne, 2007).

Motivation: There is no single blueprint for knowing whether a particular plant will be more efficient under the CRD or the RCBD. Hence, the suitable design is governed by the notion of “fitness for purpose”. The purpose of this study is to determine the effect of blocking in tomato crop so as to know the suitability of either the CRD or the RCBD. More precisely, to determine the relative efficiency of the CRD to the RCBD to the tomato crop so as to choose a more suitable design for it.

### METHODOLOGY

Using the concept of uniformity trials, that is pooling the treatment sums of squares with appropriate error sums of squares from in the ANOVA table for the respective designs, we have the respective ANOVAs for uniformity
trial which shall be used to estimate the relative efficiency of the RCBD (Table 1). However, we need to outline the variance component of CRD as a basis of comparison as follows:

\[ \text{MSE} = \frac{\text{SSR}}{n(n-1)} - \frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{n} x_{ij}^2 - 2 \sum_{j=1}^{n} x_{i} \right)}{n(n-1)} \]

\[ E(\text{MSE}) = \frac{1}{p(n-1)} \left[ \sum_{i=1}^{p} \sum_{j=1}^{n} x_{ij}^2 - 2 \sum_{j=1}^{n} x_{i} \right] \]

\[ E(\text{MSE}) = \frac{1}{p(n-1)} \left[ \sum_{i=1}^{p} \sum_{j=1}^{n} x_{ij}^2 - 2 \sum_{j=1}^{n} x_{i} \right] \]

In every designed experiment, we have:

\[ \sum_{i=1}^{p} x_{i} = 0 \quad E(\epsilon_i) = 0 \quad E(\epsilon_i^2) = \sigma^2 \]

Now;

\[ y_{ij} = \mu + t_i + e_{ij} \]

\[ \sum_{j=1}^{n} y_{ij} = n \left( \mu + t_i + e_{ij} \right) \]

\[ y_{i} = n \mu + n t_i + e_{i} \]

\[ y_{i}^2 = \left( \mu + t_i + e_{i} \right)^2 = \mu^2 + t_i^2 + e_{i}^2 + 2 \mu t_i + 2 \mu e_i + 2 t_i e_i \]

\[ \sum_{i=1}^{p} \sum_{j=1}^{n} x_{ij}^2 = \sum_{i=1}^{p} \sum_{j=1}^{n} E(\epsilon_i) + 2 \sum_{i=1}^{p} \sum_{j=1}^{n} t_i E(\epsilon_i) + \sum_{i=1}^{p} \sum_{j=1}^{n} t_i^2 E(\epsilon_i) \]

\[ E(\epsilon_i) = 2 \mu \sum_{i=1}^{p} \sum_{j=1}^{n} t_i E(\epsilon_i) + \sum_{i=1}^{p} \sum_{j=1}^{n} t_i^2 E(\epsilon_i) \]

\[ E(\epsilon_i^2) = \mu^2 + t_i^2 + e_{i}^2 + 2 \mu t_i + 2 \mu e_i + 2 t_i e_i \]

\[ \sum_{i=1}^{p} \sum_{j=1}^{n} x_{ij}^2 = \sum_{i=1}^{p} \sum_{j=1}^{n} E(\epsilon_i) + 2 \sum_{i=1}^{p} \sum_{j=1}^{n} t_i E(\epsilon_i) + \sum_{i=1}^{p} \sum_{j=1}^{n} t_i^2 E(\epsilon_i) \]

\[ E(\epsilon_i^2) = \mu^2 + t_i^2 + e_{i}^2 + 2 \mu t_i + 2 \mu e_i + 2 t_i e_i \]

\[ y_{i}^2 = \left( \mu + t_i + e_{i} \right)^2 = \mu^2 + t_i^2 + e_{i}^2 + 2 \mu t_i + 2 \mu e_i + 2 t_i e_i \]

\[ y_{i}^2 / n = \left( \mu + t_i + e_{i} \right)^2 / n = \mu^2 + t_i^2 + e_{i}^2 + 2 \mu t_i / n + 2 \mu e_i / n + 2 t_i e_i \]

\[ y_{i}^2 / n = \mu^2 + t_i^2 + e_{i}^2 / n + 2 \mu t_i / n + 2 \mu e_i / n + 2 t_i e_i \]

\[ \sum_{i=1}^{p} \left( y_{i}^2 / n \right) = \mu^2 + \sum_{i=1}^{p} \sum_{j=1}^{n} x_{ij}^2 / n + \sum_{j=1}^{n} e_{ij}^2 / n + 2 \mu \sum_{i=1}^{p} \sum_{j=1}^{n} t_i / n + 2 \mu \sum_{i=1}^{p} \sum_{j=1}^{n} e_i / n + 2 \sum_{i=1}^{p} \sum_{j=1}^{n} t_i / n \]

\[ E\left[ \sum_{i=1}^{p} \left( y_{i}^2 / n \right) \right] = np \mu^2 + n \sum_{i=1}^{p} t_i^2 + \frac{1}{n} \sum_{i=1}^{p} E(e_i^2) + 2 \mu \sum_{i=1}^{p} E(e_i) \]

\[ = E\left[ \sum_{i=1}^{p} \left( y_{i}^2 / n \right) \right] = np \mu^2 + n \sum_{i=1}^{p} t_i^2 + np \sigma^2 \quad (2) \]

This gives us a simple basis for comparison of RCBD and CRD as follows:

**Comparison of RCBD and CRD:** If the blocks were not used the estimated error variance would be:

\[ \text{SSR} / \text{SSR} = \frac{(b-1)\text{MSB} + b(1-1)\text{MSR}}{bt-1} \]

The estimated error variance with blocks is, of course, MS(R), so that:

\[ \text{ERE} / \text{CRD} = \frac{(b-1)\text{MSB} + b(1-1)\text{MSR}}{(bt-1)\text{MSR}} \]

Since we have carried out an experiment with real treatments and not dummy treatments we do not know MSR. Instead we only know MSE. Hence substituting MSE for MSR in the equation above we have:

\[ \text{ERE} / \text{CRD} = \frac{(b-1)\text{MSB} + b(1-1)\text{MSE}}{(bt-1)\text{MSE}} \]

where; MSE and MSB are obtained from the ANOVA for the completed experiment using an RCBD.

**Data presentation and analysis:** The following data were obtained from Kwara State Agricultural Development Project for three farm trials. The experiment is for testing the effect of fertilizer levels on tomato yield in kilogram/hectare using sites as blocking factor. The result is displayed in the Table 2, 3, 4 for the three farm trials.

The Microsoft Excel 2007 spreadsheet was used and produced the following ANOVA tables for the respective analyses (Table 5 to 10).

**Analysis 1 (Farm trial 1):**

\[ \text{ERE} / \text{CRD} = \frac{2(210508) + (3)(155431)}{(10)(155431)} = 10644 \]
Table 8: ANOVA table for trial 2

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>591.183</td>
<td>3</td>
<td>197.0608</td>
<td>9.03</td>
<td>0.0121</td>
</tr>
<tr>
<td>Blocks</td>
<td>23.582</td>
<td>2</td>
<td>11.7908</td>
<td>0.54</td>
<td>0.6086</td>
</tr>
<tr>
<td>Error</td>
<td>130.985</td>
<td>6</td>
<td>21.8308</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>745.749</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

t: 4; b: 3; MSB: 11.7908; MSE: 21.8308

Table 9: Randomized block descriptives for trial 3

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>n</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>11.8667</td>
<td>3</td>
<td>2.5580</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>18.2000</td>
<td>3</td>
<td>5.4342</td>
</tr>
<tr>
<td>Treatment 3</td>
<td>15.9000</td>
<td>3</td>
<td>4.4710</td>
</tr>
<tr>
<td>Treatment 4</td>
<td>8.2667</td>
<td>3</td>
<td>5.0836</td>
</tr>
<tr>
<td>Block 1</td>
<td>11.6500</td>
<td>4</td>
<td>6.2282</td>
</tr>
<tr>
<td>Block 2</td>
<td>14.8750</td>
<td>4</td>
<td>6.1646</td>
</tr>
<tr>
<td>Block 3</td>
<td>14.1500</td>
<td>4</td>
<td>5.2981</td>
</tr>
<tr>
<td>Total</td>
<td>13.5583</td>
<td>12</td>
<td>5.5390</td>
</tr>
</tbody>
</table>

t: 4; b: 3; MSB: 11.4508; MSE: 23.4853

Table 10: ANOVA table for trial 3

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>173.676</td>
<td>3</td>
<td>57.8919</td>
<td>2.47</td>
<td>0.1599</td>
</tr>
<tr>
<td>Blocks</td>
<td>22.902</td>
<td>2</td>
<td>11.4508</td>
<td>0.49</td>
<td>0.6365</td>
</tr>
<tr>
<td>Error</td>
<td>140.912</td>
<td>6</td>
<td>23.4853</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>337.489</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

t: 4; b: 3; MSB: 11.4508; MSE: 23.4853

Analysis 2 (Farm trial 2):

\[
ERE_{RCBD \text{ to } CRD} = \frac{2(117908) + (3)(3)(218308)}{(11)(218308)} = 0.9164
\]

Analysis 3 (Farm trial 3):

\[
ERE_{RCBD \text{ to } CRD} = \frac{2(114508) + (3)(3)(234853)}{(11)(234853)} = 0.9068
\]

DISCUSSION

The result showed that ERE (RCBD to CRD) are (1.064, 0.9164 and 0.9068) respectively for the three trials in Table 6 through Table 8 while the respective descriptive statistics were depicted in Table 2 through Table 4. Since the values are very close to 1, Completely Randomized Designs (CRD) may have had better precision with which to estimate treatment effects compared with RCBD. Combined descriptive statistics depicted in Table 9 and the corresponding ANOVA in Table 10. Hence, in particular in the area under study, for tomato crop, the CRD has a better precision.

CONCLUSION AND RECOMMENDATIONS

Hence future study on the effect of fertilizer levels, (same levels) on tomato yield using the same sites as the blocking factors may not be blocked. In a general note, this case of tomato yield has provided us with evidence that not in all cases that blocking could minimize experimental error. It is therefore important to investigate at the design stage whether blocking should be introduced or not. It is therefore recommended that experimenters and agronomist should subject individual crops to field experiments before choosing a suitable design for it.

REFERENCES


Bisgaard, S. and D.M. Steinberg, 2007. The design and analysis of \(2^k-p\times s\) prototype experiments, Technometrics, 39: 52-62.


