CHAPTER 4

Randomized Blocks, Latin Squares, and Related Designs
HomeWork Assignment
Due Sunday 2/5/2010

- Solve the following problems at the end of chapter 4:
  4-1
  4-7
  4-12
  4-14
  4-16
The Randomized Complete Block Design (RCBD)

Concepts to be discussed:

- Blocking and nuisance factors
- The randomized complete block design (RCBD)
- Extension of the ANOVA to the RCBD
- Other blocking scenarios...Latin square designs
The Blocking Principle

- **Blocking** is a technique for dealing with **nuisance factors**
- A **nuisance** factor is a factor that probably has some effect on the response, but it’s of no interest to the experimenter. However, the variability it transmits to the response needs to be minimized

- Typical nuisance factors:
  - Batches of raw material
  - Operators
  - Pieces of test equipment
  - Time (shifts, days, etc.)
  - Different experimental units

- Many industrial experiments involve blocking (or should)
- Failure to block is a common flaw in designing an experiment
The Blocking Principle

- If the nuisance variable is known and controllable, we use blocking.
- If the nuisance factor is known and uncontrollable, sometimes we can use the analysis of covariance (Chapter 15) to remove the effect of the nuisance factor from the analysis.
- If the nuisance factor is unknown and uncontrollable, we hope that randomization balances out its impact across the experiment.
- Sometimes several sources of variability are combined in a block, so the block becomes an aggregate variable.
The Hardness Testing Example

- We wish to determine whether 4 different tips produce different (mean) hardness reading on a Rockwell hardness tester
- Assignment of the tips to an experimental unit; that is, a test coupon
- Structure of a completely randomized experiment
- The test coupons are a source of nuisance variability
- Alternatively, the experimenter may want to test the tips across coupons of various hardness levels
- The need for blocking
The Hardness Testing Example

- To conduct this experiment as a RCBD, assign all 4 tips to each coupon
- Each coupon is called a “block”; that is, it’s a more homogenous experimental unit on which to test the tips
- Variability between blocks can be large, variability within a block should be relatively small
- In general, a block is a specific level of the nuisance factor
- A complete replicate of the basic experiment is conducted in each block
- A block represents a restriction on randomization
- All runs within a block are randomized
Suppose that we use $b = 4$ blocks:

We are interested in testing the equality of treatment means, but now we have to remove the variability associated with the nuisance factor (the blocks).
Extension of the ANOVA to the RCBD

- Suppose that there are $a$ treatments (factor levels) and $b$ blocks

- A **statistical model** (effects model) for the RCBD is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

\[ \begin{align*}
  i &= 1, 2, \ldots, a \\
  j &= 1, 2, \ldots, b
\end{align*} \]

- The relevant (fixed effects) hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \quad \text{where} \quad \mu_i = (1/b) \sum_{j=1}^{b} (\mu + \tau_i + \beta_j) = \mu + \tau_i$$
Extension of the ANOVA to the RCBD
ANOVA partitioning of total variability:

\[ \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \overline{y}_{..})^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} [(\overline{y}_{i.} - \overline{y}_{..}) + (\overline{y}_{.j} - \overline{y}_{..})] + \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2 \]

\[ = b \sum_{i=1}^{a} (\overline{y}_{i.} - \overline{y}_{..})^2 + a \sum_{j=1}^{b} (\overline{y}_{.j} - \overline{y}_{..})^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2 \]

\[ SS_T = SS_{Treatments} + SS_{Blocks} + SS_E \]
Extension of the ANOVA to the RCBD

- The degrees of freedom for the sums of squares in

\[ SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E \]

are as follows:

\[ ab - 1 = a - 1 + b - 1 + (a - 1)(b - 1) \]

- Ratios of sums of squares to their DOF result in mean squares
- The ratio of the mean square for treatments to the error mean square is an $F$ statistic that can be used to test the hypothesis of equal treatment means
ANOVA Display for the RCBD

Table 4-2  Analysis of Variance for a Randomized Complete Block Design

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>$F_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>$SS_{Treatment}$</td>
<td>$a - 1$</td>
<td>$SS_{Treatments}$</td>
<td>$a - 1$</td>
</tr>
<tr>
<td>Blocks</td>
<td>$SS_{Blocks}$</td>
<td>$b - 1$</td>
<td>$SS_{Blocks}$</td>
<td>$b - 1$</td>
</tr>
<tr>
<td>Error</td>
<td>$SS_E$</td>
<td>$(a - 1)(b - 1)$</td>
<td>$SS_E$</td>
<td>$(a - 1)(b - 1)$</td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T$</td>
<td>$N - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

see Equations (4-9) – (4-12)

Example 4-1: p131
Multiple Comparisons

- Significant difference in treatment means
- Use methods in CH3 for comparison between treatment means
- Replace the number of replicates in the single factor completely randomized design (n) by the number of blocks (b).
- The number of error DOF for the randomized block is (a-1)(b-1) instead of a(n-1)
Model Adequacy Checking - Residual

- The residual can be calculated from Eq 4-13
- See Table in p135 for the calculated residuals for example 4-1.
- Figure 4-4: normality plot (P136)
- No severe indication of non-normality, no outliers
- Figure 4-5: plot of residuals by tip type and by block
- If there is more scatter in the residuals for a particular tip, that could indicate that this tip produces more erratic hardness readings than the others
- More scatter in residuals for a particular coupon indicate that the coupon is not of uniform hardness
- Figure 4-6: residuals vs. fitted values
The Randomized Complete Block Design

Missing Value Problem

• Sometimes an observation in one of the blocks is missing due to:
  1. Carelessness
  2. Error
  3. Reasons beyond our control (damage of experimental unit)

• Two general approaches
  1. The approximate analysis
  2. The exact analysis
The Randomized Complete Block Design

Missing Value Problem - Approximate

- Missing observation is estimated and the ANOVA is performed just as if the estimated observation were real data, with error DOF reduced by 1.0

\[ y_{ij} = \text{missing observation for treatment i in block j} \]

\[ y_{..} = \text{grand total with one missing observation} \]

\[ y_{i.} = \text{total for the treatment with one missing observation} \]

\[ y_{.j} = \text{total for the block with one missing observation} \]
The Randomized Complete Block Design
Missing Value Problem - Approximate

- See table 7-4
- The missing observation \( x \) is estimated such that it will have min contribution to the error sum of squares

\[
SS_E = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2
\]

\[
= x^2 - \frac{1}{b} (y_{i.} + x)^2 - \frac{1}{a} (y_{.j} + x)^2 + \frac{1}{ab} (y_{..} + x)^2 + R
\]

- \( R \) includes all terms not involving \( x \)
The Randomized Complete Block Design
Missing Value Problem - Approximate

\[ x = \frac{ay_i' + by_j' - y'_..}{(a-1)(b-1)} \]

- Example: p140
The Randomized Complete Block Design
Missing Value Problem – Several Observation

Two missing values

1. Arbitrarily estimate the 1st missing value
2. Use this value along with the real data and Eq 4-16 to estimate the 2nd value.
3. Re-estimate the 1st missing value using eq 4-16
4. Repeat
5. steps (2) and (3) until convergence is obtained

- The error DOF are reduced by one for each missing value
Other Aspects of the RCBD
See Text, Section 4-1.3, pg. 130

- The RCBD utilizes an additive model – no interaction between treatments and blocks
- Treatments and/or blocks as random effects
- Missing values
- What are the consequences of not blocking if we should have?
- Sample sizing in the RCBD? The OC curve approach can be used to determine the number of blocks to run..see page 131
The Latin Square Design

- These designs are used to simultaneously control (or eliminate) two sources of nuisance variability
- A significant assumption is that the three factors (treatments, nuisance factors) do not interact
- If this assumption is violated, the Latin square design will not produce valid results
- Latin squares are not used as much as the RCBD in industrial experimentation
The Latin Square Design
Example: p144

- Study the effects of 5 different formulations of a rocket propellant on the observed burning rate. Each formulation is mixed from a batch of raw material that is only large enough for 5 formulations. Formulations are prepared by 5 different operators, whose skills may be substantially different in skills and experience.
- Two nuisance factors: batches of raw material and operators.
- Test each formulation exactly one in each batch of raw material and for each formulation to be prepared exactly once by each of the 5 operators.
- See table 4-9
Statistical Analysis of the Latin Square Design

- The statistical (effects) model is

\[ y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk} \]

\[ i = 1, 2, \ldots, p \]

\[ j = 1, 2, \ldots, p \]

\[ k = 1, 2, \ldots, p \]

\[ SS_T = SS_{Rows} + SS_{Columns} + SS_{Treatments} + SS_E \]

\[ p^2 - 1 = (p - 1) + (p - 1) + (p - 1) + (p - 2)(p - 1) \]

- The statistical analysis (ANOVA) is much like the analysis for the RCBD.
- See the ANOVA table, (Table 4-10)
- Example 4-3: p146
The Rocket Propellant Problem – A Latin Square Design

Table 4-8 Latin Square Design for the Rocket Propellant Problem

<table>
<thead>
<tr>
<th>Batches of Raw Material</th>
<th>Operators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>A = 24</td>
</tr>
<tr>
<td>2</td>
<td>B = 17</td>
</tr>
<tr>
<td>3</td>
<td>C = 18</td>
</tr>
<tr>
<td>4</td>
<td>D = 26</td>
</tr>
<tr>
<td>5</td>
<td>E = 22</td>
</tr>
</tbody>
</table>

- This is a 5×5 Latin square design
The Latin Square Design
Model Adequacy Checking

- Residuals
  \[ e_{ijk} = y_{ijk} - \hat{y}_{ijk} \]
  \[ \hat{y}_{ijk} = \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - 2\bar{y}_{...} \]

- Normality plot
- Plot of residuals
- Table 4-13: Standard Latin Squares and Number of Latin Squares of various sizes
For a a $p \times p$ latin square, the missing value may be estimated by:

$$y_{ijk} = \frac{p(y_{i..} + y_{.j.} + y_{..k}) - 2y_{...}}{(p - 2)(p - 1)}$$