Randomized Block Experiments

Chapter 9, Section 4
(Not Tested in Exam 2)
For Exam 2

• 6:30-7:30pm, this Wed.
• Prepare your **own crib sheet** (one piece of paper, two-sided, your own handwritten)
• Bring a **Calculator** (See policy from course website) and **pencils/pens**. Show your **student ID** in the end.

• Lab 5, next Wed
Overview

• One-way ANOVA has only one factor
  – “Factor” is a **categorical** variable (e.g. blood type)
  – Interested in whether there are differences between the levels/groups of that one factor

• Many times there are more factors to study:
  – 2 factors: 2-way ANOVA
  – 3 factors: 3-way ANOVA
  – etc.
  We will look at the case where there are two factors and one is a block factor ...
Block Factor (Always Categorical)

Nuisance factor may be present in the study

- Such factor has effect on the response but its effect is not of interest
- If it is unknown or not measurable, we can try to randomize the study or experiment to balance out its effect
- If known and measurable but uncontrollable, can use analysis of covariance
- If known and controllable, we can use blocking and it becomes a block factor
An Example

An experiment was designed to study the performance of four different detergents in cleaning clothes. The following cleanliness readings (higher=cleaner) were obtained with specially designed equipment for three different types of common stains. Is there a difference between the detergents?

<table>
<thead>
<tr>
<th></th>
<th>Stain 1</th>
<th>Stain 2</th>
<th>Stain 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detergent 1</td>
<td>45</td>
<td>43</td>
<td>51</td>
</tr>
<tr>
<td>Detergent 2</td>
<td>47</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>Detergent 3</td>
<td>48</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Detergent 4</td>
<td>42</td>
<td>37</td>
<td>49</td>
</tr>
</tbody>
</table>
Summarize the experiment:

In the example, we are interested in the detergents, the three type of stains may have effect but we are not interested in it. So:

- stain is a block factor
- all detergents are applied to the stains, this is called "complete"
- There are variations between blocks and within blocks that we need to take into consideration
Complete Randomized Block Experiment

- $b$ blocks, each consists of $a$ treatment levels
- $a$ treatment levels are randomly assigned within each block
- Results in restriction on randomization – randomization is done only in blocks
- Data within block are dependent on each other
- Introducing the block factor cost degree of freedom on SSE
Hypothesis Tests

Given a significance level $\alpha$:

For a complete randomized block experiment, we are usually interested in the following hypotheses:

- $H_0$ : There is no treatment effect
- $H_a$ : There is treatment effect

OR:

- $H_0$ : There is no block effect
- $H_a$ : There is block effect

How to test these hypotheses? Look at the variation sources first
Variation Sources (SS)

Suppose there are $a$ factor levels, and $b$ block levels.

Total variation:

$$SST = \sum_{i=1}^{a} \sum_{j=1}^{b} (x_{ij} - \bar{x})^2,$$
here $\bar{x}$ is the grand mean

Variation due to the differences in treatment levels:

$$SSTr = b \sum_{i=1}^{a} (\bar{A}_i - \bar{x})^2,$$
here $\bar{A}_i$ denotes the mean within each treatment

Variation due to the differences in block levels

$$SSB = a \sum_{j=1}^{b} (\bar{B}_j - \bar{x})^2,$$
here $\bar{B}_j$ denotes the mean within each block

Variation cannot be explained by treatment and block:

$$SSE = SST - SSTr - SSB$$
Convenient Formulas to Calculate SS

\[ x_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} x_{ij} \quad \text{sum of all data} \]

\[ A_i = \sum_{j=1}^{b} x_{ij} = b \bar{A}_i \quad \text{sum of all data in the } i\text{th factor level} \]

\[ B_j = \sum_{i=1}^{a} x_{ij} = a \bar{B}_j \quad \text{sum of all data in the } j\text{th block level} \]

We then have:

\[ SST = \sum_{i=1}^{a} \sum_{j=1}^{b} x_{ij}^2 - \frac{x_{..}^2}{n} \]

\[ SST_r = \frac{1}{b} \sum_{i=1}^{a} A_i^2 - \frac{x_{..}^2}{n} \]

\[ SSB = \frac{1}{a} \sum_{j=1}^{b} B_j^2 - \frac{x_{..}^2}{n} \]
Decomposing the df

For the variations we have:

\[ \text{SST} = \text{SSTR} + \text{SSB} + \text{SSE} \]

We may decompose the degrees of freedom accordingly:

\[ \text{SS} : \quad \text{SST} = \text{SSTR} + \text{SSB} + \text{SSE} \]

\[ \text{df} : \quad (n-1) = (a-1) + (b-1) + (a-1)(b-1) \]

notice here: \( n = ab \)

Comparing to the case when block factor is not included, the df of SSE is reduced.
Randomized Block ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A (treatment)</td>
<td>$a - 1$</td>
<td>SSA</td>
<td>MSA</td>
</tr>
<tr>
<td>Factor B (block)</td>
<td>$b - 1$</td>
<td>SSB</td>
<td>MSB</td>
</tr>
<tr>
<td>Error</td>
<td>$(a - 1)(b - 1)$</td>
<td>SSE</td>
<td>MSE</td>
</tr>
<tr>
<td>Total</td>
<td>$ab - 1$</td>
<td>SST</td>
<td></td>
</tr>
</tbody>
</table>
Two $F$ tests— 1. Blocking effective?

- $H_0$: There is no block effect
- $H_a$: There is a significant block effect

$$F = \frac{\text{variation from block}}{\text{variation due to error}} = \frac{MSB}{MSE} = \frac{SSB / (b-1)}{SSE / (a-1)(b-1)} \sim F_{df=(b-1,(a-1)(b-1))}$$

- If we **reject the null** hypothesis, it shows that the **blocking variable is important**
- **BUT**, since we don’t care about the blocking variable, we can interpret this as “the blocking was effective!”
Two $F$ tests—2. Treatment effect?

- $H_0$: There is no treatment effect
- $H_a$: There is a significant treatment effect

\[
F = \frac{\text{variation from treatment}}{\text{variation due to error}} = \frac{\text{MSTr}}{\text{MSE}} = \frac{\text{SSTr} / (a-1)}{\text{SSE} / (a-1)(b-1)}
\]

\[
\sim F(\ a-1, \ (a-1)(b-1))
\]

- If we **reject the null** hypothesis, it shows that **the treatment (or Factor A) is significant**

→ Results and interpretations are similar to One-Way ANOVA

- **Remark:** If the treatment effect is significant, we can follow up with multiple comparisons to see exactly which groups are significantly different.
Before You Run ANOVA: Assumptions

• **Same assumptions as ANOVA:**
  1. Constant variance
  2. Each of the $k$ populations follows a normal distribution

• **One additional assumption**
  3. There is no interaction between the treatment and blocking variables
    - Can assess just using common sense (Just ask: Do/should they interact?)
    - OR checking a Two-way ANOVA model briefly for a significant interaction (Will see how this works later)
## Example Revisited

<table>
<thead>
<tr>
<th></th>
<th>Stain 1</th>
<th>Stain 2</th>
<th>Stain 3</th>
<th>$A_i$</th>
<th>$B_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detergent 1</td>
<td>45</td>
<td>43</td>
<td>51</td>
<td>139</td>
<td>182</td>
</tr>
<tr>
<td>Detergent 2</td>
<td>47</td>
<td>46</td>
<td>52</td>
<td>145</td>
<td>176</td>
</tr>
<tr>
<td>Detergent 3</td>
<td>48</td>
<td>50</td>
<td>55</td>
<td>153</td>
<td>207</td>
</tr>
<tr>
<td>Detergent 4</td>
<td>42</td>
<td>37</td>
<td>49</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$x_{..} = 565$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We have: $a = 4, b = 3, n = 12$

\[
SST = \sum \sum x_{ij}^2 - \frac{x_{..}^2}{12} = 26867 - \frac{565^2}{12} = 265
\]

\[
SSTr = \frac{(139^2 + 145^2 + 153^2 + 128^2)}{3} - \frac{565^2}{12} = 111
\]

\[
SSB = \frac{(182^2 + 175^2 + 207^2)}{4} - \frac{565^2}{12} = 135
\]

\[
SSE = 265 - 111 - 135 = 19
\]
SAS Output

• SAS code

```sas
proc glm data=detergent alpha=0.05;
class stain soap;
model clean = stain soap;
means soap / tukey lines;
run;
```
The GLM Procedure

Dependent Variable: clean

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>246.0833333</td>
<td>49.2166667</td>
<td>15.68</td>
<td>0.0022</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>18.8333333</td>
<td>3.1388889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>264.9166667</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square      | Coeff Var    | Root MSE     | clean Mean |
--------------|--------------|--------------|------------|
0.928908      | 3.762883     | 1.771691     | 47.083333  |

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>stain</td>
<td>2</td>
<td>135.1666667</td>
<td>67.5833333</td>
<td>21.53</td>
<td>0.0018</td>
</tr>
<tr>
<td>soap</td>
<td>3</td>
<td>110.9166667</td>
<td>36.9722222</td>
<td>11.78</td>
<td>0.0063</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>stain</td>
<td>2</td>
<td>135.1666667</td>
<td>67.5833333</td>
<td>21.53</td>
<td>0.0018</td>
</tr>
<tr>
<td>soap</td>
<td>3</td>
<td>110.9166667</td>
<td>36.9722222</td>
<td>11.78</td>
<td>0.0063</td>
</tr>
</tbody>
</table>
Tukey's Studentized Range (HSD) Test for clean

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha 0.05
Error Degrees of Freedom 6
Error Mean Square 3.138889
Critical Value of Studentized Range 4.89559
Minimum Significant Difference 5.0076

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Tukey Grouping</th>
<th>Mean</th>
<th>N</th>
<th>soap</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>51.000</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>48.333</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>46.333</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>B A</td>
<td>46.333</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Which groups are significant, different according to Tukey’s result?

3 and 4
2 and 4

Notice: We are running Tukey’s comparison because of the significant result in the previous ANOVA table!
After Class...

- Review Section 9.4 in the text
- Read Section 10.2 for Friday’s Class
- Look at Hw#9, after Wed’s exam
- Lab 5: Next Wednesday