Randomized Complete Block Design

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

March 22, 2007

The Big Picture

A **blocking variable** is a categorical variable that is not the primary variable of interest where observations within each level ought to be homogeneous except for treatment.

In a **randomized complete block design**, each treatment is applied to individuals selected at random within each block.

In a **randomized incomplete block design**, treatments are assigned at random within blocks, but every treatment may not be represented in every block.
# Seed Germination Experiment

- In an experiment, four sites were selected where the soil and climate conditions were expected to be very similar within the site.
- Here we will treat each site as a block.
- Within each block, five plots were identified.
- The treatment was applying a seed disinfectant to seeds. There were four different treatments (brands) plus a control.
- The researchers planted 100 seeds from a single treatment in each plot.
- The response is the number of seeds that germinated.
- We will also analyze this data using a generalized linear model later in the semester. The response variable is better modeled as binomial than as normal.

## Data

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>86</td>
<td>90</td>
<td>88</td>
<td>87</td>
</tr>
<tr>
<td>Arasan</td>
<td>98</td>
<td>94</td>
<td>93</td>
<td>89</td>
</tr>
<tr>
<td>Spergon</td>
<td>96</td>
<td>90</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>Semesan</td>
<td>97</td>
<td>95</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>Fermate</td>
<td>91</td>
<td>93</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

*Lines show treatment*

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>86</td>
<td>90</td>
<td>88</td>
<td>87</td>
</tr>
<tr>
<td>Arasan</td>
<td>98</td>
<td>94</td>
<td>93</td>
<td>89</td>
</tr>
<tr>
<td>Spergon</td>
<td>96</td>
<td>90</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>Semesan</td>
<td>97</td>
<td>95</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>Fermate</td>
<td>91</td>
<td>93</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

*Lines show blocking*
Crossed Variables

- A categorical variable $A$ is *crossed* with a categorical variable $B$ if there is one or more observations for each combination of levels of $A$ and $B$.
- Here, treatment and blocking are *crossed*.
- Both treatment and blocking are *nested* within the grand mean.
- Individual error is *nested* within each other variable.
- Block could be random or fixed.

Model

- A model is
  
  $$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

  where:
  - $\mu$ is an overall mean,
  - $\alpha_i$ is the effect of treatment $i$ where $\sum_i \alpha_i = 0$, for $i = 1, \ldots, 5$,
  - $\beta_j$ is the effect in block $j$ where $\sum_j \beta_i = 0$, for $j = 1, \ldots, 4$,
  - and $e_{ij} \sim \text{iid } N(0, \sigma_e^2)$ is individual random error.

  Note that $\mu_{ij} = \mu + \alpha_i + \beta_j$.

  Instead of the sum constraints, we could instead assume $\alpha_1 = \beta_1 = 0$ (the default in R).
Seed Germination Experiment

Data

```r
> seed = read.table("seed.txt", header = T)
> attach(seed)
> str(seed)
'data.frame': 20 obs. of 3 variables:
$ count : int 86 90 88 87 98 94 93 89 96 90 ...
$ treatment: Factor w/ 5 levels "AControl","Arasan",..: 1 1 1 1 2 2 2 2 5 5 ...
$ block : Factor w/ 4 levels "b1","b2","b3",..: 1 2 3 4 1 2 3 4 1 2 ...
> seed
count treatment block
1 86 AControl b1
2 90 AControl b2
3 88 AControl b3
4 87 AControl b4
5 98 Arasan b1
6 94 Arasan b2
7 93 Arasan b3
8 89 Arasan b4
9 96 Spergon b1
10 90 Spergon b2
11 91 Spergon b3
12 92 Spergon b4
13 97 Semesan b1
14 95 Semesan b2
15 91 Semesan b3
16 92 Semesan b4
17 91 Fermate b1
18 93 Fermate b2
19 95 Fermate b3
20 95 Fermate b4
```

Plot of Data

```r
> library(lattice)
> fig1 = xyplot(count ~ treatment | block, pch = 16)
> print(fig1)

- The block effect seems pretty weak.
```
Seed Germination Experiment

Alternative Plot of Data

- It looks like each treatment improves germination rate over the control.
- The spread is similar within treatments.

Linear Model Analysis

```
> library(lattice)
> fig1 = xyplot(count ~ treatment, pch = 16)
> print(fig1)

> seed.lm = lm(count ~ treatment + block)
> anova(seed.lm)
```

```
Analysis of Variance Table

Df Sum Sq Mean Sq F value Pr(>F)
---
treatment 4 102.300 25.575 3.5979 0.03775 *
block 3 18.950 6.317 0.8886 0.47480
Residuals 12 85.300 7.108
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(seed.lm)

Call:
  lm(formula = count ~ treatment + block)

Residuals:
    Min     1Q Median     3Q    Max
-3.950 -1.113  0.325  1.850  3.050

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              89.200      1.686 52.899 1.37e-15 ***
treatmentArasan         5.750       1.885  3.050  0.01008 *
treatmentFermate        5.750       1.885  3.050  0.01008 *
treatmentSemesan       6.000       1.885  3.183  0.00788 **
treatmentSpergon       4.500       1.885  2.387  0.03433 *
blockb2                -1.200       1.686  -0.712  0.49028
blockb3                -2.000       1.686  -1.186  0.25854
blockb4                -2.600       1.686  -1.542  0.14904
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.666 on 12 degrees of freedom
Multiple R-Squared: 0.587, Adjusted R-squared: 0.3461
F-statistic: 2.437 on 7 and 12 DF,  p-value: 0.08395
```
Residual Plot

> plot(fitted(seed.lm), residuals(seed.lm))
> abline(h = 0)

- Residual plot looks okay.
- Normal approximation to binomial is still not too bad as there are sufficient numbers of successes and failures in each treatment group.

Random Effects Model

- There are very few differences in a random effects model (treating block as random).
- The variance of the block variable is essentially 0.
- It is more difficult to carry out hypothesis tests.
Comparison

```r
> library(lme4)
> seed.lmer = lmer(count ~ treatment + (1 | block))
> summary(seed.lmer)

Linear mixed-effects model fit by REML
Formula: count ~ treatment + (1 | block)
    AIC BIC logLik MLdeviance REMLdeviance
90.58 96.56 -39.29  89.78  78.58
Random effects:
  Groups   Name        Variance Std.Dev.
   block   (Intercept) 3.475e-09 5.8949e-05
   Residual 6.950e+00 2.6363e+00
number of obs: 20, groups: block, 4
Fixed effects:
                  Estimate Std. Error t value
(Intercept)    87.75000   1.31800  66.570
  treatmentArsan  5.75000   1.86400   3.080
  treatmentFermate  5.75000   1.86400   3.080
  treatmentSemesan  6.00000   1.86400   3.220
  treatmentSpergon  4.50000   1.86400   2.410

Correlation of Fixed Effects:
                        (Intr) trtmntArsn trtmntFtrtmntSm
  treatmntArs  -0.707
  treatmntFrm -0.707  0.500
  treatmntSmsn-0.707  0.500  0.500
  trtmntSprgn-0.707  0.500  0.500  0.500

> summary(seed.lm)

Call:
  lm(formula = count ~ treatment + block)
Residuals:
    Min     1Q   Median     3Q    Max
-3.950 -1.113   0.325  1.850  3.050
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    89.20000   1.68600  52.899 1.37e-15 ***
treatmentArsan  5.75000   1.88500   3.050  0.01008 *
treatmentFermate  5.75000   1.88500   3.050  0.01008 *
treatmentSemesan  6.00000   1.88500   3.183  0.00788 **
treatmentSpergon  4.50000   1.88500   2.387  0.03433 *
blockb2 -1.20000   1.68600  -0.712  0.49028
blockb3 -2.00000   1.68600  -1.186  0.25854
blockb4 -2.60000   1.68600  -1.542  0.14904

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.666 on 12 degrees of freedom
Multiple R-Squared: 0.587, Adjusted R-squared: 0.3461
F-statistic: 2.437 on 7 and 12 DF,  p-value: 0.08395
```