# Restricted Randomization-Block Designs

Tim Robinson

# Randomized Block Designs

- Experimental error makes inference difficult because as $\sigma^2_{error}$ increases, confidence intervals are less precise and power to detect treatment differences declines.

- Recall that experimental error variance refers to the amount of unexplained variation from one experimental unit to the next.

- We would prefer to conduct our experiments with experimental units that are as homogenous as possible.

- A block of units is a set of units that are homogenous in some sense.

    - field plots, stream habitats, samples analyzed at about the same time, tissue samples that are analyzed at the same lab, etc.

- Units within the same block may have similar responses…in the two sample setting, in your intro stats course you learned of the paired t-test…in the paired t-test, there are two experimental units within each block.

# Comparing Two Population Means (Paired Data)

- The Wyoming Department of Environmental Quality sends sage grouse tissue specimens to two labs (lab A and lab B) for contaminant analysis. The head of the department is concerned about the high readings she is consistently receiving from lab A. To verify her suspicions, each of 10 specimens recently collected was sent to both labs for separate estimates of a certain type of contaminant in the specimen.

| Specimen | Lab A | Lab B |
|---|---|---|
| 1 | 30.1 | 30.0 |
| 2 | 45.2 | 43.3 |
| 3 | 26.7 | 26.4 |
| 4 | 51.8 | 50.9 |
| 5 | 32.6 | 31.5 |
| 6 | 23.4 | 23.2 |
| 7 | 25.3 | 25.1 |
| 8 | 48.2 | 48 |
| 9 | 45.5 | 45.8 |
| 10 | 36.9 | 36.6 |

- Blocking designs are not completely randomized designs.
- There are many, many blocking designs and we will only cover *complete* block designs…meaning that every treatment is used in every block. *Incomplete* block designs are extremely useful.
- In a Randomized Complete Block Design, there are $g$ treatments, and each treatment is assigned to $b$ experimental units for a total of $N=gb$ experimental units. The $N$ units are partitioned into $b$ blocks of $g$ units. We anticipate that the units within a given block will have similar responses. Within each block, we randomly assign the $g$ treatments to the $g$ units.
- It is important to keep in mind that blocks exist at the time of the randomization of treatments to units…we cannot impose a blocking structure on a completely randomized design…either the randomization was blocked or it was not.

**Structure of Paired Data Example:**

| Specimen | LabA | LabB | contam | Lab |
|---|---|---|---|---|
| 1 | 30.1 | 30 | 30.1 | A |
| 2 | 45.2 | 43.3 | 45.2 | A |
| 3 | 26.7 | 26.4 | 26.7 | A |
| 4 | 51.8 | 50.9 | 51.8 | A |
| 5 | 32.6 | 31.5 | 32.6 | A |
| 6 | 23.4 | 23.2 | 23.4 | A |
| 7 | 25.3 | 25.1 | 25.3 | A |
| 8 | 48.2 | 48 | 48.2 | A |
| 9 | 45.5 | 45.8 | 45.5 | A |
| 10 | 36.9 | 36.6 | 36.9 | A |
| 1 | | | 30 | B |
| 2 | | | 43.3 | B |
| 3 | | | 26.4 | B |
| 4 | | | 50.9 | B |
| 5 | | | 31.5 | B |
| 6 | | | 23.2 | B |
| 7 | | | 25.1 | B |
| 8 | | | 48 | B |
| 9 | | | 45.8 | B |
| 10 | | | 36.6 | B |

## Equal-variance t-test = F-test for comparing two groups

**R Code for Equal Variance t-:**

```
contaminant <- read.csv("WYDEQ.csv",header=TRUE)\
names(contaminant)

t.test(contam~Lab, alternative='greater', conf.level=.95, var.equal=TRUE,
       data=contaminant)
```

**R Output:**

```
Two Sample t-test

data:  contam by Lab

t = 0.106, df = 18, p-value = 0.4584

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:
 -7.527087       Inf
sample estimates:
mean in group A mean in group B
         36.57          36.08
```

**R Code for Completely Randomized Design:**

```
m1 <- lm(contam~Lab,data=contaminant)
anova(m1)
```

**R Output:**

Analysis of Variance Table

Response: contam

| | Df | SSq | MSq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Lab | 1 | 1.2 | 1.201 | 0.0112 | 0.9168 |
| Residuals | 18 | 1923.7 | 106.874 | | |

If you were to calculate $s_p^2$ for the two-sample t-test, you would get the value for MSE above

The p-value in the two-sample t-test was a one-sided p-value and if you multiply the p-value by 2, you would get the F p-value above

# Paired t-test/Randomized block analysis of sage grouse data

**R Code:**

```
t.test(contaminant$LabA, contaminant$LabB, alternative='greater',
  conf.level=.95, paired=TRUE)
```

**R Output:**

```
Paired t-test

data:  contaminant$LabA and contaminant$LabB

t = 2.4402, df = 9, p-value = 0.01868

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:
 0.1219037       Inf
sample estimates:
mean of the differences
                  0.49
```

**R Code for Randomized Complete Block Design:**

```
contaminant$Specimen <- as.factor(contaminant$Specimen)
m2 <- lm(contam~Specimen+Lab,data=contaminant)
anova(m2)
```

**R Output:**

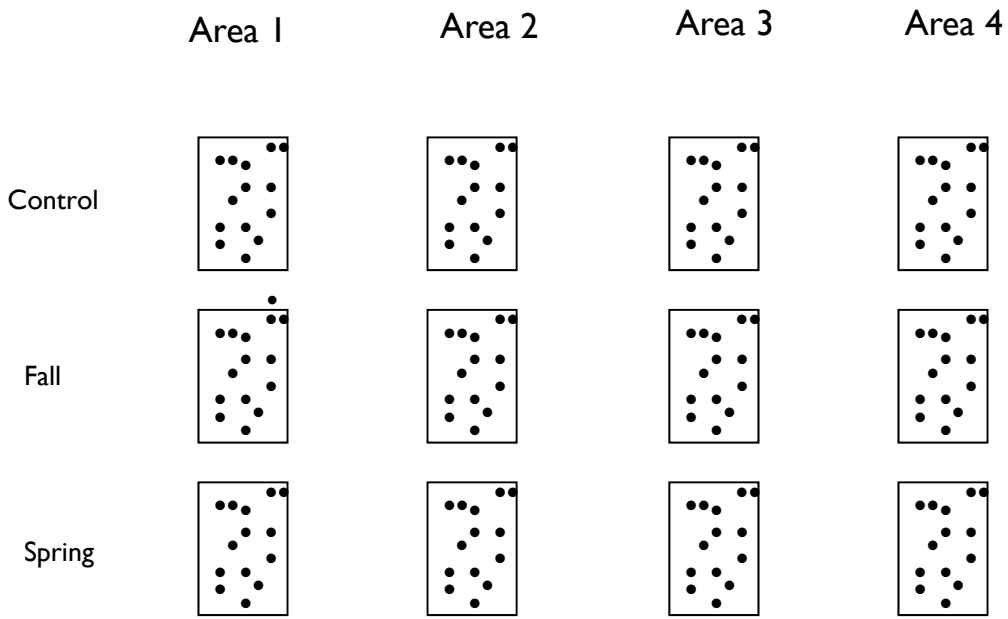
```
Analysis of Variance Table

Response: contam
             Df     SSq          MSq          F value
      Pr(>F)
Specimen     9      1921.92      213.547      1059.2023
      1.605e-12 ***
Lab          1      1.20         1.201        5.9545
      0.03735 *
Residuals    9      1.81         0.202
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value in the paired t-test was a one-sided p-value and if you multiply the p-value by 2, you would get the F p-value above

# Example – Prescribed Burn

A study is conducted to determine the effect of prescribed fire in sagebrush-bitterbrush on the availability of foods to sage grouse chicks. The treatments are a control (no burn), fall burn, and spring burn. Four study areas were selected and treated as blocks do to the varying degree of sagebrush-bitterbrush cover within each area. Within each block, 50x50m treatment plots were established and randomly assigned to Control, fall burn, or spring burn treatments. The response was the relative abundance of insects. Relative abundance was estimated using 15 pitfall traps within each treatment plot.



| Blocks (Area) | Treatment (burn times) | | |
|---|---|---|---|
| | Control | Fall | Spring |
| 1 | 10.17 | 8.54 | 8.26 |
| 2 | 9.29 | 6.9 | 7.2 |
| 3 | 9.4 | 8.42 | 6.05 |
| 4 | 10.81 | 7.98 | 8.04 |

$H_o$: The mean relative abundance of insects is the same for each of the treatment types

$H_1$: The mean relative abundance of insects is not the same for each of the treatment types

Letting $y_{ij}$ denote the relative abundance for the $i^{th}$ treatment in the $j^{th}$ block, the RCB model is given by

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

The ANOVA table for the *unreplicated* RCB is

| Source | Df | SS | MS | F |
|---|---|---|---|---|
| Blocks | b-1 | SS Blocks | MS Blocks | |
| Treat | g-1 | SS Treat | MS Treat | $MS_{Trt}$ / $MS_{Err}$ |
| Error | (b-1)(g-1) | SS Error | MS Error | |
| Total | rb -1 | SS Total | | |

**There is no test for blocks**

The block*treatment interaction serves as the error term when there is no replication within blocks.

# Partitioning the variability in the RCBD:

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

$$+(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})]^2$$

$$=b\sum_{i=1}^{g}(\bar{y}_{i.}-\bar{y}_{..})^2+g\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2$$

$$+\sum_{i=1}^{g}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$$

$$SS_T=SS_{Treatments}+SS_{Blocks}+SS_E$$

Note the expression for the $SS_{error}$ is the same as that for the SS for the two factor interaction in the 2-way CRD



# ANOVA Display for the RCBD

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

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ |
|---|---|---|---|---|
| Treatments | $SS_{Treatments}$ | $a-1$ | $\frac{SS_{Treatments}}{a-1}$ | $\frac{MS_{Treatments}}{MS_E}$ |
| Blocks | $SS_{Blocks}$ | $b-1$ | $\frac{SS_{Blocks}}{b-1}$ | |
| Error | $SS_E$ | $(a-1)(b-1)$ | $\frac{SS_E}{(a-1)(b-1)}$ | |
| Total | $SS_T$ | $N-1$ | | |

Montgomery uses '*a*' to denote the number of levels for the treatment but we are using '*g*' to denote the number of treatment levels.

**R Code for Randomized Complete Block Design:**

```
mburn <- lm(relabund~Blocks+Treatment,data=burndata)
anova(mburn)
```

**R Output:**

```
> anova(mburn)
Analysis of Variance Table

Response: relabund
           Df  SSq     MSq     F value  Pr(>F)
Blocks     1   0.0001  0.0001  0.0001   0.993198
Treatment  2   14.0806 7.0403  9.0771   0.008754 **
Residuals  8   6.2049  0.7756
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



# Completely Randomized Analysis

**R Code for Completely Randomized Design:**

```
mburncrd <- lm(relabund~Treatment,data=burndata)
anova(mburncrd)
```

**R Output:**

```
> anova(mburncrd)
Analysis of Variance Table

Response: relabund
           Df  SSq     MSq     F value  Pr(>F)
Treatment  2   14.0806 7.0403  10.212   0.004842 **
Residuals  9   6.2049  0.6894
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Relative efficiency

- How well did the blocking work? Should we use blocking in our next similar experiment? For instance in the DEQ example, our design was to send part of each sage grouse wing to each of the two labs. A CRD would have sent 10 independent specimens to lab A and 10 different, independent (i.e. different birds) specimens to lab B

- Which is the better design? How do we measure this?

- DON'T judge which design is better by testing for blocks

- Relative efficiency answers the following question: using the same population of units (ex. sage grouse), by what factor would we need to increase our sample size in the CRD to get the same power as what we have in the RCBD?

# Relative efficiency continued

- This is mostly an issue of how the error variance changes, but there is also a minor effect due to the fact that fitting blocks uses up degrees of freedom for error.
- In CRD, bigger estimate of $\sigma^2_{error}$ hurts, but larger $df_{error}$ helps; usually the first factor dominates, however.
  - We can see the difference in $df_{error}$ as 18 in the CRD and it is 9 for the RCB…implication can be seen when looking at critical values for F(0.95,df1=1, df2=18) vs. F(0.95,df=1,df2=9)
- We first estimate what error variance would have been if we had used a CRD instead of RCB, then we make a minor df adjustment.

# Relative efficiency continued

- We estimate $\sigma^2_{error}$ for the RCB using the MSE from the RCB

- We estimate $\sigma^2_{error}$ from using a CRD via a weighted average of the MSE from the RCB and the $MS_{Block}$ weighted by df.

$$\hat{\sigma}^2_{error,CRD} = \frac{df_{block} MS_{block} + \left(df_{TRT} + df_{error}\right) MS_{error}}{df_{block} + df_{TRT} + df_{error}}$$

# Relative efficiency continued

- The relative efficiency of the RCB to the CRD is given by

$$RE_{RCB:CRD} = \left[\frac{(b-1)(g-1)+1}{(b-1)(g-1)+3}\right]\left[\frac{g(b-1)+3}{g(b-1)+1}\right]\frac{\hat{\sigma}^2_{error,CRD}}{\hat{\sigma}^2_{error,RCB}}$$

- For interpretation, let's suppose the RE above is 1.7, this means that the CRD needs 70% more experimental units than the RCB to have the same power as the CRD

# Relative efficiency for DEQ Lab Data

- We have

$$\hat{\sigma}^2_{error,RCB} = MSE = 0.202$$

- We estimate $\sigma^2_{error}$ from using a CRD via a weighted average of the MSE from the RCB and the $MS_{Block}$ weighted by df.

$$\hat{\sigma}^2_{error,CRD} = \frac{9*213.547 + (1+9)*0.202}{9+1+9} = \frac{1923.943}{19} = 101.26$$

# Relative Efficiencey DEQ Lab Data

- The relative efficiency of the RCB to the CRD is given by

$$RE_{RCB:CRD} = \left[\frac{(10-1)(2-1)+1}{(10-1)(2-1)+3}\right]\left[\frac{2(10-1)+3}{2(10-1)+1}\right]\frac{101.26}{0.202}$$

$$= \frac{10}{12} * \frac{21}{19} * 501.28$$

$$= 461.71$$

- Recall there were 10 experimental units in the RCB (i.e. 10 sage grouse). This RE values implies that you would need upwards of 462*10 = 4,620 sage grouse in the CRD to have as much power to detect a lab difference as the RCB has