Generalizations

Part 3: Analysis of the RCB Design

STAT 705: Regression and Analysis of Variance

Additive Block Effects

- In the analysis of the randomized complete block design (RCB) we will assume that the effect of the blocking factor is additive.
- In other words, we assume that, in going from one block to another the responses (on average) will either increase the same amount or decrease the same amount as a result of the changing block conditions, regardless of the effects of the treatments.
- Another way to say this is that there is no interaction between the blocks and the treatments.

Agronomy Study, Re-visited

- Recall the agronomy example from the previous lesson
- Suppose that, as we move from west to east, the soil moisture of the field becomes more favorable for growing wheat. If this causes the yields to increase the same amount (on average) regardless of the effects of the treatments, then the blocking factor is additive.

West → East

1 Trt = 2	3 Trt = 1	5 Trt = 3	7 Trt = 3	9 Trt = 4	11 Trt = 1
2 Trt = 3	4 Trt = 4	6 Trt = 1	8 Trt = 2	10 Trt = 3	12 Trt = 2

The Model

- The mathematical model for an RCB design can be expressed symbolically as

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

where

Y_{ij} is the response for the i^{th} EU in the j^{th} block

μ is the overall mean response (over all EUs and all blocks)

α_i is the effect of the i^{th} treatment

β_j is the effect of the j^{th} block

ε_{ij} is the random error for the i^{th} EU in the j^{th} block

The Model

- In words, the mathematical model for an RCB design can be expressed as:

  The response is . . .

  an overall effect . . .

  plus an effect due to treatment . . .

  plus an effect due to block . . .

  plus some random error

- This is the additive model for ANOVA where one of the factors is Treatment and the other factor is Block.

Generic SAS code for RCB

- Suppose that "trt" denotes the treatment variable and "blk" denotes the blocking variable.
- The Proc GLM statements for the RCB are the same as that for two-way ANOVA, except that the interaction terms (trt*blk) are not included.

```
proc glm;
 class trt blk;
 model response = trt blk / ss3;
 lsmeans trt blk / stderr pdiff;
 run;
```

Degrees of Freedom for RCB

- Suppose that there are "t" treatments and "b" blocks.
- Because treatments appear once in each block, the total number of observations in an RCB is tb.
  - Total df = tb – 1
  - df treatments = # treatments – 1 = t – 1
  - df blocks = # blocks – 1 = b – 1
  - df error = Total df – df treatments – df blocks

    = (tb – 1) – (t – 1) – (b – 1)

    = (t – 1)(b – 1)

The MSE in an RCB

- For every ANOVA, we must have a measure of experimental error, that is, we must have an MSE.
- Because we assume that the RCB has only one observation for each block by treatment combination, we cannot use the observations within blocks to compute MSE.
- However, because there is no interaction between blocks and treatments, the MS for block by treatment interaction is affected only by random error. Thus we may use the block by treatment mean square as the MSE in an RDB design.

Example

- Suppose that the observations in our agronomy example turned out to be as shown below.

West | East

1 Trt = 2 39.4	3 Trt = 1 40.5	5 Trt = 3 43.0	7 Trt = 3 42.0	9 Trt = 4 46.1	11 Trt = 1 48.3
2 Trt = 3 38.3	4 Trt = 4 38.1	6 Trt = 1 45.4	8 Trt = 2 44.1	10 Trt = 3 46.2	12 Trt = 2 47.0

Example: Data

- Here is the same data in a table
- Notice how much the responses vary in going from west to east for each treatment.
  - Treatment 1: 48.3 - 40.5 = 7.8
  - Treatment 2: 47.0 - 39.4 = 7.6
  - Treatment 3: 46.2 - 38.3 = 7.9
  - Treatment 4: 46.1 - 38.1 = 8.0

Blocks

	West	**Middle**	**East**
Trt 1	40.5	45.4	48.3
Trt 2	39.4	44.1	47.0
Trt 3	38.3	43.0	46.2
Trt 4	38.1	42.0	46.1

- The difference is about the same for each treatment, so an additive model (i.e. no block by treatment interaction) seems appropriate

Example: Ignore Blocking

- Suppose we ignore the blocking factor and do a one-way ANOVA on the data.
- Because of the large variability of the observations within treatments, it appears that there is not a significant difference between the treatments (p = 0.8442)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	12.65	4.22	0.27	0.8442
Error	8	124.09	15.51		
Corrected Total	11	136.74			

Example: Incorporate Blocking

- When we account for the blocking factor and do an analysis of the blocked design, then we see that treatment is highly significant (p = 0.0002)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Treatment	3	12.65	4.22	46.13	0.0002
Block	2	123.55	61.77	675.93	<.0001
Error	6	0.55	0.091		
Corrected Total	11	136.74			

Example: Compare Results

- Why is there such a big difference in the results?
- When blocks are ignored, the SS for block goes into SS Error.
  - This makes SS Error larger
  - SS Error goes into the denominator of the F statistic
  - This makes the F statistic smaller
  - F statistics are significant when they are larger (not smaller)
  - Larger SS Error makes it more difficult to find significant differences in the treatments
- For the agronomy example, SS Block is 123.55, which is almost all of the SS Error (124.09) in the one-way analysis
- The RCB analysis extracts this from SS Error, leaving us with the “correct” SS Error of 0.54 and a “large” F statistic of 46.13.

Factorial Treatments for RCB

- Suppose there are two factors A and B that comprise the factorial treatment combinations.

  For example, the four fertilizer treatments in our agronomy example might consist of combinations of Nitrogen (0, 10) and Phosphorus (0, 5).

- We may assume that the factors A and B interact with each other, so that an A*B term may be included in the model.
- However, there would be no interactions between the block and A or the blocks and B.

Generic SAS Code for RCB with Factorial Treatments

- Suppose the treatments consists of the factorial combinations of A and B, and the blocking variable is denoted 'blk'.
- The SAS GLM statements would look like this

```
proc glm;
 class A B blk;
 model response = A B A*B blk / ss3;
 lsmeans A B A*B blk / stderr pdiff;
 run;
```

- This could also be written

```
proc glm;
 class A B blk;
 model response = A|B blk / ss3;
 lsmeans A|B blk / stderr pdiff;
 run;
```

KANSAS STATE UNIVERSITY

Other Cases of Blocking

- Incomplete Blocks

  It is possible to have fewer experimental units per block than there are treatments. For instance, there may be 4 treatments but only 3 experimental units per block. I recommend consulting your friendly statistician if you plan to use incomplete blocks in a study.

- Multiple Experimental Units per Block

  There may be more experimental units than there are treatments per block. For instance, there may be 4 treatments and 8 experimental units per block. Generally speaking, you can use the same GLM statements that you would use with an RCB.

What You Should Know

- Know the reasons for blocking
- Understand how the randomization (assigning experimental units to treatments) is fundamentally different for CRD vs. RCB designs
- Understand how blocks and treatments are incorporated into the linear model and be able to
  - identify the degrees of freedom for blocks, treatments and error
  - write the SAS code
  - interpret the SAS output
  - write a report of the results