Lecture 29

Randomized Complete Block Design (RCBD)

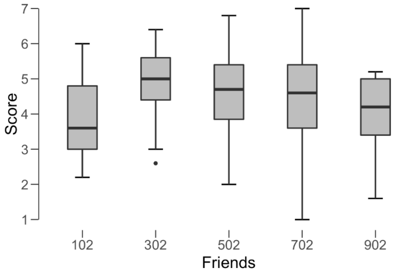
STAT 8010 Statistical Methods I
November 4, 2019

Whitney Huang
Clemson University

Facebook Friends Example Revisited

Participants examined one of five stimuli, each containing a Facebook profile mockup. Elements of these stimuli (e.g. photographs, wall posts, etc.) remained constant over the five versions, with the exception of the number of friends which appeared on the profile as 102, 302, 502, 702, or 902....

Tong et al., Journal of Computer-Mediated Communication, 2008



This is an example of **Completely Randomized Design**

Completely Randomized Design (CRD)

A CRD has

- N experimental units
- J different treatments
- J known treatment group sizes $n_1, n_2, \cdots, n_J$ with $\sum_{j=1}^{J} n_j = N$
- **Completely random assignment** of treatments to units

A key assumption of CRD is that all experimental units are (relatively) homogeneous
Question: What if this assumption is violated?

Randomized Complete Block Design (RCBD)

- The population of experimental units is divided into a number of relatively homogeneous sub-populations (blocks), and it is assumed that all experimental units within a given block are homogeneous

Randomized Complete Block Design (RCBD)

- The population of experimental units is divided into a number of relatively homogeneous sub-populations (blocks), and it is assumed that all experimental units within a given block are homogeneous

Randomized Complete Block Design (RCBD)

- The population of experimental units is divided into a number of relatively homogeneous sub-populations (blocks), and it is assumed that all experimental units within a given block are homogeneous
- Within each block, treatments are randomly assigned to experimental units such that each treatment occurs equally often (usually once) in each block

Randomized Complete Block Design (RCBD)

- The population of experimental units is divided into a number of relatively homogeneous sub-populations (blocks), and it is assumed that all experimental units within a given block are homogeneous
- Within each block, treatments are randomly assigned to experimental units such that each treatment occurs equally often (usually once) in each block

Randomized Complete Block Design (RCBD)

- The population of experimental units is divided into a number of relatively homogeneous sub-populations (blocks), and it is assumed that all experimental units within a given block are homogeneous
- Within each block, treatments are randomly assigned to experimental units such that each treatment occurs equally often (usually once) in each block
- A key assumption in the analysis is that the effect of each level of the treatment is the same for each level of the blocking factor.

RCBD Notation

- J is the number of treatments; B is the number of blocks
- X_{ij} is the measurement on the unit in block i that received treatment j
- $N = B \times J$ is the total number of experimental units
- $\bar{X}_{.j} = \sum_{i=1}^{B} \frac{X_{ij}}{B}$ is the average of all measurements for units receiving treatment j
- $X_{i.} = \sum_{j=1}^{J} \frac{X_{ij}}{J}$ is the average of all measurements for units in the i_{th} block
- $\bar{X} = \sum_{i=1}^{B} \sum_{j=1}^{J} \frac{X_{ij}}{N}$ is the average of all measurements

CLEMSON UNIVERSITY

RCBD Model and Assumptions

- The model for an RCBD is:

$$X_{ij} = \underbrace{\mu + \tau_j}_{\mu_j} + \beta_i + \varepsilon_{ij}, \quad i = 1, \cdots, B, \quad j = 1, \cdots, J$$

where μ is the overall mean, τ_j is the effect of treatment j, β_i is the effect of block i, and $\varepsilon_{ij} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$ are random errors

- The effect of each level of the treatment is the same across blocks $\Rightarrow$ no interaction between $\tau' s$ and $\beta' s$

RCBD Sums of Squares

- **Total sum of square**:

$$SS_{tot} = \sum_{i=1}^{B} \sum_{j=1}^{J} (X_{ij} - \bar{X})^2$$

- **Treatment sum of square**:

$$SS_{trt} = \sum_{j=1}^{J} B(\bar{X}_{.j} - \bar{X})^2$$

- **Block sum of square**:

$$SS_{blk} = \sum_{i=1}^{B} J(\bar{X}_{i.-\bar{X}})^2$$

- **Error sum of square**:

$$SS_{err} = \sum_{i=1}^{B} \sum_{j=1}^{J} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})^2$$

ANOVA Table and F Test

Source	df	SS	MS	F statistic
Treatment	$J-1$	SS_{trt}	$MS_{trt} = \frac{SS_{trt}}{J-1}$	$F_{trt} = \frac{MS_{trt}}{MS_{err}}$
Block	$B-1$	SS_{blk}	$MS_{blk} = \frac{SS_{blk}}{B-1}$	$F_{blk} = \frac{MS_{blk}}{MS_{err}}$
Error	$(B-1)(J-1)$	SS_{err}	$MS_{err} = \frac{SS_{err}}{(B-1)(J-1)}$	
Total	$N-1$	$SStot$		

There are two hypothesis tests in an RCBD:

- H_0 : The means of all treatments are equal
  H_a : At least one of the treatments has a different mean
  Test Statistic: $F_{trt} = \frac{MS_{trt}}{MS_{err}}$. Under H_0, $F_{trt} \sim F_{df_1=J-1, df_2=df_{err}}$

- H_0 : The means of all blocks are equal
  H_a : At least one of the blocks has a different mean
  Test Statistic: $F_{blk} = \frac{MS_{blk}}{MS_{err}}$. Under H_0, $F_{blk} \sim F_{df_1=B-1, df_2=df_{err}}$

Example

Suppose you are manufacturing concrete cylinders for, say, bridge supports. There are three ways of drying green concrete (say A, B, and C), and you want to find the one that gives you the best compressive strength.The concrete is mixed in batches that are large enough to produce exactly three cylinders, and your production engineer believes that there is substantial variation in the quality of the concrete from batch to batch.

You have data from $B = 5$ batches on each of the $J = 3$ drying processes. Your measurements are the compressive strength of the cylinder in a destructive test. (So there is an economic incentive to learn as much as you can from a well-designed experiment.)

Example: Data Set

The data are:

	Batch					
Treatment	1	2	3	4	5	Trt Sum
A	52	47	44	51	42	236
B	60	55	49	52	43	259
C	56	48	45	44	38	231
Batch Mean	168	150	138	147	123	726

The primary null hypothesis is that all three drying techniques are equivalent, in terms of compressive strength.

The secondary null is that the batches are equivalent (but if they are, then we have wasted power by controlling for an effect that is small or non-existent).

Example: ANOVA Table


```
Analysis of Variance Table

Response: x
          Df Sum Sq Mean Sq F value
trt        2   89.2   44.60  7.6239
blk        4  363.6   90.90 15.5385
Residuals  8   46.8    5.85
             Pr(>F)
trt       0.0140226 *
blk       0.0007684 ***
```

Interpretation?

Example: CRD Analysis

Suppose we had not blocked for batch. Then the data would be:

Treatment		Trt Sum
A	52, 47, 44, 51, 42	236
B	60, 55, 49, 52, 43	259
C	56, 48, 45, 44, 38	231

This is the same as before except now we ignore which batch the observation came from.

Example: CRD ANOVA

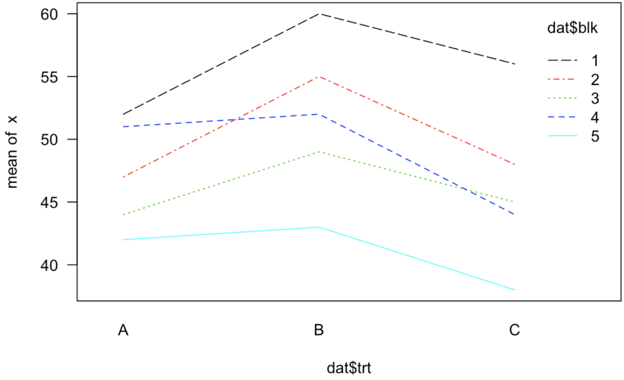
```
Analysis of Variance Table

Response: x
          Df Sum Sq Mean Sq F value Pr(>F)
trt        2   89.2    44.6  1.3041 0.3073
Residuals 12  410.4    34.2
```

We fail to reject the null $H_0 : \mu_A = \mu_B = \mu_C$ if we **ignore the block effect**

$\Rightarrow$ Using blocks gave us a more powerful test!

Assessing the Additivity Assumption: Interaction Plot



Parallel lines ⇒ No interaction occurs

Summary

In this lecture, we learned Randomized Complete Block Design (RCBD)

In next lecture we will learn statistical inference for proportions