Randomized Complete Block Designs (RCBD)

Defn: A Randomized Complete Block Design is a variant of the completely randomized design that we recently learned. In this design, blocks of experimental units are chosen where the units within are block are more similar to each other (homogeneous) than to units in other blocks. In a complete block design, there are at least *t* experimental units in each block.

Examples of blocks:

1) a litter of animals could be considered a block since they all have similar genetic structure, similar prenatal/parental care, etc.

2) a field or pasture that can be divided into quadrants since soil properties, environmental conditions, etc are similar within a field

3) a greenhouse with multiple benches since environmental conditions are usually more similar within a greenhouse than between greenhouses

4) a year in which the experiment is performed since environmental conditions are similar within a year

Example of a CRBD: A nutritionist is interested in comparing the effect of three diets on weight gain in piglets. In order to perform the experiment, the researcher chooses 10 litters, each with at least three healthy and similarly sized piglets that have just been weaned. In each litter, three piglets are selected and one treatment is randomly assigned to each piglet. Diets are labeled A, B or C.

Litter	**Piglet**		
	1	**2**	**3**
1	A	C	B
2	B	C	A
…			
10	C	B	A

In a design without blocking, the researcher would pick 30 piglets from different litters and randomly assign treatments to them. This is known as unrestricted randomization. Blocking designs have restricted randomization since the treatments are randomly assigned WITHIN each block.

An RCBD has two factors: the factor of interest that includes the treatments to be studied and the "Blocking Factor" that identifies the blocks used in the experiment.

There are several forms of Blocking Designs:

1) the RCBD that we will study

2) incomplete block designs in which not every block has *t* experimental units

3) block designs in which the blocks have more than *t* experimental units that are used in the experiment

4) Latin square designs which have very specific forms of randomization of treatments within blocks (example is usually relates to time ordering of treatments)

Assumptions of the RCBD:

1) *Sampling*:
   a. The blocks are independently sampled

   b. The treatments are *randomly* assigned to the experimental units within a block.

2) *Homogeneous Variance*: the treatments all have the same variability, i.e. they all have the same variance

3) *Approximate Normality*: each population is normally distributed

Hypotheses

As we will see, the blocking factor is included in the study only as a way of explaining some of the variation in responses (Y) of the experimental units. As such, we are not interested in testing hypotheses about the blocking factor. Instead, just like in a one-way ANOVA, we restrict our attention to the other factor ("research" factor).

So, hypothesis testing proceeds similar to the techniques we learned for the one-way ANOVA. The two differences are the calculation of the error variance (MSE) and a calculation of the effect of the blocking factor (MSB).

Notation

t the number of treatments of interest in the "research" factor

b the number of blocks containing t experimental units

N $= t \times b$, the total sample size

y_{ij} observed value for the experimental unit in the j[th] block assigned to the i[th] treatment, $j = 1,2,\ldots,b$ and $i = 1,2,\ldots,t$

$\bar{y}_{i\bullet} = \dfrac{\sum_{j=1}^{b} y_{ij}}{b}$, the sample mean of the i[th] treatment

$\bar{y}_{\bullet j} = \dfrac{\sum_{i=1}^{t} y_{ij}}{t}$, the sample mean of the j[th] block

$\bar{y}_{\bullet\bullet} = \dfrac{\sum_{i=1}^{t}\sum_{j=1}^{b} y_{ij}}{tb}$, the overall sample mean of the combined treatments

Example: piglet diet experiment with three litters

	Diet			**Block**
Litter	**A**	**B**	**C**	**Mean**
1	$y_{A1} = 54.3$	$y_{B1} = 53.1$	$y_{C1} = 59.7$	$\bar{y}_{\bullet 1} = 55.7$
2	$y_{A2} = 53.6$	$y_{B2} = 52.4$	$y_{C2} = 59.7$	$\bar{y}_{\bullet 2} = 55.2$
3	$y_{A3} = 55.2$	$y_{B3} = 57.1$	$y_{C3} = 67.2$	$\bar{y}_{\bullet 3} = 62.2$
Treatment Mean	$\bar{y}_{A\bullet} = 54.4$	$\bar{y}_{B\bullet} = 55.2$	$\bar{y}_{C\bullet} = 59.8$	**Grand Mean** $\bar{y}_{\bullet\bullet} = 56.9$

Model:

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

where

- μ is the overall (grand) mean,
- α_i is the effect due to the i^{th} treatment,
- β_j is the effect due to the j^{th} block, and,
- ε_{ij} is the error term where the error terms, are independent observations from an approximately Normal distribution with mean = 0 and constant variance σ_ε^2

Total variability of all of the $Y_{ij,}$ is

$$TSS = \sum_i \sum_j (y_{ij} - \bar{y}_{\bullet\bullet})^2$$

which can be broken up into three parts: ***TSS = SST + SSB + SSE***

$SST = b\sum_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 = b\sum_i \hat{\alpha}_i^2$ is the "sum of squares treatments"

$SSB = t\sum_j (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2 = t\sum_j \hat{\beta}_j^2$ is the "sum of squares blocks"

$SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2 = \sum_i \sum_j \hat{\varepsilon}_{ij}^2$ is the "sum of squares error".

Like before, we are interested in the Mean Squares:

$$MST = \frac{SST}{t-1}, \text{ the Mean Square Treatments}$$

$$MSB = \frac{SSB}{b-1}, \text{ the Mean Square Blocks}$$

$$MSE = \frac{SSE}{(t-1)(b-1)}, \text{ the Mean Square Error}$$

Here $E(MST) = \sigma_\varepsilon^2 + b\sum_i \frac{\alpha_i{}^2}{(t-1)}$ and $E(MSE) = \sigma_\varepsilon^2$.

ANOVA Table for a Randomized Complete Block Design

Source	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F-stat*
Treatment	*SST*	*t – 1*	*MST*	*F*=MST/MSE*
Block	*SSB*	*b – 1*	*MSB*	
Error	*SSE*	*(t – 1)(b – 1)*	*MSE*	
Total	*TSS*	*tb – 1*		

Again, the test of a treatment effect

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_t$$
$$H_A: \text{at least one mean differs}$$

uses the statistic

$$F^*=MST/MSE.$$

If the null hypothesis is true then F* has an F-Distribution on numerator degrees of freedom t – 1 and denominator degrees of freedom (t – 1)(b – 1).

In addition to the similarity of the F-test of equality of treatment means, the tests and comparisons of treatment means are done exactly the same as before as well.

Example: piglet experiment.

```
data pigsblocked;
input litter diet$ gain @@;
datalines;
1 I   54.3   2 I   53.6
3 I   55.2   1 II  53.1
2 II  52.4   3 II  57.1
1 III 59.7   2 III 59.7
3 III 67.2
run;
proc glm data=pigsblocked;
 class diet litter;
 model gain = diet litter;
quit;
```

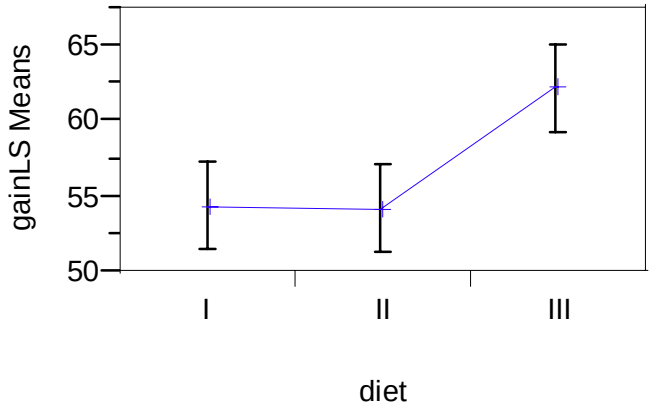
```
Dependent Variable: gain

                 Sum of
Source     DF   Squares   Mean Square  F Value  Pr > F
diet        2    125.39      62.69     19.02     0.0091
litter      2     38.46      19.23      5.83     0.0652
Error       4     13.18       3.30
CTotal      8    177.04
```

(Means comparisons using JMP v.5)

Least Squares Means Table

Level	Least Sq Mean	Std Error
I	54.366667	1.0481907
II	54.200000	1.0481907
III	62.200000	1.0481907



LSMeans Differences Tukey HSD

Alpha= 0.050 Q= 3.564

Mean[i]-Mean[j] Std Err Dif	I	II	III
I	0 0	0.16667 1.48237	-7.8333 1.48237
II	-0.1667 1.48237	0 0	-8 1.48237
III	7.83333 1.48237	8 1.48237	0 0

Level*			Least Sq Mean
III	A		62.200000
I		B	54.366667
II		B	54.200000

*Levels not connected by same letter are significantly different

Same experiment ignoring the litter effect:

```
proc glm data=pigsblocked;
 class diet litter;
 model gain = diet;
quit;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	125.39	62.69	7.28	0.0248
Error	6	51.65	8.61		
CTotal	8	177.04			

Least Squares Means Table

Level	Least Sq Mean	Std Error
I	54.366667	1.69389
II	54.200000	1.69389
III	62.200000	1.69389

LSMeans Differences Tukey HSD

Alpha= 0.050 Q= 3.06815

Mean[i]-Mean[j] Std Err Dif	I	II	III
I	0 0	0.16667 2.39552	-7.8333 2.39552
II	-0.16667 2.39552	0 0	-8 2.39552
III	7.83333 2.39552	8 2.39552	0 0

Level			Least Sq Mean
III	A		62.200000
I		B	54.366667
II		B	54.200000

Levels not connected by same letter are significantly different

Advantages of the RCBD as compared to the CRD:

1) reduce the error variance by "explaining" or identifying one source of some of the variability in the observations
    a. book refers to this as "filtering" out some of the variation
2) the design is easy to construct, i.e. when there are natural or obvious blocks with at least *t* experimental units, the restricted randomization is easy to achieve

Disadvantages

1) need homogeneous blocks in order for the blocking factor to be effective
2) the effect of the treatments in the Factor under study must be the same in every block, i.e. the effect of a treatment cannot depend on which block it is being applied to.

e.g. experiment to compare the unused red light time for five different traffic light signal sequences during morning rush hour. Traffic engineer chose several intersections and performed the different sequences at each intersection in random order. Suppose the effect of a particular sequence depends on which intersection you are studying, e.g. in intersections with heavy traffic, the average unused red light time is greater than the average time at intersections with lighter traffic maybe. This is known as interaction of factors.

Choosing Variables On Which To Block:

We want experimental units within each block to be as similar as possible to each other with respect to any characteristic which can effect or influence the response variable (Y). So, if a study relates to weight gain, we want each block to have similar characteristics with respect to growth such as starting weight, metabolic rates, etc.

Which is better, a RCBD or a CRD?

Can check using "Relative Efficiency" which compares the variance of the estimate of the i^{th} treatment mean

$$\hat{\mu}_i = \bar{y}_{i\bullet}$$

under the two different experiment designs. Efficiency is calculated as the number of observations that would be required if the experiment had been conducted as a CRD without any blocking.

$$\begin{aligned} RE(RCBD, CRD) &= \frac{MSE_{CRD}}{MSE_{RCBD}} \\ &= \frac{(SSB_{RCBD} + SSE_{RCBD})/t(b-1)}{MSE_{RCBD}} \\ &= \frac{(b-1)MSB_{RCBD} + b(t-1)MSE_{RCBD}}{(bt-1)MSE_{RCBD}} \end{aligned}$$

If the blocking was not helpful, then the relative efficiency equals 1. The larger the relative efficiency is, the more efficient the blocking was at reducing the error variance. The value can be interpreted as the ratio $\frac{r}{b}$ where *r* is the number of experimental units that would have to be assigned to each treatment if a CRD had been performed instead of a RCBD.

Example: in the piglet experiment, $SSB_{RCBD} = 38.46$, $SSE_{RCBD} = 13.18$, $t = 3$, $b = 3$, $MSE_{RCBD} = 3.30$

$$RE(RCBD, CRD) = \frac{(38.46 + 13.18)/3(2)}{3.30}$$
$$= \frac{8.61}{3.30} = 2.61$$

This implies that it would have taken more than 2.5 times as many experimental units/treatment to get the same MSE as we got using the litters as blocks. I.e. we would have needed approximately 8 (≈ 2.61*3) piglets per treatment in a CRD experiment testing the three diets.