# ANOVA Designs - Part I

Randomized Complete Block Design (RCB)
- Design
- Linear Model
- Computation
- Example
- NCSS

Latin Square Design (LS)
- Design
- Linear Model
- Computation
- Example
- NCSS

# Randomized Complete Block Design

An experimenter chooses an RCB design if they are interested in a set of treatments but need to *control for an extraneous source of variability*.

For example, an agricultural scientists wants to study the effect of 4 different fertilizers (A,B,C,D) on corn productivity. He has three fields (1,2,3) ranging in size from 4-6 ha. Since this is a large experiment, 1 ha is devoted to each fertilizer type in each field. But, the fields have different crop histories, herbicide use, etc. Field identity is an extraneous variable to control for.

# Randomized Complete Block Design

Field-1

| | |
|---|---|
| A | B |
| C | D |

Field-2

| | |
|---|---|
| A | D |
| C | B |

Field-3

| | |
|---|---|
| C | A |
| B | D |

Note: that in an RCB design, every treatment is represented in every block.

# Randomized Complete Block Design

If the extraneous variability associated with field is not removed prior to testing for a treatment effect, it will show up in the MSE term, making it more difficult to detect treatment effects via an *F*-test.

In other words, the denominator in the *F*-test will be larger than needed because it contains variability associated with field.

# Randomized Complete Block Design

-Linear Model-

The additive model for an RCB design is:

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

where:

$\mu$ : constant; overall mean

$\alpha_i$ : constant for ith treatment group; deviation from mean of i

$\beta_j$ : constant for the jth block; deviation from the mean of j

$\varepsilon_{ij}$ : random deviation associated with each observation

# Randomized Complete Block Design

-Computation-

Data for an RCB are arranged in *i* treatments (columns) and *j* blocks (rows):

Treatment (*i*)

Block (*j*)

| | A | B | C | D | Total |
|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{21}$ | $y_{31}$ | $y_{41}$ | $T_{\cdot 1}$ |
| 2 | $y_{12}$ | $y_{22}$ | $y_{32}$ | $y_{42}$ | $T_{\cdot 2}$ |
| 3 | $y_{13}$ | $y_{23}$ | $y_{33}$ | $y_{43}$ | $T_{\cdot 3}$ |
| | $T_{1\cdot}$ | $T_{2\cdot}$ | $T_{3\cdot}$ | $T_{4\cdot}$ | $T_{\cdot\cdot}$ |

# Randomized Complete Block Design

-Computation-

We can use the "T-dot" notation to refer to treatment totals (e.g., $T_{i\cdot}$), indicating that observation were summed over *j* blocks (and vice versa for treatment totals).

The uncorrected sum of squares, corrected sum of squares, and ANOVA procedures are straightforward. In a block design, the error sum of squares is sometimes referred to as the residual sum of squares.

# Randomized Complete Block Design

-Computation-

Uncorrected Sums of Squares

| Sum of Squares | Formula | Symbol | No. of Totals | Obs. per Total |
|---|---|---|---|---|
| Uncorrected Total | $\Sigma\Sigma\, y_{ij}^2$ | T | ab | 1 |
| Uncorrected Trt. | $\Sigma T_{i\cdot}^2/b$ | A | a | b |
| Ucorrected Blk. | $\Sigma T_{\cdot j}^2/a$ | B | b | a |
| Residual | $T_{\cdot\cdot}^2/ab$ | CF | 1 | ab |

# Randomized Complete Block Design

-Computation-

Corrected Sums of Squares

| Sum of Squares | df | Symbol | Definition | Computation |
|---|---|---|---|---|
| Total | ab-1 | $SS_t$ | $\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2$ | T-CF |
| Treatment | a-1 | $SS_a$ | $b\sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ | A-CF |
| Block | b-1 | $SS_b$ | $a\sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$ | B-CF |
| Residual | (a-1)(b-1) | $SS_e$ | $\sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$ | T-A-B+CF |

# Randomized Complete Block Design
-Computation-

To test the main null hypothesis:

$$H_o : \alpha_1 = \alpha_2 = ...\alpha_a \quad or \quad H_o : \sigma_A^2 = 0$$
$$H_a : \text{at least one inequality} \quad or \quad H_o : \sigma_A^2 > 0$$

we can generate the final ANOVA table in the usual manner...

# Randomized Complete Block Design
-Computation-

| Sum of Squares | df | SS | MS | F |
|---|---|---|---|---|
| Among treatments | a-1 | $SS_a$ | $MS_a$ | $MS_a/MS_e$ |
| Among blocks | b-1 | $SS_b$ | $MS_b$ | $MS_b/MS_e$ |
| Residual | (a-1)(b-1) | $SS_e$ | $MS_e$ | |
| Total | ab-1 | $SS_t$ | | |

# Randomized Complete Block Design
-Example-

Consider a situation in which 4 genetic families of green beans are treated with 3 fertilizers:

Fertilizer ($i$)

| Family ($j$) | | 1 | 2 | 3 | $T_{.j}$ |
|---|---|---|---|---|---|
| | a | 4.7 | 9.4 | 6.3 | 20.4 |
| | b | 3.5 | 7.6 | 5.1 | 16.2 |
| | c | 0.1 | 5.3 | 1.8 | 7.2 |
| | d | 1.6 | 6.2 | 3.6 | 11.4 |
| | $T_{i.}$ | 9.9 | 28.5 | 16.8 | T..= 55.2 |

# Randomized Complete Block Design

-Example-

For this experiment,

a = 3
b = 4
T = 331.46 (sum of the $y_{ij}^2$)
A = 298.13 [=$(9.9^2/4)+(28.5^2/4)+(16.8^2/4)$]
B = 286.80 [=$(20.4^2/3)+(16.2^2/3)+(7.2^2/3)+(11.4^2/3)$]
CF = 253.92 [=$(55.2^2/12)$]

Now the ANOVA table can be generated as...

# Randomized Complete Block Design

-Example-

| Source | df | SS | MS | $F_{calc}$ | $F_{table}$ |
|---|---|---|---|---|---|
| Among Trt | 2 | 44.205 (A-CF) | 22.102 | 290.8 | 5.143 |
| Among Blk | 3 | 32.880 (B-CF) | 10.96 | 144.2 | 4.757 |
| Residual | 6 | 0.455 (T-A-B+CF) | 0.076 | | |

# Randomized Complete Block Design

It may not always be advantageous to use the RCB design.

When a block design is appropriate, along with the reduction of the error SS is an associated reduction in the df. If blocking is used when it is not needed, the reduction of the error SS may be such that it reduces the power of the test and yields a NS result when in fact there really is a significant difference (i.e., a Type-I error).

A block design is best used where an environmental gradient is likely to be present (e.g., greenhouses, growth chambers, field plots, etc.)

## Randomized Complete Block Design

-Example in R -

```
> corn<-read.csv("C:/TEMPR/Corn.csv")
> corn
   Block Fert Yield
1      a    1   4.7
2      a    2   9.4
3      a    3   6.3
4      b    1   3.5
5      b    2   7.6
6      b    3   5.1
7      c    1   0.1
8      c    2   5.3
9      c    3   1.8
10     d    1   1.6
11     d    2   6.2
12     d    3   3.6
```

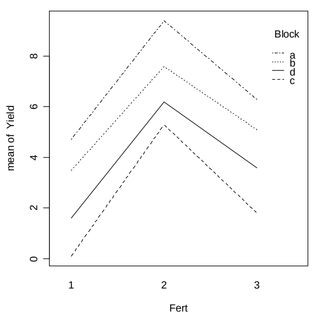
## Randomized Complete Block Design

-Example in R -

```
> attach(corn)
> xtabs(Yield ~ Fert +
Block)

    Block
Fert   a   b   c   d
   1 4.7 3.5 0.1 1.6
   2 9.4 7.6 5.3 6.2
   3 6.3 5.1 1.8 3.6

> interaction.plot(Fert,
Block,Yield)
```




## Randomized Complete Block Design

-Example in R -

```
> anova(lm(Yield~Block+Fert))

Analysis of Variance Table

Response: Yield
          Df Sum Sq Mean Sq F value    Pr(>F)
Block      3 32.880  10.960  144.53 5.534e-06 ***
Fert       2 44.205  22.102  291.46 1.057e-06 ***
Residuals  6  0.455   0.076
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1
```

NB: Same results as before; Block & Fert very sig.

# Randomized Complete Block Design

-Example in R -

```
> anova(lm(Yield~Fert))

Analysis of Variance Table

Response: Yield
          Df Sum Sq Mean Sq F value  Pr(>F)
Fert       2 44.205  22.102  5.9674 0.02240 *
Residuals  9 33.335   3.704
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
‘.’ 0.1 ‘ ’ 1
```

NB: What happens when block is not included in model?

RCB Example
Check Assumptions

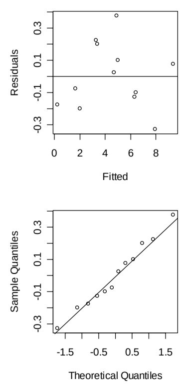
```
> Model<-
(lm(Yield~Block+Fert))

> plot(fitted(Model),
+   residuals(Model),
+   xlab="Fitted",ylab=
+   "Residuals")

> abline(h=0)

> qqnorm(residuals(Model))

> qqline(residuals(Model))
```




# Latin Square Design

Sometimes an investigator may be aware of two sources of extraneous variability. In this case, the RCB design is no longer effective.

The Latin Square (LS) design is a simple extension of the RCB design that permits blocking in two directions.

The LS design was originally conceived of for agriculture experiments to deal with gradients of moisture and soil nutrients.

# Latin Square Design

Consider the example where the seeds of 4 varieties of corn (1-4) are stored under 4 conditions (1-4) and then planted in a field and subjected to 4 different fertilizers (A-D):

Storage

| Variety | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | A | B | C | D |
| 2 | C | A | D | B |
| 3 | B | D | A | C |
| 4 | D | C | B | A |

NB: each treatment appears exactly once in each row and column (by design).

# Latin Square Design

There are many applications for this sort of design in biological experimentation.

Besides field experiments, the LS design is very appropriate in greenhouse and phytotron experiments:

In *greenhouses*, you may be forced to run a large experiment on multiple benches and multiple rooms. In *phytotrons*, plants may have to be placed on different shelves and different chambers. Both of these are common situations in experimental plant biology.

# Latin Square Design

-Model-

Note the similarity of the model with the RCB design. We have simply added one additional term to account for the second source of extraneous variation:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

where:

$\mu$ : constant; overall mean

$\alpha_i$ : constant for ith treatment group; deviation from mean of i

$\beta_j$ : constant for the jth source of variation; deviation from the mean of j

$\gamma_k$ : constant for the kth source of variation; deviation from the mean of k

$\varepsilon_{ijk}$ : random deviation associated with each observation

# Latin Square Design
-Computations-

| Source | df | SS | MS | F |
|---|---|---|---|---|
| Among $\alpha$ trt | a-1 | $SS_a$=A-CF | $MS_a=SS_a/(a-1)$ | $MS_a/MS_e$ |
| Among $\beta$ effects | a-1 | $SS_b$=B-CF | $MS_b=SS_b/(a-1)$ | $MS_b/MS_e$ |
| Among $\gamma$ effects | a-1 | $SS_c$=C-CF | $MS_c=SS_c/(a-1)$ | $MS_c/MS_e$ |
| Residual | (a-1)(a-2) | $SS_e$=T-A-B-C+2CF | MSe=SSe/(a-1)(a-2) | |
| Total | $a^2$-1 | $SS_t$=T-CF | | |

# Latin Square Design
-Computations-

$$T = \sum_j \sum_k y_{ijk}^2$$
$$A = \sum_i T_{i..}^2 / a$$
$$B = \sum_j T_{.j.}^2 / a$$
$$C = \sum_k T_{..k}^2 / a$$
$$CF = T_{...}^2 / a^2$$

Computations are performed in virtually the same way as we did for the RCB design, only now we add another set of totals for the second effect (gamma).

# Latin Square Design
-Example-

An ethnobotanist is estimating the strength of cloth woven from hemp grown by 4 farmers (A,B,C,D,). Four (4) weavers are selected and asked to weave one piece of cloth from hemp grown by each farmer each day. Each woven piece requires ca. 12 hours to complete. To account for differences among weavers (skill level) and among days (fatigue) the experimenter blocks both of the factors and conducts an LS analysis of the data...

# Latin Square Design
-Example-

| | Day | | | | |
|---|---|---|---|---|---|
| Weaver | 1 | 2 | 3 | 4 | Totals |
| 1 | B 810 | C 1080 | A 700 | D 910 | 3500 |
| 2 | C 1100 | D 880 | B 780 | A 600 | 3360 |
| 3 | D 840 | A 540 | C 1055 | B 830 | 3265 |
| 4 | A 650 | B 740 | D 1025 | C 900 | 3315 |
| Totals | 3400 | 3240 | 3560 | 3240 | 13440 |

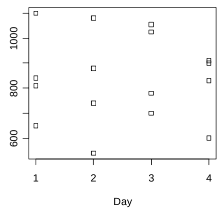
# Latin Square Design
-Example-

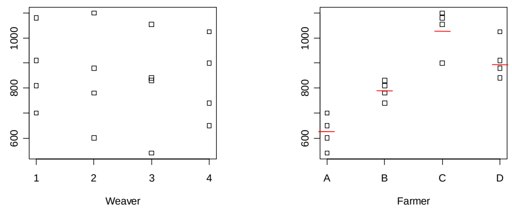
The resulting ANOVA table for these data would be:

| Source | df | SS | MS | F | P |
|---|---|---|---|---|---|
| Farmer | 3 | 371138 | 123712 | 19.93 | 0.005 |
| Weaver | 3 | 7662 | 2554 | 0.41 | ns |
| Days | 3 | 17600 | 5867 | 0.95 | ns |
| Residual | 6 | 37250 | 6208 | | |
| Total | 15 | 433650 | | | |

```
> ethno<-read.csv("C:/TEMPR/Ethno.csv")
> ethno
   Day Weaver Farmer Strength
1    1      1      B      810
2    2      1      C     1080
3    3      1      A      700
4    4      1      D      910
5    1      2      C     1100
6    2      2      D      880
7    3      2      B      780
8    4      2      A      600
9    1      3      D      840
10   2      3      A      540
11   3      3      C     1055
12   4      3      B      830
13   1      4      A      650
14   2      4      B      740
15   3      4      D     1025
16   4      4      C      900
> attach(ethno)
```

```
> windows (height=4, width=4)
>
with(ethno,stripchart(Strength~Day,
xlab="Day",vert=T))
>
with(ethno,stripchart(Strength~Weav
er,xlab="Weaver",vert=T))
>
with(ethno,stripchart(Strength~Farm
er,xlab="Farmer",vert=T))
```




## Latin Square Design

-Example in R -

```
> anova(lm(Strength ~ Farmer+Weaver+Day))

Analysis of Variance Table

Response: Strength
          Df Sum Sq Mean Sq F value   Pr(>F)
Farmer     3 371138  123713 19.9268 0.001602 **
Weaver     3   7662    2554  0.4114 0.750967
Day        3  17600    5867  0.9450 0.475896
Residuals  6  37250    6208
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
‘.’ 0.1 ‘ ’ 1
```