Factorial Experiments

Norberto E. Milla, Jr.

2025-11-12

Review

Design structure

- refers to the grouping of the experimental units into homogeneous groups or blocks
  - RCBD
  - Latin Square Design

Treatment structure

- refers to the set of treatments, treatment combinations, or populations that the experimenter has selected to study and/or compare
  - one-way treatment structure
  - two-way treatment structure

Examples of two-way treatment structures

Consider a researcher who is interested in determining whether a new mathematics curriculum is better at helping students develop spatial visualization skills. Furthermore, he wonders whether if there is a difference between boys and girls, because it is known that males tend to be better at spatial visualization than females. The researcher has the following two-way (two-by-two) factorial design.

Factors:

- curriculum (new, conventional)
- gender(male. female)

Treatments:

- $T_1 =$ (new, male)
- $T_2 =$ (new, female)
- $T_3 =$ (conventional, male)
- $T_4 =$ (conventional, female)

Examples of two-way treatment structures

An experiment is conducted to assess the effect of temperature and shading on seedling growth

Factors:

- Temperature (**H**igh, **M**edium, **L**ow)
- Shading (**F**ull, **P**artial)

Treatments:

- T_1=(H, F)
- T_2=(H, P)
- T_3=(M, F)
- T_4=(M, P)
- T_5=(L, F)
- T_6=(L, P)

Factorial experiment

- It is an experiment where the treatments consist of all possible combinations of the levels of two or more factors
- It is useful when the researcher is interested on the effect of the two or more factors on a single experiment
- Treatments are formed by combining the levels of two (or more) factors

Advantages:

- The combined effect (interaction) of one factor in the presence of another factor can be evaluated in one run of the experiment
- It provides test for independence of factors
- An estimate of the interaction of the factors may be obtained

Types of treatment effects in a factorial experiment

Simple effect- difference between each level of one factor at each level of the other factor

Main effect- average of the simple effects of one factor over all levels of the other factor

Interaction effect- difference in the simple effect of one factor at two different levels of the other factor

EXAMPLE:

Consider an experiment with two factors, A and B, each at two levels: A_1 and A_2 for factor A and B_1 and B_2 for B. We assume $A_2 > A_1$ and $B_2 > B_1$. The mean response for each treatment is shown in the following table.

Types of treatment effects in a factorial experiment

	B_1	B_2	Simple effect of B
A_1	$\overline{y}_1$	$\overline{y}_2$	$\overline{y}_2 - \overline{y}_1$
A_2	$\overline{y}_3$	$\overline{y}_4$	$\overline{y}_4 - \overline{y}_3$
Simple effect of A	$\overline{y}_3 - \overline{y}_1$	$\overline{y}_4 - \overline{y}_2$	

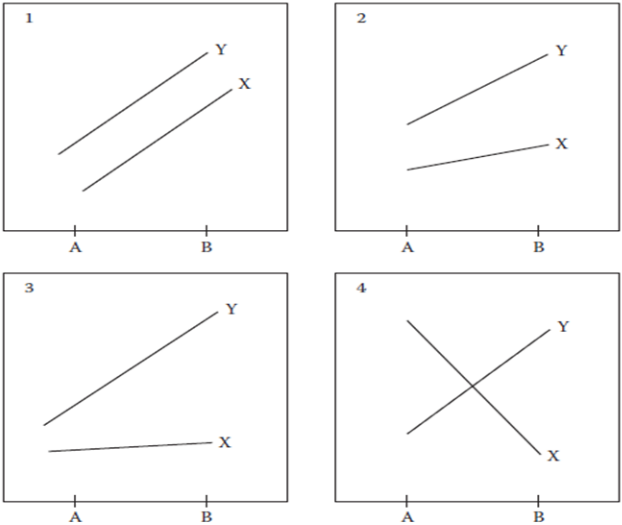
main effect of A $= \frac{(\overline{y}_3 - \overline{y}_1) + (\overline{y}_4 - \overline{y}_2)}{2}$

main effect of B $= \frac{(\overline{y}_2 - \overline{y}_1) + (\overline{y}_4 - \overline{y}_3)}{2}$

interaction effect$= (\overline{y}_4 - \overline{y}_2) - (\overline{y}_3 - \overline{y}_1)$

interaction effect$= (\overline{y}_4 - \overline{y}_3) - (\overline{y}_2 - \overline{y}_1)$

Forms of interaction effects



Two-factor factorial experiment in CRD

- Two factors to be tested
  - A with a levels
  - B with b levels
  - $r =$ no. of replications
- Treatments are assigned completely at random to homogeneous independent experimental subjects

Linear model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where:

- Y_{ijk} is the response of the k^{th} experimental unit given the $(ij)^{th}$ treatment combination; $i = 1, 2, 3, \cdots, a$; $j = 1, 2, 3, \cdots, b$; $k = 1, 2, 3, \cdots, r$
- μ is the overall mean response
- α_i is the effect of i^{th} level of A
- β_j is the effect of the j^{th} level of B
- $(\alpha\beta)_{ij}$ is the interaction effect
- ϵ_{ijk} is random variation

Two-factor factorial experiment in CRD

Assumptions:

1. $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$
2. Model I (both A and B are *fixed* factors)
   $\Rightarrow \alpha_i, \beta_j, (\alpha\beta)_{ij}$ are fixed with $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, and $\sum_i \sum_j (\alpha\beta)_{ij} = 0$
3. Model II (both A and B are *random* factors)
   $\Rightarrow \alpha_i, \beta_j, (\alpha\beta)_{ij}$ are random variables with $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_j \sim N(0, \sigma_\beta^2)$, and $(\alpha\beta)_{ij} \sim N(0, \sigma_{\alpha\beta}^2) = 0$
4. Model III (A fixed, B random)
   $\Rightarrow \sum_i \alpha_i = 0, \beta_j \sim N(0, \sigma_\beta^2), \sum_i \sum_j (\alpha\beta)_{ij} = 0$
5. Model III (A random, B fixed)
   $\Rightarrow \alpha_i \sim N(0, \sigma_\alpha^2), \sum_j \beta_j = 0, \sum_i \sum_j (\alpha\beta)_{ij} = 0$

Two-factor factorial experiment in CRD

Estimates of parameters:

$$\begin{aligned}
\hat{\mu} &= \overline{Y}_{...} \\
\hat{\alpha}_i &= \overline{Y}_{i...} - \overline{Y}_{...} \\
\hat{\beta}_j &= \overline{Y}_{.j.} - \overline{Y}_{...} \\
(\hat{\alpha\beta})_{ij} &= \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...} \\
\hat{\epsilon}_{ijk} &= Y_{ijk} - \overline{Y}_{ij.}
\end{aligned}$$

Two-factor factorial experiment in CRD

Analysis of Variance Table

Source of Variation	df	SS	MS	F
A	$a-1$	SSA	MSA	
B	$b-1$	SSB	MSB	
AB	$(a-1)(b-1$	SSAB	MSAB	
Experimental Error	$ab(r-1)$	SSE	MSE	
TOTAL	$n-1$	SST		

Two-factor factorial experiment in CRD

Computing formulas of the sums of squares:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{Y_{...}^2}{n}; n = abr$$

$$SSTr = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}^2 - \frac{Y_{...}^2}{n}$$

$$SSA = \frac{1}{br}\sum_{i=1}^{a} Y_{i..}^2 - \frac{Y_{...}^2}{n}$$

$$SSB = \frac{1}{ar}\sum_{j=1}^{b} Y_{.j.}^2 - \frac{Y_{...}^2}{n}$$

$$SSAB = SSTr - SSA - SSB$$

$$SSE = SST - SSTr$$

Expected Mean Squares

1. Prepare a two way table with the terms of the model as the row labels.
2. Write the subscripts in the model as column headings. Over each subscript, write the number of levels of the factors associated with that subscript and whether the factor is fixed (F) or random (R). Replicates are always treated as random.
3. In each row, write 1 if one of the dead subscripts in the row components matches the subscript in the column.
4. In each row, if any of the live subscripts on the row component match the subscript in the column, write 0 if the column is headed by a fixed factor and 1 if the column is headed by a random factor.

Expected Mean Squares

5. In the remaining cells, write the number of levels shown above the column heading.
6. To obtain the EMS for any model component, first cover all columns headed by live subscripts on that component. Then in each row that contains at least the same subscripts as those on the component being considered, take the product of the visible numbers and multiply by the appropriate fixed or random factor. The sum of these quantities is the EMS of the model component being considered.

Tests of hypothesis of the interaction (AB)

Fixed model

- $H_0 : (\alpha\beta)_{ij} = 0, \forall(i,j)$
- $H_1 : (\alpha\beta)_{ij} \neq 0, \exists(i,j)$

Random model

- $H_0 : \sigma^2_{\alpha\beta} = 0$
- $H_1 : \sigma^2_{\alpha\beta} > 0$

$$F = \frac{MS_{AB}}{MS_E}$$

Test of hypothesis on Factor A

<u>Fixed model</u>

- $H_0 : \alpha_i = 0, \forall i$
- $H_1 : \alpha_i \neq 0, \exists i$

<u>Random model</u>

- $H_0 : \sigma^2_\alpha = 0$
- $H_1 : \sigma^2_\alpha > 0$

$$F = \frac{MS_A}{MS_E}, \text{ if A and B are fixed}$$

$$F = \frac{MS_A}{MS_{AB}}, \text{ if A and B are random}$$

$$F = \frac{MS_A}{MS_{AB}}, \text{ if A is fixed and B is random}$$

$$F = \frac{MS_A}{MS_E}, \text{ if A is random and B is fixed}$$

Test of hypothesis on Factor B

Fixed model

- $H_0 : \beta_j = 0, \forall j$
- $H_1 : \beta_j \neq 0, \exists j$

Random model

- $H_0 : \sigma^2_\beta = 0$
- $H_1 : \sigma^2_\beta > 0$

$$F = \frac{MS_B}{MS_E}, \text{ if A and B are fixed}$$

$$F = \frac{MS_B}{MS_{AB}}, \text{ if A and B are random}$$

$$F = \frac{MS_B}{MS_E}, \text{ if A is fixed and B is random}$$

$$F = \frac{MS_B}{MS_{AB}}, \text{ if A is random and B is fixed}$$

Sequential Test of hypothesis: important guidelines

- First, do the test on the interaction effects
    - If the interaction term is significant, that tells us that the effect of A is different at each level of B. Or you can say it the other way, the effect of B differs at each level of A
    - Therefore, when we have significant interaction, it is not very sensible to even be talking about the main effect of A and B, because these change depending on the level of the other factor
    - If the interaction is significant then we want to estimate and focus our attention on the cell means.
- If the interaction is not significant, then we can test the main effects and focus on the main effect means.

Two-factor factorial in CRD: an example

An experiment was conducted to determine the effects of four different pesticides (P1, P2, P3, P4) on the yield of fruit from three different varieties (V1 ,V2 ,V3) of a citrus tree. Eight trees from each variety were randomly selected from an orchard. The four pesticides were then randomly assigned to two trees of each variety and applications were made according to recommended levels. Yields of fruit (in bushels per tree) were obtained after the test period.

	Pesticide, A			
Variety, B	**1**	**2**	**3**	**4**
1	49	50	43	53
	39	55	38	48
2	55	67	53	85
	41	58	42	73
3	66	85	69	85
	68	92	62	99

Two-factor factorial in CRD: an example

	V1	V2	V3	P Totals
P1	88	96	134	318
P2	105	125	177	407
P3	81	95	131	307
P3	101	158	184	443
V Totals	375	474	626	**1475**

Two-factor factorial in CRD: an example

$$
\begin{aligned}
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{Y_{...}^2}{n} \\
&= (49^2 + 39^2 + \cdots + 99^2) - \frac{1475^2}{24} \\
&\approx 7187.958
\end{aligned}
$$

$$
\begin{aligned}
SSTr &= \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}^2 - \frac{Y_{...}^2}{n} \\
&= \left(\frac{88^2}{2} + \frac{105^2}{2} + \cdots + \frac{184^2}{2}\right) - \frac{1475^2}{24} \\
&\approx 6680.458
\end{aligned}
$$

Two-factor factorial in CRD: an example

$$
\begin{aligned}
SSA &= \frac{1}{br}\sum_{i=1}^{a} Y_{i..}^2 - \frac{Y_{...}^2}{n} \\
&= \left(\frac{138^2}{6} + \frac{407^2}{6} + \cdots + \frac{443^2}{6}\right) - \frac{1475^2}{24} \\
&\approx 2227.458
\end{aligned}
$$

$$
\begin{aligned}
SSB &= \frac{1}{ar}\sum_{j=1}^{b} Y_{.j.}^2 - \frac{Y_{...}^2}{n} \\
&= \left(\frac{375^2}{8} + \frac{474^2}{8} + \frac{3626^2}{8}\right) - \frac{1475^2}{24} \\
&\approx 3996.083
\end{aligned}
$$

Two-factor factorial in CRD: an example

$$
\begin{aligned}
SSAB &= SSTr - SSA - SSB \\
&= 6680.458 - 2227.458 - 3996.083 \\
&= 456.917
\end{aligned}
$$

$$
\begin{aligned}
SSE &= SST - SSTr \\
&= 7187.958 - 6680.458 \\
&= 507.50
\end{aligned}
$$

Two-factor factorial in CRD: an example

ANOVA Table (Fixed Model)

Source of variation	df	SS	MS	F	p
Pesticide (A)	3	2227.458	742.486	17.56	0.0001
Variety (B)	2	3996.083	1998.042	47.24	$p<0.0001$
Interaction(AB)	6	456.917	76.153	1.8	0.1818
Exptl error	12	507.500	42.292		
TOTAL	23	7187.958			

Two-factor factorial in CRD: an example

Test of hypotheses (Fixed model)

A. Test of significance of AB interaction effect

- $H_0 : (\alpha\beta)_{ij} = 0, \forall(ij)$
- $H_1 : (\alpha\beta)_{ij} \neq 0, \exists(ij)$
- Test statistic: $F = 1.8; p = 0.1818$
- Decision: Do not reject H_0.
- Conclusion: At $\alpha = 0.05$, the effect of the different pesticides on the yield is the same for all varieties (No significant interaction effect.)

Two-factor factorial in CRD: an example

Test of hypotheses (Fixed model)

B. Test of significance of the main effect of Pesticide (A)

- $H_0 : \alpha_i = 0, \forall i$
- $H_1 : \alpha_i \neq 0, \exists i$
- Test statistic: $F = 17.56, p = 0.0001$
- Decision: Reject H_0.
- Conclusion: At $\alpha = 0.05$, the mean yield are significantly different for different pesticides. In other words, use of different pesticides significantly affect yield.

NOTE: Proceed to post-hoc test of the pesticide means.

Two-factor factorial in CRD: an example

Test of hypotheses (Fixed model)

C. Test of significance of the main effect of Variety (B)

- $H_0 : \beta_j = 0, \forall j$
- $H_1 : \beta_j \neq 0, \exists j$
- Test statistic: $F = 47.24, p < 0.0001$
- Decision: Reject H_0.
- Conclusion: At $\alpha = 0.05$, the different varieties have significantly different mean yields.

NOTE: Proceed to post-hoc test of the variety means.

Two-factor factorial in CRD: an example

ANOVA Table (Random Model)

Source of variation	df	SS	MS	F	p
Pesticide (A)	3	2227.458	742.486	9.75	0.01007
Variety (B)	2	3996.083	1998.042	26.24	0.00108
Interaction(AB)	6	456.917	76.153	1.8	0.1818
Exptl error	12	507.500	42.292		
TOTAL	23	7187.958			

Two-factor factorial in CRD: an example

Test of hypotheses (Random model)

A. Test of significance of $\sigma^2_{\alpha\beta}$

- $H_0 : \sigma^2_{\alpha\beta} = 0$
- $H_1 : \sigma^2_{\alpha\beta} > 0$
- Test staistic: $F = 1.8, p = 0.1818$
- Decision: Do not reject H_0.
- Conclusion: At $\alpha = 0.05$, there is no significant variation in the yield of the different combinations of pesticides and varieties.

Two-factor factorial in CRD: an example

Test of hypotheses (Random model)

B. Test of significance of σ_α^2

- $H_0 : \sigma_\alpha^2 = 0$
- $H_1 : \sigma_\alpha^2 > 0$
- Test staistic: $F = 9.75, p = 0.01007$
- Decision: Reject H_0.
- Conclusion: At $\alpha = 0.05$, there is significant variation in the yield of the trees applied with different pesticides.

Two-factor factorial in CRD: an example

Test of hypotheses (Random model)

C. Test of significance of σ^2_β

- $H_0 : \sigma^2_\beta = 0$
- $H_1 : \sigma^2_\beta > 0$
- Test staistic: $F = 26.24, p = 0.00108$
- Decision: Reject H_0.
- Conclusion: At $\alpha = 0.05$, there is significant variation in the yield of the different varieties of trees.

NOTE: Post hoc tests are not conducted in random (and/or mixed) models, instead we estimate the variance components.

Two-factor factorial in CRD: an example

Estimation of the variance components

- For a (balance) two-factor factorial random model, the expected means squares are:

$$
\begin{aligned}
E[MS_A] &= br\sigma^2_{\alpha} + r\sigma^2_{\alpha\beta} + \sigma_{\epsilon} \\
E[MS_B] &= ar\sigma^2_{\alpha} + r\sigma^2_{\alpha\beta} + \sigma_{\epsilon} \\
E[MS_{AB}] &= r\sigma^2_{\alpha\beta} + \sigma_{\epsilon} \\
E[MS_E] &= \sigma_{\epsilon}
\end{aligned}
$$

- In this example, we have $a = 4, b = 3,$ and $r = 2$
- Hence,$\hat{\sigma}^2_{\epsilon} = MSE = 42.292$

Two-factor factorial in CRD: an example

Estimation of the variance components

- Based on the EMS, we have $MS_{AB} = r\hat{\sigma}^2_{\alpha\beta} + \hat{\sigma}^2_{\epsilon}$

$$\begin{aligned}
\Rightarrow 76.153 &= 2\hat{\sigma}^2_{\alpha\beta} + 42.292 \\
\Rightarrow \hat{\sigma}^2_{\alpha\beta} &= \frac{76.153 - 42.292}{2} \\
\Rightarrow \hat{\sigma}^2_{\alpha\beta} &= 16.9305
\end{aligned}$$

- Similarly, $MS_B = ar\hat{\sigma}^2_{\beta} + r\hat{\sigma}^2_{\alpha\beta} + \hat{\sigma}^2_{\epsilon} = ar\hat{\sigma}^2_{\beta} + MS_{AB}$

$$\begin{aligned}
\Rightarrow 1998.042 &= 4(2)\hat{\sigma}^2_{\beta} + MS_{AB} \\
\Rightarrow \hat{\sigma}^2_{\beta} &= \frac{1998.042 - 76.153}{8} \\
\Rightarrow \hat{\sigma}^2_{\beta} &= 240.236125
\end{aligned}$$

Two-factor factorial in CRD: an example

- Finally, $MS_A = br\hat{\sigma}^2_\alpha + r\hat{\sigma}^2_{\alpha\beta} + \hat{\sigma}^2_\epsilon = br\hat{\sigma}^2_\alpha + MS_{AB}$

$$\begin{aligned}\Rightarrow 742.486 &= 3(2)\hat{\sigma}^2_\alpha + MS_{AB} \\ \Rightarrow \hat{\sigma}^2_\alpha &= \frac{742.486 - 76.153}{6} \\ \Rightarrow \hat{\sigma}^2_\alpha &= 111.056\end{aligned}$$

- Therefore,

$$\begin{aligned}\hat{V}(Y_{ijk}) &= \hat{\sigma}^2_\alpha + \hat{\sigma}^2_\beta + \hat{\sigma}^2_{\alpha\beta} + \hat{\sigma}^2_\epsilon \\ &= 111.056 + 240.236125 + 16.9305 + 42.292 \\ &= 410.514625\end{aligned}$$

Two-factor factorial in RCBD

- Two factors to be tested
  - A with a levels
  - B with b levels
  - r blocks of ab homogeneous experimental units
- The ab treatments combinations are assigned completely at random to the homogeneous units in each block

Linear model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \rho_k + \epsilon_{ijk}$$

where:

- Y_{ijk} is the response of the k^{th} experimental unit given the $(ij)^{th}$ treatment combination; $i = 1, 2, 3, \cdots, a$; $j = 1, 2, 3, \cdots, b$; $k = 1, 2, 3, \cdots, r$
- μ is the overall mean response
- α_i is the effect of i^{th} level of A
- β_j is the effect of the j^{th} level of B
- $(\alpha\beta)_{ij}$ is the interaction effect
- ρ_k is the block effect
- ϵ_{ijk} is random variation

Two-factor factorial in RCBD

Estimates of parameters:

$$\hat{\mu} = \overline{Y}_{...}$$

$$\hat{\alpha}_i = \overline{Y}_{i...} - \overline{Y}_{...}$$

$$\hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$$

$$(\hat{\alpha\beta})_{ij} = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}$$

$$\hat{\rho}_k = \overline{Y}_{..k} - \overline{Y}_{...}$$

$$\hat{\epsilon}_{ijk} = Y_{ijk} - \overline{Y}_{ij.}$$

Assumptions:

- Same as in two-factor factorial in CRD
- Additional assumption:
  - Block and treatment effects are additive and have no interaction

Two-factor factorial in RCBD

Analysis of Variance Table

Source of Variation	df	SS	MS	F
Block	$r-1$	SSR	MSR	
A	$a-1$	SSA	MSA	
B	$b-1$	SSB	MSB	
AB	$(a-1)(b-1$	SSAB	MSAB	
Experimental Error	$(ab-1)(r-1)$	SSE	MSE	
TOTAL	$n-1$	SST		

Two-factor factorial in RCBD: Sums of squares

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{Y_{...}^2}{n}; n = abr$$

$$SSTr = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}^2 - \frac{Y_{...}^2}{n}$$

$$SSR = \frac{1}{ab}\sum_{k=1}^{r} Y_{..k}^2 - \frac{Y_{...}^2}{n}$$

$$SSA = \frac{1}{br}\sum_{i=1}^{a} Y_{i..}^2 - \frac{Y_{...}^2}{n}$$

$$SSB = \frac{1}{ar}\sum_{j=1}^{b} Y_{.j.}^2 - \frac{Y_{...}^2}{n}$$

$$SSAB = SSTr - SSA - SSB$$

$$SSE = SST - SSTr - SSR$$

Two-factor factorial in RCBD: an example

An experiment was conducted in an RCB design structure (with days as blocks) to aid in developing a product that can be used as a substrate for making ribbons. The treatment structure is two-way with one factor consisting of three different base (B) polymers (mylar, nylon and polyethylene). The second factor consisted of five different additives (A) that could be included to enhance the formulation. The additives are denoted by: c1, c2, c3, c4, and c5. The variable of primary interest is the tensile strength of the resulting ribbon. The data is contained in the file *two_factor_rcbd.csv*.

Two-factor factorial in RCBD: an example

Table of totals

	Mylar	Nylon	Peth	A Totals
c1	25.8	22.8	25.8	**74.4**
c2	25.7	23.2	34.4	**83.3**
c3	26.6	29.2	27.6	**83.4**
c4	32.4	23.6	26.3	**82.3**
c5	29.9	26.2	25.7	**81.8**
B Totals	**140.4**	**125**	**139.8**	**405.2**

Day	Totals
1	144.7
2	138.6
3	121.9

Two-factor factorial in RCBD: an example

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Day	2	18.576444	9.2882222	11.848182	0.0001870
A	4	6.325778	1.5814444	2.017312	0.1192491
B	2	10.145778	5.0728889	6.471046	0.0048959
A:B	8	33.100889	4.1376111	5.277993	0.0004341
Residuals	28	21.950222	0.7839365	NA	NA

- **TASK**: Verify the values of the sums squares in the above table using the computing formulas