ANOVA: Residuals

- Residuals are the difference between what is ACTUALLY observed (Experiment) vs. what is PREDICTED from a model that is used to adequately describe the data

$$\epsilon_{ij} = y_{ij} - \widetilde{y_{ij}}$$

- In One-way ANOVA, what is the model?

$$\underline{y_{ij}} = \mu + \underline{\tau_i} + \epsilon_{ij}$$

μ = grand mean

τ_j = treatment mean

ε_{ij} = error

- What is the prediction?

$$\widetilde{y_{ij}} = \mu + \tau_i$$

"Effects Model"

- **Remember,** we had assumed that the residuals (or errors) are random and **normally distributed.**

So is that assumption valid IF we use the particular model? -> Model Adequacy Check!

$$\varepsilon_{ij} = y_{ij} - \bar{y}_i$$

$\bar{y}$ $\quad (\bar{y}_i - \bar{y})$

$$\varepsilon_{ij} = y_{ij} - \underbrace{\mu - \tau_i}_{} = \bar{y}_i$$

ANOVA: Model Adequacy Checking

e_1 e_2 e_3

$\frac{i-0.5}{20}$

Normality Assumption can be checked using several methods

- A dot diagram
- Histogram of residuals
- Normal probability plot

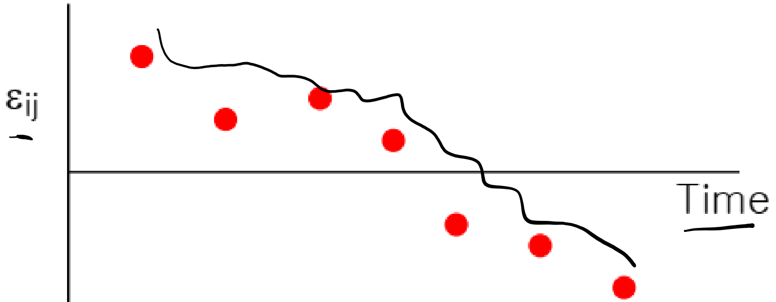
Etch Rate Data and Residuals from Example 3.1[a]

	Observations (j)						
Power (w)	**1**	**2**	**3**	**4**	**5**	$\hat{y}_{ij} = \bar{y}_{i\cdot}$	
160	23.8 575 (13)	–9.2 542 (14)	–21.2 530 (8)	–12.2 539 (5)	18.8 570 (4)	551.2	$\bar{y}_1$
180	–22.4 565 (18)	5.6 593 (9)	2.6 590 (6)	–8.4 579 (16)	22.6 610 (17)	587.4	$\bar{y}_2$
200	–25.4 600 (7)	25.6 651 (19)	–15.4 610 (10)	11.6 637 (20)	3.6 629 (1)	625.4	$\bar{y}_3$
220	18.0 725 (2)	–7.0 700 (3)	8.0 715 (15)	–22.0 685 (11)	3.0 710 (12)	707.0	$\bar{y}_4$

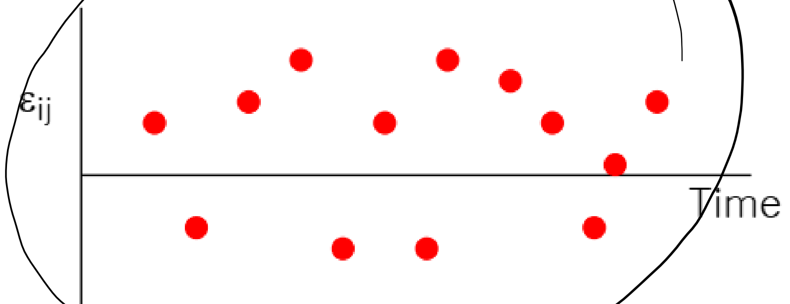
[a]The residuals are shown in the box in each cell. The numbers in parentheses indicate the order in which each experimental run was made.

Model Adequacy Checking

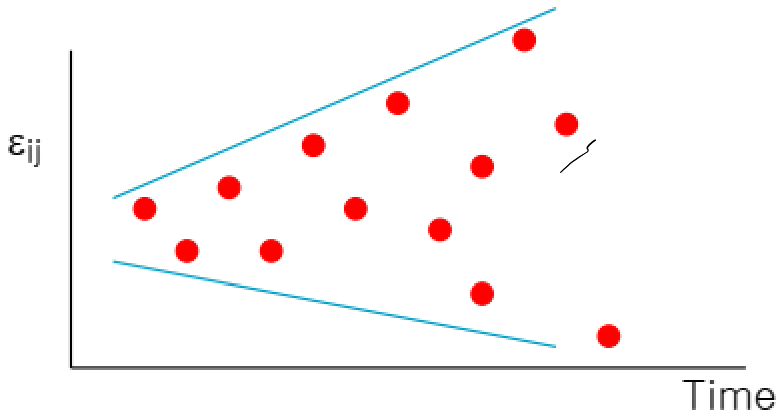
Dot Diagram of Residuals in Time Sequence



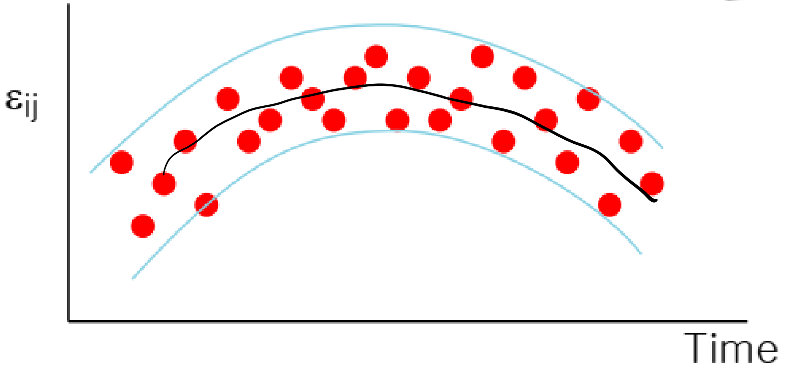
positive correlation



Independence



The variance is not constant, but increases with time.

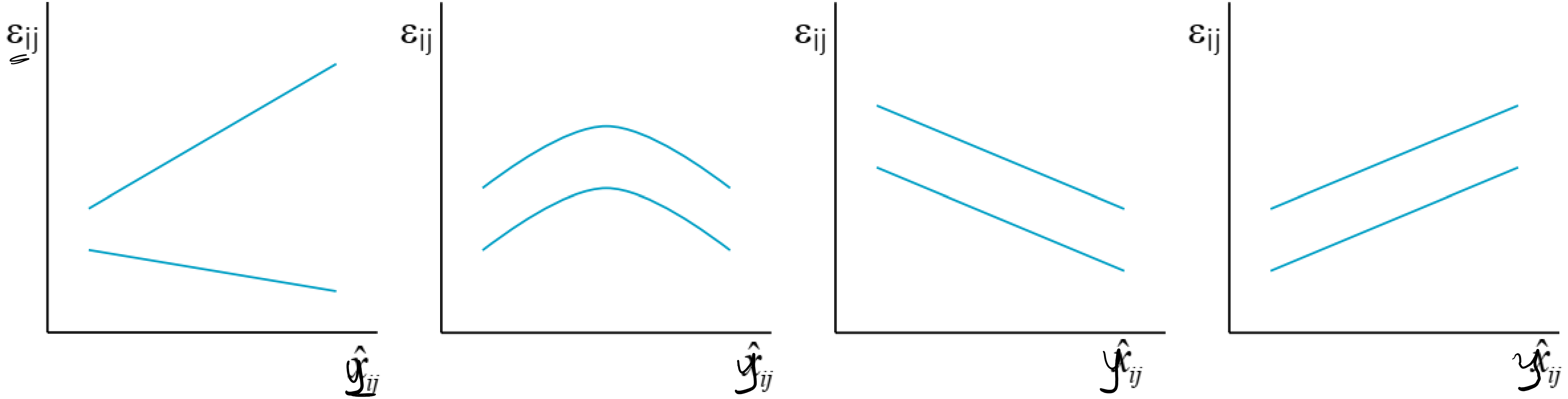


Model is not adequate, a quadratic term (may be interaction term) is needed in the model.

If the model is adequate and the assumptions are satisfied, the errors or residuals, ϵ_{ij} , should have NO structure.

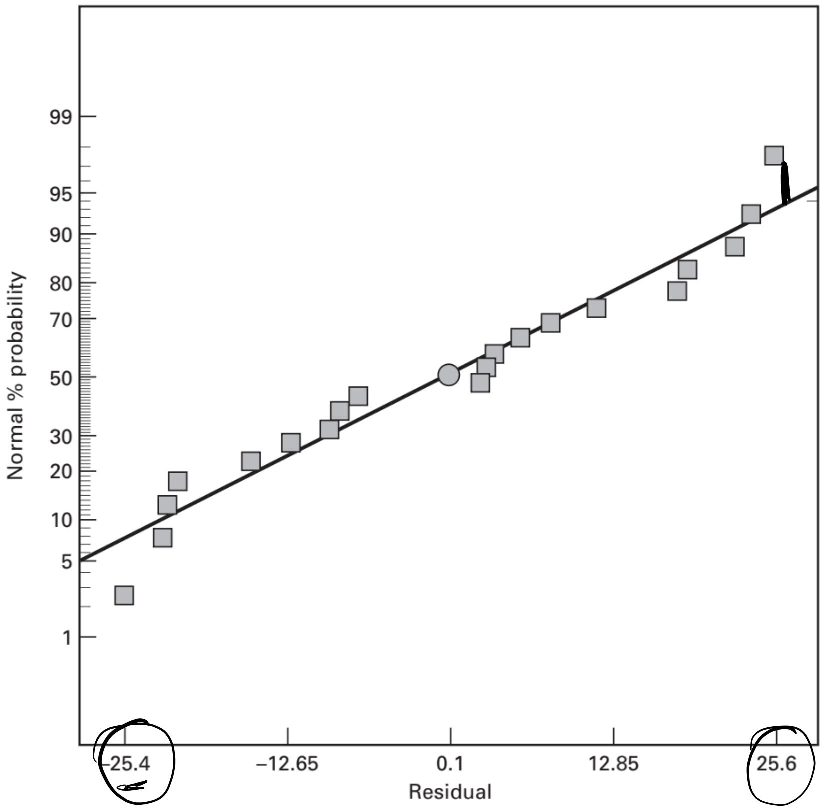
Model Adequacy Checking

Dot Diagram of Residuals (Errors) vs Model Predictions



If the model is adequate and the assumptions are satisfied, the errors or residuals, ϵ_{ij}, should be INDEPENDENT of observations

ANOVA: Model Adequacy Checking



Etch Rate Data and Residuals from Example 3.1[a]

Power (w)	Observations (j) 1	2	3	4	5	$\hat{y}_{ij} = \bar{y}_{i.}$
	23.8	−9.2	−21.2	−12.2	18.8	
160	575 (13)	542 (14)	530 (8)	539 (5)	570 (4)	551.2
	−22.4	5.6	2.6	−8.4	22.6	
180	565 (18)	593 (9)	590 (6)	579 (16)	610 (17)	587.4
	−25.4	25.6	−15.4	11.6	3.6	
200	600 (7)	651 (19)	610 (10)	637 (20)	629 (1)	625.4
	18.0	−7.0	8.0	−22.0	3.0	
220	725 (2)	700 (3)	715 (15)	685 (11)	710 (12)	707.0

[a]The residuals are shown in the box in each cell. The numbers in parentheses indicate the order in which each experimental run was made.

A rough check for outliers may be made by examining the **standardized residuals**

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}} \tag{3.18}$$

If the errors ϵ_{ij} are $N(0, \sigma^2)$, the standardized residuals should be approximately normal with mean zero and unit variance. Thus, about 68 percent of the standardized residuals should fall within the limits ± 1, about 95 percent of them should fall within ± 2, and virtually all of them should fall within ± 3. A residual bigger than 3 or 4 standard deviations from zero is a potential outlier.

For the tensile strength data of Example 3.1, the normal probability plot gives no indication of outliers. Furthermore, the largest standardized residual is

$$d_1 = \frac{e_1}{\sqrt{MS_E}} = \frac{25.6}{\sqrt{333.70}} = \frac{25.6}{18.27} = 1.40$$

which should cause no concern.

Example (DIY)

An article in Nature describes an experiment to investigate the effect of consuming chocolate on cardiovascular health ("Plasma Antioxidants from Chocolate," Nature, Vol. 424, 2003, pp. 1013).

The experiment consisted of using three different types of chocolates: 100 g of dark chocolate, 100 g of dark chocolate with 200 mL of full-fat milk, and 200 g of milk chocolate. Twelve subjects were used, 7 women and 5 men, with an average age range of 32.2 ±1 years, an average weight of 65.8 ± 3.1 kg, and a body-mass index of 21.9 ± 0.4 kgm^{-2}. On different days a subject consumed one of the chocolate-factor levels and one hour later the total antioxidant capacity of their blood plasma was measured in an assay.

Data similar to that summarized in the article are shown in the Table below.

Blood Plasma Levels One Hour Following Chocolate Consumption

	Subjects (Observations)											
Factor	1	2	3	4	5	6	7	8	9	10	11	12
DC	118.8	122.6	115.6	113.6	119.5	115.9	115.8	115.1	116.9	115.4	115.6	107.9
DC+MK	105.4	101.1	102.7	97.1	101.9	98.9	100.0	99.8	102.6	100.9	104.5	93.5
MC	102.1	105.8	99.6	102.7	98.8	100.9	102.8	98.7	94.7	97.8	99.7	98.6

Chapter 2.3

Classical Design of Experiments

Two-Factor ANOVA

Statistical model: What is 'means model' and what is 'effects model' for SINGLE factor EXP?

Single Factor
(multiple settings/levels)

Process/Product

Response

Recap

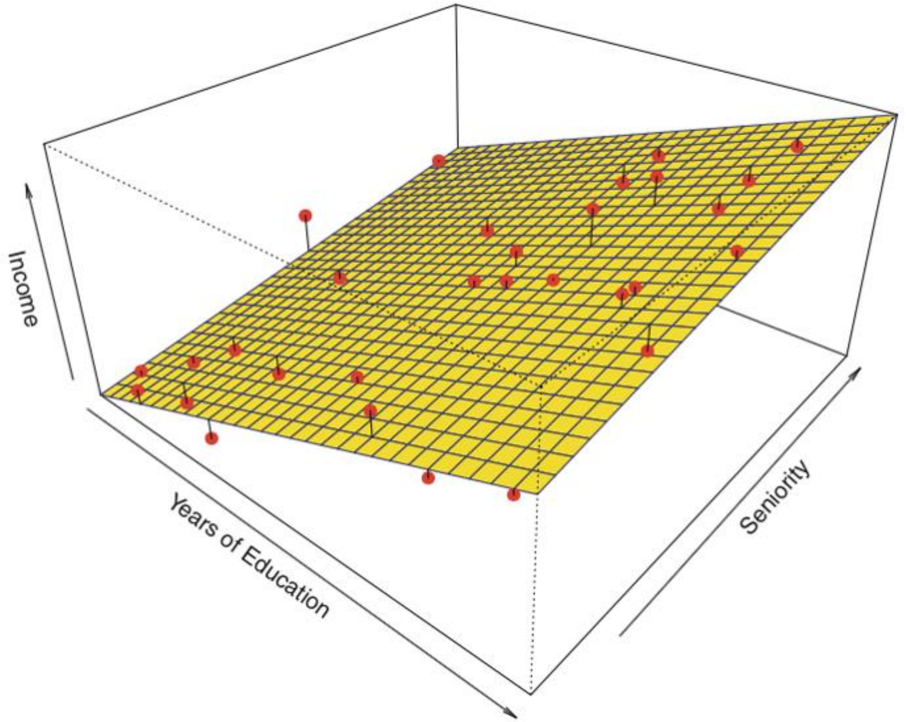
Statistical model: What is ‘means model’ and what is ‘effects model’ for TWO factor EXP?

Factor 1
(multiple settings/levels)

Factor 2
(multiple settings/levels)

Process/Product

Response



Example

We wish to compare four processes which de-ink newspaper. We want about five tests for each of the four processes. Five batches of pulp are prepared. We assume that all chemicals used will be homogeneous. A batch of pulp can run only four tests, and the amount of ink with a particular batch varies greatly.

		Process (Chemicals) (Factor 1: 4 levels)				
		1	2	3	4	Average
Batch (Factor 2: 5 levels)	A	89 (1)	88 (3)	97 (2)	94 (4)	92
	B	84 (4)	77 (2)	92 (3)	79 (1)	83
	C	81 (2)	87 (1)	87 (4)	85 (3)	85
	D	87 (1)	92 (3)	89 (2)	84 (4)	88
	E	79 (3)	81 (4)	80 (1)	88 (2)	82
	Average	84	85	89	86	

Question is WHICH factor(s) have significant effect on the response?
Factor 1? Factor 2? Both? None?

Two-Factor ANOVA: Model

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (\bar{y}_j - \bar{\bar{y}}) + \underbrace{(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{\bar{y}})}_{\text{Residual Error or Intrinsic Variation}}$$

Obs value; Grand mean; Effect of first factor; Effect of second factor; Residual Error or Intrinsic Variation

$$SS_T = SS_{mean} + SS_1 + SS_2 + SS_{error}$$

$N = ab$

DOF	N	1	$a-1$	$b-1$	$N-a-b+1 = (a-1)(b-1)$

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

Obs. — Factor 1 effect — Factor 2 effect — error

" No interaction effect assumed"

Decomposition of Observations

Decomposition of X_{ij}

$$89 = 86 + (84-86) + (92-86) + (89-84-92+86)$$

$$= 86 + (-2) + (6) + (-1)$$

$$X_{ij} = \bar{\bar{X}} + (\bar{X}_{i.} - \bar{\bar{X}}) + (\bar{X}_{.j} - \bar{\bar{X}}) + (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{\bar{X}})$$

$$\begin{bmatrix} 89 & 88 & 97 & 94 \\ 84 & 77 & 92 & 79 \\ 81 & 87 & 87 & 85 \\ 87 & 92 & 89 & 84 \\ 79 & 81 & 80 & 88 \end{bmatrix} = \begin{bmatrix} 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \end{bmatrix} + \begin{bmatrix} 6 & 6 & 6 & 6 \\ -3 & -3 & -3 & -3 \\ -1 & -1 & -1 & -1 \\ 2 & 2 & 2 & 2 \\ -4 & -4 & -4 & -4 \end{bmatrix} + \begin{bmatrix} -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \end{bmatrix} + \begin{bmatrix} -1 & -3 & 2 & 2 \\ 3 & -5 & 6 & -4 \\ -2 & 3 & -1 & 0 \\ 1 & 5 & -2 & -4 \\ -1 & 0 & -5 & 6 \end{bmatrix}$$

SSQ:	148,480	147,920	264	70	226
DOF:	20	1	4	3	12

Two-Factor ANOVA Table

When Factor 1 has 'a' levels and Factor 2 has 'b' levels and all possible combinations (N = ab) are tested.

$F_{a-1,(a-1)(b-1)}$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor 1	$SS_{\text{Factor 1}}$	$a-1$	$\frac{SS_{\text{Factor 1}}}{a-1}$	$\frac{MS_{\text{Factor 1}}}{MS_E}$
Factor 2	$SS_{\text{Factor 2}}$	$b-1$	$\frac{SS_{\text{Factor 2}}}{b-1}$	$\frac{MS_{\text{Factor 2}}}{MS_E}$
Error	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
C Total	C SS_T	$N-1$		

$F_{b-1,(a-1)(b-1)}$

- **Sum of Squares**

1. $$\sum_{i=1}^{N_b}\sum_{j=1}^{N_t} X_{ij}^2 = 89^2 + 84^2 + \cdots = 148,480$$

N_b = # of batches (5)
N_t = # of processes (4)

2. Mean: $$N \cdot \bar{\bar{X}}^2 = 20(86)^2 = 147,920$$

3. Process/ technique: $$\sum_{j=1}^{N_t} N_b(\bar{X}_{\cdot j} - \bar{\bar{X}})^2 = 5(84-86)^2 + 5(85-86)^2 + 5(89-86)^2$$
$$+5(86-86)^2 = 70$$

4. Block/Batch:
$$\sum_{i=1}^{N_b} N_t(\bar{X}_{i\cdot} - \bar{\bar{X}})^2 = 4(92-86)^2 + 4(83-86)^2 + \cdots + 4(82-86)^2 = 264$$

5. Residual: $$\sum\sum[X_{ij} - \bar{\bar{X}} - (\bar{X}_{i\cdot} - \bar{\bar{X}}) - (\bar{X}_{\cdot j} - \bar{\bar{X}})]^2 = 226$$

ANOVA table

Sources	SSQ	DoF	MS	Ratio		F_{th}
Technique (Process)	70	3	23.3	1.24	$F_{3,12}$	3.490
Block (Batch)	264	4	66	3.51	$F_{4,12}$	3.259
Grand Mean	147,920	1				
Residual	226	12	18.8			
Total	148,480	20				

Critical values for F-distribution

- $F_{0.95}(4,12) = 3.259$
- $F_{0.95}(3,12) = 3.490$

TWO Factor ANOVA

How do we statistically determine WHICH effect is significant?

When Factor 1 has ‘a’ levels and Factor 2 has ‘b’ levels and all possible combinations (N = ab) are tested.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor 1	$SS_{\text{Factor 1}}$	$a-1$	$\frac{SS_{\text{Factor 1}}}{a-1}$	$\frac{MS_{\text{Factor 1}}}{MS_E}$
Factor 2	$SS_{\text{Factor 2}}$	$b-1$	$\frac{SS_{\text{Factor 2}}}{b-1}$	$\frac{MS_{\text{Factor 2}}}{MS_E}$
Error	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$N-1$		

Randomized Complete Block Design (RCBD)

- In any experiment, variability arising from a nuisance factor can affect the results. *Generally, we define a nuisance factor as a design factor that probably has an effect on the response, but we are not interested in that effect.*
- Sometimes a nuisance factor is unknown and uncontrolled; that is, we don't know that the factor exists, and it may even be changing levels while we are conducting the experiment. Randomization is the design technique used to guard against such a "lurking" nuisance factor.
- In other cases, the nuisance factor is known but uncontrollable. If we can at least observe the value that the nuisance factor takes on at each run of the experiment, we can compensate for it in the statistical analysis by using the ANOVA
- When the **nuisance source of variability is known and controllable**, a design technique called **blocking** can be used to systematically eliminate its effect on the statistical comparisons among treatments.

RCBD: Example

Does Virat Kohli's performance in IPL change based on which team he plays against?

What kind of (exp) data do we need?

Opposition Team	Match 1	Match 2
MI		
CSK		
RR		

RCBD: Example

Do **Single Factor ANOVA** to find out

Opposition Team	Match 1	Match 2
MI		
CSK		
RR		

RCBD: Example

.. **But wait** …a sports analyst claims Virat's performance also depends on whether is he is playing at home (Bengaluru) or away! **Did you BLOCK the effect of that?**

Main Factor: Opposition team
(3 levels: MI vs CSK vs RR)

Block Factor: Location
(2 levels: Home vs Away)

What kind of (exp) data do we need?

Opposition Team	Match 1	Match 2
MI		
CSK		
RR		

RCBD: Example

Do **TWO Factor ANOVA** to find out significance of effects

Opposition Team	Match 1	Match 2
MI		
CSK		
RR		

$$SS_T = SS_{mean} + SS_1 + SS_2 + SS_{error} \qquad N = ab$$

DOF $\quad N \quad 1 \quad a-1 \quad b-1 \quad N-a-b+1$

$= (a-1)(b-1)$

Example 2

Consider a hardness testing experiment. Suppose we wish to determine whether or not **four different tips** produce different readings on a hardness testing machine.

The machine operates by pressing the tip into a metal test coupon, and from the depth of the resulting depression, the hardness of the coupon can be determined.

Let's say we want to obtain four observations for each tip.

Note that here is only one factor, i.e., 'tip type'. So a completely randomized single-factor design would consist of 4 x 4 = 16 experimental trials. We need 16 metal coupons. For each trial we can randomly use ANY ONE of the 16 metal coupons.

Do you see any potential problem in this experimental design?

Example 2

- **What if the metal coupons differ slightly in their hardness, as** might happen if they are taken from ingots that are produced in different heats? Then the coupons will contribute to the variability observed in the hardness data. (Serious Problem)
- **As a result, the experimental error will reflect both random error and variability between coupons.**

Randomized Complete Block Design for the Hardness Testing Experiment

Test Coupon (Block)			
1	**2**	**3**	**4**
Tip 3	Tip 3	Tip 2	Tip 1
Tip 1	Tip 4	Tip 1	Tip 4
Tip 4	Tip 2	Tip 3	Tip 2
Tip 2	Tip 1	Tip 4	Tip 3

- We would like to make the experimental error as small as possible; that is, we would like to remove the variability between coupons from the experimental error.
- A design that would accomplish this requires the experimenter to test each tip once on each of four coupons. This design is called a randomized complete block design (RCBD). The word "complete" indicates that each block (coupon) contains all the treatments (tips).
- Effectively, this design strategy improves the accuracy of the comparisons among tips by eliminating the variability among the coupons.
- **Note that within a block, the order in which the four tips are tested is randomly determined.**

Randomized Complete Block Design

- In many experimental situations, one would like to block the variability arising from extraneous sources. In this section, the principle of paired comparisons is extended to the comparison of more than two treatments (techniques), using randomized designs.
- In blocked designs two kinds of effects are studied:

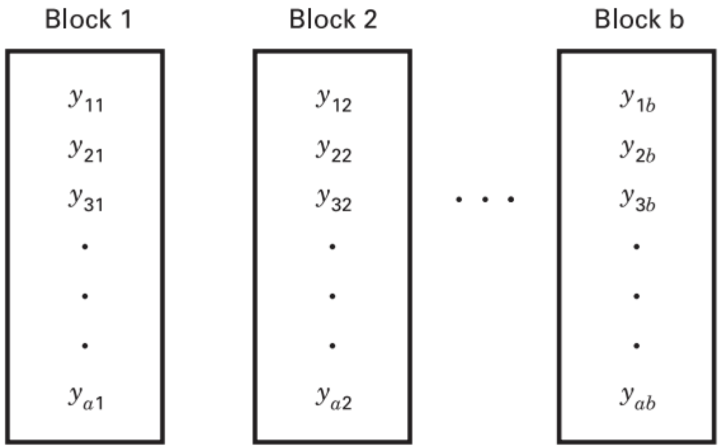
  1. Treatment effects, which are of major interest to the experimenter
  2. Block effects, which are desired to be eliminated.

-> Two- factor (treatment and blocks) ANOVA

RCBD: Statistical Model

In general, ‘**a**’ is the number treatments that are to be compared and ‘**b**’ is the number of blocks. There is **one observation per treatment in each block,** but the order in which the treatments are run *within each block* is determined randomly.

Because the only randomization of treatments is within the blocks, we often say that the blocks represent a restriction on randomization.



Effects Model for RCBD

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases}$$

where μ is an overall mean, τ_i is the effect of the ith treatment, β_j is the effect of the jth block, and ϵ_{ij} is the usual NID $(0, \sigma^2)$ random error term.

We usually think of the treatment and block effects as deviations from the overall mean so that

$$\sum_{i=1}^{a} \tau_i = 0 \quad \text{and} \quad \sum_{j=1}^{b} \beta_j = 0$$

RCBD: Statistical Model

Effects Model for RCBD

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases}$$

where μ is an overall mean, τ_i is the effect of the ith treatment, β_i is the effect of the jth block, and ϵ_{ij} is the usual NID $(0, \sigma^2)$ random error term.

$$\sum_{i=1}^{a} \tau_i = 0 \quad \text{and} \quad \sum_{j=1}^{b} \beta_j = 0$$

Block 1	Block 2		Block b
y_{11}	y_{12}		y_{1b}
y_{21}	y_{22}	. . .	y_{2b}
y_{31}	y_{32}		y_{3b}
$\vdots$	$\vdots$		$\vdots$
y_{a1}	y_{a2}		y_{ab}

What is the hypothesis test?

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1: \text{at least one } \mu_i \neq \mu_j$$

Treatment means

Because the ith treatment mean $\mu_i = (1/b)\Sigma_{j=1}^{b} (\mu + \tau_i + \beta_j) = \mu + \tau_i$, an equivalent way to write the above hypotheses is in terms of the treatment effects, say

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \text{ at least one } i$$

ANOVA for RCBD (Two factor ANOVA)

$N = ab$ be the total number of observations

Block 1	Block 2		Block b
y_{11}	y_{12}		y_{1b}
y_{21}	y_{22}		y_{2b}
y_{31}	y_{32}	$\cdots$	y_{3b}
$\vdots$	$\vdots$		$\vdots$
y_{a1}	y_{a2}		y_{ab}

Define sums,

$$y_{i.} = \sum_{j=1}^{b} y_{ij} \qquad i = 1, 2, \ldots, a$$

$$y_{.j} = \sum_{i=1}^{a} y_{ij} \qquad j = 1, 2, \ldots, b$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij} = \sum_{i=1}^{a} y_{i.} = \sum_{j=1}^{b} y_{.j}$$

Define avg. or means, $\bar{y}_{i.} = y_{i.}/b \quad \bar{y}_{.j} = y_{.j}/a \quad \bar{y}_{..} = y_{..}/N$

ANOVA for RCBD (Two Factor ANOVA)

Block 1	Block 2		Block b
y_{11}	y_{12}		y_{1b}
y_{21}	y_{22}		y_{2b}
y_{31}	y_{32}	$\cdots$	y_{3b}
$\vdots$	$\vdots$		$\vdots$
y_{a1}	y_{a2}		y_{ab}

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j}-\bar{y}_{i.}+\bar{y}_{..})^2$$

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

What are the degrees of freedom?

$N = ab$ be the total number of observations

F-tests

Therefore, to test the equality of treatment means, we would use the test statistic

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$$

reject H_0 if $F_0 > F_{\alpha,a-1,(a-1)(b-1)}$

We may also be interested in comparing block means because, if these means do not differ greatly, blocking may not be necessary in future experiments. From the expected mean squares, it seems that the hypothesis $H_0{:}\beta_j = 0$ may be tested by comparing the statistic $F_0 = MS_{\text{Blocks}}/MS_E$ to $F_{\alpha,b-1,(a-1)(b-1)}$. However, recall that randomization has been applied only to treatments *within* blocks; that is, the blocks represent a **restriction on randomization**. What effect does this have on the statistic $F_0 = MS_{\text{Blocks}}/MS_E$? Some differences in treat-

$F_0 = MS_{\text{Blocks}}/MS_E$ to $F_{\alpha,b-1,(a-1)(b-1)}$

ANOVA

Block 1	Block 2	...	Block b
y_{11}	y_{12}		y_{1b}
y_{21}	y_{22}		y_{2b}
y_{31}	y_{32}	$\cdots$	y_{3b}
$\vdots$	$\vdots$		$\vdots$
y_{a1}	y_{a2}		y_{ab}

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j}-\bar{y}_{i.}+\bar{y}_{..})^2$$

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_{\text{Treatments}} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$$

$$SS_{\text{Blocks}} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N}$$

Analysis of Variance for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$\frac{SS_{\text{Treatments}}}{a-1}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$\frac{SS_{\text{Blocks}}}{b-1}$	MS_{Blocks}/MS_E
Error	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$N-1$		

Example

THREE analysts each measures the melting point of a particular liquid with each of FOUR different thermometers,

	Thermometer			
Analyst	A	B	C	D
1	2.0	1.0	−0.5	1.5
2	1.0	0.0	−1.0	−1.0
3	1.5	1.0	1.0	0.5

We wish to know:

1. Are there significant differences among the analysts?
2. Are there significant differences among the thermometers?

Example

	Thermometer			
Analyst	A	B	C	D
1	2.0	1.0	−0.5	1.5
2	1.0	0.0	−1.0	−1.0
3	1.5	1.0	1.0	0.5

	Thermometer			
Analyst	A	B	C	D
1	2.0	1.0	−0.5	1.5
2	1.0	0.0	−1.0	−1.0
3	1.5	1.0	1.0	0.5

ANOVA Table

Source	SSQ	DoF	M.S.	F
Analyst	4.17	2	2.09	5.35
Thermometer	4.44	3	1.48	3.79
Grand mean	4.04	1		
Residual	2.35 (By subtraction)	6	0.39	
Total	15.00	12		

- $F_{0.95}(2, 6) = 5.143 < 5.35$
- $F_{0.95}(3, 6) = 4.757 > 3.79$
- We conclude that there are significant differences in analysts, but not in thermometers

Estimation of Missing Value

- When using the RCBD, sometimes an observation in one of the blocks is missing.
- This may happen because of carelessness or error or for reasons beyond our control, such as unavoidable damage to an experimental unit.
- A missing observation introduces a new problem into the analysis because treatments are no longer orthogonal to blocks; that is, every treatment does not occur in every block.
- There are two general approaches to the missing value problem.
  1. **Approximate Analysis**: Missing observation is estimated and the usual analysis of variance is performed *just as if the estimated observation were real data*, with the error degrees of freedom reduced by 1.
  2. **Exact Analysis:** We estimate the missing observation such that it's influence in the error estimation is minimum

Exact Analysis to Find Missing Value

Extrusion Pressures (PSI)	Batch of Resin (Block) 1	2	3	4	5	6	
8500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8700	92.5	89.5	90.6	x	87.0	95.8	455.4
8900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9100	82.5	89.5	85.6	87.4	78.9	90.7	514.6
Block totals	350.8	359.0	364.0	267.5	341.3	377.8	$y'_{..} = 2060.4$

so that x will have a minimum contribution to the error sum of squares. Because $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$, this is equivalent to choosing x to minimize

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{b}\sum_{i=1}^{a}\left(\sum_{j=1}^{b} y_{ij}\right)^2 - \frac{1}{a}\sum_{j=1}^{b}\left(\sum_{i=1}^{a} y_{ij}\right)^2 + \frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}\right)^2$$

or

$$SS_E = x^2 - \frac{1}{b}(y'_{i.} + x)^2 - \frac{1}{a}(y'_{.j} + x)^2 + \frac{1}{ab}(y'_{..} + x)^2 + R \qquad \textbf{(4.20)}$$

where R includes all terms not involving x. From $dSS_E/dx = 0$, we obtain

not involving x. From $dSS_E/dx = 0$, we obtain

$$x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}$$

Once you have the missing value, proceed with the usual ANOVA with ONE less DOF in error

Exact Analysis to Find Missing Value

Randomized Complete Block Design for the Vascular Graft Experiment with One Missing Value

Extrusion Pressures (PSI)	Batch of Resin (Block) 1	2	3	4	5	6	
8500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8700	92.5	89.5	90.6	x	87.0	95.8	455.4
8900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9100	82.5	89.5	85.6	87.4	78.9	90.7	514.6
Block totals	350.8	359.0	364.0	267.5	341.3	377.8	$y'_{..} = 2060.4$

$$x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}$$

$$x \equiv y_{24} = \frac{4(455.4) + 6(267.5) - 2060.4}{(3)(5)} = 91.08$$

Approximate Analysis of Variance for Example 4.1 with One Missing Value

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Extrusion pressure	166.14	3	55.38	7.63	0.0029
Batches of raw material	189.52	5	37.90		
Error	101.70	14	7.26		
Total	457.36	23			

where it can be accessed by others.