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		

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

$N = ab$

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.

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}

Effects Model for RCBD

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

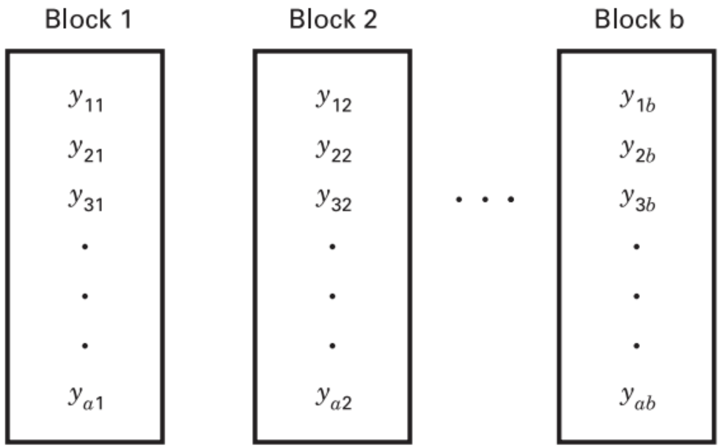
where μ is an overall mean, τ_i is the effect of the ith treatment, β_j is the effect of the jth block, and ϵ_{ij} is the usual NID $(0, \sigma^2)$ random error term.

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_j is the effect of the jth block, and ϵ_{ij} is the usual NID $(0, \sigma^2)$ random error term.

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

What is the hypothesis test?

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

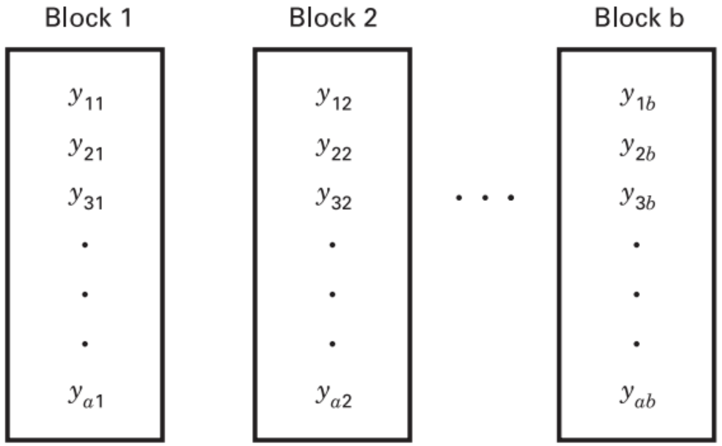
Treatment means

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



Define sums,

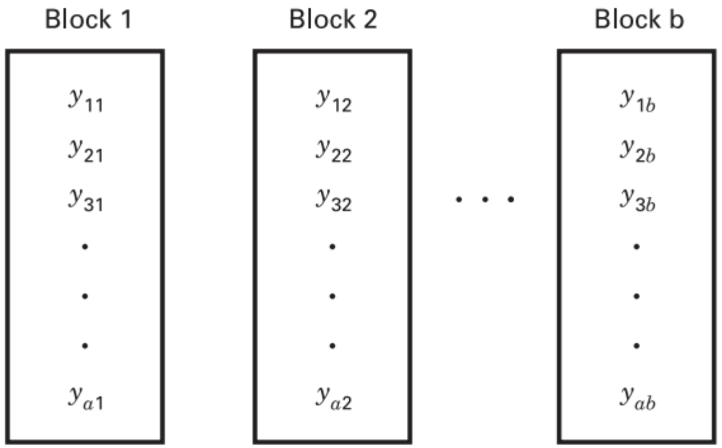
$$y_{i.} = \sum_{j=1}^{b} y_{ij} \qquad i = 1, 2, \ldots, a$$

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

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

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)



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

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

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

ny of the platforms where it can be accessed by others.

ANOVA

Block 1	Block 2		Block b
y_{11}	y_{12}		y_{1b}
y_{21}	y_{22}	$\cdots$	y_{2b}
y_{31}	y_{32}		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	$SS_{\text{Treatments}}$	$a-1$	$\dfrac{SS_{\text{Treatments}}}{a-1}$	$\dfrac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$\dfrac{SS_{\text{Blocks}}}{b-1}$	MS_{Blocks}/MS_E
Error	SS_E	$(a-1)(b-1)$	$\dfrac{SS_E}{(a-1)(b-1)}$	
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.

ANOVA: Interaction Effects

	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

- In this example, it is not unreasonable to expect that an analyst is accustomed to working with a particular type or brand of a thermometer. Therefore, his reading may be influenced by the thermometer he is using.
- Hence, there is a possibility that there may be an **interaction effect** between analysts and thermometers.
- The residual sum of squares we calculated could be the sum of squares due to the interaction effect plus the 'actual' residual sum of squares.
- In our previous problem there was no reason to suspect interaction between processes and batches. Therefore, the sum of squares due to interaction will be negligible and the residual can be calculated by subtraction.
- However, in this problem, when they are lumped together, it will decrease the F-ratios which are calculated. We might conclude that certain factors aren't significant when they really are. **What we need then is an estimate of the residual error.**
- How do we do this? By running more than one test under the same conditions, we can obtain an estimate of the residual error.

Interaction Effects

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

We can find column/row means after finding mean of repeated trials first

	Thermometer				
Analyst	A	B	C	D	Mean
1	2.0 1.75 1.5	1.0 1.25 1.5	−0.5 0.00 0.5	1.5 1.50 1.5	1.125
2	1.0 1.00 1.0	0.0 0.50 1.0	−0.5 -0.50 -0.5	−1.0 -0.50 0.0	0.125
3	1.5 1.25 1.0	1.0 1.25 1.5	1.0 1.00 1.0	0.5 0.75 1.0	1.063
Mean	1.333	1.000	0167	0.583	0.771

Mathematical Model

$$x_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

Observed Value — Grand mean — Effect of factor 1 — Effect of factor 2 — "Effect of Interaction" — Residual Error

here, $i = 1, 2, \ldots, a$ (number of factor 1) $j = 1, 2, \ldots, b$ (number of factor 2) $k = 1, 2, \ldots, r$ (number of replicates)

$$x_{ijk} = \bar{x}_{\ldots} + (\bar{x}_{i\cdot} - \bar{x}_{\ldots}) + (\bar{x}_{\cdot j} - \bar{x}_{\ldots}) + \underbrace{(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} - \bar{x}_{\ldots})}_{\text{"Interaction term"}} + (x_{ijk} - \bar{x}_{\ldots})$$

Mean Values

Grand Mean $= \bar{x}_{\cdots} = \dfrac{\sum_i \sum_j \sum_k x}{N}$ here, $N = abr$

Means of replicates $= \bar{x}_{ij\cdot} = \sum_k x_{ij} / r$

Means of factor 1 $= \bar{x}_{i\cdot} = \sum_j \bar{x}_{ij\cdot} / b$

Means of factor 2 $= \bar{x}_{\cdot j} = \sum_i \bar{x}_{ij\cdot} / a$

	Thermometer				
Analyst	A	B	C	D	Mean
1	2.0 1.75 1.5	1.0 1.25 1.5	−0.5 0.00 0.5	1.5 1.50 1.5	1.125
2	1.0 1.00 1.0	0.0 0.50 1.0	−0.5 -0.50 -0.5	−1.0 -0.50 0.0	0.125
3	1.5 1.25 1.0	1.0 1.25 1.5	1.0 1.00 1.0	0.5 0.75 1.0	1.063
Mean	1.333	1.000	0167	0.583	0.771

Sum of Squares

$SS_T = SS_M + SS_1 + SS_2 + SS_{interaction} + SS_{Error}$

found by subtraction

Total $SS_T = \sum_i \sum_j \sum_k x_{ijk}$

Grand Mean $SS_M = N \bar{x}_{...}^2 = abr\,\bar{x}_{...}^2$

Factor 1 $SS_1 = (rb) \sum_{i=1}^{a} (\bar{x}_{i.} - \bar{x}_{...})^2$

Factor 2 $SS_2 = (ra) \sum_{j=1}^{b} (\bar{x}_{.j} - \bar{x}_{...})^2$

Pure error $SS_E = \sum_i \sum_j \sum_{k=1}^{r} (x_{ijk} - \bar{x}_{ij.})$

ANOVA Table

$$SS_{analyst} = 2\times4[(1.125-0.771)^2+(0.125-0.771)^2+(1.063-0.771)^2]$$
$$=5.021$$

$$SS_{therm} = 6[(1.333-0.771)^2+(1.000-0.771)^2+(0.167-0.771)^2+(0.583-0.771)^2]$$
$$=4.615$$

$$SS_E = [(2.0-1.75)^2+(1.5-1.75)^2]+[(1.0-1.25)^2+(1.5-1.25)^2]+\cdots$$
$$+[(0.5-0.75)^2+(1.0-0.75)^2]$$

	Thermometer				
Analyst	A	B	C	D	Mean
1	2.0 1.75 1.5	1.0 1.25 1.5	−0.5 0.00 0.5	1.5 1.50 1.5	1.125
2	1.0 1.00 1.0	0.0 0.50 1.0	−0.5 -0.50 -0.5	−1.0 -0.50 0.0	0.125
3	1.5 1.25 1.0	1.0 1.25 1.5	1.0 1.00 1.0	0.5 0.75 1.0	1.063
Mean	1.333	1.000	0167	0.583	0.771

Source	SSQ	DoF	M.S.	*F*
Analyst	5.021	3-1=2	2.51	15.06
Thermometer	4.615	4-1=3	1.54	9.24
Interaction	3.354 (By subtraction)	6	0.56	3.354
Grand mean	14.26	1		
Pure Error	2.00	12	0.167	
Total	29.25	24		

- $F_{0.95}(2, 12) = 3.89$
- $F_{,0.95}(3, 12) = 3.49$
- $F_{0.95}(6, 12) = 3.00$

Conclusions

- Significant differences in analysts and thermometers
- Significant interaction effect