





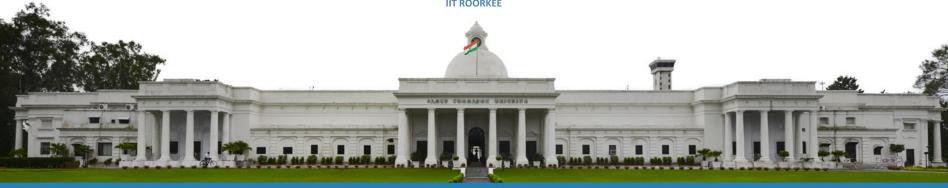


RBD

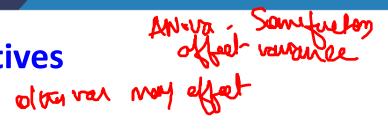


Dr. A. Ramesh

DEPARTMENT OF MANAGEMENT **IIT ROORKEE**



Everning Objectives



- Estimate variance components in an experiment involving random factors
- Understand the blocking principle and how it is used to isolate the effect of nuisance factors \(\)
- Design and conduct experiments involving the randomized complete block

design

ar Remone effort of other vor.
too juin > PED.

copon so bornant and po

Randomized Block Design

Blocking I

- A completely randomized design (CRD) is useful when the experimental units are homogeneous
- If the experimental units are heterogeneous, blocking is often used to form homogeneous groups



My RBD?

- A problem can arise whenever differences due to extraneous factors (ones not considered in the experiment) cause the MSE term in this ratio to become large.
- In such cases, the F value in equation can become small, signaling no difference among treatment means when in fact such a difference exists.

due to the descent of the descent o



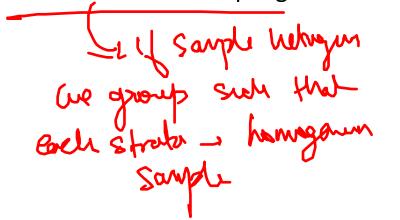




Randomized block design

- Experimental studies in business often involve experimental units that are highly heterogeneous; as a result, randomized block designs are often employed.
- Blocking in experimental design is similar to stratification in sampling.









Randomized block design

- Its purpose is to control some of the extraneous sources of variation by removing such variation from the MSE term.
- This design tends to provide a better estimate of the true error variance and leads to a more powerful hypothesis test in terms of the ability to detect differences among treatment means.







Air Traffic Controller Stress Test

- A study measuring the fatigue and stress of air traffic controllers resulted in proposals for modification and redesign of the controller's work station
- After consideration of several designs for the work station, three specific alternatives are selected as having the best potential for reducing controller stress
- The key question is: To what extent do the three alternatives differ in terms of their effect on controller stress?









Air Traffic Controller Stress Test

- In a completely randomized design, a random sample of controllers would be assigned to each work station alternative.
- However, controllers are believed to differ substantially in their ability to handle stressful situations.
- What is high stress to one controller might be only moderate or even low stress to another.
- Hence, when considering the within-group source of variation (MSE), we must realize that this variation includes both random error and error due to individual controller differences.
- In fact, managers expected controller variability to be a major contributor to the MSE term.







A randomized block design for the air traffic controller stress test

Treatments

	System A	System B	System C M
Controller 1	15	15	18 45
Controller 2	14 كون	14	14
Controller 3	14 V2 00	11	15
Controller 4	13	12	17
Controller 5	16	13	16
Controller 6	13	13	13

Blocks



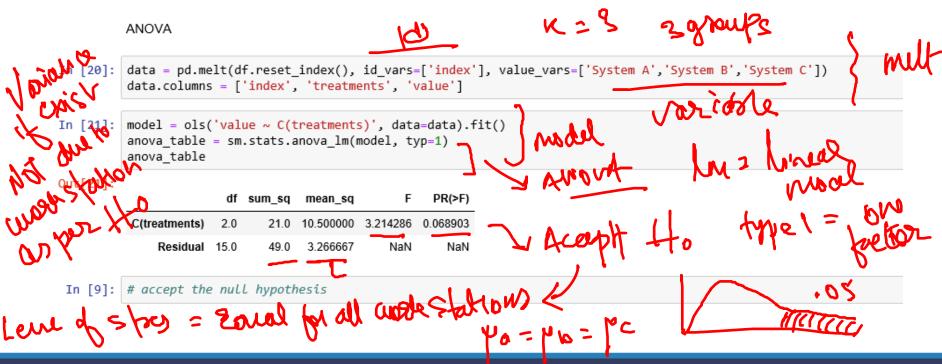
Solving this example using ANOVA in python

```
In [1]:
        import pandas as pd
                                               y state models
        import numpy as np
        import scipy
        import statsmodels.api as sm
        from statsmodels.formula.api import ols
In [4]: df = pd.read excel('RBD.xlsx')
Out[4]:
           System A System B System C
                15
                         15
                                 18
                14
                         14
                                 14
                                                      12
        2
                                 15
                10
                         11
         3
                13
                         12
                                 17
                16
                         13
                                 16
                                            39
                13
                         13
                                 13
```





Solving this example using ANOVA in python









Summary of stress data for the air traffic controller stress test

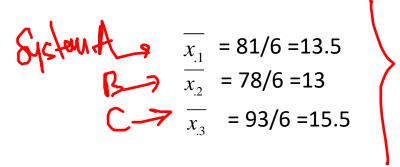
Treatments ⇒ Blocks ↓	System A	System B	System C	Block total	Block means
Controller 1	15	15	18	48	$\overline{x_{1.}}$ =16
Controller 2	14	14	14	42	$\overline{x_{2.}}$ =14
Controller 3	10	11	15	36	$\overline{x_{3.}}$ =12
Controller 4	13	12	17	42	$\overline{x_{4.}}$ =14
Controller 5	16	13	16	45	$\overline{x_{5.}}$ =15
Controller 6	13	13	13	39	${x_{6.}}$ =13
Column Total	81	78	93	252	$\frac{3}{x}$ 252/18 = 14





Summary of stress data for the air traffic controller stress test

Treatment means







ANOVA TABLE FOR THE RANDOMIZED BLOCK DESIGN WITH k (k-1)-(k-1)-(k-1)-(k-1)TREATMENTS AND b BLOCKS

	Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P- value
	Treatments	SS Treatments	k-1 Came	MS Treatments = SSTR/k-1	MS Treatmen	
)	Blocks	SS block	(b-1) V	MSBL = SSBL/b-	ts / MSE	
	Error	SSE	(k-1)(b-1)	MSE= SSE/(k-1)(b-1)		
	Total	SST	nt-1 Sawl			





RBD Problem

```
x_{ij} = \text{value of the observation corresponding to treatment } j \text{ in block } i
\bar{x}_{\cdot j} = \text{sample mean of the } j \text{th treatment}
\bar{x}_{i\cdot} = \text{sample mean for the } i \text{th block}
\bar{\bar{x}} = \text{overall sample mean}
```







RBD Problem

Step 1. Compute the total sum of squares (SST).

SST =
$$\sum_{i=1}^{b} \sum_{j=1}^{k} (x_{ij} - \bar{x})^2$$

Step 1. SST =
$$(15 - 14)^2 + (15 - 14)^2 + (18 - 14)^2 + \dots + (13 - 14)^2 = 70$$

Step 2. Compute the sum of squares due to treatments (SSTR).

SSTR =
$$b \sum_{j=1}^{k} (\bar{x}_{\cdot j} - \bar{\bar{x}})^2$$
 Same Sympletic Solution

Step 2. SSTR =
$$6[(13.5 - 14)^2 + (13.0 - 14)^2 + (15.5 - 14)^2] = 21$$





RBD Problem

Step 3. Compute the sum of squares due to blocks (SSBL).

$$SSBL = k \sum_{i=1}^{b} (\bar{x}_i - \bar{x})^2$$

Step 3. SSBL =
$$3[(16-14)^2 + (14-14)^2 + (12-14)^2 + (14-14)^2 + (15-14)^2 + (13-14)^2] = 30 - 55$$
 (BL) Lue to William

Step 4. Compute the sum of squares due to error (SSE).

$$SSE = SST - SSTR - SSBL$$

Step 4.
$$SSE = 70 - 21 - 30 = 19$$





ANOVA table for the air traffic controller stress test

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F .9	P- value
Treatments	21	2	10.5 .	10.5/1.9	0.024
Blocks	30	5	6.0	=5.53	
Error	19	10	1.9		
Total	70	17			

$$F_{.025} = 5.46$$
 and $F_{.01} = 7.56$.
Reject the null hypothesis







Solving RBD example using python

```
In [1]:
        import pandas as pd
         import numpy as np
         import scipy
         import statsmodels.api as sm
         from statsmodels.formula.api import ols
        df = pd.read_excel('RBD.xlsx')
Out[4]:
            System A System B System C
                 15
                          15
                                    18
                 14
                          14
                                    14
                 10
                          11
                                   15
                 13
                          12
                                   17
                 16
                          13
                                    16
                 13
                          13
                                    13
```







Solving RBD example using python

```
data = pd.melt(df.reset_index(), id_vars=['index'], value_vars=['System A','System B','System C'])
         data.columns = ['blocks', 'treatments', 'value']
         model = ols('value ~ C(block)+ C(treatments)', data=data).fit()
In [22]:
          anova table = sm.stats.anova lm(model, typ=1)
          anova table
Out[22]:
                                                      PR(>F)
                           sum sq mean sq
                                        6.0 3.157895 0.057399
           C(block)
                              30.0
          C(treatments)
                                       10.5 5.526316 0.024181
                       2.0
                              21.0
              Residual 10.0
                              19.0
                                        1.9
                                               NaN
                                                        NaN
In [23]: # reject the null hypothesis
```







Conclusion

- Finally, note that the ANOVA table shown in Table provides an F value to test for treatment effects but *not* for blocks.
- The reason is that the experiment was designed to test a single factor—work station design.
- The blocking based on individual stress differences was conducted to remove such variation from the MSE term.
- However, the study was not designed to test specifically for individual differences in stress.







- An experiment was performed to determine the effect of four different chemicals on the strength of a fabric.
- These chemicals are used as part of the permanent press finishing process.
- Five fabric samples were selected, and a randomized complete block design was run by testing each chemical type once in random order on each fabric sample.
- The data are shown in Table.
- We will test for differences in means using an ANOVA with alpha = 0.01.
 - -99







• Table: Fabric Strength Data—Randomized Complete Block Design

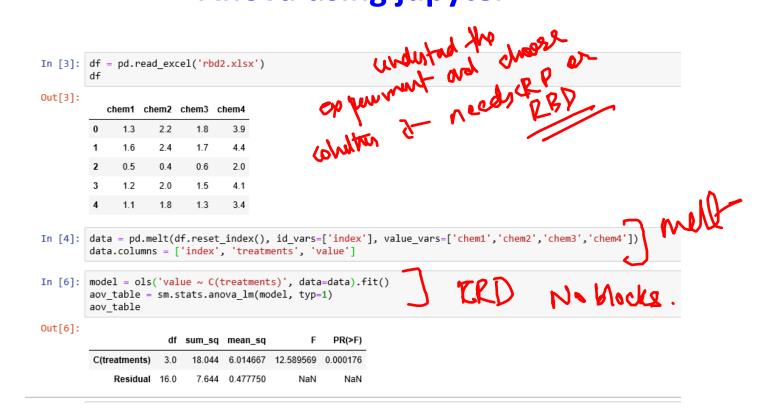
		Fabric Sample					Treatment Averages
Chemical Type	1	2	3	4	5	${\cal Y}_i.$	$ar{y}_i$.
1	1.3	1.6	0.5	1.2	1.1	5.7	1.14
2	2.2	2.4	0.4	2.0	1.8	8.8	1.76
3	1.8	1.7	0.6	1.5	1.3	6.9	1.38
4	3.9	4.4	2.0	4.1	3.4	17.8	3.56
Block totals y.,	9.2	10.1	3.5	8.8	7.6	39.2(y)	
Block averages $\overline{y}_{\cdot j}$	2.30	2.53	0.88	2.20	1.90		$1.96(\bar{y})$







Anova using jupyter









The sums of squares for the analysis of variance are computed as follows:

$$SS_{T} = \sum_{i=1}^{4} \sum_{j=1}^{5} y_{ij}^{2} - \frac{y_{..}^{2}}{ab}$$

$$= (1.3)^{2} + (1.6)^{2} + \dots + (3.4)^{2} - \frac{(39.2)^{2}}{20} = 25.69$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{4} \frac{y_{i}^{2}}{b} - \frac{y_{..}^{2}}{ab}$$

$$= \frac{(5.7)^{2} + (8.8)^{2} + (6.9)^{2} + (17.8)^{2}}{5} - \frac{(39.2)^{2}}{20} = 18.04$$







$$SS_{\text{Blocks}} = \sum_{j=1}^{5} \frac{y_{j}^{2}}{a} - \frac{y_{.}^{2}}{ab}$$

$$= \frac{(9.2)^{2} + (10.1)^{2} + (3.5)^{2} + (8.8)^{2} + (7.6)^{2}}{4} - \frac{(39.2)^{2}}{20} = 6.69$$

$$SS_{E} = SS_{T} - SS_{\text{Blocks}} - SS_{\text{Treatments}}$$

$$= 25.69 - 6.69 - 18.04 = 0.96$$

$$SS_{E} = SS_{T} - SS_{\text{Blocks}} - SS_{\text{Treatments}}$$

$$= 25.69 - 6.69 - 18.04 = 0.96$$







Analysis of Variance for the Randomized Complete Block Experiment

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F .08	P- value
Chemical types (Treatments)	18.04	3	6.01	75.13	4.79 E-8
Fabric samples (Blocks)	6.69	4	1.67	1 1/5	
Error	0.96	12	0.08 -> 2	ulu it	
Total	25.69	19	0		







Conclusion

- The ANOVA is summarized in the previous table
- Since $f_0 = 75.13 > f_{0.01,3,12} = 5.95$ (the *P*-value is 4.79 x E-8), we conclude that there is a significant difference in the chemical types so far as their effect on strength is concerned.









Python code for problem 2

```
In [2]:
        import pandas as pd
                                             Regular
(mjort data
         import statsmodels.api as sm
         from statsmodels.formula.api import ols
         from statsmodels.stats.anova import anova_lm
In [3]: df = pd.read_excel('RBD2.xlsx')
In [4]:
Out[4]:
            chem1 chem2 chem3 chem4
         0
               1.3
                      2.2
                            1.8
                                   3.9
                            1.7
               1.6
                      2.4
                                   4.4
                            0.6
                                   2.0
         2
               0.5
                      0.4
         3
               1.2
                      2.0
                            1.5
                                   4.1
                      1.8
                                   3.4
         4
               1.1
                            1.3
```







Python code for problem 2

Out[7]:

	Fabric sam	ples	Chemical types	value
0		0	chem1	1.3
1		1	chem1	1.6
2		2	chem1	0.5
3		3	chem1	1.2
4		4	chem1	1.1
5		0	chem2	2.2
6		1	chem2	2.4
7		2	chem2	0.4
8		3	chem2	2.0
9	L	4	chem2	1.8
10		0	chem3	1.8
11		1	chem3	1.7
12		2	chem3	0.6
13		3	chem3	1.5
14		4	chem3	1.3
15		0	chem4	3.9
16		1	chem4	4.4
17		2	chem4	2.0
18		3	chem4	4.1
40		A	aham4	2 4

data Propostion







Python code for problem 2

```
In [11]: model = ols('value ~ C(Fabric) + C(Chemical)', data=data).fit()
anova_table = sm.stats.anova_lm(model, typ=1)
anova_table
```

Out[11]:

	df	sum_sq	mean_sq	F	PR(>F)
C(Fabric)	4.0	6.693	1.673250	21.113565	2.318913e-05
C(Chemical)	3.0	18.044	6.014667	75.894848	4.518310e-08
Residual	12.0	0.951	0.079250	NaN	NaN



