IIT ROORKEE

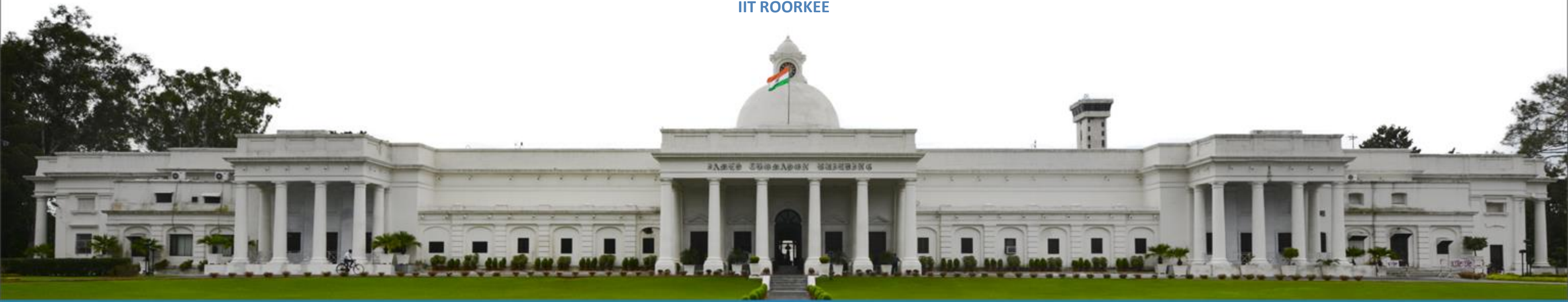
NPTEL ONLINE CERTIFICATION COURSE

Two Way ANOVA

Dr. A. Ramesh

DEPARTMENT OF MANAGEMENT

IIT ROORKEE



Fctorial design approach also called 2 way ANOVA

Learning objectives

- Design and conduct engineering experiments involving several factors using the factorial design approach
- Understand how the ANOVA is used to analyze the data from these experiments
- Know how to use the two-level series of factorial designs

Δ response → Prod. by Δ level of factor

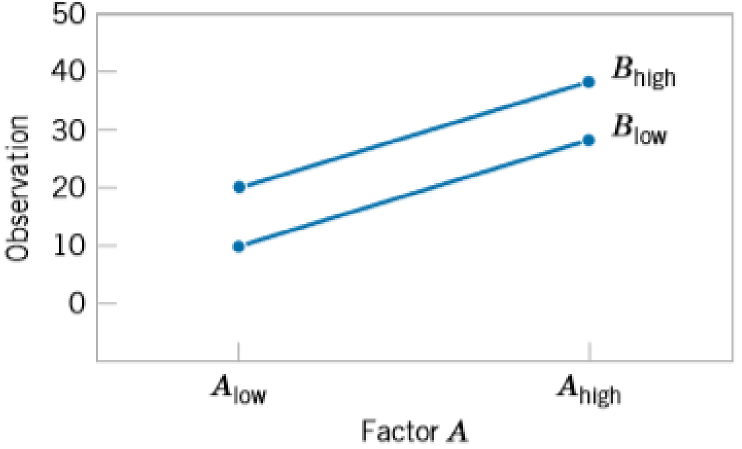
Factorial Experiment

effect of 2 var

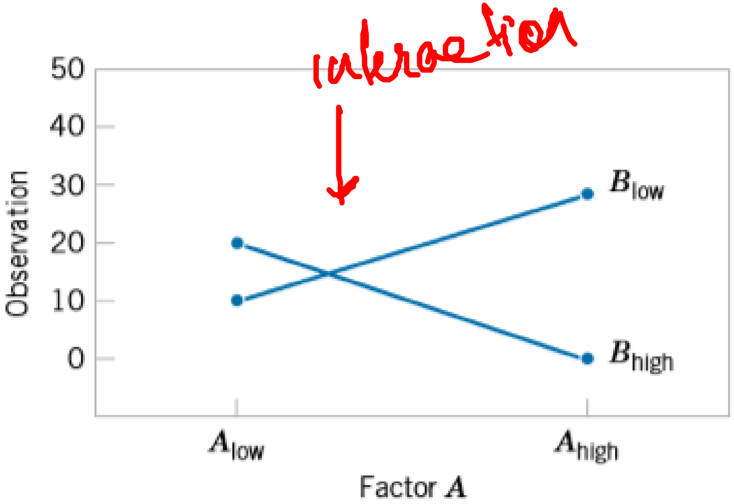
- A **factorial experiment** is an experimental design that allows simultaneous conclusions about two or more factors.
- The term factorial is used because the experimental conditions include all possible combinations of the factors.
- The effect of a factor is defined as the change in response produced by a change in the level of the factor. It is called a main effect because it refers to the primary factors in the study
- For example, for a levels of factor A and b levels of factor B, the experiment will involve collecting data on ab treatment combinations.
- Factorial experiments are the only way to discover interactions between variables.

main effect

Factorial Experiment



Factorial Experiment, no interaction



Factorial Experiment, with interaction

Two-factor Factorial Experiments

- The simplest type of factorial experiment involves only two factors, say, A and B.
- There are a levels of factor A and b levels of factor B.
- This two-factor factorial is shown in next table .
- The experiment has n replicates, and each replicate contains all ab treatment combinations.

Two-factor Factorial Experiments

2 factor is also a CRD.

a x b level of combinations

Data Arrangement for a Two-Factor Factorial Design

		Factor B				Totals	Averages
		1	2	...	b		
Factor A	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$	$y_{1\cdot\cdot}$	$\bar{y}_{1\cdot\cdot}$
	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$	$y_{2\cdot\cdot}$	$\bar{y}_{2\cdot\cdot}$
	$\vdots$						
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$	$y_{a\cdot\cdot}$	$\bar{y}_{a\cdot\cdot}$
Totals		$y_{\cdot 1\cdot}$	$y_{\cdot 2\cdot}$		$y_{\cdot b\cdot}$	$y_{\cdot\cdot\cdot}$	
Averages		$\bar{y}_{\cdot 1\cdot}$	$\bar{y}_{\cdot 2\cdot}$		$\bar{y}_{\cdot b\cdot}$		$\bar{y}_{\cdot\cdot\cdot}$

Two-factor Factorial Experiments

- The observation in the *ij*th cell for the *k*th replicate is denoted by y_{ijk}
- In performing the experiment, the *abn* observations would be run in random order.
- Thus, like the single factor experiment, the two-factor factorial is a completely randomized design.

Example

- As an illustration of a two-factor factorial experiment, we will consider a study involving the Common Admission test (CAT), a standardized test used by graduate schools of business to evaluate an applicant's ability to pursue a graduate program in that field.
- Scores on the CAT range from 200 to 800, with higher scores implying higher aptitude.

Three CAT preparation programs.

3 type

3 Prog

- In an attempt to improve students' performance on the CAT, a major university is considering offering the following three CAT preparation programs.

1. A three-hour review session covering the types of questions generally asked on the CAT.
2. A one-day program covering relevant exam material, along with the taking and grading of a sample exam.
3. An intensive 10-week course involving the identification of each student's weaknesses and the setting up of individualized programs for improvement.

Factor - 1 , 3 treatment

- One factor in this study is the CAT preparation program, which has three treatments:
  - Three-hour review,
  - One-day program, and
  - 10-week course.

1st prog has three learning method

- Before selecting the preparation program to adopt, further study will be conducted to determine how the proposed programs affect CAT scores.

Factor 2 : 3 Treatment

UG

- The CAT is usually taken by students from three colleges:
- the College of Business,
- the College of Engineering, and
- the College of Arts and Sciences.

Some IITs may not allow

Student Background Undergrad

- Therefore, a second factor of interest in the experiment is whether a student's undergraduate college affects the CAT score.
- This second factor, undergraduate college, also has three treatments:
  - Business,
  - Engineering, and
  - Arts and sciences.

UG college background

Nine Treatment Combinations for The Two-factor **CAT** Experiment

2nd factor
3

Factor A: Preparation Program	Factor B: College		
	Business	Engineering	Arts and sciences
Three-hour review	1	2	3
One-day program	4	5	6
10-Week course	7	8	9

1st factor
3

3x3 = 9
Possibilities

Replication

- In experimental design terminology, the sample size of two for each treatment combination indicates that we have two **replications**.

CAT SCORES FOR THE TWO-FACTOR EXPERIMENT

2 observation for each possibility is taken i.e $2 \times 9 = 18$ obs

Factor A: Preparation Program	Factor B: College		
	Business	Engineering	Arts and sciences
Three-hour review	500 580	540 460	480 400
One-day program	460 540	560 620	420 480
10-Week course	560 600	600 580	480 410

is there diff in score of people taking diff prog?

The analysis of variance computations answers the following questions.

- **Main effect (factor A):** Do the preparation programs differ in terms of effect on CAT scores?
- **Main effect (factor B):** Do the undergraduate colleges differ in terms of effect on CAT scores?
- **Interaction effect (factors A and B):** Do students in some colleges do better on one type of preparation program whereas others do better on a different type of preparation program?

is there a diff in CAT score card just bcoz of their background?

Interaction

both factors

- The term **interaction** refers to a new effect that we can now study because we used a factorial experiment.
- If the interaction effect has a significant impact on the CAT scores, we can conclude that the effect of the type of preparation program depends on the undergraduate college.

ANOVA Table for the Two-factor Factorial Experiment with *r* Replications

denominator = MSE

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P- value
Factor A	SSA	(a -1)	SSA/a-1	MSA / MSE	
Factor B	SSB	(b-1)	SSB/b-1	MSB/ MSE	
Interaction	SSAB	(a-1)(b-1)	MSAB = SSAB/(a-1)(b-1)	MSAB / MSE	
Error	SSE	ab(r-1)	MSE= SSE/(ab)(r-1)		
Total	SST	n_T-1			

Abbreviation

$$a = \text{number of levels of factor A}$$
$$b = \text{number of levels of factor B}$$
$$r = \text{number of replications}$$
$$n_T = \text{total number of observations taken in the experiment; } \boxed{n_T = abr}$$

no. of each possibility observation.

Eg:

$$n_T = a \times b \times r = 3 \times 3 \times 2 = 18$$

ANOVA Procedure

- The ANOVA procedure for the two-factor factorial experiment requires us to partition the sum of squares total (SST) into four groups:
  - sum of squares for factor A (SSA),
  - sum of squares for factor B (SSB),
  - sum of squares for interaction (SSAB), and
  - sum of squares due to error (SSE).
- The formula for this partitioning follows.

$$SST = \underbrace{SSA}_{A} + \underbrace{SSB}_{B} + \underbrace{SSAB}_{AB} + \underbrace{SSE}_{\text{Error}}$$

Computations and Conclusions

x_{ijk} = observation corresponding to the kth replicate taken from treatment i of factor A and treatment j of factor B

$\bar{x}_{i\cdot}$ = sample mean for the observations in treatment i (factor A)

$\bar{x}_{\cdot j}$ = sample mean for the observations in treatment j (factor B)

$\bar{x}_{ij}$ = sample mean for the observations corresponding to the combination of treatment i (factor A) and treatment j (factor B)

$\bar{\bar{x}}$ = overall sample mean of all n_T observations

A B

i

j

① Each possibility mean   for col 1   for col 2   ij. mean ( )

CAT Summary Data for The Two-factor Experiment

Factor A: Preparation Program	Factor B: College			Row totals
	Business	Engineering	Arts and sciences	(Sum Σ)
Three-hour review	500, 580, $\bar{x}_{11}=540$	540, 460, $\bar{x}_{12}=500$	480, 400, $\bar{x}_{13}=440$	2960
One-day program	460, 540, $\bar{x}_{21}=500$	560, 620, $\bar{x}_{22}=590$	420, 480, $\bar{x}_{23}=450$	3080
10-Week course	560, 600, $\bar{x}_{31}=580$	600, 580, $\bar{x}_{32}=590$	480, 410, $\bar{x}_{33}=445$	3230
Column totals	3240	3360	2670	Overall total= 9270; $\bar{\bar{x}} = 515$

CAT Summary Data for The Two-factor Experiment

- Factor A means

row 1 $\overline{x}_{1.} = 493.33$ (all colleges Prog 1)

row 2 $\overline{x}_{2.} = 513.33$ (All Colleges Prog 2)

row 3 $\overline{x}_{3.} = 538.33$

- Factor B means

$\overline{x}_{.1} = 540$ (All Prog Engg)

$\overline{x}_{.2} = 560$ (All Prog Business Eg)

$\overline{x}_{.3} = 445$ (All Prog A&S sci)

CAT Example:

Step 1. Compute the total sum of squares.

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

Step 1. $\text{SST} = (500 - 515)^2 + (580 - 515)^2 + (540 - 515)^2 + \cdots + (410 - 515)^2 = 82{,}450$

CAT Example:

515 = overall mean $\bar{\bar{x}}$

Step 2. Compute the sum of squares for factor A.

factor 1 levels

each possibility replication

$$SSA = br\sum_{i=1}^{a}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2$$

row1

row2

Step 2. $SSA = (3)(2)[(493.33 - 515)^2 + (513.33 - 515)^2 + (538.33 - 515)^2] = 6100$

row3 mean

CAT Example:

Step 3. Compute the sum of squares for factor B.

column mean

$$SSB = ar\sum_{j=1}^{b}(\bar{x}_{\cdot j} - \bar{\bar{x}})^2$$

Step 3. $SSB = (3)(2)[(540 - 515)^2 + (560 - 515)^2 + (445 - 515)^2] = 45{,}300$

factor 2 levels

replication

overall mean $\bar{\bar{x}}$

column 2, 3 mean

CAT Example:

observation mean value

factor B

Factor 1 P_1

mean

Step 4. Compute the sum of squares for interaction.

$$SSAB = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{\bar{x}})^2$$

r row col overall

Step 4. $SSAB = 2[(540 - 493.33 - 540 + 515)^2 + (500 - 493.33 - 560 + 515)^2 + \cdots + (445 - 538.33 - 445 + 515)^2] = 11{,}200$

CAT Example:

Step 5. Compute the sum of squares due to error.

$$SSE = SST - SSA - SSB - SSAB$$

Step 5. $SSE = 82{,}450 - 6100 - 45{,}300 - 11{,}200 = 19{,}850$

No affect different program

ANOVA Table for the **CAT** two-factor design

No effect → Interaction

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P- value
Factor A	6100	2	3050	1.38	0.299
Factor B	45300	2	22650	10.27	0.005
Interaction	11200	4	2800	1.27	0.350
Error	19850	9	2206		
Total	82450	17			

$\frac{6100}{2}$ $\frac{45300}{2}$ $\frac{11200}{4}$ $\frac{19850}{9}$

Accept H_0

Reject H_0

Accept H_0

Jupyter Code

```
In [15]: df2 = pd.read_excel('2way.xlsx')
```

```
In [16]: df2
```

Jupyter code

Out[16]:

	Value	prep_pro	college
0	500	three_hr	Business
1	580	three_hr	Business
2	540	three_hr	Engineering
3	460	three_hr	Engineering
4	480	three_hr	Artsandscience
5	400	three_hr	Artsandscience
6	460	One-day	Business
7	540	One-day	Business
8	560	One-day	Engineering
9	620	One-day	Engineering
10	420	One-day	Artsandscience
11	480	One-day	Artsandscience
12	560	10-Week	Business
13	600	10-Week	Business
14	600	10-Week	Engineering
15	580	10-Week	Engineering
16	480	10-Week	Artsandscience
17	410	10-Week	Artsandscience

→ Replication is 2 student per each possibility

Jupyter Code

In [20]:
```
formula = 'Value ~C(college)+C(prep_pro)+C(college):C(prep_pro)'
model = ols(formula, df2).fit()
aov_table = anova_lm(model, typ=2)


print(aov_table)
```

```
                          sum_sq   df          F    PR(>F)
C(college)               45300.0  2.0  10.269521  0.004757
C(prep_pro)               6100.0  2.0   1.382872  0.299436
C(college):C(prep_pro)   11200.0  4.0   1.269521  0.350328
Residual                 19850.0  9.0        NaN       NaN
```

Thank You