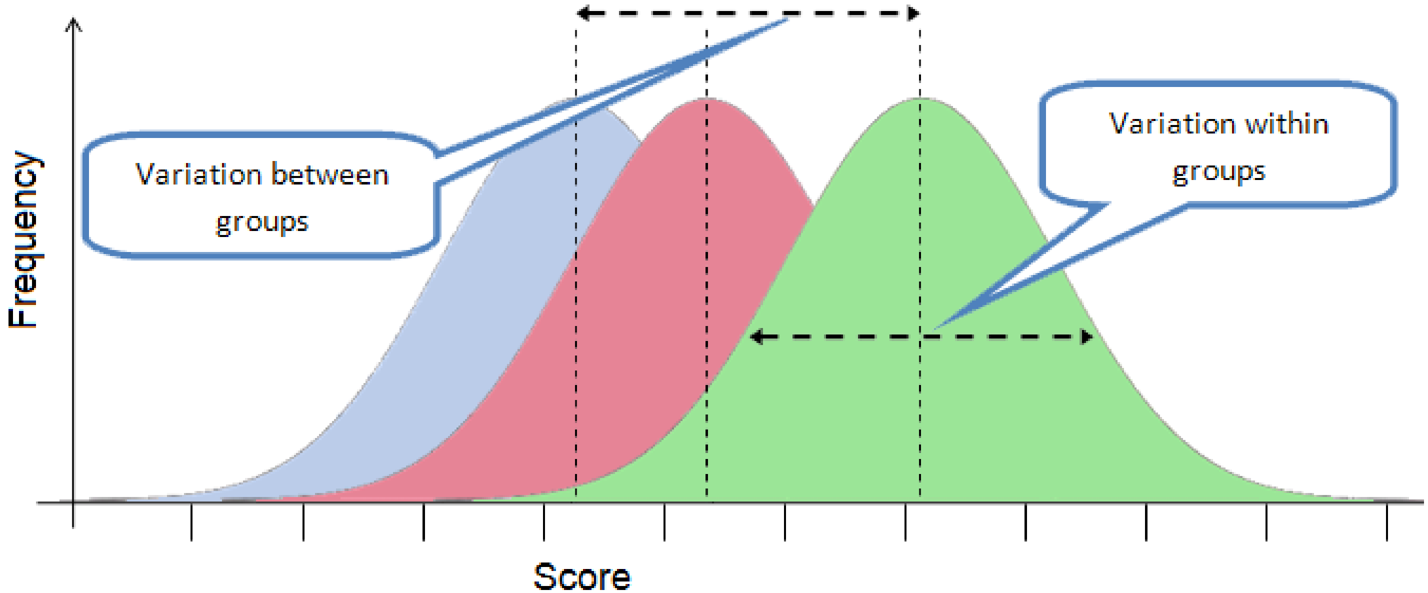
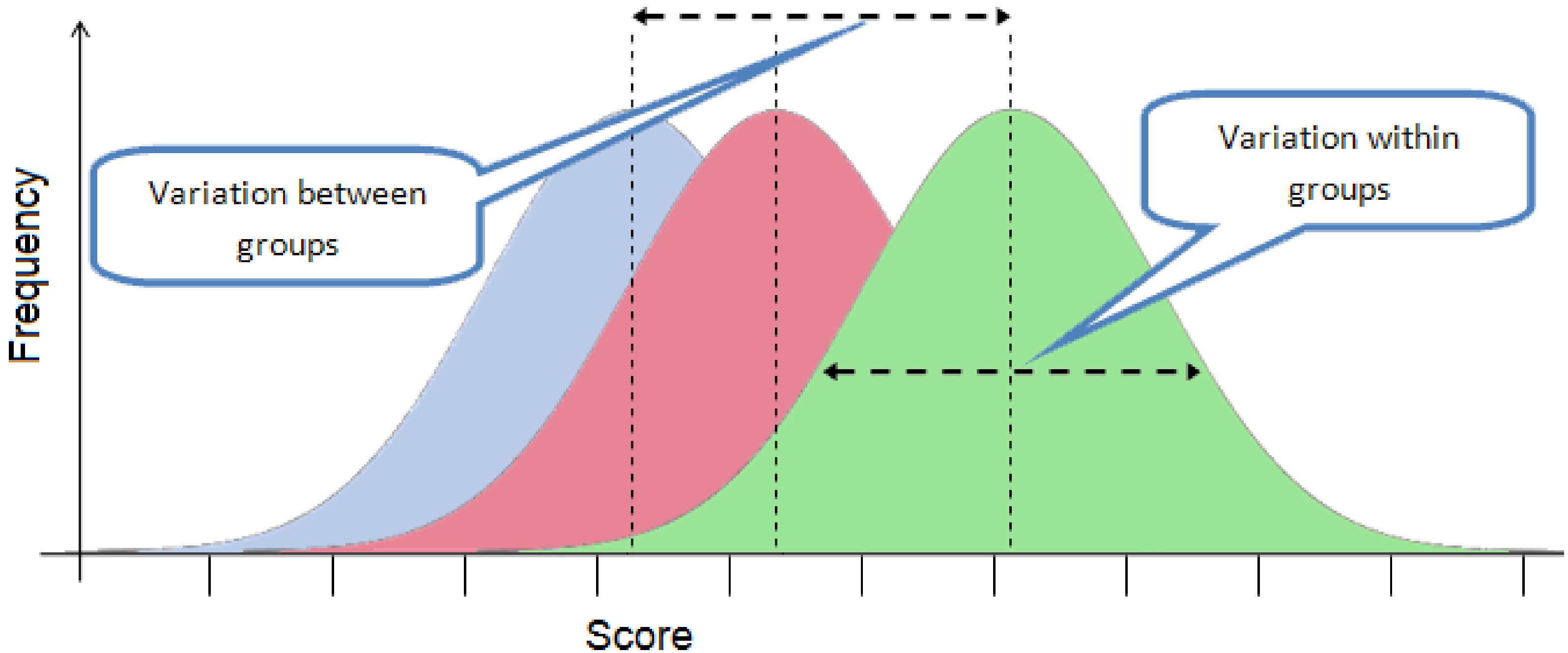
ANOVA
(Analysis of Variance)



ICT 242_2

Introduction
What is ANOVA?

Analysis of Variance (ANOVA) is a statistical technique used to determine whether there are significant differences between the means of three or more independent groups. It does this by analyzing the variances within and between groups.

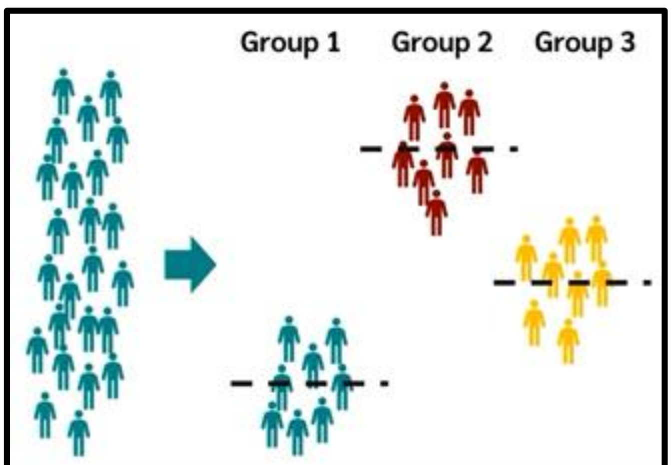


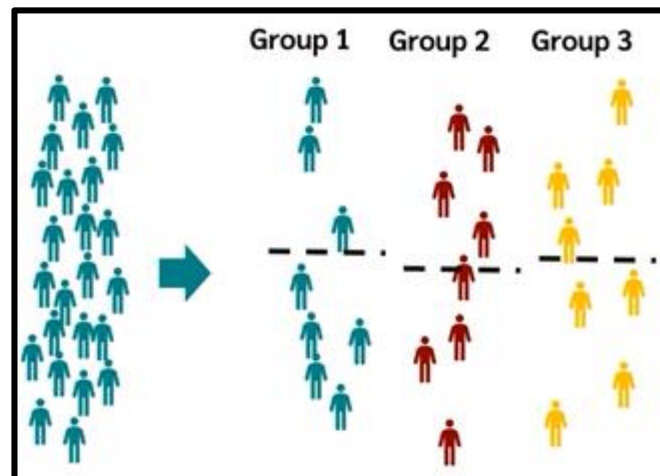
Introduction

Why Analyze Variance to Compare Means?

ANOVA compares variances to achieve the ultimate goal comparing means because:

- If group means are different, there will be more variability between groups than within groups.
- ANOVA partitions the total variability into between-group variance and within-group variance.
- A larger between-group variance relative to within-group variance suggests that the group means are significantly different.





Small ← Variance within the groups → Large

Large ← Variance between group means → Small

Types of ANOVA

Type	Description	Example
One-Way ANOVA	Tests for differences in means across **one factor** with multiple levels	Comparing stress test results across 4 materials
Two-Way ANOVA	Tests the effect of **two factors** and their interaction	Comparing strength across 3 materials and 2 curing methods
Repeated Measures ANOVA	Same subjects measured under different conditions	Measuring vibration response of the same bridge over time
MANOVA (Multivariate ANOVA)	Multiple dependent variables	Analyzing strength and ductility across treatments

One-way ANOVA vs Two-way ANOVA

A botanist wants to know whether or not plant growth is influenced by sunlight exposure and watering frequency

.

	Sunlight Exposure			
Watering Frequency	**None**	**Low**	**Medium**	**High**
Daily	4.8	5	6.4	6.3
	4.4	5.2	6.2	6.4
	3.2	5.6	4.7	5.6
	3.9	4.3	5.5	4.8
	4.4	4.8	5.8	5.8
Weekly	4.4	4.9	5.8	6
	4.2	5.3	6.2	4.9
	3.8	5.7	6.3	4.6
	3.7	5.4	6.5	5.6
	3.9	4.8	5.5	5.5

ONE-WAY ANOVA

One-Way Analysis of variance (ANOVA)

- One-Way ANOVA is a statistical method used to determine whether there are significant differences between the means of three or more independent groups based on one independent variable (or "factor").

- It is called "one-way" because the analysis considers only one factor that categorizes the data into groups.

Use One-Way ANOVA when there:

- One categorical independent variable (e.g., machine type, material, tool).
- One continuous dependent variable (e.g., strength, defect count, efficiency).
- To compare 3 or more group means.
- Independent groups, and data are roughly normally distributed with equal variances.

One-Way ANOVA Examples

1. Material Testing:

Comparing the tensile strength of different alloys (Alloy A, B, and C).

	Material (Independent Variable)		
Tensile Strength (MPa) (Dependent)	A	B	C
	450	470	490
	455	465	488
	448	472	495

Material (Independent Variable)	Tensile Strength (MPa) (Dependent Variable)
Alloy A	450
Alloy A	455
Alloy A	448
Alloy B	470
Alloy B	465
Alloy B	472
Alloy C	490
Alloy C	488
Alloy C	495

One-Way ANOVA Examples

2. Manufacturing Processes:

Evaluating the effect of different machine settings (Speed 1, Speed 2, Speed 3) on product defects.

	Machine Speed (Independent Variable)		
	Speed 1	Speed 2	Speed 3
Number of Defects (Dependent)	5	8	10
	6	9	11
	4	7	12

One-Way ANOVA Examples

3. Quality Control:

Comparing surface roughness resulting from different cutting tools.

	Tool Type (Independent Variable)		
	Tool A	Tool B	Tool C
Surface Roughness (μm)	3.2	2.8	2.5
	3.1	2.7	2.6
	3.3	2.9	2.4

One-Way ANOVA Examples

4. Civil Engineering:

Comparing concrete mix designs based on compressive strength.

	Mix Type (Independent)		
	Mix 1	Mix 2	Mix 3
Compressive Strength (MPa)	30	35	38
	32	36	39
	31	34	40

One-ay Analysis of variance (ANOVA)
Assumptions

Before applying one-way ANOVA, the following assumptions must be satisfied:

- **Independence**: Observations are independent within and across groups. (Independence of observations: the data were collected using statistically valid sampling methods, and there are no hidden relationships among observations.)

- **Normality**: The response variable is normally distributed within each group (Normally-distributed response variable: The values of the dependent variable follow a normal distribution.)

- **Homogeneity of Variance** (Homoscedasticity): Variances are equal across groups (Homogeneity of variance: The variation within each group being compared is similar for every group. If the variances are different among the groups, then ANOVA probably isn’t the right fit for the data).

Homogeneity of Variance (Homoscedasticity)

What Does "Equal Variance Across Groups" Mean?

- The assumption of homogeneity of variance means that the variability (spread) within each group (not between them) is approximately the same.

- So for testing three materials for tensile strength, it assumes that the variability of strength measurements within each material group (A, B, C) is similar—not necessarily that all groups have the same mean.

Why This Assumption “Homoscedasticity”, If We're Analyzing Variance?

ANOVA analyzes differences in group means using a ratio of two kinds of variance:

- Between-group variance (due to differences in group means)
- Within-group variance (random error or natural variability)

The F-statistic used in ANOVA is calculated as:

$$F = \frac{Variance\ Between\ Groups}{Variance\ Within\ Groups}$$

If the within-group variances are wildly different, the F-test becomes unreliable because:

- The denominator (within-group variance) no longer represents a consistent baseline.
- Groups with higher variability might inflate or mask differences in means.

So, for the test to be valid, we need to assume that the within-group variability is consistent otherwise, it causes incorrect conclusions.

How Do We Check This Assumption "Homoscedasticity"?

There are statistical tests and visual methods to verify homogeneity of variances:

- Levene's Test
- Bartlett's Test
- Boxplots: Visually inspect group spreads

Examples…

1) An experiment was done to determine whether four specific firing temperatures (100, 125, 150, 175, in Celsius) affect the density of a certain type of brick. Five specimens (bricks) at each level of firing temperature were tested.

2) Three brands of smartphone batteries are under study. It is suspected that the lives (in hours) of the three brands are different. Five randomly selected batteries of each brand are tested.

One-Way Analysis of Variance

- Suppose we have a treatments or different levels of a single factor that we wish to compare.

- The treatments are randomly assigned to a set of homogeneous experimental units.

- Where the treatments are assigned completely at random, so that each experimental unit has the same chance of receiving any one treatment.

Layout of the Design

- The diagram that shows the arrangement of treatments in experimental units is called as the layout of the design.
- Consider an experiment with 4 treatments and 5 replicates for each treatment:
  - Determine the total number of experimental units (n) as the number of treatments (t) and number of replications (r).

$$n = t * r = 4(5) = 20$$

  - We need n, 20 experimental units to carryout this experiment with 4 treatments and 5 replicates.

Layout of the Design…

- The 20 units are numbered as follows, Layout of the design by using CRD Experiment (before randomized).

E1	E2	E3	E4	E5
E6	E7	E8	E9	E10
E11	E12	E13	E14	E15
E16	E17	E18	E19	E20

1(T1)	2(T1)	3(T1)	4(T1)	5(T1)
6(T2)	7(T2)	8(T2)	9(T2)	10(T2)
11(T3)	12(T3)	13(T3)	14(T3)	15(T3)
16(T4)	17(T4)	18(T4)	19(T4)	20(T4)

- Assign the treatments to the experimental units by randomly,

8(T2) E1	12(T3) E2	16(T4) E3	1(T1) E4	20(T4) E5
3(T1) E6	11(T3) E7	19(T4) E8	15(T3) E9	14(T3) E10
6(T2) E11	18(T4) E12	13(T3) E13	9(T2) E14	4(T1) E15
10(T2) E16	17(T4) E17	2(T1) E18	5(T1) E19	7(T2) E20

What does this test do?

- The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other.

Specifically, it tests the null hypothesis:

Null Hypothesis (H_0): All group means are equal

$$H0\colon \mu_1 = \mu_2 = \cdots = \mu_k$$

Alternative Hypothesis (H_1): At least one group mean is different

$$H1\colon At\ least\ one \mu_i \neq \mu_j$$

Hypotheses to be tested…

- In the means model: We are interested to testing the equality of the t treatment means

$$\mathrm{H_o}: \mu_1 = \mu_2 = \cdots = \mu_t$$

$$\mathrm{H_A}: \mu_i \neq \mu_j \text{ for at least one pair } (i, j)$$

- In the effects model: We are interested to testing that the treatment effects are zero (There is no treatments/ factor effects)

$$H_o: \boldsymbol{\alpha}_i = 0 \; for \; \forall \; i$$

$$H_A: \boldsymbol{\alpha}_i \neq 0 \text{ for at least one } i$$

✓ **The null hypothesis will be that all population means are equal**

✓ **The alternative hypothesis is that at least one mean is different.**

Models for the Data

- The statistical model of the CRD for the single factor experiment with t treatments and r replicates, is of the form (Total number of observations, $\boldsymbol{n} = \boldsymbol{tr}$).

  – Means Model

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad ; i = 1, 2, \dots, \mathrm{s} : j = 1, 2, \dots, \mathrm{n}$$

$$\mu_i = \mu + \boldsymbol{\alpha}_i : i = 1, 2, \dots, \mathrm{s}$$

  – Effect Model

$$y_{ij} = \mu + \boldsymbol{\alpha}_i + \varepsilon_{ij} \qquad ; i = 1, 2, \dots, \mathrm{s} \quad : j = 1, 2, \dots, \mathrm{n}$$

Where,

$\boldsymbol{y_{ij}}$ = *the j^{th} observation on the i^{th} treatment*

$\boldsymbol{\mu}$ = *grand (overall) mean*

$\boldsymbol{\alpha}_i$ = effect of i^{th} treatment

$\boldsymbol{\varepsilon_{ij}}$ = random error

Model Assumptions

1. Errors are normally distributed with mean zero and variance (σ^2),

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

2. We shall assume that random variable y_{ij} which are all independent, have the normal distribution with mean μ_i and common variance σ^2 .

- This implies that the observations;

$$y_{ij} \sim N(\mu_i, \sigma^2)$$

3. The observations are mutually independent.

ε_{ij} are independent.

ANOVA Table for CRD of a Single Factor Experiment: One Way ANOVA Table

Source Of Variance	Sum of Square	Degree of the Freedom	Mean of Square	F-Value
Treatment	SS_{Tr}	S-1	$\frac{SS_{Tr}}{s-1} = \alpha$	$\frac{\alpha}{\beta} = f_0$
Error	SS_E	N-S	$\frac{SS_E}{N-s} = \beta$	
Total	SS_T	N-1		

1 less than number of groups

1 less than number of individuals (just like other situations)

Notations

$\xrightarrow{Tretment(i)}$

1	2	3	.	.	.	s	
y_{11}	y_{21}	y_{31}	.	.	.	y_{s1}	
y_{12}	y_{22}	y_{32}	.	.	.	y_{s2}	
y_{13}	y_{23}	y_{33}	.	.	.	y_{s3}	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
y_{1n_1}	y_{2n_2}	y_{3n_3}	.	.	.	y_{sn_k}	
$T_{1.}$	$T_{2.}$	$T_{3.}$				$T_{s.}$	$T_{..}$

- $T_{i.} = \ ith\ Treatment\ Total = \sum_{j=1}^{n_i} y_{ij}$
- $T_{..} = Grand\ Total = \sum_{i=1}^{s} \sum_{j=1}^{n_i} y_{ij}$
- $\bar{T}_{..} = Grand\ Total\ Mean = \dfrac{\sum_{i=1}^{s} \sum_{j=1}^{n_i} y_{ij}}{N}$, $N = \sum_{i=1}^{s} n_i$

SS_T - Total Sum of Squares

$$SS_T = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(y_{ij} - \overline{y_{..}})^2$$

$$= \sum_{i=1}^{s}\sum_{j=1}^{n_i} y_{ij}^{\,2} - \frac{y_{..}^{\,2}}{N}$$

SS_{Tr} - Total Variance due to treatment

- Treatment Sum of Squares

$$SS_{Tr} = \sum_{i=1}^{s}\sum_{j=1}^{n_i}\left(\overline{T}_{i.} - \overline{T}_{..}\right)^2$$

$$= \sum_{i=1}^{s}\frac{T_{i.}^{\,2}}{n_i} - \frac{T_{..}^{\,2}}{N}$$

$$\boxed{SS_T = SS_{Tr} + SS_E}$$

$\boxed{SS_E - Error\ sum\ of\ square}$

One Way ANOVA Table

Source Of Variance	Sum of Square	Degree of the Freedom	Mean of Square	F-Value
Treatment	SS_{Tr}	S-1	$\frac{SS_{Tr}}{s-1} = \alpha$	$\frac{\alpha}{\beta} = f_0$
Error	SS_E	N-S	$\frac{SS_E}{N-s} = \beta$	
Total	SS_T	N-1		

Decision Rule

- Compute the F-statistic.
- Compare to the critical value from F-distribution table at significance level αα.
- If $F_{calc} > F_{critical}$, reject H_0.

Example:

A mechanical engineer tests 3 types of lubricants (A, B, and C) to see which reduces friction the most. Each type is tested 5 times, and the friction force is recorded (in N).

Lubricant A	Lubricant B	Lubricant C
12	15	20
14	17	19
13	16	22
15	14	21
11	18	23

Steps:

1. Compute group means and overall mean.
2. Calculate SSB, SSW, SST.
3. Fill the ANOVA table.
4. Compute F-ratio.
5. Compare with F-critical (from F-table with df1 = 2, df2 = 12).

Solution:

Lubricant A	Lubricant B	Lubricant C
12	15	20
14	17	19
13	16	22
15	14	21
11	18	23

One-way ANOVA: Lubricant A, Lubricant B, Lubricant C

Source	DF	SS	MS	F	P
Factor	2	163.33	81.67	32.67	0.000
Error	12	30.00	2.50		
Total	14	193.33			

S = 1.581   R-Sq = 84.48%   R-Sq(adj) = 81.90%

Example (01)

Imagine that you manufacture paper bags and that you want to improve the tensile strength of the bag. You suspect that changing the concentration of hardwood in the bag will change the tensile strength. You measure the tensile strength in pounds per square inch (PSI). So, you decide to test this at 5%, 10%, 15% and 20% hardwood concentration levels. These "levels" are also called "treatments“. Stating your assumptions, construct the One way analysis of variance (ANOVA) table and, obtain the relevant conclusions at 0.05 significance level.

Hardwood Concentration (%)			
A(5%)	B (10%)	C (15%)	D (20%)
12	18	15	24
14	12	18	25
15	17	10	24
11	13	19	15
10	18	16	13
	19	18	
	15		

Example

An experiment was conducted to study the performance of five tomato varieties in a greenhouse using CRD. Tomato varieties were assigned to experimental units (plots) randomly. The experiment contained 25 plots and the yield values recorded from each plot are given in the table below.

Variety	Yield(kg/plot)				
1	1.6	1.9	1.5	2.1	1.7
2	1.9	2.4	2.3	2.2	2.1
3	2.3	2.6	2.4	2.5	2.7
4	1.2	0.8	1.0	0.9	0.8
5	2.6	2.8	3.0	2.9	2.8

Table : Tomato yield (kg/plot) recorded from 25 plots

Example

The data in the following table represent the number of hours of relief provided by 4 different brands of headache tablets administered to 22 subjects experiencing fevers of or more.

Tablet			
A	B	C	D
5.2	**9.1**	**3.2**	**2.4**
4.7	**7.1**	**5.8**	**3.4**
8.1	**8.2**	**2.2**	**4.1**
6.2	**6.0**	**3.1**	**1.0**
	9.1	**7.2**	**4.0**
		6.2	**7.6**
		3.4	

Perform the analysis of variance and test the hypothesis at the 0.1 level of significance that the mean number of hours of relief provided by the tablets is the same for all 4 brands. Discuss the results.