Definition of ANOVA:

According to R.A. Fisher, Analysis of Variance (ANOVA) is the "Seperation of Variance Ascribable to one group of causes from the variance ascribable to other group." By this technique the total variation in the sample data is expressed as the sum of its non-negative components where each of these components is a measure of the variation due to some specific independent source or factor or cause.

The Analysis of Variance, popularly known as ANOVA, can be used in cases where there are more than 2 groups.

Assumptions:

For the validity of the F test in ANOVA, the following assumptions are made:

(1) The observations are independent

(2) Parent popl^n from which observations are taken is normal &

(3) Various treatment and environmental effects are additive in nature.

Errors

Variation is inherent in nature. The total variation in any set of numerical data is due to the no: of causes which are classified as: (1) Assignable causes → Detected & rectified

(2) Chance causes → beyond the human control and cannot be traced seperately

Objectives of ANOVA:

1. It identifies the causes of variation and sort out corresponding components of variation with associated degrees of freedom.

2. It provides test of significance based on F-test.

H_0:

H_1: cal $F <$ tab $F_{(n-1, m-1, \alpha)}$

Accept H_0

cal $F >$ tab $F_{(n-1, m-1, \alpha)}$

Reject H_0

① One way ANOVA:

Two groups → based on one factor (independent variable)

One-way ANOVA → compares means betⁿ the groups that you are interested in and determines whether any of those means differ significantly from each other:

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_K$; $H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_K$

where μ_i = i^{th} group mean, K = no: of groups. If, however, the one-way ANOVA returns (gives) a statistically significant result, we accept the alternative hypothesis (H_1) which means that there are at least two group means which are significantly different from each other.

Mathematical model:

Let there are N units and k treatments. These treatments are applied such that i^{th} treatment is given to n_i units. So

$N = \sum n_i$, $i = 1, 2 \ldots k$.

$$x_{ij} = \mu_i + \varepsilon_{ij} \quad ; \quad i = 1, 2 \ldots k \; ; \; j = 1, 2 \ldots, n_i$$

$$\mu_i = \mu + \gamma_i$$

$$x_{ij} = \mu + \gamma_i + \varepsilon_{ij} \; ; \quad i = 1, 2 \ldots k \; ; \; k = 1, 2 \ldots, n_i$$

μ = common parameter for all treatments

γ_i = i^{th} treatment effect

ε_{ij} = random error, $\varepsilon_{ij} \sim N(0, \sigma^2)$

ε_{ij} = random error, $\varepsilon_{ij} \sim N(0, \sigma^2)$

$\Rightarrow x_{ij} \sim N(\mu+\gamma_i, \sigma^2)$

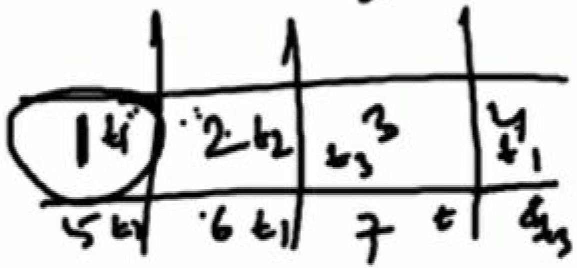
$$\boxed{x_{ij} = \mu + \gamma_i + \varepsilon_{ij}}$$

The total variation in the observation x_{ij} can be split into the following two components:

1) The variation betn classes or the variation due to different bases of classification known as treatments.
2) The variation within the classes, i.e. the inherent variation of the random variable within the observation of a class.

→ Assignable causes

→ Chance causes



C_1	C_2	C_3		. – –	C_n
[illegible]	$\cancel{t_1}$		$\cancel{t}_2$ –		t_K

$n \rightarrow$ no: of cows.
$K \rightarrow$ no: of treatment (food of the cows).

1. Effect of the treatments (rations / food) $\rightarrow$ Assignable
2. Error due to chance causes produced by numerous causes that they are not detected and identified.

Test producerse:

1) $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ or $H_0: \tau_1 = \tau_2 = \dots = \tau_k$

against
$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_k$ or $H_0: \tau_1 \neq \tau_2 \neq \dots \neq \tau_k$

$\Rightarrow H_1: \mu_i \neq \mu_j$ for at least one pair of (i,j)
$(i = 1, 2 \dots k)$

(2) L.O.S. α.

(3) Test statistic:

(a) Grand total $= G = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}$

(b) Correction factor (C.F) $= \frac{G^2}{N}$

$\sum\sum x_{ij} = 101571$
$N = 100$
$G^2 = \frac{(101571)^2}{100}$

(c) TSS (total sum of squares) $= \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 - CF$

(d) SST (Sum of squares due to treatments) $= \dfrac{\sum T_i^2}{n_i} - CF$

x_{ij}, $i = 1 \ldots k$, $j = 1 \ldots n_i$

i : treatment = 3

j : 4

$1^2 + 2^2 + 5^2 + 7^2$

$+ 9^2 + 8^2 \ldots + 2^2$

$\dfrac{25^2 + 39^2 + 33^2}{4}$

					Total
T_1	1	2	5	7	$\sum T_1 = 25$
T_2	9	8	1	7	$\sum T_2 =$ [illegible]
T_3	2	3	6	2	$\sum T_3 =$ [illegible]

where n_i ($i = 1, 2 \ldots k$) is the no: of observations received i^{th} treatment.

f) SSE (sum of squares due to errors)

= TSS − SST → sum of squares due to treatment.

↓

Total SS

4) D.f: TSS → $(N-1)$ df.

SST → $(K-1)$ df

SSE → $(N-K)$ df. → $= N-K$

$TSS - SST = SSE$

$(N-1) - (K-1)$

$= N \cancel{-1} - K \cancel{+1}$

$= N - K$

(5) Mean Sum of Squares : MSE $= \frac{SS}{df}$

Mean sum of squares for treatments is $MST = \frac{SST}{K-1}$

Mean sum of squares for error is $MSE = \frac{SSE}{N-K}$

(6) $F = \frac{MST}{MSE}$

(7) tab $F(K-1, N-K, \alpha)$

Calculation of variance ratio $F = \frac{MST}{MSE}$.

Critical value of F or table value of F for $(K-1, N-K)$ df at α% level of significance

$\boxed{F_{(3,7,0.05)}}$

(7 Inference.

If cal $F <$ tab $F_{(K-1, N-1, \alpha)}$, we accept H_0 and say that there is no significant difference between the treatments. If cal $F >$ tab F, we reject H_0 and say that there is difference bet^n^ the treatments or the difference bet^n^ treatments is significant.

ANOVA Table for one way classification

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean sum of squares $MSS = \frac{SS}{df}$	F-calculated	FCV (critical value)
Between Treatments	SST	$K-1$	$MST = \frac{SST}{K-1}$	$F = \frac{MST}{MSE}$	tab $F(K-1, N-K)$
Error	$TSS - SST = SSE$	$N-K$	$MSE = \frac{SSE}{N-K}$		
Total	TSS	$N-1$			

(2) ANOVA is an omnibus test statistic and cannot tell us which specific groups are statisticelly significantly different from each other.

Post hoc test

T_1	T_2	T_3
T_3	T_1	T_2
T_2	T_1	T_3

$F_{cal} > F_{tab}$

reject H_0 accept H_1

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

$H_1 : \mu_1 \neq \mu_2 \neq \dots \neq \mu_k$

$H_1 : \mu_i \neq \mu_j$

Ex: Three processes A, B and C are tested to see whether their outputs are equivalent. The following observations of outputs are made.

A	20	22	23	21	20	24	25	23
B	19	21	20	22	23			
C	21	20	20	25	24	22	23	

Carry out the ANOVA and state your conclusion.

Solⁿ: Null hypo: $H_0 : \mu_1 = \mu_2 = \mu_3$; i.e. there is no significant difference between the three processes

Alternative hypo: $H_1 = \mu_1 \neq \mu_2 \neq \mu_3$

Level of significance: $\alpha = 0.05$ (5% l.o.s)

Ex: treatments (K)

Process	Observations								Total
A	20	22	23	21	20	24	25	23	178
B	19	21	20	22	23				105
C	21	20	20	25	24	22	23		135

Carry out the ANOVA and state your conclusion.

Solⁿ: Null hypo: $H_0 : \mu_1 = \mu_2 = \mu_3$; i.e. there is no significant difference between the three processes

Alternative hypo: $H_1 = \mu_1 \neq \mu_2 \neq \mu_3$

Level of significance: $\alpha = 0.05$ (5% l.o.s)

Let us construct the following table:

Here, $K = 3$,

$n_1 = 8, n_2 = 5, n_3 = 7$

$N = \sum n_i = 8+5+7$
$= 20$

Group totals: $T_1 = 178$, $T_2 = 105$, $T_3 = 135$

$\therefore G = 418$

$$C.F = \frac{G^2}{N} = \frac{(418)^2}{19} = 9196$$

							Total
C	21	20	25	24	22	23	135
A							178
B							105

$n_3 = 6$

$N = 19$

$K = 3$

$$TSS = \sum_i \sum_j x_{ij}^2 - CF$$

$$= (20^2 + 22^2 + \dots + 23^2) - 9196$$

$$= 9254 - 9196 = 58 \text{ with } (N-1)\,df = (19-1) = 18\,df$$

$K-1$
$= 3-1$
$= 2$

$$SST = \sum \frac{T_i^2}{n_i} - CF = \frac{135^2 + 178^2 + 105^2}{3} - 9196 = 9203 - 9196$$

$= 7$ with $(k-1)\,df$

$SST = 7$ with $\underline{2df}$

$SSE = TSS - SST = 58 - 7 = 51$ with $(N-K)$ df

$= 19 - 3 = 16$ df.

$\therefore SSE = 51$ with $\underline{16}$ df.

$$MST = \frac{SST}{df} = \frac{7}{2} = 3.5$$

$$MSE = \frac{SSE}{df} = \frac{51}{16} = 3.1875$$

ANOVA table for one-way classification.

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Sum of Squares $MSS = \frac{SS}{df}$	Fcal	FCV (critical value)
Between Processes	SST = 7	K-1 = 3-1 = 2 ✓	$MST = \frac{SST}{K-1} = \frac{7}{2} = 3.5$	$F = \frac{MST}{MSE} = \frac{3.5}{3.1875} = 1.098$	F(2,16,0.05) = 3.63
Error	SSE = 51	N-K = 19-3 = 16 ✓	$MSE = \frac{SSE}{16} = \frac{51}{16} = 3.1875$		
Total	58	18			

Decision Rule:

Since F calculated is less than tab value of F at 5% level of significance, we accept the null hypothesis H_0 and conclude that there is no significant difference between the processes.

$$F_{cal} < F_{tab}$$

$$1.098 < 3.63$$

ACCEPT H_0.

ANOVA table for one-way classification.

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Sum of Squares $MSS = \frac{SS}{df}$	Fcal	FCV (critical value)
Between Processes	$SST = 7$	$K-1 = 3-1 = 2$ ✓	$MST = \frac{SST}{K-1} = \frac{7}{2} = 3.5$	$F = \frac{MST}{MSE} = \frac{3.5}{3.1875} = 1.098$	$F(2, 16, 0.05) = 3.63$
Error	$SSE = 51$	$N-K = 19-3 = 16$ ✓	$MSE = \frac{SSE}{16} = \frac{51}{16} = 3.1875$		
Total	58	18			

$x_{ij} = \underline{\underbrace{\mu_{ij}}_{=} + \varepsilon_{ij}}$; $i = 1,2 \ldots k$; $j = 1,2 \ldots n$.

Let $\mu_{ij} = \underline{\mu + \gamma_i + \beta_j}$

$i = 1,2 \ldots, k$; $j = 1,2 \ldots n$

$\therefore\ x_{ij} = \mu + \gamma_i + \beta_j + \varepsilon_{ij}$;

x_{ij} = observation in j^{th} block receiving i^{th} treatment

μ = common parameter for all treatments

γ_i = i^{th} treatment effect

β_j = j^{th} block effect

ε_{ij} = random error ; $\varepsilon_{ij} \sim N(\underline{0}, \underline{\sigma}^2)$

$\Rightarrow x_{ij} \sim N(\mu + \gamma_i + \beta_j, \sigma^2)$

$x_{ij} \rightarrow$ split into 3 components:

(i) The variation bet^{n} the treatments (rations) } assignable
(ii) The variation bet^{n} the varieties (blocks / breed) } cause

(iii) The inherent variation within the observations of treatments and within the observations of varieties (block) → chance causes.

Test procedure for two-way ANOVA :

(1) $H_{01}: \mu_1 = \mu_2 = \cdots = \mu_K$ or $H_{01}: \gamma_1 = \gamma_2 = \cdots = \gamma_K$

(2) $H_{02}: \mu_1 = \mu_2 = \cdots = \mu_n$ or $H_{02}: \beta_1 = \beta_2 = \cdots = \beta_n$

→ There is no significant difference bet^{n} the treatments.

→ " " " " " varieties (blocks).

(2) L. o. s. Let α = L.O.S.

(3) (a) $G = \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}$

$$CF = \frac{G^2}{N} = \frac{\left(\sum \sum x_{ij}\right)^2}{N}$$

(b) $TSS = \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}^2 - CF$

(c) $SST = \sum_{i=1}^{k} \frac{T_{i}^2}{n} - \left(\frac{G^2}{N}\right) CF$

(d) $SSB = \sum_{j=1}^{n} \frac{B_{.j}^2}{K} - CF$

x_{11} x_{12} x_{13}

x_{nk}

$x_{11}^2 + x_{12}^2 + x_{13}^2 + \dots + x_{nk}^2$

$i = 1, 2 \dots k$

$j = 1, 2 \dots n$

B_j^2

$$\frac{B_{.j}^2}{K}$$

$SSE = TSS - SST - SSB.$

(4) Degrees of freedom:

(a) $TSS \rightarrow N-1 = nK-1$

(b) $SST \rightarrow K-1$

(c) $SSB \rightarrow n-1$

(d) $SSE = (K-1)(n-1) = \underbrace{nK}_{N} - K - n + 1 = N - K - n + 1$

(5) Mean Sum of Squares: (MSS)

(a) $MST = \frac{SST}{df} = \frac{SST}{K-1}$

(b) $MSB = \frac{SSB}{df} = \frac{SSB}{n-1}$

(c) $MSE = \frac{SSE}{df} = \frac{SSE}{(n-1)(K-1} = \frac{SSE}{N-K-n+1}$

ANOVA → 2 way classification

Source of variation	Sum of squared (SS)	Degrees of freedom (df)	$MSS = \frac{SS}{df}$	F cal	Fcv
Between treatment	SST	$K-1$	$\frac{SST}{K-1} = MST$	$F_T = \frac{MST}{MSE}$	Fcv1 $= \text{TabF}(K-1, N-K-n+1), \alpha/2)$
Between Blocks	SSB	$n-1$	$\frac{SSB}{n-1} = MSB$	$F_B = \frac{MSB}{MSE}$	Fcv2 $= \text{TabF}((n-1, N-K-n+1), \alpha/2)$
Error	$SSE = TSS - SST - SSB$	$(n-1)(K-1) = (N-K-n+1)$	$\frac{SSE}{(n-1)(K-1)} = MSE$		
Total	TSS	$N-1 = nK-1$			

(7) C. value or table: $\xrightarrow{}$ for treatment)

$$F_{CV_1} = \text{tab } F\left[(k-1),(k-1)(n-1), \alpha/2\right].$$ at α level of significance

F_{CV_2} = tab F of F for blocks

$$= \text{tab } F\left[(n-1),(k-1)(n-1), \alpha/2\right]$$ at α level of significance

(8) (i) If cal $F_T <$ tab $F\left[(k-1),(k-1)(n-1),\alpha/2\right]$, we accept H_{01}.

(ii) If cal $F_B <$ tab $F\left[(n-1),(k-1)(n-),\alpha/2\right]$, we accept H_{02}

Q. Three varieties of coal were analysed by four chemists and the ash content in the varieties was found to be as under.

Solⁿ:

Varities	Chemists.				Total
	1	2	3	4	
A	10	7	7	9	33
B	9	8	6	6	29
C	5	8	7	6	26
Total	24	23	20	21	88

Null hypo:

$H_{01}: \gamma_1 = \gamma_2 = \gamma_3$, i.e. there is no significent difference between the varieties (treatment/rows)

$H_{02}: \beta_1 = \beta_2 = \beta_3 = \beta_4$; i.e. there is no significent difference betn the results obtained by four chemists (blocks/colms)

Alternative hypo H_1:

(i) $H_{11}: \gamma_1 \neq \gamma_2 \neq \gamma_3$; i.e. all γ's are not equal.

(ii)

$n = 4, K = 3$, $N = nK = 4 \times 3 = 12$

$G = 88$

$$C.F = \frac{G^2}{N} = \frac{(88)^2}{12} = \frac{7744}{12} = 645.33$$

$$TSS = \sum\sum x_{ij}^2 - CF = (10^2 + 7^2 + 7^2 + 9^2 + 9^2 + \dots + 6^2) - CF$$
$$= 24.6667$$

$$SST = \sum_{i=1}^{K} \frac{T_i^2}{n} - CF = \frac{33^2 + 29^2 + 26^2}{4} - 645.33$$
$$= 6.166667$$

$$\underset{\text{(Chemists)}}{SSB} = \sum_{j=1}^{4} \frac{B_j^2}{K} - CF = \frac{24^2 + 23^2 + 20^2 + 21^2}{3} - 645.33 = 3.3333$$

$$SSE = TSS - SST - SSB = 24.6667 - 6.16667 - 3.3333 = 15.16667$$

ANOVA Table

Source of variation	Sum of squared (SS)	Degrees of freedom (df)	Mean Sum of squares $MSS = \frac{SS}{df}$	F-cal	FCV at $\alpha = 0.05$
Between varieties (treatments)	SST = 6.16667	K-1 = 3-1 = 2	$MST = \frac{SST}{df} = \frac{6.16667}{2} = 3.083333$	$F_T = \frac{MST}{MSE} = 1.21978$	$F_{(2,6)} = 5.14$
Between chemist (Block)	SSB = 3.3333	n-1 = 4-1 = 3	$MSB = \frac{3.3333}{3} = 1.1111$	$F_B = \frac{1.1111}{2.527778} = 0.439556 < 1$	
Error	15.16667	6	$MSE = \frac{15.16667}{6} = 2.527778$		
Total	TSS = 24.66667	11			

Inference:

(i) Since $F_T = 1.21978 \leq$ tab $F(2,6) = 5.14$ at 0.05 l.o.s, we accept the null hypothesis H_{01} and conclude that there is no significant difference between varieties.

(ii) Since $F_B = 0.48.. < 1$, we accept the null hypothesis H_{02} and conclude that there is no significant difference between the chemists.

Applications of ANOVA :

1) Pharmaceuticals & medical research → it has the ability to test more than two samples simultaneously.

2) Medical research → helps us to compare various medicine treatments and decide which one is more effective over time & cost.

3) Agricultural research → test the effectiveness of different fertilizers, seeds or other factors.

4) Business research → compare the sales of different designs based on different factors.

5) Psychology researcher → to compare the different attitude or behaviour in people and whether or not they depend on the factors.

Non-Parametric Tests.

It does not have any parameters. They are also known as <u>distribution free</u> tests. The term 'non-parametric' refers to the fact that there are no parameters involved in the traditional sense of the term 'parameter' generally used. The only assumption required in non-parametric test is parent popl^n follows continuous distribution. These tests utilize some simple aspects of sample data such as the signs of measurements, order relationships or category frequencies.

Therefore, the stretching or compressing the scale does not alter them. So, the null distribution of the non-parametric test statistic can be determined without parameters of the parent popl^n dist^n.

Difference bet^n Parametric & Non-Parametric Tests.

Parametric Tests	Non-Parametric Tests.
1. Information about popl^n is completely known	Information about pop