

Analysis of Variance [ANOVA]

ANOVA technique is developed by R. A. Fisher.

What is ANOVA ? Statistical Technique developed to study significance of difference of means of more than 2 or more Samples.

Assumptions

- Population from which samples are drawn is Normally distributed.
- Samples are independent and random.
- Each one of the population has same variance. $\sigma_1^2 = \sigma_2^2 = \dots$
- Additivity of variances

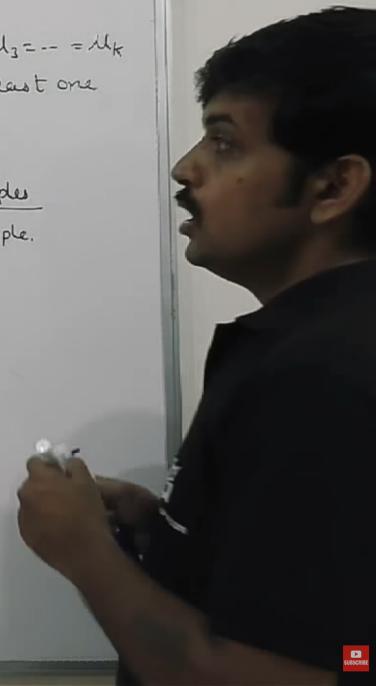
Null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$
Alternative Hypothesis, $H_1: \text{At least one mean is different.}$

Test Statistic

$$F = \frac{\text{Variation between Samples}}{\text{Variation within Sample.}}$$

Techniques of ANOVA

- One way ANOVA
- Two way ANOVA



Analysis of Variance [ANOVA]

Variation b/w samples			Variation within sample (error)
x_1	x_2	x_3	
20 21 23 16 20	18 20 17 25 15	25 28 22 28 32	
$\sum x_1 = 100$	$\sum x_2 = 95$	$\sum x_3 = 135$	
$\bar{x}_1 = 20$	$\bar{x}_2 = 19$	$\bar{x}_3 = 27$	
Grand mean, $\bar{\bar{x}} = \frac{20+19+27}{3} = 22$			
$H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3$			$\bar{x} = 22$
$H_1: \bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3$			$H_0 \text{ reject}$

ANOVA TABLE

Source of Variation	Sum of Sq.	d.f.	mean sum of Sq.	F-ratio
Explained	F_c	$> F_t$	$H_0 \text{ reject}$	
Unexplained	F_c	$< F_t$	$H_0 \text{ accept}$	

big -> reject
small -> accept

Null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$
Alternative Hypothesis, $H_1: \text{At least one mean is different.}$

Test Statistic

$$F = \frac{\text{Variation between Samples}}{(N-3)} = \frac{MSC}{MSE}$$

Techniques of ANOVA

- One way ANOVA
- Two way ANOVA

Independent factor

```

graph TD
    IF[Independent factor] --> X1[X1]
    IF --> X2[X2]
    IF --> X3[X3]
    
```

Total variation

Variation b/w samples

Variation within sample [error]

unexplained



One-way ANOVA

<p><u>One-Way ANOVA</u></p> <p>One Independent factor is influenced different Sample group (or level). Data is classified according to one criterion.</p> <p><u>Ex:</u> Mode of Payment (Independent Variable)</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th>CASH</th> <th>CHEQUE</th> <th>ONLINE</th> </tr> </thead> <tbody> <tr> <td>8</td> <td>7</td> <td>12</td> </tr> <tr> <td>10</td> <td>5</td> <td>9</td> </tr> <tr> <td>7</td> <td>10</td> <td>13</td> </tr> <tr> <td>14</td> <td>9</td> <td>12</td> </tr> <tr> <td>11</td> <td>9</td> <td>14</td> </tr> <tr> <td>$\sum x_1 = 50$</td> <td>$\sum x_2 = 40$</td> <td>$\sum x_3 = 60$</td> </tr> <tr> <td>$\bar{x}_1 = 10$</td> <td>$\bar{x}_2 = 8$</td> <td>$\bar{x}_3 = 12$</td> </tr> </tbody> </table> <p>$H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3$ $H_1: \text{At least mean is diff.}$</p>	CASH	CHEQUE	ONLINE	8	7	12	10	5	9	7	10	13	14	9	12	11	9	14	$\sum x_1 = 50$	$\sum x_2 = 40$	$\sum x_3 = 60$	$\bar{x}_1 = 10$	$\bar{x}_2 = 8$	$\bar{x}_3 = 12$	<p><u>Short-cut Method</u></p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th>x_1</th> <th>$\sum x_1^2$</th> <th>x_2</th> <th>$\sum x_2^2$</th> <th>x_3</th> <th>$\sum x_3^2$</th> </tr> </thead> <tbody> <tr> <td>$\sum x_1$</td> <td>$\sum x_1^2$</td> <td>$\sum x_2$</td> <td>$\sum x_2^2$</td> <td>$\sum x_3$</td> <td>$\sum x_3^2$</td> </tr> </tbody> </table> <ol style="list-style-type: none"> 1) Grand Total, $T = \sum x_1 + \sum x_2 + \sum x_3$ 2) Correction Factor = $\frac{T^2}{N}$, where $N = n_1 + n_2 + n_3$ 3) Sum of Square between Sample (SSC) $SSC = \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \frac{(\sum x_3)^2}{n_3} - \frac{T^2}{N}$ <p style="text-align: center;">d.f. = $C - 1$</p> 4) Total Sum of Square (SST) $SST = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 - \frac{T^2}{N}$ <p style="text-align: center;">d.f. = $N - 1$</p> 5) Sum of square within Sample (SSE) $SSE = SST - SSC$ <p style="text-align: center;">d.f. = $n - C$</p> 6. Mean sum of square between Sample (MSC) $MSC = \frac{SSC}{C - 1}$ 	x_1	$\sum x_1^2$	x_2	$\sum x_2^2$	x_3	$\sum x_3^2$	$\sum x_1$	$\sum x_1^2$	$\sum x_2$	$\sum x_2^2$	$\sum x_3$	$\sum x_3^2$
CASH	CHEQUE	ONLINE																																			
8	7	12																																			
10	5	9																																			
7	10	13																																			
14	9	12																																			
11	9	14																																			
$\sum x_1 = 50$	$\sum x_2 = 40$	$\sum x_3 = 60$																																			
$\bar{x}_1 = 10$	$\bar{x}_2 = 8$	$\bar{x}_3 = 12$																																			
x_1	$\sum x_1^2$	x_2	$\sum x_2^2$	x_3	$\sum x_3^2$																																
$\sum x_1$	$\sum x_1^2$	$\sum x_2$	$\sum x_2^2$	$\sum x_3$	$\sum x_3^2$																																

<p><u>One-Way ANOVA</u></p> <p>7. mean sum of square within Sample (MSE)</p> <p>$MSE = \frac{SSE}{n - C}$</p> <p><u>ONE-WAY ANOVA TABLE</u></p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th>Source of Variation</th> <th>Sum of squares</th> <th>degree of freedom</th> <th>mean sum of squares</th> <th>F-ratio</th> </tr> </thead> <tbody> <tr> <td>B/w Sample</td> <td>SSC</td> <td>$C - 1$</td> <td>$MSC = \frac{SSC}{C - 1}$</td> <td>$F = \frac{MSC}{MSE}$</td> </tr> <tr> <td>within Sample (error)</td> <td>SSE</td> <td>$n - C$</td> <td>$MSE = \frac{SSE}{n - C}$</td> <td></td> </tr> <tr> <td>Total</td> <td>SST</td> <td>$n - 1$</td> <td></td> <td></td> </tr> </tbody> </table> <p>Cal. F > tab F H_0 reject Cal. F < tab. F H_0 accept</p> <p>$H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3$ $H_1: \text{At least mean diff.}$</p>	Source of Variation	Sum of squares	degree of freedom	mean sum of squares	F-ratio	B/w Sample	SSC	$C - 1$	$MSC = \frac{SSC}{C - 1}$	$F = \frac{MSC}{MSE}$	within Sample (error)	SSE	$n - C$	$MSE = \frac{SSE}{n - C}$		Total	SST	$n - 1$			<p><u>Short-cut Method</u></p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th>x_1</th> <th>$\sum x_1^2$</th> <th>x_2</th> <th>$\sum x_2^2$</th> <th>x_3</th> <th>$\sum x_3^2$</th> </tr> </thead> <tbody> <tr> <td>$\sum x_1$</td> <td>$\sum x_1^2$</td> <td>$\sum x_2$</td> <td>$\sum x_2^2$</td> <td>$\sum x_3$</td> <td>$\sum x_3^2$</td> </tr> </tbody> </table> <ol style="list-style-type: none"> 1) Grand Total, $T = \sum x_1 + \sum x_2 + \sum x_3$ 2) Correction Factor = $\frac{T^2}{N}$, where $N =$ 3) Sum of square between Sample (SSC) $SSC = \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \frac{(\sum x_3)^2}{n_3} - \frac{T^2}{N}$ <p style="text-align: center;">d.f. = $C - 1$</p> 4) Total Sum of Square (SST) $SST = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 - \frac{T^2}{N}$ 5) Sum of square within Sample (SSE) $SSE = SST - SSC$ <p style="text-align: center;">d.f. = $n - C$</p> 6. Mean sum of square between Sample (MSC) $MSC = \frac{SSC}{C - 1}$ 	x_1	$\sum x_1^2$	x_2	$\sum x_2^2$	x_3	$\sum x_3^2$	$\sum x_1$	$\sum x_1^2$	$\sum x_2$	$\sum x_2^2$	$\sum x_3$	$\sum x_3^2$
Source of Variation	Sum of squares	degree of freedom	mean sum of squares	F-ratio																													
B/w Sample	SSC	$C - 1$	$MSC = \frac{SSC}{C - 1}$	$F = \frac{MSC}{MSE}$																													
within Sample (error)	SSE	$n - C$	$MSE = \frac{SSE}{n - C}$																														
Total	SST	$n - 1$																															
x_1	$\sum x_1^2$	x_2	$\sum x_2^2$	x_3	$\sum x_3^2$																												
$\sum x_1$	$\sum x_1^2$	$\sum x_2$	$\sum x_2^2$	$\sum x_3$	$\sum x_3^2$																												

Two-way ANOVA

Two-Way ANOVA

In two-way ANOVA, two independent factors influence different sample groups or levels.

[Data is classified according to two different criteria or factors].

Ex: Second Specialization | A — B — IIMs — L | Row Total

	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	Row Total
Y_1							ΣY_1
Y_2							ΣY_2
Y_3							ΣY_3
Column Total	ΣX_1	ΣX_1^2	ΣX_2	ΣX_2^2	ΣX_3	ΣX_3^2	Grand Total = T

1). Coding of data.
 2). Grand Total, $T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3$
 3). Correction factor = $\frac{T^2}{N}$, $N = n_1 + n_2 + n_3$
 4). Sum of square between Samples (SSC)
 $SSC = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - \frac{T^2}{N}$
 d.f. = $C-1$
 5). Sum of square between Rows [Blocks] (SSR)
 $SSR = \frac{(\Sigma Y_1)^2}{m_1} + \frac{(\Sigma Y_2)^2}{m_2} + \frac{(\Sigma Y_3)^2}{m_3} - \frac{T^2}{N}$
 d.f. = $R-1$
 6). Sum of square of Total (SST)
 $SST = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - \frac{T^2}{N}$
 d.f. = $N-1$

Two-Way ANOVA

7). Sum of square within Sample (SSC)

$SSC = SST - (SSC + SSR)$
 d.f. = $(C-1)(R-1)$

8). Mean sum of square b/w samples (MSC)

$MSC = \frac{SSC}{C-1}$

9). Mean sum of square b/w Rows (MSR)

$MSR = \frac{SSR}{R-1}$

10). Mean sum of square of error (MSE)

$MSE = \frac{SSG}{(C-1)(R-1)}$

11). F-ratio, $F_{\text{Column}} = \frac{MSC}{MSE}$, $F_{\text{Row}} = \frac{MSR}{MSE}$

Two-way ANOVA TABLE			
Source of Variation	Sum of Square	d.f.	Mean sum of Square
B/w Samples	SSC	$C-1$	$MSC = \frac{SSC}{C-1}$
B/w Rows	SSR	$R-1$	$MSR = \frac{SSR}{R-1}$
within Sample	SSE	$(C-1)(R-1)$	$MSE = \frac{SSG}{(C-1)(R-1)}$
Total	SST	$N-1$	

$F_{\text{cal}, R(V_1, V_2)} > F_{\text{tab}, (V_1, V_2)}$
 H_0 reject.

$F_{\text{cal}, R(V_1, V_2)} < F_{\text{tab}, (V_1, V_2)}$
 H_0 accept.

1). Coding of data.
 2). Grand Total, $T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3$
 3). Correction factor = $\frac{T^2}{N}$, $N = n_1 + n_2 + n_3$
 4). Sum of square between Samples (SSC)
 $SSC = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - \frac{T^2}{N}$
 d.f. = $C-1$
 5). Sum of square between Rows [Blocks] (SSR)
 $SSR = \frac{(\Sigma Y_1)^2}{m_1} + \frac{(\Sigma Y_2)^2}{m_2} + \frac{(\Sigma Y_3)^2}{m_3} - \frac{T^2}{N}$
 d.f. = $R-1$
 6). Sum of square of Total (SST)
 $SST = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - \frac{T^2}{N}$
 d.f. = $N-1$