⇒ Analysis of variance.

Introduced by Sir R A Fisher (1890-1962) in 1923.
(Abbreviated as ANOVA)

Analysis of variance is a technique that partitions the total variation - a term distinct from variance and measured by the sum of squares of deviations from the mean - into its component parts each of which is associated with a differente source of variation.

The analysis of variance ~~proc~~ procedure therefore compares two different estimates of variance by using F-distribution to determine wether the Population means are equal.

⇒ One-way Analysis of variance

It is also called the one-variable of classification analysis of variance.

Hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$

$H_1:$ Not all means are equal.

Data General table:

Observation	Samples (or Treatments) 1	2	...	j	...	K	Total
1	x_{11}	x_{12}		x_{1j}		x_{1K}	
2	x_{21}	x_{22}		x_{2j}		x_{2K}	
⋮	⋮	⋮		⋮		⋮	
i	x_{i1}	x_{i2}		x_{ij}		x_{iK}	
⋮	⋮	⋮		⋮		⋮	
r	x_{r1}	x_{r2}	...	x_{rj}		x_{rK}	
Total	$T_{.1}$	$T_{.2}$	...	$T_{.j}$	...	$T_{.K}$	$T_{..}$
Means	$\bar{X}_{.1}$	$\bar{X}_{.2}$		$\bar{X}_{.j}$		$\bar{X}_{.K}$	$\bar{X}_{..}$ → Grand mean

∵ K samples of equal size r

Scanned with CamScanner

Analysis of variance Table

Source of variation	d.f	Sum of squares (SS)	Mean Square (MS)	computed F
Between samples	$k-1$	$SSB = \frac{\Sigma T_{\cdot j}^2}{r} - C.F$	$S_b^2 = \frac{SSB}{k-1}$	$F = \frac{S_b^2}{S_w^2}$
Within samples (Error)	$n-k$	$SSE = TSS - SSB$	$S_w^2 = \frac{SSE}{n-k}$	
Total	$n-1$	$SST = \Sigma\Sigma X_{ij}^2 - C.F$	$S_T^2 = \frac{SST}{n-1}$	

$C.F$ (correction factor) $= \frac{T_{..}^2}{n}$

Six step

1) $H_0 : \mu_1 = \mu_2 = \dots \mu_k$
   H_1 : Not all k means are equal

2) Decide upon a significance level α

3) Test statistics: $F = \frac{S_b^2}{S_w^2}$

S_b^2 and S_w^2 are the two estimates of the common variance σ^2, if H_0 is true. has an F-dist with $v_1 = k-1$ and $v_2 = n-k$ degree of freedom.

4) Compute the necessary Sums of square and complete the analysis of variance table.

5) Determine the critical region which will consist of all values greater than or equal to $F_\alpha(k-1, n-k)$

6) Decide
   Reject H_0 if F falls in the critical region, accept H_0 otherwise.