Analysis of Variance:

Frequently, experimenters want to compare more than two treatments or results due to three or more teaching techniques; or miles per liters obtained from many different types of compact cars.

We discussed how to compare the means of two normal distributions. More generally, let us now consider m-normal distributions with unknown means $\mu_1, \mu_2, \dots, \mu_m$ and an unknown, but common, variance σ^2.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$$

H_1 : at least one is different from others.

Let $X_{i1}, X_{i2}, \dots, X_{in_i}$ represent a random sample of size n_i from the normal distributions $N(\mu_i, \sigma^2)$,

$i = 1, 2, \dots, m$.

m-Levels of treatments

					Row Means
X_1 :	X_{11}	X_{12}	$\cdots$	X_{1n_1}	$\bar{X}_{1\cdot}$
X_2 :	X_{21}	X_{22}	$\cdots$	X_{2n_2}	$\bar{X}_{2\cdot}$
$\vdots$	$\vdots$				$\vdots$
X_m :	X_{m1}	X_{m2}	$\cdots$	X_{mn_m}	$\bar{X}_{m\cdot}$
					$\bar{X}_{\cdot\cdot}$

(Sample means)

Grand Mean

Let $N = n_1 + n_2 + \cdots + n_m$

$$\bar{X}_{..} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} X_{ij} \quad \& \quad \bar{X}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad i = 1, 2, \ldots, m.$$

To determine a critical region for a test of H_0, we shall partition the sum of squares associated with the variance of the combined samples into two parts.

ANOVA SUM OF Squares Identity:

The total sum of squares (TSS) $= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2 + \sum_{i=1}^{m} n_i (\bar{X}_{i.} - \bar{X}_{..})^2$$

BSS = Between Sum of Squares $= \sum_{i=1}^{m} n_i (\bar{X}_{i.} - \bar{X}_{..})^2$

WSS = Within Sum of Squares $= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2$

$$\therefore \boxed{TSS = BSS + WSS}$$

Remark:

WSS – often called as the error sum of squares

BSS – often called as the between treatment sum of squares.

One-Way Analysis of Variance:

Source of Variation	Sum of Squares	Degrees of freedom	Mean Sum of Squares	F-ratio
Between	$BSS = \sum_{i=1}^{m} n_i(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$	$m-1$	$\frac{BSS}{m-1} = MS_B$	$\frac{BSS/m-1}{WSS/N-m} := F_{obs}$
Within	$WSS = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2$	$N-m$	$\frac{WSS}{N-m} = MS_W$	
Total	$TSS = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\cdot\cdot})^2$	$N-1$	—	

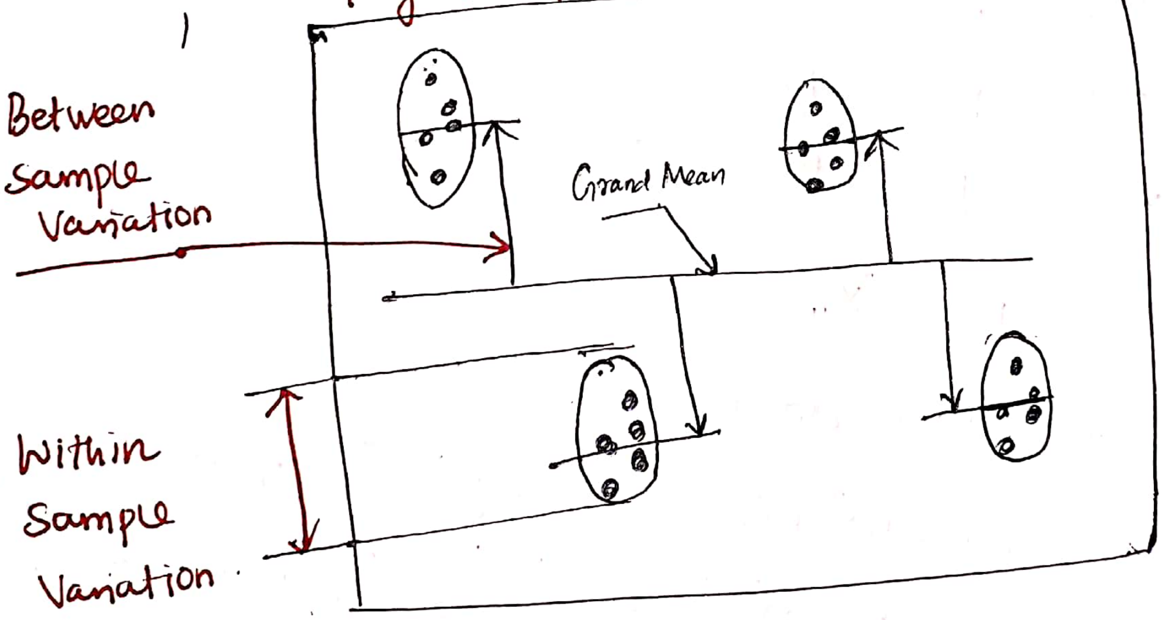
$N = n_1 + n_2 + \cdots + n_m$

$$F_{obs} = \frac{MS_B}{MS_W}$$

Accept H_0 if $F_{obs} \leq F_\alpha(m-1, N-m)$.

Reject H_0 if $F_{obs} > F_\alpha(m-1, N-m)$.

Key concepts in ANOVA:



The notions of between sample variation & within sample variation.

degrees of freedom partition: $N-1 = (m-1) + (N-m)$

There are $n_1 + n_2 + \cdots + n_m = N$ observations, thus TSS has $N-1$ degrees of freedom.

There are m levels of the factor, so BSS has $m-1$ degrees of freedom.

Finally, within any treatment, it contributes $n_i - 1$ degrees of freedom.

$\Rightarrow$ WSS has $(n_1 - 1) + (n_2 - 1) + \cdots + (n_m - 1) = N - m$

Scanned with CamScanner

(P1) Let X_1, X_2, X_3, X_4 be independent r.Vs that have normal distributions $N(\mu_i, \sigma^2)$, $i = 1, 2, 3, 4$.

We shall test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$

H_1: atleast one of the μ_i is different

On the basis of a random sample of size $n_i = 3$ from each of the four distributions.

	Observations			Means $\bar{X}_{i.}$
~~X_1~~ :	13	8	9	10
X_2 :	15	11	13	13
X_3 :	8	12	7	9
X_4 :	11	15	10	12
Grand mean $\bar{X}_{..}$				11

For the given data,

$$TSS = (13-11)^2 + (8-11)^2 + \cdots + (15-11)^2 + (10-11)^2 = 80$$

$$BSS = 3\left[(10-11)^2 + (13-11)^2 + (9-11)^2 + (12-11)^2\right] = 30$$

$$WSS = (13-10)^2 + (8-10)^2 + \cdots + (15-12)^2 + (10-12)^2 = 50.$$

Note that $TSS = BSS + WSS$, Only two of the three values need to be calculated directly from the data.

Scanned with CamScanner

The computed value of F is

$$F_{obs} = \frac{MS_B}{MS_W} = \frac{30/4-1}{50/12-4} = 1.6$$

Here $n = n_1 + n_2 + n_3 + n_4 = 12$

$m = 4$.

$$F_\alpha(m-1, n-m) = F_{0.05}(3, 8) = 4.07.$$

Since $F_{obs} < F_{0.05}(3,8)$

H_0 is accepted.