Page
Date

Unit-7

Analysis of variance :- The analysis of variance is a powerful statistical tool for tests of significance. The test of significance based on t-distⁿ is an adequate procedure only for testing the significance of the difference betⁿ two sample means. In a situation when we have three or more samples to consider at a time an alternative procedure is needed for testing the hypothesis that all the samples are drawn from the same popⁿ. i.e. they have the same mean.

eg :- Five fertilizers are applied to four plots each of wheat and yields of wheat on each of the plot is given. Now we are interested in finding out whether the effect of these fertilizers on the yields is significantly different or in other words, whether the samples have come from the same normal popⁿ.

The answer to this problem is provided by the technique of analysis of variance. Thus basic purpose of the analysis of variance is to test the homogeneity of several means.

The ~~ar~~ term "Analysis of variance" was introduced by prof. R.A. Fisher in 1930's to deal with problem in the analysis of agronomical data.

Cochran Thm :- used to justify results relating to the prob. dist^{n} of statistics that are used in ANOVA.

Page
Date

Variation is inherent in nature. The total variation in any set of numerical data is due to a no. of causes which may be classified as
i) Assignable causes, and
ii) Chance causes.
The variation due to assignable causes can be detected and measured whereas the variation due to chance causes is beyond the control of human hand and cannot be traced seperately

Definition :-

According to Prof. 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.

to credit or assign or to a cause

The ANOVA consists in the estimation of the amount of variation due to each of the indept. factors (causes) seperately and then comparing this estimates due to assignable factors (causes) with the estimates due to chance factor (causes)

Assumption :- For the validity of the F-test in ANOVA the following assumption are made :-
i) The obser^{n} are indept.
ii) Parent pop^{n} from which observation are taken as is normal and
iii) Various treatment and environmental effects are additive in nature.

Cochran's theorem :- Let $X_1, X_2, - X_n$ denote a r.s from normal pop^{n} $N(0, \sigma^2)$. Let the sum of the squares of these values are

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + - - + Q_K$$

where Q_j is a quadratic form of $X_1, X_2 \ldots X_n$ with rank (d.f) r_j; $j = 1, 2, \ldots k$: Then the r.v. $Q_1, Q_2 \ldots Q_k$ are mutually indept. and Q_j/σ^2 is a χ^2-variate with r_j degrees of freedom iff $\sum_{j=1}^{k} r_j = n$.

ANOVA for One-Way Classification :—
(with one obser^n per cell)

Let us suppose that N obser^n x_{ij} ($i = 1, 2, \ldots, k$ & ; $j = 1, 2, \ldots, n_i$) of a random variable X are grouped on some basis, into k classes of sizes $n_1, n_2, \ldots, n_k$ respectively. ($N = \sum_{i=1}^{k} n_i$) as exhibited below

				Means	Total
x_{11}	x_{12}	$\ldots$	x_{1n_1}	$\bar{x}_{1.}$	$T_{1.}$
x_{21}	x_{22}	$\ldots$	x_{2n_2}	$\bar{x}_{2.}$	$T_{2.}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
x_{k1}	x_{k2}	$\ldots$	x_{kn_k}	$\bar{x}_{k.}$	$T_{k.}$
					G

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

i) The variation bet^n the classes or the variation due to different bases of classification. commonly known as treatments.

ii) The variations within the classes i.e. the inherent variation of the r.v. within the obser^n of a class.

The first type of variation is due to assignable causes which can be detected and controlled by human endeavour and the second type of variation is due to change chance causes which are beyond the control of human hand.

The main object of analysis of variance technique is to ~~estimate~~ examine if there is significant difference between the class means ~~in view~~ of the due to chance causes i.e. variability within the separate classes.

eg ~~In~~ particular let us consider the effect of k different rations on the yield of milk of N cows (of same breed & stock) These N cows are divided into k classes of sizes $n_1, n_2, \ldots, n_k$ respectively.

$$N = \sum_{i=1}^{k} n_i$$

Here the sources of variation are
(i) Effect of the rations (treatment) t_i $i=1,2,\ldots k$
(ii) Error ε produced by numerous causes. They are not detected and identified & produce a variation of random nature.

Mathematical Model :- Let the ~~line~~ ~~us consider~~ the linear model will be

$$x_{ij} = \mu_i + \varepsilon_{ij}$$
$$= \mu + (\mu_i - \mu) + \varepsilon_{ij}$$
$$= \mu + \alpha_i + \varepsilon_{ij} \; ; \text{ where } (i = 1,2,\ldots,k \quad j = 1,2,\ldots,n_i) \quad \text{--- (1)}$$

(i) x_{ij} is the yield from the j^{th} cow, $(j = 1,2,\ldots,n_i)$ fed on the i^{th} rations $(i=1,2,\ldots k)$

(ii) μ is the general mean effect given by

$$\mu = \sum_{i=1}^{k} (n_i \mu_i)/N \qquad \text{— (2)}$$

where μ_i is the fixed effect due to the i^{th} ration, i.e. if there were no treatment differences and no chance causes then the yield of each cow will be μ.

(iii) α_i is the effect of the i^{th} ration (treatment) i.e. given by

$$\alpha_i = \mu - \mu_i \cdot (i = 1, 2 \ldots k) \qquad \text{— (3)}$$

i.e. the i^{th} treatment (rations) increase or decrease the yield by an amount α_i. On using (2), we get

$$\sum_{i=1}^{k} n_i \alpha_i = \sum_i n_i (\mu_i - \mu) = \sum n_i \mu_i - \mu \sum n_i$$

$$= N \cdot \mu - \mu \cdot N = 0$$

(iv) ε_{ij} is the error effect due to chance.

Assumption in the model

(i) All the obser^{n} x_{ij} are indept.

(ii) Different effects are additive in nature

(iii) ε_{ij} are i.i.d. $N(0, \sigma_e^2)$

under the (iii) assumption, the model (1) becom

$$E(x_{ij}) = \mu_i = \mu + \alpha_i \quad ; \quad \begin{cases} i = 1, 2, \ldots, k \\ j = 1, 2, \ldots, n_i \end{cases}$$

Null hypothesis of ~~one~~ for ANOVA for one-way classification.

Here we want to test the equality of popⁿ means i.e. the homogeneity of different rations. Hence null hypothesis is given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$$

which from ③ reduces to

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

Statistical Analysis of the Model :— Let us write

$$\overline{x_{i.}} = \text{mean of the } i^{th} \text{ class.}$$

$$= \sum_{j=1}^{n_i} \frac{x_{ij}}{n_i} \quad ; \quad (i = 1, 2, \ldots, k)$$

$$\overline{x}_{..} = \text{overall mean} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{1}{N} \sum_{i=1}^{K} n_i \overline{x_{i.}}$$

The parameters μ and α_i in ① are estimated by the principle of least squares on minimising the error (residual) sum of squares given by

$$E = \sum_i \sum_j \varepsilon_{ij}^2 = \sum_i \sum_j (x_{ij} - \mu - \alpha_i)^2 \quad \text{— ⑧}$$

The normal eqⁿˢ for estimating μ and α_i are

$$\frac{\partial E}{\partial \mu} = -2 \sum_i \sum_j (x_{ij} - \mu - \alpha_i) = 0 \quad \text{— (*)}$$

and $$\frac{\partial E}{\partial \alpha_i} = -2 \sum_{j=1}^{n_i} (x_{ij} - \mu - \alpha_i) = 0 \quad \text{— (**)}$$

Page
Date

from (*), we get

$$\sum_i \sum_j x_{ij} - N\mu - \sum n_i \alpha_i = 0$$

$$\hat{\mu} = \frac{1}{N} \sum\sum x_{ij} = \bar{x}_{..} \qquad [\because \sum n_i \alpha_i = 0]$$

& from (**), we get

$$\sum_j x_{ij} - n_i \hat{\mu} - n_i \hat{\alpha} = 0$$

$$\Rightarrow \hat{\alpha} = \frac{1}{n_i} \sum x_{ij} - \hat{\mu} = \bar{x}_{i.} - \hat{\mu}$$

$$\hat{\alpha} = \bar{x}_{i.} - \bar{x}_{..}$$

Hence substituting the value of μ & α in (1) the model becomes

$$x_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.})$$

We introduce the error component ε_{ij} so that both the sides are equal. This is the deviation within the class which is due to randomisation. Transposing $\bar{x}_{..}$ to the left squaring both sides and summing over i & j we get

$$\sum_i \sum_j [x_{ij} - \bar{x}_{i.} + \bar{x}_{i.} - \bar{x}_{..}]^2$$

$$\sum_i^k \sum_j^{n_i} (x_{ij} - \bar{x}_{..})^2 = \sum_i^k \sum_j^{n_i} \left[(\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.})\right]^2$$

$$= \sum_i^k \sum_j^{n_i} (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i^k \sum_j^{n_i} (x_{ij} - \bar{x}_{i.})^2$$

$$+ 2 \sum_i^k \cancel{\sum_j} (\bar{x}_{i.} - \bar{x}_{..}) \sum_j^{n_i} (x_{ij} - \bar{x}_{i.})$$

$$\sum_{i}^{k}\sum_{j}^{n_i}(x_{ij}-\bar{x}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-x_{i.})^2 + \sum_{i} n_i(\bar{x}_{i.}-\bar{x}_{..})^2$$

$$+\ 2\left\{\sum_{i}\left[(\bar{x}_{i.}-\bar{x}_{..})\sum_{j}(x_{ij}-\bar{x}_{i.})\right]\right\}$$

But $\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})=0$ [algebric sum of the deviation of the rations from their mean is zero]

$\therefore$ $$\sum_{i}\sum_{j}(x_{ij}-\bar{x}_{..})^2 = \sum_{i}\sum_{j}(x_{ij}-\bar{x}_{i.})^2 + \sum_{i} n_i(\bar{x}_{i.}-\bar{x}_{..})^2 \quad -(1)$$

$S_T^2 = \sum_{i}\sum_{j}(x_{ij}-\bar{x}_{..})$ is known as total sum of squares (T.S.S.)

$S_E^2 = \sum_{i}\sum_{j}(x_{ij}-\bar{x}_{i.})^2$ is called within sum of squares or error sum of squares (S.S.E.) and

$S_t^2 = \sum_{i} n_i(\bar{x}_{i.}-\bar{x}_{..})^2$ is called S.S. due to treatments (S.S.T.)

Then T.S.S. = S.S.E. + S.S.T.

Degrees of freedom for various S.S :-

S_T^2, the total S.S. will carry $(N-1)$ d.f. (one d.f. lost because of the linear constraint)

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{.})=0$$

similarly the treatment sum of squares S_t^2 will have $(k-1)$ d.f.

& S_E^2 the error s.s. will have $(N-k)$. d.f. . .

Hence we see that the

$$N-1 = (N-k) + (k-1)$$

Mean Sum of Squares (M.S.S)

Defn ~~Tho~~
The sum of squares divided by its degrees of freedom gives the corresponding variance or the mean sum of squares (M.S.S). Thus

$$\frac{S_t^2}{(k-1)} = \frac{S.S.T.}{(k-1)} = s_t^2$$ is M.S.S. due to treatment

& $$\frac{S_E^2}{N-k} = \frac{S.S.E.}{(N-k)} = s_E^2$$ is M.S.S. due to error.

Hence the test statistic for H_0 provided by the variance ratio

$$F = \frac{s_t^2}{s_E^2}$$ ~~XXX~~

If H_0 is true the F- should take the value 1, otherwise it should be greater than unity. In order to find out if an observed value of F is significantly greater than unity. we have to obtain the sampling distⁿ of the F-statistic defined as

$$F = s_t^2 / s_E^2$$

Under H_0 by Cochran's thm., $\frac{S_t^2}{\sigma_e^2}$ & $\frac{S_E^2}{\sigma_e^2}$ are independently distributed as χ^2-variates with $(K-1)$ and $(N-K)$ d.f. respectively.

Hence the statistic

$$F = \left[\frac{S_t^2}{\sigma_e^2}\cdot\frac{1}{K-1}\right] \div \left[\frac{S_E^2}{\sigma_e^2}\cdot\frac{1}{(N-K)}\right] = \frac{s_t^2}{s_E^2}$$

follows Snedecor's F (central) distⁿ with $[(K-1), (N-K)]$ d.f.

Thus if an observed value of F obtained from ~~xxx~~ is greater than tabulated value of F for $(K-1, N-K)$ d.f. at specified level of significance (usually 5% or 1%) then H_0 is refuted at that level otherwise H_0 may be retained.

The above statistical analysis is very elegantly presented in the following table

ANOVA Table for ONE-WAY CLASSIFIED DATA

Sources of Variation	Sum of Squares	d.f.	Mean Sum of Squares	Variance Ratio
Treatment (rations)	S_t^2	$K-1$	$s_t^2 = \frac{S_t^2}{K-1}$	$\frac{s_t^2}{s_E^2} = F_{K-1, N-K}$
Error	S_E^2	$N-K$	$s_E^2 = \frac{S_E^2}{N-K}$	
Total	S_T^2	$N-1$		

Page
Date

Remark: For practical point of view; various sum of squares reduced to a great extent by using following simplified formulae

$$\text{Total S.S.} = \sum_i\sum_j (x_{ij} - \bar{x}_{..})^2$$

$$= \sum\sum x_{ij}^2 - \frac{(\sum\sum x_{ij})^2}{N}$$

$$= \sum\sum x_{ij}^2 - \frac{G^2}{N}$$

where G is the grand total of all the obsern.
and $N = n_1 + n_2 + \cdots + n_k = \sum n_i$
The expression $\sum_i\sum_j x_{ij}^2$ the sum of squares of all the obsern is known as Raw sum of squares (R.S.S.) and the expression $\frac{G^2}{N}$ is called the correction factor (C.F.).

$$\text{Total S.S.} = \text{R.S.S.} - \text{C.F.}$$

$$\text{Error S.S.} = \sum\sum (x_{ij} - x_{i.})^2 = \sum_i \left\{ \sum_j (x_{ij} - \bar{x}_{i.})^2 \right\}$$

$$= \sum_i \left[\sum_j x_{ij}^2 - \frac{(\sum_j x_{ij})^2}{n_i} \right] = \sum_i \left[\sum_j x_{ij}^2 - \frac{T_{i.}^2}{n_i} \right]$$

$$S_E^2 = \sum_i\sum_j x_{ij}^2 - \sum_i \left(\frac{T_{i.}^2}{n_i} \right)$$

where $T_{i.}$ is the total yield from the units. receiving the ith treatment

$$\text{Treatment S.S.} = S_t^2 = \text{T.S.S.} - \text{S.S.E.}$$

$$= \sum \left(\frac{T_{i.}^2}{n_i} \right) - \frac{G^2}{N}$$

$$= \sum \left(\frac{T_{i.}^2}{n_i} \right) - \text{C.F.}$$

Page
Date

ANOVA for two-way Classification (with one obser's per cell :-

Let us consider the case when there are two factors which may affect the affect the variate values x_{ij}.

eg. the yield of milk may be affected by difference of treatments i.e. rations as well as the differnce in variety i.e. breed and stock of the cows. Let us & now suppose that the N cows are divided into h different groups or classes according to their breed and stock, each group containing k cows and then let us consider the effect of k treatments (i.e. rations given at random to cow in each group) on the yield of milk.

Let us suffix i refer to the treatments (rations) and suffix j refer to the varieties (breed of the cow). Then the yield of milk x_{ij} $(i=1,2,\ldots k,\ j=1,2,\ldots h)$ of $N=h\times k$ cows furnish the data for the comparison of the treatments. The yield may be expressed as varite values in the following $k\times h$ two-way table

						Mean	total
	x_{11}	x_{12} ...	x_{1j} ...	x_{1h}		$\bar{x}_{1.}$	$T_{1.}$
	x_{21}	x_{22}	x_{2j}	x_{2h}		$\bar{x}_{2.}$	$T_{2.}$
	⋮					⋮	⋮
	x_{i1}	x_{i2}	x_{ij}	x_{ih}		$\bar{x}_{i.}$	$T_{i.}$
	⋮						
	x_{k1}	x_{k2}	x_{kj}	x_{kh}		$\bar{x}_{k.}$	$T_{k.}$
Mean total	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.j}$	$\bar{x}_{.h}$		$\bar{x}_{..}$	

Total | $T_{\cdot 1}$ $T_{\cdot 2}$ $T_{\cdot j}$ $T_{\cdot h}$ | $x_{\cdot\cdot}$

Mathematical Model :- Let x_{ij} be the yield from the cow of j^{th} variety fed on the i^{th} ration ($i=1,2,\ldots,k$, $j=1,2,\ldots,h$).

Let us suppose that x_{ij} ($i=1,2,\ldots,k$; $j=1,2,\ldots,h$) are indept., normally distributed as $N(\mu_{ij}, \sigma_e^2)$

The linear mathematical model becomes

$$E(x_{ij}) = \mu_{ij}$$

or $$x_{ij} = \mu_{ij} + \varepsilon_{ij} \quad \text{--- (1)}$$

where ε_{ij} are i.i.d. $N(0, \sigma_e^2)$

μ_{ij} is further split into the following parts:

(i) The general mean effect μ given by

$$\mu = \sum_i \sum_j \mu_{ij} / N \quad \text{--- (2)}$$

(ii) The effect α_i ($i=1,2,\ldots,k$) due to the i^{th} ration given by

$$\alpha_i = \mu_{i\cdot} - \mu$$

where $\mu_{i\cdot} = \frac{1}{h}\sum_{j=1}^{h} \mu_{ij}$; ($i=1,2,\ldots,k$) --- (3)

obviously $\sum_{i=1}^{k} \alpha_i = 0$

(iii) The effect β_j, ($j=1,2,\ldots,h$) due to j^{th} variety (breed of cow) given by

$$\beta_j = \mu_{.j} - \mu$$

$$\mu_{.j} = \frac{1}{k}\sum_{i=1}^{k} \mu_{ij} \quad ; \quad (j = 1, 2, \ldots h) \qquad \text{—} (4)$$

obviously $\sum_{j=1}^{h} \beta_j = 0$

(iv) The interaction effect γ_{ij}, when the i^{th} level of first factor (rations) and j^{th} level of second factor (breed of cow) occur simultaneously and given by

$$\gamma_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu$$

$$\sum_j \gamma_{ij} = 0, \quad \forall \ i = 1, 2, \ldots, k$$

$$\sum_j \gamma_{ij} = 0, \quad \forall \ j = 1, 2, \ldots, h$$

$$\sum_{i=1}^{k} \mu_{ij} - \sum_{i=1}^{k} \mu_{i.} - k\mu_{.j} + k\mu$$

$$\cancel{k\mu_{.j}} - \cancel{k\mu} - \cancel{k\mu_{.j}} + \cancel{k\mu}$$

and Thus we have

$$\mu_{ij} = \mu + (\mu_{i.} - \mu) + (\mu_{.j} - \mu) + (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu) \qquad (5)$$

and consequently (1) becomes

$$x_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij} \qquad \text{—} (6)$$

Where ε_{ij} is the error effect due to chance &

$$\sum_{i=1}^{k} \alpha_i = 0 = \sum_{j=1}^{h} \beta_j$$

&
$$\sum_{i} \gamma_{ij} = \sum_{j} \gamma_{ij} = 0$$
$$\forall i \qquad \forall j$$

As there is only one obser^{n} in each cell, the obser^{n} corresponding to the ith level of ration & jth level of breed of cow is only one i.e. x_{ij}. But we cannot estimate by one value alone. Hence ~~in this~~ case of one obser^{n} per cell the interaction effect $\gamma_{ij}=0$ and the model (6) reduces to

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \text{—— (7)}$$

Statistical Analysis of the Model

Let us write

$$\bar{x}_{i.} = \text{Mean yield of } i\text{th ration}$$
$$= \frac{1}{h}\sum_{j=1}^{h} x_{ij} \quad (i = 1, 2, \ldots k)$$

$$\bar{x}_{.j} = \text{mean yield of } j\text{th variety}$$
$$= \frac{1}{k}\sum_{i=1}^{k} x_{ij} \quad (j = 1, 2, \ldots h)$$

$$\bar{x}_{..} = \text{Overall mean} = \frac{1}{hk}\sum_i\sum_j x_{ij}$$
$$= \frac{1}{h}\sum_j\left(\frac{1}{k}\sum_i^{k} x_{ij}\right) = \frac{1}{h}\sum_j \bar{x}_{.j}$$
$$= \frac{1}{k}\sum_i\left(\frac{1}{h}\sum_j^{h} x_{ij}\right) = \frac{1}{k}\sum_i \bar{x}_{i.}$$

The least square estimates of the parameters μ, α_i and β_j are obtained on minimising

Page ____
Date ____

the error sum of squares

$$E = \sum_{i=1}^{k}\sum_{j=1}^{h} \varepsilon_{ij}^2 = \sum_i\sum_j (x_{ij} - \mu - \alpha_i - \beta_j)^2$$

The normal eq^n: for eq estimating μ, α_i & β_j are respectively

$$\frac{\partial E}{\partial \mu} = 0 = -2\sum_i\sum_j (x_{ij} - \mu - \alpha_i - \beta_j)$$

$$\frac{\partial E}{\partial \alpha_i} = 0 = -2\sum_j (x_{ij} - \mu - \alpha_i - \beta_j)$$

$$\frac{\partial E}{\partial \beta_j} = 0 = -2\sum_i (x_{ij} - \mu - \alpha_i - \beta_j)$$

since $\sum_i \alpha_i = 0 = \sum_j \beta_j$: we get from the above eq^n.

$$\hat{\mu} = \frac{1}{hk}\sum_i\sum_j x_{ij} = \bar{x}_{..}$$

$$\hat{\alpha}_i = \frac{1}{h}\sum_j x_{ij} - \hat{\mu} = \bar{x}_{i.} - \bar{x}_{..}$$

$$\hat{\beta}_j = \frac{1}{k}\sum_i x_{ij} - \hat{\mu} = \bar{x}_{.j} - \bar{x}_{..}$$

Thus linear model (7) becomes

$$x_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

The error term ε_{ij} being so chosen that both sides are equal

Transposing $\bar{x}_{..}$ to the left side, squaring and summing both sides over i & j, we get

$$\sum_i\sum_j (x_{ij}-\bar{x}_{..})^2 = \sum_i\sum_j \left[(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..}) + (\bar{x}_{i.}-\bar{x}_{..}) + (\bar{x}_{.j}-\bar{x}_{..})\right]^2$$

$$= \sum_i\sum_j (x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2 + \sum_i\sum_j (\bar{x}_{i.}-\bar{x}_{..})^2 + \sum_i\sum_j (\bar{x}_{.j}-\bar{x}_{..})^2 + 2\sum_i\sum_j (x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})(\bar{x}_{i.}-\bar{x}_{..}) + 2\sum_i\sum_j (x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})(\bar{x}_{.j}-\bar{x}_{..}) + 2\sum_i\sum_j (\bar{x}_{i.}-\bar{x}_{..})(\bar{x}_{.j}-\bar{x}_{..})$$

since algebric sum of deviation of a set of obser^{n}: about their mean is zero.
similarly other product terms also vanishes.
Hence

$$\sum_i\sum_j (x_{ij}-\bar{x}_{..})^2 = \sum_i\sum_j (x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2 + h\sum_i (\bar{x}_{i.}-\bar{x}_{..})^2 + k\sum_j (\bar{x}_{.j}-\bar{x}_{..})^2$$

$$S_T^2 = S_E^2 + S_t^2 + S_V^2$$

T.S.S. = S.S. due to error + S.S. due to treatments + S.S. due to varieties.

~~Null~~ Null hypothesis:— We set up the null hypothesis that treatment as well as varieties are homogeneous. In other words, the null hypothesis for treatment and varieties are respectively

$$H_t = \mu_{1.} = \mu_{2.} = \cdots = \mu_{k.} = \mu$$

$$H_V = \mu_{.1} = \mu_{.2} = \cdots = \mu_{.h} = \mu$$

Q1

$H_t : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$
$H_v : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

Degrees of freedom for various S.S. :— The total S.S., ~~S_T^2~~ S_T^2 being computed from $N = hk$ quantities, $(x_{ij} - \bar{x}_{..})$, which are subject to one linear constraint. $\sum\sum(x_{ij} - \bar{x}_{..}) = 0$, will carry $(N-1)$ d.f.

similarly for S_t^2 will be based on $(k-1)$ d.f. & since $\sum_i(\bar{x}_{i.} - \bar{x}_{..}) = 0$

S_v^2 will have $(h-1)$ d.f. since $\sum_j(\bar{x}_{.j} - \bar{x}_{..}) = 0$ will carry $(h-1)$ d.f.

~~...~~ Σ & S_E^2 will carry $(N-1) - (k-1) - (h-1)$
$= (h-1)(k-1)$ d.f.

Thus the partitioning of d.f. is as follows

$(hk-1) = (k-1) + (h-1) + (h-1)(k-1)$
which $\Rightarrow$ that d.f. are additive.

Test statistic :— To obtain appropriate test statistic to test the hypothesis H_t & H_v, we need the expectation of the various mean sum of squares due to each indept factors.

M.S.S. due to treatment $= \frac{S_t^2}{k-1} = s_t^2$

Page
Date

M.S.S. due to variety $= \dfrac{S_V^2}{h-1} = s_V^2$

error M.S.S. $= \dfrac{S_E^2}{(h-1)(k-1)} = s_E^2$.

Since various S.S. and d.f. are additive and since under H_t & H_V, each of s_t^2, s_V^2 & s_E^2 provides an unbiased estimates of σ_e^2, under the assumption of normal parent popn, we get by Cochran's thm

$\dfrac{S_t^2}{\sigma_e^2}$, $\dfrac{S_V^2}{\sigma_e^2}$ & $\dfrac{S_E^2}{\sigma_e^2}$ are mutually indept.

χ^2 with $(k-1)$, $(h-1)$ & $(h-1)(k-1)$ d.f. respectively. Hence under H_t and H_V respectively, we get

$$F_t = \frac{S_t^2}{\sigma_e^2(k-1)} \div \frac{S_E^2}{\sigma_e^2(h-1)(k-1)} = \frac{s_t^2}{s_E^2} \sim F_{k-1,\,(h-1)(k-1)}$$

$$F_V = \frac{S_V^2}{\sigma_e^2(h-1)} \div \frac{S_E^2}{\sigma_e^2(h-1)(k-1)} = \frac{s_V^2}{s_E^2} \sim F_{h-1,\,(h-1)(k-1)}$$

ANOVA for two-way classified Data

Sources of variation	Sum of squares	d.f.	M.S.S.	Variance Ratio
Treatments	$S_T^2 = \sum_i h(\bar{x}_{i.} - \bar{x}_{..})^2$	$k-1$	$s_t^2 = \dfrac{S_T^2}{k-1}$	$F_t = \dfrac{s_t^2}{s_E^2} \sim k-1, (h-1)(k-1)$
Varieties	$S_V^2 = \sum_j k(\bar{x}_{.j} - \bar{x}_{..})^2$	$h-1$	$s_V^2 = \dfrac{S_V^2}{h-1}$	$F_V = \dfrac{s_V^2}{s_E^2}$
Residuals	$S_E^2 = \sum_i\sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$	$(h-1)(k-1)$	$s_E^2 = \dfrac{S_E^2}{(h-1)(k-1)}$	$(h-1), (h-1)(k-1)$
Total		$N-1$		