UNIT IV

Introduction

The square of a standard normal variate is known as a chisquare variate with

1 degree of freedom.Thus if x follows normal with mean μ and variance σ^2

then z = x-μ / σ, z^2 is a chi-square variate with 1 d.f   In general if xi , I = 1,2,……n

are n independent normal variates with means μi and variances σi^2 then

χ^2 = $\sum zi^2$ is a chisquare variate with n d.f.

Conditions for the validity of chi square test

1.the sample observations should be independent.

2.constraints on the cell frequencies , if any, should be linear for example

∑ oi = ∑ Ei

3.N, the total frequency should be reasonably large , say, greater than 50

4.No theoretical cell frequency should be less than 5.

Since chi square does not involve any population parameters, it is termed as a statistic and the test is known as non parametric or Distribution free test.

Applications

To test if the hypothetical value of the population variance is σ^2 = σo^2 (say)

To test the goodness of fit

To test the independence of attributes

To test the homogeneity of independent estimates of the population variance

Inferences about a population variance

Suppose we want to test if a random sample xi, I = 1,2,…..n has been drawn from a normal population with a specified variance σ^2 = σo^2 then the statistic

χ^2= $\sum$ [$xi - \ddot{x}]^2/\sigma o^2$ = ns^2 / σo^2 follows chisquare distribution with (n-1) d.f.

It is believed that the precision of an instrument is no more than .16 write down the null and alternative hypothesis for testing this belief

2.5 2.3 2.4 2.3 2.5 2.7 2.5 2.6 2.6 2.7 2.5

Solution

H_0: σ^2 = .16

H_1: $\sigma^2 >$.16

Goodness of fit test

A very powerful test for testing the significance of the discrepancy between theory and experiment was given by prof.karl pearson and is known as chi-square test of goodness of fit.

It enables us to find if the deviation of the experiment from theory is just by Chance or is it really due to the inadequacy of the theory to fit the observed Data.

If oi i=1,2,…..n is a set of observed frequencies and Ei I = 1,2,…….n is the

Corresponding set of expected frequencies then

$\chi^2 = \sum (oi - Ei)^2 / Ei$, ($\sum oi = \sum Ei$)

Follows chi-square with (n-1) d.f.

The demand for a particular spare part in a factory was found to vary from day-to-day.

Days	Mon	Tue	Wed	Thurs	Fri	Sat
Number	1124	1125	1110	1120	1126	1115

Solution

H_0: The number of parts demanded does not depend on the day of week.

Digits:	0	1	2	3	4	5	6	7	8	9
Frequency:	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur frequently in the directory.

15·6·3. Test of Independence of Attributes—Contingency Tables. Let us consider two attributes A and B, A divided into r classes $A_1, A_2, \ldots, A_r$ and B divided into s classes $B_1, B_2, \ldots, B_s$. Such a classification in which attributes are divided into more than two classes is known as *manifold classification*. The various cell frequencies can be expressed in the following table known as $r \times s$ *manifold contingency table* where (A_i) is the number of persons possessing the attribute A_i, $(i = 1, 2, \ldots, r)$, (B_j) is the number of persons possessing the attribute B_j $(j = 1, 2, \ldots, s)$ and (A_iB_j) is the number of persons possessing both the attributes A_i and B_j, $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, s)$. Also $\sum_{i=1}^{r} (A_i) = \sum_{j=1}^{s} (B_j) = N$, where N is the total frequency.

TABLE 15·7 : $r \times s$ CONTINGENCY TABLE

B \ A	A_1	A_2	...	A_i	...	A_r	*Total*
B_1	(A_1B_1)	(A_2B_1)	...	(A_iB_1)	...	(A_rB_1)	(B_1)
B_2	(A_1B_2)	(A_2B_2)	...	(A_iB_2)	...	(A_rB_2)	(B_2)
⋮	⋮	⋮		⋮	⋮	⋮	⋮
B_j	(A_1B_j)	(A_2B_j)	...	(A_iB_j)	...	(A_rB_j)	(B_j)
⋮	⋮	⋮		⋮	⋮	⋮	⋮
B_s	(A_1B_s)	(A_2B_s)	...	(A_iB_s)	...	(A_rB_s)	(B_s)
Total	(A_1)	(A_2)	...	(A_i)	...	(A_r)	N

Example 15·17. (2 × 2 CONTINGENCY TABLE). *For the 2 × 2 table,*

a	b
c	d

, prove that chi-square test of independence gives

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}, \; N = a+b+c+d \quad \ldots(15\cdot18)$$

Solution. Under the hypothesis of independence of attributes,

$$E(a) = \frac{(a+b)(a+c)}{N}$$

$$E(b) = \frac{(a+b)(b+d)}{N}$$

$$E(c) = \frac{(a+c)(c+d)}{N}$$

and $$E(d) = \frac{(b+d)(c+d)}{N}$$

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

$$\therefore \quad \chi^2 = \frac{[a-E(a)]^2}{E(a)} + \frac{[b-E(b)]^2}{E(b)} + \frac{[c-E(c)]^2}{E(c)} + \frac{[d-E(d)]^2}{E(d)} \quad \ldots(*)$$

$$a - E(a) = a - \frac{(a+b)(a+c)}{N} = \frac{a(a+b+c+d)-(a^2+ac+ab+bc)}{N} = \frac{ad-bc}{N}$$

Similarly, we will get : $\quad b - E(b) = -\dfrac{ad-bc}{N} = c - E(c)\,; \quad d - E(d) = \dfrac{ad-bc}{N}$

Substituting in (*), we get

$$\chi^2 = \frac{(ad-bc)^2}{N^2}\left[\frac{1}{E(a)} + \frac{1}{E(b)} + \frac{1}{E(c)} + \frac{1}{E(d)}\right]$$

$$= \frac{(ad-bc)^2}{N}\left[\left\{\frac{1}{(a+b)(a+c)} + \frac{1}{(a+b)(b+d)}\right\} + \left\{\frac{1}{(a+c)(c+d)} + \frac{1}{(b+d)(c+d)}\right\}\right]$$

$$= \frac{(ad-bc)^2}{N}\left[\frac{b+d+a+c}{(a+b)(a+c)(b+d)} + \frac{b+d+a+c}{(a+c)(c+d)(b+d)}\right]$$

$$= (ad-bc)^2\left[\frac{c+d+a+b}{(a+b)(a+c)(b+d)(c+d)}\right] = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}.$$

Remark. We can calculate the value of χ^2 for 2 × 2 contingency table by using (15·18) directly. The reader is advised to obtain the value of χ^2 in *Example 15·16* by using (15·18).

Example 15·18. *Out of 8,000 graduates in a town 800 are females, out of 1,600 graduate employees 120 are females. Use χ^2 to determine if any distinction is made in appointment on the basis of sex. Value of χ^2 at 5% level for one degree of freedom is 3·84.*

5·1. INTRODUCTION

The analysis of variance is a powerful statistical tool for tests of significance. The test of significance based on t-distribution is an adequate procedure only for testing the significance of the difference between 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 population, *i.e.*, they have the same mean. For example, five fertilizers are applied to four plots each of wheat and yield of wheat on each of the plot is given. We may be 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 population. The answer to this problem is provided by the technique of analysis of variance. *The basic purpose of the analysis of variance is to test the homogeneity of several means.*

... was introduced by **Prof. R.A. Fisher** in 1920's to deal ...

factor (causes), the latter being known as experimental error or simply error.

5·1·1. Assumptions for ANOVA Test. ANOVA test is based on the test statistics F (or *Variance Ratio*).

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

(*i*) The observations are independent,

(*ii*) Parent population from which observations are taken is normal, and

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

5·2. ONE-WAY CLASSIFICATION

Let us suppose that N observations y_{ij}, $(i = 1, 2, ..., k;\ j = 1, 2, ..., n_i)$ of a random variable Y are grouped, on some basis, into k classes of sizes $n_1, n_2, ..., n_k$ respectively, $\left(N = \sum_{i=1}^{k} n_i\right)$ as exhibited in *Table 5·1*.

TABLE 5·1 : ONE-WAY CLASSIFIED DATA

Class	*Sample Observations*				*Total*	*Mean*
1	y_{11}	y_{12}	...	y_{1n_1}	$T_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	...	y_{2n_2}	$T_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{in_i}	$T_{i\cdot}$	$\bar{y}_{i\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	y_{k1}	y_{k2}	...	y_{kn_k}	$T_{k\cdot}$	$\bar{y}_{k\cdot}$

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

(*i*) The variation *between the classes* or the variation due to different bases of classification, commonly known as *treatments*.

(*ii*) The variation *within the classes, i.e.,* the inherent variation of the random variable within the observations 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 *chance causes* which are beyond the control of human hand.

The main object of analysis of variance technique is to examine if there is significant difference between the class means in view of the inherent variability within the separate classes.

In particular, let us consider the effect of k different rations on the yield in milk of N cows (of the same breed and stock) divided into k classes of sizes $n_1, n_2, ..., n_k$ respectively, $N = \sum_{i=1}^{k} n_i$. Here the sources of variation are :

5·2·1. ANOVA for Fixed Effect Model. If the factor levels under consideration are the only levels of interest, then the *fixed effect* or *parametric* model given below is used :

$$y_{ij} = \mu_i + \varepsilon_{ij}$$
$$= \mu + \alpha_i + \varepsilon_{ij} \,;\, (i = 1, 2, ..., k; j = 1, 2, ..., n_i) \qquad ...(5·3)$$

where α_i's are fixed (unknown) constants and all the symbols have been explained in (5·2) to (5·2*d*).

Assumptions in Model (5·3).

(*i*) All the observations (y_{ij}'s) are independent and $y_{ij} \sim N(\mu_i, \sigma_e^2)$.

(*ii*) Different effects are additive in nature.

(*iii*) ε_{ij} are *i.i.d.* $N(0, \sigma_0^2)$, *i.e.*, $E(\varepsilon_{ij}) = 0$ and $V(\varepsilon_{ij}) = 0 \ \forall\, i$ and j. ...(5·3*a*)

Under the third assumption, the model (5·3) becomes :

$$E(y_{ij}) = \mu_i = \mu + \alpha_i \,; \quad (i = 1, 2, ..., k; \quad j = 1, 2, ..., n_i). \qquad ...(5·4)$$

Statistical Analysis of Model (5·3)

Null Hypothesis. We want to test the equality of the population means, *i.e.*, the homogeneity of different rations. Hence, null hypothesis is given by :

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k = \mu \qquad ...(5·4a)$$

which from (5·2*b*) reduces to

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 0 \qquad ...(5·4b)$$

Alternative Hypothesis. At least two of the means $\mu_1, \mu_2, ..., \mu_k$ are different.

Let us write :

$$\left.\begin{aligned} \bar{y}_{i\cdot} &= \text{Mean of the } i\text{th class} = \sum_{j=1}^{n_i} y_{ij}/n_i \,;\, (i = 1, 2, ..., k) \\ \text{and} \quad \bar{y}_{\cdot\cdot} &= \text{Over all mean} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{N}\sum_{i=1}^{k} n_i \bar{y}_{i\cdot} \end{aligned}\right\} \qquad ...(5·4c)$$

Least Square Estimates of Parameters. The parameters μ and α_i in (5·3) are estimated by the principle of least squares on minimising the error (residual) sum of squares given by :

$$E = \sum_i \sum_j \varepsilon_{ij}^2 = \sum_i \sum_j (y_{ij} - \mu - \alpha_i)^2$$

The normal equations for estimating μ and α_i are :

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

From (*), we get

$$\sum_i \sum_j y_{ij} - N\mu - \sum_i n_i\alpha_i = 0 \ \Rightarrow \ \hat{\mu} = \frac{1}{N}\sum\sum y_{ij} = \bar{y}_{\cdot\cdot} \quad \left[\because \sum_{i=1}^{k} n_i\alpha_i = 0, \text{ using } (5·2c)\right] \qquad ...(5·5)$$

From (**), we get

$$\sum_j y_{ij} - n_i\hat{\mu} - n_i\hat{\alpha} = 0 \quad \Rightarrow \quad \hat{\alpha} = \frac{1}{n_i}\sum_j y_{ij} - \hat{\mu} = \bar{y}_{i\cdot} - \hat{\mu}, \ i.e., \ \hat{\alpha} = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} \ ...(5·5a)$$

Hence, substituting in (5·3), the model becomes :

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

↓ Observation ↓ (Grand Mean) ↓ (Deviation due to ith treatment) ↓ (Residual or error component)

Partitioning of Sum of Squares. 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 $\bar{y}_{..}$ to the left, *squaring both* sides and summing over i and j, we get

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

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

But $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.}) = 0$, since the algebraic sum of the deviations of the rations from their mean is zero.

$$\therefore \quad \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \qquad \dots (5·6)$$

$S_T^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$ is known as total sum of squares (T.S.S.) ;

$S_E^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$ is called within sum of squares or error sum of squares (S.S.E.); and

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

Then, Total S.S. = S.S.E. + S.S.T. ... (5·6a)

TABLE 5·2 : ANOVA TABLE FOR ONE-WAY CLASSIFIED DATA

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

***Remarks* 1.** We have seen above that although S_t^2 and S_E^2 add up to S_T^2 ... also additive, their mean squares ...

TABLE 5·13 : TWO-WAY CLASSIFIED DATA

Treatments (Rations)	Varieties of Cows 1	2	...	j	...	h	Row Totals $= \left(\sum_j y_{ij}\right)$	Row Means $= \left(\sum_j y_{uj}\right)/h$
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1h}	$T_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2h}	$T_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ih}	$T_{i\cdot}$	$\bar{y}_{i\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kh}	$T_{k\cdot}$	$\bar{y}_{k\cdot}$
Column Totals	$T_{\cdot 1}$	$T_{\cdot 2}$	...	$T_{\cdot j}$	...	$T_{\cdot h}$	$G = \sum\sum y_{ij}$	
Column Means $= \left(\sum_i y_{ij}\right)/k$	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	...	$\bar{y}_{\cdot j}$	...	$\bar{y}_{\cdot h}$		

5·3·1. ANOVA for Fixed Effect Model. If we assume that, in the above discussion, for both the factors the levels used are the only ones of interest, then the *fixed effect or parametric model,* is used.

Factor A : Treatments (Rations)

Factor B : Variety (Breed and Stock) of cow.

In the above illustration, the fixed effect model is :

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} \quad \Rightarrow \quad E(y_{ij}) = \mu_{ij} \; ; (i = 1, 2, ..., k \; ; \; j = 1, 2, ..., h) \qquad ... (5·48)$$

where y_{ij} are independent $N(\mu_{ij}, \sigma_e^2)$ and ε_{ij} are *i.i.d.* $N(0, \sigma_e^2)$ $\forall$ i, j.

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

(*i*) The general mean effect μ given by : $\mu = \sum_i \sum_j \mu_{ij}/N$. ... (5·49)

(*ii*) The effect α_i, ($i = 1, 2, ..., k$) due to the ith ration given by : $\alpha_i = \mu_{i\cdot} - \mu$, ... (5·50)

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

Obviously $$\sum_{i=1}^{k} \alpha_i = 0 \qquad ... (5·50a)$$

(*iii*) The effect β_j, ($j = 1, 2, ..., h$) due to the jth variety (breed of cow) given by :

$$\beta_j = \mu_{\cdot j} - \mu \text{, where } \mu_{\cdot j} = \frac{1}{k} \sum_{i=1}^{k} \mu_{ij}, (j = 1, 2, ..., h) ... (5·51)$$

Obviously, $$\sum_{j=1}^{h} \beta_j = 0 \qquad ... (5·51a)$$

(*iv*) The interaction effect γ_{ij} when the ith level of first factor (rations) and jth level of second factor (breed of cow) occur simultaneously and is given by :

$$\gamma_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu \qquad ... (5·52)$$

where $$\sum_j \gamma_{ij} = 0 \; \forall \; i = 1, 2, ..., k \quad \text{and} \quad \sum_i \gamma_{ij} = 0 \; \forall \; j = 1, 2, ..., h \qquad ... (5·52a)$$

Thus, we have

$$\mu_{ij} = \mu + (\mu_{i\cdot} - \mu) + (\mu_{\cdot j} - \mu) + (\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu)$$

and consequently the model (5·16) becomes

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij} \qquad \ldots (5·53)$$

where ε_{ij} is the error effect due to chance and

$$\sum_{i=1}^{k} \alpha_i = 0 = \sum_{j=1}^{h} \beta_j \,; \quad \sum_{i=1}^{k} \gamma_{ij} = 0 \;\forall\, j \;; \quad \sum_{j=1}^{h} r_{ij} = 0 \;\forall\, i \qquad \ldots (5·53a)$$

As there is only one observation in each cell, the observation corresponding to the ith level of ration and jth level of breed of cow is only one, *i.e.*, y_{ij}. But we cannot estimate by one value alone. Hence, in this case (one observation per cell), the interaction effect $\gamma_{ij} = 0$ and the model (5·18) reduces to

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad \ldots (5·54)$$

The model (5·54), where α_i, β_j and μ are fixed (unknown) constants and, ε_{ij} and x_{ij} are random variables is known as the *fixed effect model* for two-way classified data with one observation per cell.

Statistical Analysis of the Fixed Effect Model (5·54). Let us write :

$$\bar{y}_{i\cdot} = \text{Mean yield of the } i\text{th treatment (ration)} = \frac{1}{h} \sum_{j=1}^{h} y_{ij} = \frac{T_{i\cdot}}{h}; \quad (i = 1, 2, \ldots, k) \qquad \ldots (5·55)$$

$$\bar{y}_{ij} = \text{Mean yield of the } j\text{th variety} = \frac{1}{k} \sum_{i=1}^{k} y_{ij} = \frac{T_{\cdot j}}{k}\,; \quad j = 1, 2, \ldots, h \qquad \ldots (5·55a)$$

$$\bar{y}_{\cdot\cdot} = \text{The overall mean} = \frac{1}{hk} \sum_i \sum_j y_{ij} = \frac{G}{N} = \frac{1}{k} \sum_i \bar{y}_{i\cdot} = \frac{1}{h} \sum_j \bar{y}_{\cdot j} \qquad \ldots (5·55b)$$

Least Square Estimates of Parameters. The least square estimates of the parameters μ, α_i and β_j are obtained on minimizing the error sum of squares :

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

The normal equations for estimating μ, α_i and β_j are respectively :

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

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

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

Since $\sum_i \alpha_i = 0 = \sum_i \beta_j$, we get from the above equations :

$$\hat{\mu} = \frac{1}{hk} \sum_i \sum_j y_{ij} = \bar{y}_{\cdot\cdot} \;;\; \hat{\alpha}_i = \frac{1}{h} \sum_j y_{ij} - \hat{\mu} = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} \;;\; \hat{\beta}_j = \frac{1}{k} \sum_i y_{ij} - \hat{\mu} = \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot} \qquad \ldots (5·56)$$

Thus the linear model (5·19) becomes

$$y_{ij} = \bar{y}_{\cdot\cdot} + (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}) \qquad \ldots (5·56a)$$

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

Partitioning of the Sum of Squares. Transposing $\bar{y}_{..}$ to the left side, squaring and summing both sides over i from 1 to k and j from 1 to h, we get

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

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

$$+ 2\sum_i\sum_j(\bar{y}_{i.}-\bar{y}_{..})(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..}) + 2\sum_i\sum_j(\bar{y}_{.j}-\bar{y}_{..})(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})$$

$$+ 2\sum_i\sum_j(\bar{y}_{i.}-\bar{y}_{..})(\bar{y}_{.j}-\bar{y}_{..})$$

Now $$\sum_i\sum_j(\bar{y}_{i.}-\bar{y}_{..})(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..}) = \sum_i\left[(\bar{y}_{i.}-\bar{y}_{..})\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})\right]$$

$$= \sum_i\left[(\bar{y}_{i.}-\bar{y}_{..})\left\{\sum_j(y_{ij}-\bar{y}_{i.})-\sum_j(\bar{y}_{.j}-\bar{y}_{..})\right\}\right]=0,$$

since algebraic sum of deviations of a set of observations about their mean is zero.

Similarly it can be easily seen that other product terms also vanish.

$$\therefore \quad \sum_i\sum_j(y_{ij}-\bar{y}_{..})^2 = h\sum_i(\bar{y}_{i.}-\bar{y}_{..})^2 + k\sum_j(\bar{y}_{.j}-\bar{y}_{..})^2 + \sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2 \quad \ldots (5\cdot57)$$

or $\quad S_T^2 = S_t^2 + S_V^2 + S_E^2$

where $\quad S_T^2 = \sum_i\sum_j(y_{ij}-\bar{y})^2$ is the *Total S.S.*

$S_t^2 = h\sum_i(\bar{y}_{i.}-\bar{y}_{..})^2$ is *S.S.* due to *treatments,*

$S_V^2 = k\sum_j(\bar{y}_{.j}-\bar{y}_{..})^2$ is the *S.S. due to varieties,*

and $\quad S_E^2 = \sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$ *is the error or residual S.S.*

Null Hypotheses. We set up the null hypotheses that the treatments as well as varieties are homogeneous. In other words, the null hypotheses for treatments and varieties are respectively :

$$H_{ot}:\ \mu_{1.}=\mu_{2.}=\ldots=\mu_{k.}=\mu\ ;\ H_{ov}:\mu_{.1}=\mu_{.2}=\ldots=\mu_{.h}=\mu \quad \ldots (5\cdot58)$$

or and (5·14*b*), their equivalents :

$$H_{ot}:\ \alpha_1=\alpha_2=\ldots=\alpha_k=0\ ;\ H_{ov}:\ \beta_1=\beta_2=\ldots=\beta_h=0 \quad \ldots (5\cdot58a)$$

Alternative Hypotheses

H_{1t} : At least two of the $\mu_{i.}$'s are different ; H_{1v} : At least two of the $\mu_{.j}$'s are different. or their equivalents :

H_{1t} : At least one of the α_i's is not zero ; H_{1v} : At least one of β_j's is not zero.

Degrees of Freedom for Various S.S. The total S.S., S_T^2 being computed from $N = hk$ quantities $(y_{ij}-\bar{y}_{..})$ which are subject to one linear constraint $\sum_i\sum_j(y_{ij}-\bar{y}_{..})=0$ will carry $(N-1)$ *d.f.* Similarly, S_t^2 will have $(k-1)$ *d.f.*, since $\sum_i(\bar{y}_{i.}-\bar{y}_{..})=0$ and S_V^2 will have $(h-1)$ *d.f.*, since $\sum_j(\bar{y}_{.j}-\bar{y}_{..})=0$ and s_E^2 will carry $(N-1)-(k-1)-(h-1)=(h-1)(k-1)$ *d.f.* $(\because N=hk)$.

TABLE 5·13 : ANOVA TABLE FOR TWO-WAY CLASSIFIED DATA (ONE OBSERVATION PER CELL)

Source of Variation	*Sum of Squares*	*d.f.*	*Mean Sum of Squares*	*Variance Ratio*
Treatments (Rations)	$S_t^2 = h \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2$	$k-1$	$s_t^2 = \frac{S_t^2}{(k-1)}$	$F_t = \frac{s_t^2}{s_E^2} \sim F[k-1, (h-1)(k-1)]$
Varieties (Breeds of cows)	$S_V^2 = k\sum (\bar{y}_{.j} - \bar{y}_{..})^2$	$h-1$	$s_V^2 = \frac{S_V^2}{(h-1)}$	$F_V = \frac{s_V^2}{s_E^2} \sim F[h-1, (h-1)(k-1)]$
Residual	$S_E^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$(h-1) \times (k-1)$	$s_E^2 = \frac{S_E^2}{(h-1)(k-1)}$	
Total	$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$	$hk-1$		

Mathematical model

Factor A : Treatments (Rations)

Factor B : variety (Breed of cow)

$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

where y_{ij} the observation corresponding to the ith level of ration and jth level of breed of cow. $i = 1, 2 .. k$, $j = 1, 2 .. h$

μ is the general mean effect i.e $\mu = \sum_i \sum_j \frac{\mu_{ij}}{N}$

α_i, β_j and ϵ_{ij}

The effect α_i, $i = 1, 2 .. k$ due to the ith ration are constants $\sum_{i=1}^{k} \alpha_i = 0$

The effect β_j, $j = 1, 2 .. h$ due to the jth variety (breed of cow) are constants. $\sum_{j=1}^{h} \beta_j = 0$

ϵ_{ij} are i.i.d $N(0, \sigma_0^2)$

Assumptions.

1. All the observations are independent.
2. Different effects are additive in nature
3. $\epsilon_{ij} \sim N(0, \sigma_0^2)$

Statistical analysis

H_{01}: Treatments are same

H_{02}: Varieties are same

$\bar{y}_{i.}$ = mean yield of the ith treatment

$= \frac{1}{h} \sum_{j=1}^{h} y_{ij} = \frac{T_{i.}}{h}$ $i = 1, 2 .. k$

$\bar{y}_{.j}$ = mean yield of the j^{th} variety

$= \frac{1}{k} \sum_{i=1}^{k} y_{ij} = \frac{T_{.j}}{k} \quad j = 1, 2 \ldots h$

$\bar{Y}_{..}$ = The overall mean $= \frac{1}{hk} \sum_i \sum_j y_{ij} = \frac{G}{N}$

The parameters μ, α_i and β_j are obtained on minimizing the error sum of squares.

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

The normal equations for estimating the parameters μ, α_i and β_j are obtained respectively

$$\frac{\partial E}{\partial \mu} = 0 \Rightarrow -2 \sum_i \sum_j (y_{ij} - \mu - \alpha_i - \beta_j) \quad \text{—} (1)$$

$$\frac{\partial E}{\partial \alpha_i} = 0 \Rightarrow -2 \sum_j (y_{ij} - \mu - \alpha_i - \beta_j) \quad \text{—} (2)$$

$$\frac{\partial E}{\partial \beta_j} = 0 \Rightarrow -2 \sum_i (y_{ij} - \mu - \alpha_i - \beta_j) \quad \text{—} (3)$$

Since $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$ we get

$$\sum_i \sum_j y_{ij} - \sum_i \sum_j \mu - \sum_j \left(\sum_i \alpha_i \right) - \sum_i \left(\sum_j \beta_j \right) = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{hk} \sum_i \sum_j y_{ij} = \bar{Y}_{..}$$

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

i.e by (2) $\sum_j (y_{ij} - \mu - \alpha_i - \beta_j) = 0$

$\sum_j y_{ij} - h\mu - h\alpha_i - \sum_j \beta_j = 0$

$\Rightarrow \sum_j y_{ij} - h(\mu) = h\alpha_i$

$\Rightarrow \frac{1}{h} \sum_j y_{ij} - \hat{\mu} = \hat{\alpha}_i$

$\Rightarrow \bar{y}_{i.} - \bar{y}_{..}$

by (3) $\hat{\beta}_j = \frac{1}{k} \sum_i y_{ij} - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..}$

The linear model becomes.

$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$

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

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

T.S.S = Treatment S.S + S.S due to varieties + SSE.

Since algebraic sum of the deviations of observations about their mean is zero.

ANOVA table.

Sources of variation	d.o.f	S.S	M.S.S	F-ratio
Factor A	$k-1$	SSA	$SSA/k-1$ — ①	$F_1 = \frac{①}{③}$
Factor B	$h-1$	SSB	$SSB/h-1$ — ②	$F_2 = \frac{②}{③}$
Error	$(h-1)(k-1)$	SSE	$\frac{SSE}{(h-1)(k-1)}$ — ③	
Total	$hk-1$			

$F_1 \sim F_{(k-1),\ (h-1)(k-1)}$

$F_2 \sim F_{(h-1),\ (h-1)(k-1)}$

$$hk-1-(k-1)-(h-1) = hk-1-k+1-h+1$$
$$= k(h-1)-(h-1)$$
$$= (k-1)(h-1)$$