Two Way Analysis of Variance ①

Suppose 'N' observations are classified into 'K' Categories (or classes), say $A_1, A_2, \ldots A_K$ according to some criterion, A and into 'h' Categories, say, $B_1, B_2 \ldots B_h$ according to some criterion B, having Kh Combinations (A_i, B_j); $i = 1, 2, \ldots K$; $j = 1, 2, \ldots h$; often called Cells. This Scheme of classification according to two factors or criteria is called two-way classification and its analysis is called two way Analysis of Variance. The number of observations in each cell may be equal or different, but here we shall consider the case of one observation per cell so that $N = hK$.

The linear model for analysis of two way classified data is

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij} \quad \cdots (*)$$

where, μ = general effect

α_i = effect due to the ith level of factor 'A'

β_j = effect due to the j-th level of factor B

γ_{ij} = the interaction effect due to ith level of factor A and jth level of factor B.

e_{ij} = Errors $\sim N(0, \sigma_e^2)$ (2)

Moreover, $\sum_i \alpha_i = 0 = \sum_j \beta_j$

and $\sum_{i=1}^{K} \gamma_{ij} = \sum_{j=1}^{h} \gamma_{ij} = 0$

As there is only one observation in each cell, γ_{ij} can not be estimated ~~calculated~~ by one value alone.

Hence $\gamma_{ij} = 0$ and (*) reduces to

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

Estimation of parameters:

μ, α_i and β_j are obtained by the method of least squares by minimizing $E = \sum_i \sum_j e_{ij}^2$

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

The normal equations will give the estimates of μ, α_i and β_j

i.e. $\frac{\partial E}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \bar{y}_{..}$

$$\frac{\partial E}{\partial \alpha_i} = 0 \Rightarrow \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$$

and $$\frac{\partial E}{\partial \beta_j} = 0 \Rightarrow \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

where, $\bar{y}_{..}$ = grand mean

$\bar{y}_{i.}$ = mean of i^{th} level of factor 'A'

$\bar{y}_{.j}$ = mean of j^{th} level of factor 'B'.

Partitioning the total Sum of Squares:

After estimating μ, α_i and β_j ~~the~~ (*) becomes,

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

$[\because e_{ij} = y_{ij} - \mu - \alpha_i - \beta_j]$

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

Squaring both sides and summing over i and j we get, 4

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

[Product term vanishes as since sum of deviation from mean is zero].

$$\therefore TSS = SSA + SSB + SSE$$

where, $TSS = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$

$SSA = \sum_i h(\bar{y}_{i.} - \bar{y}_{..})^2$

$SSB = \sum_j k(\bar{y}_{.j} - \bar{y}_{..})^2$

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

Partitioning the degrees of freedom of Sum of Squares:

The total d.f. is partitioned as Total d.f = d.f due to factor 'A' + d.f. due to factor 'B' + d.f. due to Error.

or, $(hK-1) = (K-1) + (h-1) + (K-1)(h-1)$

~~Mean Square of~~ Squares

Null Hypothesis:

$H_{0A}: \alpha_1 = \alpha_2 = \cdots = \alpha_K = 0$
i.e. there is no significant difference between different levels of factor 'A'

$H_{0B}: \beta_1 = \beta_2 = \cdots = \beta_h = 0$
i.e. there is no significant difference between different levels of factor 'B'

Alternative Hypothesis:

H_{1A}: at least two of the levels are different for factor 'A'

H_{1B}: at least two of the levels are different for factor 'B'

Mean Squares:

Mean squares for factor 'A' $(MSA) = \dfrac{SSA}{K-1}$

Mean Squares for factor 'B' $(MSB) = \dfrac{SSB}{h-1}$

Mean Squares of Errors $(MSE) = \dfrac{SSE}{(K-1)(h-1)}$

Test statistic:

$$F_A = \frac{MSA}{MSE} \sim F_{(k-1),(k-1)(h-1)} \text{ d.f}$$

$$F_B = \frac{MSB}{MSE} \sim F_{(h-1),(k-1)(h-1)} \text{ d.f}.$$

Decision:

If $F_A \geq F_{\alpha, k-1, (h-1)(k-1)}$ we reject H_{0A} otherwise we accept H_{0A}:

If $F_B \geq F_{\alpha, (h-1), (h-1)(k-1)}$ we reject H_{0B}; otherwise we accept H_{0B}.

ANOVA TABLE

Sources of variation	df	SS	MS	F
Due to factor 'A'	$(k-1)$	SSA	MSA	$F_A = \frac{MSA}{MSE}$
Due to factor 'B'	$(h-1)$	SSB	MSB	$F_B = \frac{MSB}{MSE}$
Error	$(k-1)(h-1)$	SSE	MSE	
Total	$(hk-1)$	TSS	—	—