ST-302
Design, Planning and Analysis of Experiments

Dr. (Mrs.) Kirtee Kiran Kamalja
Department of Statistics, School of Mathematical Sciences,
Kavayitri Bahinbai Chaudhari North Maharashtra University,
Jalgaon

Topic 2: Two-way Classification Model

About the experiments for two-way classification model

Objective: To study the effect of two factors in the response variable.

Factor	No. of Levels	Levels
A	p	$A_1, A_2, \dots, A_p$
B	q	$B_1, B_2, \dots, B_q$

Experiments: Corresponding to all possible combinations of levels of factor A and B are to be performed. That is the combination of $(A_i, B_j), i = 1,2, \dots, p, j = 1,2, \dots, q$ each is allocated randomly to experimental units and pq experiments are performed in random order.

Data: The results of the experiment can be arranged in two-way table

Data…

Factor	B_1	B_2	…	B_q	Sums	Averages
A_1	y_{11}	y_{12}	…	y_{1q}	$y_{1.}$	$\bar{y}_{1.}$
A_2	y_{21}	y_{22}	…	y_{2q}	$y_{2.}$	$\bar{y}_{2.}$
:	:	:	…	:	…	…
A_p	y_{p1}	y_{p2}	…	y_{pq}	$y_{p.}$	$\bar{y}_{p.}$
Sums	$y_{.1}$	$y_{.2}$	…	$y_{.q}$	$y_{..}$	
Averages	$\bar{y}_{.1}$	$\bar{y}_{.2}$	…	$\bar{y}_{.q}$		$\bar{y}_{..}$

Example 1: Two-way classification model

A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The

	Bolt				
Chemical	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

Two-way classification model

Model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1,2, \dots, p; j = 1,2, \dots, q$$

y_{ij}: i th observation receiving j th treatment

μ: common effect

α_i: effect due i th treatment

β_j: effect due j th block

ε_{ij}: Random error component

Assumptions: $\varepsilon_{ij} \sim IIDNormal(0, \sigma^2)$

Assumptions

- Mean error is zero i.e $E(\varepsilon_{ij}) = 0$
- Variance of errors is constant i.e $var(\varepsilon_{ij}) = \sigma^2 \; for \; all \; i, j$
- Covariances between errors is zero $var(\underline{\varepsilon}) = \sigma^2 I_n$
- $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$

Implications

- $E(y_{ij}) = \mu + \alpha_i + \beta_j$
- $var(y_{ij}) = \sigma^2$
- y_{ij} are independently distributed but not identical
- $y_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$

Model details

- Number of observations=$n = pq$
- Number of parameters=$p + q + 1$

$$\mu,$$

$$\alpha_1, \alpha_2, \dots, \alpha_p,$$

$$\beta_1, \beta_2, \dots, \beta_q$$

- Here $pq > 1 + p + q$ ($n > p$ assumption in GLM)

Derivation of normal equations

- $\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$ is called fitted values (by model)

- Define errors as:

  Residuals/error, $e_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$

- Obtain $\hat{\mu}, \hat{\alpha}_i,\ \hat{\beta}_j$ such that the errors/error sum of squares is minimum.

- It will lead to $\hat{\mu}, \hat{\alpha}_i,\ \hat{\beta}_j$ such that y_{ij} (observed value) and $\hat{y}_{ij}$ (fitted value) close to each other.

Derivation of normal equations…

Minimize function ϕ with respect to $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j$

$$\phi = \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right)^2$$

$$\frac{d\phi}{d\hat{\mu}} = -2 \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right) \qquad (1)$$

$$\frac{d\phi}{d\hat{\alpha}_i} = -2 \sum_{j=1}^{q} \left(y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right) \qquad i = 1,2,\ldots,p \quad \text{(A)}$$

$$\frac{d\phi}{d\hat{\beta}_j} = -2 \sum_{i=1}^{p} \left(y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right) \qquad j = 1,2,\ldots,q \quad \text{(B)}$$

Derivation of normal equations…

(1)⇒ $$y_{..} = pq\hat{\mu} + q\sum_{i=1}^{p}\hat{\alpha}_i + p\sum_{j=1}^{q}\hat{\beta}_j \quad (2)$$

(A)⇒ $$y_{i.} = q\hat{\mu} + q\hat{\alpha}_i + \sum_{j=1}^{q}\hat{\beta}_j, \qquad i = 1,2,\dots,p \quad (C)$$

(B)⇒ $$y_{.j} = p\hat{\mu} + \sum_{i=1}^{p}\hat{\alpha}_i + p\hat{\beta}_j, \qquad j = 1,2,\dots,q \quad (D)$$

- Observe that (2), (C) and (D) are $1 + p + q$ equations in $1 + p + q$ variables
- These equations are called normal equations.
- But all are not linearly independent.
- Hence the solution to normal equations would not be unique.

Solution of normal equations

- Only $(1+p+q)-2$ of these are linearly independent as
  - ❖ $\sum_{i=1}^{p}(C)=(2)$ i.e. addition of p equations in (C) gives (2)
  - ❖ $\sum_{j=1}^{q}(D)=(2)$ i.e. addition of q equations in (D) gives (2)
- Hence we need additional two equations which are linearly independent with (1), (C) and (D).
- Let these two equations be:
  - ❖ $\sum_{i=1}^{p}\hat{\alpha}_i=0 \qquad (3)$
  - ❖ $\sum_{j=1}^{q}\hat{\beta}_j=0 \qquad (4)$

Solution of normal equations…

Using (3) and (4) in (2), (C) and (D),

(2)⇒ $\hat{\mu} = \bar{y}_{..}$

(C)⇒ $\hat{\alpha}_i = \bar{y}_{i.} - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..}$, $i = 1,2,\ldots,p$

(D)⇒ $\hat{\beta}_j = \bar{y}_{.j} - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..}$, $j = 1,2,\ldots,q$

Note:

- These are known as solutions of normal equations and not estimates of the respective parameters.
- Since the model in NFRM, individual parameters are not estimable.
- Only few linear parametric functions are estimable

Rank of Estimation Space

- Estimation space: It is collection of all *lpfs* which are estimable.

  $\rho(estimation\ space)$

  =number of linearly independent normal equations

  = Number of *linearly independent estimable lpfs*

  $= p + q - 1$

- It means that there would be only $p + q - 1$ *epfs* which would be *l.i.*

Rank of error space

- Error space: This is the space which is orthogonal to estimation space and contain all unbiased estimators of zero (representing errors)

  $\rho(Error\ space)$

  $= n - \rho(estimation\ space)$

  $= pq - (p + q - 1)$

  $= (p - 1)(q - 1)$

- Here $n = pq$ which are number of observations.

Estimability of linear parametric function in $\alpha's$…

Some Definitions:

- **Contrasts**: The parametric function whose coefficients add to zero is called contrast.
- **Elementary contrasts**: The parametric function which involve only two parameters with unit coefficient with opposite signs is called elementary contrasts.

Note:

- Every elementary contrasts is contrast but the converse is not true.
- Every contrasts can be expressed as linear combination of elementary contrasts.
- For example, $\alpha_1 - \alpha_3$ is elementary contrast and $\alpha_1 - 2\alpha_2 + \alpha_3$ is contrast. Observe that : $\alpha_1 - 2\alpha_2 + \alpha_3 =(\alpha_1 - \alpha_2)$-$(\alpha_2 - \alpha_3)$

Estimability condition of linear parametric functions

Equations (2), (C) and (D) can also be written as follows.

$$E(y_{..}) = pq\mu + q\sum_{i=1}^{p}\alpha_i + p\sum_{j=1}^{q}\beta_j \qquad (2)$$

$$\mathrm{E}(y_{i.}) = q\mu + q\alpha_i + \sum_{j=1}^{q}\beta_j, \quad i = 1,2,\dots,p \qquad (C)$$

$$\mathrm{E}(y_{.j}) = p\mu + \sum_{i=1}^{p}\alpha_i + p\beta_j, \quad j = 1,2,\dots,q \qquad (D)$$

(2)$\Rightarrow$ $E(\bar{y}_{..}) = \mu + \frac{1}{p}\sum_{i=1}^{p}\alpha_i + \frac{1}{q}\sum_{j=1}^{q}\beta_j$

$\Rightarrow$ $\mu + \frac{1}{p}\sum_{i=1}^{p}\alpha_i + \frac{1}{q}\sum_{j=1}^{q}\beta_j$ is estimable parametric function

$\Rightarrow$ $\mu + \bar{\alpha} + \bar{\beta}$ is an estimable parametric functions

Estimability of linear parametric functions in $\alpha's$

(C)$\Rightarrow$ $\quad E(\bar{y}_{i.}) = \mu + \alpha_i + \frac{1}{q}\sum_{j=1}^{q}\beta_j, \quad i = 1,2,\dots,p$

$$E(\bar{y}_{i.}) = \mu + \alpha_i + \bar{\beta}, \quad i = 1,2,\dots,p$$

Consider the pair of equations from (C) for $i \neq u$ as follows.

$$E(\bar{y}_{i.}) = \mu + \alpha_i + \bar{\beta}$$

$$E(\bar{y}_{u.}) = \mu + \alpha_u + \bar{\beta}$$

Subtracting these we get:

$$E(\bar{y}_{i.} - \bar{y}_{u.}) = \alpha_i - \alpha_u, \qquad\qquad i \neq u$$

- Thus $\alpha_i - \alpha_u$, is estimable for all $i \neq u$.

Estimability of linear parametric function in $\alpha's$…

- All elementary contrasts in $\alpha's$ are estimable.
- All contrasts in $\alpha's$ are estimable
- Thus, $\sum_{i=1}^{p} c_i\alpha_i$ is estimable if $\sum_{i=1}^{p} c_i = 0$
- $\alpha_1 - 2\alpha_2 + \alpha_3$ and $\alpha_1 - 2\alpha_2 - \alpha_3 + 2\alpha_4$ are estimable.
- While $\alpha_1 + \alpha_2$, $\alpha_1 - 2\alpha_3$ are not estimable.

Estimability of linear parametric functions in $\beta's$

(D)⇒ $E(\bar{y}_{.j}) = \mu + \frac{1}{p}\sum_{i=1}^{p}\alpha_i + \beta_j, \quad j = 1,2,\dots,q$

$$E(\bar{y}_{.j}) = \mu + \bar{\alpha} + \beta_j, \qquad j = 1,2,\dots,q$$

Consider the pair of equations from (D) for $j \neq v$ as follows.

$$E(\bar{y}_{.j}) = \mu + \bar{\alpha} + \beta_j$$

$$E(\bar{y}_{.v}) = \mu + \bar{\alpha} + \beta_v$$

Subtracting these we get:

$$E(\bar{y}_{.j} - \bar{y}_{.v}) = \beta_j - \beta_v, \qquad j \neq v$$

- Thus$\beta_j - \beta_v$, is estimable for all $j \neq v$.

Estimability of linear parametric function in $\beta's$…

- All elementary contrasts in $\beta's$ are estimable.
- All contrasts in $\beta's$ are estimable
- Thus, $\sum_{j=1}^{q} d_j\beta_j$ is estimable if $\sum_{j=1}^{q} d_j = 0$
- $\beta_1 - 2\beta_2 + \beta_3$ and $\beta_1 - 2\beta_2 - \beta_3 + 2\beta_4$ are estimable.
- While $\beta_1 + \beta_2$, $\beta_1 - 2\beta_3$ are *not* estimable.

Summary of estimability conditions of l. parametric functions

1. $\mu + \bar{\alpha} + \bar{\beta}$ is estimable (A *lpf* which involve all parameters)
2. All contrasts in $\alpha's$ are estimable (A *lpf* which involve only $\alpha's$)
3. All contrasts in $\beta's$ are estimable (A *lpf* which involve only $\beta's$)

- Only $p-1$ contrasts in $\alpha's$ are linearly independent.
- Only $q-1$ contrasts in βs are linearly independent.
- Thus there are only $1+(p-1)+(q-1)$ linearly independent estimable parametric function
- Justify estimablity condition (2) and (3).

One set of linearly independent e.p.f.

- $\mu + \bar{\alpha} + \bar{\beta}$ $\quad (1)$
- $\alpha_1 - \alpha_2$
- $\alpha_1 - \alpha_3$
- $\vdots$ $\quad (p-1)$
- $\alpha_1 - \alpha_p$
- $\beta_1 - \beta_2$
- $\beta_1 - \beta_3$
- $\vdots$ $\quad (q-1)$
- $\beta_1 - \beta_q$

linearly independent e.p.fs. $(p+q-1)$

BLUEs and Variance(BLUE) of *epf*

Result: LHS of normal equations are BLUE of expected value of their RHS

- In GLM $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ normal equations are

$$X'\underline{Y} = X'X\underline{\hat{\beta}}$$

- $X'\underline{Y}$ =LHS of normal equations and
- $X'X\underline{\hat{\beta}}$ =RHS of normal equations and
- E(RHS of normal equations)=$E(X'\underline{Y}) = X'X\underline{\beta}$
- Thus $X'\underline{Y} = X'X\underline{\hat{\beta}}$ is BLUE of its expected value i.e. $X'X\underline{\beta}$

BLUEs and Variance(BLUE) of *epf*

- BLUE of $\mu + \bar{\alpha} + \bar{\beta}$ which is estimable.

Hence $\widehat{\mu + \bar{\alpha} + \bar{\beta}} = \bar{y}_{..}$ as $E(\bar{y}_{..}) = \mu + \bar{\alpha} + \bar{\beta}$

- Variance

Variance(BLUE)

$= \mathrm{var}\left(\widehat{\mu + \bar{\alpha} + \bar{\beta}}\right)$

$= \mathrm{var}(\bar{y}_{..})$

$= \dfrac{\sigma^2}{pq}$

BLUEs and Variance(BLUE) of *epf*

- $\beta_j - \beta_v$, is estimable for all $j \neq v$.
- Further $E(\bar{y}_{.j} - \bar{y}_{.v}) = \beta_j - \beta_v$, $\qquad j \neq v$
- Hence for $j \neq v$

BLUE of $\beta_j - \beta_v$

$= \widehat{\beta_j - \beta_v}$

$= \bar{y}_{.j} - \bar{y}_{.v}$

BLUEs and Variance(BLUE) of *epf*

- Variance(BLUE)

$$= \mathrm{var}(\widehat{\beta_j - \beta_v})$$

$$= \mathrm{var}\left(\bar{y}_{.j} - \bar{y}_{.v}\right)$$

$$= \mathrm{var}\left(\bar{y}_{.j}\right) + \mathrm{var}(\bar{y}_{.v}) - 2cov\left(\bar{y}_{.j}, \bar{y}_{.v}\right)$$

$$= \frac{\sigma^2}{p} + \frac{\sigma^2}{p} - 2 \times 0$$

$$= \frac{2\sigma^2}{p}$$

BLUEs and Variance(BLUE) of *epf*

- In general $\sum_{j=1}^{q} d_j\beta_j$ is estimable if $\sum_{j=1}^{q} d_j = 0$
- BLUE of $\sum_{j=1}^{q} d_j\beta_j$

$$= \widehat{\sum_{j=1}^{q} d_j}\beta_j$$

$$= \sum_{j=1}^{q} d_j\, \hat{\beta}_j$$

$$= \sum_{j=1}^{q} d_j \left(\bar{y}_{.j} - \bar{y}_{..}\right)$$

$$= \sum_{j=1}^{q} d_j\, \bar{y}_{.j}$$

BLUEs and Variance(BLUE) of *epf*

- Variance(BLUE of $\sum_{j=1}^{q} d_j\beta_j$)

$$= \operatorname{var}\left(\widehat{\sum_{j=1}^{q} d_j\beta_j}\right)$$

$$= \operatorname{var}\left(\sum_{j=1}^{q} d_j\, \bar{y}_{.j}\right)$$

$$= \sum_{j=1}^{q} d_j^2 \operatorname{var}\left(\bar{y}_{.j}\right)$$

$$= \sum_{j=1}^{q} d_j^2\, \frac{\sigma^2}{p}$$

$$= \frac{\sigma^2}{p}\sum_{j=1}^{q} d_j^2$$

BLUEs and Variance(BLUE) of *epf*

- $\alpha_i - \alpha_u$, is estimable for all $i \neq u$.
- Further $\mathrm{E}(\bar{y}_{i.} - \bar{y}_{u.}) = \alpha_i - \alpha_u, \qquad\qquad i \neq u$
- Hence for $i \neq u$

BLUE of $\alpha_i - \alpha_u$

$= \widehat{\alpha_i - \alpha_u}$

$= \bar{y}_{i.} - \bar{y}_{u.}$

BLUEs and Variance(BLUE) of *epf*

- Variance(BLUE)

$$= \operatorname{var}(\widehat{\alpha_i - \alpha_u})$$

$$= \operatorname{var}(\bar{y}_{i.} - \bar{y}_{u.})$$

$$= \operatorname{var}(\bar{y}_{i.}) + \operatorname{var}(\bar{y}_{u.}) - 2cov(\bar{y}_{i.}, \bar{y}_{u.})$$

$$= \frac{\sigma^2}{q} + \frac{\sigma^2}{q} - 2 \times 0$$

$$= \frac{2\sigma^2}{q}$$

BLUEs and Variance(BLUE) of *epf*

- In general $\sum_{i=1}^{p} c_i\alpha_i$ is estimable if $\sum_{i=1}^{p} c_i = 0$
- BLUE of $\sum_{i=1}^{p} c_i\alpha_i$

$$= \widehat{\sum_{i=1}^{p} c_i}\alpha_i$$

$$= \sum_{i=1}^{p} c_i\hat{\alpha}_i$$

$$= \sum_{i=1}^{p} c_i(\bar{y}_{i.} - \bar{y}_{..})$$

$$= \sum_{i=1}^{p} c_i\bar{y}_{i.}$$

BLUEs and Variance(BLUE) of *epf*

- Variance(BLUE of $\sum_{i=1}^{p} c_i\alpha_i$)

$$= \mathrm{var}\left(\widehat{\sum_{i=1}^{p} c_i}\alpha_i\right)$$

$$= \mathrm{var}\left(\sum_{i=1}^{p} c_i\bar{y}_{i.}\right)$$

$$= \sum_{i=1}^{p} c_i^2 \,\mathrm{var}(\bar{y}_{i.})$$

$$= \sum_{i=1}^{p} c_i^2 \frac{\sigma^2}{q}$$

$$= \frac{\sigma^2}{q}\sum_{i=1}^{p} c_i^2$$

Summary of BLUEs and Variance(BLUE) of *epf*

Estimable parametric functions	BLUE	Variance(BLUE)
$\mu + \bar{\alpha} + \bar{\beta}$	$\bar{y}_{..}$	$\frac{\sigma^2}{pq}$
$\sum_{i=1}^{p} c_i \alpha_i$ with $\sum_{i=1}^{p} c_i = 0$	$\sum_{i=1}^{p} c_i \bar{y}_{i.}$	$\frac{\sigma^2}{q} \sum_{i=1}^{p} c_i^2$
$\sum_{j=1}^{q} d_j \beta_j$ with $\sum_{j=1}^{q} d_j = 0$	$\sum_{j=1}^{q} d_j \, \bar{y}_{.j}$	$\frac{\sigma^2}{p} \sum_{j=1}^{q} d_j^2$
$\alpha_i - \alpha_u,\ i \neq u$	$\bar{y}_{i.} - \bar{y}_{u.}$	$\frac{2\sigma^2}{q}$
$\beta_j - \beta_v,\ j \neq v$	$\bar{y}_{.j} - \bar{y}_{.\mathrm{v}}$	$\frac{2\sigma^2}{p}$

Model value and error

- **Model value of $\boldsymbol{y_{ij}}$**

$$\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$$

$$= \bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} + \bar{y}_{.j} - \bar{y}_{..}$$

$$= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

- **Error sum of squares**

$$\text{SSE} = \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \hat{y}_{ij}\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \left(\bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}\right)\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2$$

Other way to express SSE

- **Error sum of squares**

SSE

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \hat{y}_{ij}\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \left(\bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} + \bar{y}_{.j} - \bar{y}_{..}\right)\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left((y_{ij} - \bar{y}_{..}) - \left(\bar{y}_{i.} - \bar{y}_{..}\right) - (\bar{y}_{.j} - \bar{y}_{..})\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{..}\right)^2 - \sum_{i=1}^{p} \sum_{j=1}^{q} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$- \sum_{i=1}^{p} \sum_{j=1}^{q} \left(\bar{y}_{.j} - \bar{y}_{..}\right)^2$$

Other way to express SSE …

Symbolically, let

$$TSS = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij} - \bar{y}_{..}\right)^2$$

$$SSA = \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{p} q(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SSB = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 = \sum_{j=1}^{q} p\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2$$

Then SSE can be expressed as,

$$SSE = SST - SSA - SSB$$

Other way to express all sum of squares

Simplified way to express sum of squares

$$TSS = \sum_{i=1}^{p}\sum_{j=1}^{q} {y_{ij}}^2 - \frac{{y_{..}}^2}{pq}$$

$$SSA = \frac{\sum_{i=1}^{p} {y_{i.}}^2}{q} - \frac{{y_{..}}^2}{pq}$$

$$SSB = \frac{\sum_{j=1}^{q} {y_{.j}}^2}{p} - \frac{{y_{..}}^2}{pq}$$

- These are specifically useful expressions while implementing the formulae in software.

Other way to express all sum of squares…

- These need to calculate the following:
    - Row sums/sum of all observations corresponding to fixed level of factor A i.e. $y_{i.}, i = 1,2, \dots, p$
    - Column sums/sum of all observations corresponding to fixed level of factor B i.e. $y_{.j}, j = 1,2, \dots, q$

Testable hypothesis

- The hypothesis which include the estimable parametric functions is called testable hypothesis.
- Examples of testable hypothesis:

$$H_0: \alpha_i = \alpha_u \text{ , } i \neq u$$

$$H_0: \beta_j = \beta_v, \ j \neq v$$

- Non-testable hypothesis:

$$H_0: \alpha_1 + \alpha_2 \quad (\alpha_1 + \alpha_2 \text{ is not estimable})$$

Interpretations of hypothesis

- The effect of two levels of factor A are equal

$$H_0: \alpha_i = \alpha_u \ , i \neq u$$

- Effect of first level of factor A is same as average effect of second and third level.

$$H_0: \alpha_1 = \frac{\alpha_2 + \alpha_3}{2}$$

$$H_0: 2\alpha_1 - \alpha_2 - \alpha_3 \quad \text{(contrast in } \alpha's)$$

- $H_0: \alpha_1 - 2\alpha_2 + \alpha_3$ (contrast in $\alpha's$)

It means the interest is in testing whether the second level of factor A is equal to the average effect of first and third level.

Testing of hypothesis

- Testing equality of effect of all levels of factor A

$$H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$$

- Testing equality of effect of all levels of factor A

$$H_{02}: \beta_1 = \beta_2 = \cdots = \beta_q$$

- Testing equality of effect of any two levels of factor A

$$H_{03}: \alpha_i = \alpha_u \text{ , } i \neq u$$

- Testing equality of effect of all levels of factor A

$$H_{04}: \beta_j = \beta_v, \; j \neq v$$

Steps to develop test-statistic for testing the hypothesis

- Obtain SSE and degrees of freedom for SSE for **original model.** Let it be SSE and df_{SSE}
- Obtain SSE and degrees of freedom for SSE for **reduced model** (model subject to the null hypothesis). Let it be SSE_c and df_{SSE_c}
- Then $SSH_0 = SSE_c - SSE$ and degrees of freedom for the SSH_0 are

$$df_{SSH_0} = df_{SSE_c} - df_{SSE}$$

Steps to develop test-statistic for testing the hypothesis…

- Then SSH_0 and degrees of freedom for the SSH_0 are

$$SSH_0 = SSE_c - SSE$$

$$df_{SSH_0} = df_{SSE_c} - df_{SSE}$$

- Procedure to construct Testing Statistic is:

$$SSH_0 \sim \sigma^2 \chi^2 \text{ with } df_{SSH_0}$$

$$SSE \sim \sigma^2 \chi^2 \text{ with } df_{SSE}$$

$$SSH_0 \amalg \quad SSE$$

$$Test - Statistic = \frac{SSH_0/df_{SSH_0}}{SSE/df_{SSE}} \sim F\left(df_{SSH_0}, df_{SSE}\right)$$

Testing of hypothesis $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$

- H_{01} can be rewritten in the form of estimable parametric functions.

$$H_{01}: \left.\begin{array}{c} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \\ \vdots \\ \alpha_1 - \alpha_p \end{array}\right\} \quad (p-1)\ l.i.e.p.f.$$

- Let $\alpha_1 = \alpha_2 = \cdots = \alpha_p = \alpha\ (say)$

Testing of hypothesis $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$

- **Original model**

Model : $E(y_{ij}) = \mu + \alpha_i + \beta_j$

Solution of

Normal equations: $\hat{\mu} = \bar{y}_{..}$,

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..},$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..},$$

Fitted value : $\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$

SSE : $SSE = SST - SSA - SSB$

DF for SSE : $df_{SSE} = (p-1)(q-1)$

Testing of hypothesis $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$

- **Reduced model**

Model : $E(y_{ij}) = \mu + \alpha + \beta_j$

$= \mu^0 + \beta_j$ where $\mu^0 = \mu + \alpha$

Sol. of N.Eqs. : $\hat{\mu}_0 = \bar{y}_{..}$,

$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}, \qquad j = 1,2, \dots, q$

Fitted value : $\hat{y}_{ij} = \hat{\mu}^0 + \hat{\beta}_j = \bar{y}_{.j}$

SSE : $\sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{.j}\right)^2$

Testing of hypothesis $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$

- **SSE for Reduced model**

SSE for reduced model

$=SSE_c$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{.j}\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(\left(y_{ij} - \bar{y}_{..}\right) - \left(\bar{y}_{.j} - \bar{y}_{..}\right)\right)^2$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{..}\right)^2 - \sum_{i=1}^{p} \sum_{j=1}^{q} \left(\bar{y}_{.j} - \bar{y}_{..}\right)^2$$

$$= SST - SSB$$

Testing of hypothesis $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$

- $\boldsymbol{SSH_0}$ **for the hypothesis**

$$
\begin{aligned}
&\mathrm{SS}H_0 \\
&= \mathrm{SSE_c} - \mathrm{SSE} \\
&= (SST - SSB) - (SST - SSA - SSB) \\
&= SSA \\
&= \textstyle\sum_{i=1}^{p} \sum_{j=1}^{q} (\bar{y}_{i.} - \bar{y}_{..})^2 \\
&= \textstyle\sum_{i=1}^{p} q(\bar{y}_{i.} - \bar{y}_{..})^2 \\
&= \frac{\sum_{i=1}^{p} y_{i.}^2}{q} - \frac{y_{..}^2}{pq}
\end{aligned}
$$

Testing of hypothesis $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_p$

- **Test Statistic**

$$\mathrm{SS}H_0 = \sum_{i=1}^{p} \sum_{j=1}^{q} (\bar{y}_{i.} - \bar{y}_{..})^2 \sim \sigma^2 \chi^2_{(p-1)}$$

$$SSE = \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2 \sim \sigma^2 \chi^2_{(p-1)(q-1)}$$

$$\mathrm{MS}H_0 = \sum_{i=1}^{p} \sum_{j=1}^{q} (\bar{y}_{i.} - \bar{y}_{..})^2 / (p-1)$$

$$MSE = \sum_{i=1}^{p} \sum_{j=1}^{q} \left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2 / (p-1)(q-1)$$

$$Test - Statistic = \frac{MSH_0}{MSE} \sim F_{(p-1),(p-1)(q-1)}$$

Testing of hypothesis $H_{02}: \boldsymbol{\beta_1 = \beta_2 = \cdots = \beta_q}$

- **Test Statistic**

$$\mathrm{SS}H_0 = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 \sim \sigma^2\chi^2_{(q-1)}$$

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2 \sim \sigma^2\chi^2_{(p-1)(q-1)}$$

$$\mathrm{MS}H_0 = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 /(q-1)$$

$$MSE = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2/(p-1)(q-1)$$

$$Test - Statistic = \frac{MSH_0}{MSE} \sim F_{(q-1),(p-1)(q-1)}$$

Estimation of error variance

- Observe that

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2 \sim \sigma^2 \chi^2_{(p-1)(q-1)}$$

$$E(SSE) = \sigma^2(p-1)(q-1)$$

$$E(SSE/(p-1)(q-1)) = \sigma^2$$

$$E(MSE) = \sigma^2$$

Thus,

$$MSE = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2/(p-1)(q-1)$$

is an unbiased estimator of σ^2 i.e. Error variance

ANOVA

Source of variation	Degrees of Freedom	Sum of squares	Mean SS	F-ratio	Hypothesis
Factor A	$(p-1)$	$SSA = \frac{\sum_{i=1}^{p} y_{i.}^2}{q} - \frac{y_{..}^2}{pq}$	$MSA = \frac{SSA}{p-1}$	$\frac{MSA}{MSE}$	$H_{01}: \alpha_1 = \cdots = \alpha_p$
Factor B	$(q-1)$	$SSB = \frac{\sum_{\mathrm{j}=1}^{\mathrm{q}} y_{.j}^2}{p} - \frac{y_{..}^2}{pq}$	$MSB = \frac{SSB}{q-1}$	$\frac{MSB}{MSE}$	$H_{02}: \beta_1 = \cdots = \beta_q$
Error	$(p-1)(q-1)$	$SSE = \mathrm{SST} - \mathrm{SSA} - \mathrm{SSB}$	MSE		
Total	$(pq-1)$	$\mathrm{SST} = \sum_{i=1}^{p} \sum_{j=1}^{q} y_{ij}^2 - \frac{y_{..}^2}{pq}$			

- $\frac{MSA}{MSE} \sim F_{(p-1),(p-1)(q-1)}$ and $\frac{MS\mathrm{B}}{MSE} \sim F_{(q-1),(p-1)(q-1)}$
- Further $E(MSE) = \sigma^2$, that is MSE is unbiased estimator of σ^2

Decision about TOH

- p-value for $H_{01} = 1 - \text{P}\left(F_{(p-1),(p-1)(q-1)} \leq \frac{MSA}{MSE}\right)$
- p-value for $H_{02} = 1 - \text{P}\left(F_{(q-1),(p-1)(q-1)} \leq \frac{MS\text{B}}{MSE}\right)$

- If p-value for $H_{01} < \alpha$ then reject H_{01}

- If p-value for $H_{02} < \alpha$ then reject H_{02}

Testing the hypothesis with individual epf.

- $H_0: \alpha_1 = \alpha_2$
- Rewrite H_0 as $\alpha_1 - \alpha_2 = 0$
- $\widehat{\alpha_1 - \alpha_2} = \bar{y}_{1.} - \bar{y}_{2.} \sim N\left(\alpha_1 - \alpha_2, \mathrm{var}(\widehat{\alpha_1 - \alpha_2})\right)$
- $$\bar{y}_{1.} - \bar{y}_{2.} \sim N\left(\alpha_1 - \alpha_2, \frac{2\sigma^2}{q}\right)$$
- $\dfrac{(\bar{y}_{1.} - \bar{y}_{2.}) - (\alpha_1 - \alpha_2)}{\sqrt{\frac{2\sigma^2}{q}}} \sim N(0,1)$
- $\left\{\dfrac{(\bar{y}_{1.} - \bar{y}_{2.}) - (\alpha_1 - \alpha_2)}{\sqrt{\frac{2\sigma^2}{q}}}\right\}^2 \sim \chi^2_{(1)}$

Testing the hypothesis with individual epf.

- $\left\{\frac{(\bar{y}_{1.}-\bar{y}_{2.})-(\alpha_1-\alpha_2)}{\sqrt{\frac{2\sigma^2}{q}}}\right\}^2 \sim\chi^2_{(1)}$
- Under the null hypothesis $\alpha_1-\alpha_2=0$. Hence
- $\frac{(\bar{y}_{1.}-\bar{y}_{2.})^2}{\left(\frac{2\sigma^2}{q}\right)} \sim\chi^2_{(1)}$ and is independently distributed of $\frac{SSE}{\sigma^2}=$

$\sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..}\right)^2$ which has $\chi^2_{(p-1)(q-1)}$

Hence test statistic is: $\frac{(\bar{y}_{1.}-\bar{y}_{2.})^2/\left(\frac{2}{q}\right)}{MSE} \sim F_{1,(p-1)(q-1)}$

General test statistic for Testing the individual epf.

- H_0: $\underline{\lambda}'\underline{\beta} = d$

- As $\underline{\lambda}'\underline{\hat{\beta}} \sim N\left(\underline{\lambda}'\underline{\beta},\ var\left(\underline{\lambda}'\underline{\hat{\beta}}\right)\right) \Rightarrow\ SSH_0 = \frac{\left(\underline{\lambda}'\underline{\widehat{\beta}} - d\right)^2}{var\left(\underline{\lambda}'\underline{\widehat{\beta}}\right)} \sim \chi^2_{(1)}$

- $\frac{SSE}{\sigma^2} \sim \chi^2_{(p-1)(q-1)}$

- SSH_0 is independently distributed of SSE

- Hence $test - statistic \sim F_{1,(p-1)(q-1)}$ distribution and is:

$$\frac{\left(BLUE\left(\underline{\lambda}'\underline{\beta}\right) - hypothetical\ value\ of\ \underline{\lambda}'\underline{\beta}\right)^2 / \left(var\left(\underline{\lambda}'\underline{\widehat{\beta}}\right) without \sigma^2\right)}{MSE}$$

Sample questions

- Test statistic for testing the hypothesis of equality of all treatment effects in two-way classification model.
- BLUE of elementary contrast in treatment effects for one way classification model.
- What would be the rank of estimation space for two-way model with p=3 and q=4?
- For two-way classification model with p=5, q=4 and r=3, what is the variance of BLUE of contrast in block effects?
- State the test statistic for testing the hypothesis of equality of any two column effects in two-way classification model.
- State an unbiased estimator of error variance for two-way classification model.

Sample questions…

- For two-way classification model with write the linear model and answer the following.
  - What is rank of error space and estimation space?
  - Specify one complete set of linearly independent estimable parametric functions and their BLUEs and variances of BLUEs.
  - What is fitted value of y_{ij}? Hence give formula for SSE.
  - Write an ANOVA table specifying the hypothesis to be tested against each row of it.
- For RBD with 4 treatment and 5 blocks which of the following are estimable

A) $\alpha_1 - 2\alpha_2 + \alpha_3$ B) $\mu + \alpha_1 + \alpha_2 + 2\beta_1$

C) $\mu + \alpha_3 + \beta_2 + \beta_5$ D) $\mu + \alpha_1 + \beta_5$

What we have studied

- Two-way classification model
- Normal equations and their solutions
- Estimability conditions for linear parametric functions
- Estimable parametric functions, their BLUEs and variances of BLUEs.
- Rank of error space and estimation space
- A set of linearly independent estimable parametric functions
- Fitted value of y_{ij}, SSE and an unbiased estimator of error variance
- Testing of hypothesis: equality of all row/column effects
- Testing of hypothesis: single epf
- ANOVA table specifying the hypothesis to be tested against each row of it.