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 3: Two-way Classification Model with r observations per cell

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 r times***. 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 pqr 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
A_1	$y_{111}, y_{112}, \dots, y_{11r}$	$y_{121}, y_{122}, \dots, y_{12r}$	…	$y_{1q1}, y_{1q2}, \dots, y_{1qr}$
A_2	$y_{211}, y_{212}, \dots, y_{21r}$	$y_{221}, y_{222}, \dots, y_{22r}$	…	$y_{2q1}, y_{2q2}, \dots, y_{2qr}$
:	:	:	…	:
A_p	$y_{p11}, y_{p12}, \dots, y_{11r}$	$y_{p21}, y_{p22}, \dots, y_{p2r}$	…	$y_{pq1}, y_{pq2}, \dots, y_{pqr}$
Sums	$y_{.1.}$	$y_{.2.}$	…	$y_{.q.}$
Averages	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	…	$\bar{y}_{.q.}$

Example 1: Two-way classification model

The quality control department of a fabric finishing plant is studying the effect of several factors on the dying of cotton synthetic cloth used to manufacture men's shirt. Three operators and three cycle times were selected and three small specimens of cloth were dyed under each setting condition. The finished cloth is compared to standard and a numerical score is assigned. The results are as follows.

Cycle time	Operator		
	1	2	3
40	23, 24, 25	27, 28, 26	31, 32, 29
50	36, 35, 36	34, 38, 39	33, 34, 35
60	28, 24, 27	35, 35, 34	26, 27, 25

Example 1: Two-way classification model…

Factor A:

Factor B:

Replications:

Response:

Two-way classification model with r-observations per cell

Model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \quad i = 1,2,\dots,p; j = 1,2,\dots,q$$
$$k = 1,2,\dots,r$$

y_{ijk}:

μ:

α_i:

β_j:

ε_{ijk}:

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

Assumptions

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

Implications

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

Model details

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

$$\mu, \alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_p$$

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

Derivation of normal equations

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

- Define errors as:

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

- 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_{ijk} (observed value) and $\hat{y}_{ijk}$ (fitted value) close to each other.

Derivation of normal equations…

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

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

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

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

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

Derivation of normal equations…

$$(1)\Rightarrow \qquad y_{\ldots} = pqr\hat{\mu} + qr\sum_{i=1}^{p}\hat{\alpha}_i + pr\sum_{j=1}^{q}\hat{\beta}_j \qquad (2)$$

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

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

- Observe that (1), (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.

Solution of normal equations

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

Solution of normal equations…

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

(2)⇒ $\hat{\mu} = \bar{y}_{...}$

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

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

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)

$$\rho(Error\ space)$$

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

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

Here $n = pqr$ which are number of observations.

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

Some Definitions:

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

Subtracting these we get:

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

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

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

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.

- $\mu + \bar{\alpha} + \bar{\beta}$ (1)
- $\alpha_1 - \alpha_2$
- $\alpha_1 - \alpha_3$
- $\vdots$
- $\alpha_1 - \alpha_p$

$(p - 1)$

- $\beta_1 - \beta_2$
- $\beta_1 - \beta_3$
- $\vdots$
- $\beta_1 - \beta_q$

$(q - 1)$

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.

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

- Variance

Variance(BLUE)

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

$$= \operatorname{var}(\bar{y}_{\dots})$$

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

BLUEs and Variance(BLUE) of *epf*

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

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

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

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

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

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

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

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

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

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

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

BLUEs and Variance(BLUE) of *epf*

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

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

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

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

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

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

BLUEs and Variance(BLUE) of *epf*

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

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

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

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

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

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

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

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

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

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

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

BLUEs and Variance(BLUE) of *epf*

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

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

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

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

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

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

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

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

Model value and error

- **Model value of y_{ij}**

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

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

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

- **Error sum of squares**

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

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

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

Other way to express SSE

- **Error sum of squares**

SSE

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

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

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

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

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

Other way to express SSE …

Symbolically, let

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

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

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

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

$$TSS = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} {y_{ijk}}^2 - \frac{{y_{\dots}}^2}{pqr}$$

$$SSA = \frac{\sum_{i=1}^{p} {y_{i..}}^2}{qr} - \frac{{y_{\dots}}^2}{pqr}$$

$$SSB = \frac{\sum_{j=1}^{q} {y_{.j.}}^2}{qr} - \frac{{y_{\dots}}^2}{pqr}$$

- 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:

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

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

- Non-testable hypothesis:

$H_0: \alpha_1 + \alpha_2 \quad (\alpha_1 + \alpha_2$ 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.

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

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

- $H_0: \alpha_1 - 2\alpha_2 + \alpha_3$ (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

$$H_{02}: \beta_1 = \beta_2 = \cdots = \beta_q, \; j \neq v$$

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:

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

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

$$SSH_0 \ \coprod \quad SSE$$

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

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

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

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

- 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 : $E(y_{ijk}) = \mu + \alpha_i + \beta_j$

Solution of

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

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

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

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

SSE : $SSE = SST - SSA - SSB$

DF for SSE : $df_{SSE} = pqr - p - q + 1$

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

- **Reduced model**

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

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

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

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

Fitted value : $\hat{y}_{ijk} = \hat{\mu}_0 + \hat{\beta}_j = \bar{y}_{.j.}$

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

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

- **SSE for Reduced model**

SSE for reduced model

$=SSE_c$

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

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

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

$$= SST - SSB$$

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

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

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

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

- **Test Statistic**

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

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

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

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

$$MSE = SSE/(pqr - p - q + 1)$$

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

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

- **Test Statistic**

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

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

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

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

$$MSE = SSE/(pqr - p - q + 1)$$

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

Estimation of error variance

- Observe that

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

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

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

$$E\left(\frac{SSE}{pqr-p-q+1}\right) = \sigma^2$$

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

$\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\left(y_{ijk}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}\right)^2}{(pqr-p-q+1)}$ 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	$(p-1)$	$SSA = \frac{\sum_{i=1}^{p} y_{i..}^2}{q\mathrm{r}} - \frac{y_{...}^2}{pq\mathrm{r}}$	$MSA = \frac{SSA}{p-1}$	$\frac{MSA}{MSE}$	$H_{01}: \alpha_1 = \cdots = \alpha_p$
Factor B	$(q-1)$	$SSB = \frac{\sum_{\mathrm{j}=1}^{\mathrm{q}} y_{.j.}^2}{p\mathrm{r}} - \frac{y_{...}^2}{pq\mathrm{r}}$	$MSB = \frac{SSB}{q-1}$	$\frac{MSB}{MSE}$	$H_{02}: \beta_1 = \cdots = \beta_q$
Error	$(pqr - p - q + 1)$	$SSE = \mathrm{SST} - \mathrm{SSA} - \mathrm{SSB}$	MSE		
Total	$(pq-1)$	$\sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{\mathrm{k}=1}^{\mathrm{r}} y_{ij\mathrm{k}}^2 - \frac{y_{...}^2}{pq\mathrm{r}}$			

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

Decision about TOH

- p-value for $H_{01} = 1 - \mathrm{P}\left(F_{(p-1),(pqr-p-q+1)} \leq \frac{MSA}{MSE}\right)$
- p-value for $H_{02} = 1 - \mathrm{P}\left(F_{(q-1),(pqr-p-q+1)} \leq \frac{MS\mathrm{B}}{MSE}\right)$
- 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 $\alpha_1 - \alpha_2 = 0$
- $\widehat{\alpha_1 - \alpha_2} = \bar{y}_{1..} - \bar{y}_{2..} \sim N\left(\alpha_1 - \alpha_2, \text{var}(\widehat{\alpha_1 - \alpha_2})\right)$
- $$\bar{y}_{1..} - \bar{y}_{2..} \sim N\left(\alpha_1 - \alpha_2, \frac{2\sigma^2}{qr}\right)$$
- $\dfrac{(\bar{y}_{1..} - \bar{y}_{2..}) - \alpha_1 - \alpha_2}{\sqrt{\frac{2\sigma^2}{qr}}} \sim N(0,1)$
- $\left\{\dfrac{(\bar{y}_{1..} - \bar{y}_{2..}) - \alpha_1 - \alpha_2}{\sqrt{\frac{2\sigma^2}{qr}}}\right\}^2 \sim \chi^2_{(1)}$

Testing the hypothesis with individual epf.

- Under the null hypothesis $\alpha_1 - \alpha_2 = 0$.

- Hence $\frac{(\bar{y}_{1..} - \bar{y}_{2..})^2}{\left(\frac{2\sigma^2}{qr}\right)} \sim \chi^2_{(1)}$ and is independently distributed of

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

$\sigma^2 \chi^2_{(pqr-p-q+1)}$

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

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

- Derive a test for testing the hypothesis of equality of all column/block effects in two-way classification model with r observations per cell.
- Obtain normal equations and their solutions for two-way classification model with r observations per cell.
- What would be the rank of error space for two-way model with p=3, q=4, $r = 2$? Compare it with unreplicated model.
- For two-way classification model r observations per cell and p=5, q=4 and r=3, what is the $Var(\bar{y}_{1..} - \bar{y}_{2..})$?
- State the test statistic for testing the hypothesis of equality of any two row effects in two-way classification model r observations per cell.
- State an unbiased estimator of error variance for two-way classification model r observations per cell.

Sample questions…

- For two-way classification model with r observations per cell, 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_{ijk}? Hence give formula for SSE.
  - Write an ANOVA table specifying the hypothesis to be tested against each row of it.

What we have studied

- GLM for two-way classification model with r observations per cell
- 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.