

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 5

**Two-way Classification Model with r -
observations per cell and interaction**

About the experiments for two-way classification model

Objective: To study the effect of two factors on the response variable and test whether the two factors have joint effect.

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 ($r > 1$)*. 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...

Factor	B_1	B_2	...	B_q	Sums	Averages
A_1	$y_{111}, y_{112}, \dots, y_{11r}$	$y_{121}, y_{122}, \dots,$ y_{12r}	...	$y_{1q1}, y_{1q2}, \dots, y_{1qr}$	$y_{1..}$	$\bar{y}_{1..}$
A_2	$y_{211}, y_{212}, \dots, y_{21r}$	$y_{221}, y_{222}, \dots, y_{22r}$...	$y_{2q1}, y_{2q2}, \dots, y_{2qr}$	$y_{2..}$	$\bar{y}_{2..}$
:	:	:	...	:
A_p	$y_{p11}, y_{p12}, \dots, y_{p1r}$	$y_{p21}, y_{p22}, \dots, y_{p2r}$...	$y_{pq1}, y_{pq2}, \dots, y_{pqr}$	$y_{p..}$	$\bar{y}_{p..}$
Sums	$y_{.1.}$	$y_{.2.}$...	$y_{.q.}$	$y_{...}$	
Averages	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$...	$\bar{y}_{.q.}$		$\bar{y}_{...}$

Two-way classification model

Model:

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

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
- Covariances between errors is zero $var(\underline{\varepsilon}) = \sigma^2 I_n$
- $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$

Implications

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

Model details

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

$$\mu,$$

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

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

$$\gamma_{11}, \gamma_{12}, \dots, \gamma_{1q}$$

$$\gamma_{21}, \gamma_{22}, \dots, \gamma_{2q}$$

$$\vdots$$

$$\gamma_{p1}, \gamma_{p2}, \dots, \gamma_{pq}$$

Model speciality

- Here note that number of observations are less than number of parameters.

$$n = pqr > 1 + p + q + pq$$

- The assumption $n > p$ in GLM is true.

Derivation of normal equations

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

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

- Obtain $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j, \hat{\gamma}_{ij}$ such that the errors/error sum of squares is minimum.
- It will lead to $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j, \hat{\gamma}_{ij}$ such that y_{ijk} (observed value) and \hat{y}_{ijk} (fitted value) are close to each other in least square sense.

Derivation of normal equations...

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

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

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

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

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

$$\frac{d\phi}{d\hat{\gamma}_{ij}} = -2 \sum_{k=1}^r (y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij}), \quad i = 1:p, j = 1:q \quad (C)$$

Derivation of normal equations...

$$(1) \Rightarrow y_{...} = pqr\hat{\mu} + qr \sum_{i=1}^p \hat{\alpha}_i + pr \sum_{j=1}^q \hat{\beta}_j + r \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij}$$

$$(A) \Rightarrow y_{i..} = qr\hat{\mu} + qr\hat{\alpha}_i + r \sum_{j=1}^q \hat{\beta}_j + r \sum_{j=1}^q \hat{\gamma}_{ij}, \quad i = 1:p$$

$$(B) \Rightarrow y_{.j.} = pr\hat{\mu} + r \sum_{i=1}^p \hat{\alpha}_i + pr\hat{\beta}_j + r \sum_{i=1}^p \hat{\gamma}_{ij}, \quad j = 1:q$$

$$(C) \Rightarrow y_{ij.} = r\hat{\mu} + r\hat{\alpha}_i + r\hat{\beta}_j + r\hat{\gamma}_{ij}, \quad i = 1:p, j = 1:q$$

- The above equations are $1 + p + q + pq$ equations in $1 + p + q + pq$ variables.
- These equations are called normal equations.
- But all are not linearly independent.

Normal equations

$$y_{...} = pqr\hat{\mu} + qr \sum_{i=1}^p \hat{\alpha}_i + pr \sum_{j=1}^q \hat{\beta}_j + r \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij} \quad (2)$$

$$y_{i..} = qr\hat{\mu} + qr\hat{\alpha}_i + r \sum_{j=1}^q \hat{\beta}_j + r \sum_{j=1}^q \hat{\gamma}_{ij}, \quad i = 1:p \quad (D)$$

$$y_{.j.} = pr\hat{\mu} + r \sum_{i=1}^p \hat{\alpha}_i + pr\hat{\beta}_j + r \sum_{i=1}^p \hat{\gamma}_{ij}, \quad j = 1:q \quad (E)$$

$$y_{ij.} = r\hat{\mu} + r\hat{\alpha}_i + r\hat{\beta}_j + r\hat{\gamma}_{ij}, \quad i = 1:p, j = 1:q \quad (F)$$

.

Solution of normal equations

- **Observe the following inter-relations among the normal equations.**
 - ❖ $\sum_{i=1}^p (D) = (2)$ i.e. addition of p equations in (D) gives equation (1)
 - ❖ $\sum_{j=1}^q (E) = (2)$ i.e. addition of q equations in (E) gives equation (1)
 - ❖ $\sum_{i=1}^p \sum_{j=1}^q (F) = (2)$ i.e. addition of pq equations in (F) gives equation (1)
 - ❖ $\sum_{j=1}^q (F) = (D)$ for $i = 1:p$ i.e. for fixed i if we add q equations in (E) we get set of equations in (D)
 - ❖ $\sum_{i=1}^p (F) = (E)$ for $j = 1:q$ i.e. for fixed j if we add p equations in (D) we get set of equations in (E)

Solution of normal equations

- (2) is linearly dependent as $\sum_{i=1}^p (D) = (2)$
- (D) are linearly dependent as $\sum_{j=1}^q (F) = (D)$
- (E) are linearly dependent as $\sum_{i=1}^p (F) = (E)$
- Only equations in (F) are linearly independent.
- Thus only pq out of $1 + p + q + pq$ equations are linearly independent.
- Thus rank of estimation space i.e. number of linearly independent normal equations for this model is pq .

Solution of normal equations

- To solve the normal equations we need additional $1 + p + q$ equations which are linearly independent with (2), (D), (E) and (F).

- Let these two equations be:

$$\diamond \sum_{i=1}^p \hat{\gamma}_{ij} = 0, \quad j = 1:q \quad (\text{A1})$$

$$\diamond \sum_{j=1}^q \hat{\gamma}_{ij} = 0, \quad i = 1:p \quad (\text{A2})$$

$$\diamond \sum_{i=1}^p \hat{\alpha}_i = 0$$

$$\diamond \sum_{j=1}^q \hat{\beta}_j = 0$$

- These seem to be $p + q + 2$ equations.

Solution of normal equations

- Observe that

$$\diamond \sum_{i=1}^p \hat{\gamma}_{ij} = 0, \quad j = 1:q \Rightarrow \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij} = 0$$

$$\diamond \sum_{j=1}^q \hat{\gamma}_{ij} = 0, \quad i = 1:p \Rightarrow \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij} = 0$$

❖ Hence (A1) and (A2) in fact $p + q - 1$ linearly independent equations and not $p + q$.

❖ Thus (A1) and (A2) together with $\sum_{i=1}^p \hat{\alpha}_i = 0$ and $\sum_{j=1}^q \hat{\beta}_j = 0$ are $p + q + 1$ linearly independent equations

Derivation of normal equations...

$$y_{...} = pqr\hat{\mu} + qr \sum_{i=1}^p \hat{\alpha}_i + pr \sum_{j=1}^q \hat{\beta}_j + r \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij}$$

$$\Rightarrow \hat{\mu} = \bar{y}_{...}$$

$$y_{i..} = qr\hat{\mu} + qr\hat{\alpha}_i + r \sum_{j=1}^q \hat{\beta}_j + r \sum_{j=1}^q \hat{\gamma}_{ij}, \quad i = 1:p$$

$$\Rightarrow \hat{\alpha}_i = \bar{y}_{i..} - \hat{\mu} = \bar{y}_{i..} - \bar{y}_{...}, \quad i = 1, 2, \dots, p$$

$$y_{.j.} = pr\hat{\mu} + r \sum_{i=1}^p \hat{\alpha}_i + pr\hat{\beta}_j + r \sum_{i=1}^p \hat{\gamma}_{ij}, \quad j = 1:q$$

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

Derivation of normal equations...

$$y_{ij.} = r(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}), \quad i = 1:p, j = 1:q \quad (\text{F})$$

$$\begin{aligned} \Rightarrow \quad \hat{\gamma}_{ij} &= \bar{y}_{ij.} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \\ &= \bar{y}_{ij.} - \bar{y}_{...} - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}) \\ &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \end{aligned}$$

Solution of normal equations...

$$\hat{\mu} = \bar{y}_{...}$$

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}, \quad i = 1, 2, \dots, p$$

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

$$\hat{\gamma}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}, \quad i = 1:p, j = 1:q$$

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.

$\rho(\text{estimation space})$

=number of linearly independent normal equations

= Number of *linearly independent estimable lpfs*

= pq

- It means that there would be only pq *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)

$$\begin{aligned}\rho(\text{Error space}) \\ &= n - \rho(\text{estimation space}) \\ &= pqr - pq \\ &= pq(r - 1)\end{aligned}$$

- Here $n = pqr$ represent number of observations.

Normal equations

$$y_{...} = pqr\hat{\mu} + qr \sum_{i=1}^p \hat{\alpha}_i + pr \sum_{j=1}^q \hat{\beta}_j + r \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij} \quad (2)$$

$$y_{i..} = qr\hat{\mu} + qr\hat{\alpha}_i + r \sum_{j=1}^q \hat{\beta}_j + r \sum_{j=1}^q \hat{\gamma}_{ij}, \quad i = 1:p \quad (D)$$

$$y_{.j.} = pr\hat{\mu} + r \sum_{i=1}^p \hat{\alpha}_i + pr\hat{\beta}_j + r \sum_{i=1}^p \hat{\gamma}_{ij}, \quad j = 1:q \quad (E)$$

$$y_{ij.} = r\hat{\mu} + r\hat{\alpha}_i + r\hat{\beta}_j + r\hat{\gamma}_{ij}, \quad i = 1:p, j = 1:q \quad (F)$$

Expectations of normal equations:

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

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

$$E(y_{i..}) = qr\mu + qr\alpha_i + r \sum_{j=1}^q \beta_j + r \sum_{j=1}^q \gamma_{ij}, \quad i = 1:p$$

$$E(y_{.j.}) = pr\mu + r \sum_{i=1}^p \alpha_i + rp\beta_j + r\gamma_{ij}, \quad j = 1:q$$

$$E(y_{ij.}) = r\mu + r\alpha_i + r\beta_j + r\gamma_{ij}, \quad i = 1:p, j = 1:q$$

Estimability of parametric functions in terms of all parameters

Using first normal equation,

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

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

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

$$\Rightarrow \mu + \frac{1}{p} \sum_{i=1}^p \alpha_i + \frac{1}{q} \sum_{j=1}^q \beta_j + \frac{1}{pq} \sum_{i=1}^p \sum_{j=1}^q \gamma_{ij} \text{ is an } epf$$

$$\Rightarrow \mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma}_{..} \text{ is an estimable parametric function}$$

Estimability of linear parametric functions in α' s

$$E(y_{i..}) = qr\mu + qr\alpha_i + r \sum_{j=1}^q \beta_j + r \sum_{j=1}^q \gamma_{ij}, \quad i = 1:p$$

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

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

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

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

Subtracting these we get:

$$\begin{aligned} E(\bar{y}_{i..} - \bar{y}_{u..}) &= (\alpha_i + \bar{\gamma}_i) - (\alpha_u + \bar{\gamma}_u) \quad i \neq u \\ &= \alpha_i^* - \alpha_u^* \quad \text{where } \alpha_i^* = (\alpha_i + \bar{\gamma}_i) \end{aligned}$$

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

Estimability of linear parametric function in α' s...

- All elementary contrasts in α^* 's are estimable.
- All contrasts in α^* '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 β' s

$$(E) \Rightarrow E(y_{.j.}) = pr\mu + r \sum_{i=1}^p \alpha_i + rp\beta_j + r\gamma_{ij}, \quad j = 1:q$$

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

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

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

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

Subtracting these we get:

$$E(\bar{y}_{.j.} - \bar{y}_{.v.}) = (\beta_j + \bar{\gamma}_{.j}) - (\beta_v + \bar{\gamma}_{.v}), \text{ where } \beta_j^* = \beta_j + \bar{\gamma}_{.j}$$

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

Estimability of linear parametric function in β' s...

- All elementary contrasts in β^* 's are estimable.
- All contrasts in β^* '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.

Estimability of parametric function in interaction terms

Performing $(2) - (D) - (E) + (F)$ as follows.

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

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

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

$$+ E(\bar{y}_{ij.}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

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

Estimability of parametric function in interaction terms...

- Thus $\gamma_{ij}^* = \gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..}$ are all estimable for all $i = 1:p, j = 1:q$
- Further all pq γ_{ij}^* are not linearly independent as the following relations exists between them.
 - $\sum_{j=1}^q \gamma_{ij}^* = 0$, for $i = 1:p$
 - $\sum_{i=1}^p \gamma_{ij}^* = 0$, for $j = 1:q$
 - $\sum_{i=1}^p \sum_{j=1}^q \gamma_{ij}^* = 0$
- Thus only $(p-1)(q-1)$ of the γ_{ij}^* are linearly independent.

Estimability of parametric function in interaction terms...

- What is meaning of $\gamma_{ij}^* = 0$

$$\gamma_{ij}^* = 0$$

$$\Rightarrow \gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..} = 0$$

$$\Rightarrow \gamma_{ij} = \bar{\gamma}_{i.} + \bar{\gamma}_{.j} - \bar{\gamma}_{..}$$

Summary of estimability conditions of l. parametric functions

1. $\mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma}_{..}$ is estimable (A *lpf* which involve all parameters)
2. All contrasts in α^{*} 's are estimable.
3. All contrasts in β^{*} 's are estimable.
4. $\gamma_{ij}^{*} = \gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..}$ are all estimable for all $i = 1:p, j = 1:q$

One set of linearly independent e.p.f.

- $\mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma}_{..} \quad (1)$

- $\alpha_1^* - \alpha_2^*$
- $\alpha_1^* - \alpha_3^*$
- \vdots
- $\alpha_1^* - \alpha_p^*$

(p - 1)

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

(q - 1)

$\gamma_{ij}^*, i = 1:p, j = 1:q$ are linearly independent e.p.fs.
(p - 1)(q - 1)

In all this is set of pq linearly independent e.p.fs.
 pq

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
- $E(\text{RHS of normal equations}) = E(X'\underline{Y}) = X'X\underline{\beta}$
- 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 $\mu + \bar{\alpha} + \widehat{\bar{\beta}} + \bar{\gamma}_{..}$ which is estimable.

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

- Variance

Variance(BLUE)

$$= \text{var}(\mu + \bar{\alpha} + \widehat{\bar{\beta}} + \bar{\gamma}_{..})$$

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

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

BLUES and Variance(BLUE) of *epf*

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

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

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

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

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

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

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

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

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

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

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

BLUES and Variance(BLUE) of *epf*

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

$$= \text{var}(\widehat{\alpha_i^* - \alpha_u^*})$$

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

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

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

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

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

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

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

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

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

BLUES and Variance(BLUE) of *epf*

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

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

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

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

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

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

BLUE of parametric function in interaction terms

- $\gamma_{ij}^* = \gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..}$ are all estimable
- BLUE of γ_{ij}^*

$$= \hat{\gamma}_{ij}^*$$

$$= \gamma_{ij} - \widehat{\bar{\gamma}_{i.}} - \bar{\gamma}_{.j} + \bar{\gamma}_{..}$$

$$= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \quad \text{as } E(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) = \gamma_{ij}^*$$

$$\hat{\gamma}_{ij}^* = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

Variance of BLUE of parametric function in interaction terms

$$\begin{aligned} & \text{var}(\hat{\gamma}_{ij}^*) \\ &= \text{var}(\gamma_{ij} - \widehat{\bar{\gamma}_{i.}} - \widehat{\bar{\gamma}_{.j}} + \bar{\gamma}_{..}) \\ &= \text{var}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) \\ &= \text{var}(\bar{y}_{ij.}) + \text{var}(\bar{y}_{i..}) + \text{var}(\bar{y}_{.j.}) + \text{var}(\bar{y}_{...}) - \\ & \quad 2\text{cov}(\bar{y}_{ij.}, \bar{y}_{i..}) - 2\text{cov}(\bar{y}_{ij.}, \bar{y}_{.j.}) + 2\text{cov}(\bar{y}_{ij.}, \bar{y}_{...}) + \\ & \quad 2\text{cov}(\bar{y}_{i..}, \bar{y}_{.j.}) - 2\text{cov}(\bar{y}_{i..}, \bar{y}_{...}) - 2\text{cov}(\bar{y}_{.j.}, \bar{y}_{...}) \\ &= \frac{\sigma^2}{r} + \frac{\sigma^2}{qr} + \frac{\sigma^2}{pr} + \frac{\sigma^2}{pqr} - 2\frac{r\sigma^2}{r.qr} - 2\frac{r\sigma^2}{r.pr} + 2\frac{r\sigma^2}{rpqr} + 2\frac{r\sigma^2}{qrpr} - \\ & \quad 2\frac{qr\sigma^2}{qr.pqr} - 2\frac{pr\sigma^2}{prpqr} \end{aligned}$$

Variance of BLUE of parametric function in interaction terms

$$\begin{aligned} & \text{var}(\hat{y}_{ij}^*) \\ &= \frac{\sigma^2}{r} + \frac{\sigma^2}{qr} + \frac{\sigma^2}{pr} + \frac{\sigma^2}{pqr} - 2 \frac{r\sigma^2}{r.qr} - 2 \frac{r\sigma^2}{r.pr} + 2 \frac{r\sigma^2}{rpqr} + 2 \frac{r\sigma^2}{qrpr} \\ & \quad - 2 \frac{qr\sigma^2}{qr.pqr} - 2 \frac{pr\sigma^2}{pr.pqr} \\ &= \frac{\sigma^2}{r} - \frac{\sigma^2}{qr} - \frac{\sigma^2}{pr} + \frac{\sigma^2}{pqr} \\ &= \sigma^2 \left(\frac{1}{r} - \frac{1}{qr} - \frac{1}{pr} + \frac{1}{pqr} \right) \\ &= \frac{\sigma^2}{pqr} (pq - p - q + 1) = \frac{\sigma^2}{pqr} (p - 1)(q - 1) \end{aligned}$$

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

Estimable parametric functions	BLUE	Var (BLUE)
$\mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma}_{..}$	$\bar{y}_{...}$	$\frac{\sigma^2}{pqr}$
$\alpha_i^* - \alpha_u^*, i \neq u$	$\bar{y}_{i..} - \bar{y}_{u..}$	$\frac{2\sigma^2}{qr}$
$\beta_j^* - \beta_v^*, j \neq v$	$\bar{y}_{.j.} - \bar{y}_{.v.}$	$\frac{2\sigma^2}{pr}$
$\gamma_{ij}^* = \gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..}$	$\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$	$\frac{\sigma^2(p-1)(q-1)}{pqr}$

Model value and error

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

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij} \\ &= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) \\ &= \bar{y}_{ij.}\end{aligned}$$

Fitted observed value in a $(i, j)^{th}$ cell is just cell average.

- **Error sum of squares**

$$SSE = \sum_{i,j,k} (y_{ijk} - \hat{y}_{ijk})^2 = \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij.})^2$$

Some notations:

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

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

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

$$SS_{\text{cells}} = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{...})^2$$

$$SS_{AB} = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$SSE = \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij.})^2$$

Total sum of squares

$$\begin{aligned} \text{SST} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk} - \bar{y}_{...})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk}^2 - 2y_{ijk}\bar{y}_{...} + \bar{y}_{...}^2) \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r \left[y_{ijk}^2 - 2y_{ijk} \frac{y_{...}}{pqr} + \left(\frac{y_{...}}{pqr} \right)^2 \right] \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r y_{ijk}^2 - 2y_{...} \frac{y_{...}}{pqr} + \frac{y_{...}^2}{pqr} \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r y_{ijk}^2 - \frac{y_{...}^2}{pqr} \end{aligned}$$

Sum of squares due to Factor A

$$\begin{aligned} \text{SSA} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{i..}^2 - 2\bar{y}_{i..}\bar{y}_{...} + \bar{y}_{...}^2) \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r \left[\left(\frac{y_{i..}}{qr} \right)^2 - 2 \frac{y_{i..}}{qr} \frac{y_{...}}{pqr} + \left(\frac{y_{...}}{pqr} \right)^2 \right] \\ &= \sum_{i=1}^p \frac{y_{i..}^2}{qr} - 2y_{...} \frac{y_{...}}{pqr} + \frac{y_{...}^2}{pqr} \\ &= \sum_{i=1}^p \frac{y_{i..}^2}{qr} - \frac{y_{...}^2}{pqr} \end{aligned}$$

Sum of squares due to Factor B

$$\begin{aligned} \text{SSB} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{.j.}^2 - 2\bar{y}_{.j.}\bar{y}_{...} + \bar{y}_{...}^2) \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r \left[\left(\frac{y_{.j.}}{pr} \right)^2 - 2 \frac{y_{.j.}}{pr} \frac{y_{...}}{pqr} + \left(\frac{y_{...}}{pqr} \right)^2 \right] \\ &= \sum_{j=1}^q \frac{y_{.j.}^2}{pr} - 2y_{...} \frac{y_{...}}{pqr} + \frac{y_{...}^2}{pqr} \\ &= \sum_{j=1}^q \frac{y_{.j.}^2}{pr} - \frac{y_{...}^2}{pqr} \end{aligned}$$

Sum of squares due to cells

$$\begin{aligned}SS_{cells} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{...})^2 \\&= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.}^2 - 2\bar{y}_{ij.}\bar{y}_{...} + \bar{y}_{...}^2) \\&= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r \left[\left(\frac{y_{ij.}}{r} \right)^2 - 2 \frac{y_{ij.}}{r} \frac{y_{...}}{pqr} + \left(\frac{y_{...}}{pqr} \right)^2 \right] \\&= \sum_{i=1}^p \sum_{j=1}^q \frac{y_{ij.}^2}{r} - 2y_{...} \frac{y_{...}}{pqr} + \frac{y_{...}^2}{pqr} \\&= \sum_{i=1}^p \sum_{j=1}^q \frac{y_{ij.}^2}{r} - \frac{y_{...}^2}{pqr}\end{aligned}$$

Sum of squares due to Interaction due to Factor A and B

$SSAB$

$$\begin{aligned} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r \left((\bar{y}_{ij.} - \bar{y}_{...}) - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}) \right)^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{...})^2 - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &= SS_{cells} - SSA - SSB \end{aligned}$$

Error sum of squares

SSE

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

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

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

Error sum of squares...

SSE

$$\begin{aligned} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \end{aligned}$$

Error sum of squares...

SSE

$$\begin{aligned} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{..})^2 \\ &= SST - SSA - SSB - (SS_{cells} - SSA - SSB) \\ &= SST - SS_{cells} \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk} - \bar{y}_{...})^2 - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{ij.} - \bar{y}_{...})^2 \end{aligned}$$

Other way to express SSE ...

Symbolically, let

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

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

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

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

Thus SSE can be also be expressed as,

$$SSE = SST - SSA - SSB - SSAB$$

Other way to express all sum of squares

Expressions for implementing the formulae in software

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

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

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

$$SS_{cells} = \frac{\sum_{i=1}^p \sum_{j=1}^q y_{ij.}^2}{r} - \frac{y_{...}^2}{pqr}$$

$$SSAB = SS_{cells} - SSA - SSB$$

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. $y_{i..}$, $i = 1:p$
 - Column sums/sum of all observations corresponding to fixed level of factor B i.e. $y_{.j.}$, $j = 1:q$
 - Cell sums i.e. sum of all observations corresponding to $(i,j)^{th}$ cell i.e. $y_{ij.}$, $i = 1:p, j = 1:q$

First test to be performed

- The hypothesis of no interaction is to be tested first. It ultimately tests the validity of the model.
- **What is validity of the model?**

There are pq interaction terms.

- If the interaction terms are significant then the model is correct.
- If the interactions are not significant then the model is **overparametrized** and need to be modified by dropping the interaction terms.

Hypothesis of no interactions

- There is no interaction or the interactions are not significant.
- As γ_{ij} are not estimable we cannot test the hypothesis $H_0: \gamma_{ij} = 0$.

- Rather γ_{ij}^* are estimable and $H_0: \gamma_{ij}^* = 0$ can be tested.

$$\text{Further } \gamma_{ij}^* = 0 \quad \Rightarrow \gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..} = 0$$

$$\Rightarrow \gamma_{ij} = \bar{\gamma}_{i.} + \bar{\gamma}_{.j} - \bar{\gamma}_{..}$$

- Under $H_0: \gamma_{ij}^* = 0$ the model really reduces to model with non interaction terms.

Model under H_0 : $\gamma_{ij}^* = 0$

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \bar{\gamma}_{i.} + \bar{\gamma}_{.j} - \bar{\gamma}_{..}$$

$$\text{as } \gamma_{ij} = \bar{\gamma}_{i.} + \bar{\gamma}_{.j} - \bar{\gamma}_{..}$$

$$E(y_{ijk}) = (\mu - \bar{\gamma}_{..}) + (\alpha_i + \bar{\gamma}_{i.}) + (\beta_j + \bar{\gamma}_{.j})$$

$$**E(y_{ijk}) = \mu^0 + \alpha_i^* + \beta_j^***$$

where

$$\mu^0 = (\mu - \bar{\gamma}_{..})$$

$$\alpha_i^* = (\alpha_i + \bar{\gamma}_{i.})$$

$$\beta_j^* = (\beta_j + \bar{\gamma}_{.j})$$

Testing of hypothesis

- Testing the hypothesis of no interaction

$$H_0: \gamma_{ij}^* = 0$$

- 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 B

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

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

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

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

$$H_{02}: \beta_1^* = \beta_2^*, j \neq v$$

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:

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

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

$$SSH_0 \perp\!\!\!\perp SSE$$

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

Testing of hypothesis $H_0: \gamma_{ij}^* = 0$

- Testing the hypothesis of no interaction

$$H_0: \gamma_{ij}^* = 0$$

- H_0 contains $(p - 1)(q - 1)$ linearly independent estimable parametric functions.
- Hence degrees of freedom associated with SSH_0 are $(p - 1)(q - 1)$.
- **Original model** : $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$
- **Reduced model** : $E(y_{ijk}) = \mu^0 + \alpha_i^* + \beta_j^*$

where $\mu^0 = (\mu - \bar{\gamma}_{..})$,

$$\alpha_i^* = (\alpha_i + \bar{\gamma}_{i.}), \quad \beta_j^* = (\beta_j + \bar{\gamma}_{.j})$$

Testing of hypothesis $H_0: \gamma_{ij}^* = 0 \dots$

- SSE for original model:

$$SSE(OM)$$

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

$$= SST - SS_{cells}$$

Testing of hypothesis $H_0: \gamma_{ij}^* = 0 \dots$

- SSE for reduced model:

$$E(y_{ijk}) = \mu^0 + \alpha_i^* + \beta_j^*$$

- Solution of normal equations

$$\hat{\mu}^0 = \bar{y}_{...}$$

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

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

- Fitted/model value

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu}^0 + \hat{\alpha}_i^* + \hat{\beta}_j^* \\ &= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...})\end{aligned}$$

Testing of hypothesis $H_0: \gamma_{ij}^* = 0 \dots$

$$\begin{aligned} SSE(RM) &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk} - \hat{y}_{ijk})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (y_{ijk} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 \\ &\quad - \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &= SST - SSA - SSB \end{aligned}$$

Testing of hypothesis $H_0: \gamma_{ij}^* = 0 \dots$

$$SSH_0 = SSE_c - SSE$$

$$= SSE(RM) - SSE(OM)$$

$$= SST - SSA - SSB - (SST - SS_{cells})$$

$$= SS_{cells} - SSA - SSB$$

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

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

$$= SS_{AB}$$

$$SS_{AB} = SS_{cells} - SSA - SSB$$

Testing of hypothesis $H_0: \gamma_{ij}^* = 0 \dots$

$$df(SSH_0)$$

$$= df(SS_{AB})$$

$$= df(SSE_c) - df(SSE)$$

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

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

$$SS_{AB} \sim \sigma^2 \chi_{(p-1)(q-1)}^2$$

$$SSE \sim \sigma^2 \chi_{pq(r-1)}^2$$

Further SS_{AB} and SSE are independently distributed.

Testing of hypothesis $H_0: \gamma_{ij}^* = 0 \dots$

- **Test Statistic**

$$MS_{AB} = SS_{AB} / (p - 1)(q - 1)$$

$$MSE = SSE / pq(r - 1)$$

$$\text{Test - Statistic} = \frac{MS_{AB}}{MSE} \sim F_{(p-1)(q-1), pq(r-1)}$$

The testing of hypothesis will be continued...