Mean Model with k treatment (One-way ANOVA)

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i=1,\dots,k; \quad j=1,\dots,n_i$$

μ_i : mean of i-th treatment

Y_{ij} : response ~~of~~ of j-th replicate at i-th trt level, all independent.

ε_{ij} : error of j-th replicate at i-th trt ~~lev~~ level, $\varepsilon_{ij} \overset{iid}{\sim} N(0,\sigma^2)$

k : # of treatment levels

n_i : # of ~~replication~~ observations at i-th trt level

$N = \sum_{i=1}^{k} n_i$: total # of observations

$$X = \begin{pmatrix} 1_{n_1} & & 0 \\ & 1_{n_2} & \\ 0 & & 1_{n_k} \end{pmatrix}$$

In general

$$(X'X) = \begin{pmatrix} n_1 & & 0 \\ & n_2 & \\ 0 & & n_k \end{pmatrix}$$

$\Rightarrow$ Least Square Estimators : $\vec{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$

$$\hat{\vec{\mu}} = (X'X)^{-1} X'Y = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \right)_{i=1,\dots,k} = \begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \vdots \\ \bar{y}_{k\cdot} \end{pmatrix}$$

i.e. the estimate of i-th trtment mean is

$$\hat{\mu}_i = \bar{y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad i=1,\dots,k.$$

← average over replications

$H_0: \mu_1 = \cdots = \mu_k = \mu$ vs H_1: not all μ_i's are the same.

Effect Model with k treatment (fixed)

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0,\sigma^2)$$

$i = 1,\dots,n_i$

effects : $\sum_{i=1}^{k} n_i \tau_i = 0$ $\left[\overset{\text{balanced}}{\sum_{i=1}^{k} \tau_i = 0} \right]$ $H_0: \tau_1 = \cdots = \tau_k = 0$ ← $\left\{ \begin{array}{l} \text{treatment} \\ \text{effects} \end{array} \right.$

$\mu_i = \mu + \tau_i, \quad i=1,\dots,k$ vs. H_1: not all $\tau_i = 0$.

$$Y_{ij} = \mu_i + \varepsilon_{ij} \sim N(\mu_i, \sigma^2) \quad i = 1, \cdots, k. \quad Y_{ij} \text{ are independent.}$$

$$E(Y_{ij}) = \mu_i \quad Var(Y_{ij}) = Var(\varepsilon_{ij}) = \sigma^2$$

LS Criterion:

$$Q_i = \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 \qquad \frac{\partial Q_i}{\partial \mu_i} = -2 \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)$$

Let $\frac{\partial Q_i}{\partial \mu_i} = 0$, solution is $\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \bar{Y}_{i\cdot}$ least square estimate.

Fitted value $\hat{Y}_{i1} = \hat{Y}_{i2} = \cdots = \hat{Y}_{in_i} = \hat{\mu}_i = \bar{Y}_{i\cdot}$

Similarly $Q = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \mu - \tau_i)^2$

$$\frac{\partial Q}{\partial \mu} = 0 \Rightarrow \hat{\mu}_{LSE} = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij} = \bar{Y}, \quad \hat{\tau}_i = \bar{Y}_{i\cdot} - \bar{Y}$$

$$\underbrace{Y_{ij} - \bar{Y}}_{\text{Deviation from overall mean}} = \underbrace{Y_{ij} - \bar{Y}_{i\cdot}}_{\text{Deviation from sample mean (within treatment)}} + \underbrace{\bar{Y}_{i\cdot} - \bar{Y}}_{\text{Difference between sample mean and overall mean (between treatment)}}$$

$$Y_{ij} - \bar{Y}_{i\cdot} = Y_{ij} - \hat{Y}_{ij} \text{ (residual)}$$

$$SSTO = SSE + SSTR$$

$$\sum_i \sum_j (Y_{ij} - \bar{Y})^2 = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2 + \sum_i \sum_j (\bar{Y}_{i\cdot} - \bar{Y})^2$$

Special case (1) $SSE = 0 \Leftrightarrow Y_{ij} = \bar{Y}_{i\cdot}, \ \forall i, j \Leftrightarrow Y_{i1} = Y_{i2} = \cdots = Y_{in_i} \ \forall i$

(2) $SSTR = 0 \Leftrightarrow \bar{Y}_{i\cdot} = \bar{Y} \ \forall i \Leftrightarrow \bar{Y}_1 = \bar{Y}_2 = \cdots = \bar{Y}_K$

$$DF(SSTO) = N-\cancel{k}1$$

$$DF(SSTR) = k-1 \qquad \left(\sum_{i=1}^{k} \tau_i = 0\right)$$

$$DF(SSE) = N-k$$

Equivalence of two system $H_0: \mu_1 = \cdots = \mu_k$

$$\begin{cases} \mu_1 = \mu_2 \\ \mu_1 = \mu_3 \\ \vdots \\ \mu_1 = \mu_k \end{cases} \qquad (k-1) \text{ linearly independent equations.}$$

$$SSE = \sum_{i=1}^{k} S_i^2 \cdot (n_i - 1), \qquad S_i^2 : i\text{-th sample variance.}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \qquad Y_{ij} \overset{ind.}{\sim} N(\mu_i, \sigma^2) \xRightarrow{i\text{-th level}} \frac{(n_i-1) S_i^2}{\sigma^2} \sim \chi^2(n_i - 1)$$

$$\xRightarrow{\text{additivity of } \chi^2} \sum_{i=1}^{k} \frac{(n_i-1) S_i^2}{\sigma^2} \sim \chi^2\left(\sum_{i=1}^{k} (n_i - 1)\right)$$

$$\Rightarrow \frac{SSE}{\sigma^2} \sim \chi^2(N-k)$$

$$E\left[\frac{SSE}{\sigma^2}\right] = N-k, \Rightarrow E\left[\frac{SSE}{N-k}\right] = \sigma^2$$

$$\therefore \hat{\sigma}^2 = MSE = \frac{SSE}{N-k}$$

$n_1 = n_2 = \cdots = n_k = n$ (balanced case)

$$E[MSTR] = E\left[\sum_{i=1}^{k} n(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right] / (k-1)$$

$$= \frac{n}{k-1} E\left[\sum_{i=1}^{n} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right]$$

$$\bar{Y}_{i\cdot} = \mu + \tau_i + \bar{\varepsilon}_{i\cdot}$$

$$\bar{Y}_{\cdot\cdot} = \mu + \bar{\tau} + \bar{\varepsilon}_{\cdot\cdot}$$

$$\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = \tau_i - \bar{\tau} + \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}$$

$$\sum_{i=1}^{k} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \underbrace{\sum_{i=1}^{k} (\tau_i - \bar{\tau})^2}_{(\mathrm{I})} + \underbrace{\sum_{i=1}^{k} (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}_{(\mathrm{II})} + \underbrace{2\sum_{i=1}^{k} (\tau_i - \bar{\tau})(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})}_{(\mathrm{III})}$$

$H_0: \tau_1 = \cdots = \tau_k = 0$

$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, \dots, k$, $j = 1, \dots, n$

$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

$\bar{\varepsilon}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{ij} \sim N(0, \frac{\sigma^2}{n})$

$\bar{\tau} = \frac{1}{k} \sum_{i=1}^{k} \tau_i$

$\bar{\varepsilon}_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} \varepsilon_{ij}$, $N = n \cdot k$

$\sim N(0, \frac{\sigma^2}{N})$

(I) $\bar{\tau} = \frac{1}{k} \sum_{i=1}^{k} \tau_i \xrightarrow[\sum_{i=1}^{k} n_i \tau_i = 0]{\text{restriction}} 0$, $\therefore E\left[\sum_{i=1}^{k} (\tau_i - \bar{\tau})^2\right] = \sum_{i=1}^{n} \tau_i^2$

(II) $E\left[\sum_{i=1}^{k} (\tau_i - \bar{\tau})(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})\right] = \sum_i (\tau_i - \bar{\tau}) \cdot E(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) = 0$

(III) $E\left[\sum_{i=1}^{k} (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2\right]$

$= E\left[\sum_{i=1}^{k} (X_i - \bar{X})^2\right]$

$= (k-1)\sigma^2/n$

Let $X_i = \bar{\varepsilon}_{i\cdot}$, $\bar{X} = \frac{1}{k} \sum_{i=1}^{k} X_i = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} \varepsilon_{ij} = \bar{\varepsilon}_{\cdot\cdot}$

$X_i \overset{iid}{\sim} N(0, \frac{\sigma^2}{n})$, $\frac{\sum_{i=1}^{k} (X_i - \bar{X})^2}{\sigma^2/n} \sim \chi^2(k-1)$

$\therefore E\left[\sum_{i=1}^{k} (X_i - \bar{X})^2\right] = (k-1)\sigma^2/n$

$$\therefore E\left[\sum_{i=1}^{k} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right] = \sum_{i=1}^{k} \tau_i^2 + (k-1)\sigma^2$$

$$E[MSTR] = \frac{n}{k-1} \cdot E\left(\sum_{i=1}^{k} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right)$$

$$= \frac{n}{k-1} \left(\sum_{i=1}^{k} \tau_i^2 + (k-1) \cdot \frac{\sigma^2}{n}\right)$$

$$= \frac{1}{k-1} \sum_{i=1}^{k} n \tau_i^2 + \sigma^2$$

In general, $E[MSTR] = \sigma^2 + \frac{1}{k-1} \sum_{i=1}^{k} n_i \cdot \tau_i^2$