Analysis of Variance - RCBD

(1) Parameter Estimator : Least Square / Maximum Likelihood

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

$i = 1, \cdots, k$, $j = 1 \cdots, b$

$\sum_{i=1}^{k} \tau_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$ (both fixed factors)

$\Rightarrow \hat{\mu} = \bar{Y}_{\cdot\cdot}$, $\hat{\tau}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$, $\hat{\beta}_j = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}$

Then fitted value: $\hat{Y}_{ij} = \hat{\mu}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j$

$$\hat{Y}_{ij} = \bar{Y}_{i\cdot} + \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}$$

Residual, $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}$

(2) S.S. Decomposition

Note that: $\sum_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) = 0$, $\sum_j (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}) = 0$

$\sum_j (Y_{ij} - \bar{Y}_{i\cdot}) = 0 \quad \forall i = 1 \cdots k \quad \Rightarrow \sum_j e_{ij} = 0, \ \forall i = 1 \cdots k$

$\sum_i (Y_{ij} - \bar{Y}_{\cdot j}) = 0 \quad \forall j = 1 \cdots b \quad \Rightarrow \sum_i e_{ij} = 0, \ \forall j = 1 \cdots b$

$\sum_i \sum_j e_{ij} = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}) = 0$

$$\left[\sum_i \sum_j Y_{ij} = b \sum_i \bar{Y}_{i\cdot} = k \cdot \sum_j \bar{Y}_{\cdot j} = bk \cdot \bar{Y}_{\cdot\cdot} \right]$$

Deviation Decomposition

$$Y_{ij} - \bar{Y}_{\cdot\cdot} = (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) + (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}) + (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})$$

$$Y_{ij} - \bar{Y}_{\cdot\cdot} = \underset{(\hat{\tau}_i)}{A_i} + \underset{(\hat{\beta}_j)}{B_j} + e_{ij}$$

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i \sum_j (A_i + B_j + e_{ij})^2$$

$$= \sum_i \sum_j A_i^2 + \sum_i \sum_j B_j^2 + \sum_i \sum_j e_{ij}^2$$

$$+ 2\left(\sum_i \sum_j A_i e_{ij} + \sum_i \sum_j B_j e_{ij} + \sum_i \sum_j A_i B_j\right)$$

It can be shown that

$$\sum_i \sum_j A_i e_{ij} = \sum_i A_i \cdot \left(\sum_j e_{ij}\right) = 0$$

$$\sum_i \sum_j B_j e_{ij} = \sum_j B_j \cdot \left(\sum_i e_{ij}\right) = 0$$

$$\sum_i \sum_j A_i B_j = \left(\sum_i A_i\right)\left(\sum_j B_j\right) = 0$$

$$\therefore \sum_{i=1}^{k} \sum_{j=1}^{b} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{k} \sum_{j=1}^{b} (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{b} e_{ij}^2$$

$$\sum_{i=1}^{k} \sum_{j=1}^{b} (Y_{ij} - \bar{Y}_{..})^2 = b \sum_{i=1}^{k} (\bar{Y}_{i.} - \bar{Y}_{..})^2 + k \sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{b} e_{ij}^2$$

$$SSTO = SSTR + SSB + SSE$$

And $DF(SSTO) = kb - 1 = 0$

$DF(SSB) = b - 1$

$DF(SSTR) = k - 1$

$DF(SSE) = (b-1)(k-1)$

restriction

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..}) = 0$$

$$\sum_j (\bar{Y}_{.j} - \bar{Y}_{..}) = 0$$

$$\sum_i (\bar{Y}_{i.} - \bar{Y}_{..}) = 0$$

$$\begin{cases} \sum_i e_{ij} = 0 & \forall j = 1, \dots, b \\ \sum_j e_{ij} = 0 & \forall i = 1, \dots, k \end{cases}$$

Based on Cochran's Theorem,

$\frac{SSTR}{\sigma^2}$, $\frac{SSB}{\sigma^2}$, $\frac{SSE}{\sigma^2}$ are independent χ^2 random variables.

i.e. $\frac{SSTR}{\sigma^2} \sim \chi^2(k-1)$, $\frac{SSB}{\sigma^2} \sim \chi^2(b-1)$, $\frac{SSE}{\sigma^2} \sim \chi^2((k-1)(b-1))$

and $SSTR \perp\!\!\!\perp SSB \perp\!\!\!\perp SSE$ (mutually independent)

Under $H_0: \tau_1 = \cdots = \tau_k = 0$, H_1: not all $\tau_i = 0$

$$F_{TR} = \frac{MSTR}{MSE} \sim F(k-1, (k-1)(b-1))$$

Under $H_0: \beta_1 = \cdots = \beta_b = 0$ H_1: not all $\beta_j = 0$

$$F_B = \frac{MSB}{MSE} \sim F(b-1, (k-1)(b-1))$$

And $$E[MSE] = E\left[\frac{SSE}{(k-1)(b-1)}\right] = \sigma^2$$

$$E[MSTR] = E\left[\frac{SSTR}{k-1}\right] = \sigma^2 + b\cdot\left(\frac{1}{k-1}\sum_{i=1}^{k}\tau_i^2\right)$$

$$E[MSB] = E\left[\frac{SSB}{b-1}\right] = \sigma^2 + k\cdot\left(\frac{1}{b-1}\sum_{j=1}^{b}\beta_j^2\right)$$

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

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

$$\Rightarrow Y_{ij} \overset{ind.}{\sim} N(\mu + \tau_i + \beta_j, \sigma^2)$$

Treatment mean at i-th level

$$\bar{Y}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b} Y_{ij}$$

mean $E(\bar{Y}_{i\cdot}) = \frac{1}{b}\sum_{j=1}^{b}(\mu+\tau_i+\beta_j) \overset{\sum_j \beta_j = 0}{=} \mu + \tau_i$

variance $Var(\bar{Y}_{i\cdot}) \overset{ind}{=} \frac{1}{b^2}\sum_{j=1}^{b} Var(Y_{ij}) = \frac{\sigma^2}{b}$

$$\bar{Y}_{i\cdot} \overset{ind.}{\sim} N\left(\mu+\tau_i, \frac{\sigma^2}{b}\right)$$

Under $H_0: \tau_1 = \cdots = \tau_k = 0$, $\bar{Y}_{i\cdot} \overset{H_0}{\sim} N\left(\mu, \frac{\sigma^2}{b}\right)$.

Grand mean $\bar{Y}_{\cdot\cdot} = \frac{1}{kb}\sum_i\sum_j Y_{ij} = \frac{1}{k}\sum_{i=1}^{k}\bar{Y}_{i\cdot} \sim N\left(\mu, \frac{\sigma^2}{kb}\right)$

Under H_0: $\bar{Y}_{\cdot\cdot}$ is an average of $\bar{Y}_{i\cdot}$ (k independent observations with same mean).

Following Student's t-theorem

$$\frac{\sum_{i=1}^{k}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2}{\left(\frac{\sigma^2}{b}\right)} \sim \chi^2(k-1)$$

$X_i = \bar{Y}_{i\cdot}$ $\quad \bar{X} = \bar{Y}_{\cdot\cdot}$

$Var(X_i) = \frac{\sigma^2}{b}$, $i = 1, \cdots, b$

$$E\left\{\frac{1}{k-1}\left[b\cdot\sum_{i=1}^{k}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})\right]^2\right\} \overset{H_0: \tau_i = 0\ \forall i}{=\!=\!=} \sigma^2$$

i.e. $E[MSTR] = \sigma^2$ if $\tau_1 = \cdots = \tau_k = 0$ is true.

For a general τ_i (may not be 0).

$$\bar{Y}_{i\cdot} - \tau_i \overset{\text{ind.}}{\sim} N\left(\mu, \frac{\sigma^2}{b}\right)$$

$$\bar{Y}_{\cdot\cdot} = \frac{1}{k}\sum_{i=1}^{k}(\bar{Y}_{i\cdot} - \tau_i) \sim N\left(\mu, \frac{\sigma^2}{kb}\right)$$

Then $\frac{1}{k-1} E[SSTR]$

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

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

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

Note that τ_i is a parameter (constant) and

$$E(\bar{Y}_{i\cdot} - \tau_i - \bar{Y}_{\cdot\cdot}) = \mu - \mu = 0$$

$\therefore$

$$E[MSTR] = E\left[\frac{SSTR}{k-1}\right]$$

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

Similarly we can show that

$$E[MSB] = \sigma^2 + \frac{k}{b-1}\sum_{j=1}^{b}\beta_j^2 \quad \text{(Homework Problem)}$$