One way Classification model

- There are a total of n observations
- There are k treatments
- Suppose there are n_i observations for treatment i. $(n_1 + n_2 + \cdots + n_k = n)$

A linear model

$$\underbrace{y_{ij}}_{(i,j)^{th} \text{ observation}} = \underbrace{\mu}_{\text{overall mean}} + \underbrace{\mu_i}_{\text{effect of treatment } i} + \underbrace{\varepsilon_{ij}}_{(i,j)^{th} \text{ error term}}$$

$1 \le i \le k$

$1 \le j \le n_i$

Note: "i" $\equiv$ stands for i^{th} treatment

"j" $\equiv$ stands for the replications made for application of the i^{th} treatment

Assumptions:

$1 \le i \le k, \ 1 \le j \le n_i$

(i) $E[\varepsilon_{ij}] = 0$

(ii) $Var(\varepsilon_{ij}) = \sigma^2$

(iii) $Cov(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0 \quad \forall\, i \ne i'$ or $j \ne j'$

This linear model is called One-way Classification model.

Estimation

- How many estimable linearly independent linear parametric functions are there?

(Ex) Ans = k

e.g. (one such)

$$\begin{pmatrix} \mu + \sum_{i=1}^{k} \mu_i \\ \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \\ \vdots \\ \mu_{k-1} - \mu_k \end{pmatrix}$$

← Overall mean effect

$k-1$ constructs

Estimates

(Ex.)

$$\widehat{(\mu + \mu_i)}_{L.S.} = \bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \frac{T_{i.}}{n_i}$$

$$\widehat{(\mu_i - \mu_{i'})}_{L.S} = \bar{y}_{i.} - \bar{y}_{i'.}$$

$$Var\left(\widehat{(\mu_i - \mu_{i'})}_{LS}\right) = \left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\sigma^2$$

$$\widehat{\left(\mu + \sum_{i=1}^{k} \mu_i\right)}_{LS} = \bar{y}_{..} = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{y}_{i.} = \frac{T_{..}}{n}$$

$$Var\left(\widehat{\left(\mu + \sum_{i=1}^{k} \mu_i\right)}_{LS}\right) = \frac{1}{n^2} \sum_{i=1}^{k} n_i \sigma^2 = \frac{\sigma^2}{n}$$

$$R_0^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i}$$

<u>Test</u>: $H_0 : \mu_1 = \mu_2 = \dots \mu_k$

$$R_1^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1} \frac{T_{..}^2}{n_i}$$

$$\bar{y}_{..} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}}{n} \qquad \bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

$$\underbrace{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2}_{SS_{Total}} =$$

(Cross terms are 0)

$$= \underbrace{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2}_{SS_W} + \underbrace{\sum_{i=1}^{k} n_i (\bar{y}_{i.} - \bar{y}_{..})^2}_{SS_B}$$

<u>Assume</u>: · $n_i = J \quad \forall\ 1 \le i \le k$

· ε_{ij} are Normally distributed

(Ex.) $\cdot \quad \dfrac{SS_W}{\sigma^2} \sim \chi^2_{k(J-1)} \equiv \chi^2_{n-k}$

- $$\frac{SS_B}{\sigma^2} \sim \chi^2_{k-1}$$

- SS_B is independent of SS_W

$$\Rightarrow \quad \frac{\frac{SS_B}{k-1}}{\frac{SS_W}{n-k}} \sim F(k-1, n-k)$$

- Test: - Decide $\alpha = 0.05$ (say)

  - p-value $:=$

$$\mathbb{P}\left(F(k-1, n-k) > \frac{SS_B}{k-1} \Big/ \frac{SS_W}{n-k} \right)$$

- If p-value $< \alpha$ then you reject the null hypothesis.

General case $\quad 1 \leq i \leq k \qquad 1 \leq j \leq n_i$

$$SS_W = \sum_{i=1}^{k} \underbrace{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2}_{\text{degrees of freedom} = n_i - 1}$$

$$\boxed{\sum_{i=1}^{k} n_i - 1 = n - k}$$

Ex:

- $E\left[\frac{SS_W}{n-k}\right] = \sigma^2$

- $\frac{SS_W}{\sigma^2} \sim \chi^2_{n-k}$

- $MS_W = \frac{SS_W}{n-k}$

$$SS_B = \sum_{i=1}^{k} n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$MS_B = \frac{SS_B}{k-1}$$

$$\frac{SS_B}{\sigma^2} \sim \chi^2_{k-1}$$

- $F(k-1, n-k) \sim \frac{MS_B}{MS_W}$

Test: Reject if

$$\mathbb{P}\left(F(k-1, n-k) > \frac{MS_B}{MS_W} \right) < \alpha$$

for given α.

Anova table

Source	Sum of Squares	Degrees of freedom	MS	F
Treatment ($\mu_1 = \mu_2 \cdots = \mu_k$)	SS_B	$k-1$	$\frac{SS_B}{k-1}$	$\frac{SS_B}{k-1} \Big/ \frac{SS_W}{n-k}$
Error Residual	SS_W	$n-k$	$\frac{SS_W}{n-k}$	
Total	SS_T	$n-1$		

Two-way classification model

Model with No interaction and one observation / cell

Treatment \ Block	1st Block	.. ..	J^{th} Block	
1st treatment				$y_{1.}$
⋮		$\boxed{y_{ij}}$		
I^{th} treatment	$y_{.1}$		$y_{.J}$	$I \times J$ $y_{..}$

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad 1 \le i \le I, \quad 1 \le j \le J$$

$n = IJ$

- $m = 1 + I + J$

- $r = 1 + (I-1) + (J-1) = I + J - 1$

Estimable linearly independent linear parametric functions

$$\mu + \bar{\alpha} + \bar{\beta} = \mu + \frac{1}{I}\sum_{i=1}^{I} \alpha_i + \frac{1}{J}\sum_{j=1}^{J} \beta_j$$

$$\alpha_i - \alpha_{i'} \quad \forall\ 1 \le i \ne i' \le I$$

$$\beta_j - \beta_{j'} \quad \forall\ 1 \le j \ne j' \le J$$

$$E(\mu, \underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_{i=1}^{I}\sum_{j=1}^{J} (y_{ij} - \mu - \alpha_i - \beta_j)^2$$

Use calculus or otherwise

- $\hat{\mu} + \hat{\bar{\alpha}} + \hat{\bar{\beta}} = \bar{y}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij} = \frac{T_{..}}{IJ}$

- $\hat{\mu} + \hat{\alpha}_i + \hat{\bar{\beta}} = \bar{y}_{i.} = \frac{T_{i.}}{I} \quad \forall\ 1 \le i \le I$

- $\hat{\mu} + \hat{\bar{\alpha}} + \hat{\beta}_j = \bar{y}_{.j} = \frac{T_{.j}}{J} \quad \forall\ 1 \le j \le J$

LSE:

- $\mu + \hat{\bar{\alpha}} + \hat{\bar{\beta}} = \bar{y}_{..}$

- $\widehat{\alpha_i - \alpha_{i'}} = \bar{y}_{i.} - \bar{y}_{i'.}$

- $\widehat{\beta_j - \beta_{j'}} = \bar{y}_{.j} - \bar{y}_{.j'}$

- $\widehat{\mu + \alpha_i + \beta_j} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$

<u>Testing</u> :

$$R_0^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \sim \sigma^2 \chi^2_{(I-1)(J-1)}$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{..})^2 - J \sum_{i=1}^{I} (\bar{y}_{i.} - \bar{y}_{..})^2 - I \sum_{j=1}^{J} (\bar{y}_{.j} - \bar{y}_{..})^2$$

$H_0^\alpha := \alpha_1 = \alpha_2 = \cdots = \alpha_I$ (Treatment)

$$R_{1\alpha}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{.j})^2$$

(Ex) $R_{1\alpha}^2 - R_0^2 = J \sum_{i=1}^{I} (\bar{y}_{i.} - \bar{y}_{..})^2$

$$H_0^{\beta} := \beta_1 = \beta_2 = \cdots = \beta_J$$

$$R_{1\beta}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{i\cdot})^2$$

(Ex) $$R_{1\beta}^2 - R_0^2 = I \sum_{j=1}^{J} (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$$

ANOVA Table

Source	SS	df	MS	
Treatments $\alpha_1 = \alpha_2 \cdots = \alpha_I$	$S_\alpha = J \sum_{i=1}^{I} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$I-1$	$\frac{S_\alpha}{I-1}$	$F_\alpha = (J-1)\,\frac{S_\alpha}{S_c}$
Blocks $\beta_1 = \beta_2 = \cdots \beta_J$	$S_\beta = I \sum_{j=1}^{J} (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$	$J-1$	$\frac{S_\beta}{J-1}$	$F_\beta = (I-1)\,\frac{S_\beta}{S_c}$
Cell H_0^α & H_0^β	$S_c = S_\alpha + S_\beta$	$I+J-2$	$S_c / I+J-2$	
Error (Residual)	$S_e = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} \; \bar{y}_{\cdot\cdot})^2$	$(I-1)(J-1)$	$S_e/(I-1)(J-1)$	
Total	$S_t = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_{\cdot\cdot})^2$	$IJ-1$		

$x = 1:100$

$y = c \qquad y = c + \varepsilon$

$\ln(y - x)$

$x = 1:100$

$y = 3x + \varepsilon$

$\ln(y \sim x)$

.. 0 0