Two-way ANOVA

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \qquad \text{means model}$$

$$= \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \text{main effect interaction model}$$

$\alpha_i = \mu_{i.} - \mu_{..}$ effect due to factor A

$\beta_j = \mu_{.j} - \mu_{..}$ effect due to factor B

$\mu_{..}$ overall effect

$$i = 1, \cdots a, \quad j = 1, \cdots b, \quad k = 1, \cdots n_{ij}$$

$$n_T = \sum_i \sum_j n_{ij}$$

If none levels of factor A interact the levels of factor B

$$\mu_{ij} - \mu_{im} = \mu_{\ell j} - \mu_{\ell m} \quad \text{for all } i, j, \ell, m$$

no interaction.

If factors A and B have no interaction, then

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j \quad \text{for all } i, j$$

proof: $\mu_{ij} - \mu_{..} - \alpha_i - \beta_j = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$

$$= \mu_{ij} - \frac{1}{b}\sum_{m=1}^{b} \mu_{im} - \frac{1}{a}\sum_{l=1}^{a} \mu_{lj} + \frac{1}{ab}\sum_{l}\sum_{m} \mu_{lm}$$

$$= \frac{1}{ab}\sum_{l}\sum_{m} \mu_{ij} - \frac{1}{ab}\sum_{l=1}^{a}\sum_{m} \mu_{im} - \frac{1}{ab}\sum_{l}\sum_{m} \mu_{lj} + \frac{1}{ab}\sum_{l}\sum_{m} \mu_{lm}$$

$$= \frac{1}{ab}\sum_{l}\sum_{m} (\underbrace{\mu_{ij} - \mu_{im} - \mu_{lj} + \mu_{lm}}_{0})$$

$$= 0$$

Fixed effects Model

$$y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

$E\,\varepsilon_{ijk} = 0$,

Normal, $Var(\varepsilon_{ijk}) = \sigma^2$, $cov(\varepsilon_{ijk}, \varepsilon_{lmn}) = 0$

Estimation

Least square $\sum_i \sum_j \sum_k (y_{ijk} - \mu_{ij})^2 \longrightarrow \min\{\mu_{ij}\}$

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

or $\sum_i \sum_j \sum_k (y_{ijk} - \mu_{..} - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2$

subject to $\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$

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

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

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

$$\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

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

Fitted value

$$\hat{y}_{ijk} = \bar{y}_{ij.}$$

ANOVA (Balanced case) $(n_{ij} \equiv n)$

$$SSTO = SSTR + SSE \quad \text{where}$$

$$SSTO = \sum_i \sum_j \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{...})^2 \qquad (n_T - 1)\ df$$

$$SSTR = \sum_i \sum_j \sum_{k=1}^{n_{ij}} (\bar{y}_{ij.} - \bar{y}_{...})^2$$

$$= \sum_i \sum_j n_{ij} (\bar{y}_{ij.} - \bar{y}_{...})^2$$

with 1 constraint $\sum_i \sum_j n_{ij}(\bar{y}_{ij.} - \bar{y}_{...}) = 0$

$(ab - 1)\ df$

$$SSE = \sum_i \sum_j \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij\cdot})^2$$

$$= \sum_i \sum_j (n_{ij}-1) S_{ij}^2 \qquad (n_T - ab)\ df$$

proof: $$SSTO = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij\cdot} + \bar{y}_{ij\cdot} - \bar{y}_{\cdots})^2$$

$$= \sum\sum\sum (y_{ijk} - \bar{y}_{ij\cdot})^2 + \sum\sum\sum (\bar{y}_{ij\cdot} - \bar{y}_{\cdots})^2$$

$$+ 2 \sum\sum\sum (y_{ijk} - \bar{y}_{ij\cdot})(\bar{y}_{ij\cdot} - \bar{y}_{\cdots})$$

Note $$\sum_i \sum_j \sum_k \underbrace{(y_{ijk} - \bar{y}_{ij\cdot})}(\bar{y}_{ij\cdot} - \bar{y}_{\cdots})$$

$$= \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{\cdots}) \underbrace{\sum_k (y_{ijk} - \bar{y}_{ij\cdot})}_{0}$$

$$= 0$$

- Partition of SSTR (balanced $n_{ij} \equiv n$)

$$SSTR = \sum_i \sum_j n (\bar{y}_{ij\cdot} - \bar{y}_{\cdots})^2$$

$$= n \sum_i \sum_j \{ (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}) \}^2$$

$$= n \sum_i b (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2 + n a \sum_j (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$$

$$+ n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2 + \text{cross products}$$

$$= nb \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + na \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$+ n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$= SSA + SSB + SSAB$$

· Cross products part

(1) $\sum_i \sum_j (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...}) = 0$

(2) $\sum_i \sum_j \{ (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) \}$

$$= \sum_i (\bar{y}_{i..} - \bar{y}_{...}) \sum_{j=1}^{b} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})$$

$$b(\bar{y}_{i..} - \bar{y}_{i..} - \bar{y}_{...} + \bar{y}_{...}) = 0$$

$= 0$

(3) similarly $= 0$

Expected Mean Square (balanced)

$$MSE = \frac{\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2}{n_T - ab} = \frac{\sum_i \sum_j (n_{ij} - 1) S_{ij}^2}{\sum_i \sum_j (n_{ij} - 1)}$$

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

$$MSA = \frac{\sum\sum\sum(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2}{a-1} = \frac{bn\sum_i \hat{\alpha}_i^2}{a-1}$$

Consider a 1-way ANOVA model

$$\bar{y}_{i\cdot k} = \mu_i^* + \varepsilon_{ik}^*, \qquad \varepsilon_{ik}^* \sim N(0, \frac{\sigma^2}{b})$$

$$Y \sim N(0, \sigma^2)$$

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

$$SSTR = SSA + SSB + SSAB$$

$$= nb\sum_i(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 + na\sum_j(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$$

$$+ n\sum_i\sum_j(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot})^2$$

$$(ab-1) = a-1 + b-1 + (a-1)(b-1)$$

two way ANOVA data

y_{111}	...	...	y_{1b1}	$\bar{y}_{1\cdot 1}$
$\vdots$			$\vdots$	$\vdots$
y_{11n}	...	...	y_{1bn}	$\bar{y}_{1\cdot n}$
$\vdots$	$\ddots$	$\ddots$	$\vdots$	
y_{a11}	...	...	y_{ab1}	$\bar{y}_{a\cdot 1}$
$\vdots$			$\vdots$	$\vdots$
y_{a1n}	...	...	y_{abn}	$\bar{y}_{a\cdot n}$

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

$$\bar{y}_{i\cdot k} = \mu_{i\cdot} + \bar{\varepsilon}_{i\cdot k}$$

$$\bar{\varepsilon}_{i\cdot k} \sim N(0, \frac{\sigma^2}{b})$$

$$SS_{TR}^{*} = \sum_i n(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$$

$$E\,MS_{TR}^{*} = \frac{\sigma^2}{b} + \frac{n\sum_i(\mu_{i\cdot} - \mu_{\cdot\cdot})^2}{a-1}$$

$$E\,SS_{TR}^{*} = (a-1)\frac{\sigma^2}{b} + n\sum_i(\mu_{i\cdot} - \mu_{\cdot\cdot})^2$$

$$SS_{TR}^{*} = \sum_j n(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$$

$$E\,SS_{TR}^{*} = (b-1)\frac{\sigma^2}{a} + n\sum_j(\mu_{\cdot j} - \mu_{\cdot\cdot})^2$$

$$SS_{TR}^{*} = \sum_j n(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2 \qquad i = 1, \cdots a$$

$$E\,SS_{TR}^{*} = (b-1)\sigma^2 + n\sum_j(\mu_{ij} - \mu_{i\cdot})^2$$

Two way ANOVA

$$E\,MSA = \frac{1}{a-1}E\,nb\sum_i(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$$

$$= \frac{b}{a-1}\left[(a-1)\frac{\sigma^2}{b} + n\sum_i(\mu_{i\cdot} - \mu_{\cdot\cdot})^2\right] = \sigma^2 + nb\frac{\sum(\mu_{i\cdot} - \mu_{\cdot\cdot})^2}{a-1}$$

$$E\,MSB = \sigma^2 + na\frac{\sum(\mu_{\cdot j} - \mu_{\cdot\cdot})^2}{b-1}$$

$$E\,MSAB = \frac{1}{(a-1)(b-1)} E\, n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2$$

$$= \frac{\cancel{E}\ 1}{(a-1)(b-1)} E \Big\{ n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2 + na \sum_j (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$$

$$- 2n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \Big\}$$

$$= \frac{1}{(a-1)(b-1)} E \Big\{ n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2 - na \sum_j (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2 \Big\}$$

$$= \frac{1}{(a-1)(b-1)} \Big[\sum_i \Big\{ (b-1)\sigma^2 + n \sum_j (\mu_{ij} - \mu_{i\cdot})^2 \Big\} - a(b-1)\frac{\sigma^2}{a} - an \sum_j (\mu_{\cdot j} - \mu_{\cdot\cdot})^2 \Big]$$

$$= \frac{1}{(a-1)(b-1)} \Big[(a-1)(b-1)\sigma^2 + n \sum_i \sum_j (\mu_{ij} - \mu_{i\cdot})^2 - an \sum_j (\mu_{\cdot j} - \mu_{\cdot\cdot})^2 \Big]$$

$$\overset{(*)}{=} \sigma^2 + \frac{n}{(a-1)(b-1)} \sum_i \sum_j (\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot})^2$$

$$= \sigma^2 + \frac{n \sum_i \sum_j (\alpha\beta)_{ij}^2}{(a-1)(b-1)}$$

(*) because $\sum_i \sum_j (\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot})^2$

$$= \sum_i \sum_j (\mu_{ij} - \mu_{i\cdot})^2 + \sum_i \sum_j (\mu_{\cdot j} - \mu_{\cdot\cdot})^2$$

$$-2\sum_i\sum_j(\mu_{ij}-\mu_{i\cdot})(\mu_{\cdot j}-\mu_{\cdot\cdot})$$

$$=\sum_i\sum_j\left\{(\mu_{ij}-\mu_{i\cdot})^2-(\mu_{\cdot j}-\mu_{\cdot\cdot})^2\right\}$$

Testing A effects, B effects, and interaction

So we test. $H_0: \alpha_i \equiv 0$, with $F_A = \dfrac{MSA}{MSE}$

$H_0: \beta_j \equiv 0$ with $F_B = \dfrac{MSB}{MSE}$

$H_0: (\alpha\beta)_{ij} \equiv 0$ with $F_{AB} = \dfrac{MSAB}{MSE}$

all $\overset{H_0}{\sim}$ F distr. with suitable d.f.

But: Multipul comparisons (3 test at the same stage)

(1). Bonfemonni, if 3 tests are done with sig. levels $\alpha_1, \alpha_2, \alpha_3$, then

$$FWER = \alpha \leq \alpha_1 + \alpha_2 + \alpha_3$$

(2). Improvement, Kimball inequality

$$\text{FWER } \alpha \le 1 - (1-\alpha_1)(1-\alpha_2)(1-\alpha_3)$$

If $\alpha_1 = \alpha_2 = \alpha_3 = .1$, $\alpha_{Bonf} \le .3$

$\alpha_{Kimball} \le .271$

$\alpha_1 = \alpha_2 = \alpha_3 = .05$, $\alpha_{Bonf} \le .15$

$\alpha_{Kimball} \le .1426$

So if $\alpha_1 = \alpha_2 = \alpha_3$, then

$$\text{FWER } \alpha \le 1-(1-\alpha_i)^3$$

We choose $\alpha_i = 1-(1-\alpha)^{1/3}$

· Kimball inequality

Let $\xi_0, \xi_1, \xi_2, \xi_3$ ~ independent, >0

(There are MSE, MSA, MSB, MSAB under H_0)

F_1, F_2, F_3 are some positive numbers (critical values).

Then: $$P\left\{\bigcap_{k=1}^{3}\left(\frac{\xi_k}{\xi_0} \le F_k\right)\right\} \ge \prod_{k=1}^{3} P\left\{\frac{\xi_k}{\xi_0} \le F_k\right\}$$

Pf: Let $\zeta_k := P\{\frac{\xi_k}{\xi_0} \le F_k \mid \xi_0\}$; $\eta_k = 1_{\{\frac{\xi_k}{\xi_0} \le F_k\}}$

all ζ_k are non-decreasing functions of ξ_0

If both f, g monotone in the same direction

$$(f(x_2) - f(x_1))(g(x_2) - g(x_1)) \ge 0$$

Then $E(f(X)g(X)) \ge Ef(X)\,Eg(X)$

or $Cov(f(X), g(X)) \ge 0$ (positively correlated)

Pf: X_1, X_2 iid r.v.s. of X

$$0 \le E\left[\{f(X_2) - f(X_1)\}\{g(X_2) - g(X_1)\}\right]$$

$$= E\{f(X_2)g(X_2) - f(X_1)g(X_2) - f(X_2)g(X_1) + f(X_1)g(X_1)\}$$

$$= 2\{E(f(X)g(X)) - Ef(X)\,Eg(X)\}$$

Conditional on ξ_0, $\{\eta_k\}$ are indep.

Then $E(\eta_1 \eta_2 \eta_3) = P\left\{ \bigcap_{k=1}^{3} \left(\frac{\xi_k}{\xi_0} \le F_k \right) \right\}$

$$= E\left(E(\eta_1 \mid \xi_0)\, E(\eta_2 \mid \xi_0)\, E(\eta_3 \mid \xi_0) \right)$$

$$= E(\zeta_1 \zeta_2 \zeta_3)$$

$$\ge E(\zeta_1)\, E(\zeta_2 \zeta_3) \ge E\zeta_1\, E\zeta_2\, E\zeta_3$$

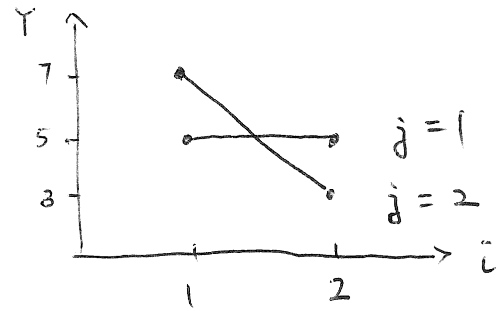
$$= \prod_{k=1}^{3} P\left\{ \frac{\xi_k}{\xi_0} \le F_k \right\}$$

Ex: A clinical trial is used to find the optimal combination of 2 antibiotics A & B. Each antibiotics can be in 2 forms: powder or liquid. Each combinations was given to 2 patients as following

Y_{ijk}		$i=1$		$i=2$	
j	1	4	6	4	6
	2	6	8	2	4

$\bar{Y}_{ij\cdot}$		$i=1$	$i=2$
j	1	5	5
	2	7	3

a). plot the data to show the interaction.



b) Test on the interaction. ($H_0 : (\alpha\beta)_{ij} \equiv 0$

$$F = \frac{SSAB/1}{SSE/4} = \frac{8}{8/4} = 4 \quad \text{with d.f } (1,4)$$

P-value $> .1$

Fail to reject H_0. no significant interaction