· Estimation

$$\hat{\mu}_{ijk} = \bar{y}_{ijk\cdot}$$

$$\hat{\alpha}_i = \bar{y}_{i\cdots} - \bar{y}_{\cdots\cdot}, \quad \hat{\beta}_j = \quad , \quad \hat{\gamma}_k =$$

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

$$\widehat{(\alpha\beta\gamma)}_{ijk} = \bar{y}_{ijk\cdot} - \bar{y}_{ij\cdot\cdot} - \bar{y}_{i\cdot k\cdot} - \bar{y}_{\cdot jk\cdot} + \bar{y}_{i\cdots} + \bar{y}_{\cdot j\cdot\cdot} + \bar{y}_{\cdot\cdot k\cdot} - \bar{y}_{\cdots\cdot}$$

· ANOVA (Balanced $n_{ijk} \equiv n$)

$$\begin{aligned} SSTO &= SSTR + SSE \\ &= \underset{(a-1)}{SSA} + \underset{(b-1)}{SSB} + \underset{(c-1)}{SSC} + \underset{(a-1)(b-1)}{SSAB} + \underset{(b-1)(c-1)}{SSBC} + \underset{(a-1)(c-1)}{SSAC} + \underset{(a-1)(b-1)(c-1)}{SSABC} + SSE \end{aligned}$$

· Higher order interactions.

4-way interactions

Hierarchy.

Since k-way interactions are defined on (k-1)-way interactions, we cannot skip orders.

Eg: when no AB interactions then no higher order interactions that involve both A and B like ABD or ABCD

Eg: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$ is hierarchical

$y_{ijkl} = \mu + \alpha_i + \gamma_k + (\alpha\beta)_{ij} + \varepsilon_{ijkl}$ is not

y_{ijkl} include a term $(\alpha\beta\gamma)_{ijk}$, must include $\alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}$

- Unless there is a specific reason, we stick to hierarchical models.

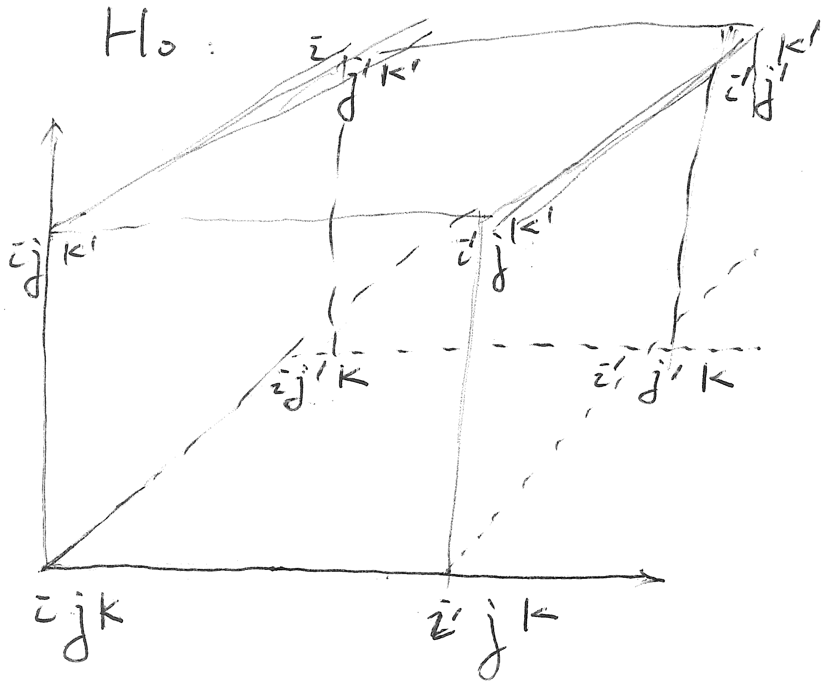
- advantage of the factorial design.
  - It can test the effects of ~~a~~ several factors at one time more efficiently.
  - interaction

- disadvantage.
  - If there are many factors or many levels, the size of model grows rapidly.
    (fractional factorial design remedy)

$$y_{ijkl} = \mu_{ijk} + \varepsilon_{ijkl}, \quad i=1,2, \quad j=1,2, \quad k=1,2$$

$\exists$ a higher order interaction.

H_0 :



$$H_0: \mu_{ijk} - \mu_{i'jk} - \mu_{ijk'} + \mu_{i'jk'} - \mu_{ij'k} + \mu_{i'j'k} + \mu_{ij'k'} - \mu_{i'j'k'} = 0$$

Chapter 23 Unequal sample sizes (n_{ij}) in a 2-way ANOVA.

Reasons: costs, lack of control

problems: SSTR doesn't split nicely
(in balanced case: SSTR = SSA + SSB + SSAB)

LS equations aren't as simple as equal case

① Use regression approach – always possible.

Test H_0: by $F = \dfrac{SS\,\text{extra} / df\,\text{extra}}{MSE}$

Regression approach

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

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

Development of Regression model

A: 3 levels, B: 3 levels.

$$X_{ijk1} = \begin{cases} 1 & \text{if level 1 for } A \\ -1 & \text{if level 3 for } A \\ 0 & \text{o.w.} \end{cases}$$

$$X_{ijk2} = \begin{cases} 1 & \text{if level 2 for } A \\ -1 & \text{if level 3 for } A \\ 0 & \text{o.w.} \end{cases}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 0$$

Similarly for X_{ijk3}, X_{ijk4} .. for B

$$\begin{aligned} Y_{ijk} = \mu_{..} &+ \alpha_1 X_{ijk1} + \alpha_2 X_{ijk2} + \beta_1 X_{ijk3} + \beta_2 X_{ijk4} \\ &+ (\alpha\beta)_{11} X_{ijk1} X_{ijk3} + (\alpha\beta)_{12} X_{ijk1} X_{ijk4} \\ &+ (\alpha\beta)_{21} X_{ijk2} X_{ijk3} + (\alpha\beta)_{22} X_{ijk2} X_{ijk4} + \varepsilon_{ijk} \end{aligned}$$

A	B	Freq	X_1	X_2	X_3	X_4	X_1X_3	X_1X_4	X_2X_3	X_2X_4
1	1	4	1	0	1	0	1	0	0	0
1	2	2	1	0	0	1	0	1	0	0
⋮			⋮	⋮						
3	2	3	-1	-1	0	1	0	-1	0	-1
3	3	2	-1	-1	-1	-1	1	1	1	1

- Inference about factor effects.

$$\hat{\mu}_{i\cdot} = \frac{1}{b}\sum_j \bar{y}_{ij\cdot}, \quad S^2(\hat{\mu}_{i\cdot}) = \frac{MSE}{b^2}\sum_j \frac{1}{n_{ij}}$$

$$L = \sum c_i \mu_i, \quad \hat{L} = \sum c_i \hat{\mu}_{i\cdot}$$

$$S^2(\hat{L}) = \frac{MSE}{b^2}\sum_i c_i^2 \sum_j \frac{1}{n_{ij}}$$

Tukey,

Scheffé : $\hat{L} \pm S(\hat{L})\sqrt{(a-1)F(1-\alpha, a-1, n_T - ab)}$

- Main differences against analysis of balanced design ($n_{ij} \equiv n$)

① Least square estimates for μ_{ij} are

$$\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$$

but $\hat{\mu}_{i\cdot} \neq \bar{y}_{i\cdot\cdot}$ ($\bar{y}_{i\cdot\cdot}$ are not unbiased for $\mu_{i\cdot}$)

$$\hat{\mu}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b}\hat{\mu}_{ij} = \frac{1}{b}\sum_{j=1}^{b}\bar{y}_{ij\cdot} = \frac{1}{b}\sum_j \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} y_{ijk}$$

$$\hat{\mu}_{\cdot j} = \frac{1}{a}\sum_i \frac{1}{n_{ij}}\sum_k y_{ijk}$$

$$\hat{\sigma}^2 = \frac{1}{N_T - ab} \sum_{ijk} (y_{ijk} - \bar{y}_{ij\cdot})^2$$

② . $\{\hat{\mu}_{i\cdot} - \mu_{i\cdot}\}_{i=1,\dots a}$ are not identically distributed.

Equal $n_{ij} (\equiv n)$, $\hat{\mu}_{i\cdot} \sim N(\mu_{i\cdot}, \frac{\sigma^2}{bn})$

Unequal. $\hat{\mu}_{i\cdot} \sim N(\mu_{i\cdot}, \underbrace{\frac{1}{b^2} \sum_j \frac{1}{n_{ij}} \sigma^2}_{\text{depends on } i})$

$\{\hat{\mu}_{ij\cdot} - \mu_{ij}\}$ are not identically distributed

$$\hat{\mu}_{ij} \sim N(\mu_{ij}, \frac{\sigma^2}{n_{ij}})$$

③ Equal: $SS_{TR} = SS_A + SS_B + SSAB$

Unequal: $\neq$

④ . t-test for a contrast $H_0 : \sum c_i \mu_i = 0$

equal: $t = \frac{\sum c_i \hat{\mu}_{i\cdot}}{\frac{s}{\sqrt{bn}} \sqrt{\sum_i c_i^2}}$ unequal: $t = \frac{\sum c_i \hat{\mu}_{i\cdot}}{\frac{s}{\sqrt{b^2}} \sqrt{\sum_i c_i^2 \sum_j \frac{1}{n_{ij}}}}$

Type I ANOVA table

Source	df	sequentials
A	$a-1$	$SS(A \mid 1)$
B	$b-1$	$SS(B \mid 1, A)$
AB	$(a-1)(b-1)$	$SS(AB \mid 1, A, B)$

SS (a term | some other terms)

= SSE (some other terms) − SSE (the term, some other terms)

order: A, B, AB

"sequential" sum of squares

. For balanced data,

$$SS(A \mid 1, B) = SS(A \mid 1)$$

R: anova (lm (y ~ A * B))

⟶ type I by default

Type II

SS_A : take the biggest hierarchical model without effect A, and compare it to the same model with A added.

Type II SS for AB (A, B, C)

$$SS(AB \mid 1, A, B, C, AC, BC)$$

$$= SSE(1, A, B, C, AC, BC) - SSE(1, A, B, C, AC, BC, AB$$

$$SS(C \mid 1, A, B, AB)$$

$$= SSE(1, A, B, AB) - SSE(1, A, B, AB, C)$$

$$SS(C \mid 1, A, AB)$$

Source

A	$(a-1)$	$SS(A \mid 1, B) = SSE(1, B) - SSE(1, B, A)$
B	$(b-1)$	$SS(B \mid 1, A) = SSE(1, A) - SSE(1, A, B)$
AB	$(a-1)(b-1)$	$SS(AB \mid 1, A, B) = SSE(1, A, B) - SSE(1, A, B, AB$

Type III:

source		
A	$(a-1)$	$SS(A \mid 1, B, AB) = SSE(1, B, AB) - SSE(1, A, B, AB$
B	$(b-1)$	$SS(B \mid 1, A, AB) = SSE(1, A, AB) - SSE(1, A, B, AB)$
AB	$(a-1)(b-1)$	$SS(AB \mid 1, A, B) = SSE(1, A, B) - SSE(1, A, B, AB$

- If interactions are not significant, type II gives a more powerful test.
- When balanced, I, II, III give the same result.

Chapter 25

Random models and mixed models

Def: Random factor: its levels consists of a random sample of levels from a population of all levels.

Fixed factor: its levels are selected by a non-random process.

Def: Fixed model: all factors fixed.

Random model: all factors random.

Mixed model: some fixed, some random.

· Random models (ANOVA model II)

1-way $\quad y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \begin{matrix} \varepsilon_{ij} \sim N(0, \sigma^2) \\ \tau_i \sim N(0, \sigma_\mu^2) \end{matrix} \Big\}$ indep

$i=1, \dots a, \quad j=1, \dots n_i$

· $E\, y_{ij} = \mu$

· $Var(y_{ij}) = Var(\tau_i) + Var(\varepsilon_{ij}) = \sigma_\mu^2 + \sigma^2$

$\{y_{ij}\} \longrightarrow$ identical distr, but not iid.

$$\mathrm{Cov}(y_{ij}, y_{ij'}) = \mathrm{Cov}(\tau_i + \varepsilon_{ij}, \tau_i + \varepsilon_{ij'})$$

$$= \mathrm{Cov}(\tau_i, \tau_i) + \mathrm{Cov}(\tau_i, \varepsilon_{ij'}) + \mathrm{Cov}(\varepsilon_{ij}, \tau_i) + \mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{ij'})$$

$$= \sigma_\mu^2$$

- $H_0: \sigma_\mu^2 = 0$. Use F-test for H_0.

$n_i = n$

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 = \sum_i \sum_j (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2$$

$$= \sum_i (n-1) S_{\varepsilon,i}^2$$

$$E(SSE) = a(n-1)\sigma^2$$

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

$$SSTR = \sum_{i=1}^{a} n (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\bar{y}_{i.} = \frac{1}{n} \sum_j y_{ij} = \frac{1}{n} \sum_j (\mu + \tau_i + \varepsilon_{ij}) = \mu + \tau_i + \bar{\varepsilon}_{i.}$$

$$\bar{y}_{i.} \sim \mathrm{Normal}\left(\mu, \sigma_\mu^2 + \frac{\sigma^2}{n}\right)$$

$$E\,SSTR = n E \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= n(a-1)\, E \frac{1}{(a-1)} \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= n(a-1)\left(\sigma_\mu^2 + \frac{\sigma^2}{n}\right)$$

$$= (a-1)(n\sigma_\mu^2 + \sigma^2)$$

$$E(MSTR) = n\sigma_\mu^2 + \sigma^2$$

Use $F = \dfrac{MSTR}{MSE}$ to test $H_0 : \sigma_\mu^2 = 0$

· Estimation of μ.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \cdots a$$

$$E y_{ij} = \mu \Rightarrow \hat{\mu} = \bar{y}_{..} \text{ is unbiased.}$$

$$Var(\bar{y}_{..}) = \frac{Var(\bar{y}_{i.})}{a} = \frac{1}{n^2 a} Var\left(\sum_{j=1}^{n} y_{ij}\right)$$

$$= \frac{1}{an^2}\left(\sum_{j=1}^{n} Var(y_{ij}) + \sum_{j \neq k}\sum Cov(y_{ij}, y_{ik})\right)$$

$$= \frac{1}{an^2}\left(n(\sigma^2 + \sigma_\mu^2) + n(n-1)\sigma_\mu^2\right)$$

$$= \frac{n\sigma^2 + n^2\sigma_\mu^2}{an^2} = \frac{\sigma^2 + n\sigma_\mu^2}{an}$$

$$\widehat{Var}(\bar{y}_{..}) = \frac{MSTR}{an}$$

$$\Rightarrow \quad \bar{y}_{..} \pm t_{\frac{\alpha}{2}, a-1} \sqrt{\frac{MSTR}{an}}$$

- Estimation of τ_i

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

$$Var(\hat{\tau}_i) = Var(\bar{y}_{i.}) + Var(\bar{y}_{..}) - 2\,cov(\bar{y}_{i.}, \bar{y}_{..})$$

$$cov(\bar{y}_{i.}, \bar{y}_{..})$$

$$= cov\left(\bar{y}_{i.}, \frac{1}{a}\sum_{i'=1}^{a} \bar{y}_{i'.}\right)$$

$$= \frac{1}{a} Var(\bar{y}_{i.})$$

$$Var(\hat{\tau}_i) = \frac{\sigma^2 + n\sigma_\mu^2}{n} + \frac{\sigma^2 + n\sigma_\mu^2}{an} - 2\,\frac{\sigma^2 + n\sigma_\mu^2}{an}$$

$$= \left(1 - \frac{1}{a}\right) \frac{\sigma^2 + n\sigma_\mu^2}{n}$$

Confidence interval is then trivial.

$SS_{TOT} = SS_{TR} + SS_E$

	df	SS	MS	E(MS)
Treat	$a-1$	SS_{TR}	MS_{TR}	$\sigma^2 + n\sigma_\mu^2$
error	$N_T - a$	SSE	MSE	σ^2
total	$N_T - 1$			

$\Rightarrow \hat{\sigma}^2 = MSE$, $\quad \hat{\sigma}_\mu^2 = \dfrac{MSTR - MSE}{n}$

MOM

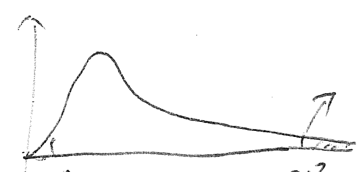
method of moment

$\hat{\sigma}_{\mu^2} = \dfrac{MSTR - MSE}{n} \vee 0$, biased

C.I for Variance components

- $\dfrac{MSE}{\sigma^2}(N_T - a) \sim \chi^2_{N_T-a}$

$$P\left(\chi^2_{N_T-a,.025} < \frac{SSE}{\sigma^2} < \chi^2_{N_T-a,.975}\right) = .95$$



95% C.I for σ^2 $\left[\dfrac{SSE}{\chi^2_{N_T-a,.975}}, \dfrac{SSE}{\chi^2_{N_T-a,.025}}\right]$

$$P\left(\frac{SSE}{\chi^2_{N_T-a,.975}} < \sigma^2 < \frac{SSE}{\chi^2_{N_T-a,.025}}\right) = .95$$

- No known exact C.I for σ_μ^2.

but exact C.I for $\psi = \dfrac{\sigma_\mu^2}{\sigma^2}$ and $\rho = \dfrac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$

$$Y_i = (Y_{i1}, \cdots Y_{in})^T$$

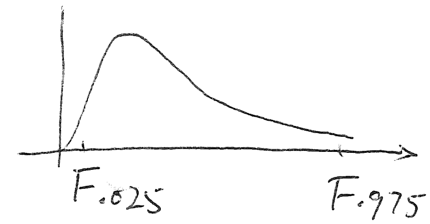
$$Var(Y_i) = \begin{pmatrix} \sigma_\mu^2 + \sigma^2 & & \sigma_\mu^2 \\ & \sigma_\mu^2 + \sigma^2 & \\ & & \ddots & \\ \sigma_\mu^2 & & & \sigma_\mu^2 + \sigma^2 \end{pmatrix}$$

$$= (\sigma^2 + \sigma_\mu^2) \begin{pmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{pmatrix}$$

$$\frac{(a-1)\, MSTR}{\sigma^2 + n\sigma_\mu^2} \sim \chi^2_{a-1} \perp\!\!\!\perp MSE$$

$$\frac{\dfrac{MSTR}{\sigma^2 + n\sigma_\mu^2}}{\dfrac{MSE}{\sigma^2}} = \frac{MSTR}{MSE} \cdot \frac{\sigma^2}{\sigma^2 + n\sigma_\mu^2} \sim F_{a-1,\, N_T - a}$$

95% CI for $\frac{\sigma^2 + n\sigma_\mu^2}{\sigma^2}$: $\left[\frac{MSTR/MSE}{F_{.975, a-1, N_T - a}} , \frac{MSTR/MSE}{F_{.025, a-1, N_T - a}} \right]$



$\psi = \frac{\sigma_\mu^2}{\sigma^2}$: $\left[\frac{1}{n}\left(\frac{MSTR/MSE}{F_{.975}} - 1 \right) , \frac{1}{n}\left(\frac{MSTR/MSE}{F_{.025}} - 1 \right) \right]$

$\rho = \frac{\sigma_\mu^2}{\sigma^2 + \sigma_\mu^2}$: $\left[\frac{L_\psi}{L_\psi + 1} , \frac{U_\psi}{U_\psi + 1} \right]$

Balanced case : algorithm for computing $E(MS)$.

$$y_{ijk} = \mu + \underbrace{\tau_i}_{\text{fixed}} + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad \begin{matrix} i = 1, \dots a \\ j = 1, \dots b \\ k = 1, \dots n \end{matrix}$$

random

$E(MS_\tau)$

$E(MS_\beta)$

$E(MS_\gamma)$

$E(MSE) = \sigma^2$

Main effect	F	R	R
	a	b	n
	i	j	k
τ_i	0	b	n
β_j	a	1	n
γ_{ij}	0	1	n

← prepare the table

1. put 0 if at least one row index matches column index and Fixed
2. put 1 if - and Random

in the remaining put # of levels.