앞에서 한건 고정효과

- 반복이 있는 2요인 변량효과 모형에서의 관심문제

조합의 반복이 있으면
조합에 대한 상호작용효과를 모델에 반영

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

- ○ $\alpha_i \sim$ iid $N(0,\sigma^2_\alpha)$, $\beta_j \sim$ iid $N(0,\sigma^2_\beta)$, $(\alpha\beta)_{ij} \sim$ iid $N(0,\sigma^2_{(\alpha\beta)})$ → 모두 rv
- ○ 상호작용이 있는가? ⇨ $\sigma^2_{(\alpha\beta)} > 0$
- ○ A 요인의 주효과가 있는가? ⇨ $\sigma^2_\alpha > 0$
- ○ B 요인의 주효과가 있는가? ⇨ $\sigma^2_\beta > 0$

→ 없는경우 0

- ○ 분산요소 σ^2_α, σ^2_β, $\sigma^2_{(\alpha\beta)}$, σ^2의 추정

↳ 모두 모수니까 추정 가능!
α_i, β_j, $(\alpha\beta)_{ij}$는 rv 이므로 추정의미 없음

- 반복이 있는 2요인 혼합효과 모형에서의 관심문제(A: 고정, B: 변량)

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

하나라도 변량이면 상호작용도 변량

  - $\alpha_i = \mu(A_i) - \mu = \mu_{i.} - \mu$ / $\beta_j \sim$ iid $N(0, \sigma^2_\beta)$ / $(\alpha\beta)_{ij} \sim$ iid $N(0, \sigma^2_{(\alpha\beta)})$
  - 상호작용이 있는가? ⇨ $\sigma^2_{(\alpha\beta)} > 0$
  - A 요인의 주효과가 있는가? ⇨ 하나 이상의 α_i가 0이 아니다.

    $H_0: \alpha_1 = \cdots = \alpha_a = 0$
    $H_1:$ not H_0
  - B 요인의 주효과가 있는가? $\sigma^2_\beta > 0$
  - 분산요소 $\sigma^2_\beta, \sigma^2_{(\alpha\beta)}, \sigma^2$의 추정 +) $\varepsilon_{ijk} \sim N(0, \sigma^2)$
  - 고정수준 요인의 효과 추정과 비교

우리가 한것: unrestricted ver.
restricted ver. → $\Sigma(\alpha\beta)_{ij} = 0$

$F = \frac{MSB}{MS(AB)}$: SAS의 default구현

mixed에서 A, (AB)는 같고 B가 (A가 고정 B가 변량)
unrestricted → $\sigma^2 + na\sigma_\beta^2 + n\sigma_{(\alpha\beta)}^2$
restricted → $\sigma^2 + na\sigma_\beta^2$

$F = \frac{MSB}{MSE}$

- 분산분석표
  - 변량효과모형과 혼합효과모형의 제곱합, 자유도, 평균제곱은 고정효과모형의 경우와 같음
  - EMS는 고정수준의 경우와 다르고 이에 따라 검정통계량도 달라짐

    Expected Mean Square

① 상호작용 → $\sigma_{(\alpha\beta)}^2$가 고정이 아닌경우 (rv) → 변량 or 혼합효과모형

- $H_0 : \sigma^2_{(\alpha\beta)} = 0$ vs $H_1 : \sigma^2_{(\alpha\beta)} > 0$
- $F = MS(AB)/MSE \sim F_{(a-1)(b-1),ab(n-1)}$
- 상호작용의 강약
  - $Var[(\widehat{\alpha\beta})_{ij}] > \hat{\sigma}^2$ 이면 강한 것으로 보고 아니면 약한 것으로 봄

$\sigma_{(\alpha\beta)}^2$가 error에 의한 변동의 추정치보다 클경우 강하다
매우 intuitive함!
$(\alpha\beta)_{ij}$라는 rv의 변동성이 sampling의 것보다 클경우

② 주효과 검정

$H_0 : \sigma_\alpha^2 = 0$ vs $H_1 : \sigma_\alpha^2 > 0$ (변량효과모형) → β_j가 rv

$H_0 : \alpha_1 = \cdots = \alpha_a = 0$ vs H_1 : not H_0 (혼합모형)

↳ α가 고정인경우 고정효과모형처럼

변인	자유도	SS
A	a-1	$nb\sum(\overline{Y}_{i..} - \overline{Y}_{...})^2$
B	b-1	$na\sum(\overline{Y}_{.j.} - \overline{Y}_{...})^2$
(AB)	(a-1)(b-1)	$n\sum\sum(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2$
Error	ab(n-1)	$\sum\sum\sum(Y_{ijk} - \overline{Y}_{ij.})^2$

↳ 간편식까지 모두같고 F값만 다름!

변인	EMS		
	Fixed	Random	A고정, B변량인 Mixed
A MSA의 기댓값	$\sigma^2+\dfrac{nb}{a-1}\sum_i \alpha_i^2$ ①	$\sigma^2+n\sigma_{(\alpha\beta)}^2+nb\sigma_\alpha^2$ ③	$\sigma^2+n\sigma_{(\alpha\beta)}^2+\frac{nb}{a-1}\Sigma\alpha_i^2$
B MSB의 기댓값	똑같이! $\sigma^2+\dfrac{na}{b-1}\sum_j \beta_j^2$	똑같이 $\sigma^2+n\sigma_{(\alpha\beta)}^2+na\sigma_\beta^2$	$\sigma^2+na\sigma_\beta^2$
(AB) MS(AB)의기댓값	$\sigma^2+\dfrac{n}{(a-1)(b-1)}\sum_{ij}(\alpha\beta)_{ij}^2$ ②	$\sigma^2+n\sigma_{(\alpha\beta)}^2$ ④	$\sigma^2+n\sigma_{(\alpha\beta)}^2$
Error MSE의 기댓값	σ^2	σ^2	σ^2

고정효과모형에서

① $H_0: \alpha_1=\cdots=\alpha_a=0 \leftrightarrow \Sigma\alpha_i^2=0$

$F=\frac{MSA}{MSE}$ → H_0가 사실이라면 $\frac{nb}{a-1}\Sigma\alpha_i^2=0$

아니라면 MSA > MSE

② $H_0: \beta_1=\cdots=\beta_b \leftrightarrow \Sigma\beta_j^2=0$, 따라서 $F=\frac{MSB}{MSE}$

③ $H_0: (\alpha\beta)_{11}=\cdots=(\alpha\beta)_{ab}=0 \leftrightarrow \Sigma\Sigma(\alpha\beta)_{ij}^2=0$

⇒ 즉, 적절한 검정통계량이다

반복이 있는 2요인 고정효과모형

① $E(SSA) = E\left[nb\sum_i(\bar{Y}_{i..} - \bar{Y}_{...})^2\right]$

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

$\bar{Y}_{i..} = \dfrac{\sum_j\sum_k Y_{ijk}}{nb} = \mu + \alpha_i + 0 + 0 + \bar{\varepsilon}_{i..}$

$\bar{Y}_{...} = \dfrac{\sum\sum\sum Y_{ijk}}{abn} = \mu + 0 + 0 + 0 + \bar{\varepsilon}_{...}$

$= E\left[nb\sum_i\left((\mu+\alpha_i+\bar{\varepsilon}_{i..}) - (\mu+\bar{\varepsilon}_{...})\right)^2\right]$

$= E\left[nb\sum_i(\alpha_i + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2\right]$

$= E\left[nb\sum_i\alpha_i^2\right] + E\left[nb\sum_i(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2\right] \leftarrow \bar{\varepsilon}_{i..} \sim N(0, \frac{\sigma^2}{nb})$

$= nb\sum_i\alpha_i^2 + (a-1)\sigma^2$

$E(MSA) = E\left(\dfrac{SSA}{a-1}\right) = \sigma^2 + \dfrac{nb\sum\alpha_i^2}{a-1}$

② $E(SSAB) = E\left[n\sum_i\sum_j(\bar{Y}_{ij.} - \bar{Y}_{...})^2\right]$

이때 $\bar{Y}_{ij.} = \dfrac{\sum_k Y_{ijk}}{n} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \bar{\varepsilon}_{ij.}$

$= nE\left[\sum_i\sum_j\left((\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\bar{\varepsilon}_{ij.}) - (\mu+\bar{\varepsilon}_{...})\right)^2\right]$

$= nE(b\sum\alpha_i^2) + nE(a\sum\beta_j^2) + nE(\sum\sum(\alpha\beta)_{ij}^2) + E(n\sum\sum(\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{...})^2) \leftarrow \bar{\varepsilon}_{ij.} \sim N(0, \frac{\sigma^2}{n})$

$= nb\sum\alpha_i^2 + na\sum\beta_j^2 + n\sum\sum(\alpha\beta)_{ij}^2 + \sigma^2(ab-1)$

$E(SSA) = nb\sum\alpha_i^2 + (a-1)\sigma^2$

$E(SSB) = na\sum\beta_j^2 + (b-1)\sigma^2$

$E(SS(AB)) = E(SSAB) - E(SSA) - E(SSB) = n\sum\sum(\alpha\beta)_{ij}^2 + (a-1)(b-1)\sigma^2$

$E(MS(AB)) = E\left(\dfrac{SS(AB)}{(a-1)(b-1)}\right) = \sigma^2 + \dfrac{n\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$

반복이 있는 2요인 혼합효과모형

① $E(SSA) = E\left[nb\sum_i(\bar{Y}_{i..} - \bar{Y}_{...})^2\right]$

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

$\bar{Y}_{i..} = \dfrac{\sum_j\sum_k Y_{ijk}}{bn} = \mu + \alpha_i + \bar{\beta} + \overline{(\alpha\beta)}_{i.} + \bar{\varepsilon}_{i..}$

$\bar{Y}_{...} = \dfrac{\sum\sum\sum Y_{ijk}}{abn} = \mu + \bar{\alpha} + \bar{\beta} + \overline{(\alpha\beta)}_{..} + \bar{\varepsilon}_{...}$

$\bar{\varepsilon}_{i..} \sim N(0, \frac{\sigma^2}{nb})$

$\overline{\alpha\beta}_{i.} \sim N(0, \frac{\sigma_{(\alpha\beta)}^2}{b})$

$= E\left[nb\sum_i\left((\mu+\alpha_i+\bar{\beta}+\overline{(\alpha\beta)}_{i.}+\bar{\varepsilon}_{i..}) - (\mu+\bar{\alpha}+\bar{\beta}+\overline{(\alpha\beta)}_{..}+\bar{\varepsilon}_{...})\right)^2\right]$

$= E\left[nb\sum_i(\alpha_i - \bar{\alpha} + \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{..} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2\right]$

$= E\left[nb\sum_i(\alpha_i - \bar{\alpha})^2\right] + E\left[nb\sum(\overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{..})^2\right]$

$+ E\left[nb\sum_i(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2\right]$

$= nb\sigma_\alpha^2(a-1) + n(a-1)\sigma_{(\alpha\beta)}^2 + (a-1)\sigma^2$

$E(MSA) = E\left(\dfrac{SSA}{a-1}\right) = \sigma^2 + nb\sigma_\alpha^2 + n\sigma_{(\alpha\beta)}^2$

11/27 녹화

④ $E(SSAB) = E\left[n \sum_i \sum_j (\overline{Y_{ij\cdot}} - \overline{Y_{\cdots}})^2\right]$

$= E\left[n\sum\sum\left(\left(\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\overline{\varepsilon_{ij\cdot}}\right)-\left(\mu+\bar{\alpha}+\bar{\beta}+\overline{(\alpha\beta)_{\cdot\cdot}}+\overline{\varepsilon_{\cdots}}\right)\right)^2\right]$

$= E\left[n\sum\sum(\alpha_i-\bar{\alpha}+\beta_j-\bar{\beta}+(\alpha\beta)_{ij}-\overline{(\alpha\beta)_{\cdot\cdot}}+\overline{\varepsilon_{ij\cdot}}-\overline{\varepsilon_{\cdots}})^2\right]$

$= E\left[n\sum_i\sum_j(\alpha_i-\bar{\alpha})^2\right]+E\left[n\sum_i\sum_j(\beta_j-\bar{\beta})^2\right]+E\left[n\sum\sum((\alpha\beta)_{ij}-\overline{(\alpha\beta)_{\cdot\cdot}})^2\right]+E\left[n\sum\sum(\overline{\varepsilon_{ij\cdot}}-\overline{\varepsilon_{\cdots}})^2\right]$

$= nb(a-1)\sigma_\alpha^2 + na(b-1)\sigma_\beta^2 + n(ab-1)\sigma_{(\alpha\beta)}^2 + (ab-1)\sigma^2$

$E(SS(AB)) = E(SSAB) - E(SSA) - E(SSB)$

이 때 $\overline{Y_{ij\cdot}} = \dfrac{\sum_k Y_{ijk}}{n} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\overline{\varepsilon_{ij\cdot}}$

$= nb(a-1)\sigma_\alpha^2 + na(b-1)\sigma_\beta^2 + n(ab-1)\sigma_{(\alpha\beta)}^2 + (ab-1)\sigma^2$

$-\left(nb\sigma_\alpha^2(a-1) + n(a-1)\sigma_{(\alpha\beta)}^2 + (a-1)\sigma^2\right)$

$-\left(na\sigma_\beta^2(b-1) + n(b-1)\sigma_{(\alpha\beta)}^2 + (b-1)\sigma^2\right)$

$= (a-1)(b-1)\sigma^2 + n(a-1)(b-1)\sigma_{(\alpha\beta)}^2$

$E(MS(AB)) = E\left(\dfrac{SS(AB)}{(a-1)(b-1)}\right) = \sigma^2 + n\sigma_{(\alpha\beta)}^2$

왜 F통계량이 아래처럼 construct 되는지 알아야됨!
EMS로부터 유도하기
+) 고정효과모형에서는 $\frac{\square}{MSE}$ 형태였음을 기억!

가설	Random에서의 검정통계량
$H_0: \sigma_\alpha^2 = 0$ vs $H_1: \sigma_\alpha^2 > 0$	MSA / MS(AB)
$H_0: \sigma_\beta^2 = 0$ vs $H_1: \sigma_\beta^2 > 0$	MSB / MS(AB)
$H_0: \sigma_{(\alpha\beta)}^2 = 0$ vs $H_1: \sigma_{(\alpha\beta)}^2 > 0$	MS(AB) / MSE

+) Fixed에서

가설	검정통계량
$H_0: \alpha_1 = \cdots = \alpha_a = 0$, H_1: not H_0	MSA / MSE
$H_0: \beta_1 = \cdots = \beta_b = 0$, H_1: not H_0	MSB / MSE
$H_0: (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{ab} = 0$, H_1: not H_0	MS(AB) / MSE

+) A고정, B 변량인 Mixed에서

가설	검정통계량
$H_0: \alpha_1 = \cdots = \alpha_a = 0$, H_1: not H_0	MSA / MS(AB)
$H_0: \sigma_\beta^2 = 0$, $H_1: \sigma_\beta^2 > 0$	MSB / MSE
$H_0: \sigma_{(\alpha\beta)}^2 = 0$, $H_1: \sigma_{(\alpha\beta)}^2 > 0$	MS(AB) / MSE

← ?? 이해안됨...

◉ 비료(고정)와 농작물(랜덤)에 따른 산출량

비료	농작물	수확량					
A	1	22.1	24.1	19.1	22.1	25.1	18.1
	2	27.1	15.1	20.6	28.6	15.1	24.6
	3	22.3	25.8	22.8	28.3	21.3	18.3
	4	19.8	28.3	26.8	27.3	26.8	26.8
	5	20	17	24	22.5	28	22.5
B	1	13.5	14.5	11.5	6	27	18
	2	16.9	17.4	10.4	19.4	11.9	15.4
	3	15.7	10.2	16.7	19.7	18.2	12.2
	4	15.1	6.5	17.1	7.6	13.6	21.1
	5	21.8	22.8	18.8	21.3	16.3	14.3
C	1	19	22	20	14.5	19	16
	2	20	22	25.5	16.5	18	17.5
	3	16.4	14.4	21.4	19.9	10.4	21.4
	4	24.5	16	11	7.5	14.5	15.5
	5	11.8	14.3	21.3	6.3	7.8	13.8

```
data<- scan(what=list("","","",1))
A 1 22.1 A 1 24.1 A 1 19.1 A 1 22.1 A 1 25.1 A 1 18.1
A 2 27.1 A 2 15.1 A 2 20.6 A 2 28.6 A 2 15.1 A 2 24.6
A 3 22.3 A 3 25.8 A 3 22.8 A 3 28.3 A 3 21.3 A 3 18.3
A 4 19.8 A 4 28.3 A 4 26.8 A 4 27.3 A 4 26.8 A 4 26.8
A 5 20 A 5 17 A 5 24 A 5 22.5 A 5 28 A 5 22.5
B 1 13.5 B 1 14.5 B 1 11.5 B 1 6 B 1 27 B 1 18
B 2 16.9 B 2 17.4 B 2 10.4 B 2 19.4 B 2 11.9 B 2 15.4
B 3 15.7 B 3 10.2 B 3 16.7 B 3 19.7 B 3 18.2 B 3 12.2
B 4 15.1 B 4 6.5 B 4 17.1 B 4 7.6 B 4 13.6 B 4 21.1
B 5 21.8 B 5 22.8 B 5 18.8 B 5 21.3 B 5 16.3 B 5 14.3
C 1 19 C 1 22 C 1 20 C 1 14.5 C 1 19 C 1 16
C 2 20 C 2 22 C 2 25.5 C 2 16.5 C 2 18 C 2 17.5
C 3 16.4 C 3 14.4 C 3 21.4 C 3 19.9 C 3 10.4 C 3 21.4
C 4 24.5 C 4 16 C 4 11 C 4 7.5 C 4 14.5 C 4 15.5
C 5 11.8 C 5 14.3 C 5 21.3 C 5 6.3 C 5 7.8 C 5 13.8
names(data) <- c("fertil", "variety", "yield")
df <- data.frame(data)
```

random component가 뭔지

```
library(sasLM)
RanTest(yield~fertil+variety+fertil*variety, df, Random="variety")
```

요인	df	SS	MS	F	p
Fertil	2	953.16	476.58	10.18	0.00633
variety	4	11.38	2.845	0.0608	0.9918
Fertil * variety	8	374.49	46.811	2.3822	0.02409
Error	25	1473.77	19.650		
total	89	2812.8			

→ unrestricted ver. 구현