■ 변량효과(random effect)모형과 혼합효과(mixed effect)모형

○ 고정효과 모형(fixed effect model) → 여태배운모형

- ◦ 요인의 수준이 실험자의 의도에 의해 조정 또는 결정되는 경우
- ◦ 처리효과에 대한 가설검정의 결과는 분석에서 **고려된 요인의 수준에 대해서만 적용**할 수 있음

○ 변량효과 모형(random effect model)

- ◦ 고려할 수 있는 요인의 수준에서 random하게 선택된 경우
- ◦ 처리효과는 모수가 아니고 **확률변수로 취급**
- ◦ 처리효과의 분산에 대해 $\sigma_\tau^2 = 0$인지를 검정함
- ◦ 선택된 수준뿐만 아니라 **고려했던 모든 요인의 수준에 대해서도 결과를 확장**시킬 수 있음

ex) 온도100, 120, 140°C 에서 처리효과가 없다는결론
→ 160, 180°C 에서도 그러하다고 확장

일원배치) $H_0: \tau_1 = \cdots = \tau_i = 0$
이원배치) $H_0: \alpha_1 = \cdots = \alpha_a = 0$
→ fixed값이므로 가능
확률변수가 0이냐아니냐는 의미없음! → 변수의 Var로검정

○ **모형의 분류**

- 모든 요인이 고정수준을 가지면 고정효과모형
- 모든 요인이 변량수준을 가지면 변량효과모형
- 일부 요인은 고정수준, 나머지는 변량수준을 가지면 혼합효과모형

□ 1 요인 변량효과 모형

○ 두 종류의 모집단

① 수준들의 모집단 Ω : 비교대상인 많은 수준들의 집합

◉ 초등학생들의 학업성취도 비교 ⇨ 전국에 있는 모든 초등학교

○ 표준상황: Ω에서 수준평균은 $N(\mu, \sigma_\mu^2)$ 분포를 따름

- 모든 학교의 평균 성적의 분포는 $N(\mu, \sigma_\mu^2)$

○ μ: 수준평균들의 평균(상수) ⇨ 모든 학교의 전체평균

○ σ_μ^2: 수준평균들의 분산

② 각 수준별 관측단위들의 모집단

○ 표준상황: 각 수준에서 반응변수는 정규분포를 따름

○ 그 평균은 다를 수 있으나 ⇦ 학교마다 평균을 다를 수 있음

○ 분산은 모든 수준에서 σ^2

○ 두 단계 추출

① 수준 추출 → 전국초등학교중 p개 추출

- ◦ 수준들의 모집단 Ω 에서 p개의 수준을 무작위 추출
- ◦ μ_i: i번째 추출되는 수준들의 평균반응

$$\Rightarrow \mu_i \sim \text{iid } N(\mu, \sigma_\mu^2)$$

② 관측단위 추출 → i번째 초등학교에서 n_i명의 성적뽑기

- ◦ 추출되는 각 수준에서 n_i개의 관측단위를 무작위로 추출(무작위 배정)
- ◦ $N = \sum_{i=1}^{p} n_i$

- i번째 수준이 **추출되었다**고 할 때 그 평균반응이 μ_i^*라고 하면, 그 수준에 대한 관측단위의 모집단에서 반응변수는 $N(\mu_i^*, \sigma^2)$를 따름 ← 등분산가정

≠ 고정효과

○ 1 요인 변량효과 분산분석 모형

ex) i번째 학교에서 j번째 학생의 성적

- Y_{ij}: i 번째 추출 수준(학교)에서 j 번째 추출하는 관측단위(학생)의 반응변수

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 1,\ ..., p,\ \ j = 1, ..., n_i$$

  ○ $\mu_i \sim$ iid $N(\mu, \sigma_\mu^2)$: i 번째 추출수준의 수준평균 → μ_i는 확률변수로 본다

  ○ $\varepsilon_{ij} \sim$ iid $N(0, \sigma^2)$ → 고정효과와 동일

  ○ $\tau_i = \mu_i - \mu$라고 하면

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = 1,\ ..., p,\ \ j = 1, ..., n_i$$

  - $\tau_i \sim$ iid $N(0, \sigma_\mu^2)$ → 고정효과모형에서는 고정된 값이었음!
  - 모든 τ_i와 ε_{ij}는 서로 독립

$$\begin{cases} E(\tau_i) = E(\mu_i - \mu) = E(\mu_i) - E(\mu) = \mu - \mu = 0 \\ Var(\tau_i) = Var(\mu_i - \mu) = Var(\mu_i) = \sigma_\mu^2 \end{cases}$$

* 고정효과모형에서는 $\varepsilon_{ij} \sim N(0, \sigma^2)$

$E(\mu_i + \varepsilon_{ij}) = \mu_i + 0 = \mu_i$

$Var(\mu_i + \varepsilon_{ij}) = Var(\varepsilon_{ij}) = \sigma^2$ ($\because$ μ_i가 상수)

⇒ 고정효과모형의 var이 변동효과모형의 var보다 작다

- 모형의 특징
  - $E(Y_{ij}) = E(\mu_i + \varepsilon_{ij}) = \mu$ → $E(\mu_i + \varepsilon_{ij}) = \mu + 0 = \mu$
  - $Var(Y_{ij}) = Var(\mu_i + \varepsilon_{ij}) = \sigma_\mu^2 + \sigma^2$ → $Var(\mu_i + \varepsilon_{ij}) = Var(\mu_i) + Var(\varepsilon_{ij}) = \sigma_\mu^2 + \sigma^2 + 0$ ← 독립

    ↳ 어떤 학교가 추출될지 모르는 것에 대한 변동
  - $Y_{ij} \sim N(\mu, \sigma_\mu^2 + \sigma^2)$
  - $Cov(Y_{ij}, Y_{ik}) = Cov(\mu_i + \varepsilon_{ij}, \mu_i + \varepsilon_{ik}) = Var(\mu_i) = \sigma_\mu^2$ = $Var(\mu_i) + 0 + 0 + 0$

    ⇨ 추출된 한 학교에서 추출된 두 학생의 성적의 공분산은 $\sigma_\mu^2 (\geq 0)$ → 양의 상관관계

    ⇒ 한 학생의 성적이 높으면 다른 학생의 성적이 높은 경향이 있음

    ↳ 같은학교학생이면 성적이 비슷한 경향
  - 분산 σ_μ^2와 σ^2를 variance components라고 함

    ⇨ components of variance 또는 random effects 모형이라고 부름

○ 관심문제

- ○ 특정 수준의 평균에는 관심이 없음
- ○ 수준에 따라 평균반응이 다른가? → 변동이 있다 = 수준마다 다른값을 갖는다

  ⇒ μ_i는 확률변수로 $\sigma_\mu^2 > 0$라는 것은 μ_i가 다른 값을 가질 수 있음을 의미

- ○ 가설: $H_0 : \sigma_\mu^2 = 0$ vs $H_1 : \sigma_\mu^2 > 0$

- σ_μ^2의 추정
- μ의 추정 → μ_i는 확률변수 (고정된 모수X) 이므로 추정하지 않음. → $\mu_i \sim iid\ N(\mu, \sigma_\mu^2)$ 이므로
- σ^2의 추정 → $\epsilon_{ij} \sim iid\ N(0, \sigma^2)$ 이므로
- $\dfrac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$의 추정

$$Var(Y_{ij}) = \sigma_\mu^2 + \sigma^2$$

어느 학교인가? ↲ ↳ 그 학교의 어느 학생인가?

$$Cov(Y_{ij}, Y_{ik}) = \sigma_\mu^2$$

$\sigma_\mu^2 + \sigma^2$	전체 분산	한 수준을 추출하고 그 수준에서 한 관측단위를 추출해서 반응변수 값을 얻을 때,
σ_μ^2	학교 간의 차이의 분산	수준평균의 분산이 차지하는 비율

⇓

$$\rho = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij})}\sqrt{Var(Y_{ik})}} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$$

○ ρ가 크다는 것은

① 전체분산 중 수준평균의 분산이 차지하는 비율이 높음

② 한 수준의 두 관측값의 상관관계가 높음

⇒ 즉 학교별 차이의 의미가 크며 학교내의 성적 차이는 거의 없다

상위권 학교학생은 대체로 공부를 잘한다

학교출신이 중요!

○ 분산분석

고정효과모형과의 차이점

- 가정: $\varepsilon_{ij} \overset{iid}{\sim} N(0,\sigma^2)$, $\tau_i \overset{iid}{\sim} N(0,\sigma_\tau^2)$, τ_i와 ε_{ij}는 독립
- $H_0\ :\ \sigma_\tau^2 = 0$ VS $H_1\ :\ \sigma_\tau^2 > 0$
  - $\sigma_\tau^2 = 0$이라면, 모든 처리는 동일

$\tau_i = 0$
항상 그 값만 취한다

  - $\sigma_\tau^2 > 0$이라면, 처리들 간에 변동이 있다는 것을 의미
- 제곱합 등식 $TSS = SSTR + SSE$는 계속 사용

변인	자유도	제곱합(SS)	평균제곱(MS)	F
처리(모형)	$p-1$	SSTR	MSTR	MSTR/MSE
오차	$N-p$	SSE	MSE	
전체	$N-1$	TSS		

○ $E(MSTR) = \sigma^2 + n'\sigma_\mu^2$, $\quad n' = (N - \sum n_i^2/N)/(p-1)$

- 모든 $n_i = n$ 이면 $n' = n$
- $n'\sigma_\mu^2 = E(MSTR) - E(MSE)$
- $\sigma_\mu^2 = \dfrac{E(MSTR) - E(MSE)}{n'} \xrightarrow{\text{추정}} \hat{\sigma}_\mu^2 = \dfrac{MSTR - MSE}{n'}$
- $MSTR < MSE$이면 $\hat{\sigma}_\mu^2 < 0 \Rightarrow \hat{\sigma}_\mu^2 = 0$으로 고쳐 사용

음수가 될수 없으므로

$$E(SSTR) = E\left(\sum n_i (\bar{y}_{i.} - \bar{y}_{..})^2\right)$$
$$= E\left(\sum_i n_i ((\mu + \tau_i + \bar{e}_{i.}) - (\mu + \bar{e}_{..}))^2\right) \quad \because y_{ij} = \mu + \tau_i + e_{ij}$$
$$= E\left(\sum n_i (\tau_i + (\bar{e}_{i.} - \bar{e}_{..}))^2\right)$$
$$= \sum n_i \tau_i^2 + E\left(\sum n_i (\bar{e}_{i.} - \bar{e}_{..})^2\right) + E\left(2\sum n_i \tau_i (\bar{e}_{i.} - \bar{e}_{..})\right)$$
$$= \sum n_i \tau_i^2 + (p-1)\sigma^2$$
$$E(MSTR) = E\left(\frac{SSTR}{p-1}\right) = \sigma^2 + \frac{\sum n_i \tau_i^2}{p-1}$$

$$n' = \frac{N - \sum \frac{n_i^2}{N}}{p-1}$$

$$n_i = n \rightarrow n' = \frac{N - \frac{n^2 p}{N}}{p-1} = \frac{\frac{(np)^2 - n^2 p}{np}}{p-1} = \frac{\frac{np(p-1)n}{np}}{p-1} = n$$

↳ 고정효과모형에서는 $F = \dfrac{MSTR}{MSE}$

변량효과모형에서는?

↑ 사진판서확인 11/13

11/6

$$E(SSTR) = E\left[\Sigma n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right]$$

$$= E\left[\Sigma n_i\left((\mu + \tau_i + \bar{e}_{i\cdot}) - (\mu + \bar{\tau} + \bar{e}_{\cdot\cdot})\right)^2\right]$$

$$= E\left[\Sigma n_i\left((\tau_i - \bar{\tau}) + (\bar{e}_{i\cdot} - \bar{e}_{\cdot\cdot})\right)^2\right]$$

$$= E\left[\Sigma n_i(\tau_i - \bar{\tau})^2\right] + E\left[\Sigma n_i(\bar{e}_{i\cdot} - \bar{e}_{\cdot\cdot})^2\right] + E\left[\Sigma 2n_i(\tau_i - \bar{\tau})(\bar{e}_{i\cdot} - \bar{e}_{\cdot\cdot})\right]$$

$$= E\left[\Sigma n_i(\tau_i^2 - 2\tau_i\bar{\tau} + \bar{\tau}^2)\right] + (p-1)\sigma^2$$

$$= E\left[\Sigma n_i\tau_i^2 - 2\bar{\tau}N\bar{\tau} + \bar{\tau}^2 N\right] + (p-1)\sigma^2$$

이때 $Y_{ij} = \mu + \tau_i + e_{ij}$

$$= E\left[\Sigma n_i\tau_i^2 - N\bar{\tau}^2\right] + (p-1)\sigma^2$$

$$\bar{Y}_{i\cdot} = \frac{\sum_j Y_{ij}}{n_i} = \frac{1}{n_i}\sum_j(\mu + \tau_i + e_{ij}) = \mu + \tau_i + \bar{e}_{i\cdot}$$

$$= N\sigma_\mu^2 - NE(\bar{\tau}^2) + (p-1)\sigma^2$$

$$E\left[\frac{\Sigma n_i(\bar{e}_{i\cdot} - \bar{e}_{\cdot\cdot})^2}{p-1}\right] = \sigma^2$$

$$= N\sigma_\mu^2 - NE\left(\frac{\Sigma n_i\tau_i}{N}\right)^2 + (p-1)\sigma^2$$

$e_{ij} \sim N(0, \sigma^2)$, $\tau_i \sim N(0, \sigma_\mu^2)$

$$= N\sigma_\mu^2 - \frac{\sigma_\mu^2 \Sigma n_i^2}{N} + (p-1)\sigma^2$$

$$\bar{\tau} = \frac{\Sigma n_i \tau_i}{N}$$

$$= \left(\frac{N^2 - \Sigma n_i^2}{N}\right)\sigma_\mu^2 + (p-1)\sigma^2$$

$E(\tau_i^2) =$

$$E(MSTR) = E\left(\frac{SSTR}{p-1}\right) = \sigma^2 + n'\sigma_\mu^2$$

이때 $n' = \dfrac{\left(N - \frac{\Sigma n_i^2}{N}\right)}{p-1}$

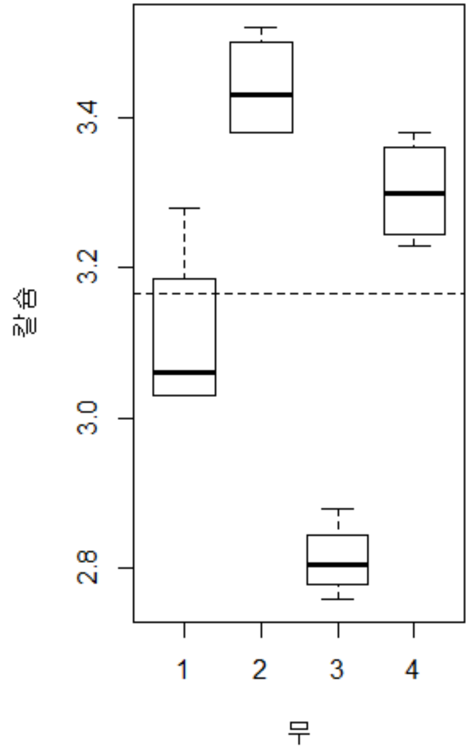
↗ 변량효과모형

◉ 무 4개를 무작위로 추출하여 잎에 포함된 칼슘함량을 4회씩 측정

→ 무에 따른 칼슘함량 차이가 있는지?

각 수준별 boxplot

종류	1	2	3	4	
	3.28	3.52	2.88	3.34	
	3.09	3.48	2.80	3.38	
	3.03	3.38	2.81	3.23	
	3.03	3.38	2.76	3.26	
합	12.43	13.76	11.25	13.21	50.65
평균	3.11	3.44	2.81	3.30	3.17



↓

무에 따라 칼슘함량의 차이가 있음을 graphically 확인 가능

random-effect 모델 사용

∵ 무 4개를 무작위추출했으므로 변량효과모형 사용

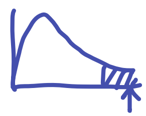
〈검정〉

↳ 무에따른 칼슘함량의 차이가 없다

- $H_0 \ : \ \sigma_\tau^2 = 0$ VS $H_1 \ : \ \sigma_\tau^2 > 0$ ($\sigma_\tau^2 \neq 0$)
  - $CT = 50.65^2/16 = 160.3389$
  - $TSS = 3.28^2 + 3.52^2 + \cdots + 3.26^2 - CT = 0.9676$
  - $SSTR = \dfrac{1}{4}(12.43^2 + \cdots + 13.21^2) - CT = 0.8884$
  - $SSE = TSS - SSTR = 0.9676 - 0.8884 = 0.0792$
  - 분산분석표

변인	자유도	제곱합(SS)	평균제곱(MS)	F
처리(모형)	p-1 3	0.8884	0.2961	44.9
오차	N-p 12	0.0792	0.0066	
전체	N-1 15	0.9676		

- $F_{0.01,3,12} = 5.95$ < 44.9 ⇨ 1% 유의수준에서 H_0 기각



- 무에 따라 칼슘함량에 차이가 있음

〈추정〉

- μ의 추정
  - $\mu = E(\mu_i)$
    - $\hat{\mu}_i = \overline{Y}_{i.}$
    - $\hat{\mu} = \overline{Y}_{i.}$ 의 평균 ⇨ 균형 자료이므로 $\hat{\mu} = \overline{Y}_{..}$ ⇦ 3.17

      p개의 처리가 있을때 $\frac{1}{p}\Sigma \overline{Y}_{i.} = \frac{1}{p}\frac{1}{n}\Sigma Y_{ij} = \overline{Y}_{..}$

- σ^2와 σ^2_μ의 추정
  - $\hat{\sigma}^2 = MSE$ ⇦ 0.0066
  - $\hat{\sigma}^2_\mu = \dfrac{MSTR - MSE}{n'}$ ⇦ $\dfrac{0.2961 - 0.0066}{4} = 0.0724$

    $n_i = n$ 이면 $n' = n$

    슬라이드 p10

슬라이드 p8 참고

- $\dfrac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2}$의 추정 ⇦ $\dfrac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2_\mu + \hat{\sigma}^2} = \dfrac{0.0724}{0.0724 + 0.0066} = 0.916$

  ① 칼슘 함량의 정도가 대부분 무 간의 차이에 의해 발생 → 즉 같은 무면 칼슘함량의 정도가 비슷함 (어떤부위에서 추출하냐에 따른 변동은 있을수있음)

  ② 동일 무에서 추출된 칼슘의 함량의 상관계수는 0.916

  굉장히크다!

↓ 하나는 고정, 하나는 변량효과모형

□ 2 요인 변량효과 모형과 혼합효과 모형

- 반복이 없는 2요인 변량효과 모형

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

  - $\alpha_i \sim$ iid $N(0, \sigma_\alpha^2)$, $\beta_j \sim$ iid $N(0, \sigma_\beta^2)$, $\varepsilon_{ij} \sim$ iid $N(0, \sigma^2)$
  - α_i, β_j, ε_{ij} 서로독립
  - $Var(Y_{ij}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$
  - $Cov(Y_{ij}, Y_{ik}) = Cov(\alpha_i, \alpha_i) = \sigma_\alpha^2$, $Cov(Y_{ij}, Y_{kj}) = Cov(\beta_j, \beta_j) = \sigma_\beta^2$

$Var(Y_{ij}) = Var(\mu + \alpha_i + \beta_j + \varepsilon_{ij})$

$= Var(\mu) + Var(\alpha_i) + Var(\beta_j) + Var(\varepsilon_{ij}) + Cov$

$= 0 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$

(cf) 반복이 없는 2요인 고정효과모형의 경우

$Var(Y_{ij}) = Var(\mu + \alpha_i + \beta_j + \varepsilon_{ij}) = Var(\varepsilon_{ij}) = \sigma^2$

$Cov(Y_{ij}, Y_{ik}) = Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij},\ \mu + \alpha_i + \beta_k + \varepsilon_{ik})$

μ는 상수
서로 iid여서 0

$= Cov(\alpha_i, \alpha_i) = \sigma_\alpha^2$

$Cov(Y_{ij}, Y_{kj}) = Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij},\ \mu + \alpha_k + \beta_j + \varepsilon_{kj})$

$= Cov(\beta_j, \beta_j) = \sigma_\beta^2$

(cf) 반복이 없는 2요인 고정효과모형의 경우

$Cov(Y_{ij}, Y_{ik}) = Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij},\ \mu + \alpha_i + \beta_k + \varepsilon_{ik}) = 0$

$Cov(Y_{ij}, Y_{kj}) = Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij},\ \mu + \alpha_k + \beta_j + \varepsilon_{kj}) = 0$

- 반복이 없는 2요인 혼합효과 모형(A: 고정, B: 변량)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

  - $\alpha_i = \mu_{i.} - \mu$: 요인 A의 i번째 처리효과 ⇨ $H_0 : \alpha_1 = \cdots = \alpha_a = 0$ → 기존방식
  - $\beta_j \sim$ iid $N(0, \sigma_\beta^2)$, $\varepsilon_{ij} \sim$ iid $N(0, \sigma^2)$, β_j와 ε_{ij}는 서로독립
  - $E(Y_{ij}) = \mu + \alpha_i = \mu_{i.}$ $= E(\mu + \alpha_i + \beta_j + \varepsilon_{ij})$ → A가 변량, B가 고정인 경우 $E(Y_{ij}) = \mu + \beta_j$
    A, B 둘다 고정인 경우 $E(Y_{ij}) = \mu + \alpha_i + \beta_j$
  - $Var(Y_{ij}) = \sigma_\beta^2 + \sigma^2$
  - $Cov(Y_{ij}, Y_{ik}) = 0$, $Cov(Y_{ij}, Y_{kj}) = Cov(\beta_j, \beta_j) = \sigma_\beta^2$

$Var(Y_{ij}) = Var(\mu + \alpha_i + \beta_j + \varepsilon_{ij})$
$= Var(\beta_j + \varepsilon_{ij})$
$= Var(\beta_j) + Var(\varepsilon_{ij}) + Cov$
$= \sigma_\beta^2 + \sigma^2$

cf) 반복이 없는 2요인 고정효과모형의 경우 $Var(Y_{ij}) = \sigma^2$
반복이 없는 2요인 변량효과모형의 경우 $Var(Y_{ij}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$

$Cov(Y_{ij}, Y_{ik}) = Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij}, \mu + \alpha_i + \beta_k + \varepsilon_{ik}) = 0$
⇒ α의 처리가 같아도 β가 다르면 독립
$Cov(Y_{ij}, Y_{kj}) = Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij}, \mu + \alpha_k + \beta_j + \varepsilon_{kj}) = Cov(\beta_j, \beta_j) = \sigma_\beta^2$
⇒ β의 처리가 같으면 α가 달라도 연관성이 있음

Expected Mean Square

변인	EMS		
	고정 Fixed	변량 Random	혼합 Mixed
A	① $\sigma^2 + \frac{b}{a-1}\sum_i \alpha_i^2$	② $\sigma^2 + b\sigma_\alpha^2$	③ $\sigma^2 + \frac{b}{a-1}\sum_i \alpha_i^2$
B	$\sigma^2 + \frac{a}{b-1}\sum_j \beta_j^2$	$\sigma^2 + a\sigma_\beta^2$	$\sigma^2 + a\sigma_\beta^2$
Error	σ^2	σ^2	σ^2

→ $F = \frac{MSA}{MSE}$

→ $F = \frac{MSB}{MSE}$

→ MSE의 기댓값은 어느 모형이든 동일함
(σ^2에 대한 추정치로 MSE를 쓰는 이유)

11/8, 11/13

① 반복이 없는 2요인 고정효과모형

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

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

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

$= b\sum\alpha_i^2 + (a-1)\sigma^2$

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

이때 $Y_{ij} = \mu+\alpha_i+\beta_j+\varepsilon_{ij}$

$\bar{Y}_{i.} = \frac{\sum_j Y_{ij}}{b} = \frac{1}{b}\sum_j(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+\alpha_i+0+\bar{\varepsilon}_{i.}$

$\bar{Y}_{..} = \frac{\sum\sum Y_{ij}}{ab} = \frac{1}{ab}\sum\sum(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+0+0+\bar{\varepsilon}_{..}$

$E\left[\frac{b\sum_i(\bar{\varepsilon}_{i.}-\bar{\varepsilon}_{..})^2}{a-1}\right] = \sigma^2$

② 반복이 없는 2요인 변량효과모형

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

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

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

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

$= b(a-1)\sigma_\alpha^2 + (a-1)\sigma^2$

$\rightarrow E(MSA) = E\left(\frac{SSA}{a-1}\right) = b\sigma_\alpha^2 + \sigma^2$

이때 $Y_{ij} = \mu+\alpha_i+\beta_j+\varepsilon_{ij}$

$\bar{Y}_{i.} = \frac{\sum_j Y_{ij}}{b} = \frac{1}{b}\sum_j(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+\alpha_i+\bar{\beta}+\bar{\varepsilon}_{i.}$

$\bar{Y}_{..} = \frac{\sum\sum Y_{ij}}{ab} = \frac{1}{ab}\sum\sum(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+\bar{\alpha}+\bar{\beta}+\bar{\varepsilon}_{..}$

$E\left[\frac{\sum(\alpha_i-\bar{\alpha})^2}{a-1}\right] = \sigma_\alpha^2$

$E\left[\frac{b\sum_i(\bar{\varepsilon}_{i.}-\bar{\varepsilon}_{..})^2}{a-1}\right] = \sigma^2$

③ 반복이 없는 2요인 mixed (A가 고정, B가 변량인 경우) 모형

이때 $Y_{ij} = \mu+\alpha_i+\beta_j+\varepsilon_{ij}$

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

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

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

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

$= b\sum\alpha_i^2 + (a-1)\sigma^2$

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

$\bar{Y}_{i.} = \frac{\sum_j Y_{ij}}{b} = \frac{1}{b}\sum_j(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+\alpha_i+\bar{\beta}+\bar{\varepsilon}_{i.}$

$\bar{Y}_{..} = \frac{\sum\sum Y_{ij}}{ab} = \frac{1}{ab}\sum\sum(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+0+\bar{\beta}+\bar{\varepsilon}_{..}$

$E\left[\frac{b\sum_i(\bar{\varepsilon}_{i.}-\bar{\varepsilon}_{..})^2}{a-1}\right] = \sigma^2$

$E(SSB) = E\left[\sum_i\sum_j (\bar{Y}_{.j}-\bar{Y}_{..})^2\right] = E\left[a\sum_j (\bar{Y}_{.j}-\bar{Y}_{..})^2\right]$

$= E\left[a\sum_j\left((\mu+\beta_j+\bar{\varepsilon}_{.j})-(\mu+\bar{\beta}+\bar{\varepsilon}_{..})\right)^2\right]$

$= E\left[a\sum_j\left((\beta_j-\bar{\beta})+(\bar{\varepsilon}_{.j}-\bar{\varepsilon}_{..})\right)^2\right]$

$= E\left[a\sum_j(\beta_j-\bar{\beta})^2\right] + E\left[a\sum_j(\bar{\varepsilon}_{.j}-\bar{\varepsilon}_{..})^2\right]$

$= a(b-1)\sigma_\beta^2 + (b-1)\sigma^2$

$\rightarrow E(MSB) = E\left(\frac{SSB}{b-1}\right) = a\sigma_\beta^2 + \sigma^2$

$\bar{Y}_{.j} = \frac{\sum_i Y_{ij}}{a} = \frac{1}{a}\sum_i(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \mu+0+\beta_j+\bar{\varepsilon}_{.j}$

$E\left[\frac{\sum\sum(\bar{\varepsilon}_{.j}-\bar{\varepsilon}_{..})^2}{b-1}\right] = \sigma^2$

11/15

- 반복이 있는 2요인 변량효과 모형에서의 관심문제

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

  - $\alpha_i \sim$ iid $N(0, \sigma_\alpha^2)$, $\beta_j \sim$ iid $N(0, \sigma_\beta^2)$, $(\alpha\beta)_{ij} \sim$ iid $N(0, \sigma_{(\alpha\beta)}^2)$
  - 상호작용이 있는가? ⇨ $\sigma_{(\alpha\beta)}^2 > 0$
  - A 요인의 주효과가 있는가? ⇨ $\sigma_\alpha^2 > 0$
  - B 요인의 주효과가 있는가? ⇨ $\sigma_\beta^2 > 0$
  - 분산요소 $\sigma_\alpha^2, \sigma_\beta^2, \sigma_{(\alpha\beta)}^2, \sigma^2$의 추정

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

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

  - ○ $\alpha_i = \mu(A_i) - \mu = \mu_{i.} - \mu, \beta_j \sim$ 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_i 가 0이 아니다.
  - ○ B 요인의 주효과가 있는가? $\sigma^2_\beta > 0$
  - ○ 분산요소 $\sigma^2_\beta, \sigma^2_{(\alpha\beta)}, \sigma^2$의 추정
  - ○ 고정수준 요인의 효과 추정과 비교

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

① 상호작용

- ○ $H_0 : \sigma^2_{(\alpha\beta)} = 0$ 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

② 주효과 검정

- $H_0 : \sigma_\alpha^2 = 0$ vs $H_1 : \sigma_\alpha^2 > 0$ (변량효과모형)
- $H_0 : \alpha_1 = \cdots = \alpha_a = 0$ vs $H_1 :$ not H_0 (혼합모형)

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)	$\sum\sum\sum(Y_{ijk} - \overline{Y}_{ij.})^2$

변인	EMS	
	Fixed	Random
A	$\sigma^2 + \frac{nb}{a-1}\sum_i \alpha_i^2$	$\sigma^2 + n\sigma^2_{(\alpha\beta)} + nb\sigma^2_{\alpha}$
B	$\sigma^2 + \frac{na}{b-1}\sum_j \beta_j^2$	$\sigma^2 + n\sigma^2_{(\alpha\beta)} + na\sigma^2_{\beta}$
(AB)	$\sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{ij} (\alpha\beta)_{ij}^2$	$\sigma^2 + n\sigma^2_{(\alpha\beta)}$
Error	σ^2	σ^2

가설	검정통계량
$H_0 : \sigma_{\alpha}^2 = 0$ vs $H_1 : \sigma_{\alpha}^2 > 0$	
$H_0 : \sigma_{\beta}^2 = 0$ vs $H_1 : \sigma_{\beta}^2 > 0$	
$H_0 : \sigma_{(\alpha\beta)}^2 = 0$ vs $H_1 : \sigma_{(\alpha\beta)}^2 > 0$	