■ 반복측정 분산분석(Repeated Measures ANOVA)

- 한 실험개체에서 2회 이상 측정을 수행하는 것

- 반복측정의 장점
  - 어떤 실험에서는 충분한 수의 실험개체를 확보하기 어려울 때 적용
  - 반복측정에서는 실험개체 스스로가 대조(control)의 역할을 함

    ⇨ 실험개체가 블록이 됨

- 반복측정의 단점
  - 각 개체를 여러 번 실험처리하기 때문에 시간이 많이 소요
  - 이월효과(carry-over effect, 잔류효과)가 발생할 수 있음

    ↳ 앞의 처리 효과가 사라질 때까지 기다려야함
    기다려도 잔류효과 있을 수도!

□ 단일요인 반복측정 분산분석

● 반복측정 자료의 예 → 개체별로 처리 (일변량)

개체	처리 1	처리 2	처리 3	처리 4
1	30	28	16	34
2	14	18	10	22
3	24	20	18	30
4	38	34	20	44
5	26	28	14	30
평균	26.4	25.6	15.6	32.0

○ 일변량 분산분석

- 통계모형

$$Y_{ij} = \mu + \underset{\text{개체효과}}{\rho_i} + \underset{\text{처리효과}}{\tau_j} + \varepsilon_{ij}, \qquad \overset{\text{개체}}{i = 1,\ \ldots,\ n}, \quad \overset{\text{처리(반복)}}{j = 1,\ \ldots,\ p}$$

  ○ ρ_i: 개체 i의 효과(subject effect)

  ☆ 동일한 개체의 자료들 간 상관관계가 존재할 수 있음

  ⇨ Y_{ij}와 Y_{ik}는 독립이라고 보기 어려움

  ⇒ ρ_i는 랜덤효과(변량효과): $\rho_i \sim N(0, \sigma_s^2)$

  ↳ 고정효과로 둔다면 추정해야할 개수가 커짐!

11/27

*과제5 설명 11/29

- 변동분해

$\sum\sum(Y_{ij}-\overline{Y}_{..})=0$ $\sum(\overline{Y}_{i.}-\overline{Y}_{..})=0$ $\sum\sum(Y_{ij}-\overline{Y}_{i.})=0$, $\sum_j(Y_{ij}-\overline{Y}_{i.})=0$

자유도 np-1 자유도 n-1 자유도 np-n

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

TSS SSB SSW

  - SSB(SS due to Between subject): 개체 간 제곱합
  - SSW(SS due to Within subject): 개체 내 제곱합
  - 개체 내 제곱합 분해

$$\sum\sum(Y_{ij}-\overline{Y}_{i.})^2=\sum\sum(\overline{Y}_{.j}-\overline{Y}_{..})^2+\sum\sum(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})^2$$

SSW SSTR SSE

자유도 np-n 자유도 p-1 자유도 (n-1)(p-1)

$\sum_j(\overline{Y}_{.j}-\overline{Y}_{..})=0$

  - EMS

① $E(MSB)=\sigma^2+p\sigma_s^2$

② $E(MSTR)=\sigma^2+n\sum\tau_j^2/(p-1)$

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

① $E(SSB) = E\left[\sum\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right]$

$$= E\left[\sum\sum\left((\mu+\rho_i+\bar{\varepsilon}_{i\cdot}) - (\mu+\bar{\rho}+\bar{\varepsilon}_{\cdot\cdot})\right)^2\right]$$

$$= E\left[\sum\sum\left((\rho_i-\bar{\rho})+(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})\right)^2\right]$$

$$= E\left[\sum\sum(\rho_i-\bar{\rho})^2\right] + E\left[\sum\sum(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2\right]$$

$$= E\left[p\sum(\rho_i-\bar{\rho})^2\right] + (n-1)\sigma^2$$

$$= p(n-1)\sigma_S^2 + (n-1)\sigma^2$$

$$E(MSB) = E\left(\frac{SSB}{n-1}\right) = \frac{E(SSB)}{n-1} = p\sigma_S^2+\sigma^2$$

이때 $Y_{ij} = \mu+\rho_i+\tau_j+\varepsilon_{ij}$

$$\bar{Y}_{i\cdot} = \frac{1}{p}\sum_j Y_{ij} = \mu+\rho_i+0+\bar{\varepsilon}_{i\cdot}$$

$$\bar{Y}_{\cdot\cdot} = \frac{1}{np}\sum\sum Y_{ij} = \mu+\bar{\rho}+0+\bar{\varepsilon}_{\cdot\cdot}$$

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

$$= E\left[\sum\sum\left((\mu+\bar{\rho}+\tau_j+\bar{\varepsilon}_{\cdot j}) - (\mu+\bar{\rho}+\bar{\varepsilon}_{\cdot\cdot})\right)^2\right]$$

$$= E\left[\sum\sum(\tau_j+\bar{\varepsilon}_{\cdot j}-\bar{\varepsilon}_{\cdot\cdot})^2\right]$$

$$= E\left[\sum\sum\tau_j^2\right] + E\left[\sum\sum(\bar{\varepsilon}_{\cdot j}-\bar{\varepsilon}_{\cdot\cdot})^2\right]$$

$$= n\sum_j\tau_j^2 + (p-1)\sigma^2$$

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

이때 $Y_{ij} = \mu+\rho_i+\tau_j+\varepsilon_{ij}$

$$\bar{Y}_{\cdot j} = \frac{1}{n}\sum_i Y_{ij} = \mu+\bar{\rho}+\tau_j+\bar{\varepsilon}_{\cdot j}$$

$$\bar{Y}_{\cdot\cdot} = \frac{1}{np}\sum\sum Y_{ij} = \mu+\bar{\rho}+0+\bar{\varepsilon}_{\cdot\cdot}$$

F가 왜 이런형태?

: MSTR과 MSE를 비교하면 H_0, H_1에 대한 목적 달성 가능!

$\Sigma \tau_j^2 = 0$,

하나라도 0이 아니면 MSTR > MSE

- 가설검정
  - $H_0 : \tau_1 = \tau_2 = \cdots = \tau_p = 0$
  - $$F = \frac{\frac{SSTR}{p-1}}{\frac{SSE}{(n-1)(p-1)}} = \frac{MSTR}{MSE} \sim F_{p-1,(n-1)(p-1)}$$ → 처리간의 효과검정

⇨ 개체 간 변동에 영향을 받지 않고 순수하게 개체 내 변동만 평가하기 때문에 처리평균 간 비교를 정밀화할 수 있음

SSTR과 SSE가 SSW 내에서 이루어짐

SSW는 개체 내 제곱합

→ 개체간의 변동의 영향 X

Sigma matrix

$$\begin{matrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_p) \\ Cov(X_2,X_1) & Var(X_2) & & \vdots \\ \vdots & \vdots & \ddots & \\ & Cov(X_p,X_{p-1}) & & Var(X_p) \end{matrix}$$

벡터적인 방법으로 접근

○ 다변량 분산분석과 자유도 수정

- 다변량적 접근방법 두껍게

$$Y_i = (Y_{i1},\ Y_{i2},\ \ldots,\ Y_{ip})^T, \qquad i = 1,\ \ldots, n$$

  - ○ $Y_i \sim N_p(\mu,\ \Sigma),\ \ \mu = (\mu_1, \mu_2,\ \ldots, \mu_p)^T$

    벡터가 normal

Σ matrix

$$\begin{pmatrix} \sigma_s^2+\sigma^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2+\sigma^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2+\sigma^2 \end{pmatrix}$$

- 공분산행렬 → 조건 ⟨ 대칭, positive semidefinite (음수X)

  - ○ 복합대칭성(compound symmetry)

    = 구형성

$$\sigma_{jk} = Cov(Y_{ij}, Y_{ik}) = \begin{cases} \sigma_s^2 + \sigma^2, & j = k \ \text{①} \\ \sigma_s^2, & j \neq k \ \text{②} \end{cases}$$

repeated measure model을 다변량으로 접근? →

- 일변량적 접근방법은 복합대칭성 또는 구형성(sphericity)과 같은 특수한 조건을 만족하는 경우 타당 → 단변량과 같음?! ≠ : 일변량 분산분석으로 진행

✓ 복합대칭성 또는 구형성 가정을 위배하는 경우 다변량 검정 또는 자유도 수정이 필요 (앞의 가정이 틀린 것이므로)

12/4

* 복합대칭성이 성립하는 경우

① diagnal ($j=k$)

$$Cov(Y_{ij}, Y_{ik}) = Cov(Y_{ij}, Y_{ij})$$

$$= Var(Y_{ij})$$

상수

$$= Var(\mu + \rho_i + \tau_j + \varepsilon_{ij})$$

$$= Var(\rho_i + \varepsilon_{ij})$$

$$= \sigma_s^2 + \sigma^2 + 0$$

② not diagonal ($j \neq k$)

iid

$$Cov(Y_{ij}, Y_{ik}) = Cov(\mu + \rho_i + \tau_j + \varepsilon_{ij}, \mu + \rho_i + \tau_k + \varepsilon_{ik})$$

$$= Cov(\rho_i, \rho_i)$$

$$= Var(\rho_i)$$

$$= \sigma_s^2$$

다변량 ⊃ 단변량 (아무때나 쓰면 안됨!) : 다변량이 더 broad한 개념!

↑ 공분산에 대한 가정 X

- 다변량 분산분석(MANOVA, Multivariate ANOVA)
  - $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = 0$
  - $\mu_j = \mu + \tau_j \Rightarrow H_0 : \tau = 0 \rightarrow \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_p \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

    공통평균

11/29

- Hotelling's T^2 통계량

  i번째 사람의
  p번째 관측치

  - $W_i = (W_{i1}, \ldots, W_{i,p-1})^T \overset{\text{정의}}{\equiv} (Y_{i1} - Y_{ip}, \ldots, Y_{i,p-1} - Y_{ip})^T$
  - $H_0 : \tau = 0 \Rightarrow H_0 : \mu_W = 0$

    $\begin{pmatrix} \overline{W_1} \\ \vdots \\ \overline{W_{p-1}} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} W_{i1} \\ \vdots \\ \frac{1}{n}\sum W_{ip} \end{pmatrix}$

    $\begin{pmatrix} W_{11}-\overline{W_1} \\ \vdots \\ W_{1p-1}-\overline{W_{p-1}} \end{pmatrix} \cdot (W_{11}-\overline{W_1} \cdots W_{1p-1}-\overline{W_{p-1}})$

  - 검정통계량 : $T^2 = n\,\overline{W}^T S_W^{-1} \overline{W}$

    ⇒ $1\times(p-1) \cdot (p-1)\times(p-1) \cdot (p-1)\times 1 = 1\times 1$ 차원 (스칼라)

    $S_W = \frac{1}{n-1}\sum_{i=1}^{n} \underbrace{(W_i - \overline{W})}_{\mathbb{R}^{(p-1)\times 1}} \underbrace{(W_i - \overline{W})^T}_{\mathbb{R}^{1\times(p-1)}}$

    ⇒ $(p-1)\times(p-1)$한 matrix를 componentwise하게 $\frac{1}{n-1}\Sigma$ 계산

$$F_t = \frac{n-p+1}{(n-1)(p-1)} T^2 \sim F_{p-1,\, \boxed{n-p+1}}$$

$n \geq p$ 인 경우에만 적용가능

- 단변량 분모의 자유도 = $(n-1)(p-1)$
- 다변량 분모의 자유도 = $n-p+1$ < $(n-1)(p-1)$

  $= np-n-p+1$

  ↳ 정리하면 $n < n(p-1)$

→ 분모의 자유도가 커질수록 꼬리쪽의 확률이 작아짐 (p-value가 커짐?)

n은 샘플수, p는 차원수
고전통계학은 $n \geq p$라고 가정
최신통계학은 그렇지 않은 경우 有 (ex. 유전학) → 새로운 방법론 필요

단변량적방법은 위의조건을 만족해야 쓸수있음

- 자유도 수정방법
  - 다변량적 방법에서는 공분산행렬에 대한 가정이 없음

    ⇨ 구형성 가정을 어느 정도 만족하는 경우 자유도 손실이 큼

    단변량적 방법을 쓸수 있는데 쓰지 않았음

    → 분모의 자유도가 작아짐 ; (n-1)(p-1)이 아니라 n-p+1

    → 꼬리쪽의 확률이 커짐 (p-value 값이 커짐) ??
  - 자유도 수정계수 ϵ 이용
    - $\epsilon \in [0, 1]$
    - $\epsilon \rightarrow 1$ 구형성 만족
    - 복합대칭성(구형성) 가정 하에서 $F \sim F_{p-1,(n-1)(p-1)}$ ←단변량
    - 실제 F는 $F_{\epsilon(p-1),\epsilon(n-1)(p-1)}$에 근사
  - ϵ의 추정
    - Greenhouse and Geisser ϵ_{GG}, Huynh and Feldt ϵ_{HF}
    - $\epsilon_{GG} \leq \epsilon_{HF}$, $(\because \epsilon_{HF} = \min(\epsilon_{HF}, 1))$
  - ϵ이 충분히 크면 일변량적 방법 이용

ex) $\epsilon_{GG} = 0.6049$라면?

→ 자유도수정 : $(p-1) \times 0.6049$, $(n-1)(p-1) \times 0.6049$

□ 다요인 반복측정 분산분석

○ 반복요인과 분류요인이 하나씩인 경우

◉ 반복측정 2요인 실험자료: 반복요인이 B인 경우 → 한개체에대해 B를여러번반복

요인 A	개체	요인 B				평균
		1	2	3	4	
1	1	0	0	5	3	2.00
	2	3	1	5	4	4.25
	3	4	3	6	2	4.75
2	1	4	2	7	8	5.20
	2	5	4	6	6	5.20
	3	7	5	8	9	5.80
평균		3.83	2.50	6.17	5.33	4.46

○ 개체(subject)는 일종의 블록요인이고 요인 A에 지분되어(nested) 있음

- 통계모형식

i가 j에 nested

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + s_{i(j)} + \varepsilon_{ijk},$$

$$\underbrace{i = 1, \ldots, s}_{\text{개체}},\ \underbrace{j = 1,\ \ldots, a}_{\text{요인A}},\ \underbrace{k = 1,\ \ldots, b}_{\text{요인B}}$$

subject effect

- $s_{i(j)}$: 개체효과로 일반적으로 랜덤(변량)요인 → 추정해야할 모수의 개수가 매우 줄어듦! (분포에 관여하는 모수만 추정)
- $E(Y_{ijk}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$
- $Var(Y_{ijk}) = \sigma^2 + \sigma_s^2$
- $Cov(Y_{ijk}, Y_{ijk'}) = \sigma_s^2, \quad k \neq k'$
- $Cov(Y_{ijk}, Y_{i'j'k'}) = 0, \quad i \neq i$ or/and $j \neq j'$

$E(Y_{ijk}) = E(\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \underbrace{s_{i(j)} + \varepsilon_{ijk}}_{})$ → random effect니까 0

$Var(Y_{ijk}) = Var(\underbrace{\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}}_{\text{상수니까 } 0} + \underbrace{s_{i(j)}}_{\sigma_s^2} + \underbrace{\varepsilon_{ijk}}_{\sigma^2})$

요인B가 같지 않을경우 →

상수는 모두 0

$Cov(Y_{ijk}, Y_{ijk'}) = Cov(\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + s_{i(j)} + \varepsilon_{ijk},\ \mu + \alpha_j + \beta_{k'} + (\alpha\beta)_{jk'} + s_{i(j)} + \varepsilon_{ijk'})$

$= Cov(s_{i(j)} + \varepsilon_{ijk},\ s_{i(j)} + \varepsilon_{ijk'}) = \sigma_s^2 + \underbrace{0 + 0 + 0}_{iid} = \sigma_s^2$

$Cov(Y_{ijk}, Y_{i'j'k'}) = Cov(\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + s_{i(j)} + \varepsilon_{ijk},\ \mu + \alpha_{j'} + \beta_{k'} + (\alpha\beta)_{j'k'} + s_{i'(j')} + \varepsilon_{i'j'k'})$

$Cov(s_{i(j)} + \varepsilon_{ijk},\ s_{i'(j')} + \varepsilon_{i'j'k'}) = \underbrace{0 + 0 + 0 + 0}_{iid} = 0$

- 변동분해

nested

$$TSS = \underbrace{SSA + SS_{subject(A)}}_{SS_{between}} + \underbrace{SSB + SS(AB) + SSE}_{SS_{within}}$$

i와 k에 대한 Σ가 모두 상수 $\quad \because \sum(\bar{Y}_{.j.}-\bar{Y}_{...})=0$

- $SSA = bs\sum_j(\bar{Y}_{.j.}-\bar{Y}_{...})^2 \ : \ a-1 \ \Rightarrow \ E(MSA)=\sigma^2+b\sigma_s^2+bs\sum\alpha_j^2/(a-1)$
- $SSB = as\sum_k(\bar{Y}_{..k}-\bar{Y}_{...})^2 \ : \ b-1 \ \Rightarrow \ E(MSB)=\sigma^2+as\sum\beta_k^2/(b-1)$
- $SS(AB) = s\sum_j\sum_k(\bar{Y}_{.jk}-\bar{Y}_{.j.}-\bar{Y}_{..k}+\bar{Y}_{...})^2 \ : \ (a-1)(b-1)$

$(\bar{Y}_{.jk}-\bar{Y}_{...})-(\bar{Y}_{.j.}-\bar{Y}_{...})-(\bar{Y}_{..k}-\bar{Y}_{...})$

$$\Rightarrow \ E(MS(AB))=\sigma^2+s\sum\sum(\alpha\beta)_{jk}^2/((a-1)(b-1))$$

- $SSS(A) = b\sum_i\sum_j(\bar{Y}_{ij.}-\bar{Y}_{.j.})^2 \ : \ a(s-1) \ \Rightarrow \ E(MSS(A))=\sigma^2+b\sigma_s^2$
- $SSE = \sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{.jk}-\bar{Y}_{ij.}+\bar{Y}_{.j.})^2 \ : \ a(s-1)(b-1) \ \Rightarrow \ \boxed{E(MSE)=\sigma^2}$

나머지

$$(abs-1)-\underbrace{(a-1)}_{SSA}-\underbrace{(b-1)}_{SSB}-\underbrace{(a-1)(b-1)}_{SS(AB)}-\underbrace{a(s-1)}_{SSS(A)}$$

$SS_{between} = b\sum_i\sum_j(\bar{Y}_{ij.}-\bar{Y}_{...})^2 \rightarrow$ 자유도 $as-1$

$(as-1)-(a-1)=a(s-1)$

★기말고사 질의응답 다시듣기

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

$$= E\left[bs\sum_j\left((\mu+\alpha_j+\bar{S}_{.(j)}+\bar{\varepsilon}_{.j.})-(\mu+\bar{S}_{.(.)}+\bar{\varepsilon}_{...})\right)^2\right]$$

$$= E\left[bs\sum_j(\alpha_j+\bar{S}_{.(j)}-\bar{S}_{.(.)}+\bar{\varepsilon}_{.j.}-\bar{\varepsilon}_{...})^2\right]$$

$$= bs\sum_j\alpha_j^2 + bE\left[s\sum_j(\bar{S}_{.(j)}-\bar{S}_{.(.)})^2\right] + bsE\left[\sum_j(\bar{\varepsilon}_{.j.}-\bar{\varepsilon}_{...})^2\right]$$

$$= bs\sum_j\alpha_j^2 + b(a-1)\sigma_S^2 + (a-1)\sigma^2$$

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

이때 $Y_{ijk} = \mu+\alpha_j+\beta_k+(\alpha\beta)_{jk}+S_{i(j)}+\varepsilon_{ijk}$

$$\bar{Y}_{.j.} = \frac{1}{\overset{bs?}{\cancel{ab}}}\sum_i\sum_k Y_{ijk} = \mu+\alpha_j+0+0+\bar{S}_{.(j)}+\bar{\varepsilon}_{.j.}$$

$$\bar{Y}_{...} = \frac{1}{abs}\sum_i\sum_j\sum_k Y_{ijk} = \mu+0+0+0+\bar{S}_{.(.)}+\bar{\varepsilon}_{...}$$

$$\bar{Y}_{..k} = \frac{1}{as}\sum_i\sum_j Y_{ijk} = \mu+0+\beta_k+0+\bar{S}_{.(.)}+\bar{\varepsilon}_{..k}$$

$$\bar{Y}_{.jk} = \frac{1}{s}\sum_i Y_{ijk} = \mu+\alpha_j+\beta_k+(\alpha\beta)_{jk}+\bar{S}_{.(j)}+\bar{\varepsilon}_{.jk}$$

$$\bar{Y}_{ij.} = \frac{1}{b}\sum_k Y_{ijk} = \mu+\alpha_j+0+0+S_{i(j)}+\bar{\varepsilon}_{ij.}$$

12/4

$$E(SSB) = E\left[as\sum_k(\bar{Y}_{..k}-\bar{Y}_{...})^2\right]$$

$$= E\left[as\sum_k\left((\mu+\beta_k+\bar{S}_{.(.)}+\bar{\varepsilon}_{..k})-(\mu+\bar{S}_{.(.)}+\bar{\varepsilon}_{...})\right)^2\right]$$

$$= E\left[as\sum_k(\beta_k+\bar{\varepsilon}_{..k}-\bar{\varepsilon}_{...})^2\right]$$

$$= E\left[as\sum_k\beta_k^2\right] + E\left[as\sum_k(\bar{\varepsilon}_{..k}-\bar{\varepsilon}_{...})^2\right]$$

$$= as\sum_k\beta_k^2 + (b-1)\sigma^2$$

$$E(MSB) = E\left(\frac{SSB}{b-1}\right) = \frac{as\sum_k\beta_k^2}{b-1} + \sigma^2$$

$$E(SSS(A)) = E\left[b\sum_i\sum_j(\bar{Y}_{ij.}-\bar{Y}_{.j.})^2\right]$$

$$= E\left[b\sum_i\sum_j\left((\mu+\alpha_j+S_{i(j)}+\bar{\varepsilon}_{ij.})-(\mu+\alpha_j+\bar{S}_{.(j)}+\bar{\varepsilon}_{.j.})\right)^2\right]$$

$$= E\left[b\sum_i\sum_j(S_{i(j)}-\bar{S}_{.(j)}+\bar{\varepsilon}_{ij.}-\bar{\varepsilon}_{.j.})^2\right]$$

$$= E\left[b\sum_i\sum_j(S_{i(j)}-\bar{S}_{.(j)})^2\right] + E\left[b\sum\sum(\bar{\varepsilon}_{ij.}-\bar{\varepsilon}_{...})^2\right]$$

$$= a(s-1)b\sigma_S^2 + (s-1)a\sigma^2$$

$$E(MSS(A)) = E\left(\frac{SSS(A)}{\underline{a(s-1)}}\right) = b\sigma_S^2+\sigma^2$$

$$E(SSAB) = E\left[s\sum_j\sum_k(\bar{Y}_{.jk}-\bar{Y}_{...})^2\right]$$

$$= E\left[s\sum_j\sum_k\left((\mu+\alpha_j+\beta_k+(\alpha\beta)_{jk}+\bar{S}_{.(j)}+\bar{\varepsilon}_{.jk})-(\mu+\bar{S}_{.(.)}+\bar{\varepsilon}_{...})\right)^2\right]$$

$$= E\left[s\sum_j\sum_k(\alpha_j+\beta_k+(\alpha\beta)_{jk}+\bar{S}_{.(j)}-\bar{S}_{.(.)}+\bar{\varepsilon}_{.jk}-\bar{\varepsilon}_{...})^2\right]$$

$$= E\left[s\sum_j\sum_k\alpha_j^2\right] + E\left[s\sum_j\sum_k\beta_k^2\right] + E\left[s\sum_j\sum_k(\alpha\beta)_{jk}^2\right] + \underline{E\left[s\sum_j\sum_k(\bar{S}_{.(j)}-\bar{S}_{.(.)})^2\right]} + E\left[s\sum_j\sum_k(\bar{\varepsilon}_{.jk}-\bar{\varepsilon}_{...})^2\right]$$

$$= sb\sum_j\alpha_j^2 + sa\sum_k\beta_k^2 + s\sum_j\sum_k(\alpha\beta)_{jk}^2 + b(a-1)\sigma_S^2 + (ab-1)\sigma^2$$

$E\left[bs\sum_j(\bar{S}_{.(j)}-\bar{S}_{.(.)})^2\right]$

$= E\left[b\sum_j\sum_k(\bar{S}_{.(j)}-\bar{S}_{.(.)})^2\right]$, 이때 $\sum_j(\bar{S}_{.(j)}-\bar{S}_{.(.)})=0$

자유도 $a-1$

$\sum_j\sum_k(\bar{\varepsilon}_{.jk}-\bar{\varepsilon}_{...})=0$

자유도 $ab-1$

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

$$= sb\sum_j\alpha_j^2 + sa\sum_k\beta_k^2 + s\sum_j\sum_k(\alpha\beta)_{jk}^2 + b(a-1)\sigma_S^2 + (ab-1)\sigma^2 - \left(bs\sum_j\alpha_j^2 + b(a-1)\sigma_S^2 + (a-1)\sigma^2\right) - \left(as\sum_k\beta_k^2 + (b-1)\sigma^2\right)$$

$$= s\sum_j\sum_k(\alpha\beta)_{jk}^2 + (a-1)(b-1)\sigma^2$$

$$E(MS(AB)) = E\left(\frac{SS(AB)}{\underline{(a-1)(b-1)}}\right) = \sigma^2 + \frac{s\sum_j\sum_k(\alpha\beta)_{jk}^2}{(a-1)(b-1)}$$

자유도의 합관계

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

* SSE의 자유도 = SST의 자유도 − SSA 자유도 − SSB 자유도 − SS(AB) 자유도 − SSS(A) 자유도

$= abs-1-(a-1)-(b-1)-(a-1)(b-1)-a(s-1)$

$= a(s-1)(b-1)$

MSE의 기댓값 → σ^2

※ 무엇과 비교해야 내 목적을 위한 F를 만들 수 있을지!

cf) $\hat{\sigma}_s^2 = ?$

$\frac{MSS(A) - MSE}{b}$ 이용

- 유의성검정
  - 요인 A의 주효과: $H_0 : \alpha_1 = \cdots = \alpha_a = 0$

$$\begin{cases} E(MSA) = \sigma^2 + b\sigma_s^2 + \frac{bs\sum \alpha_j^2}{a-1} \\ E(MSS(A)) = \sigma^2 + b\sigma_s^2 \end{cases}$$

$$F_A = \frac{SSA/(a-1)}{SS_{subject(A)}/(a(s-1))} \sim F_{a-1,a(s-1)}$$

  - 요인 B의 주효과: $H_0 : \beta_1 = \cdots = \beta_b = 0$

$$\begin{cases} E(MSB) = \sigma^2 + \frac{as\sum \beta_k^2}{b-1} \\ E(MSE) = \sigma^2 \end{cases}$$

$$F_B = \frac{SSB/(b-1)}{SSE/(a(b-1)(s-1))} \sim F_{b-1,a(b-1)(s-1)}$$

  - 상호작용 (A*B)의 효과: $H_0 : (\alpha\beta)_{ij} = 0, \quad \forall\, i, j$

$$F_{(AB)} = \frac{SS(AB)/(a-1)(b-1)}{SSE/(a(b-1)(s-1))} \sim F_{(a-1)(b-1),a(b-1)(s-1)}$$

$$\begin{cases} E(MS(AB)) = \sigma^2 + \frac{s\sum\sum (\alpha\beta)_{jk}^2}{(a-1)(b-1)} \\ E(MSE) = \sigma^2 \end{cases}$$

기말