실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

이원배치법

이원배치 분산분석 모형(고정효과)

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

$$i = 1,2,\ldots,a;\ j = 1,2,\ldots,b;\ k = 1,2,\ldots,k$$

$$\sum_{i=1}^{a} \alpha_i = 0$$

$$\sum_{j=1}^{b} \beta_i = 0$$

$$\sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$$

$\epsilon_{ijk} \sim N(0, \sigma^2)$ 이고 모두 독립

- α_i: 요인 A에 의한 차이
- β_j: 요인 B에 의한 차이
- $(\alpha\beta)_{ij}$: 요인 A와 B의 상호작용에 의한 차이

제곱합의 분해와 분산분석표

제곱합의 분해

$$y_{ijk}-\bar{y}_{...}=(\bar{y}_{i..}-\bar{y}_{...})+(\bar{y}_{.j.}-\bar{y}_{...})+(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})+(y_{ijk}-\bar{y}_{ij.})$$

에서 양변을 제곱하여 합하면

$$\sum_{i,j,k}(y_{ijk}-\bar{y}_{...})^2=\sum_{i,j,k}(\bar{y}_{i..}-\bar{y}_{...})^2+\sum_{i,j,k}(\bar{y}_{.j.}-\bar{y}_{...})^2+\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2+\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij.})^2$$

$$+2\sum_{i,j,k}(\bar{y}_{i..}-\bar{y}_{...})(\bar{y}_{.j.}-\bar{y}_{...})$$ 등의 교차항

- 교차항은 모두 0임을 증명할 수 있음.
- $CT=\dfrac{y_{...}^2}{abn}=abn\cdot\bar{y}_{...}^2$ 라 하면
  - $\sum_{i,j,k}(y_{ijk}-\bar{y}_{...})^2=\sum_{i,j,k}y_{ijk}^2-abn\times\bar{y}_{...}^2=\sum_{i,j,k}y_{ijk}^2-CT$

제곱합의 분해와 분산분석표

- $\sum_{i,j,k}(\bar{y}_{i..}-\bar{y}_{...})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2 = bn\sum_{i=1}^{a}\bar{y}_{i..}^2 - \frac{y_{...}^2}{abn} = bn\sum_{i=1}^{a}\bar{y}_{i..}^2 - CT$
- $\sum_{i,j,k}(\bar{y}_{.j.}-\bar{y}_{...})^2 = \sum_{j=1}^{b}an(\bar{y}_{.j.}-\bar{y}_{...})^2 = an\sum_{j=1}^{b}\bar{y}_{.j.}^2 - CT$
- $\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2 = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$

각 제곱합은 다음과 같음.

- 전체 제곱합 SST: $\sum_{i,j,k}(y_{ijk}-\bar{y}_{...})^2 = \sum_{i,j,k}y_{ijk}^2 - CT$
- 요인 A에 의한 제곱합 SSA: $\sum_{i,j,k}(\bar{y}_{i..}-\bar{y}_{...})^2 = bn\sum_{i=1}^{a}\bar{y}_{i..}^2 - CT$
- 요인 B에 의한 제곱합 SSB: $\sum_{i,j,k}(\bar{y}_{.j.}-\bar{y}_{...})^2 = an\sum_{j=1}^{b}\bar{y}_{.j.}^2 - CT$

제곱합의 분해와 분산분석표

- 요인 A, B의 상호작용 제곱합 SSAB:

$$\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2 = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$$

$$= n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij.}^2 - bn\sum_{i=1}^{a}\bar{y}_{i..}^2 - an\sum_{j=1}^{b}\bar{y}_{.j.}^2 + CT$$

- 오차제곱합 SSE: $\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij.})^2$

따라서

SST = SSA + SSB + SSAB + SSE

실제 계산에서

SSE = SST - SSA - SSB - SSAB 로 SSE 계산

제곱합의 분해와 분산분석표

각 제곱합의 자유도

- SST = $\sum_{i,j,k}(y_{ijk}-\bar{y}_{...})^2$의 자유도: $abn-1$
- SSA = $\sum_{i,j,k}(\bar{y}_{i..}-\bar{y}_{...})^2$ 의 자유도: $a-1$
- SSB = $\sum_{i,j,k}(\bar{y}_{.j.}-\bar{y}_{...})^2$ 의 자유도: $b-1$
- SSAB = $\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$의 자유도: $ab-a-b+1=(a-1)(b-1)$
- SSE = $\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij.})^2$의 자유도: $abn-ab=ab(n-1)$

자유도도 다음이 성립:

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

제곱합의 분해와 분산분석표

분산분석표

요인	제곱합	자유도	평균제곱(MS)	F	유의확률
A	SSA	$a-1$	$\mathrm{MSA}=\dfrac{\mathrm{SSA}}{a-1}$	$F_0=\dfrac{\mathrm{MSA}}{\mathrm{MSE}}$	$\Pr[F_{a-1,ab(n-1)}>F_0]$
B	SSB	$b-1$	$\mathrm{MSB}=\dfrac{\mathrm{SSB}}{b-1}$	$F_0=\dfrac{\mathrm{MSB}}{\mathrm{MSE}}$	$\Pr[F_{b-1,ab(n-1)}>F_0]$
A×B	SSAB	$(a-1)(b-1)$	$\mathrm{MSAB}=\dfrac{\mathrm{SSAB}}{(a-1)(b-1)}$	$F_0=\dfrac{\mathrm{MSAB}}{\mathrm{MSE}}$	$\Pr[F_{(a-1)(b-1),ab(n-1)}>F_0]$
오차	SSE	$ab(n-1)$	$\mathrm{MSE}=\dfrac{\mathrm{SSE}}{ab(n-1)}$		
전체	SST	$abn-1$			

제곱합의 분해와 분산분석표

- 귀무가설 $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ (요인 A에 의한 차이 없음)은

  $F_0 = \dfrac{\text{MSA}}{\text{MSE}} > F_{a-1, ab(n-1);\alpha}$ 이거나 유의확률이 유의수준 α보다 작으면 기각

- 귀무가설 $H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$ (요인 B에 의한 차이 없음)은

  $F_0 = \dfrac{\text{MSB}}{\text{MSE}} > F_{b-1, ab(n-1);\alpha}$ 이거나 유의확률이 유의수준 α보다 작으면 기각

- 귀무가설 $H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{ab} = 0$ (요인 A와 B에 의한 상호작용 없음)은

  $F_0 = \dfrac{\text{MSAB}}{\text{MSE}} > F_{(a-1)(b-1), ab(n-1);\alpha}$ 이거나 유의확률이 유의수준 α보다 작으면 기각

제곱합의 분해와 분산분석표

평균제곱의 기댓값(모수모형)

$$E(\mathrm{MSA}) = \sigma^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i^2$$

$$E(\mathrm{MSB}) = \sigma^2 + \frac{an}{b-1}\sum_{j=1}^{b}\beta_j^2$$

$$E(\mathrm{MSAB}) = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2$$

$$E(\mathrm{MSE}) = \sigma^2$$

- σ^2에 대한 추정치는 분산분석표의 MSE
- 각 귀무가설이 참이면 해당 평균제곱의 기댓값이 모두 σ^2