실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

제곱합의 분해와 분산분석-Nested Design

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(A) = $\sum_{i,j,k}(\bar{y}_{ij.}-\bar{y}_{i..})^2$ 의 자유도: $ab-a=a(b-1)$

  SSB의 자유도와 SSAB의 자유도 합 $(b-1)+(a-1)(b-1)$과 같음
- SSE = $\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij.})^2$의 자유도: $abn-ab=ab(n-1)$

자유도도 다음이 성립:

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

제곱합의 분해와 분산분석-Nested Design

분산분석표(두 요인 A, B가 모두 고정효과)

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

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

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

- 귀무가설 $H_0:\beta_{1(1)}=\beta_{2(1)}=\cdots=\beta_{b(a)}=0$ (요인 B(A)에 의한 차이 없음)은

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

제곱합의 분해와 분산분석-Nested Design

고정효과와 다름

분산분석표(두 요인 A, B(A)가 모두 임의효과)

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

- A의 주효과 검정: 귀무가설 $H_0:\sigma_\alpha^2=0$ 대 대립가설 $H_1:\sigma_\alpha^2\neq 0$

  $F_0=\dfrac{\mathrm{MSA}}{\mathrm{MSB(A)}}>F_{a-1,a(b-1);\alpha}$ 이면 H_0 기각

- B(A)의 효과 검정: 귀무가설 $H_0:\sigma_\beta^2=0$ 대 대립가설 $H_1:\sigma_\beta^2\neq 0$

  $F_0=\dfrac{\mathrm{MSB(A)}}{\mathrm{MSE}}>F_{a(b-1),ab(n-1);\alpha}$ 이면 H_0 기각

수치예제 및 SAS-Nested Design

예: 학교나 담임선생님에 따라 학생들의 학업성취도가 차이나는지 알아보기 임의로 3개의 학교를 선정하고 각 학교에서 2개의 반을 선택하여 성취도 테스트를 한 결과이다. (학교를 요인 A, 학급(담임)을 요인 B)

A1		A2		A3	
B1	B2	B1	B2	B1	B2
20, 18, 14	19, 20, 20	14, 18, 14	12, 12, 9	13, 16, 13	9, 4, 4

$\sum_{ijk} y_{ijk}^2 = 3873$, $\sum_{i,j,k} y_{ijk} = 249$, $CT = \dfrac{249^2}{3 \cdot 2 \cdot 3} = 3444.5$ 이고 필요한 수준별 평균 $\bar{y}_{i..}$ 및 $\bar{y}_{ij.}$ 는 다음과 같다.

	B1	B2	평균	전체 평균
A1	17.333	19.667	18.500	
A2	15.333	11.000	13.167	13.83333
A3	14.000	5.667	9.833	

$\bar{y}_{ij.}$ $\bar{y}_{i..}$ $\bar{y}_{...}$

각 제곱합은

수치예제 및 SAS-Nested Design

$$\mathrm{SST} = \sum_{i,j,k} y_{ijk}^2 - CT = 3873 - 3444.5 = 428.5$$

$$\mathrm{SSA} = bn\sum_{i=1}^{a} \bar{y}_{i..}^2 - CT = 2 \cdot 3(18.5^2 + 13.167^2 + 9.8333^2) - 3444.5 = 229.333$$

1번째 수준

$$\mathrm{SSB(A)} = n\sum_{i,j}^{a} \bar{y}_{ij.}^2 - bn\sum_{i,j}^{a} \bar{y}_{i..}^2$$

$$= 3(17.333^2 + 19.667^2 + 15.333^2 + 11^2 + 14^2 + 5.667^2) - 2 \cdot 3(18.5^2 + 13.167^2 + 9.8333^2)$$

$$= 140.500$$

$$\mathrm{SSE} = \mathrm{SST} - \mathrm{SSA} - \underbrace{\mathrm{SSB\text{-}SSAB}}_{SS\,B(A)} = 428.5 - 229.333 - 140.500 = 58.666$$

으로 계산된다.

참고로

$$\mathrm{SSB} = an\sum_{j=1}^{b} \bar{y}_{.j.}^2 - CT = 3 \cdot 3(15.556^2 + 12.111^2) - 3444.5 = 53.388$$

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

수치예제 및 SAS-Nested Design

$$= 3(17.333^2 + 19.667^2 + 15.333^2 + 11^2 + 14^2 + 5.667^2)$$

$$- 2 \cdot 3(18.500^2 + 13.167^2 + 9.833^2) - 3 \cdot 3(15.556^2 + 12.111^2) - 3444.5$$

$$= 3814.333 - 3673.833 - 3497.889 + 3444.5 = 87.111$$

이므로 SSB(A) = SSB + SSAB $= 53.388 + 87.111 = 140.499$ 로 얻을 수도 있다.

이 경우 학교 및 담임선생님이 모두 임의효과이므로 분산분석표는

요인	제곱합	자유도	평균제곱(MS)	F	유의확률
A	229.333	2	114.666	2.448	0.2342
B(A)	140.500	3	46.833	9.58	0.0017
오차	58.666	12	4.888		
전체	428.5	17			

- $F_0 = 2.448 < F_{2,3;0.05} = 9.55$ 이므로 귀무가설 $H_0 : \sigma_\alpha^2 = 0$을 기각하지 못함(따라서 학교 차이 없음)
-

수치예제 및 SAS-Nested Design

- $F_0 = 9.58 > F_{3,12;0.05} = 3.4903$ 이므로 귀무가설 $H_0 : \sigma_\beta^2 = 0$을 기각(따라서 담임선생님 차이 있음)

위의 유의확률은 다음과 같음

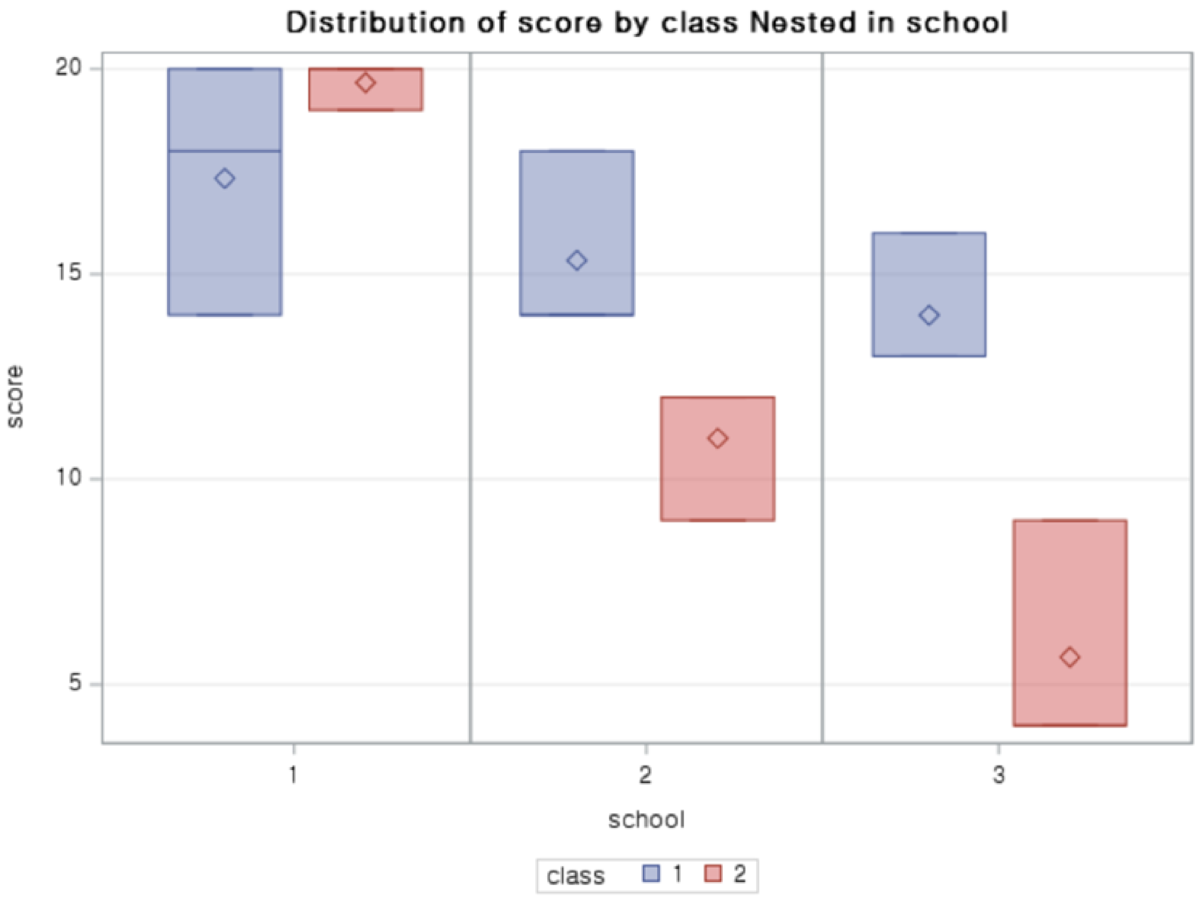
```
> 1-pf(2.448, 2,3)
[1] 0.2341915
> 1-pf(9.58,3,12)
[1] 0.001655408
```

수치예제 및 SAS-Nested Design

SAS를 사용한 내포설계 분석

```
data a; /* nested1.sas */
input school class score@@;
cards;
1 1 20    1 1 18    1 1 14
1 2 19    1 2 20    1 2 20
2 1 14    2 1 18    2 1 14
2 2 12    2 2 12    2 2 9
3 1 13    3 1 16    3 1 13
3 2 9     3 2 4     3 2 4
;
proc glm data=a;
  class school class;
  model score = school class(school);
  random school class(school)/test;
run;
```

수치예제 및 SAS-Nested Design



수치예제 및 SAS-Nested Design

고정효과 모형일때의 ANVA table:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
school	2	229.3333333	114.6666667	23.45	<.0001
class(school)	3	140.5000000	46.8333333	9.58	0.0017

MSA, MSB(A) 의 기댓값:

Source	Type III Expected Mean Square
school	Var(Error) + 3 Var(class(school)) + 6 Var(school)
class(school)	Var(Error) + 3 Var(class(school))

Dependent Variable: score

Source	DF	Type III SS	Mean Square	F Value	Pr > F
school	2	229.333333	114.666667	2.45	0.2342
Error	3	140.500000	46.833333		
Error: MS(class(school))					

수정된 분산분석표:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
class(school)	3	140.500000	46.833333	9.58	0.0017
Error: MS(Error)	12	58.666667	4.888889		

수치예제 및 SAS-Nested Design

위의 기댓값을 보면 school에 검정은 class(school)로 나누어야 함. 따라서 TEST 문을 사용

```
/* nested2.sas */
data step 생략
proc glm data=a;
  class school class;
  model score = school class(school);
  test H = school E= class(school);
run;
```

검정결과

Tests of Hypotheses Using the Type III MS for class(school) as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
school	2	229.3333333	114.6666667	2.45	0.2342

수치예제 및 SAS-Nested Design

다중비교:

- 이 경우 class(school)이 유의하므로 다중비교는 이를 기준으로 실행.
- 이 경우 주효과(school)는 실제 유의하지도 않고 유의하더라도 이 주효과에 대한 다중비교는 상대적으로 의미없음.

```
/* nested3.sas */
data step 생략

proc glm data=a;
  class school class;
  model score = school class(school);
  random school class(school)/test;
  lsmeans class(school) / adjust=tukey lines;
run;
```

수치예제 및 SAS-Nested Design

모든 가능한 조합의 비교

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

class	school	score LSMEAN	LSMEAN Number
1	1	17.3333333	1
2	1	19.6666667	2
1	2	15.3333333	3
2	2	11.0000000	4
1	3	14.0000000	5
2	3	5.6666667	6

Least Squares Means for effect class(school)
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: score

i/j	1	2	3	4	5	6
1		0.7838	0.8691	0.0389	0.4748	0.0003
2	0.7838		0.2299	0.0045	0.0721	<.0001
3	0.8691	0.2299		0.2299	0.9728	0.0018
4	0.0389	0.0045	0.2299		0.5780	0.0975
5	0.4748	0.0721	0.9728	0.5780		0.0061
6	0.0003	<.0001	0.0018	0.0975	0.0061	

수치예제 및 SAS-Nested Design