

5장 | 다중비교

SAS를 이용한 실험 계획과 분산 분석 (자유아카데미)

다중비교의 필요성

- ANOVA test 후 귀무가설을 기각했을 때 후속적인 결과 분석
 - 차이는 어디에서 발생하는가?
 - 어느 실험 조건을 선택할 것인가?
- 예) 송아지의 사료에 따라 체중이 다르다면 과연 어느 사료를 써야 하는가?
- 군집분석에 유용함

ONE POSSIBLE WAY

$$y_{1j} \overset{i.i.d.}{\sim} N(\mu_1, \sigma^2), \quad y_{2j} \overset{i.i.d.}{\sim} N(\mu_2, \sigma^2)$$

$$\bar{y}_{1.} \sim N(\mu_1, \sigma^2/n_1), \quad \bar{y}_{2.} \overset{i.i.d.}{\sim} N(\mu_2, \sigma^2/n_2)$$

$$\bar{y}_{1.} - \bar{y}_{2.} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \sigma^2\right)$$

$$\frac{(\bar{y}_{1.} - \bar{y}_{2.}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \hat{\sigma}^2}} \sim t_{n_1+n_2-2}$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = MSE$$

95% Confidence Interval for $\mu_i - \mu_j$:

$$P\left(-t_{n_1+n_2-2}(0.025) \leq \frac{(\bar{y}_{1.} - \bar{y}_{2.}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) MSE}} \leq t_{n_1+n_2-2}(0.025)\right) = 0.95$$

$$(\bar{y}_{1.} - \bar{y}_{2.}) \pm t_{n_1+n_2-2}(0.025) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) MSE}$$

$$\text{If } n_1 = n_2 = n, \text{ then } n_1 + n_2 - 2 = 2(n - 1)$$

$$\text{If } n_1 = n_2 = \dots = n_a = n, \text{ then } n_1 - 1 + n_2 - 1 + \dots + n_a - 1 = a(n - 1)$$

여러가지 다중 비교

- Bonferroni Correction (본페로니 수정)*
- Duncan's Multiple Range Test (던칸의 다중범위 검정)
- Scheffe's Test (쉐페의 검정)
- Tukey's Test (튜키의 검정)
- Dunnett's Test (듀넛의 검정)*
- LSD Test (최소유의차검정)*

다중비교의 검정력

- 검정력이란 귀무가설이 거짓일때 귀무가설을 기각할 확률 – 그룹간 차이를 구분하는 능력

높은 검정력			낮은 검정력
LSD	Duncann	Tukey/Dunnett	Scheffe
<—————Bonferroni—————>			

최소 유의차 검정: LEAST SIGNIFICANT DIFFERENCE (LSD)

$$C.I. \text{ for } \tau_i - \tau_j \text{ is } \left\{ (\bar{y}_{i.} - \bar{y}_{j.}) \pm t_{0.025, a(n-1)} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \right\}$$

$$LSD_{(i,j)} = t_{0.025, a(n-1)} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

모든 (i, j) 에 대하여 LSD_(i,j) 를 계산한 다음,

$$|\bar{y}_{i.} - \bar{y}_{j.}| > LSD_{(i,j)} \text{ 이면, } \tau_i \neq \tau_j$$

(LSD = “유의한 차이를 보이는 최소의 거리”)

예

$$LSD_{(i,j)} = 11.005, \text{ for any } i, j$$

$$\bar{y}_{1.} = 74.4$$

$$\bar{y}_{2.} = 78.2$$

$$\bar{y}_{3.} = 68.6$$

$$\bar{y}_{4.} = 79.8$$

1) 그룹평균을 계산하여 크기 순으로 나열한다

2) LSD 보다 작은 차이는 같은 line 상에 존재하도록 underline 을 긋는다.

$\bar{y}_{4.}$	$\bar{y}_{2.}$	$\bar{y}_{1.}$	$\bar{y}_{3.}$
79.8	78.2	74.4	68.6

79.8 78.2 74.4 68.6

SAS CODE

```
data a;  
  
input treat $ y @@; cards;  
  
1 74 1 87 2 64 2 53 ... 4 38 4 55  
  
;  
  
proc glm data=a;  
  
    class treat;  
  
    model y=treat;  
  
    means treat / lsd lines;   혹은 / t lines;   혹은 / pdiff lines;  
  
run;
```


SAS OUTPUT

The GLM Procedure

t Tests (LSD) for y

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	16
Error Mean Square	67.375
Critical Value of t	2.11991
Least Significant Difference	11.005

Means with the same letter are not significantly different.

t	Grouping	Mean	N	treat
	A	79.800	5	4
	A			
B	A	78.200	5	2
B	A			
B	A	74.400	5	1
B				
B		68.600	5	3

본페로니 수정

최소유의차검정에서 LSD를 계산할 때 유의수준 α 를 사용하였다. 그러나 여러 개의 귀무가설을 각각 유의수준 α 로 검정하면 전체적인 ‘집단오류율(familywise error rate: FWER)’이 α 보다 커지는 현상이 발생한다. 예를 들어 다음과 같은 m 개의 귀무가설을 동시에 검정한다고 가정할 때

$$\left\{ \begin{array}{ll} H_{01} : & \mu_1 = \mu_2 \\ H_{02} : & \mu_1 = \mu_3 \\ \vdots & \vdots \\ H_{0m} : & \mu_{a-1} = \mu_a \end{array} \right. \quad (5.4)$$

본페로니 수정

m 개의 귀무가설을 각각 유의수준 α 로 검정한다면(E_i 를 H_{0i} 에 대한 제1종 오류라고 할 때) 본페로니 부등식(Bonferroni's inequality)³에 의해

$$\begin{aligned}\text{집단오류율} &= P(\text{제1종 오류}) = P\left(\bigcup_{i=1}^m E_i\right) \\ &\leq \sum_{i=1}^m P(E_i) = m \times \alpha\end{aligned}\tag{5.5}$$

가 되므로 전체적인 유의수준(집단오류율: FWER)은 $m\alpha$ 까지 증가할 수 있다.

그러나 만일 우리가 각 귀무가설을 α 대신 $\frac{\alpha}{m}$ 로 검정한다면 집단오류율(FWER)은 α 로 유지될 것이다. 이를 ‘본페로니 수정(Bonferroni correction)’이라고 한다.

본페로니 수정

- 본페로니 수정은 귀무가설의 수가 많아지면 m (가능한 짝의 조합 수)이 증가하기 때문에 아주 ‘보수적인 검정 (conservative test)’이 될 수 있다. 이는 두 그룹이 웬만한 차이가 아니고는 서로 다르다고 인정하지 않는다는 의미이다.

본페로니 수정 (예)

- 기름의 종류 (fat1, fat2, fat3, fat4)에 따라 생성되는 트랜스 지방의 양을 측정하는 실험에서 본페로니 수정을 통해 트랜스 지방이 많은 기름과 적은 기름을 나누어 보자.

[2] 도넛(doughnut)을 기름에 튀길 때 생성되는 트랜스지방의 양을 측정하는 실험을 하였다. 기름의 종류는 4가지(fat1, fat2, fat3, fat4)이고 이 가운데 fat1과 fat2는 동물성 기름이며 fat3과 fat4는 식물성 기름이다.

fat1	fat2	fat3	fat4
164	178	175	155
172	191	193	166
168	197	178	149
177	182	171	164
156	185	163	170
195	177	176	168

표 4.5: 여러 종류의 기름에 대해 도넛 튀김에 발생하는 트랜스지방의 양

본페로니 수정의 예

$$t_{\frac{0.025}{6}, 20} \sqrt{100.9 \left(\frac{1}{6} + \frac{1}{6} \right)} = 16.975 \quad (5.6)$$

가 되고 LSD인 12.097보다 큰 값이 되어 LSD보단 보수적인 검정이 된다. 이를 이용하면

$$\begin{aligned} |\bar{y}_{1.} - \bar{y}_{2.}| &= 13.0 < 16.975, & |\bar{y}_{1.} - \bar{y}_{3.}| &= 4.0 < 16.975 \\ |\bar{y}_{1.} - \bar{y}_{4.}| &= 10.0 < 16.975, & |\bar{y}_{2.} - \bar{y}_{3.}| &= 9.0 < 16.975 \\ |\bar{y}_{2.} - \bar{y}_{4.}| &= 23.0 > 16.975, & |\bar{y}_{3.} - \bar{y}_{4.}| &= 14.0 < 16.975 \end{aligned}$$

$\bar{y}_{2.}$	$\bar{y}_{3.}$	$\bar{y}_{1.}$	$\bar{y}_{4.}$
185.0	176.0	172.0	162.0

$$(5.7)$$

SAS CODE

```
proc glm data=a;  
    class fat;  
    model trans = fat ;  
    means fat / bon lines ;  
run;
```

or

```
lsmeans fat / adjust=bon lines ;
```

SAS OUTPUT

The GLM Procedure

Bonferroni (Dunn) t Tests for trans

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9
Critical Value of t	2.92712
Minimum Significant Difference	16.976

Means with the same letter are not significantly different.				
Bon Grouping		Mean	N	fat
	A	185.000	6	2
	A			
B	A	176.000	6	3
B	A			
B	A	172.000	6	1
B				
B		162.000	6	4

듀넛의 검정

- 대부분의 실험 계획은 한 개의 대조군 (control group)과 여러 개의 처리군 (treatment group)으로 이루어지는 경우가 많다.
- 여기서 대조군은 기존의 실험조건이나 위약 (placebo)등이 배정되는 그룹이며, 처리군은 관심있는 새로운 방법이나 실험 조건이 배정되는 그룹을 의미한다.

듀넛의 검정

만일 ANOVA F 검정에서 그룹 간 차이가 유의하다고 결론이 나왔다면, 이제 우리의 관심은 처리군 중 어느 그룹이 대조군과 유의하게 차이나는가에 있게 된다. 듀넛의 검정은 이와 같은 상황에 최적화된 다중비교 방법이다. 총 k 개의 처리그룹이 존재할 때 각 그룹평균이 대조군(control group)의 평균과 유의하게 다른지를 알아보기 위해 아래와 같은 신뢰구간을 사용한다. 여기서 $d_{\alpha, k, \text{df(MSE)}}$ 는 듀넛이 만든 표(Appendix 2 참조)에서 얻을 수 있으며 n_i 는 i 번째 그룹의 관측치 개수이고 n_c 는 대조군의 관측치 개수이다.

$$(\bar{y}_{i\cdot} - \bar{y}_{\text{control}}) \pm d_{\alpha, k, \text{df(MSE)}} \sqrt{\text{MSE} \left(\frac{1}{n_i} + \frac{1}{n_c} \right)} \quad (5.8)$$

듀넛의 검정의 예

- 도넛을 튀길 때 사용하는 기름의 종류 (fat1, fat2, fat3, fat4)에 따라 생성되는 트랜스 지방의 양을 조사함에 있어서 fat2가 그 동안 사용해온 기름이라고 하자.
- 이때 어느 기름이 트랜스 지방 생성량에 있어 기존 기름과 다른지 듀넛의 검정을 통해 살펴보자

듀넛의 검정의 예

우리는 기존의 기름이 fat2 이고 $k = 3$ 이고 $df(MSE) = 20$ 그리고 MSE는 100.9 임을 알고 있다. 따라서 듀넛의 검정 결과는 아래와 같다.

$$\begin{cases} H_{01} : \mu_1 = \mu_2, & H_{11} : \mu_1 \neq \mu_2 \\ H_{02} : \mu_3 = \mu_2, & H_{12} : \mu_3 \neq \mu_2 \\ H_{03} : \mu_4 = \mu_2, & H_{13} : \mu_4 \neq \mu_2 \end{cases} \quad (5.9)$$

$$\frac{|\bar{y}_{1.} - \bar{y}_{2.}|}{\sqrt{MSE \left(\frac{1}{6} + \frac{1}{6} \right)}} = \frac{13.0}{5.799} = 2.241 < 2.54 = d_{0.05, 3, 20}$$

$$\frac{|\bar{y}_{3.} - \bar{y}_{2.}|}{\sqrt{MSE \left(\frac{1}{6} + \frac{1}{6} \right)}} = \frac{9.0}{5.799} = 1.551 < 2.54 = d_{0.05, 3, 20}$$

$$\frac{|\bar{y}_{4.} - \bar{y}_{2.}|}{\sqrt{MSE \left(\frac{1}{6} + \frac{1}{6} \right)}} = \frac{23.0}{5.799} = 3.966 > 2.54 = d_{0.05, 3, 20}$$

fat1 과 fat3는 기존의 기름인 fat2에 비해 트랜스지방이 비슷하게 생성되었으나, fat4는 기존의 기름과 트랜스지방 생성량이 유의하게 달랐다.

SAS CODE

```
proc glm data=a;  
  class fat;  
  model trans = fat ;  
  means fat / dunnett('2') ; or lsmeans fat / adjust=dunnett pdiff=control('2') ;  
run;
```

SAS OUTPUT

The GLM Procedure

Dunnett's t Tests for trans

Note: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9
Critical Value of Dunnett's t	2.54035
Minimum Significant Difference	14.733

Comparisons significant at the 0.05 level are indicated by ***.

fat Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	-9.000	-23.733	5.733	
1 - 2	-13.000	-27.733	1.733	
4 - 2	-23.000	-37.733	-8.267	***

MEANS VS. LSMEANS

- Means (평균)
 - 흔히 산술 평균이라고 부르는 개념으로 그룹내 관측치의 합을 관측치 개수로 나눈 값이다.
- LSMEANS (Least Squares Means/최소제곱평균)
 - 모형에 존재하는 다른 요인의 효과를 보정한 평균값으로서 반복수가 같은 균형 자료 (balanced data)인 경우에는 MEANS와 값이 항상 일치한다.
 - 공분산 분석 (ANCOVA)와 같이 공변량 (covariate)이 존재할 때는 MEANS대신 LSMEANS를 사용한다.

MEANS VS. LSMEANS

- (예: 이원배치법의 반복수가 일정하지 않을 때)

저장시간과 온도에 따른		온도(Temp)	
박테리아 수		20°C	30°C
시간(Time)	3개월	2, 5	9, 12, 15
	6개월	6, 6, 7, 7	16, 15

MEANS VS. LSMEANS

만일 MEANS 를 사용하여 저장시간(Time)별 박테리아 수의 평균값을 계산한다면

(Time = 3 개월) 때의 mean:

$$\frac{2 + 5 + 9 + 12 + 15}{5} = 8.6 \quad (5.10)$$

(Time = 6 개월) 때의 mean:

$$\frac{6 + 6 + 7 + 7 + 16 + 15}{6} = 9.5 \quad (5.11)$$

가 되어 두 저장시간에 따른 박테리아 수의 평균 차이는 0.9가 된다.

MEANS VS. LSMEANS

그러나 만일 LSMEANS를 사용하여 저장시간(Time)별 박테리아 수의 평균값을 계산한다면 온도(Temp)라는 요인을 추가로 고려해서

(Time = 3 개월) 때의 lsmean:

$$\begin{aligned} & \frac{1}{2} \{ (20^{\circ}\text{C에서 3 개월 보관한 그룹평균}) + (30^{\circ}\text{C에서 3 개월 보관한 그룹평균}) \} \\ &= \frac{1}{2} \left\{ \frac{2+5}{2} + \frac{9+12+15}{3} \right\} = 7.75 \end{aligned} \quad (5.12)$$

(Time = 6 개월) 때의 lsmean:

$$\begin{aligned} & \frac{1}{2} \{ (20^{\circ}\text{C에서 6 개월 보관한 그룹평균}) + (30^{\circ}\text{C에서 6 개월 보관한 그룹평균}) \} \\ &= \frac{1}{2} \left\{ \frac{6+6+7+7}{4} + \frac{16+15}{2} \right\} = 11.0 \end{aligned} \quad (5.13)$$

이 되어, 두 저장시간의 차이에 따른 박테리아 수의 평균 차이는 0.9가 아닌 2.25가 된다.

SUMMARY: MEANS VS. LSMEANS

- **LSmeans** consider other variables in a regression model

>> Mean Adjusted to the Model (회귀모형을 고려한 평균)

- Balanced data : LSmeans = Means
- If missing value exists, **LSmeans** should be used

SAS CODE (예)

```
proc glm data=a;  
    class temp time;  
    model y=temp time;  
    lsmeans temp / pdiff lines cl adjust=tukey ;  
run;
```

SAS CODE

```
lsmeans fat / pdiff cl lines
```

```
lsmeans fat / pdiff cl adjust = tukey lines;
```

[2] 도넛(doughnut)을 기름에 튀길 때 생성되는 트랜스지방의 양을 측정하는 실험을 하였다. 기름의 종류는 4가지(fat1, fat2, fat3, fat4)이고 이 가운데 fat1과 fat2는 동물성 기름이며 fat3과 fat4는 식물성 기름이다.

fat1	fat2	fat3	fat4
164	178	175	155
172	191	193	166
168	197	178	149
177	182	171	164
156	185	163	170
195	177	176	168

표 4.5: 여러 종류의 기름에 대해 도넛 튀김에 발생하는 트랜스지방의 양

DIFFOGRAM

- 다중 비교에서 그룹 간 차이를 시각화하는 방법으로 비슷한 그룹평균을 밑 줄로 연결하는 방법이 존재
- 하지만 그룹의 개수가 많거나 불균형 자료의 경우에는 이와 같은 밑줄로 표현 할 수 없는 경우가 발생.
- 이 문제의 보완을 위해 평균-평균 산점도로 표현되는 시각화 기법.

DIET DATA (저칼로리, 저지방, 저탄수화물, 대조군)

```
proc glm data=a;  
  
    class treat;  
  
    model weight = treat;  
  
    lsmeans treat / adjust=tukey;  
  
run;
```

Cal	Carb	Fat	Cont.
6.6	3.4	3.0	1.2

SAS 시스템

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

treat	weight LSMEAN	LSMEAN Number
Cal	6.60000000	1
Carb	3.40000000	2
Cont	1.20000000	3
Fat	3.00000000	4

Least Squares Means for effect treat
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: weight

i/j	1	2	3	4
1		0.0425	0.0007	0.0205
2	0.0425		0.2199	0.9823
3	0.0007	0.2199		0.3769
4	0.0205	0.9823	0.3769	

SAS OUTPUT

- 평균값
- 두 평균 차이의 신뢰구간이 대각선을 지나는지 여부 관찰

if 대각선 지남 (빨강)

=> 차이가 유의하지 않음

if 대각선 안 지남 (파랑)

=> 차이가 유의함

