실험계획과 분산분석(001) 과제 1

소프트웨어학부 컴퓨터과학과 2016133 이유진

#1 $p=4, N=12$

1)

	자유도	제곱합	평균제곱	F-통계량
처리	4-1=3	1.9788	0.6596	31.1868
오차	12-4=8	0.1692	0.02115	
전체	12-1=11	2.148		

	A_1	A_2	A_3	A_4
합	25.08	26.1	28.44	26.58
평균	8.36	8.7	9.48	8.86

$$SSTR = (25.08^2 + 26.1^2 + 28.44^2 + 26.58^2)/3 - (25.08 + 26.1 + 28.44 + 26.58)^2/12$$

$$= 941.8488 - 939.87 = 1.9788$$

$$TSS = (8.44^2 + 8.59^2 + \cdots + 8.74^2) - (25.08 + 26.1 + 28.44 + 26.58)^2/12$$

$$= 942.018 - 939.87 = 2.148$$

$$SSE = TSS - SSTR = 0.1692$$

$$MSTR = SSTR/3 = 0.6596$$

$$MSE = SSE/8 = 0.02115$$

$$F = MSTR/MSE = 31.1868$$

2) $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ (A1의 분산 σ_1^2, A_2의분산 σ_2^2, .. 라고할때)

H_1: not H_0

$F = 31.1868 > F_{0.05,(3,8)} = 4.07$ 이므로 귀무가설 기각

따라서 5% 유의수준에서 반응온도에 따른 강도의 변화가 있다고 할수있다

37 ① 최소유의차

$$|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}| > t_{\frac{0.05}{2},8} \times \sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$ 이면 이 쌍의 차이는 유의함

$$= 2.306 \times \sqrt{0.02115}\sqrt{\frac{1}{3}+\frac{1}{3}} = 0.273$$

② Bonferroni

$$|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}| > t_{\frac{0.025}{6},8} \times \sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$ 이면 이 쌍의 차이는 유의함

$$= 3.479 \times \sqrt{0.02115}\sqrt{\frac{1}{3}+\frac{1}{3}} = 0.413$$

③ Scheffe

$$|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}| > \sqrt{3 \times F_{0.05,3,8}} \times \sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$ 이면 이 쌍의 차이는 유의함

$$= 4.066 \times \sqrt{0.02115}\sqrt{\frac{1}{3}+\frac{1}{3}} = 0.483$$

④ Tukey

$$|\bar{y}_{i\cdot}-\bar{y}_{j\cdot}| > \left(q(0.05,4,8)/\sqrt{2}\right) \times \sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$ 이면 이 쌍의 차이는 유의함

$$= (4.53/\sqrt{2}) \times \sqrt{0.02115}\sqrt{\frac{1}{3}+\frac{1}{3}} = 0.380$$

0는차이가있음을의미

종류	평균	종류	평균	차이	최소유의차	Bonferroni	Scheffe	Tukey
1	8.36	2	8.7	0.34	O	X	X	X
		3	9.48	1.12	O	O	O	O
		4	8.86	0.5	O	O	O	O
2	8.7	3	9.48	0.78	O	O	O	O
		4	8.86	0.16	X	X	X	X
3	9.48	4	8.86	0.62	O	O	O	O

	1	2	3	4
따라서 최소유의차	a	b	c	b
본페로니	a	ac	b	c
셰페	a	ac	b	c
튜키	a	ac	b	c

와 같은 모습으로 그룹화할 수 있다.

각각 다른 기호는 서로 유의함을 의미한다.

R코드첨부

```
> rat <- scan(what=list(1,0.1))
1: 1 8.44 1 8.36 1 8.28
4: 2 8.59 2 8.91 2 8.60
7: 3 9.36 3 9.41 3 9.69
10: 4 8.92 4 8.92 4 8.74
13:
Read 12 records
> names(rat) <- c("temp", "str")
> rat <- data.frame(rat)
> attach(rat)
> temp <- as.factor(temp)
>
```

→데이터입력

```
> pairwise.t.test(str,temp,p.adj="none")

        Pairwise comparisons using t tests with pooled SD

data:  str and temp

  1       2       3
2 0.01960 -       -
3 1.1e-05 0.00015 -
4 0.00269 0.20814 0.00068

P value adjustment method: none
```

α=0.05보다 작은경우
차이가유의함

```
> pairwise.t.test(str,temp,p.adj="bonf")
```

→ 본페로니

```
        Pairwise comparisons using t tests with pooled SD

data:  str and temp

  1       2       3
2 0.11762 -       -
3 6.7e-05 0.00089 -
4 0.01615 1.00000 0.00405

P value adjustment method: bonferroni
```

α=0.05보다 작은경우
차이가유의함

```
> result <- aov(str~temp, data=rat)
> TukeyHSD(result)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = str ~ temp, data = rat)

$temp
         diff         lwr        upr     p adj
2-1  0.3400000 -0.03421721  0.7142172 0.0755830
3-1  1.1266667  0.75244946  1.5008839 0.0000518
4-1  0.5000000  0.12578279  0.8742172 0.0115108
3-2  0.7866667  0.41244946  1.1608839 0.0006721
4-2  0.1600000 -0.21421721  0.5342172 0.5496123
4-3 -0.6266667 -1.00088388 -0.2524495 0.0029968
```

→ Tukey

이용 α=0.05보다 작은경우
차이가유의함

```
> library(agricolae)
> df <- result$df.residual
> mse <- sum(result$residuals^2)/df
> print(LSD.test(str,temp,df,mse))
$statistics
    MSerror Df     Mean       CV  t.value       LSD
  0.02048333  8 8.851667 1.616871 2.306004 0.2694727

$parameters
        test p.ajusted name.t ntr alpha
  Fisher-LSD      none   temp   4  0.05

$means
       str       std r      LCL      UCL  Min  Max   Q25  Q50   Q75
1 8.360000 0.0800000 3 8.169454 8.550546 8.28 8.44 8.320 8.36 8.400
2 8.700000 0.1819341 3 8.509454 8.890546 8.59 8.91 8.595 8.60 8.755
3 9.486667 0.1778576 3 9.296121 9.677213 9.36 9.69 9.385 9.41 9.550
4 8.860000 0.1039230 3 8.669454 9.050546 8.74 8.92 8.830 8.92 8.920

$comparison
NULL

$groups
       str groups
3 9.486667      a
4 8.860000      b
2 8.700000      b
1 8.360000      c

attr(,"class")
[1] "group"
```

→ LSD

→ (1,2), (1,3), (1,4) 의 차이가 유의함
(2,3)
(3,4)

```
> print(scheffe.test(str,temp,df,mse))
$statistics
    MSerror Df        F     Mean       CV  Scheffe CriticalDifference
  0.02048333  8 4.066181 8.851667 1.616871 3.492641          0.4081395

$parameters
     test name.t ntr alpha
  Scheffe   temp   4  0.05

$means
       str       std r  Min  Max   Q25  Q50   Q75
1 8.360000 0.0800000 3 8.28 8.44 8.320 8.36 8.400
2 8.700000 0.1819341 3 8.59 8.91 8.595 8.60 8.755
3 9.486667 0.1778576 3 9.36 9.69 9.385 9.41 9.550
4 8.860000 0.1039230 3 8.74 8.92 8.830 8.92 8.920

$comparison
NULL

$groups
       str groups
3 9.486667      a
4 8.860000      b
2 8.700000     bc
1 8.360000      c

attr(,"class")
[1] "group"
> |
```

→ scheffe

(1,3), (1,4) 의 차이가 유의함
(2,3)
(3,4)

4) 잔차 $e_{13} = Y_{13} - \overline{Y_{1.}} = 8.28 - 8.36 = -0.08$

studentized 잔차

$$e_{13} = \frac{n_1 - 1}{n_1} Y_{13} - \frac{1}{n_1} \sum_{k \neq 3} Y_{1k} = \frac{3-1}{3} \times 8.28 - \frac{1}{3}(8.44 + 8.36)$$

$$= -0.08$$

#2

H₀ : 동쪽호수와 서쪽호수간의 산소량의 차이가 없다

H_1 : not H_0

처리-대조비교 : 대조(호수1)와 처리(호수2, 호수3)로 동-서 비교

Dunnett 방법

: $|\bar{y}_c - \bar{y}_i| > t'(\alpha, p-1, N-p)\sqrt{MSE}\sqrt{1/n_c + 1/n_i}$ 시 차이가 유의함 → H_0 기각가능

종류	평균	종류	평균	차이	Dunnett
1	2.2	2	4.6	2.4	X
		3	21.1	18.9	O

따라서 동쪽호수와 서쪽호수2와는 산소량에 차이가 없고

서쪽호수 3과는 산소량에 차이가 있다고 할수있다.

데이터 입력

```
> rat <- scan(what=list(1,1))
1: 1 0 1 2 1 1 1 3 1 1 1 2 1 3 1 4 1 1 1 5
11: 2 1 2 3 2 4 2 6 2 8 2 7 2 5 2 3 2 4 2 5
21: 3 14 3 26 3 25 3 18 3 19 3 22 3 21 3 16 3 20 3 30
31:
Read 30 records
> names(rat) <- c("lake","ox")
> rat <- data.frame(rat)
> rat$lake <- as.factor(rat$lake)
> attach(rat)
> library(DescTools)
> DunnettTest(x=rat$ox, g=rat$lake)

  Dunnett's test for comparing several treatments with a control :
    95% family-wise confidence level

$`1`
    diff     lwr.ci    upr.ci     pval
2-1  2.4 -0.9057288  5.705729   0.1767
3-1 18.9 15.5942712 22.205729  4.2e-13 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

설치

$\alpha = 0.05$보다 작은경우 차이가유의함

#3

$$\begin{cases} H_0: \sigma_A^2 = \sigma_B^2 \\ H_1: \sigma_A^2 \neq \sigma_B^2 \end{cases}$$

검정통계량 $H^* = \dfrac{\max(S_i^2)}{\min(S_i^2)} \sim H(p, n-1)$ 이므로 계산하면

$$\begin{cases} S_A^2 = \dfrac{1}{6-1}\left(\Sigma x_i^2 - \dfrac{1}{n}(\Sigma x_i)^2\right) = 1.1 \\ S_B^2 = \dfrac{1}{6-1}\left(\Sigma x_i^2 - \dfrac{1}{n}(\Sigma x_i)^2\right) = 15.7667 \end{cases}$$

$H = \dfrac{15.1667}{1.1} = 14.3333 > H(2,5) = 7.15$ 이므로 H_0 기각

따라서 A와 B는 등분산이라고 할수없다.