

숙제4

통계학전공 3학년 안효준 (5463312)

Problem 1.

a) 댐요인과 지역요인이 모두 고정효과라는 가정하에서 분산분석을 실시해 보아라
(유의수준 0.05)

```

33 * 1번 ;
34 data a;
35   input Dam Area Pollution @@;
36   cards;
37 1 1 11 1 1 12 1 2 13 1 2 11 1 3 10 1 3 11
38 2 1 17 2 1 20 2 2 20 2 2 21 2 3 15 2 3 14
39 ;
40 run;
41
42 proc print data = a;
43 run;
44
45 * a), b) ;
46 proc glm data = a;
47   class Dam Area;
48   model Pollution = Dam Area(Dam); *지역은 댐의 내포요인;
49   *내포모형은 교호작용이 내포요인으로 들어가서 없으므로 Dam | Area(Dam)와 같이 쓰지 않는다;
50   random Dam Area(Dam) / test; *댐요인과 지역요인을 모두 random effect라고 가정;
51   *test: 앞에서 지정해준 random effect에 대한 분산분석을 진행;
52   lsmeans Area(Dam) / adjust = tukey lines;
53   *adjust = tukey: 튜키의 사후검정을 진행;
54   *lines: 모든 평균 쌍 간의 비교 결과를 표시;
55 run;

```

The GLM Procedure					
Dependent Variable: Pollution					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	166.4166667	33.28333333	23.49	0.0007
Error	6	8.5000000	1.41666667		
Corrected Total	11	174.9166667			

R-Square	Coeff Var	Root MSE	Pollution Mean
0.951405	8.161632	1.190238	14.58333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Dam	1	126.7500000	126.7500000	89.47	<.0001
Area(Dam)	4	39.6666667	9.9166667	7.00	0.0191

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Dam	1	126.7500000	126.7500000	89.47	<.0001
Area(Dam)	4	39.6666667	9.9166667	7.00	0.0191

H_0 : 댐에 따른 오염도에 차이가 없다. H_1 : 댐에 따른 오염도에 차이가 있다.
유의확률이 <.0001이므로 귀무가설 기각 (즉, 댐에 따른 오염도에 차이가 있다)

H_0 : 지역에 따른 오염도에 차이가 없다. H_1 : 지역에 따른 오염도에 차이가 있다.
유의확률이 0.0191이므로 귀무가설 기각 (즉, 지역에 따른 오염도에 차이가 있다)

b) 댐요인과 지역요인이 모두 임의효과라는 가정하에서 분산분석을 실시해 보아라
(유의수준 0.05)

(a의 코드를 그대로 사용)

Dependent Variable: Pollution					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Dam	1	126.750000	126.750000	12.78	0.0233
Error: MS(Area(Dam))	4	39.666667	9.916667		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Area(Dam)	4	39.666667	9.916667	7.00	0.0191
Error: MS(Error)	8	8.500000	1.416667		

(고정효과와 다르게 Dam의 F값을 MS(Error)가 아닌 MS(Area(Dam))으로 나누어 계산)

H_0 : 댐에 따른 오염도에 차이가 없다. H_1 : 댐에 따른 오염도에 차이가 있다.

유의확률이 0.0233이므로 귀무가설 기각 (즉, 댐에 따른 오염도에 차이가 있다)

H_0 : 지역에 따른 오염도에 차이가 없다. H_1 : 지역에 따른 오염도에 차이가 있다.

유의확률이 0.0191이므로 귀무가설 기각 (즉, 지역에 따른 오염도에 차이가 있다)

c) 어떤 환경학자는 수질의 오염도를 조사함에 있어서, 수심의 깊이에 따라 오염도가 달라질 수 있다고 생각하였다. 그래서 두 개의 댐의 중앙 위치에서 수심에 따라 1m, 5m, 10m로 각각 2병씩의 물을 담아왔다. 그 결과, 표 8.11을 얻었을 때 이에 해당하는 모형식과 분산분석표를 작성하여라(유의수준 0.05)

```

25 * c) ;
26 data b;
27     input Dam Depth Pollution @@;
28     cards;
29 1 1 11 1 1 12 1 2 13 1 2 11 1 3 10 1 3 11
30 2 1 17 2 1 20 2 2 20 2 2 21 2 3 15 2 3 14
31 ;
32 run;
33
34 proc print data = b;
35 run;
36
37 proc glm data = b;
38     class Dam Depth;
39     model Pollution = Dam Depth(Dam);
40     random Dam Depth(Dam) / test;
41     lsmeans Depth(Dam) / adjust = tukey lines;
42 run;

```

모형식: $y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$ (단, $i = 1, 2$, $j = 1, 2, 3$, $k = 1, 2$)

$\alpha_i \sim i.i.d. N(0, \sigma_\alpha^2)$, $\beta_{j(i)} \sim i.i.d. N(0, \sigma_\beta^2)$, $\epsilon_{ijk} \sim i.i.d. N(0, \sigma^2)$ (단, α_i , $\beta_{j(i)}$, ϵ_{ijk} 는 서로 독립)

The GLM Procedure					
Tests of Hypotheses for Random Model Analysis of Variance					
Dependent Variable: Pollution					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Dam	1	126.750000	126.750000	12.78	0.0233
Error	4	39.666667	9.916667		
Error: MS(Depth(Dam))					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Depth(Dam)	4	39.666667	9.916667	7.00	0.0191
Error: MS(Error)	8	8.500000	1.416667		

(분산분석표)

요인	자유도	SS	MS	F값	유의확률
Dam	1	126.75	126.75	12.78	0.0233
Depth(Dam)	4	39.66	9.91	7.00	0.0191
Error	6	8.50	1.41		
Total	11	174.91			

Problem 2. 컴퓨터 키보드의 종류(A, B, C, D)에 따라 타자속도에 차이가 있는지 알아보려고 한다. 임의로 8명의 비서를 실험에 참여시켜가지 키보드 종류에 대하여 각각 10분 동안 입력한 글자 수를 측정하고 표 10.11을 얻었다 (단위: 분당글자 수)

a) 이 실험에서 랜덤화 과정이 왜 필요하며, 어떻게 실시할지 기술하라.

비서가 키보드를 입력하는 것을 반복할수록 손이 익숙해지므로 비교적 처음 입력한 키보드보다 마지막에 입력한 키보드의 글자 수가 많을 수 있다. 따라서 각각의 비서가 입력하는 키보드의 순서(A, B, C, D)를 랜덤하게 섞는 과정이 필요하다.

b) RCBD 임을 고려하여 분산분석표를 작성하라.

```

44 * 2번 ;
45 data c;
46     input Key Sec Num @@;
47     cards;
48 1 1 79 1 2 80 1 3 77 1 4 75 1 5 82 1 6 77 1 7 78 1 8 76
49 2 1 74 2 2 79 2 3 73 2 4 70 2 5 76 2 6 78 2 7 72 2 8 74
50 3 1 82 3 2 86 3 3 80 3 4 79 3 5 79 3 6 81 3 7 80 3 8 80
51 4 1 79 4 2 81 4 3 77 4 4 78 4 5 82 4 6 77 4 7 77 4 8 78
52 ;
53 run;
54
55 proc print data = c;
56 run;
57
58 proc glm data = c;
59     class Key Sec;
60     model Num = Sec Key; * 비서가 block이므로 treat인 Key보다 먼저 입력 ;
61 run;

```

The GLM Procedure					
Dependent Variable: Num					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	271.2500000	27.1250000	9.70	<.0001
Error	21	58.7500000	2.7976190		
Corrected Total	31	330.0000000			

R-Square	Coeff Var	Root MSE	Num Mean
0.821970	2.144370	1.672608	78.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sec	7	104.0000000	14.8571429	5.31	0.0013
Key	3	167.2500000	55.7500000	19.93	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sec	7	104.0000000	14.8571429	5.31	0.0013
Key	3	167.2500000	55.7500000	19.93	<.0001

(분산분석표)

요인	자유도	SS	MS	F값	유의확률
비서	7	104.00	14.85		
키보드	3	167.25	55.75	19.93	<.0001
Error	21	58.75	2.79		
Total	31	330.00			

c) 비서를 고정 효과로 간주하고 S A S ® 를 이용하여 키보드 종류에 대한 다중 비교를 튜키의 검정을 통해 실시하라.

```

64 * c) ;
65 proc glm data = c;
66     class Key Sec;
67     model Num = Sec Key;
68     lsmeans Key / adjust = tukey lines;
69 run;

```

The GLM Procedure					
Dependent Variable: Num					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	271.2500000	27.1250000	9.70	<.0001
Error	21	58.7500000	2.7976190		
Corrected Total	31	330.0000000			

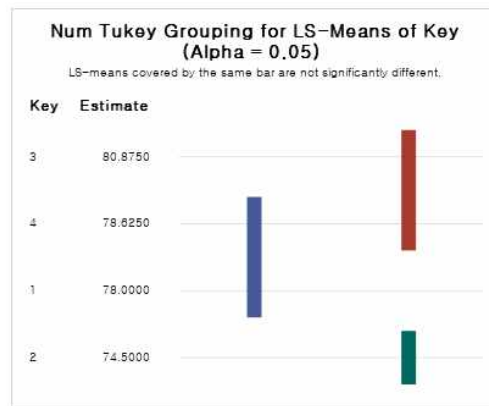
R-Square	Coeff Var	Root MSE	Num Mean
0.821970	2.144370	1.672608	78.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sec	7	104.0000000	14.8571429	5.31	0.0013
Key	3	167.2500000	55.7500000	19.93	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sec	7	104.0000000	14.8571429	5.31	0.0013
Key	3	167.2500000	55.7500000	19.93	<.0001

H_0 : 키보드에 따른 글자 수에 차이가 없다. H_1 : 키보드에 따른 글자 수에 차이가 있다.

분산분석표에서 키보드의 유의확률이 <.0001으로 귀무가설을 기각하게 되어 키보드의 종류에 따른 글자 수에 차이가 있음을 알 수 있으며, 어느 키보드에서 차이가 발생하는지 확인하기 위해 튜키의 사후검정을 실시



집단을 총 3개(빨강, 파랑, 초록)로 나눌 수 있으며, 서로 다른 집단끼리는 입력한 글자 수에 차이가 있다고 할 수 있다.

ex)

- 키보드 C와 키보드 A는 입력한 글자 수에 차이가 있다.
- 키보드 C와 키보드 B는 입력한 글자 수에 차이가 있다.
- 키보드 D와 키보드 B는 입력한 글자 수에 차이가 있다.
- 키보드 A와 키보드 B는 입력한 글자 수에 차이가 있다.