



실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

<http://jupiter.hallym.ac.kr>

제공합의 분해와 분산분석표

보기: 공장에서 생산되는 불량품의 수의 영향을 미치는 요인으로 기계(machine)에 의한 효과와 작업자(employee)에 의한 효과를 고려하여 이원배치법으로 실험을 하였다. 세 종류의 기계와 두 명의 작업자에 의한 차이를 알아보기 위해 이들 수준의 조합으로 실험을 행하여 다음의 자료를 얻었다.

	employee1	employee2
machine1	20, 18, 14	19, 20, 20
machine2	14, 18, 14	12, 12, 9
machine3	13, 16, 13	9, 4, 4

풀이:

$\sum_{ijk} y_{ijk}^2 = 3873$, $\sum_{i,j,k} y_{ijk} = 249$, $CT = \frac{249^2}{3 \cdot 2 \cdot 3} = 3444.5$ 이고, 각 수준별 평균을 계산해보면 다음과 같다.

	employee1	employee2	평균
machine1	17.333	19.667	18.500
machine2	15.333	11.000	13.167
machine3	14.000	5.667	9.833
평균	15.556	12.111	13.83333

제공합의 분해와 분산분석표

따라서 각 제공합은

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

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

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

$$\begin{aligned} SSAB &= n \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{ij.}^2 - bn \sum_{i=1}^a \bar{y}_{i..}^2 - an \sum_{j=1}^b \bar{y}_{.j.}^2 + CT \\ &= 3(17.333^2 + 19.667^2 + 15.333^2 + 11^2 + 14^2 + 5.667^2) \\ &\quad - 2 \cdot 3(18.500^2 + 13.167^2 + 9.833^2) - 3 \cdot 3(15.556^2 + 12.111^2) - 3444.5 \\ &= 3814.333 - 3673.833 - 3497.889 + 3444.5 = 87.111 \end{aligned}$$

$$SSE = SST - SSA - SSB - SSAB = 428.5 - 229.333 - 53.388 - 87.111 = 58.666$$

으로 계산된다.

따라서 분산분석표는

제곱합의 분해와 분산분석표

요인	제곱합	자유도	평균제곱(MS)	F	유의확률
machine	229.333	2	114.666	23.45	7.15e-05
employee	53.388	1	53.388	10.92	0.0063
m * e	87.111	2	43.555	8.91	0.0042
오차	58.666	12	4.888		
전체	428.500	17			

로 얻는다. 참고로 위의 유의확률은 R에서 다음과 같이 얻었다.

```
> 1-pf(23.45,2,12)
```

```
[1] 7.151476e-05
```

```
> 1-pf(10.92,1,12)
```

```
[1] 0.006287623
```

```
> 1-pf(8.91,2,12)
```

```
[1] 0.004246603
```

참고: 위의 제곱합 계산에서 인쇄된 것보다 더 많은 자리수를 사용하여 계산한 것으로 반올림을 중간중간에서 할 경우 값이 약간 차이날 수 있음

제공합의 분해와 분산분석표

위의 모든 유의확률이 모두 유의수준 0.05보다 작으므로 machine의 의한 주효과, employee에 의한 주효과 및 이 둘의 상호작용이 모두 유의하다.

기각역에 의한 검정은

- machine의 $F = 23.45 > F_{2,12;0.05} = 3.885$ 이므로 유의
 - employee의 $F = 10.92 > F_{1,12;0.05} = 4.747$ 이므로 유의
 - 상호작용의 $F = 8.91 > F_{2,12;0.05} = 3.885$ 이므로 유의
- 로 같은 결론을 얻는다.

SAS를 이용한 이원배치 분산분석

앞의 자료를 SAS를 사용하여 분석해보자.

```
data a; /* twoway1.sas */
input machine employee defects @@;
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 machine employee;
  model defects = machine employee machine * employee;
run;
```

SAS를 이용한 이원배치 분산분석

일원배치법과 기본적으로 같음.

자료의 구조는 요인을 표시하는 변수, 반응값을 표시되는 변수들로 열을 구성(시각적으로 보이는 표로 자료를 입력하지 않음)

요인이 두 개이므로

- class 문에 이 두 개의 요인이 설정되어야 함.
- model 문에서 이 두 개의 요인과 관련된 상호작용을 설정(상호작용은 *로 표시)
|을 사용하면 모든 가능한 상호작용이 적용됨. 즉,
MODEL Y = A|B 는 Y = A B A*B와 같고
MODEL Y = A|B|C 는 Y = A B C A*B A*C B*C A*B*C 와 같음

SAS를 이용한 이원배치 분산분석

Dependent Variable: defects

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	369.8333333	73.9666667	15.13	<.0001
Error	12	58.6666667	4.8888889		
Corrected Total	17	428.5000000			

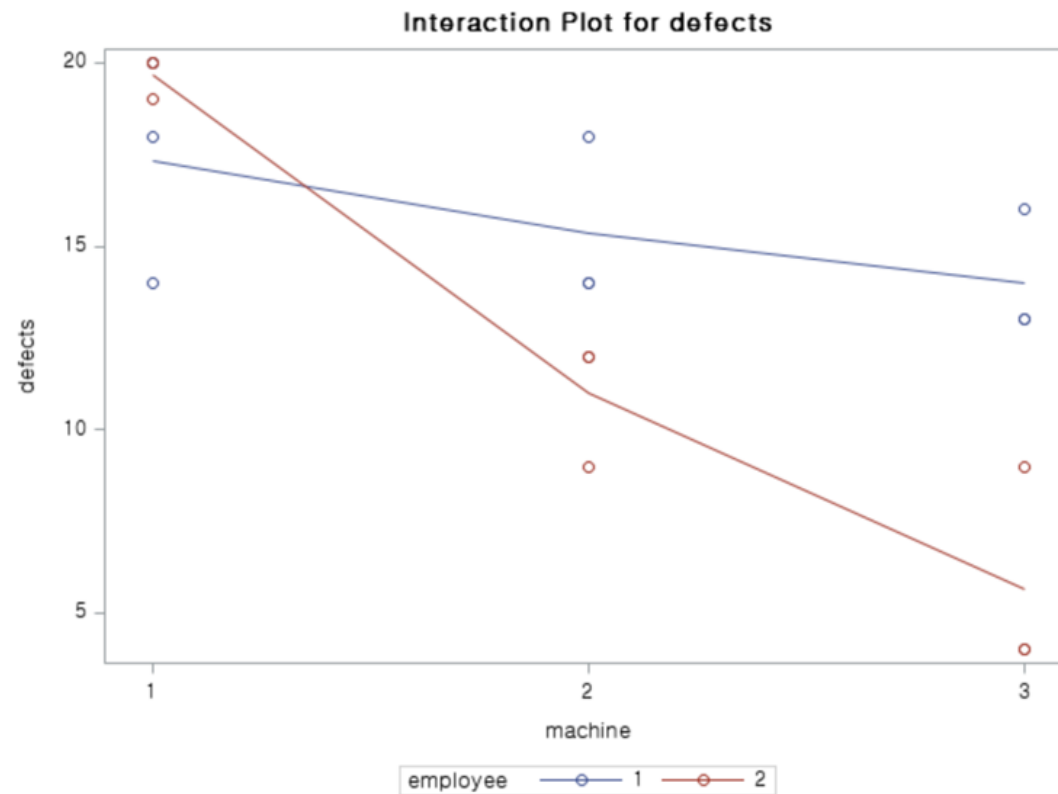
R-Square	Coeff Var	Root MSE	defects Mean
0.863088	15.98373	2.211083	13.83333

TYPE III Sum of Squares를 확인

Source	DF	Type III SS	Mean Square	F Value	Pr > F
machine	2	229.3333333	114.6666667	23.45	<.0001
employee	1	53.3888889	53.3888889	10.92	0.0063
machine*employee	2	87.1111111	43.5555556	8.91	0.0042

SAS를 이용한 이원배치 분산분석

상호작용에 대한 그래프



SAS를 이용한 이원배치 분산분석

data step 은 생략

```
proc glm data=a; /* twoway2.sas */  
  class machine employee;  
  model defects = machine employee machine * employee;  
  lsmeans machine*employee / slice= machine;  
  lsmeans machine*employee / slice= employee;  
run;
```

LSMEANS는

- 주어진 요인에 따른 평균을 계산
- MEANS와 유사하나 처리수준별 자료수가 같지 않은 것 등을 보정함.
- SLICE 에 의해서 특정 요인의 수준별 분석이 가능

SAS를 이용한 이원배치 분산분석

machine의 각 수준에서 employee에 따른 차이가 있는지 검정.

The GLM Procedure
Least Squares Means

machine*employee Effect Sliced by machine for defects					
machine	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	8,166667	8,166667	1,67	0,2205
2	1	28,166667	28,166667	5,76	0,0335
3	1	104,166667	104,166667	21,31	0,0006

그룹별 평균

The GLM Procedure
Least Squares Means

machine	employee	defects LSMEAN
1	1	17,3333333
1	2	19,6666667
2	1	15,3333333
2	2	11,0000000
3	1	14,0000000
3	2	5,6666667

SAS를 이용한 이원배치 분산분석

employee의 각 수준에서 machine에 따른 차이가 있는지 검정.

The GLM Procedure
Least Squares Means

machine*employee Effect Sliced by employee for defects					
employee	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	2	16.888889	8.444444	1.73	0.2192
2	2	299.555556	149.777778	30.64	<.0001