



실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

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

임의효과와 고정효과

보기: 앞의 기계와 작업자에서 기계 3개는 이 세 종류의 기계에 따른 차이를 알고자 한다고 가정하고, 작업자에 따른 차이를 알기 위해 작업자 중 임의로 2명을 뽑았다고 가정하면 혼합모형이 된다. (기계: 고정효과, 작업자: 임의효과)

따라서 분산분석표는

요인	제곱합	자유도	평균제곱(MS)	F	유의확률
machine	229.333	2	114.666	2.63	0.2753
employee	53.388	1	53.388	1.23	0.3836
m * e	87.111	2	43.555	8.91	0.0042
오차	58.666	12	4.888		
전체	428.500	17			

로 얻어지며,

유의확률을 볼 때 machine의 주효과, employee의 주효과 모두 유의하지 않으며 상호작용만 유의하다.

기각역을 사용한 검정결과는

임의효과와 고정효과

- 요인 A: $F_0 = 2.63 \leq F_{2,2;0.05} = 19.000$ 이므로 귀무가설 $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$ 을 기각하지 못한다.
- 요인 B: $F_0 = 1.23 \leq F_{1,2;0.05} = 18.51282$ 이므로 귀무가설 $H_0 : \sigma_{\beta}^2 = 0$ 을 기각하지 못한다.
- 상호작용: $F_0 = 8.91 > F_{2,12;0.05} = 3.8853$ 이므로 귀무가설 $H_0 : \sigma_{\alpha\beta}^2 = 0$ 을 기각한다.

임의효과와 고정효과

SAS를 사용한 분석

- 임의효과인 요인을 RANDOM 문에 설정 \Rightarrow 각 제곱합의 기댓값 출력됨
- RANDOM 문의 옵션으로 TEST를 사용 \Rightarrow 수정된 분산분석표 얻음
- RANDOM 문의 기댓값을 사용하여 TEST 문에 H에 분자에 사용할 제곱합, E에는 분모에 사용할 제곱합을 설정하여 수정된 가설검정 얻음.

data step은 생략

```
proc glm data=a; /* twoway3.sas */  
  class machine employee;  
  model defects = machine employee machine * employee;  
  random employee machine*employee /test;  
run;
```

MSA
MSAB

임의효과와 고정효과

각 평균제곱합의 기댓값

Source	Type III Expected Mean Square
machine	$\text{Var}(\text{Error}) + 3 \text{Var}(\text{machine} \times \text{employee}) + Q(\text{machine})$
employee	$\text{Var}(\text{Error}) + 3 \text{Var}(\text{machine} \times \text{employee}) + 9 \text{Var}(\text{employee})$
machine*employee	$\text{Var}(\text{Error}) + 3 \text{Var}(\text{machine} \times \text{employee})$

σ^2

σ^2

σ^2

수정된 분산분석표

Source	DF	Type III SS	Mean Square	F Value	Pr > F
machine	2	229.333333	114.666667	2.63	0.2753
employee	1	53.388889	53.388889	1.23	0.3836
Error	2	87.111111	43.555556		
Error: MS(machine*employee)					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
machine*employee	2	87.111111	43.555556	8.91	0.0042
Error: MS(Error)	12	58.666667	4.888889		

임의효과와 고정효과

```
proc glm data=a;  
  class machine employee;  
  model defects = machine employee machine * employee;  
  random employee machine*employee;  
  test h=machine e=machine*employee;  
  test h=employee e=machine*employee;
```

run: $\frac{h}{e}$ ↗

두 TEST 문의 결과

The GLM Procedure

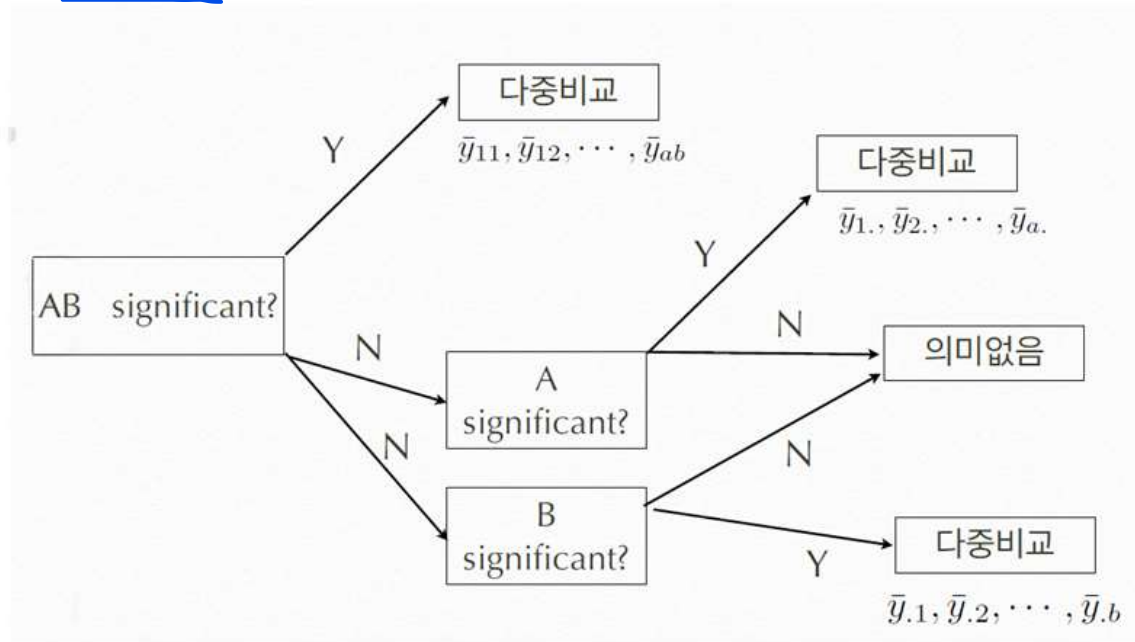
Dependent Variable: defects

Tests of Hypotheses Using the Type III MS for machine*employee as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
machine	2	229,3333333	114,6666667	2,63	0,2753
employee	1	53,3888889	53,3888889	1,23	0,3836

다중비교

이원배치법의 경우 요인이 두 개이므로 주효과에 대한 다중비교를 하거나 상호작용에 의한 다중비교를 하게 된다. 큰 원칙은

- 상호작용이 유의하면 두 효과의 모든 수준 조합에 대한 다중비교를 하고,
- 상호작용이 유의하지 않을 때만 주효과에 대한 다중비교를 한다.



다중비교

SAS 사용보기(Tukey 방법)

앞의 수치보기의 연속

data step 생략

```
proc glm data=a; /* twoway5.sas */
```

```
class machine employee;
```

```
model defects = machine employee machine * employee;
```

```
random employee machine*employee;
```

```
lsmeans machine*employee / adjust=tukey lines;
```

```
run;
```

- LSMEAN 문의 옵션에 사용할 다중비교의 이름 설정(LSD, BON, TUKEY 등)
- 이 자료의 경우 주효과가 유의하지 않으므로 상호작용(두 수준조합)에 대한 다중비교 적용

다중비교

각 수준별 평균값 및 차이에 대한 유의성 검정결과

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

machine	employee	defects LSMEAN	LSMEAN Number
1	1	17.3333333	1
1	2	19.6666667	2
2	1	15.3333333	3
2	2	11.0000000	4
3	1	14.0000000	5
3	2	5.6666667	6

Least Squares Means for effect machine*employee
Pr > |t| for H0: LSMean(i)=LSMean(j)

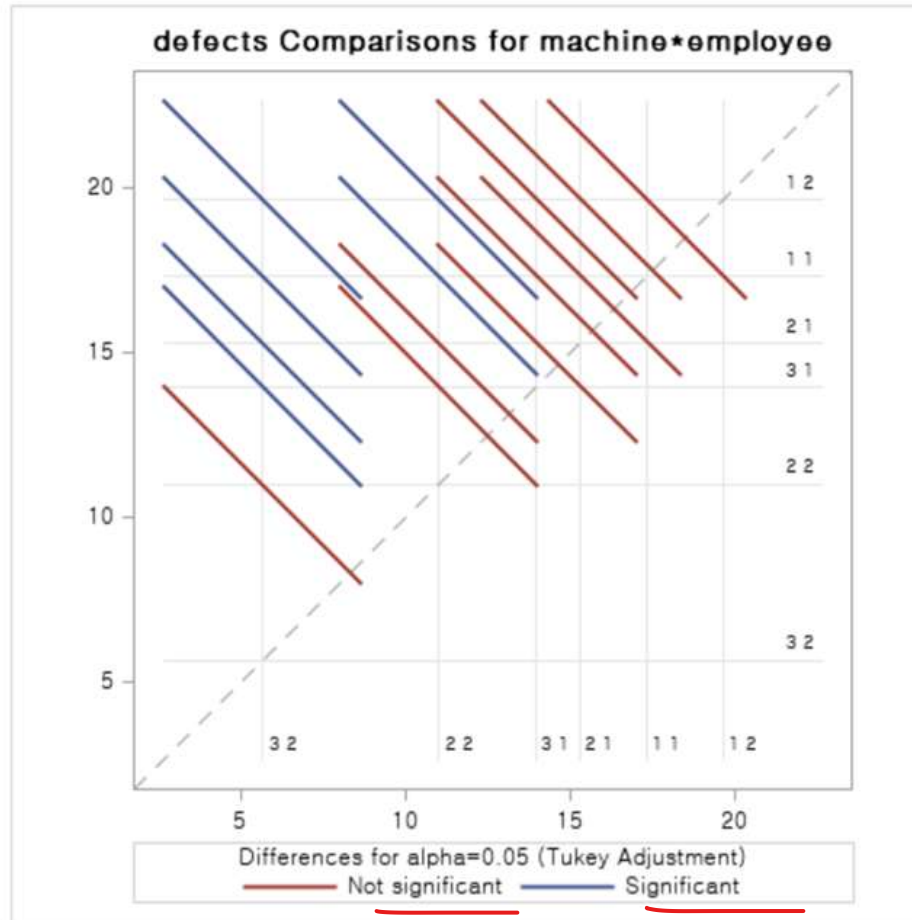
Dependent Variable: defects

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	

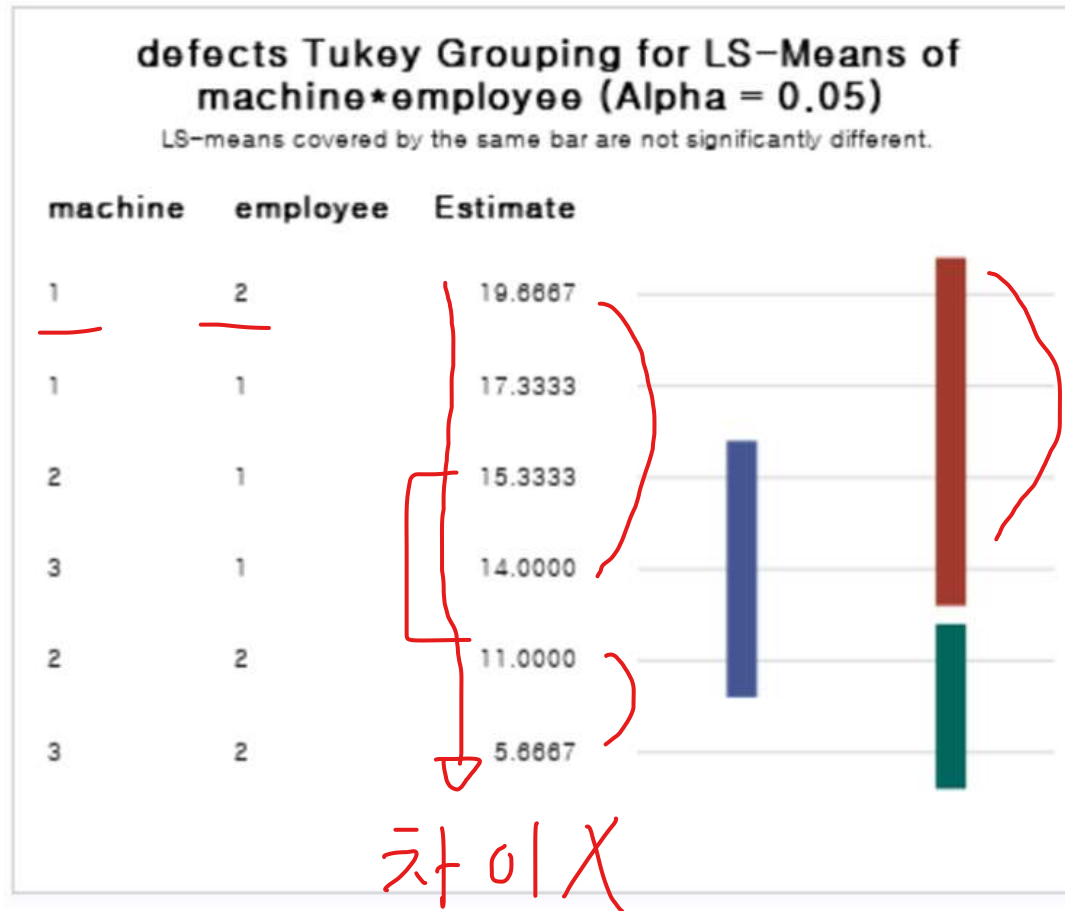
유의확률

다중비교

각 수준조합의 평균차에 대한 비교



다중비교



상호작용의 풀링(pooling)

오차항의 풀링

둘 이상의 주효과가 있는 경우(이원배치, 삼원배치, 다원배치)에는 상호작용을 먼저 검정하고 상호작용이 유의하지 않으면 상호작용의 제곱합 및 자유도를 각각 오차항의 제곱합 및 자유도에 합해 주는 것을 상호작용의 풀링(pooling)이라고 함.

이원배치인 경우 오차항의 풀링

요인	제곱합	자유도
A	SSA	$a - 1$
B	SSB	$b - 1$
A*B	SSAB	$(a - 1)(b - 1)$
오차	SSE	$ab(n - 1)$
전체	SST	$abn - 1$

→

요인	제곱합	자유도
A	SSA	$a - 1$
B	SSB	$b - 1$
오차	$SSE' = SSAB + SSE$	$abn - a - b + 1$
전체	SST	$abn - 1$

주효과의 검정은 새로 얻은 SSE' 에서 $F_0 = \frac{MSA}{MSE'}$ 와 $F_0 = \frac{MSB}{MSE'}$ 로 하며 각각 F 분포의 자유도는 $(a - 1, abn - a - b + 1)$ 및 $(b - 1, abn - a - b + 1)$ 이다.

상호작용의 풀링(pooling)

상호작용의 풀링은 반드시 해야 하는 것은 아니며, 학자들 사이에서도 풀링의 적용여부에 대한 의견은 다른 경우도 있다. ✓

다원배치법

- 삼원배치 및 그 이상의 요인이 있는 경우 제곱합의 분해나 자유도의 계산은 이원배치법에서 사용한 이론이 그대로 적용된다.
- 이론적으로 제곱합의 분해 등을 해보는 것도 의미있지만 실무적으로 계산은 SAS 등 통계 패키지를 사용하게 된다.

- 다원배치는 GLM 프로시저에서

MODEL Y = A B C A*B A*C B*C A*B*C;

와 같이 (이 경우 $Y = A|B|C$ 과 같음) MODEL 문에 모형에서 필요한 요인을 적절하게 구현하고

LSMEANS 이 MEANS 문을 사용하여 평균에 대한 추론을 하고,

임의효과가 있으면 RANDOM 문을 해당 효과(및 파생된 상호작용)에 적용하면 된다.

수고하셨습니다^^