

# 13장 | 분할구 설계

SAS를 이용한 실험 계획과 분산 분석 (자유아카데미)

# 실험 단위와 실험 차

- 실험 단위 (experimental unit): 처리수준/처리조합이 배정되는 최소의 단위
- 실험 오차 (experimental error): 실험 단위와 실험 단위 사이에 발생하는 측정 오차

예1) 12마리의 생쥐에 대해 안약 조율에 따른 치료효과를 비교한다고 하자. 서로 다른 안약을 안약 1, 안약 2라고 할 때 12마리 중 임의로 6마리를 선택하여 안약 1을, 나머지는 안약 2를 투여한다.

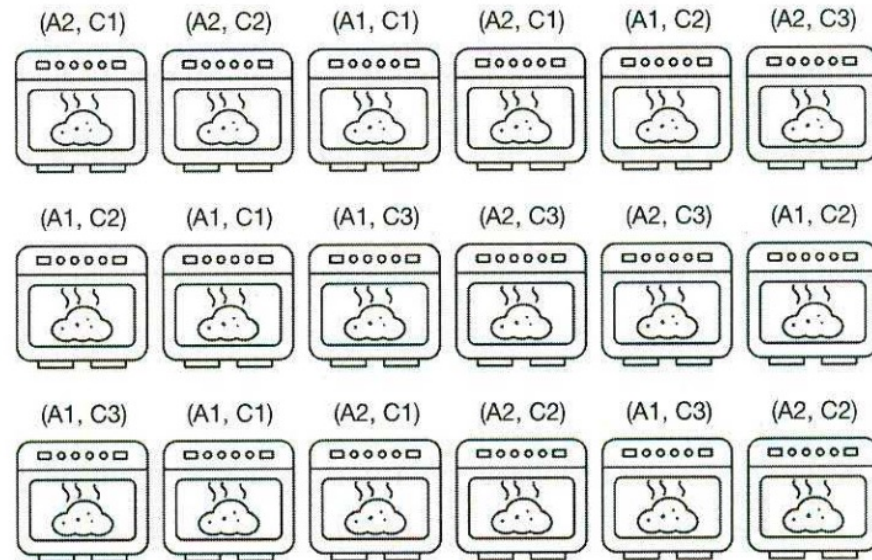
예2) 물고기가 10마리씩 들어있는 수조가 2개 있다고 가정하고 한 수조에는 먹이 1을 다른 수조에는 먹이 2를 주었다.

# 분할구 설계 (SPLIT-PLOT DESIGN)

- 실험에서 실험 조건을 실험단위에 완전히 랜덤 배치하는것이 실제로 불가능한 경우가 종종 발생
  - 온도 (A1,A2)와 밀가루 반죽 비율 (C1, C2, C3)을 달리하며 과자를 굽는 실험 (밀가루는 18개 과자를 만들 수 있는 분량)
  - 총  $2 \times 3 = 6$ 개의 처리조합이 발생

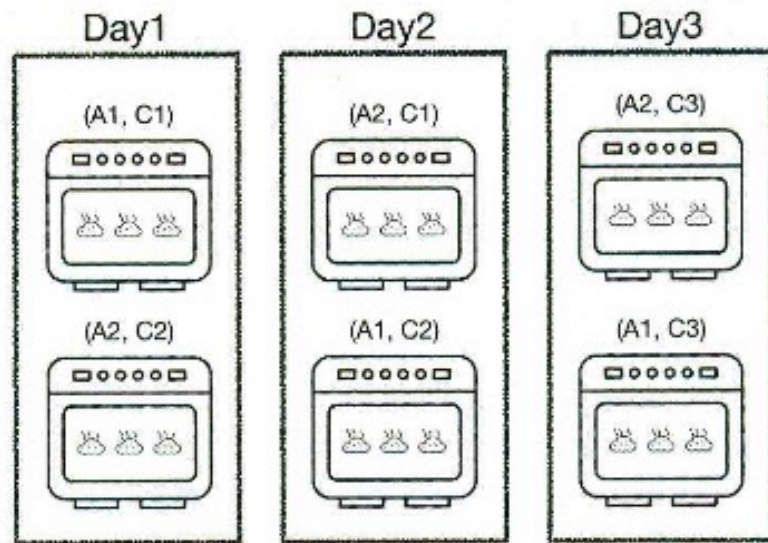


**Two-Way ANOVA**



# 분할구 설계 (SPLIT-PLOT DESIGN)

- 만약 물리적 시간적 제약으로 인해 하루에 오븐을 단 2회만 사용할 수 있다고 하면? 총 9일 소요
- 모든 실험을 3일만에 마칠 수 있는 실험 계획?
- 만일 (온도:A1, 반죽:C1) 처리조합에 대해 오븐일 1회 사용할 때 3개의 쿠키를 한꺼번에 굽게 되면 총 18개의 쿠키를 3일만에 구울 수 있다.



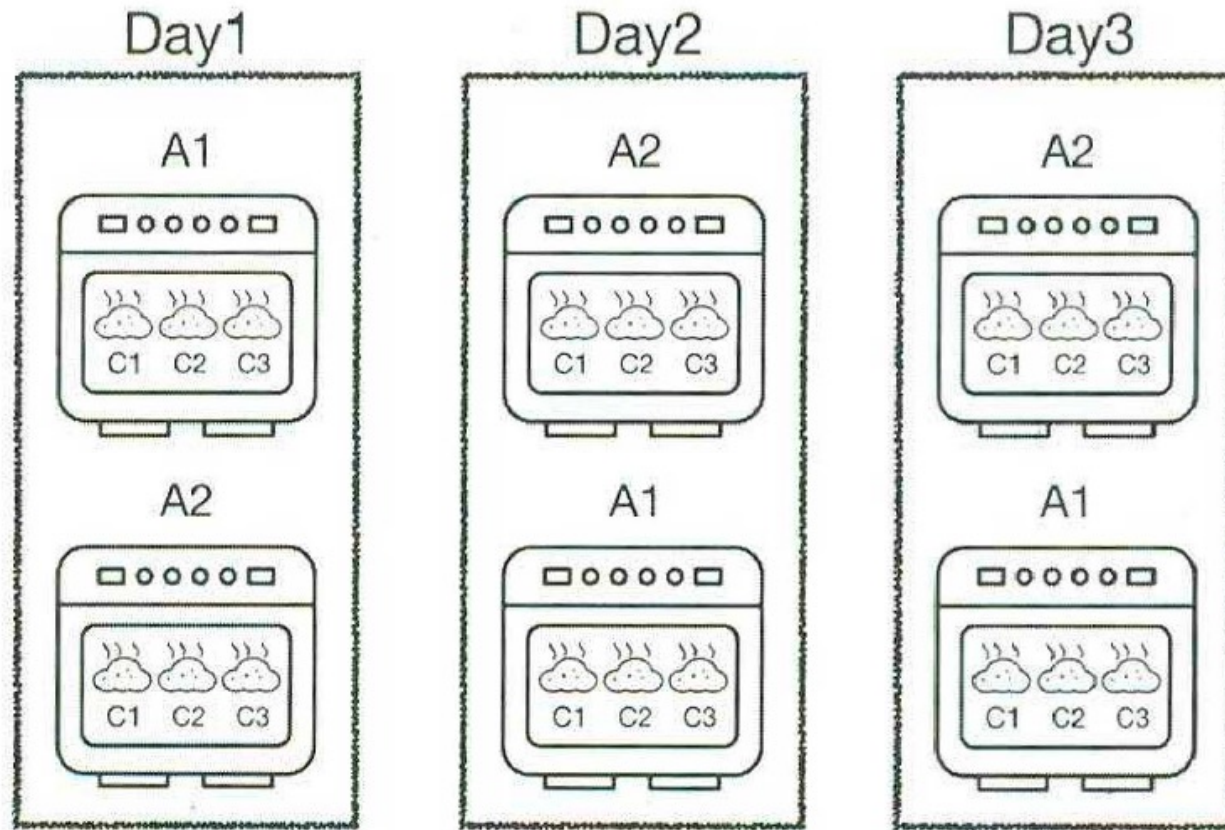
# 분할구 설계 (SPLIT-PLOT DESIGN)

- 만약 날씨에 따라 쿠키의 맛이 달라진다면?
- Day 1과 Day 3의 쿠키 맛이 달라진다. (반죽 차이 vs 날씨 차이)
- 하루에 두 번 오븐 사용
  - 한번은 온도를 A1으로 맞추고 다른 한번은 온도를 A2에 맞추
  - 오븐 안에 반죽의 비율을 달리한 3개의 쿠키 (C1, C2, C3)를 한꺼번에 굽는 실험



분할구 설계

# 분할구 설계 (SPLIT-PLOT DESIGN)



# 분할구 설계의 용어

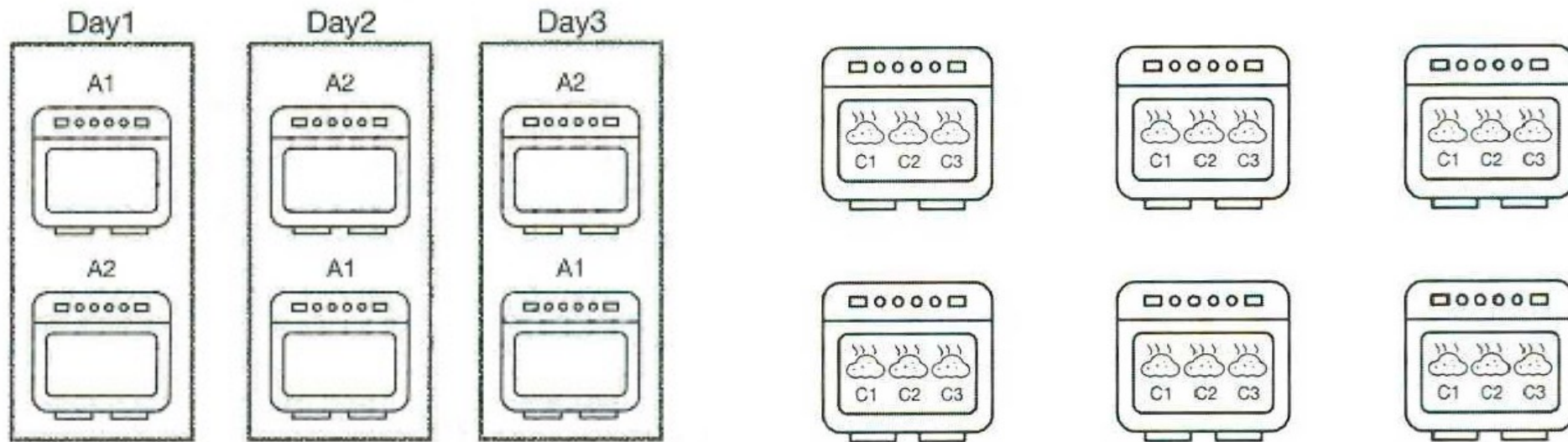
- 온도 ( $A_1, A_2$ )가 랜덤 배치 되는 실험단위는 ‘3개의 쿠키를 담은 오븐’  
→ 주구 (whole-plot)
- 반죽 ( $C_1, C_2, C_3$ )이 랜덤 배치 되는 실험단위는 ‘쿠키 한 개 분량의 밀가루 반죽’  
→ 세구 (subplot)
- 주구에 배정되는 온도 ( $A_1, A_2, A_3$ )  
→ 주구처리 (whole-plot treatment)
- 세구에 배정되는 반죽 ( $C_1, C_2, C_3$ )  
→ 세구처리 (subplot treatment)
- 온도에 대한 실험오차는 오븐과 오븐 차이에서 나옴  
→ 주구 오차 (whole-plot error)
- 반죽에 대한 실험 오차는 쿠키와 쿠키사이에서 발생  
→ 세구오차 (subplot error)

# 분할구 설계의 용어

- 주구 처리 (whole-plot treatment): 실험단위가 큰 ‘주구(whole-plot)’에 랜덤 배치되는 처리
- 주구 오차 (whole-plot error): 주구처리를 받은 실험단위 사이에 발생하는 실험 오차
- 세구 처리 (subplot treatment): 실험 단위가 작은 ‘세구(subplot)’에 랜덤 배치 되는 처리
- 세구 오차 (subplot error): 세구 처리를 받은 실험단위 사이에 발생하는 실험 오차



# SPLIT-PLOT DESIGN



기본적으로 RCBD의 형태를 띄고 있음

# SPLIT-PLOT DESIGN

Source	d.f.	SS
Day	3-1	SSday
온도	2-1	SStemp
Error	(3-1)(2-1)	SSday*temp
주구	$3 \times 2 - 1$	SSwhole

표 13.1: 주구(whole-plot)에 대한 분산분석표

Source	d.f.	SS
주구	$3 \times 2 - 1$	SSwhole
반죽비율	3-1	SSratio
Error	(6-1)(3-1)	SSwhole*ratio
Total	$3 \times 6 - 1$	SStotal

표 13.2: 세구(subplot)에 대한 분산분석표

기본적으로 RCBD의 형태를 띄고 있음

# ANOVA

Source	d.f.	SS	$F_0$
Day	3-1	SSday	
온도	2-1	SStemp	$\frac{MStemp}{MSday*temp}$
Whole-plot Error	(3-1)(2-1)	SSday*temp	
반죽비율	3-1	SSratio	$\frac{MSratio}{MSE}$
온도*반죽비율	(2-1)(3-1)	SStemp*ratio	$\frac{MStemp*ratio}{MSE}$
Subplot Error	2(3-1)(3-1)	SSE	
Total	$3 \times 6 - 1$	SStotal	

표 13.3: 분할구설계(split-plot design)에 대한 분산분석표

## 모형식

$$y_{ijk} = \mu + \rho_i + \tau_j + (\rho\tau)_{ij} + \gamma_k + (\tau\gamma)_{jk} + \epsilon_{ijk}$$

$$\sum_{i=1}^3 \rho_i = 0, \quad \sum_{j=1}^2 \tau_j = 0, \quad \sum_{k=1}^3 \gamma_k = 0,$$

$$\sum_{i=1}^3 (\rho\tau)_{ij} = \sum_{j=1}^2 (\rho\tau)_{ij} = 0,$$

$$\sum_{j=1}^2 (\tau\gamma)_{jk} = \sum_{k=1}^3 (\tau\gamma)_{jk} = 0,$$

$$\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2),$$

# SAS 예시

종이의 인장력에 펄프 가열온도가 영향을 주는지 조사하기 위하여 다음과 같은 실험을 했다. 실험은 3일에 걸쳐 실시하는데 하루에 할 수 있는 실험의 수는 총 12 회이다. 매일 펄프를 만드는 방법을 달리해 가면서 3 종류의 펄프(펄프1, 펄프2, 펄프3)를 준비한 후, 준비된 펄프를 사 등분 하여 각각 200°C, 225°C, 250°C, 275°C 로 가열하여 종이를 생산하도록 한다. 표 13.4는 펄프종류와 가열온도에 따른 종이의 인장력이다<sup>2</sup>.

온도	Day1			Day2			Day2		
	펄프1	펄프2	펄프3	펄프1	펄프2	펄프3	펄프1	펄프2	펄프3
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

표 13.4: 펄프 제조방법과 가열온도에 따른 종이의 인장력 자료

# SAS CODE

```
proc glm data=a;  
  
    model y = day pulp day*pulp temp pulp*temp;  
  
    random day day*pulp /test;  
  
run;
```

Source	Type III Expected Mean Square
day	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{day*pulp}) + 12 \text{Var}(\text{day})$
pulp	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{day*pulp}) + Q(\text{pulp}, \text{pulp*temp})$
day*pulp	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{day*pulp})$
temp	$\text{Var}(\text{Error}) + Q(\text{temp}, \text{pulp*temp})$
pulp*temp	$\text{Var}(\text{Error}) + Q(\text{pulp*temp})$



## SAS 결과

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	day	2	77.555556	38.777778	4.28	0.1016
*	pulp	2	128.388889	64.194444	7.08	0.0485
	Error: MS(day*pulp)	4	36.277778	9.069444		

\* This test assumes one or more other fixed effects are zero.

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	day*pulp	4	36.277778	9.069444	2.28	0.1003
*	temp	3	434.083333	144.694444	36.43	<.0001
	pulp*temp	6	75.166667	12.527778	3.15	0.0271
	Error: MS(Error)	18	71.500000	3.972222		

\* This test assumes one or more other fixed effects are zero.

## SAS 결과 응용

Source	d.f.	SS	MS	$F_0$	p-value
Day	2	77.555	38.777		
펄프	2	128.388	64.194	7.08	0.0485
주구오차	4	36.277	9.069		
온도	3	434.083	144.694	36.43	< 0.0001
펄프*온도	6	75.166	12.527	3.15	0.027
세구오차	18	71.500	3.972		
Total	35	822.972			



# SAS CODE & OUTPUT

```
proc glm data=a;  
  class car tire ;  
  model wear = car tire ;  
  
run;
```

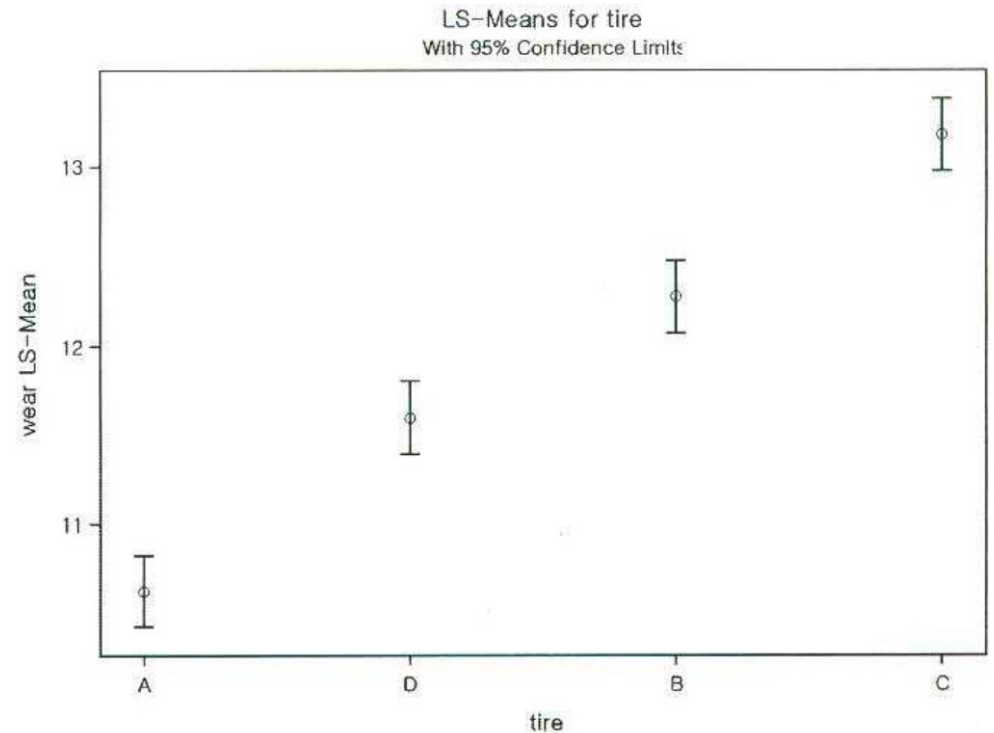
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	14.19375000	2.36562500	73.26	<.0001
Error	9	0.29062500	0.03229167		
Corrected Total	15	14.48437500			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
car	3	0.27187500	0.09062500	2.81	0.1005
tire	3	13.92187500	4.64062500	143.71	<.0001

# SAS CODE & OUTPUT

```
proc glm data=a plots=( diffplot(center)
                        meanplot(cl ascending) );

  class car tire ;
  model wear = car tire ;
  lsmeans tire / adjust=tukey lines ;
run;
```



# RCBD 안의 이원배치법

예를 들어 타이어 마모도에 대한 실험에서 브랜드(Brand)로는 ‘외국산(foreign)’과 ‘국산(domestic)’을, 타이어 종류(Type)로는 ‘사계절(all season)’과 ‘겨울용(winter)’을 비교하려고 한다. 따라서 전체 처리조합은  $2 \times 2 = 4$  개가 되고 4대의 자동차에 4 개의 처리조합을 랜덤하게 배치함으로써 RCBD 를 수행하였다. 주행 후 마모도를 측정한 결과 표 10.7을 얻게 되었다.

Brand	Type	자동차1	자동차2	자동차3	자동차4
domestic	all	10.4	10.9	10.5	10.7
domestic	winter	12.4	12.4	12.3	12.0
foreign	all	11.8	11.8	11.4	11.4
foreign	winter	13.1	13.4	12.9	13.3

# 모형식

$$y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk},$$

$$i = 1, 2, 3, 4, \quad j = 1, 2, \quad k = 1, 2,$$

$$\sum_{i=1}^4 \rho_i = 0, \quad \sum_{j=1}^2 \alpha_j = 0, \quad \sum_{k=1}^2 \beta_k = 0,$$

$$\sum_{j=1}^2 (\alpha\beta)_{jk} = 0, \quad \sum_{k=1}^2 (\alpha\beta)_{jk} = 0,$$

$$\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$$

여기서  $\rho_i$ 는 블록효과이고  $\alpha_j$ 는 브랜드(외국산, 국산)효과,  $\beta_k$ 는 타이어 종류(사계절, 겨울용)효과를 의미한다. 블록과 처리 간 상호작용은 RCBD 정의상 존재하지 않는데, 이는 자동차 내 브랜드나 타이어 종류의 차이가 자동차가 바뀌어도 일정하게 유지된다는 의미로 해석된다.  $(\alpha\beta)_{jk}$ 는 브랜드와 타이어의 상호작용효과

# RCBD 제곱합에 대한 간편한 공식

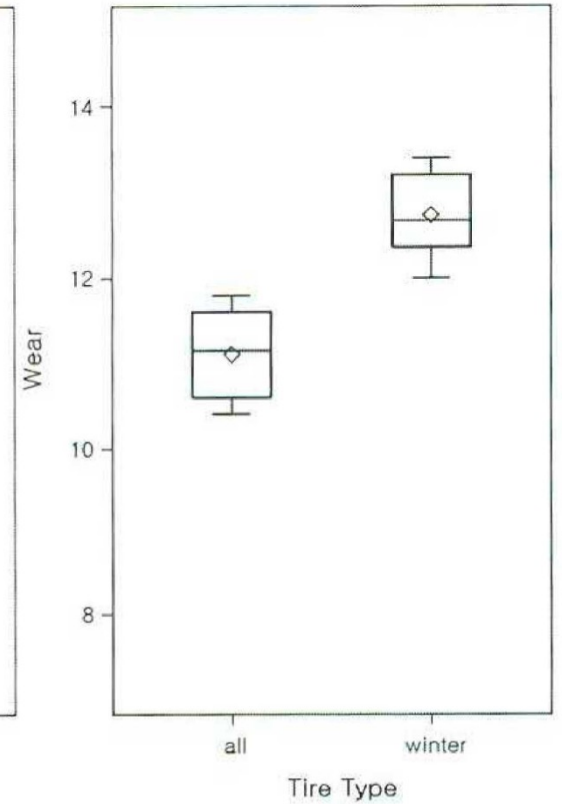
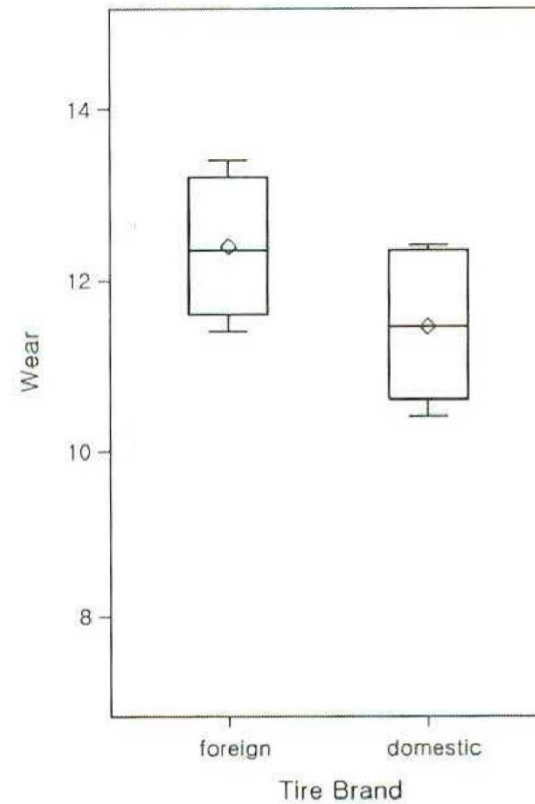
Source	d.f.	SS
Car	$4 - 1$	$\sum_{i=1}^4 \frac{y_{i..}^2}{4} - CT, \quad CT = \frac{y_{...}^2}{16}$
Brand	$2 - 1$	$\sum_{j=1}^2 \frac{y_{.j.}^2}{8} - CT$
Type	$2 - 1$	$\sum_{k=1}^2 \frac{y_{..k}^2}{8} - CT$
Brand*Type	$(2-1)(2-1)$	$\sum_{j=1}^2 \sum_{k=1}^2 \frac{y_{.jk}^2}{4} - \sum_{j=1}^2 \frac{y_{.j.}^2}{8} - \sum_{k=1}^2 \frac{y_{..k}^2}{8} + CT$
Error	$3+3+3$	$SST - SS_{car} - SS_{brand} - SS_{type} - SS_{brand*type}$
Total	$16 - 1$	$\sum_{i=1}^4 \sum_{j=1}^2 \sum_{k=1}^2 y_{ijk}^2 - CT$

표 10.8: RCBD 안의 이원배치법에 대한 자유도와 제곱합

# SAS CODE & OUTPUT

```
proc glm data=a;  
  class car brand type ;  
  model wear = car brand type brand*type ;  
run;
```

Source	d.f.	SS	MS	$F_0$	p-value
Car	3	0.2718	0.0906	2.81	0.1005
Brand	1	3.5156	3.5156	108.87	< 0.0001
Type	1	10.4006	10.4006	322.08	< 0.0001
Brand*Type	1	0.0056	0.0056	0.17	0.6862
Error	9	0.2906	0.0322		
Total	15	14.484			



# SUMMARY

- RCBD 는 블록안에서 랜덤화
- 블록을 이용해서 외생변수의 영향을 제어함
- 블록과 블록사이의 SSblock 클수록 SSE 줄어듦
- 블록 수가 증가하면 SSE 의 d.f. 줄어듦
- SSE 는 작을수록, d.f. 는 클수록 좋음
- 적은 블록 수+효과적인 블록킹(blocking) 요구