



# Ch08. 분산분석(ANOVA)2

---



# Two-Way ANOVA

## ■ 문제의 정의

- K치킨회사의 연구부장인 이 대리는 맛있는 치킨을 만드는 연구를 하고 있다.
- 치킨의 맛을 결정하는 두 가지 요인은 튀길 때의 기름 온도와 튀기는 시간이다.
- 이 두 가지 요인이 정말로 치킨의 맛에 영향을 미치는가를 조사하기 위하여 다음과 같은 실험을 하였다.
- 튀길 때의 온도를 1,000도와 1,500도로 하고, 튀기는 시간을 10분과 20분으로 하여 치킨을 튀긴 후에 사람들에게 맛을 평가하도록 하였다.
- 과연 온도와 시간이 맛을 결정하는데 중요한 요인인가? 이 두 가지 요인들 간의 상호작용효과는 없었는가?
- (6.TWA.sav).

# Two-Way ANOVA

## ■ 가설1

- 귀무가설(  $H_0$  ): 튀기는 온도에 따라 맛은 차이가 없다.

$$H_0 : \mu_1 = \mu_2$$

- 연구가설(  $H_1$  ): 온도에 따라 맛은 차이가 있다.

$$H_1 : \mu_1 \neq \mu_2$$

## ■ 가설2

- 귀무가설(  $H_0$  ): 튀기는 시간에 따라 맛은 차이가 없다.

$$H_0 : \mu_1 = \mu_2$$

- 연구가설(  $H_1$  ): 튀기는 시간에 따라 맛은 차이가 있다.

$$H_1 : \mu_1 \neq \mu_2$$

## ■ 가설3=가장중요함.

- 귀무가설(  $H_0$  ): 튀기는 온도와 시간사이에는 상호작용이 없다.

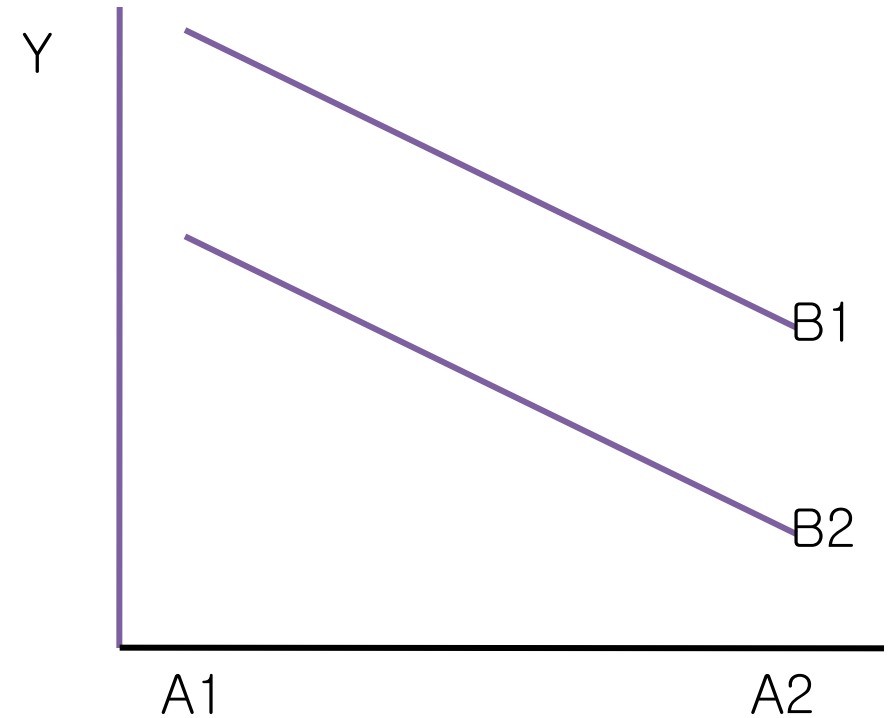
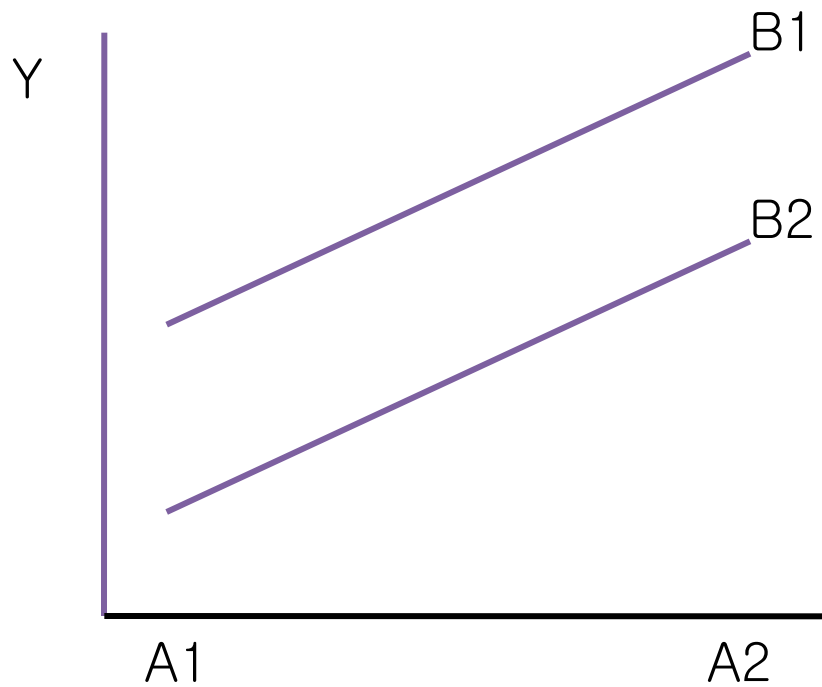
$$H_0 : \alpha\beta_{ij} = 0$$

- 연구가설(  $H_1$  ): 튀기는 온도와 시간사이에는 상호작용이 있다.

$$H_1 : \alpha\beta_{ij} \neq 0$$

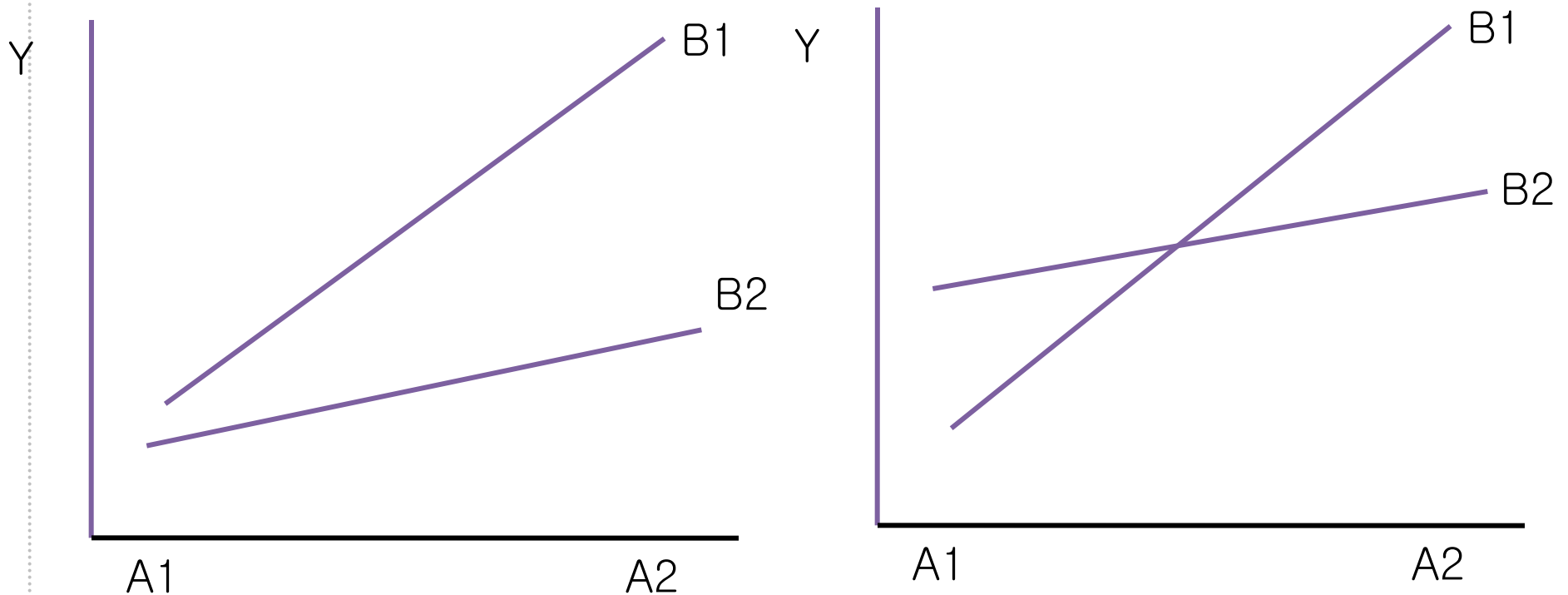
# Two-Way ANOVA

- 평균반응 프로파일(average response profile)
  - 상호작용 없는 경우



# Two-Way ANOVA

- 평균반응 프로파일(average response profile)
  - 상호작용 있는 경우





# Two-Way ANOVA

요인 A \ 요인 B		온도								평균	
		1,000도				1,500도					
시간	10분	90	88	88	91	86	91	88	87	88	90.73
	20분	90	91	88	87	91	86	88	88	92	
평균		90.79				89.97					

# Two-Way ANOVA

- ##### 이원 분산분석(Two-way ANOVA) #####
- # 01.데이터 불러오기
- twaData <- read.csv("./data/06.TWA.csv",
- header=TRUE,
- na.strings = ".")
- )
- twaData\$time <- factor(twaData\$time,
- levels=c(1:2),
- labels=c("10분","20분"))
- twaData\$temp <- factor(twaData\$temp,
- levels=c(1:2),
- labels=c("1,000도","1,500도"))
- str(twaData)

# Two-Way ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ↗

```
> ##### 이원 분산분석(Two-way ANOVA) #####
>
> # 01.데이터 불러오기
> twaData <- read.csv("./data/06.TWA.csv",
+                      header=TRUE,
+                      na.strings = ".")
> )
> twaData$time <- factor(twaData$time,
+                        levels=c(1:2),
+                        labels=c("10분", "20분"))
> twaData$temp <- factor(twaData$temp,
+                        levels=c(1:2),
+                        labels=c("1,000도", "1,500도"))
>
> str(twaData)
'data.frame': 60 obs. of 3 variables:
 $ time : Factor w/ 2 levels "10분","20분": 1 1 1 1 1 1 1 1 1 1 ...
 $ temp : Factor w/ 2 levels "1,000도","1,500도": 2 2 2 2 2 2 2 2 2 2 ...
 $ taste: int 95 93 94 98 97 94 96 92 91 94 ...
> |
```



# Two-Way ANOVA

- `attach(twaData)` # 객체연결 (`attach` -> `detach`)
- # 02.기본통계치 확인: `describe(psych패키지 이용)`
- `library(psych)`
- `describeBy(taste, time:temp, mat=T)` #그룹핑 변수가 2개여서 2개 들어옴.

# Two-Way ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ↗

```
> attach(twaData) # 객체연결 (attach -> detach)
>
> # 02.기본통계치 확인: describe(psych패키지 이용)
> library(psych)
> describeBy(taste, time:temp, mat=T)
```

	item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	ra
X11	1	10분:1,000도	1	14	87.35714	1.598420	87.5	87.41667	2.2239	85	89	
X12	2	10분:1,500도	1	16	93.68750	2.982588	94.0	93.85714	2.2239	87	98	
X13	3	20분:1,000도	1	15	94.00000	2.299068	94.0	94.00000	2.9652	90	98	
X14	4	20분:1,500도	1	15	86.00000	1.195229	86.0	86.00000	1.4826	84	88	
	skew		kurtosis		se							
X11	-0.33624392		-1.5466792		0.4271956							
X12	-0.84663287		-0.1256755		0.7456471							
X13	-0.09874741		-1.0767470		0.5936168							
X14	-0.23426481		-1.1706667		0.3086067							

```
> |
```

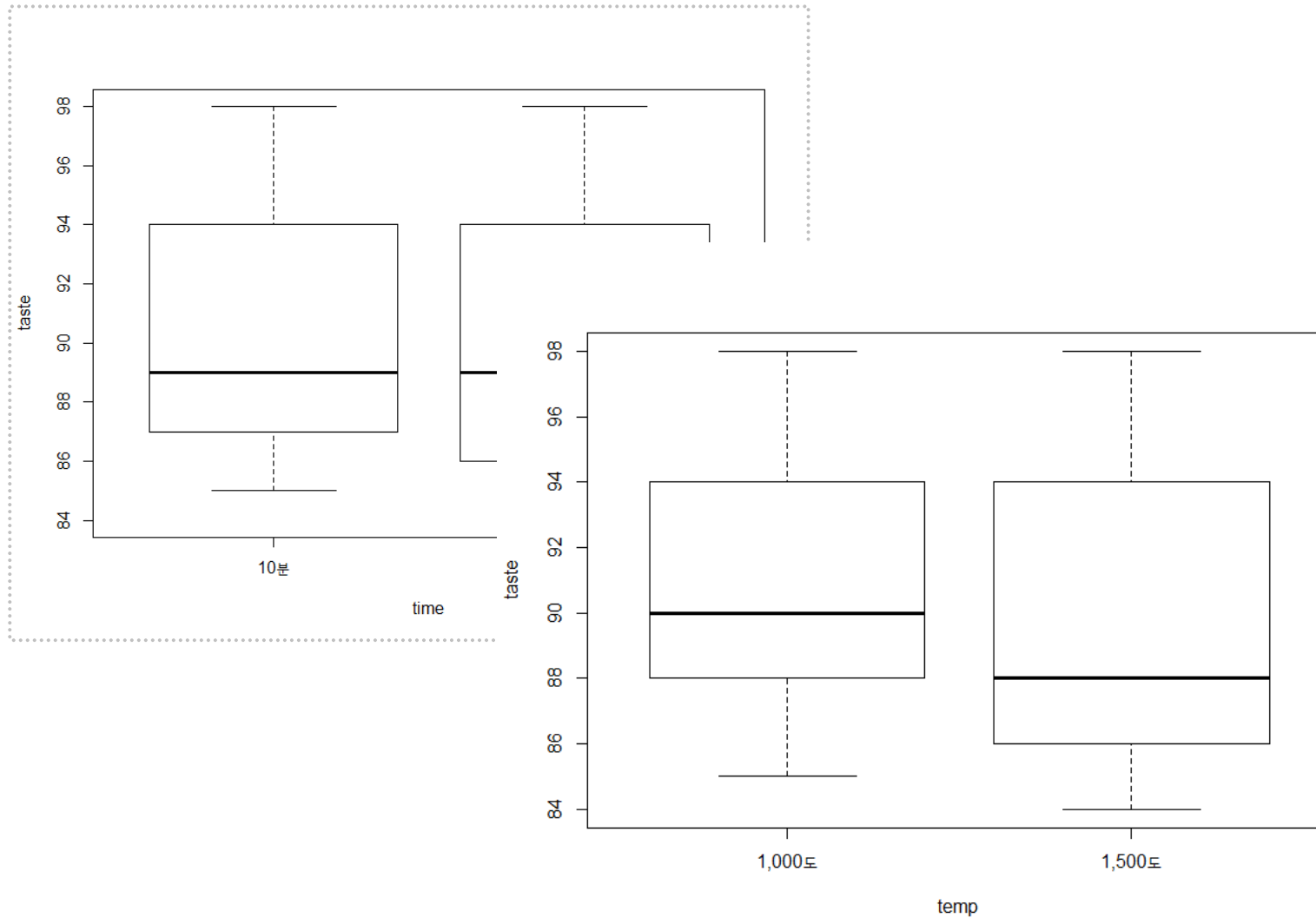
# Two-Way ANOVA

- # 03.그래프 그리기(박스그래프,히스토그램)
- `plot(taste ~ time + temp, data=twaData)`#별로 2개 찍기
- `boxplot(taste ~ time*temp, data=twaData)`#한눈에 보여줘 = 시간과 온도 모두의 상호작용이 있는지를 알아보는 과정

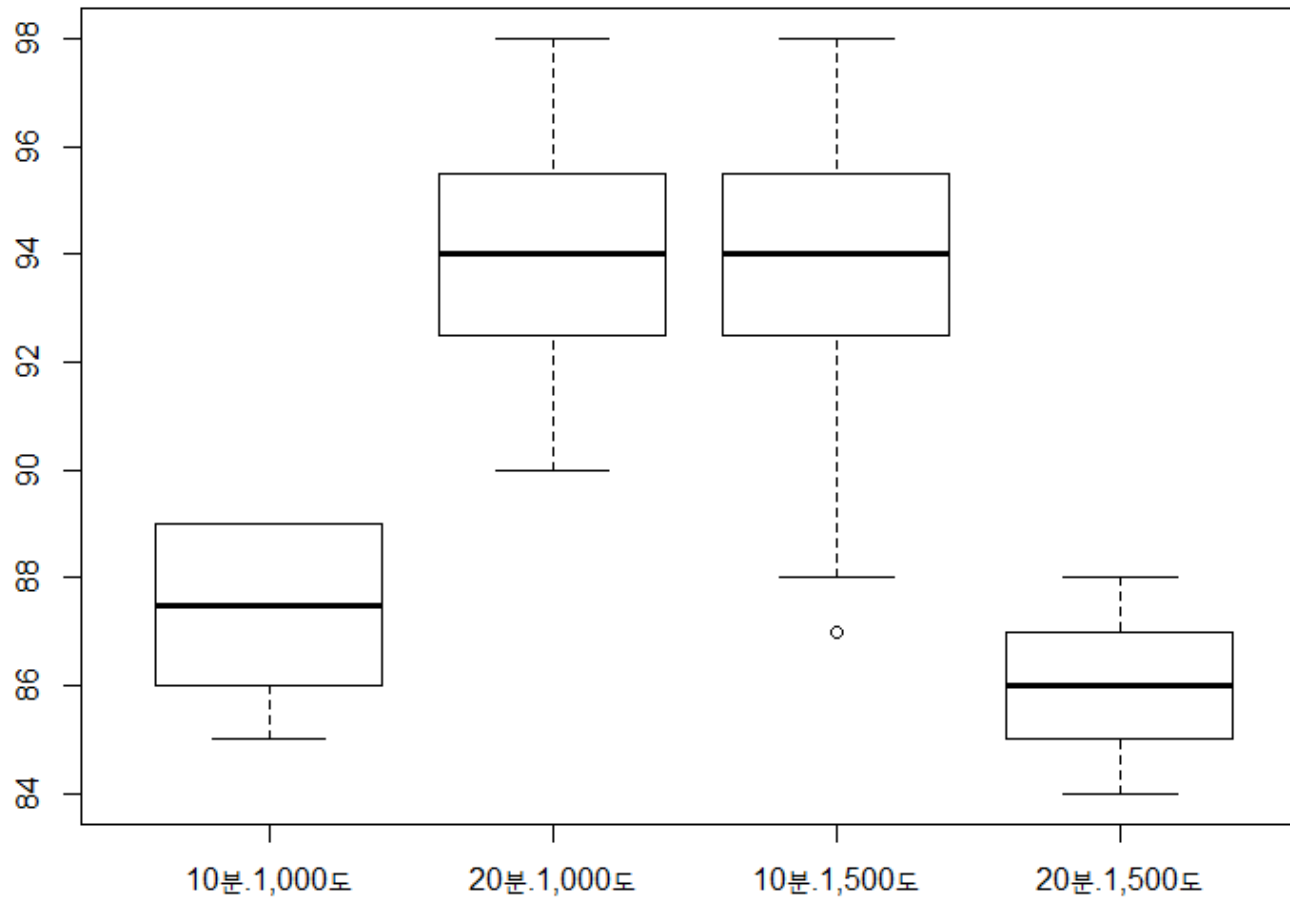
- `library(ggplot2)`
- `ggplot(twaData, aes(x = temp, y = taste)) +`
- `geom_boxplot(outlier.colour="red") +`
- `facet_wrap(~time) +`
- `ggtitle("시간*온도에 따른 맛")`

잘 안됨

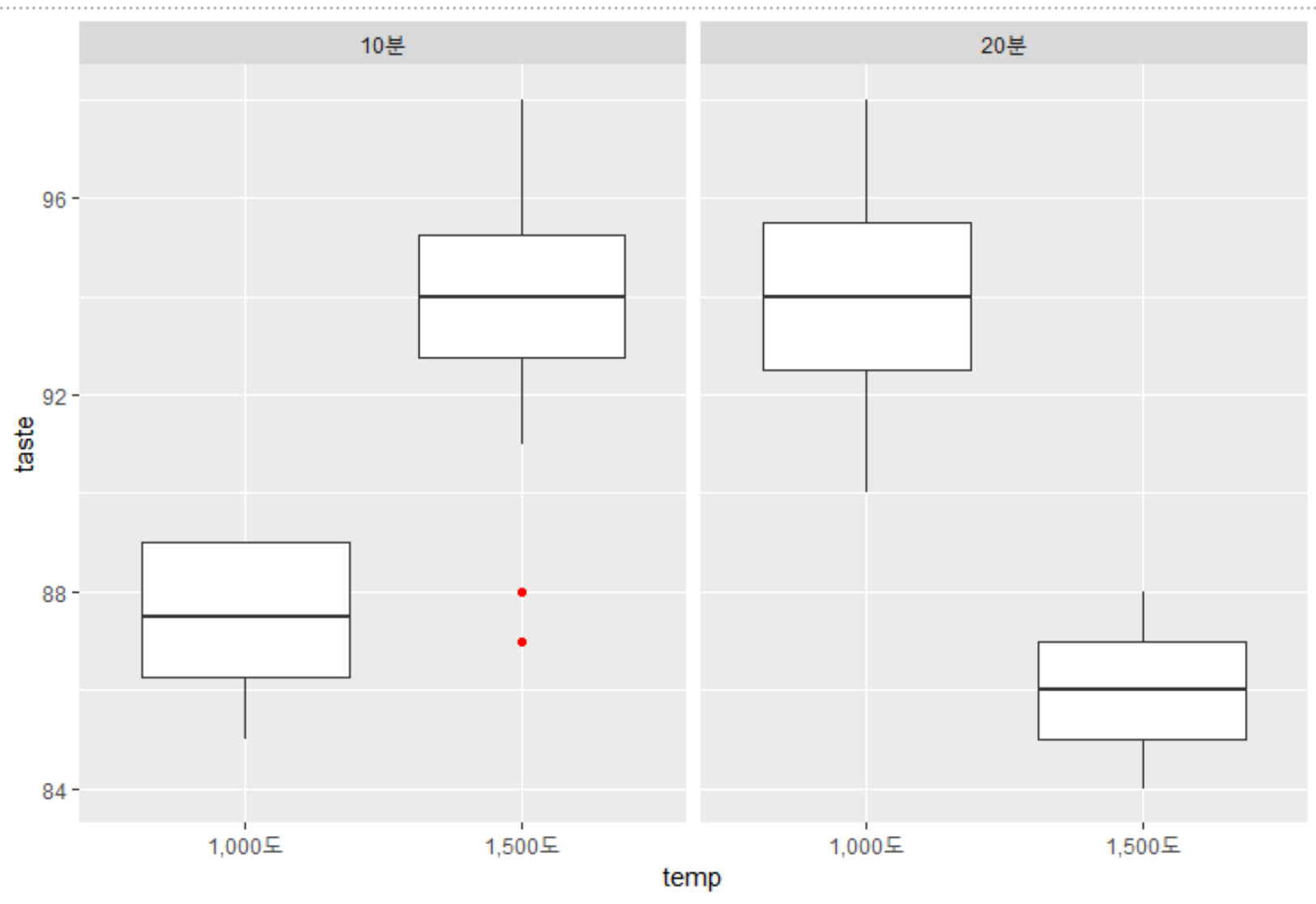
# Two-Way ANOVA



# Two-Way ANOVA



# Two-Way ANOVA



# Two-Way ANOVA

- # 04.통계분석
- `twaResult <- aov(taste ~ time + temp + time:temp, data=twaData)`
- `summary(twaResult)`
- #상호작용효과 그래프
- `interaction.plot(time, temp, taste)`

# Two-Way ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/

```
> # 04.통계분석
> twaResult <- aov(taste ~ time + temp + time:temp, data=twaData)
> summary(twaResult)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
time	1	8.1	8.1	1.733	0.193
temp	1	10.8	10.8	2.327	0.133
time:temp	1	768.4	768.4	165.084	<2e-16 ***
Residuals	56	260.7	4.7		

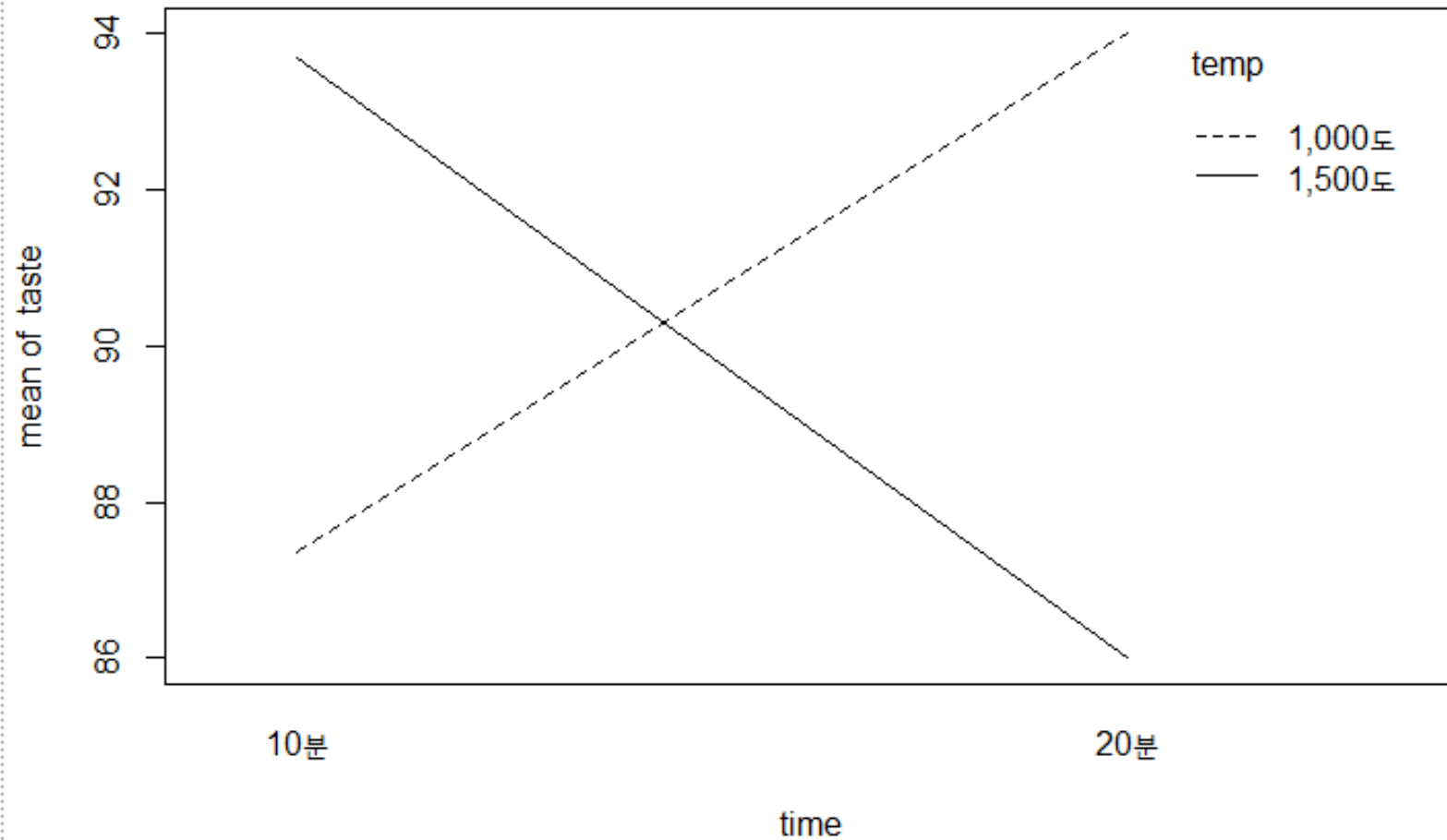
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

무조건 time:temp부터 해석해서 위의 time과 temp를 해석한다.

상호작용 컴비네이션을 살펴보고 나서 <0.05이기 때문에 차이가 있는 상호작용이 있었다.따라서 time과 temp를 동시에 해석한다.



# Two-Way ANOVA



# Two-Way ANOVA

- # 사후검정(Multicamparison test )
- # 상호작용이 있을 경우 : **그룹별로 나누어서 분석**
- # 상호작용이 없을 경우 : 각 변수별 주효과 분석(t-test, ANOVA분석)#굳이 하지마라#미리 구한것이 있으므로

```
> twaResult <- aov(taste ~ time + temp + time:temp, data=twaData)
> summary(twaResult)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
time	1	8.1	8.1	1.733	0.193
temp	1	10.8	10.8	2.327	0.133
time:temp	1	768.4	768.4	165.084	<2e-16 ***
Residuals	56	260.7	4.7		

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

무조건 time:temp부터  
해석해서 위의 time과  
temp를 해석한다.

- # 상호작용이 있을 경우 : 그룹별로 나누어서 분석
- `tw10 <- twaData[twaData$time=="10분",]`
- `tw20 <- twaData[twaData$time=="20분",]`



# Two-Way ANOVA

- `t.test(taste ~ temp,`
  - `data=tw10,`
  - `alternative = c("two.sided"),`
  - `var.equal = TRUE,`
  - `conf.level = 0.95)`
- `t.test(taste ~ temp,`
  - `data=tw20,`
  - `alternative = c("two.sided"),`
  - `var.equal = TRUE,`
  - `conf.level = 0.95)`
- `detach(twaData)`

# Two-Way ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ↗

## Two Sample t-test

```
data: taste by temp
t = -7.0903, df = 28, p-value = 1.029e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.159210 -4.501504
sample estimates:
mean in group 1,000도 mean in group 1,500도
      87.35714          93.68750
```

```
> t.test(taste ~ temp,
+       data=tw20,
+       alternative = c("two.sided"),
+       var.equal = TRUE,
+       conf.level = 0.95)
```

## Two Sample t-test

```
data: taste by temp
t = 11.957, df = 28, p-value = 1.62e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  6.629527 9.370473
sample estimates:
mean in group 1,000도 mean in group 1,500도
      94          86
```

# Two-Way ANOVA

- 통닭을 튀기는 시간과 온도에 따라 맛이 차이가 있는지를 분석한 결과, 튀기는 시간과 온도에 따른 상호작용효과가 있었다( $F = 165.084$ ,  $p = 0.000$ ). 튀기는 시간이 10분일 경우에는 온도가 1500도 일 경우에 맛이 제일 좋았으며, 20분일 경우에는 1,000도로 튀겼을 경우에 맛이 좋았다.

		온도			F	P
		1,000도	1,500도			
시간	10분	87.36	93.69	시간	1.733	0.193
				온도	2.327	0.133
	20분	94.00	86.00	시간 * 온도	165.084	0.000

# Two-Way Repeated Measures ANOVA

- 문제의 정의 = **변수 2개중 하나가 시간을 나타냄.**
  - K 병원에서는 이번에 새롭게 아로마테라피 치료를 개발하였다.
  - 이 치료가 통증에 효과가 있는지를 검증하기 위해
    - 새롭게 개발한 치료제로 향기요법을 처치 받는 실험군과
    - 일반 향기치료제로 가짜 향기요법을 처치 받는 대조군을 나누고,
    - 치료전과 후에 통증이 차이가 있는지를 검증하였다.
  - 7.TWRMA.sav
- 가설3
  - 귀무가설(  $H_0$  ): 실험군과 대조군은 시점에 따라서 통증의 차이가 없다.
  - 연구가설(  $H_1$  ): 실험군과 대조군은 시점에 따라서 통증의 차이가 있다.

#여기서의 시간은 쌍으로 묶여 있어서 자신을 2번 측정하는것.



## Two-Way Repeated Measures ANOVA

- 문제의 정의

- K 병원에서는 이번에 새롭게 아로마테라피 치료를 개발하였다.
- 이 치료가 통증에 효과가 있는지를 검증하기 위해
  - 새롭게 개발한 치료제로 향기요법을 처치 받는 실험군과
  - 일반 향기치료제로 가짜 향기요법을 처치 받는 대조군을 나누고,
  - 치료전과 후에 통증이 차이가 있는지를 검증하였다.
- 7.TWRMA.sav

- 가설3

- 귀무가설( $H_0$ ): 실험군과 대조군은 시점에 따라서 통증의 차이가 없다.
- 연구가설( $H_1$ ): 실험군과 대조군은 시점에 따라서 통증의 차이가 있다.

# Two-Way Repeated Measures ANOVA

## ■ 이원 반복측정 분산분석

		요인A (시점)		평균
요인 B	개체	치료전	치료후	
대조군	1	12.32	23.09	45.34
	2	12.32	23.09	
	3	16.62	44.97	
	4	6.3	40.78	
	...	...	...	
실험군	1	13.54	40.55	42.35
	2	15.33	33.70	
	3	15.95	35.42	
	4	19.35	33.12	
	...	...	...	
평균		46.12	39.71	



# Two-Way Repeated Measures ANOVA

```
## 이원 반복측정 분산분석(Two-way Repeated Measures ANOVA) ##  
twrmaData <- read.csv("./data/07.twrma.csv",  
                      header=TRUE,  
                      na.strings = ".")  
twrmaData$group <- factor(twrmaData$group,  
                          levels=c(1,2),  
                          labels=c("대조군","실험군"))  
twrmaData$time <- factor(twrmaData$time,  
                         levels=c(1:2),  
                         labels=c("사전","사후"))  
str(twrmaData)
```

# Two-Way Repeated Measures ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ➤

```
> ##### 이원 반복측정 분산분석(Two-way Repeated Measures ANOVA) #####
> twrmaData <- read.csv("./data/07.twrma.csv",
+                       header=TRUE,
+                       na.strings = ".")
> twrmaData$group <- factor(twrmaData$group,
+                           levels=c(1,2),
+                           labels=c("대조군", "실험군"))
> twrmaData$time <- factor(twrmaData$time,
+                           levels=c(1:2),
+                           labels=c("사전", "사후"))
> str(twrmaData)
'data.frame': 284 obs. of 4 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ group   : Factor w/ 2 levels "대조군","실험군": 1 1 1 1 1 1 1 1 1 1 ...
 $ time    : Factor w/ 2 levels "사전","사후": 1 1 1 1 1 1 1 1 1 1 ...
 $ paincore: num 12.3 12.3 16.6 18.6 20.1 ...
> |
```



# Two-Way Repeated Measures ANOVA

- `attach(twrmaData) # 객체연결 (attach -> detach)`
- `# 02.기본통계치 확인: describe(psych패키지 이용)`
- `library(psych)`
- **describeBy**(painscore, time:group, mat=T)

# Two-Way Repeated Measures ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ↗

```
> # 02.기본통계치 확인: describe(psych패키지 이용)
> library(psych)
> describeBy(painscore, time:group, mat=T)
```

	item	group1	vars	n	mean	sd	median	trimmed	mad	min
X11	1	사전:대조군	1	71	45.79549	22.714636	43.37	42.98000	12.009060	12.32
X12	2	사전:실험군	1	71	45.28268	22.459870	44.48	43.22351	19.659276	5.08
X13	3	사후:대조군	1	71	45.05972	16.620947	43.56	43.86491	12.335232	12.57
X14	4	사후:실험군	1	71	26.47352	6.932682	26.15	26.73368	7.694694	8.62
	max	range	skew	kurtosis	se					
X11	146.44	134.12	1.8821859	5.2924556	2.695731					
X12	131.24	126.16	1.2314591	2.5827325	2.665496					
X13	104.46	91.89	0.8352065	1.4257983	1.972544					
X14	37.94	29.32	-0.3038318	-0.4139521	0.822758					

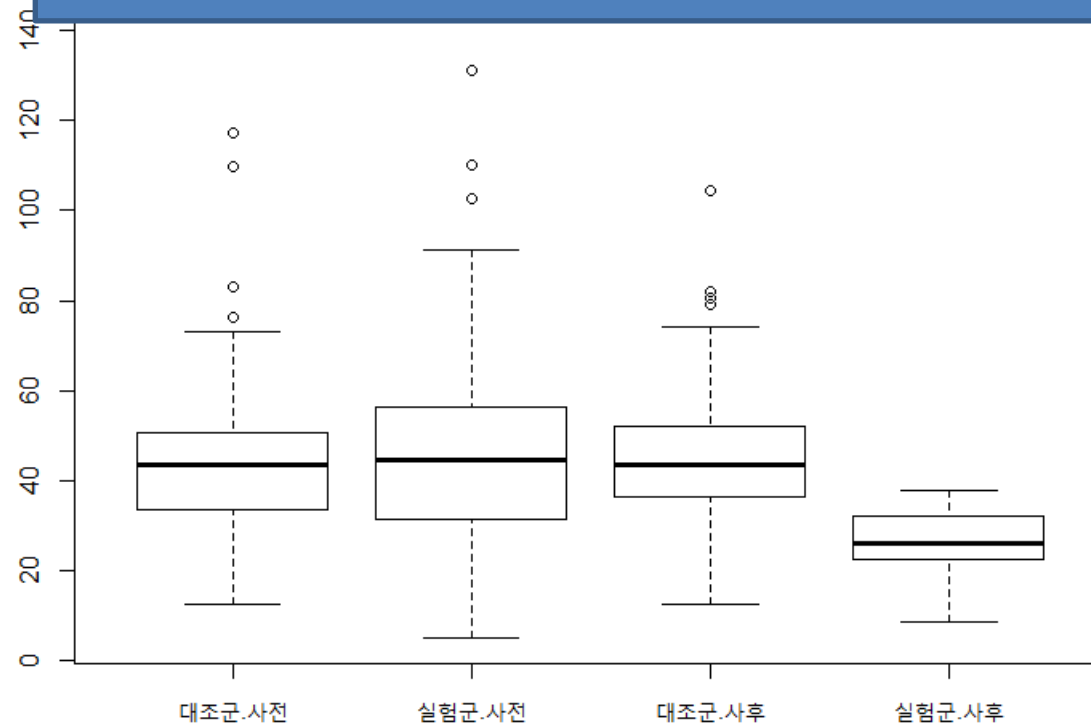
```
> |
```

# Two-Way Repeated Measures ANOVA

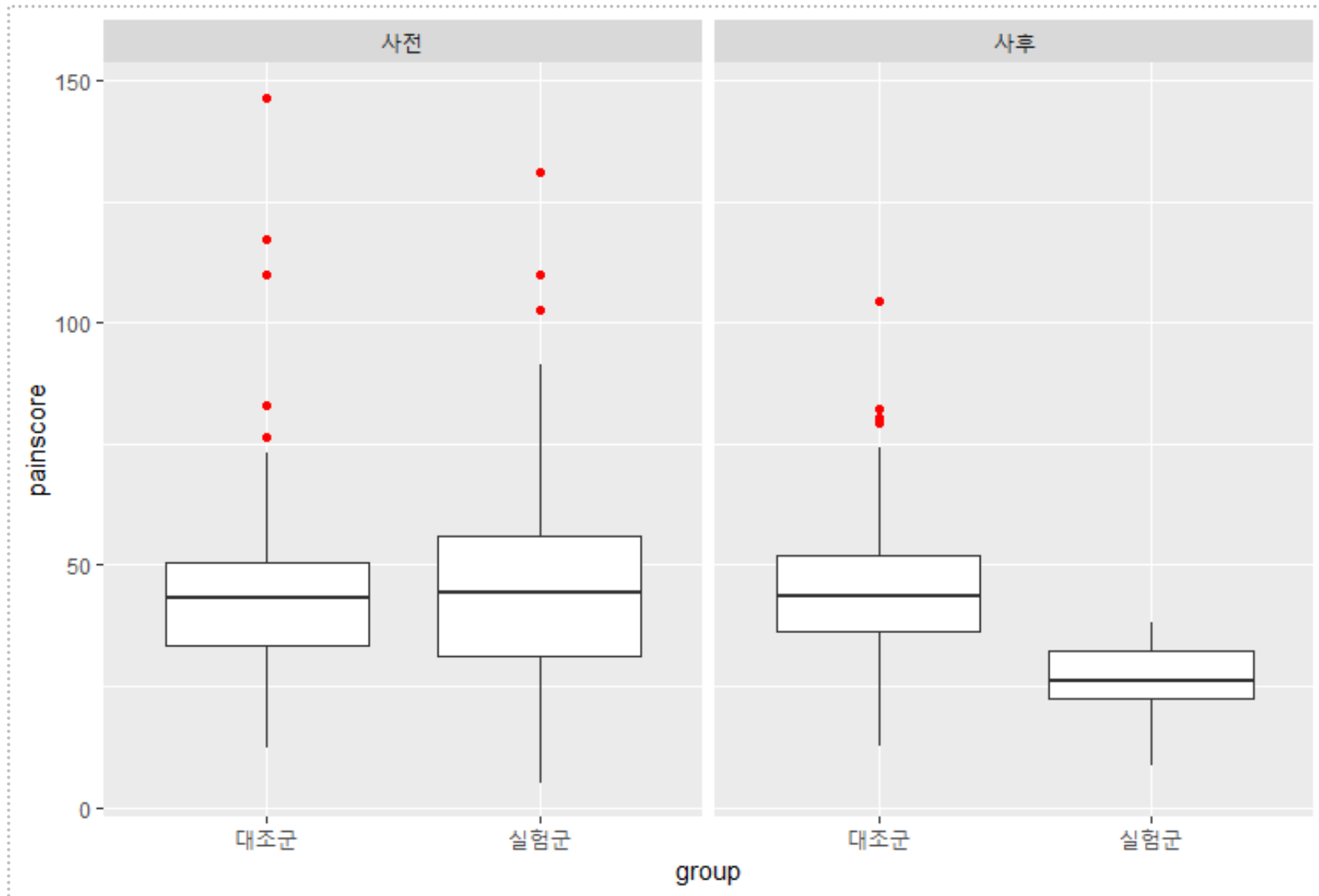
- # 03.그래프 그리기(박스그래프,히스토그램)
- `boxplot(painscore ~ group*time, data=twrmaData)`
- `library(ggplot2)`
- `ggplot(twrmaData, aes(x = group, y = painscore)) +`
- `geom_boxplot(outlier.colour="red") +`
- `facet_wrap(~time)`
- `ggtitle("실험군과 대조군의 실험전후 통증") +`
- `theme_classic() + # ggplot2 테마`

# Two-Way Repeated Measures ANOVA

이상치 제거/처리 동영상  
복습해라



# Two-Way Repeated Measures ANOVA



# Two-Way Repeated Measures ANOVA

- # 04.통계분석
- # 구형성(sphericity)검정: Mauchly's test.**#3개이상 되야지 상관 관계를말할수 있다.**
- # **본 예제는 변수가 2개 이기 때문에 구형성 검정 안됨**
- require(car)
- twrmaMatrix <- cbind(painscore[time=="사전"], painscore[time=="사후"])
- twrmaModelLm <- lm(twrmaMatrix ~ 1)
- timeF <- factor(c("사전","사후"))
- options(contrasts=c("contr.sum", "contr.poly"))
- twrmaResultMt <- Anova(twrmaModelLm, idata=data.frame(timeF),  
#Anova 대문자
- idesign=~timeF, type="III")
- summary(twrmaResultMt, multivariate=F)



# Two-Way Repeated Measures ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ➔

```
> # 구형성(sphericity)검정: Mauchly's test.  
> # 본 예제는 변수가 2개 이기 때문에 구형성 검정 안됨  
> require(car)  
> twrmaMatrix <- cbind(painscore[time=="사전"], painscore[time=="사후"])  
> twrmaModelLm <- lm(twrmaMatrix ~ 1)  
> timeF <- factor(c("사전", "사후"))  
> options(contrasts=c("contr.sum", "contr.poly"))  
> twrmaResultMt <- Anova(twrmaModelLm, idata=data.frame(timeF), #Anova 대문자  
+                          idesign=~timeF, type="III")  
> summary(twrmaResultMt, multivariate=F)
```

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

	Sum Sq	num Df	Error SS	den Df	F value	Pr(>F)
(Intercept)	469354	1	62219	141	1063.641	< 2.2e-16 ***
timeF	6781	1	44184	141	21.638	7.505e-06 ***

---

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

```
> |
```

# Two-Way Repeated Measures ANOVA

- # ANOVA 검정 (방법1=방법2)=2개 방법으로 결과는 같다.
- `twrmaResult <- aov(painscore ~ time*group + Error(id),  
data=twrmaData)`
- `summary(twrmaResult)`
  
- `twrmaResult <- aov(painscore ~ time+group+time:group,  
data=twrmaData)`
- `summary(twrmaResult)`
  
- #상호작용효과 그래프
- `interaction.plot(time, group, painscore)`

# Two-Way Repeated Measures ANOVA

```
Console C:/Users/leecho/Desktop/RStat/Ch08/
> twrmaResult <- aov(painscore ~ time*group)
> summary(twrmaResult)

Error: id
      Df Sum Sq Mean Sq
group  1  10.9    10.9

Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
time    1  6781    6781  24.73 1.15e-06 ***
group    1 24094   24094  87.87 < 2e-16 ***
time:group 1  5798    5798  21.15 6.46e-06 ***
Residuals 279 76500      274
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

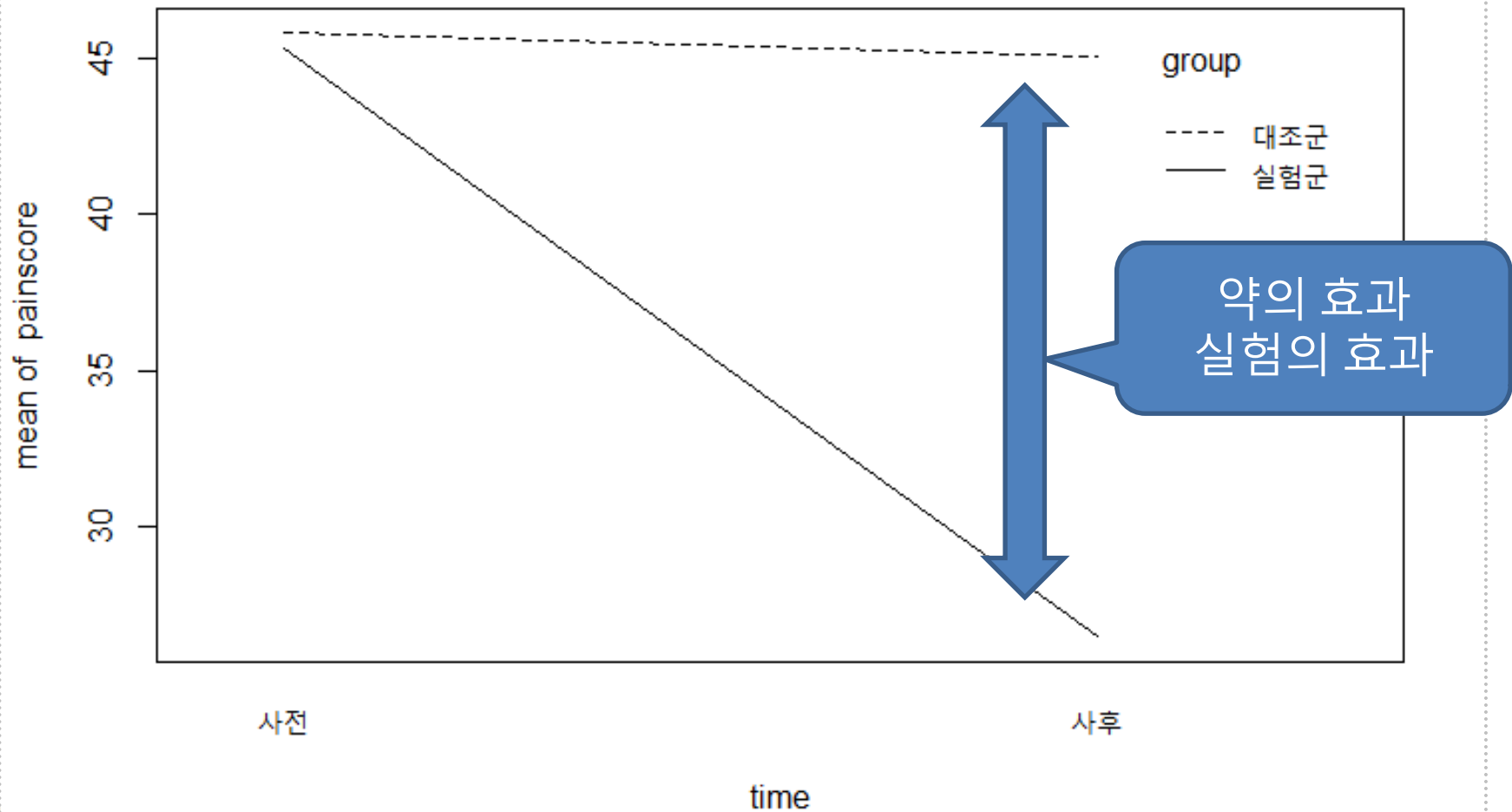
> twrmaResult <- aov(painscore ~ time+group+time:group,
> summary(twrmaResult)

      Df Sum Sq Mean Sq F value Pr(>F)
time    1  6781    6781  20.17 1.04e-05 ***
group    1  6475    6475  19.26 1.62e-05 ***
time:group 1  5798    5798  17.25 4.36e-05 ***
Residuals 280 94130      336
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

TIME과 그룹간의 연관성이있다고  
나옴/그리고 위의 TIME과 그룹의  
수치도 차이가 있다고 p값이 나타남.

만약에 상관관계가  
없다면 TIME/GROUP따로  
해석해주면 된다.

# Two-Way Repeated Measures ANOVA



# Two-Way Repeated Measures ANOVA

- # 사후검정(Multicamparison test )
- # 상호작용이 있을 경우 : 그룹별로 나누어서 분석#위에서 구했으니 카피해서 써라
- # 상호작용이 없을 경우 : 각 변수별 주효과 분석(t-test, ANOVA분석)
- # 사전-사후에 따른 그룹 비교=#차이검정
- `pre <- twrmaData[twrmaData$time=="사전",]`
- `post <- twrmaData[twrmaData$time=="사후",]`



# Two-Way Repeated Measures ANOVA

- `t.test(painscore ~ group,`
  - `data = pre,`
  - `alternative = c("two.sided"),`
  - `var.equal = TRUE,`
  - `conf.level = 0.95)`
- `t.test(painscore ~ group,`
  - `data = post,`
  - `alternative = c("two.sided"),`
  - `var.equal = TRUE,`
  - `conf.level = 0.95)`

# Two-Way Repeated Measures ANOVA

```
Console C:/Users/leecho/Desktop/RStat/Ch08/
Two Sample t-test

data: painscore by group
t = 0.13527, df = 140, p-value = 0.8926
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.982234  8.007868
sample estimates:
mean in group 대조군 mean in group 실험군
      45.79549      45.28268

> t.test(painscore ~ group,
+       data = post,
+       alternative = c("two.sided"),
+       var.equal = TRUE,
+       conf.level = 0.95)

Two Sample t-test

data: painscore by group
t = 8.6963, df = 140, p-value = 8.319e-15
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 14.36073 22.81167
sample estimates:
mean in group 대조군 mean in group 실험군
      45.05972      26.47352
```

# Two-Way Repeated Measures ANOVA

- # 그룹에 따른 사전-사후 비교
- `controlG <- twrmaData[twrmaData$group=="대조군",]`
- `treatG <- twrmaData[twrmaData$group=="실험군",]`
- `t.test(painscore ~ time,`
  - `data = controlG,`
  - `alternative = c("two.sided"),`
  - `var.equal = TRUE,`
  - `conf.level = 0.95)`
- `t.test(painscore ~ time,`
  - `data = treatG,`
  - `alternative = c("two.sided"),`
  - `var.equal = TRUE,`
  - `conf.level = 0.95)`
- `detach(twrmaData)`



# Two-Way Repeated Measures ANOVA

Console C:/Users/leecho/Desktop/RStat/Ch08/ ↗

## Two Sample t-test

```
data: painscore by time
t = 0.22027, df = 140, p-value = 0.826
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.868264  7.339813
sample estimates:
mean in group 사전 mean in group 사후
      45.79549      45.05972
```

```
> t.test(painscore ~ time,
+       data = treatG,
+       alternative = c("two.sided"),
+       var.equal = TRUE,
+       conf.level = 0.95)
```

## Two Sample t-test

```
data: painscore by time
t = 6.7426, df = 140, p-value = 3.762e-10
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 13.29399 24.32432
sample estimates:
mean in group 사전 mean in group 사후
      45.28268      26.47352
```