

## 임상시험 자료분석 II

2<sup>nd</sup> Assignment

조현선 | 162STG26 | 2016/10/10

## 문제

HW-data-2.csv 는 새로 개발된 마취제의 효과를 4 가지 처리수준에서 실험한다. 19 마리의 개들에게 처리의 순서를 랜덤화하여 마취제를 놓고 장박동간격(RRinterval)을 밀리초로 측정하여 반응값으로 기록해 놓은 것이다. 이를 분석하시오. (SAS 나 R 어떤 것을 사용해도 무방하나 결과제출은 자료탐색과정, 자료분석과정, 결과 및 해석이 포함되어 있어야 함.)

## 데이터 소개

데이터는 19 마리의 개에게 한 가지 마취제를 4 개의 처리 수준에서 실험한 데이터로, 단일요인 반복측정 자료이다.

이 자료 분석의 목적은 과연 처리에 따라 다른 효과를 보이는가 하는 것을 알아보는 데 있다. 반복 측정 설계가 의학 및 행동과학 등의 분야에서 빈번히 활용되는 이유는, 실험에서 충분한 수의 실험 개체를 확보하기 어려운 경우에 시행하는 경우도 있고, 개체 스스로가 대조의 역할을 하기 때문에, 실험 개체간 변동이 실험 처리의 비교에 큰 영향을 줄 수 있는 경우에, 개체 스스로가 블록이 된다는 장점도 있다.

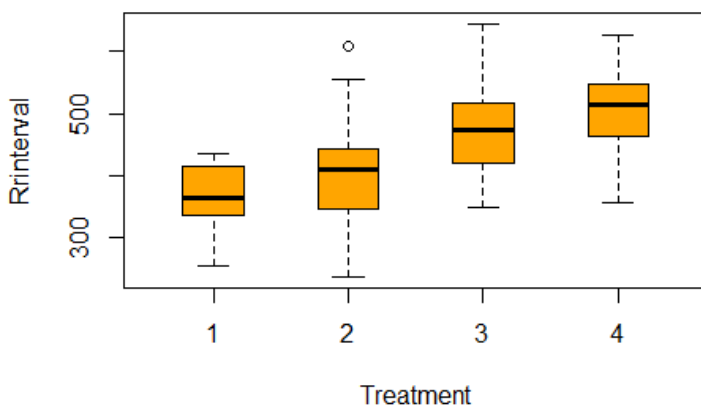
이번 과제에서 다룰 방법은 일변량적 접근방법인데, 이와 같은 일변량적 접근방법은 공분산행렬의 구형성(Sphericity) 같은 특수한 조건이 만족되는 경우에만 타당성을 갖는다. 따라서 공분산행렬에 대해 이와 같은 가정을 할 수 있는 지를 파악하고, 할 수 있다면 일변량 분산분석을 진행한다. 만약 이런 가정을 할 수 없는 경우에는 보다 일반적인 형태의 분석방법인 MANOVA 혹은 자유도 수정방법등을 실행한다.

	Dog	Trt1	Trt2	Trt3	Trt4
1	1	426	609	556	600
2	2	253	236	392	395
3	3	359	433	349	357
4	4	432	431	522	600
5	5	405	426	513	513
6	6	324	438	507	539
7	7	310	312	410	456
8	8	326	326	350	504
9	9	375	447	547	548
10	10	286	286	403	422
11	11	349	382	473	597
12	12	429	410	488	547
13	13	348	377	447	514
14	14	412	473	472	446
15	15	347	326	455	468
16	16	434	458	637	524
17	17	364	367	432	469
18	18	420	395	508	531
19	19	397	556	645	625

## 자료탐색

본격적인 자료분석을 시작하기에 앞서 데이터의 전반적인 추세를 먼저 보고자 한다. 처리 별로 반응변수를

**Boxplot Rinterval vs Trt**



Rinterval 로 두어 Box Plot 을 그려보았더니 왼쪽의 그림과 같았다. Treatment 4 에서 평균이 제일 높아 보이고 treatment 1 에서 제일 낮은 것을 확인 할 수 있었다. 그림으로 보기에 Treatment 별로 차이가 있어보였다.

R 에서 summary 함수를 이용하여 Treatment 별로 수치적으로 요약통계량을 확인해 본 결과를 다음의

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
summary_trt1	253	336.5	364	368.2	416.0	434
summary_trt2	236	346.5	410	404.6	442.5	609
summary_trt3	349	421.0	473	479.3	517.5	645
summary_trt4	357	462.0	514	508.2	547.5	625

표에 정리해놓았다. 숫자 값으로 보기에  
 Treatment1 < Treatment2 < Treatment3 <  
 Treatment4 별로 median 과 mean 이 높았음을  
 확인하였다.

### 자료분석

이 데이터에서는 반복요인이 하나만 존재하는 경우이므로 이때의 반복측정 분산분석을 해보고자 한다. 단일 요인 반복 측정 설계로  $y_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij}$ ,  $\epsilon_{ij} \sim N(0, \sigma^2)$ , ( $i=1,2,\dots,19$ ,  $j=1,2,3,4$ ) 모형을 생각하자. 처리  $\tau_j$ 에 대하여 다음과 같은 가설은 ' $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 (= 0)$ '이고 개체효과를 고정효과가 아닌 랜덤효과로 간주하고 분석을 시작하였다.

위와 같은 일변량적 접근방법은 구형성의 특수한 조건이 만족되는 경우에만 타당성을 가지므로, 이에 대한 검증을 먼저 한다.

Mauchly Tests for Sphericity	구형성 검정의 결과를 살펴보면, p 값이 0.83755 로 충분히 크므로 구형성을 만족한다고 할 수 있다. 따라서 이 자료에 대해서는 혼합모형에 의한 일변량 분산분석을 수행해도 무방하다.
Test statistic p-value Trt 0.88279 0.83755	
Greenhouse-Geisser and Huynh-Feldt Corrections for Departure from Sphericity	또한 E의 추정치를 보면 $E_{GG}=0.92092$ , $E_{HF}=1.104589$ (즉, 1)이다. 그린하우스-가이어 방식에서는 자유도가 조금 바뀌지만, 훈-펠트 방식에서는 수정계수가 1로 간주되어 아무런 수정이 이루어지지 않는다. F 검정 결과 자유도 수정에 관계없이 처리 효과는 유의하다는 것을 알 수 있다.
GG eps Pr(>F[GG]) Trt 0.92092 8.893e-13 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 HF eps Pr(>F[HF]) Trt 1.104589 1.086689e-13	
Multivariate Tests: Trt	
Df test stat approx F num Df den Df Pr(>F)	
Pillai 1 0.855869 31.66997 3 16 5.8053e-07 ***	
Wilks 1 0.144131 31.66997 3 16 5.8053e-07 ***	
Hotelling-Lawley 1 5.938120 31.66997 3 16 5.8053e-07 ***	
Roy 1 5.938120 31.66997 3 16 5.8053e-07 ***	
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	
다요인 반복측정 분산분석의 경우에는 위와 같은 통계량들이 사용될 수 있다. 위 예시는 단일요인 반복측정 분산분석이므로 이들 통계량의 값이 모두 동일함을 알 수 있다.	

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

	SS	num Df	Error SS	den Df	F	Pr(>F)
(Intercept)	14718000	1	305063	18	868.424	< 2.2e-16 ***
Trt	239243	3	108461	54	39.704	1.087e-13 ***

---

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

처리 효과에 관한 F-검정 결과를 살펴보면 p 값이 1.09e-13 으로, 처리효과는 유의하다고 볼 수 있다.

```
> aov.out = aov(Rrinterval ~ factor(Trt) + Error(factor(Dog)/factor(Trt)))
> summary(aov.out)
```

Error: factor(Dog)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	18	305063	16948		

Error: factor(Dog):factor(Trt)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(Trt)	3	239243	79748	39.7	1.09e-13 ***
Residuals	54	108461	2009		

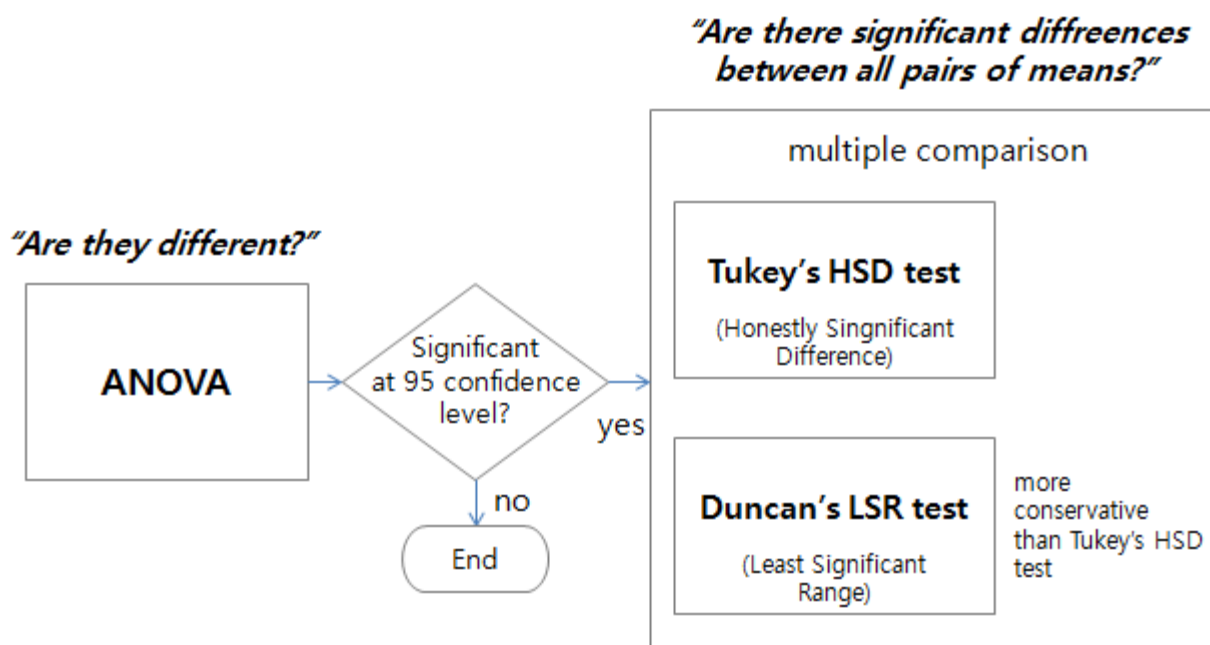
---

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

R 에서 aov 함수를 이용한 결과 위의 표와 일치하는 결과를 보여준다.

## 결과 및 해석

ANOVA 검정 시 집단 간 평균에 차이가 있다고 판단되므로 사후분석으로 Tukey's HSD(honestly significant difference) test 와 Duncan's LSR(Least Significant Range) 를 이용하여 다중비교를 해보았다.



```
> TukeyHSD(aov_model)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = Rrinterval ~ Trt, data =
group_df)

$Trt
      diff      lwr      upr      p adj
2-1  36.42105 -28.24675 101.08885 0.4540140
3-1 111.05263  46.38483 175.72043 0.0001391
4-1 139.94737  75.27957 204.61517 0.0000015
3-2  74.63158   9.96378 139.29938 0.0172213
4-2 103.52632  38.85852 168.19412 0.0004152
4-3  28.89474 -35.77306  93.56254 0.6443651
```

왼쪽의 결과를 보면 'Trt 3'와 'Trt1'은 평균 차이가 -111.05263 이고 adj. P-value 가 0.0001391 이고, 'Trt4'와 'Trt 1'의 평균 차이는 **139.94737** 이고 adj. P-value 가 0.0000015 으로서, 유의수준 0.05 보다 훨씬 작으므로 이들 group (수준) 간에는 평균의 차이가 있다고 유의수준 0.05 기준 하에 판단할 수 있다. 반면에, 'Trt2'와 'Trt1'는 평균 차이가 -36.42105 이고 adjusted P-value 가 0.4540140 로서 유의수준 0.05 보다 크므로 귀무가설 H0 '두 집단 간 평균차이는 없다'를 채택하게 됩니다. 이는 과제 초반에 시행한 자료탐색부분의 Boxplot 을 보면 좀더 쉽게 이해가 될 것 같다.

```
> duncan.test(aov_model, "Trt", alpha = 0.05, console = TRUE)
```

```
Study: aov_model ~ "Trt"
```

```
Duncan's new multiple range test
for Rrinterval
```

```
Mean Square Error: 5743.386
```

```
Trt, means
```

	Rrinterval	std	r	Min	Max
1	368.2105	53.09695	19	253	434
2	404.6316	89.23640	19	236	609
3	479.2632	82.77268	19	349	645
4	508.1579	73.07398	19	357	625

```
alpha: 0.05 ; Df Error: 72
```

```
Critical Range
```

	2	3	4
	49.01517	51.57070	53.26122

```
Means with the same letter are not significantly different.
```

```
Groups, Treatments and means
```

a	4	508.2
a	3	479.3
b	2	404.6
b	1	368.2

사후분석으로 "어느 집단 간에 차이가 있는가?"를 알아보기 위해 Duncan's LSR test 를 해보니, Treatment 의 조건에 Rrinterval 의 평균은 위의 첫번째 결과와 같았고, Trt level '4'와 '3'은 서로 차이가 없고, Trt level '1'과 '2'가 짝을 이룬 수준평균간에 차이가 있다고 나왔다.

## INDEX → R code

```
data<-read.table("C://Users//samsung//Desktop//대학원//임상//hw-data-21.csv",header=T,sep=",")
data2<-read.table("C://Users//samsung//Desktop//대학원//임상//HW-data-2.csv",header=T,sep=",")
View(data2)
attach(data2)
head(data2)
data2$Dog<-as.factor(Dog)
data2$Trt<-as.factor(Trt)
str(data2)
```

### #자료탐색과정

```
boxplot(Rinterval~Trt,col="orange",main="Boxplot Rinterval vs Trt",boxwex=0.45,xlab="Treatment",ylab="Rinterval")
summary_trt1<-summary(Rinterval[Trt=='1'])
summary_trt2<-summary(Rinterval[Trt=='2'])
summary_trt3<-summary(Rinterval[Trt=='3'])
summary_trt4<-summary(Rinterval[Trt=='4'])
rbind(summary_trt1,summary_trt2,summary_trt3,summary_trt4)
```

### #자료분석과정

```
library(nlme)
am2<-lme(Rinterval~factor(Trt),random=~1|Dog/Trt)
summary(am2)
anova(am2)

aov.out = aov(Rinterval ~ factor(Trt) + Error(factor(Dog)/factor(Trt)))
summary(aov.out)
```

### #구형성 확인 및 분석

```
library(car)
lm<-lm(cbind(Trt1,Trt2,Trt3,Trt4)~1,data=data)
trt<-factor(c("Trt1","Trt2","Trt3","Trt4"))
Trt.data<-data.frame(Trt=trt)
ano<-Anova(lm,idata=Trt.data,idesign=~Trt)
summary(ano)
summary(ano,multivariate = FALSE)
```

### #사후분석

```
install.packages("agricolae")
```

```
library(agricolae)
group_df<-data.frame(Rinterval,Trt)
sapply(group_df, class)
group_df <- transform(group_df, Trt= factor(Trt))
sapply(group_df, class)
aov_model <- aov(Rinterval ~ Trt, data = group_df)
summary(aov_model)

duncan.test(aov_model, "Trt", alpha = 0.05, console = TRUE)
attach(group_df)
boxplot(Rinterval~Trt)
TukeyHSD(aov_model)
```