
임상시험 자료분석II HW4

Repeated Measures ANOVA



과목명	임상시험 자료분석II
제출일자	2018.10.13
학과	통계학과
학번	182STG01
이름	고명지

1. 데이터 소개

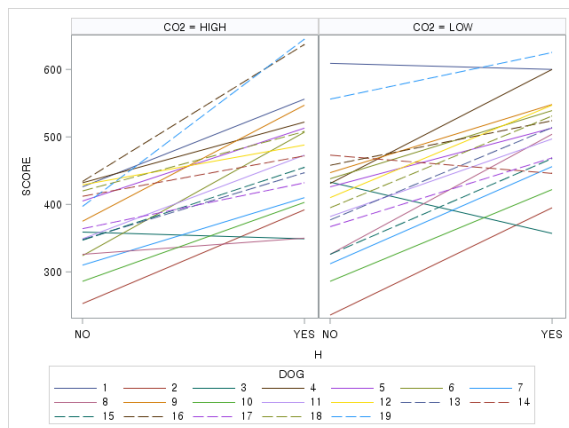
개선된 마취제는 종종 동물에 대한 영향을 먼저 연구함으로써 개발된다. 한 연구에서, 19마리의 개에게 처음으로 약물 pentobarbital을 투여하였다. 각각의 개는 두 개의 압력 수준의 이산화탄소 환경에서 실험되었고, halothane(H)가 추가된 후 다시 두개의 압력 수준에서 이산화탄소 환경에서 실험되었다. 반응변수는 심장박동간격(단위:밀리초)이다.

- Treatment1 : High CO2 pressure without H - Treatment2 : Low CO2 pressure without H
- Treatment3 : High CO2 pressure with H - Treatment4 : Low CO2 pressure with H

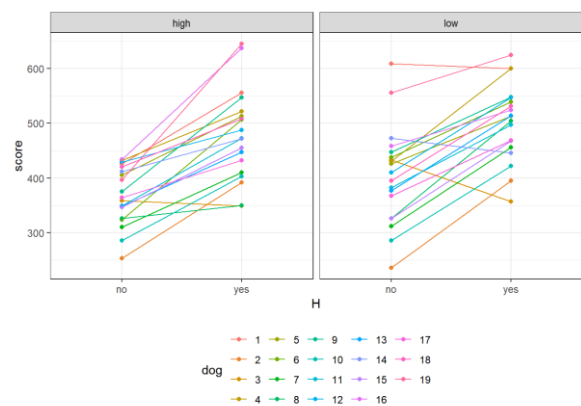
처리는 H의 여부, CO2의 압력 크기로 2가지이고 각각에 대해 2번을 반복하여 개 한 마리당 총 4번을 반복한 자료이다. 따라서 각 처리에 대한 효과를 Repeated Measures ANOVA를 통해 알아보아야 한다.

2. 자료 탐색

(1) H의 여부



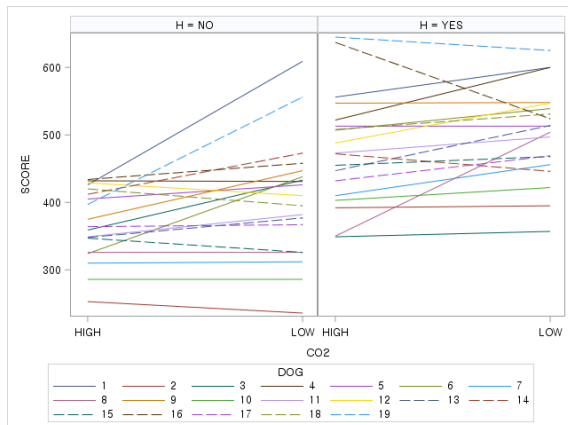
[SAS]



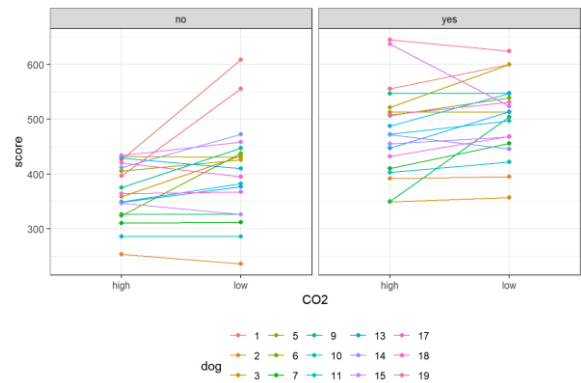
[R]

H의 효과를 먼저 그림을 통해 예상해보았다. 예외인 개도 있지만 CO2의 압력 수준에 따라 상관없이 H가 없는 환경보다 H가 추가된 환경에서 개의 심장박동간격이 더 높았다는 것을 알 수 있다. 따라서 H의 효과가 있을 것이라고 예상해 보았다.

(2) CO2 압력 수준



[SAS]



[R]

H의 여부에 상관없이 다수의 개가 CO2가 높은 수준일 때와 낮은 수준일 때의 심장박동수가 거의 비슷하다는 것을 확인하였다. 하지만 개에 따라서 H가 없을 때와 H가 있을 때 CO2의 압력 수준에 따라 심장박동수의 양상이 다르므로 Repeated measures ANOVA를 이용하여 정확한 분석을 통해 알아볼 필요가 있다.

3. 자료 분석

이 데이터에서는 두 가지의 처리가 존재하고 각각의 수준은 모두 2가지 이므로 모형을 세워보면

$y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + \varepsilon_{ijk}$, $i = 1, 2, \dots, 19, j = 1, 2, k = 1, 2$ 로 나타낼 수 있다.

(1) Method 1

The GLM Procedure					
Dependent Variable: SCORE					
Tests of Hypotheses Using the Type III MS for DOG*H as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
H	1	208112.2237	208112.2237	88.26	<.0001
Tests of Hypotheses Using the Type III MS for DOG*CO2 as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CO2	1	17130.01316	17130.01316	13.19	0.0019

[SAS]

	numDF	denDF	F-value	p-value
## (Intercept)	1	54	862.3015	<.0001
## H	1	54	112.5668	<.0001
## CO2	1	54	9.2655	0.0036
## H:CO2	1	54	0.4203	0.5196

[R]

SAS에서는 모형을 세우고 각각의 처리에 대하여 test할 때 error를 지정해주어 해줄 수 있지만 R에서는 그러지 못한다. 이 데이터에서는 개가 random 처리가 되지만 각 처리에 대한 F-value를 계산해 줄 때는 error term이 처리와 dog의 교호작용이 된다. 하지만 R에서는 dog에 대해서 random 처리를 해주면 error term이 H*dog, CO2*dog, H*dog*CO2 term이 모두 풀링된 형태가 된다. 따라서 값이 약간 다를 수 있다. H의 효과에 대한 p-value와 CO2의 효과에 대한 p-value가 SAS와 R에서 모두 0.05보다 작으므로 두 효과는 유의하다고 할 수 있다.

(2) Method 2

The GLM Procedure					
Repeated Measures Analysis of Variance					
Univariate Tests of Hypotheses for Within Subject Effects					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
H	1	208112.2237	208112.2237	88.26	<.0001
Error(H)	18	42445.0263	2358.0570		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CO2	1	17130.01316	17130.01316	13.19	0.0019
Error(CO2)	18	23381.23684	1298.95760		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
H*CO2	1	776.96053	776.96053	0.41	0.5294
Error(H*CO2)	18	34008.28947	1889.34942		

[SAS]

```
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##              Sum Sq num Df Error SS den Df  F value    Pr(>F)
## (Intercept) 14630119      1  305394      18 862.3015 < 2.2e-16 ***
## H           208112      1   42445      18  88.2558 2.315e-08 ***
## CO2         17130      1   23381      18  13.1875 0.001909 **
## H:CO2        777      1   34008      18   0.4112 0.529426
```

[R]

이 경우에는 SAS에서와 R에서 같은 값을 가짐을 알 수 있다. H의 여부와 CO2 압력 수준에 따른 p-value가 모두 0.05보다 작으므로 유의하다고 할 수 있지만 교호작용 효과는 유의하지 않다.

INDEX

(1) SAS

```

PROC IMPORT DATAFILE="C:\Users\User\Desktop\대학원\2학기\임상\HW4\sleeping-
dog.csv"
OUT=DOG;
GETNAMES=YES;
RUN;

DATA DOG2; SET DOG (RENAME=(Dog=DOG) );
    KEEP DOG H CO2 SCORE;
    LENGTH H$ 3. CO2$ 4.;
    SCORE=TRT1; H="NO"; CO2="HIGH"; OUTPUT;
    SCORE=TRT2; H="NO"; CO2="LOW"; OUTPUT;
    SCORE=TRT3; H="YES"; CO2="HIGH"; OUTPUT;
    SCORE=TRT4; H="YES"; CO2="LOW"; OUTPUT;
RUN;

/*PLOT*/
PROC SGPANEL DATA=DOG2;
    PANELBY CO2;
    SERIES X=H Y=SCORE / GROUP=DOG;
RUN;
PROC SGPANEL DATA=DOG2;
    PANELBY H;
    SERIES X=CO2 Y=SCORE / GROUP=DOG;
RUN;

/*METHOD1*/
PROC GLM DATA=DOG2;
    CLASS DOG H CO2;
    MODEL SCORE=H CO2 DOG H*CO2 H*DOG CO2*DOG;
    RANDOM DOG;
    TEST H=H E=DOG*H;
    TEST H=CO2 E=DOG*CO2;
RUN;

/*METHOD2*/
DATA DOG; SET DOG (RENAME=(Dog=DOG TRT1=NH TRT2=NL TRT3=YH TRT4=YL) );
    KEEP DOG NH NL YH YL;
RUN;
PROC GLM DATA=DOG;
    CLASS DOG;
    MODEL NH NL YH YL=;
    REPEATED H 2, CO2 2 / PRINTE SUMMARY;
    QUIT;
RUN;

```

(2) R

```

library(tidyverse);library(nlme);library(car)
setwd("C:/Users/User/Desktop/대학원/2학기/임상/HW4")
dog.t<-read.csv("sleeping-dog.csv",header=T)
colnames(dog.t)<-c("dog","NH","NL","YH","YL"); dog.t$dog<-as.factor(dog.t$dog)
dog2<-gather(dog.t, H, score, NH, NL, YH, YL)
dog2<-dog2%>%mutate("H"=substr(dog2$H,1,1),"CO2"=substr(dog2$H,2,2))
dog2$H<-ifelse(dog2$H=="N","no","yes"); dog2$CO2<-ifelse(dog2$CO2=="H","high","low")
dog2<-dog2%>%select(dog,H,CO2,score)%>%arrange(dog)

# plot
ggplot(dog2, aes(x=H,y=score,group=dog,col=dog))+geom_point()+geom_line()+facet_wrap(~CO2)+theme_bw()+theme(legend.p
osition = "bottom")
ggplot(dog2, aes(x=CO2,y=score,group=dog,col=dog))+geom_point()+geom_line()+facet_wrap(~H)+theme_bw()+theme(legend.p
osition = "bottom")

# method1
model1<-lme(score~H*CO2+H*dog+CO2*dog,random=~1|dog, data=dog2)
anova(model1)

# method2
H<-factor(c("yes","no"));CO2<-factor(c("high","low"))
HCO2.data<-data.frame(H=factor(rep(H,each=2)),CO2=factor(rep(CO2,2)))
model2.lm<-lm(cbind(NH,NL,YH,YL)~1,data=dog.t)
model2<-Anova(model2.lm,idata=HCO2.data,idesign=~H+CO2+H*CO2)
summary(model2)

```