HW2

임상시험자료분석 II

182STG27

임지연

1. R의 nlme package 에 내장되어 있는 BodyWeight 자료를 이용하시오.

1) 적절한 그림으로 자료를 살펴보시오.

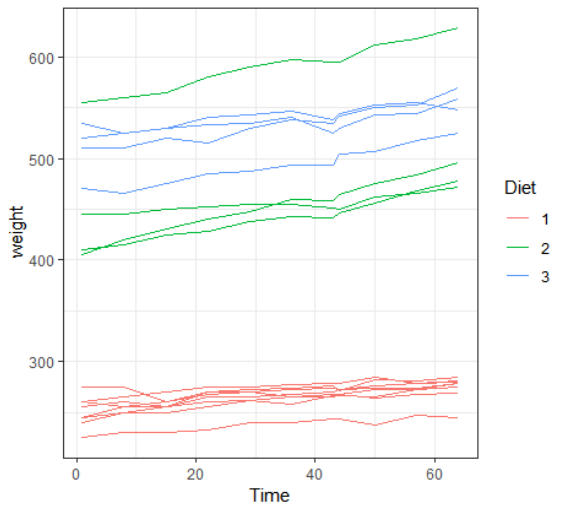
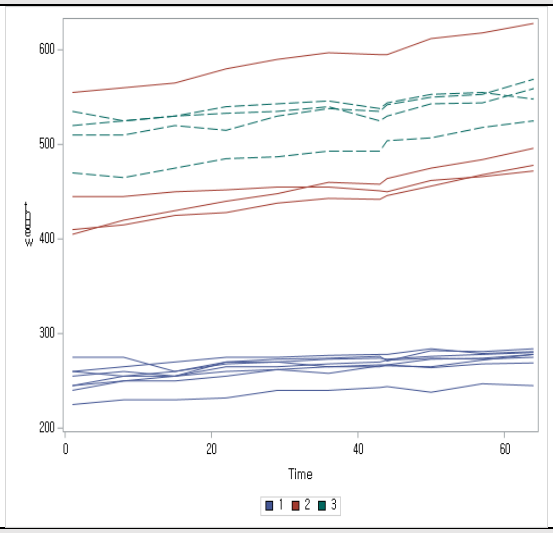
<table>
<tr><th>R</th><th>SAS</th></tr>
<tr><td colspan="2">1. CODE</td></tr>
<tr><td>

```
library(nlme); library(faraway)
library(ggplot2); library(gridExtra)
library(gmodels); library(car)
library(reshape); library(dplyr)
str(BodyWeight)
theme_set(theme_bw())
ggplot(BodyWeight, aes(x = Time, y = weight, colour = Diet)) + geom_line(aes(group = Rat))
```

</td><td>

```
data BodyWeight;
        infile "C:₩Users₩jeeyeon₩Desktop₩data₩BodyWeight.csv" delimiter=',' firstobs=2;
        input weight Time Rat Diet;
        week = compress('W'|| Time, ''); run;

proc sgplot data=BodyWeight;
    title 'BodyWeight';
    series x=time y=weight / group=Rat groupIc=Diet name='grouping';
    keylegend 'grouping' / type=linecolor;
run;
```

</td></tr>
<tr><td colspan="2">2. PLOT</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td colspan="2">3. RESULT</td></tr>
<tr><td colspan="2">BodyWeight 자료는 세 그룹의 쥐를 각각 다른식이 요법으로 처리하여 1일 ~ 7 일까지 64 일 동안 측정 된 쥐의 체중 (44 일에는 추가 측정)을 나타낸 자료이다. 이 자료의 변수설명은 다음과같다.
- weight : numeric vector -giving the body weight of the rat
- time : numeric vector -giving time at which the measurement is made
- Rat : ordered factor
- Diet : a factor level 1 to 3
즉, 이 데이터에서 Diet - Trt, Rat - Patient , Time - time, Weight – Score 에 해당한다. 우리의 목적은 각각 다른 식이요법간에 (Trt- Diet) 쥐의 Weight 차이가 있는지를 살펴보는 것이다. 따라서, 각 그룹을 Diet 1,2,3으로 나누어 그래프를 그려본 결과 1 그룹의 Weight는 300 미만으로 현저하게 낮게 나타난 반면, 2,3 그룹의 Weight는 400~ 600 사이이며, 2 그룹의 분산이 더 커보인다. 그래프로 봐서는 Diet 1 약의 효과가 있을 것이라고 예상할 수 있다.</td></tr>
</table>

2) 적절한 방법으로 분석하여 결과를 해석하시오.

R

1. CODE

```
# Anova ------------------------------------------
lm1 <- lm(weight ~ factor(Diet) + factor(Rat) + factor(Time)
+factor(Diet) * factor(Time), data = BodyWeight )
temp <- anova(lm1); temp
F.group <- temp[1,3] / temp[2,3] # F.value
P.group <- 1-pf(F.group,temp[1,1], temp[2,1])
c(F.group, P.group)
lme1 <- lme(weight ~ factor(Diet) + factor(Time)
+factor(Diet) * factor(Time), random = ~1|factor(Rat),data =
BodyWeight )
anova(lme1)
# Manova -----------------------------------------
head(BodyWeight)
BodyWeight <- BodyWeight %>% mutate(Week =
paste("W",BodyWeight$Time, sep=""))
BodyWeight_M <- cast(BodyWeight , Diet + Rat ~ Week,
value="weight")
lm2 <- lm( cbind(W1, W8, W15, W22, W29, W36, W43, W44,
W50, W57, W64) ~ Diet ,data = BodyWeight_M )
measure_time <- factor( c("W1", "W8", "W15", "W22", "W29",
"W36", "W43", "W44", "W50","W57","W64") )
measure_time_data <- data.frame(measure_time =
measure_time)
mv1 <- Anova(lm2, idata = measure_time_data,idesign = ~
measure_time)
summary(mv1)
```

SAS

1. CODE

```
/* ANOVA -------------------------*/
PROC GLM data=BodyWeight;
        CLASS Diet Rat Time;
        MODEL Weight = Diet Rat(Diet) Time
Diet*Time / ss3;
        RANDOM RAT(Diet);
        TEST H = Diet E = Rat(Diet);
        QUIT;
RUN;


/*  MANOVA -------------------------*/
proc sort data=BodyWeight; by Diet Rat; run;
proc transpose data=BodyWeight out=Bodyweight_M;
        by Diet Rat;
        id Week;
        var weight;
run;

ods exclude
        partialCorr
        ErrorSSCP;

proc glm data = Bodyweight_M;
        class Diet;
        model W1 W8 W15 W22 W29 W36 W43 W44
W50 W57 W64 = Diet / ss3;
        repeated time profile /printe summary;
        quit;
run;
```

2. TABLE

 # ANOVA TABLE

```
> temp <- anova(lm1); temp
Analysis of Variance Table

Response: weight
                     Df  Sum Sq Mean Sq    F value    Pr(>F)
factor(Diet)          2 2604050 1302025 32738.2082 < 2.2e-16 ***
factor(Rat)          13  192186   14784   371.7177 < 2.2e-16 ***
factor(Time)         10   23400    2340    58.8378 < 2.2e-16 ***
factor(Diet):factor(Time) 20    4906     245     6.1679 2.883e-11 ***
Residuals           130    5170      40
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> c(F.group, P.group)
[1] 88.07276434383193  0.00000002763486


> anova(lme1)
                            numDF denDF   F-value p-value
(Intercept)                     1   130 1759.9049  <.0001
factor(Diet)                    2    13   88.0728  <.0001
factor(Time)                   10   130   58.8378  <.0001
factor(Diet):factor(Time)      20   130    6.1679  <.0001
```

 # ANOVA TABLE

The GLM Procedure

Dependent Variable: weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	45	2824541.744	62767.594	1578.23	<.0001
Error	130	5170.205	39.771		
Corrected Total	175	2829711.949			

R-Square	Coeff Var	Root MSE	weight Mean
0.998173	1.640231	6.306410	384.4830

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Diet	2	2604049.733	1302024.866	32738.2	<.0001
Rat(Diet)	13	192185.670	14783.513	371.72	<.0001
Time	10	27000.205	2700.020	67.89	<.0001
Diet*Time	20	4906.080	245.304	6.17	<.0001

The GLM Procedure

Source	Type III Expected Mean Square
Diet	Var(Error) + 11 Var(Rat(Diet)) + Q(Diet,Diet*Time)
Rat(Diet)	Var(Error) + 11 Var(Rat(Diet))
Time	Var(Error) + Q(Time,Diet*Time)
Diet*Time	Var(Error) + Q(Diet*Time)

The GLM Procedure

Dependent Variable: weight

Tests of Hypotheses Using the Type III MS for Rat(Diet) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Diet	2	2604049.733	1302024.866	88.07	<.0001

Diet , Time 변수의 p-value < 0.0001로 두 변수가 모두 유의하다고 볼 수 있다.

 # MANOVA TABLE

```
Multivariate Tests: Diet
                 Df test stat approx F num Df den Df         Pr(>F)
Pillai            2   0.93127 88.07276      2     13 0.000000027635 ***
Wilks             2   0.06873 88.07276      2     13 0.000000027635 ***
Hotelling-Lawley  2  13.54966 88.07276      2     13 0.000000027635 ***
Roy               2  13.54966 88.07276      2     13 0.000000027635 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Multivariate Tests: measure_time
                 Df test stat approx F num Df den Df    Pr(>F)
Pillai            1   0.97831 18.04504     10      4 0.0066562 **
Wilks             1   0.02169 18.04504     10      4 0.0066562 **
Hotelling-Lawley  1  45.11260 18.04504     10      4 0.0066562 **
Roy               1  45.11260 18.04504     10      4 0.0066562 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Multivariate Tests: Diet:measure_time
                 Df test stat  approx F num Df den Df    Pr(>F)
Pillai            2   1.82897  5.346895     20     10 0.0047364 **
Wilks             2   0.00327  6.590314     20      8 0.0050459 **
Hotelling-Lawley  2  50.23337  7.535006     20      6 0.0095215 **
Roy               2  44.52489 22.262445     10      5 0.0015751 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Mauchly Tests for Sphericity

                  Test statistic          p-value
measure_time      0.000000068081 0.00000000005357
Diet:measure_time 0.000000068081 0.00000000005357
```

 # MANOVA TABLE

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	54	0.0004138	73.747132	0.0383
Orthogonal Components	54	6.8081E-8	156.22426	<.0001

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no time Effect
H = Type III SSCP Matrix for time
E = Error SSCP Matrix

S=1 M=4 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01448680	27.21	10	4	0.0030
Pillai's Trace	0.98551320	27.21	10	4	0.0030
Hotelling-Lawley Trace	68.02837375	27.21	10	4	0.0030
Roy's Greatest Root	68.02837375	27.21	10	4	0.0030

MANOVA Test Criteria and F Approximations for the Hypothesis of no time*Diet Effect
H = Type III SSCP Matrix for time*Diet
E = Error SSCP Matrix

S=2 M=3.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00327436	6.59	20	8	0.0050
Pillai's Trace	1.82896904	5.35	20	10	0.0047
Hotelling-Lawley Trace	50.23337469	10.05	20	4	0.0186
Roy's Greatest Root	44.52488954	22.26	10	5	0.0016

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

P-value = 0.0047로, Diet, Time 의 교호효과가 있다고 할 수 있다. 또한 Diet, Time 각각의 변수도 유의하다고 보여진다. 또한 구형성 검정을 해보았을 때, p-value 값은 매우 작아 구형성가정을 만족하지 않는다고 할 수 있다. 따라서 Anova 분석보다 Manova 분석이 더 적합하다.

3. RESULT

BodyWeight 자료분석 결과, p-value < 0.05 이므로 구형성가정을 만족하지 않아, Manova 분석이 더 적합하다고 판단하였다. Manova 분석 결과 Time * Diet 의 p-value < 0.05로, 효과가 유의하다고 할 수 있다. 따라서 Time, Diet 에 따라 Weight 에 효과가 있으며, Diet 그룹 1,2,3 에 대하여 사후분석을 통해 어떠한 그룹간에 차이가 있는지 살펴볼 수 있다.

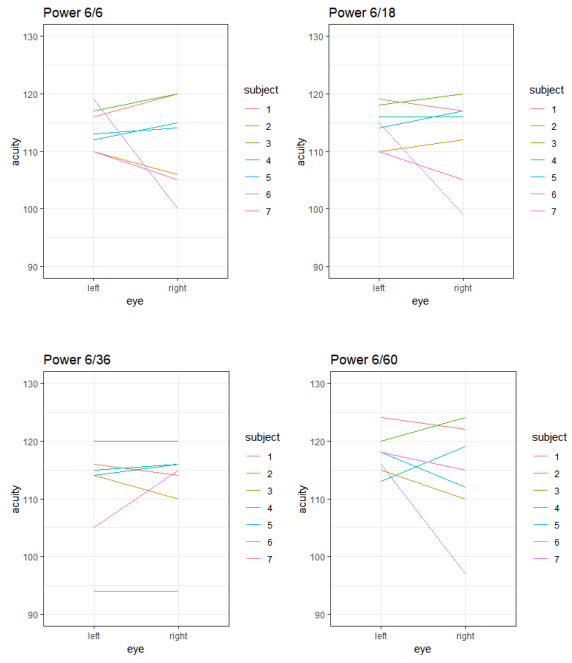
2. R의 faraway package에 내장되어 있는 vision 자료를 이용하시오.

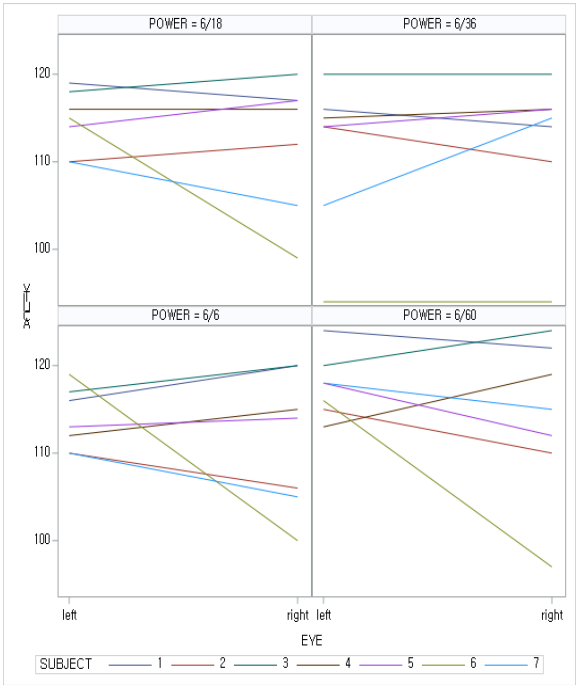
1) 적절한 그림으로 자료를 살펴보시오.

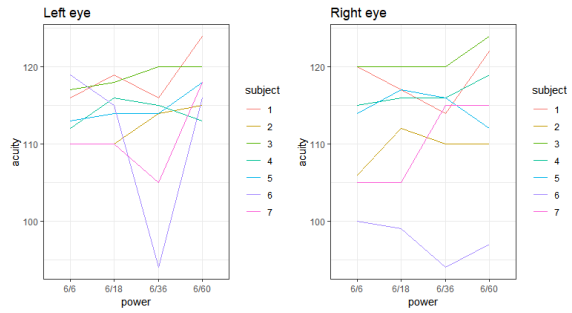
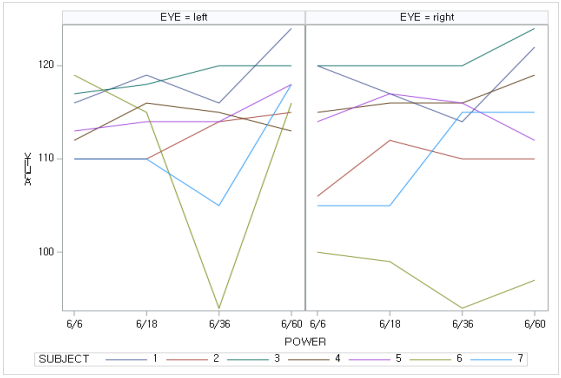
R	SAS
1. CODE	

str(vision)
theme_set(theme_bw())
g1<-ggplot(vision[vision$power=='6/6',],aes(eye,acuity))+
geom_line(aes(color=subject,group=subject))+
ggtitle('Power 6/6')+ ylim(90,130)
g2<-ggplot(vision[vision$power=='6/18',],aes(eye,acuity))+
geom_line(aes(color=subject,group=subject))+
ggtitle('Power 6/18')+ ylim(90,130)
g3<-ggplot(vision[vision$power=='6/36',],aes(eye,acuity))+
geom_line(aes(color=subject,group=subject))+
ggtitle('Power 6/36')+ ylim(90,130)
g4<-ggplot(vision[vision$power=='6/60',],aes(eye,acuity))+
geom_line(aes(color=subject,group=subject))+
ggtitle('Power 6/60')+ ylim(90,130)
grid.arrange(g1,g2,ncol=2); grid.arrange(g3,g4,ncol=2)
 	
g1 <- ggplot(vision[vision$eye=='left',], aes(x=power,y=acuity))+
geom_line(aes(color=subject,group=subject))+ggtitle('Left eye')
g2 <- ggplot(vision[vision$eye=='right',], aes(x=power,y=acuity))+
geom_line(aes(color=subject,group=subject))+ggtitle('Right eye')
grid.arrange(g1,g2,ncol=2)
	PROC SGPANEL DATA=vision;
	panelby EYE;
	title 'vision';
	series X=POWER Y=ACUITY / group=SUBJECT;
	RUN;
	
	PROC SGPANEL DATA=vision;
	panelby POWER;
	title 'vision';
	series X=EYE Y=ACUITY / group=SUBJECT;
	RUN;

2. PLOT	
# Power 수준을 나눈 right / left eye의 acuity graph	**# Power 수준을 나눈 right / left eye의 acuity graph**





# Eye 수준을 나눠 각 power별 acuity graph	# Eye 수준을 나눠 각 power별 acuity graph
	

3. RESULT

Vision 자료는 7명의 피실험자를 4가지 종류의 렌즈(power)를 왼쪽눈, 오른쪽눈으로 나누어 시력검사를 한 자료이다. 이 자료의 변수설명은 다음과같다.

- acuity : a numeric vector (예민함)
- power : a factor with levels 6/6 6/18 6/36 6/60 : power of lens의미/ 물체의거리 Diet : a factor level 1 to 3
- eye : a factor with levels : left right
- subject : a factor with levels 1 2 3 4 5 6 7

즉, 이 데이터에서 power-Trt, subject - Subject, eye- Time, acuity – Score 에 해당한다. 우리의 목적은 각각 다른 렌즈간에 (Trt- power) 피실험자의 acuity 차이가 있는지를 살펴보는 것이다. 따라서, 각 subject를 power(6/6, 6/18, 6/36, 6/60), eye(left, right)로 나누어 그래프를 그려본 결과 power에 따라 개인별 차이를 알 수 있었다. 또한 개인별로 left, right 차이를 알아보기 위하여 그래프를 그려본 결과 거의 비슷하다는 것을 직관적으로 알 수 있었다.

2) 적절한 방법으로 분석하여 결과를 해석하시오.

R

1. CODE

```
# Anova
lm1 <- lm(acuity ~ factor(power) + factor(subject) + factor(eye)
+factor(power) * factor(eye), data = vision )
temp <- anova(lm1); temp
F.group <- temp[1,3] / temp[2,3] # F.value
P.group <- 1-pf(F.group,temp[1,1], temp[2,1])
c(F.group, P.group)
lme2 <- lme(acuity ~ factor(power) + factor(eye) + factor(power) *
factor(eye), random = ~1|factor(subject),data = vision )
anova(lme2)


# Manova
vision_M <- cast(vision , power + subject ~ eye, value="acuity")
lm2 <- lm( cbind(left, right) ~ power ,data = vision_M )
measure_time <- factor( c("left","right") )
measure_time_data <- data.frame(measure_time = measure_time)
mv2 <- Anova(lm2, idata = measure_time_data,idesign = ~
measure_time)
summary(mv2)
```

SAS

1. CODE

```
/* ANOVA */
PROC GLM data=vision;
        CLASS power subject eye;
        MODEL acuity = power subject eye power*eye /
ss3;
        RANDOM subject;
        TEST H = power E = subject;
        QUIT;
RUN;


/*  MANOVA */
proc sort data=vision; by power subject; run;
proc transpose data=vision out=vision_M;
        by power subject;
        id eye;
        var acuity;
run;

ods exclude
        partialCorr
        ErrorSSCP;

proc glm data = vision_M;
        class power;
        model right left = power / ss3;
        repeated eye profile /printe summary;
        quit;
run;
```

2. TABLE

 # ANOVA TABLE

```
> temp <- anova(lm1); temp
Analysis of Variance Table

Response: acuity
                      Df  Sum Sq Mean Sq F value       Pr(>F)
factor(power)          3  140.77  46.923  2.0437       0.1223
factor(subject)        6 1379.43 229.905 10.0136 0.0000007423 ***
factor(eye)            1   46.45  46.446  2.0230       0.1623
factor(power):factor(eye)  3   40.62  13.542  0.5898       0.6251
Residuals             42  964.29  22.959
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


> c(F.group, P.group)
[1] 0.2040959 0.8899679


> anova(lme2)
                      numDF  denDF   F-value p-value
(Intercept)               1     42 3132.9123  <.0001
factor(power)             3     42    2.0437  0.1223
factor(eye)               1     42    2.0230  0.1623
factor(power):factor(eye)     3     42    0.5898  0.6251
```

 # ANOVA TABLE

vision

The GLM Procedure

Dependent Variable: ACUITY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	1607.267857	123.635989	5.39	<.0001
Error	42	964.285714	22.959184		
Corrected Total	55	2571.553571			

R-Square	Coeff Var	Root MSE	ACUITY Mean
0.625018	4.224975	4.791574	113.4107

Source	DF	Type III SS	Mean Square	F Value	Pr > F
POWER	3	140.767857	46.922619	2.04	0.1223
SUBJECT	6	1379.428571	229.904762	10.01	<.0001
EYE	1	46.446429	46.446429	2.02	0.1623
POWER*EYE	3	40.625000	13.541667	0.59	0.6251

Source	Type III Expected Mean Square
POWER	Var(Error) + Q(POWER,POWER*EYE)
SUBJECT	Var(Error) + 8 Var(SUBJECT)
EYE	Var(Error) + Q(EYE,POWER*EYE)
POWER*EYE	Var(Error) + Q(POWER*EYE)

Dependent Variable: ACUITY

Tests of Hypotheses Using the Type III MS for SUBJECT as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
POWER	3	140.7678571	46.9226190	0.20	0.8900

Power, Eye 변수의 p-value > 0.05로 두 변수가 모두 유의하지 않으므로 두 변수 모두 효과가 없다고 할 수 있다.

 # MANOVA TABLE

```
Multivariate Tests: power
                 Df test stat  approx F num Df den Df  Pr(>F)
Pillai            3 0.0742426 0.6415724      3     24 0.59574
Wilks             3 0.9257574 0.6415724      3     24 0.59574
Hotelling-Lawley  3 0.0801965 0.6415724      3     24 0.59574
Roy               3 0.0801965 0.6415724      3     24 0.59574

Multivariate Tests: measure_time
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            1 0.0731584 1.894392      1     24 0.18141
Wilks             1 0.9268416 1.894392      1     24 0.18141
Hotelling-Lawley  1 0.0789330 1.894392      1     24 0.18141
Roy               1 0.0789330 1.894392      1     24 0.18141

Multivariate Tests: power:measure_time
                 Df test stat  approx F num Df den Df  Pr(>F)
Pillai            3 0.0645811 0.5523185      3     24 0.65151
Wilks             3 0.9354189 0.5523185      3     24 0.65151
Hotelling-Lawley  3 0.0690398 0.5523185      3     24 0.65151
Roy               3 0.0690398 0.5523185      3     24 0.65151
```

 # MANOVA TABLE

vision

The GLM Procedure

Repeated Measures Analysis of Variance

Repeated Measures Level Information		
Dependent Variable	right	left
Level of eye	1	2

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no eye Effect
H = Type III SSCP Matrix for eye
E = Error SSCP Matrix

S=1 M=-0.5 N=11

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.92684162	1.89	1	24	0.1814
Pillai's Trace	0.07315838	1.89	1	24	0.1814
Hotelling-Lawley Trace	0.07893299	1.89	1	24	0.1814
Roy's Greatest Root	0.07893299	1.89	1	24	0.1814

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no eye*POWER Effect
H = Type III SSCP Matrix for eye*POWER
E = Error SSCP Matrix

S=1 M=0.5 N=11

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.93541885	0.55	3	24	0.6515
Pillai's Trace	0.06458115	0.55	3	24	0.6515
Hotelling-Lawley Trace	0.06903982	0.55	3	24	0.6515
Roy's Greatest Root	0.06903982	0.55	3	24	0.6515

P-value >0.05 로, Power, Eye 의 교호효과가 없다고 할 수 있다. 또한 Diet, Time 각각의 변수도 유의하지 않다고 보여진다. 따라서 두 변수 모두 효과가 없다.

3. RESULT

Vision 자료의 Anova, Manova 분석 결과 Eye * Power 의 p-value > 0.05로, 효과가 유의하지 않다고 할 수 있다. 따라서 Eye, Power 에 따라 Acuity에 미치는 효과가 없으며, Power levels: 6/6 6/18 6/36 6/60 와 Eye levels: left, right 가 시력에 미치는 효과가 없다.