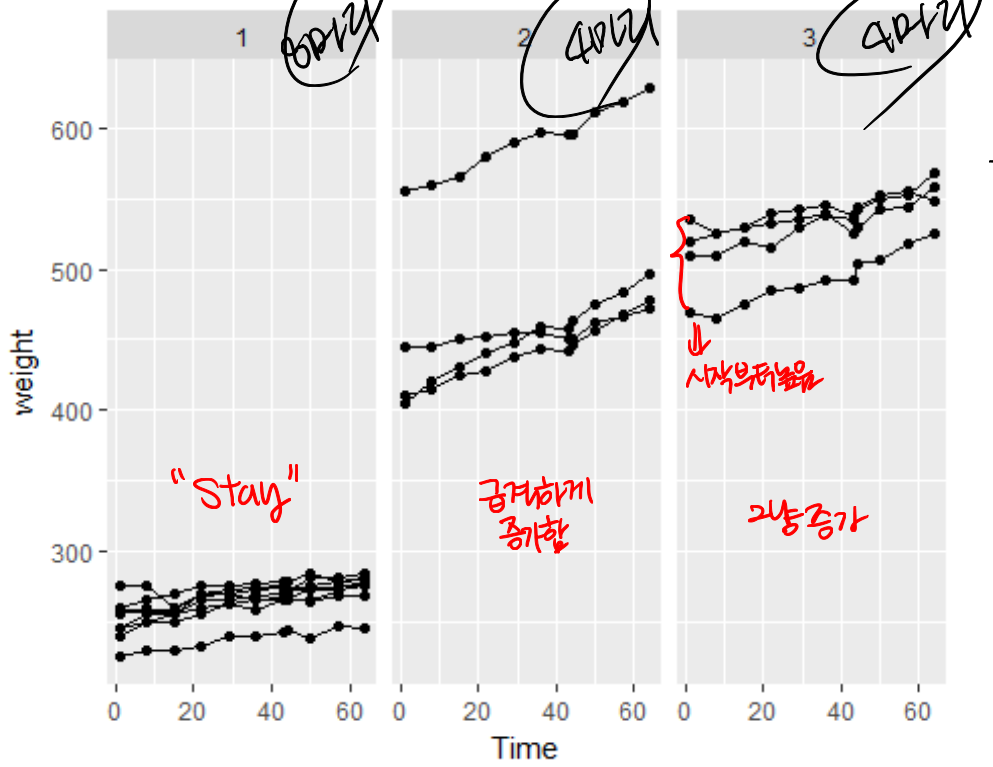
HW#2 - solution

1. R 의 nlme package 에 내장되어 있는 BodyWeight 자료를 이용하시오(자료설명은 help 이용).

1) 적절한 그림으로 자료를 살펴보시오.

```
library(nlme)
head(BodyWeight)

## Grouped Data: weight ~ Time | Rat
##   weight Time Rat Diet
## 1    240    1   1    1
## 2    250    8   1    1
## 3    255   15   1    1
## 4    260   22   1    1
## 5    262   29   1    1
## 6    258   36   1    1

library(ggplot2)
ggplot(BodyWeight,aes(Time,weight,group=Rat))+geom_line()+geom_point()+facet_
wrap(~Diet)
```



2) 적절한 방법으로 분석하여 결과를 해석하시오.

```
library(nlme)
# method 1      Mixed Effect : ANOVA                    Random effect
model1<-lme(weight~factor(Diet)*factor(Time),random=~1|factor(Rat),data = Bod
yWeight)
anova(model1)

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

```
DATA bodyweight;
 INFILE 'BodyWeight.csv' DELIMITER=',' FIRSTOBS=2;
 INPUT weight time rat diet @@;
RUN;

PROC GLM data = bodyweight;
  CLASS diet rat time;
  MODEL weight = diet rat(diet) time diet*time;
  RANDOM rat(diet);
  TEST H=diet E=rat(diet);
RUN;
```

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 I 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	23400.261	2340.026	58.84	<.0001
diet*time	20	4906.080	245.304	6.17	<.0001

다른결과나옴

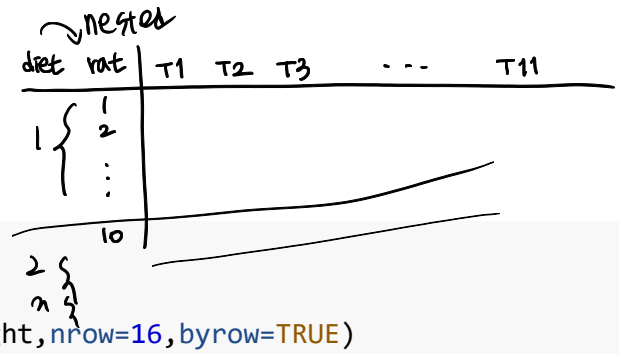
Source	DF	Type III SS	Mean Square	F Value	Pr > F
diet	2	2604049.733	1302024.866	32738.2	<.0001
rat(diet)	13	192185.670	14783.513	371.72	<.0001
time	10	27000.205	2700.020	67.89	<.0001
diet*time	20	4906.080	245.304	6.17	<.0001

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

제대로된 test




```r
# method 2  MANOVA : 다변량
library(car)

newweight<-matrix(BodyWeight$weight,nrow=16,byrow=TRUE)
colnames(newweight)<-paste("Time",names(table(BodyWeight$Time)),sep="")
BodyWeight1<-data.frame(Rat=1:16,Diet=rep(c(1,2,3),c(8,4,4)),newweight)
model2.lm<-lm(cbind(Time1,Time8,Time15,Time22,Time29,Time36,
                    Time43,Time44,Time50,Time57,Time64)~factor(Diet),data = B
odyWeight1)
time<-factor(colnames(newweight))
time.data<-data.frame(time=time)
model2<-Anova(model2.lm,idata=time.data,idesign=~time)
summary(model2)

## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1   0.99267 1759.905      1     13 2.8986e-15 ***
## Wilks             1   0.00733 1759.905      1     13 2.8986e-15 ***
## Hotelling-Lawley  1 135.37730 1759.905      1     13 2.8986e-15 ***
## Roy               1 135.37730 1759.905      1     13 2.8986e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Multivariate Tests: factor(Diet)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            2   0.93127 88.07276      2     13 2.7635e-08 ***
## Wilks             2   0.06873 88.07276      2     13 2.7635e-08 ***
## Hotelling-Lawley  2  13.54966 88.07276      2     13 2.7635e-08 ***
## Roy               2  13.54966 88.07276      2     13 2.7635e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
## Multivariate Tests: 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: factor(Diet):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 **
```

결과 비슷하게 나옴!

```
## 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
##
## Univariate Type II Repeated-Measures ANOVA Assuming Sphericity
##
##                      Sum Sq num Df Error SS den Df   F value    Pr(>F)
## (Intercept)       26017577      1   192186     13 1759.9049 2.899e-15 ***
## factor(Diet)       2604050      2   192186     13   88.0728 2.763e-08 ***
## time                 23400     10     5170    130   58.8378 < 2.2e-16 ***
## factor(Diet):time     4906     20     5170    130    6.1679 2.883e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Mauchly Tests for Sphericity
##
##                   Test statistic   p-value
## time                  6.8081e-08 5.357e-11
## factor(Diet):time     6.8081e-08 5.357e-11
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
##  for Departure from Sphericity
##
##                    GG eps Pr(>F[GG])
## time               0.18726    7.6e-10 ***
## factor(Diet):time 0.18726    0.00168 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                      HF eps   Pr(>F[HF])
## time               0.2176299 4.103351e-11
## factor(Diet):time 0.2176299 8.359944e-04
```

```
DATA bodyweight1;
 INFILE 'BodyWeight1.csv' DELIMITER=',' FIRSTOBS=2;
 INPUT rat diet Time1 Time8 Time15 Time22 Time29 Time36 Time43 Time44 Time50 Time57
Time64 @@;
RUN;

PROC GLM data = bodyweight1;
  CLASS diet;
  MODEL Time1 Time8 Time15 Time22 Time29 Time36 Time43 Time44 Time50 Time57 Time64 =
diet;
  REPEATED TIME PROFILE / PRINTE SUMMARY;
  QUIT;
RUN;
```

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.

The GLM Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
diet	2	2604049.733	1302024.866	88.07	<.0001
Error	13	192185.670	14783.513		

SAS 시스템

The GLM Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

						Adj Pr > F	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	G - G	H-F-L
TIME	10	27000.20455	2700.02045	67.89	<.0001	<.0001	<.0001
TIME*diet	20	4906.07955	245.30398	6.17	<.0001	0.0017	0.0008
Error(TIME)	130	5170.20455	39.77080				

Greenhouse-Geisser Epsilon	0.1873
Huynh-Feldt-Lecoutre Epsilon	0.2176

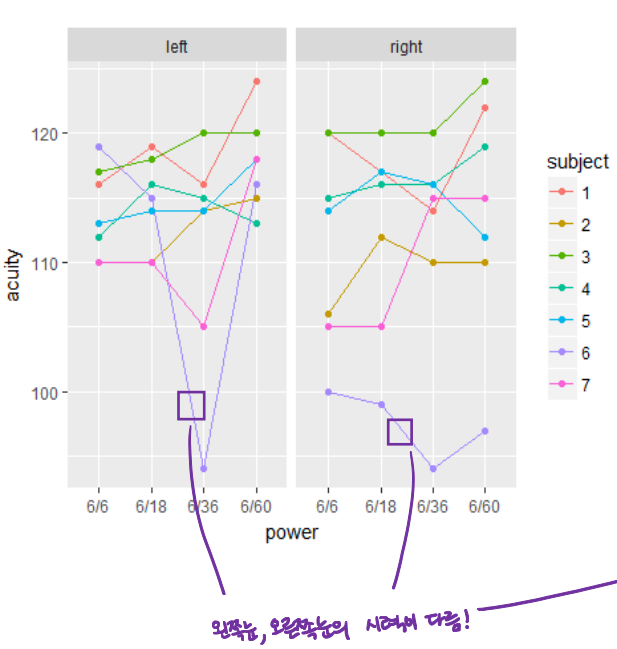
2. R 의 faraway package 에 내장되어 있는 vision 자료를 이용하시오(자료설명은 help 이용)

1) 적절한 그림으로 자료를 살펴보시오.

```
library(faraway)

## Warning: package 'faraway' was built under R version 3.4.4

##
## Attaching package: 'faraway'

## The following objects are masked from 'package:car':
##
##     logit, vif

head(vision)

##   acuity power   eye subject
## 1    116   6/6  left       1
## 2    119  6/18  left       1
## 3    116  6/36  left       1
## 4    124  6/60  left       1
## 5    120   6/6 right       1
## 6    117  6/18 right       1

ggplot(vision,aes(power,acuity,group=subject,color=subject))+geom_line()+face
t_wrap(~eye)+geom_point()

ggplot(vision,aes(power,acuity,color=eye,group=eye))+geom_line()+facet_wrap(~
subject)+geom_point()
```

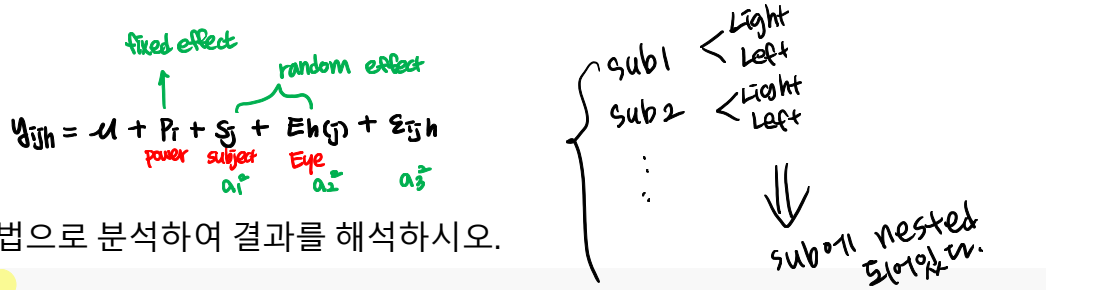
power	6	18	36	60	6	8	36	60
subject	R				L			
1								
2								
⋮								
7								

→ 한 사람에 대해서 repeated 자료



왼쪽눈, 오른쪽눈의 시력에 다름!

6/30 에서 자료 값이 갑자기 떨어짐!
초록색과 같은 분포형, 그냥 양 쪽눈의 시력이 다른 경우이지만 갑자기 한점에서 떨어지니까, 문제가 있는 자료 아닐까? 의심됨



2) 적절한 방법으로 분석하여 결과를 해석하시오.

```
library(lme4)
# method 1
model1 <- lmer(acuity~power+(1|subject)+(1|subject:eye),vision)
anova(model1)

## Analysis of Variance Table
##       Df Sum Sq Mean Sq F value
## power  3 140.77  46.923  2.8263

summary(model1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: acuity ~ power + (1 | subject) + (1 | subject:eye)
##    Data: vision
##
## REML criterion at convergence: 328.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4240 -0.3232  0.0109  0.4408  2.4662
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  subject:eye (Intercept) 10.27    3.205
##  subject     (Intercept) 21.53    4.640
##  Residual                16.60    4.075
## Number of obs: 56, groups:  subject:eye, 14; subject, 7
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 112.6429     2.2349  50.401
## power6/18     0.7857     1.5400   0.510
## power6/36    -1.0000     1.5400  -0.649
## power6/60     3.2857     1.5400   2.134
##
## Correlation of Fixed Effects:
##             (Intr) pw6/18 pw6/36
## power6/18 -0.345
## power6/36 -0.345  0.500
## power6/60 -0.345  0.500  0.500
```

```r
# method 2  MANOVA

library(nlme)
model2 <- lme(acuity~power,random=~1|subject/eye,vision)
anova(model2)

##             numDF denDF   F-value p-value
## (Intercept)     1    39 3132.9132  <.0001
## power           3    39    2.8263  0.0511

summary(model2)

## Linear mixed-effects model fit by REML
##  Data: vision
##        AIC      BIC    logLik
##   342.7098 356.3686 -164.3549
##
## Random effects:
##  Formula: ~1 | subject
##         (Intercept)
## StdDev:    4.639616
##
##  Formula: ~1 | eye %in% subject
##         (Intercept) Residual
## StdDev:    3.205242 4.074568
##
## Fixed effects: acuity ~ power
##                 Value Std.Error DF  t-value p-value
## (Intercept) 112.64286  2.234914 39 50.40143  0.0000
## power6/18     0.78571  1.540042 39  0.51019  0.6128
## power6/36    -1.00000  1.540042 39 -0.64933  0.5199
## power6/60     3.28571  1.540042 39  2.13352  0.0392
##  Correlation:
##           (Intr) pw6/18 pw6/36
## power6/18 -0.345
## power6/36 -0.345  0.500
## power6/60 -0.345  0.500  0.500
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -3.4240087 -0.3232125  0.0109485  0.4408116  2.4661858
##
## Number of Observations: 56
## Number of Groups:
##          subject eye %in% subject
##                7               14
```

subject에 nested된 eye

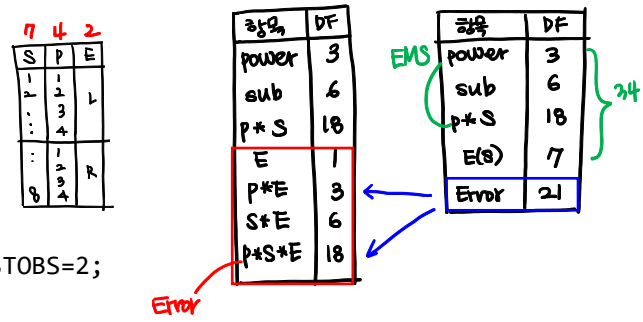
power 3 = 4-1
subject 6 = 7-1
Eye(S) 7 = 1+(7-1) = 7
Error 39 → p*S가 에러

$\sigma_1 = 4.63$

σ_2 σ_3

→ eye가 subject에 nested된것을 표현

p*S, p*E, p*S*E가 랜덤변량 값!




```
DATA vision;

 INFILE 'vision.csv' DELIMITER=',' FIRSTOBS=2;
 INPUT acuity power $ eye $ subject @@;
RUN;

PROC GLM data = vision;
  CLASS subject power eye;
  MODEL acuity = power subject power*subject eye(subject) / SS3;
  RANDOM subject eye(subject);
  TEST H=power E=subject*power;
RUN;
```

The GLM Procedure

Dependent Variable: acuity

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	34	2299.928571	67.644958	5.23	<.0001
Error	21	271.625000	12.934524		
Corrected Total	55	2571.553571			

R-Square	Coeff Var	Root MSE	acuity Mean
0.894373	3.171182	3.596460	113.4107

Source	DF	Type III SS	Mean Square	F Value	Pr > F
power	3	140.767857	46.922619	3.63	0.0297
subject	6	1379.428571	229.904762	17.77	<.0001
subject*power	18	375.857143	20.880952	1.61	0.1459
eye(subject)	7	403.875000	57.696429	4.46	0.0035

The GLM Procedure

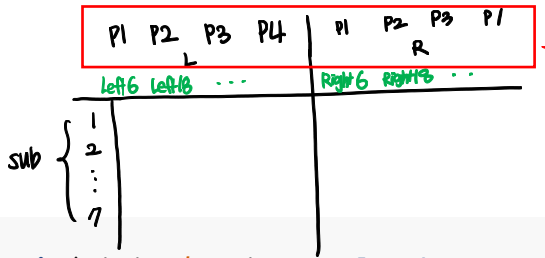
Source	Type III Expected Mean Square
power	Var(Error) + Q(power,subject*power)
subject	Var(Error) + 4 Var(eye(subject)) + 8 Var(subject) + Q(subject*power)
subject*power	Var(Error) + Q(subject*power)
eye(subject)	Var(Error) + 4 Var(eye(subject))

Dependent Variable: acuity

Tests of Hypotheses Using the Type III MS for subject*power as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
power	3	140.7678571	46.9226190	2.25	0.1177

$F(3,18) = \frac{\text{power (3)}}{\text{power*s (18)}}$ ↔ Rotate $F(3,39)$




```r
# method 3
acuity.m<-matrix(vision$acuity,ncol=8,byrow=TRUE)
colnames(acuity.m)<-paste(rep(c("L","R"),each=4),c(6,18,36,60),sep="")
vision2<-data.frame(ID=1:7,acuity.m)

power<-factor(c(6,18,36,60))
LR<-factor(c("Left","Right"))
powerRL.data<-data.frame(eye=factor(rep(LR,each=4)),power=factor(rep(power,
2)))
model3.lm<-lm(cbind(L6,L18,L36,L60,R6,R18,R36,R60)~1,data = vision2)
model3<-Anova(model3.lm,idata=powerRL.data,idesign=~power+eye+power*eye)

## Note: model has only an intercept; equivalent type-III tests substituted.

summary(model3)
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1    0.9981 3132.912      1      6 2.1841e-09 ***
## Wilks             1    0.0019 3132.912      1      6 2.1841e-09 ***
## Hotelling-Lawley  1  522.1520 3132.912      1      6 2.1841e-09 ***
## Roy               1  522.1520 3132.912      1      6 2.1841e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
## Multivariate Tests: power
##                  Df test stat approx F num Df den Df    Pr(>F)
## Pillai            1  0.941581 21.49012      3      4 0.0062731 **
## Wilks             1  0.058419 21.49012      3      4 0.0062731 **
## Hotelling-Lawley  1 16.117587 21.49012      3      4 0.0062731 **
## Roy               1 16.117587 21.49012      3      4 0.0062731 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Multivariate Tests: eye
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1  0.115002 0.7796763      1      6 0.41122
## Wilks             1  0.884998 0.7796763      1      6 0.41122
## Hotelling-Lawley  1  0.129946 0.7796763      1      6 0.41122
## Roy               1  0.129946 0.7796763      1      6 0.41122
##
## ------------------------------------------
## Multivariate Tests: power:eye
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.2929031 0.5523111      3      4 0.67335
## Wilks             1 0.7070969 0.5523111      3      4 0.67335
```

```
## Hotelling-Lawley  1 0.4142333 0.5523111      3      4 0.67335
## Roy               1 0.4142333 0.5523111      3      4 0.67335
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##             Sum Sq num Df Error SS den Df   F value    Pr(>F)
## (Intercept) 720271      1  1379.43      6 3132.9123 2.184e-09 ***
## power          141      3   375.86     18    2.2471    0.1177
## eye             46      1   357.43      6    0.7797    0.4112
## power:eye       41      3   231.00     18    1.0552    0.3925
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Mauchly Tests for Sphericity
##
##           Test statistic  p-value
## power           0.025172 0.004729
## power:eye       0.230506 0.235620
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
##  for Departure from Sphericity
##
##            GG eps Pr(>F[GG])
## power     0.49661     0.1665
## power:eye 0.54927     0.3700
##
##              HF eps Pr(>F[HF])
## power     0.6229404  0.1528236
## power:eye 0.7302526  0.3818914
```

```
DATA vision2;
INFILE 'vision2.csv' DELIMITER=',' FIRSTOBS=2;
 INPUT subject L6 L18 L36 L60 R6 R18 R36 R60  @@;
RUN;

PROC GLM data =vision2;
  CLASS subject;
  MODEL  L6 L18 L36 L60 R6 R18 R36 R60 = ;
  REPEATED EYE 2, POWER 4 / PRINTE SUMMARY;
  QUIT;
RUN;
```

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	5	0.0266065	17.125616	0.0043
Orthogonal Components	5	0.0251719	17.387347	0.0038

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

S=1 M=0.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.05841945	21.49	3	4	0.0063
Pillai's Trace	0.94158055	21.49	3	4	0.0063
Hotelling-Lawley Trace	16.11758703	21.49	3	4	0.0063
Roy's Greatest Root	16.11758703	21.49	3	4	0.0063

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
EYE	1	46.4464286	46.4464286	0.78	0.4112
Error(EYE)	6	357.4285714	59.5714286		

Source	DF	Type III SS	Mean Square	F Value	Pr > F	Adj Pr > F	
						G - G	H - F
POWER	3	140.7678571	46.9226190	2.25	0.1177	0.1665	0.1528
Error(POWER)	18	375.8571429	20.8809524				

Greenhouse-Geisser Epsilon	0.4966
Huynh-Feldt Epsilon	0.6229

Source	DF	Type III SS	Mean Square	F Value	Pr > F	Adj Pr > F	
						G - G	H - F
EYE*POWER	3	40.6250000	13.5416667	1.06	0.3925	0.3700	0.3819
Error(EYE*POWER)	18	231.0000000	12.8333333				

Greenhouse-Geisser Epsilon	0.5493
Huynh-Feldt Epsilon	0.7303