

# 바이오통계 과제

-제 6장 반복측정 자료분석-



2019314199

통계학과

김동환

## 목차

- I. 데이터 소개
- II. 분석
- III. 코드

## I. 데이터 소개

이 연구의 다음 3가지 가설을 알아보는 것이다.

1. 성별 간에 턱뼈 성장의 평균 프로파일이 유의한 차이가 있는지.
2. 8세에서 14세까지 나이에 따른 턱뼈 크기에 차이가 있는지(8세 이후로도 계속해서 턱뼈가 자라는지).
3. 성별과 나이 간에 교호작용이 있는지.

이 자료는 16명의 남자 어린이들과 11명의 여자 어린이들을 대상으로 나이에 따라 턱뼈가 커가는 정도를 기록한 자료이며 뇌하수체와 익돌상악의 틈사이의 길이를 mm단위로 측정하여 기록하였다.

4회 턱뼈크기 반복측정자료

성별	개체	나이				성별	개체	나이			
		8세	10세	12세	14세			8세	10세	12세	14세
남자	1	26.0	25.0	29.0	31.0	여자	1	21.0	20.0	21.5	23.0
	2	21.5	22.5	23.0	26.5		2	21.0	21.5	24.0	25.5
	3	23.0	22.5	24.0	27.5		3	20.5	24.0	24.5	26.0
	4	25.5	27.5	26.5	27.0		4	23.5	24.5	25.0	26.5
	5	20.0	23.5	22.5	26.0		5	21.5	23.0	22.5	23.5
	6	24.5	25.5	27.0	28.5		6	20.0	21.0	21.0	22.5
	7	22.0	22.0	24.5	26.5		7	21.5	22.5	23.0	25.0
	8	24.0	21.5	24.5	25.5		8	23.0	23.0	23.5	24.0
	9	23.0	20.5	31.0	26.0		9	20.0	21.0	22.0	21.5
	10	27.5	28.0	31.0	31.5		10	16.5	19.0	19.0	19.5
	11	23.0	23.0	23.5	25.0		11	24.5	25.0	28.0	28.0
	12	21.0	23.5	24.0	28.0						
	13	17.0	24.5	26.0	29.5						
	14	22.5	25.5	25.5	26.0						
	15	23.0	24.5	26.0	30.0						
	16	22.0	21.5	23.5	25.0						

출처 : Potthoff, R. F., Roy, S. N. (1964). A generalized multivariate analysis of variance model usefully especially for growth curve problems. Biometrika, 51(3), 313-326.

## II. 분석

### 1) 구형성 조건 검정

Partial Correlation Coefficients from the Error SSCP Matrix / Prob >  r				
df = 25	a8	a10	a12	a14
a8	1.000000	0.570699 0.0023	0.661320 0.0002	0.521583 0.0063
a10	0.570699 0.0023	1.000000	0.5663167 0.0027	0.726216 <.0001
a12	0.661320 0.0002	0.5663167 0.0027	1.000000	0.728098 <.0001
a14	0.521583 0.0063	0.726216 <.0001	0.728098 <.0001	1.000000

#### 1-1) 부분상관계수 행렬

From above Partial correlation coefficient matrix,

Partial correlation (age8,age10) = 0.57699

Partial correlation (age8,age12) = 0.661320

Partial correlation (age8,age14) = 0.521583

Partial correlation (age10,age12) = 0.5663167

Partial correlation (age10,age14) = 0.726216

Partial correlation (age12,age14) = 0.728098

나이에 따라 부분상관계수가 차이가 있긴 하지만 그 정도가 크지 않음.

따라서 나이에 대한 오차항의 공분산 행렬이 구형성 조건을 만족할 가능성이 높음.

## 1-2) 오차항의 공분산 행렬에 대한 구형성 검정(Sphericity Tests)

$H_0$ : 분산- 공분산 행렬이 구형성을 만족한다.

$H_1$ : not  $H_0$

Based on Mauchly's Criterion  $\sim \chi^2$

$$W = \frac{\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} (R_{ij} - R_{i.} R_{.j} / R_{..})^2}{(I-1)(J-1) - I}$$

$R_{ij}$  는  $i$  번째 행과  $j$  번째 열의 분산- 공분산 행렬의 원소,  $R_{i.}$  은  $i$  번째 행의 합,  $R_{.j}$  는  $j$  번째 열의 합,  $R_{..}$  는 행렬의 대각선 원소의 합

Chi - Square statistics :

Transformed Variates : 16.559181,  $P$  - value : 0.0057

Orthogonal Components: 7.2929515,  $P$  - value : 0.1997

under  $\alpha = 0.05$  cannot reject  $H_0$

검정 결과  $P$ -value가 모두 0.05보다 큼으로 구형성 가정을 만족한다. Transformed Variates일 경우  $P$ 값이 0.05에 상당히 근접하여 애매할 수 있으나, Orthogonal Components의 결과가 매우 유의하지 않음으로 구형성 가정을 만족한다고 판단한다. 따라서 일변량 분석을 진행한다.

## 2) 일변량 분석 결과

### 2-1) 성별의 효과가 유의한지 검정

$H_0$ : 성별에 따라 턱뼈 성장에 차이가 없다.

$H_1$ : not  $H_0$

$$F_{statistics} = \frac{MS_{gender}}{MSE}$$

$$F - value : \frac{140.4548569}{15.1165909} = 9.29, P - value : 0.0054$$

under  $\alpha=0.05$  cannot reject  $H_0$

따라서 유의수준 0.05하에서  $H_0$ 를 기각할 수 없다.

엄밀하게 따지면  $P$ -value가 0.05보다 큼으로 귀무가설을 기각할 수 없으나, 그 값이 0.05에 매우 근접한 것을 알 수 있다. 즉, 연구자의 해석에 따라 성별에 따라 턱뼈 성장에 차이가 있다고 판단 할 수 있다.

## 2-2) 나이의 효과가 유의한지 검정

$H_0$ : 나이에 대한 차이가 없다.

$H_1$ : not  $H_0$

$$F_{\text{statistics}} = \frac{MS_{\text{age}}}{MSE(\text{age})}$$

$$F - \text{value} : \frac{69.8123246}{1.9750379} = 35.35, P - \text{value} < .0001$$

under  $\alpha=0.05$  can reject  $H_0$

따라서 유의수준 0.05하에서  $H_0$ 를 기각한다.

검정 결과 나이에 따라서 턱뼈 크기에 차이가 난다.

## 2-3) 나이와 성별의 교호작용이 존재하는지 검정

$H_0$ : 나이와 성별에 교호작용이 없다.

$H_1$ : not  $H_0$

$$F_{\text{statistics}} = \frac{MS_{\text{age*gender}}}{MSE(\text{age})}$$

$$F - \text{value} : \frac{4.66417165}{1.9750379} = 2.36, P - \text{value} : 0.0781$$

under  $\alpha=0.05$  cannot reject  $H_0$

따라서 유의수준 0.05하에서  $H_0$ 를 기각할 수 없다.

검정 결과 나이에 따라서 성별에 대한 턱뼈 크기에 유의한 차이가 없다.

### 3) Contrast statement 출력결과

Contrast Variable: a\_1 (나이 8살 VS 나이 14살)

Source	DF	Type 3 SS	Mean Square	F Value	Pr > F
Mean	1	323.8300189	323.8300189	65.64	<.0001
gender	1	29.8300189	29.8300189	6.05	0.0212
Error	25	123.3366477	4.9334659		

Contrast Variable: a\_2 (나이 10살 VS 나이 14살)

Source	DF	Type 3 SS	Mean Square	F Value	Pr > F
Mean	1	173.3072917	173.3072917	72.38	<.0001
gender	1	30.3072917	30.3072917	12.66	0.0015
Error	25	59.8593750	2.3943750		

Contrast Variable: a\_2 (나이 12살 VS 나이 14살)

Source	DF	Type 3 SS	Mean Square	F Value	Pr > F
Mean	1	49.29629630	49.29629630	15.70	0.0005
gender	1	3.66666667	3.66666667	1.17	0.2902
Error	25	78.50000000	3.14000000		

$H_0$ : 14세의 턱뼈 크기 대비 n세의 성별에 대해 차이가 없다.

n은 8,10,12임

$H_1$ : not  $H_0$

$$F_{\text{statistics}} = \frac{MS_{\text{gender}}}{MSE}$$

$$n = 8, F - \text{value} : \frac{29.8300189}{4.9334659} = 6.05, P - \text{value} : 0.0212$$

$$n = 10, F - \text{value} : \frac{30.3072917}{2.3943750} = 12.66, P - \text{value} : 0.0015$$

$$n = 12, F - \text{value} : \frac{3.66666667}{3.14000000} = 1.17, P - \text{value} : 0.2902$$

n=8,10

under alpha=0.05 can reject  $H_0$

두 경우 (n=8,10) 유의수준 0.05 하에서  $H_0$ 를 기각할 수 있다.

n=12

under alpha=0.05 cannot reject  $H_0$

n=12 인 경우 유의수준 0.05 하에서  $H_0$ 를 기각할 수 없다.

따라서 14세와 8세의 턱뼈 크기의 차이는 성별에 따라 유의한 차이가 있다.

### III. 코드

```
data height;
input a8 a10 a12 a14 gender $ @@;
cards;
26.0 25.0 29.0 31.0 male 21.5 22.5 23.0 26.5 male
23.0 22.5 24.0 27.5 male 25.5 27.5 26.5 27.0 male
20.0 23.5 22.5 26.0 male 24.5 25.5 27.0 28.5 male
22.0 22.0 24.5 26.5 male 24.0 21.5 24.5 25.5 male
23.0 20.5 31.0 26.0 male 27.5 28.0 31.0 31.5 male
23.0 23.0 23.5 25.0 male 21.5 23.5 24.0 28.0 male
17.0 24.5 26.0 29.5 male 22.5 25.5 25.5 26.0 male
23.0 24.5 26.0 30.0 male 22.0 21.5 23.5 25.0 male
21.0 20.0 21.5 23.0 female 21.0 21.5 24.0 25.5 female
20.5 24.0 24.5 26.0 female 23.5 24.5 25.0 26.5 female
21.5 23.0 22.5 23.5 female 20.0 21.0 21.0 22.5 female
21.5 22.5 23.0 25.0 female 23.0 23.0 23.5 24.0 female
20.0 21.0 22.0 21.5 female 16.5 19.0 19.0 19.5 female
24.5 25.0 28.0 28.0 female
;
proc glm data=height;
class gender;
model a8 a10 a12 a14 = gender / nouni ss3;
repeated a 4 (8 10 12 14) contrast(4) / summary printe;
run;
```

Partial Correlation Coefficients from the Error SSCP Matrix / Prob >  r				
DF = 25	a8	a10	a12	a14
a8	1.000000	0.570699 0.0023	0.661320 0.0002	0.521583 0.0063
a10	0.570699 0.0023	1.000000	0.563167 0.0027	0.726216 <.0001
a12	0.661320 0.0002	0.563167 0.0027	1.000000	0.728098 <.0001
a14	0.521583 0.0063	0.726216 <.0001	0.728098 <.0001	1.000000

Sphericity Tests				
Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	5	0.4998695	16.449181	0.0057
Orthogonal Components	5	0.7353334	7.2929515	0.1997

Contrast Variable: a\_1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	323.8300189	323.8300189	65.64	<.0001
gender	1	29.8300189	29.8300189	6.05	0.0212
Error	25	123.3366477	4.9334659		

Contrast Variable: a\_2

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	173.3072917	173.3072917	72.38	<.0001
gender	1	30.3072917	30.3072917	12.66	0.0015
Error	25	59.8593750	2.3943750		

Contrast Variable: a\_3

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	49.29629630	49.29629630	15.70	0.0005
gender	1	3.66666667	3.66666667	1.17	0.2902
Error	25	78.50000000	3.14000000		

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	Adj Pr > F	
						G - G	H-F-L
a	3	209.4369739	69.8123246	35.35	<.0001	<.0001	<.0001
a*gender	3	13.9925295	4.6641765	2.36	0.0781	0.0878	0.0797
Error(a)	75	148.1278409	1.9750379				

Greenhouse-Geisser Epsilon	0.8672
Huynh-Feldt-Lecoutre Epsilon	0.9769

### SAS 시스템

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
gender	1	140.4648569	140.4648569	9.29	0.0054
Error	25	377.9147727	15.1165909		



R

```
sink('grow.txt')
```

```
cat('gender e t tw f  
m 26.0 25.0 29.0 31.0
```

```
m 21.5 22.5 23.0 26.5
```

```
m 23.0 22.5 24.0 27.5
```

```
m 25.5 27.5 26.5 27.0
```

```
m 20.0 23.5 22.5 26.0
```

```
m 24.5 25.5 27.0 28.5
```

```
m 22.0 22.0 24.5 26.5
```

```
m 24.0 21.5 24.5 25.5
```

```
m 23.0 20.5 31.0 26.0
```

```
m 27.5 28.0 31.0 31.5
```

```
m 23.0 23.0 23.5 25.0
```

```
m 21.5 23.5 24.0 28.0
```

```
m 17.0 24.5 26.0 29.5
```

```
m 22.5 25.5 25.5 26.0
```

```
m 23.0 24.5 26.0 30.0
```

```
m 22.0 21.5 23.5 25.0
```

```
f 21.0 20.0 21.5 23.0
```

```
f 21.0 21.5 24.0 25.5
```

```
f 20.5 24.0 24.5 26.0
```

```
f 23.5 24.5 25.0 26.5
```

```
f 21.5 23.0 22.5 23.5
```

```
f 20.0 21.0 21.0 22.5
```

```
f 21.5 22.5 23.0 25.0
```

```
f 23.0 23.0 23.5 24.0
```

```
f 20.0 21.0 22.0 21.5
```

```
f 16.5 19.0 19.0 19.5
```

```
f 24.5 25.0 28.0 28.0
```

```
')
```

```
sink()
```

```
gr<-read.table('grow.txt',sep=' ',header=T)
```

```
gr$id<-1:27; gr$id<-factor(gr$id)
```

```
gr$gender<-factor(gr$gender)
```

```
gr<-melt(gr,id.vars=c('id','gender'))
```

```
colnames(gr)<-c('id','gender','age','value')
```

```
interaction.plot(gr$age,gr$gender,gr$value,fun=mean,type='b',pch=c(2,4),legend=F,
```

```
col=c(3,6),xlab='age',ylab='Mean of value')
```

```
fit1<-aov(value~gender*age+Error(id/age),data=gr)
```

```
with(gr,pairwise.t.test(value,age,paired=T))
```

```
> summary(fit1)
```

Error: id

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender	1	140.5	140.46	9.292	0.00538 **
Residuals	25	377.9	15.12		

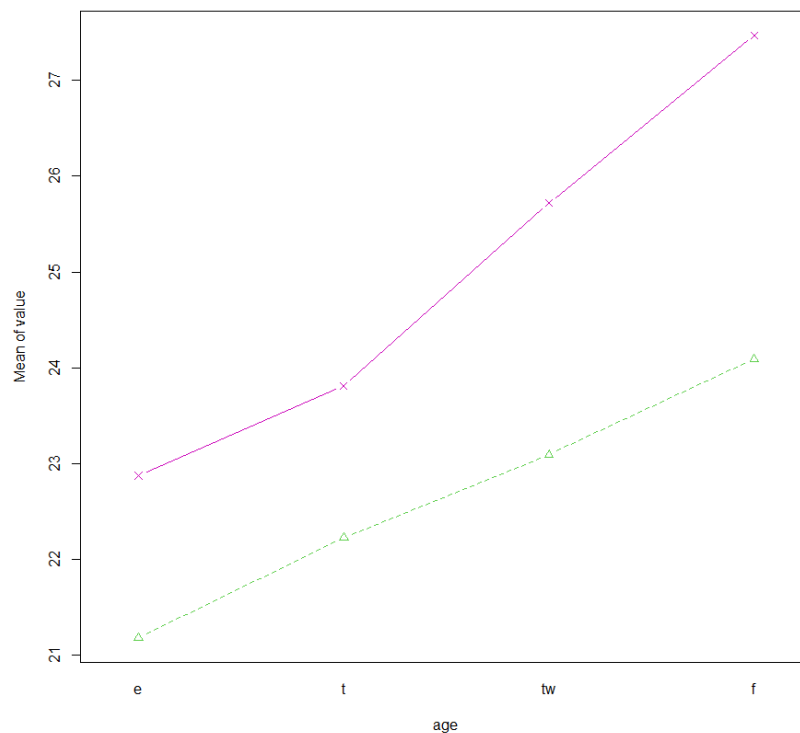
---

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

Error: id:age

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	3	237.19	79.06	40.032	1.49e-15 ***
gender:age	3	13.99	4.66	2.362	0.0781 .
Residuals	75	148.13	1.98		

```
> interaction.plot(gr$age,gr$gender,gr$value,fun=mean,type='b',pch=c(2,4),legend=F,  
+                  col=c(3,6),xlab='age',ylab='Mean of value')
```



```
> with(gr,pairwise.t.test(value,age,paired=T,p.adjust.method = 'bonferroni'))
```

Pairwise comparisons using paired t tests

data: value and age

	e	t	tw
t	0.1028	-	-
tw	6.7e-06	0.0105	-
f	2.3e-08	3.8e-08	0.0017

P value adjustment method: bonferroni