

바이오통계 과제

-제 5장 공분산 분석-



2019314199

통계학과

김동환

목차

- I. 데이터 소개
- II. 분석
- III. 코드

I. 데이터 소개

이 데이터는 편두통 예방제인 벤조디아제핀 부류의 약물 실험에 대한 내용으로, 해당 약물들의 효과를 조사하기 위해 행해진 실험의 결과이다. 실험 약물은 알프라조람(A), 트리아졸람(T), 플라시보 약물인 설탕(S)이다. 이 연구의 목적은 약물 복용 시 특정 기억에 어떤 영향이 있는지 알아보는 것이다. 기억 점수는 이미지 회상검사를 통해 산출하였다. 여기서 기억 점수 란 예를 들어, 긍정적인 기억을 가지고 있을 때, 그 기억을 얼마나 긍정적으로 느끼고 있는지에 대한 것이다. 즉 약물 복용 후 점수가 낮아졌다면, 해당 기억이 약화되었다고 볼 수 있다. 또한 복용 뒤 점수가 복용 전 초기 점수에 영향을 받았을 것이라 생각하여 초기 점수도 같이 측정하였다.

관측 번호	복용 약물	X(초기 점수)	X(복용 뒤 점수)
1	A	63.5	58.2
2	A	59.7	52.1
3	A	47	38.3
4	A	76.3	68.8
5	S	39.6	37.7
6	S	36	35.8
7	S	30.7	27.1
8	S	48.3	47
9	T	48.3	44.1
10	T	50.9	41.8
11	T	76.1	73.1
12	T	100	96

출처 : Ahn, S., et al. (2020). Memory Test on Drugged Islanders Data. Kaggle.
<https://www.kaggle.com/datasets/steveahn/memory-test-on-drugged-islanders-data>

II. 분석

1) 교호작용 효과 검정

Model : $y = \text{trt} + x + x*\text{trt}$

H_0 : 처리간 공변량 효과가 동일하다. ($\beta_{(x*\text{trt})} = 0$)

H_1 : not H_0

$$F\text{statistics} = \frac{MS_{\text{interaction}}}{MSE}$$

$$F - \text{value} : \frac{0.097352}{4.359462} = 0.02, P - \text{value} : 0.9780$$

under $\alpha=0.05$ cannot reject H_0

공변량과 처리 사이에 교호작용이 존재하지 않음으로

처리간 공변량 효과가 동일하다는 가정을 만족한다. 따라서 공분산 분석을 시행한다.

2) 모형 적합도

Model : $y = \text{trt} + x$

H_0 : 모형이 적합하지않다.

H_1 : not H_0

$$F\text{statistics} = \frac{MS_{\text{model}}}{MSE}$$

$$F - \text{value} : \frac{1380.098397}{3.293935} = 418.98, P - \text{value} < 0.0001$$

under $\alpha=0.05$ can reject H_0

따라서 유의수준 0.05하에서 모델의 유의하다.

현 모형이 자료에 적합하다.

2) 공변량의 효과가 유의한지 검정

Model : $y = \text{trt} + x$

H_0 : 공변량의 효과가 유의하지 않다. ($\beta_x = 0$)

H_1 : not H_0

$$F\text{statistics} = \frac{MS_{\text{covariate}}}{MSE}$$

$$F - \text{value} : \frac{2655.248524}{3.293935} = 806.10, P - \text{value} < 0.0001$$

under $\alpha=0.05$ can reject H_0

귀무가설을 기각하게 되어 공변량의 효과는 유의함을 알 수 있다.

공변량일 경우 제 1종 제곱합과 제 3종 제곱합에 따른 차이가 없음.

두 경우 모두 $p\text{-value} < .0001$ 로 유의함.

3) 약의 효과가 유의한지 검정

Model : $y = \text{trt} + x$

H_0 : 약의 효과가 유의하지 않다. ($\beta_{\text{trt}} = 0$)

H_1 : not H_0

$$F\text{statistics} = \frac{MS_{\text{trt}}}{MSE}$$

$$\text{제1종 제곱합 일 경우 } F - \text{value} : \frac{742.523333}{3.293935} = 225.42, P - \text{value} < 0.0001$$

$$\text{제3종 제곱합 일 경우 } F - \text{value} : \frac{32.724888}{3.293935} = 9.93, P - \text{value} : 0.0068$$

약의 효과일 경우 제1종 제곱합과 제 3종 제곱합에 따라 그 결과가 상이함.

1) 공변량의 효과를 보정하지 않았을 때 약의 효과:

F-value : 225.42, P-value < .0001으로 유의수준 0.05하에서 유의함.

2) 공변량의 효과를 보정하였을 때 약의 효과:

F-value : 9.93, P-value : 0.0068으로 유의수준 0.05하에서 유의하지 않음.

유의수준 0.05에서 초기 기억 점수를 보정하였을 때, 약의 type에 따라 복용 후 기억 점수의 모평균에는 차이가 없다.

3) 약의 type에 따른 차이에 대한 분석

parameter	Estimate	t value	Pr > t
Intercept	-4.026217587	-2.36	0.0457
A(알피아조람)	-6.878068540	-4.46	0.0021
T(트리아조람)	-5.102098725	-2.99	0.0174
S(플라시보-설탕)	0.000000000	.	.
x(초기 점수)	1.058893081	28.39	<.0001

플라시보 약물인 설탕을 0으로 놓았을 때, 플라시보와의 차이에 대한 유의성 검정 결과는 위의 표와 같다.

1) 각 모수에 대한 검정:

$$H_0 : \alpha_1 = \alpha_3$$

t-value : -4.46, p-value : 0.0021

A와 S사이에는 유의한 차이가 존재한다.

$$H_0 : \alpha_2 = \alpha_3$$

t-value : -2.99, p-value : 0.0174

T와 S사이에는 유의한 차이가 존재한다.

공분산 모형식

$$\hat{y}_{1j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}x_{1j} = -4.026 - 6.878 + 1.059x_{1j}$$

$$\hat{y}_{2j} = \hat{\beta}_0 + \hat{\alpha}_2 + \hat{\beta}x_{2j} = -4.026 - 5.102 + 1.059x_{2j}$$

$$\hat{y}_{3j} = \hat{\beta}_0 + \hat{\alpha}_3 + \hat{\beta}x_{3j} = -4.026 + 0.0 + 1.059x_{3j}$$

2) 보정 평균

trt	y LSMEAN	LSMEAN number
A	48.7819872	1
T	50.5579570	2
S	55.6600558	3

공분산 모형식으로부터 LSMEAN을 구할 수 있다.

$$\overline{y_{adj1}} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}\bar{x}_{..} = -4.026 - 6.878 + 1.059 \times 56.367 = 48.789$$

$$\overline{y_{adj2}} = \hat{\beta}_0 + \hat{\alpha}_2 + \hat{\beta}\bar{x}_{..} = -4.026 - 5.102 + 1.059 \times 56.367 = 50.564$$

$$\overline{y_{adj3}} = \hat{\beta}_0 + \hat{\alpha}_3 + \hat{\beta}\bar{x}_{..} = -4.026 + 0.0 + 1.059 \times 56.367 = 55.667$$

보정평균은 모든 처리에서 공변량의 평균이 같다고 했을 때의 반응변수의 평균값이므로, 추정된 회귀식에서 공변량 전체 평균값(56.367)을 대입하여 보정평균을 구할 수 있다

3) 각 처리 별 보정 평균의 차이

	1	2	3
1	t-value p	-1.35453 0.2126	-4.45728 0.0021
2	1.354529 0.2126	t-value p	-2.98912 0.0174
3	4.45728 0.0021	2.989118 0.0174	t-value p

$$H_0 : LSMEAN\ 1 = LSMEAN\ 3$$

t-value : -4.45728, p-value: 0.0021

A는 S와 유의한 차이가 있다.

$$H_0 : LSMEAN\ 2 = LSMEAN\ 3$$

t-value : -2.98912, p-value: 0.0174

T는 S와 유의한 차이가 있다.

Ⅲ. 코드

```
data memory;
input trt $ x y ;
cards;
A 63.5 58.2
A 59.7 52.1
A 47 38.3
A 76.3 68.8
T 48.3 44.1
T 50.9 41.8
T 76.1 73.1
T 100 96
S 39.6 37.7
S 36 35.8
S 30.7 27.1
S 48.3 47
run;

proc glm data=memory;
class trt;
model y=trt x trt*x;
run;

proc glm data=memory;
class trt;
model y=trt x/SOLUTION;
lsmeans trt/TDIFF;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	1.638483	0.819242	0.19	0.8334
x	1	1143.581888	1143.581888	262.32	<.0001
x*trt	2	0.194705	0.097352	0.02	0.9780

SAS 시스템

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4140.295190	1380.098397	418.98	<.0001
Error	8	26.351476	3.293935		
Corrected Total	11	4166.646667			

R-Square	Coeff Var	Root MSE	y Mean
0.993676	3.512748	1.814920	51.66667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	2	1485.046667	742.523333	225.42	<.0001
x	1	2655.248524	2655.248524	806.10	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	65.449777	32.724888	9.93	0.0068
x	1	2655.248524	2655.248524	806.10	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-4.026217587	B 1.70332850	-2.36	0.0457
trt A	-6.878068540	B 1.54310898	-4.46	0.0021
trt T	-5.102098725	B 1.70689107	-2.99	0.0174
trt S	0.000000000	B .	.	.
x	1.058893081	0.03729555	28.39	<.0001

SAS 시스템

The GLM Procedure Least Squares Means

trt	y LSMEAN	LSMEAN Number
A	48.7819872	1
T	50.5579570	2
S	55.6600558	3

Least Squares Means for Effect trt t for H0: LSMean(i)=LSMean(j) / Pr > |t| Dependent Variable: y

i/j	1	2	3
1		-1.35453 0.2126	-4.45728 0.0021
2	1.354529 0.2126		-2.98912 0.0174
3	4.45728 0.0021	2.989118 0.0174	

R

```
sink('drug.txt')
```

```
cat('trt before after
```

```
S 39.6 37.7
```

```
S 36 35.8
```

```
S 30.7 27.1
```

```
S 48.3 47
```

```
A 63.5 58.2
```

```
A 59.7 52.1
```

```
A 47 38.3
```

```
A 76.3 68.8
```

```
T 48.3 44.1
```

```
T 50.9 41.8
```

```
T 76.1 73.1
```

```
T 100 96
```

```
')
```

```
sink()
```

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

```
drug$trt1<-factor(drug$trt,levels=c('S','A','T'))
```

```
###교호작용 효과 확인
```

```
fit1<-aov(after~trt1+before+trt1*before,data=drug); summary(fit1)
```

```
###각 x변수 효과 확인
```

```
###type1 ss
```

```
fit2<-aov(after~trt1+before,data=drug); summary(fit2)
```

```
###type3 ss
```

```
type_3<-Anova(fit2,type='III'); type_3
```

```
###각 X변수 추정치 확인
```

```
fit3<-glm(after~trt1+before,data=drug); summary(fit3)
```

```
###전체 모델 적합도
```

```
MStotal<-sum((drug$after-mean(drug$after))^2)/(fit3$df.null-fit3$df.residual)
```

```
MSE<-deviance(fit3)/fit3$df.residual; f_value<-MStotal/MSE
```

```
p_value<-1-pf(f_value,fit3$df.null-fit3$df.residual,fit3$df.residual)
```

```
sprintf('f_value : %s p_value : %s',f_value,p_value)
```

```
###보정평균 확인
```

```
library(lsmmeans)
```

```
lm1<-lsmeans(fit3,list(pairwise~trt1)); lm1
```

```
> summary(fit1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt1	2	1485.0	742.5	170.325	5.19e-06 ***
before	1	2655.2	2655.2	609.077	2.91e-07 ***
trt1:before	2	0.2	0.1	0.022	0.978
Residuals	6	26.2	4.4		

```
---
```

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

```
> summary(fit2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt1	2	1485.0	742.5	225.4	9.24e-08 ***
before	1	2655.2	2655.2	806.1	2.56e-09 ***
Residuals	8	26.4	3.3		

```
---
```

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

```
> ###type3 ss
```

```
> type_3<-Anova(fit2,type='III'); type_3
```

```
Anova Table (Type III tests)
```

```
Response: after
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	18.40	1	5.5872	0.045691 *
trt1	65.45	2	9.9349	0.006789 **
before	2655.25	1	806.1024	2.56e-09 ***
Residuals	26.35	8		

```
---
```

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

```
> summary(fit3)
```

```
Call:
```

```
glm(formula = after ~ trt1 + before, data = drug)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-2.9693	-0.8431	-0.2088	1.6614	2.0838

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.0262	1.7033	-2.364	0.04569 *
trt1A	-6.8781	1.5431	-4.457	0.00212 **
trt1T	-5.1021	1.7069	-2.989	0.01736 *
before	1.0589	0.0373	28.392	2.56e-09 ***

```
---
```

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

```
(Dispersion parameter for gaussian family taken to be 3.293935)
```

```
Null deviance: 4166.647 on 11 degrees of freedom  
Residual deviance: 26.351 on 8 degrees of freedom  
AIC: 53.494
```

```
Number of Fisher Scoring iterations: 2
```

```

> lsmeans(fit3,~trt1)
  trt1 lsmean    SE df lower.CL upper.CL
  S      55.7 1.123  8     53.1     58.2
  A      48.8 0.928  8     46.6     50.9
  T      50.6 1.019  8     48.2     52.9

> lmlist<-lsmeans(fit3,list(pairwise~trt1)); lmlist
$`lsmeans of trt1`
  trt1 lsmean    SE df lower.CL upper.CL
  S      55.7 1.123  8     53.1     58.2
  A      48.8 0.928  8     46.6     50.9
  T      50.6 1.019  8     48.2     52.9

```

Confidence level used: 0.95

```

$`pairwise differences of trt1`
  1      estimate    SE df t.ratio p.value
S - A      6.88 1.54  8    4.457 0.0053
S - T      5.10 1.71  8    2.989 0.0412
A - T     -1.78 1.31  8   -1.355 0.4071

```

P value adjustment: tukey method for comparing a family of 3 estimates

SAS와 결과가 동일함(multiple comparison일 경우 t값은 동일한데, p값이 다름.)