HW5

임상시험자료분석 II

182STG27

임지연

Data

Trial 자료를 이용하여 분석하시오.

	TRT	CENTER	PAT	SEX	AGE	SCORE
1	A	1	101	M	55	5
2	A	1	107	F	44	21
3	A	1	112	F	31	10
4	A	1	118	F	39	25
5	A	1	123	F	57	55
6	A	1	128	F	48	8
7	A	1	135	M	27	40
8	A	2	203	M	42	22
9	A	2	210	F	35	25
10	A	2	216	F	42	15

Analysis

1) AGE가 SCORE와 선형관계가 있는지 분석하여라. (Treatment group간 차이는 고려하지 않는다)

R

1. CODE

```
library(ggplot2); library(gridExtra)
library(tidyverse); library(dplyr)
library(HH); library(lsmeans)

trial = read_csv("C:/Users/jeeyeon/Desktop/data/ex3-1.csv")
mylm1 = lm(SCORE ~ AGE , data = trial)
summary(mylm1)
anova(mylm1)
```

2. OUPUT

```
Call:
lm(formula = SCORE ~ AGE, data = trial)

Residuals:
    Min      1Q  Median      3Q     Max
-40.665 -26.164  -4.015  20.491  64.228

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.5811    10.5473   1.477    0.143
AGE           0.3919     0.2385   1.643    0.103

Residual standard error: 28.51 on 98 degrees of freedom
Multiple R-squared:  0.02682,   Adjusted R-squared:  0.01689
F-statistic: 2.701 on 1 and 98 DF,  p-value: 0.1035
```

3. RESULT

R, SAS 결과 AGE의 P-value = 0.103 > 0.05 로 AGE 변수의 효과가 없다는 것을 알 수 있다. 따라서 AGE 변수는 SCORE에 유의한 영향을 준다고 할 증거가 없다.

SAS

1. CODE

```
data TRIAL;
infile "C:₩Users₩jeeyeon₩Desktop₩data₩ex3-1.csv"
DELIMITER=',' FIRSTOBS=2;
input TRT $ CENTER PAT SEX $ AGE SCORE;run;

PROC SORT DATA= TRIAL;
BY TRT AGE SCORE;
PROC PRINT DATA=TRIAL;
VAR PAT AGE SCORE;RUN;

PROC GLM DATA=TRIAL;
MODEL SCORE = AGE / P CLM SS1; RUN;
```

2. OUPUT

The GLM Procedure

Dependent Variable: SCORE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2196.18275	2196.18275	2.70	0.1035
Error	98	79681.52725	813.07681		
Corrected Total	99	81877.71000			

R-Square	Coeff Var	Root MSE	SCORE Mean
0.026823	88.36226	28.51450	32.27000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
AGE	1	2196.182746	2196.182746	2.70	0.1035

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	15.58112135	10.54726104	1.48	0.1428
AGE	0.39194173	0.23848060	1.64	0.1035

3. RESULT

R, SAS 결과 AGE의 P-value = 0.103 > 0.05 로 AGE 변수의 효과가 없다는 것을 알 수 있다. 따라서 AGE 변수는 SCORE에 유의한 영향을 준다고 할 증거가 없다.

2) AGE와 SCORE 의 선형관계에서 Treatment group간에 slope는 같다고 가정할 때 intercept의 차이가 있는가를 분석하여라. 이 모형에서 age로 보정했을 때의 각 treatment group의 평균 반응을 추정하시오.

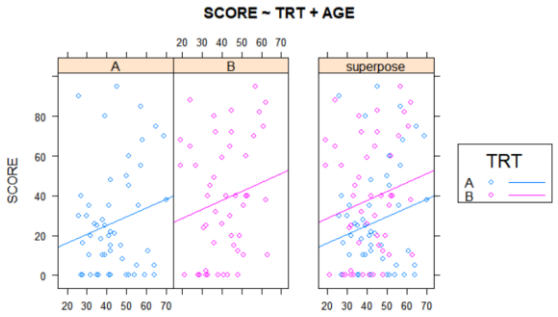
R

1. CODE

```
mylm2 = lm(SCORE ~ AGE + TRT , data = trial)
summary(mylm2)
anova(mylm2)

trial = trial %>% mutate(TRT = as.factor(TRT))
ancova(SCORE ~ TRT + AGE, data=trial)
```

2. OUPUT




```
Call:
lm(formula = SCORE ~ AGE + TRT, data = trial)

Residuals:
    Min      1Q  Median      3Q     Max
-41.301 -22.468  -4.153  21.470  71.221

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.2035    10.9739   0.656   0.5131
AGE           0.4452     0.2348   1.896   0.0609 .
TRTB         12.7276     5.6191   2.265   0.0257 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.93 on 97 degrees of freedom
Multiple R-squared:  0.07571,   Adjusted R-squared:  0.05665
F-statistic: 3.973 on 2 and 97 DF,  p-value: 0.02196



Analysis of Variance Table

Response: SCORE
          Df Sum Sq Mean Sq F value  Pr(>F)
TRT        1   3394  3393.6  4.3497 0.03964 *
AGE        1   2805  2805.3  3.5957 0.06091 .
Residuals 97  75679   780.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SAS

1. CODE

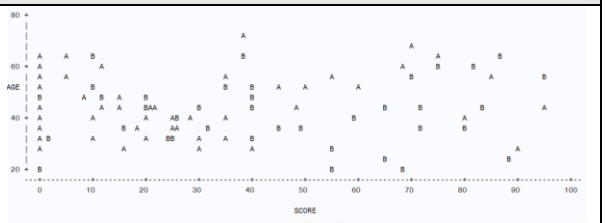
```
PROC PLOT VPERCENT= 45 DATA =   TRIAL;
PLOT AGE*SCORE = TRT;RUN;

PROC MEANS MEAN STD N DATA=TRIAL;
BY TRT;
VAR SCORE AGE;RUN;

PROC GLM DATA = TRIAL;
CLASS TRT;
MODEL SCORE = TRT AGE / SS3; RUN;

PROC GLM DATA=TRIAL;
CLASS TRT;
MODEL SCORE = TRT AGE / SOLUTION;
LSMEANS TRT / PDIFF STDERR; RUN;
```

2. OUPUT



TRT=A

변수	평균	표준편차	N
SCORE	26.6730769	26.9507329	52
AGE	43.7307692	12.2187824	52

TRT=B

변수	평균	표준편차	N
SCORE	38.3333333	29.6937083	48
AGE	41.3333333	11.7949382	48

The GLM Procedure

Dependent Variable: SCORE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6198.94404	3099.47202	3.97	0.0220
Error	97	75678.76596	780.19346		
Corrected Total	99	81877.71000			

R-Square	Coeff Var	Root MSE	SCORE Mean
0.075710	86.55700	27.93194	32.27000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	1	4002.761295	4002.761295	5.13	0.0257
AGE	1	2805.343015	2805.343015	3.60	0.0609

The GLM Procedure
Least Squares Means

TRT	SCORE LSMEAN	Standard Error	H0:LSMEAN=0 Pr > |t|	H0:LSMean1=LSMean2 Pr > |t|
A	26.1607366	3.8828755	<.0001	0.0257
B	38.8883687	4.0422402	<.0001	

3. RESULT

AGE와 SCORE 의 선형관계에서 Treatment group간에 slope는 같다고 가정할 때 intercept의 차이가 있는가를 분석해 본 결과, TRT 변수의 P-value값은 0.0257 (< 0.05) 이므로 효과가 있다고 할 수 있다. 따라서 **intercept에 차이가 있다**고 할 수 있다.

또한 age로 보정했을 때의 각 treatment group p-value < 0.05로 유의하다고 할 수 있다. 따라서 age로 보정했을 때 trt의 효과가 있다고 할 수 있다.

3) AGE와 SCORE의 선형관계에서 Treatment group 간에 slope가 같은지를 분석하시오.

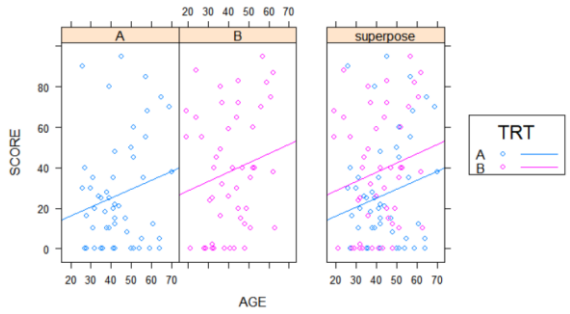
1. CODE

```
mylm3 = lm(SCORE ~ AGE * TRT, data =trial)
summary(mylm3)
anova(mylm3)
```

```
PROC PLOT VPERCENT= 45 DATA =  TRIAL;
PLOT AGE*SCORE = TRT;
RUN;

PROC GLM DATA = TRIAL;
CLASS TRT;
MODEL SCORE = TRT AGE TRT * AGE / SS3;
RUN;
```

2. OUPUT

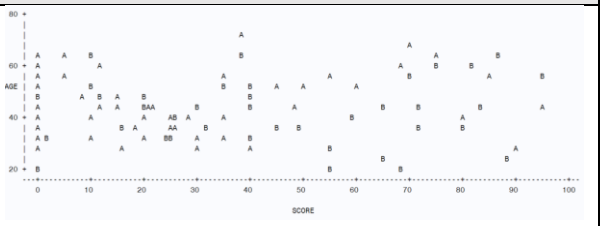



```
Call:
lm(formula = SCORE ~ AGE * TRT, data = trial)

Residuals:
   Min     1Q Median     3Q    Max
-41.36 -22.38  -4.14  21.48  71.08

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.54548   14.59968   0.517    0.606
AGE          0.43739    0.32176   1.359    0.177
TRTB        12.00917   20.86974   0.575    0.566
AGE:TRTB     0.01693    0.47338   0.036    0.972

Residual standard error: 28.08 on 96 degrees of freedom
Multiple R-squared:  0.07572,   Adjusted R-squared:  0.04684
F-statistic: 2.622 on 3 and 96 DF,  p-value: 0.05509
```



The GLM Procedure

Dependent Variable: SCORE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6199.95214	2066.65071	2.62	0.0551
Error	96	75677.75786	788.30998		
Corrected Total	99	81877.71000			

R-Square	Coeff Var	Root MSE	SCORE Mean
0.075722	87.00607	28.07686	32.27000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	1	261.029209	261.029209	0.33	0.5663
AGE	1	2797.206368	2797.206368	3.55	0.0626
AGE*TRT	1	1.008101	1.008101	0.00	0.9715

3. RESULT

Slope가 같은지 살펴보기 위해서 AGE*TRT 항을 본 결과 P-value = 0.972 (>0.05)로 효과가 없는 것으로 나타났다. 따라서 **trt 간의 slope가 같다**고 할 수 있다.

4) AGE와 study center에 대해 보정한 후 Treatment group간 차이를 분석하여라. (단, interaction은 없다고 가정한다.)

1. CODE

```
mylm4 = lm(SCORE ~ AGE + TRT + CENTER , data = trial)
summary(mylm4)
anova(mylm4)

lsmeans(mylm4, ~TRT)
```

```
PROC GLM DATA = TRIAL;
CLASS TRT;
MODEL SCORE = TRT CENTER / SS3;
RUN;

PROC GLM DATA=TRIAL;
CLASS TRT;
MODEL SCORE = TRT CENTER / SOLUTION;
LSMEANS TRT / PDIFF STDERR;
RUN;
```

2. OUPUT

```
 TRT    lsmean       SE df lower.CL upper.CL
 A    26.09517 3.888272 96 18.37701 33.81332
 B    38.95940 4.047921 96 30.92435 46.99446

Confidence level used: 0.95
```

The GLM Procedure
Least Squares Means

TRT	SCORE LSMEAN	Standard Error	H0:LSMEAN=0 Pr > |t|	H0:LSMean1=LSMean2 Pr > |t|
A	26.5260154	3.9180449	<.0001	0.0370
B	38.4926500	4.0781993	<.0001	

3. RESULT

AGE와 study center에 대해 보정한 후 Treatment group간 차이를 살펴본 결과 p-value= 0.037(<0.05)로 trt효과가 있다고 할 수 있다. 따라서 우리의 y 변수인 score는 trt 그룹에 따라 score의 차이가 있다는 결론을 낼 수 있다.