연습문제

6-1

data exam.strong;

input group $ str @@;

cards;

15% 7 15% 7 15% 15 15% 11 15% 9

20% 12 20% 17 20% 12 20% 18 20% 18

25% 14 25% 18 25% 18 25% 19 25% 19

30% 19 30% 25 30% 22 30% 19 30% 23

35% 7 35% 10 35% 11 35% 15 35% 11

;

run;

proc anova data=exam.strong;

class group;

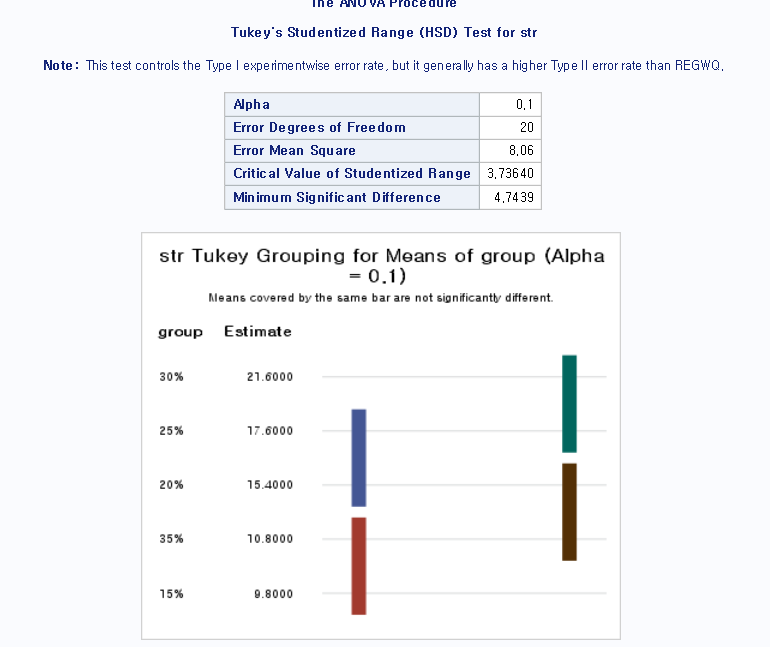
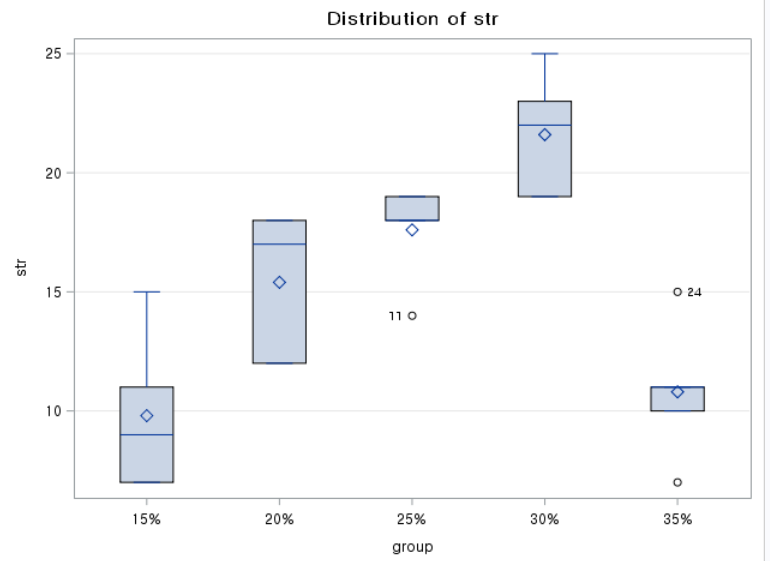
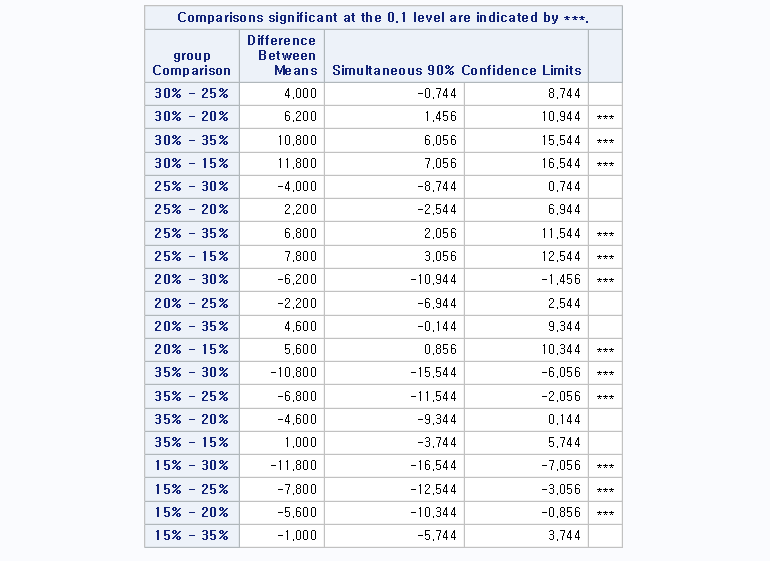
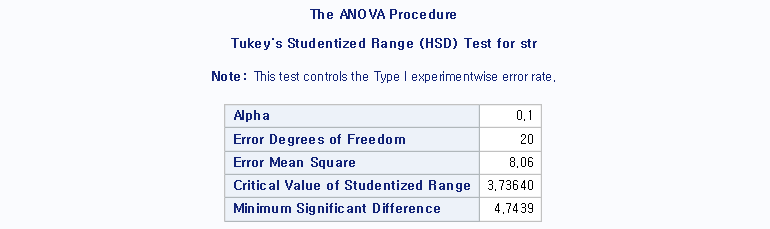
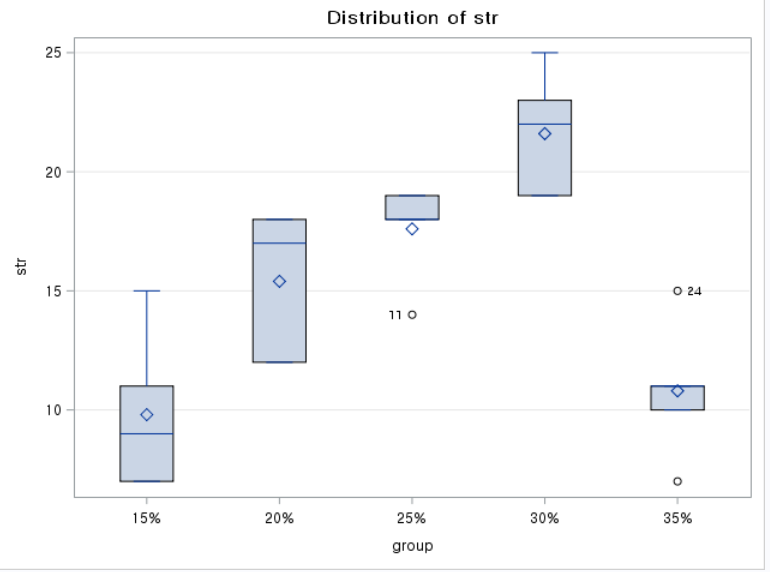
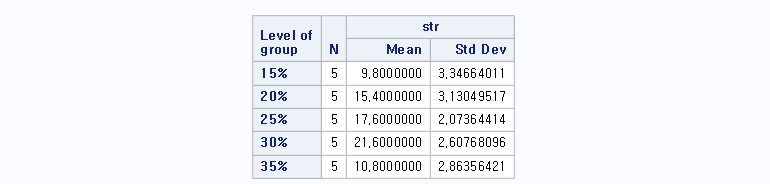
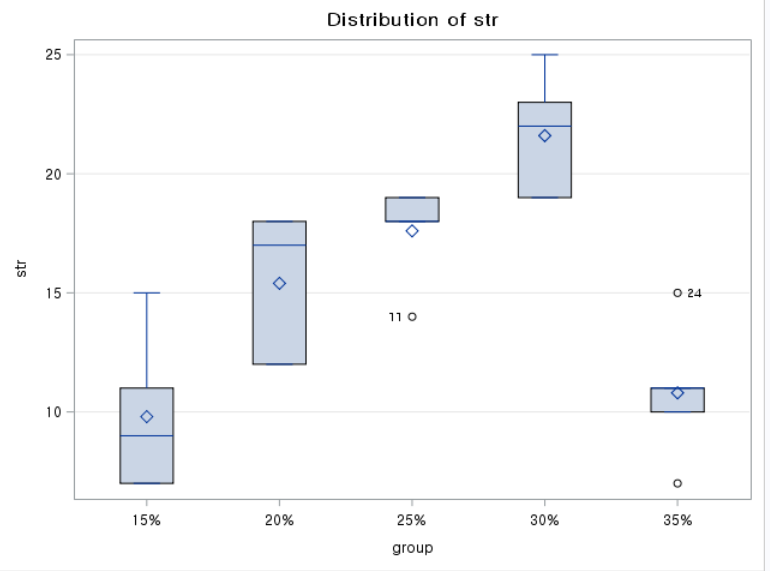
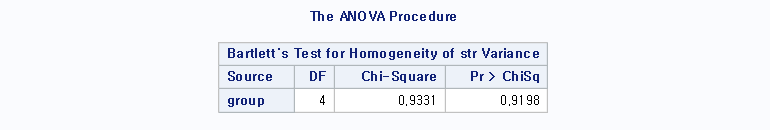
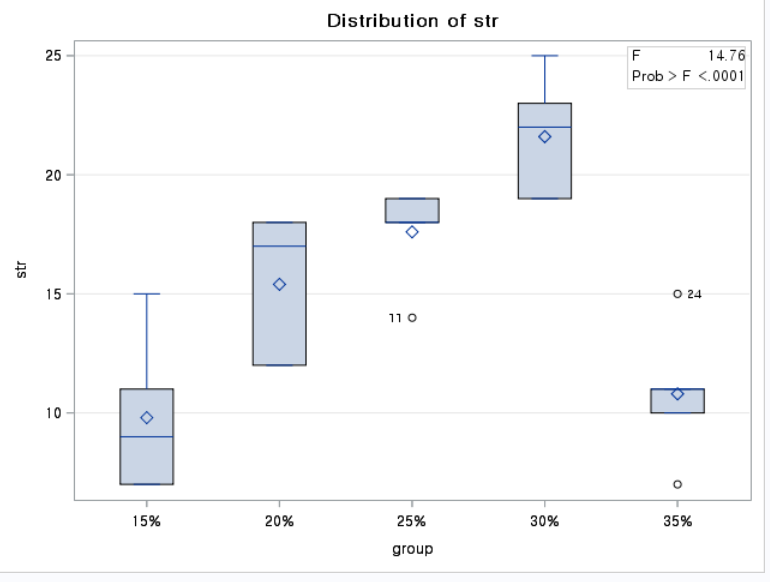
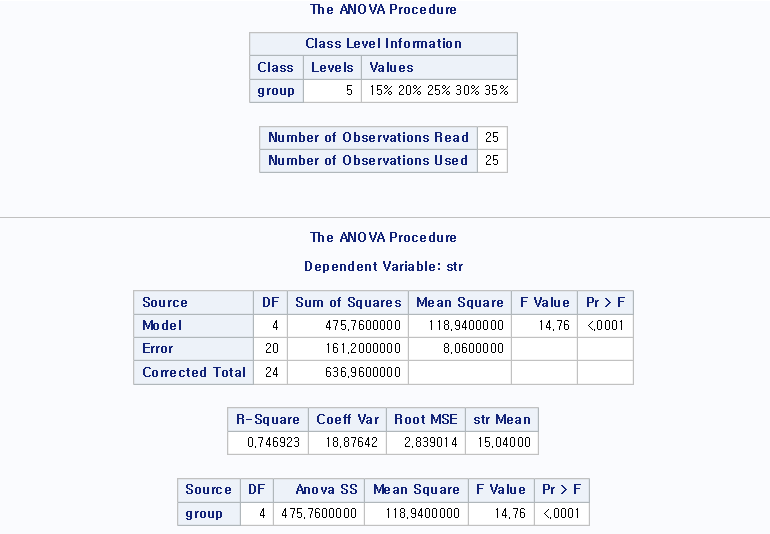
model str = group;

means group / hovtest = bartleft;

means group / tukey cldiff alpha=0.10;

means group / tukey lines alpha=0.10;

run;



해설

결과를 살펴보면 각 면 함유량에 따른 평균의 차이가 유의하다는 결론을 알 수 있다. 즉 면 함유량에 따라 평균 장력강도에 차이가 있다고 할 수 있다. 하지만 다중 비료 부분을 살펴보면 면 함유량에 따라 평균 장력강도에 차이가 있다고 할 수 있으려면 최소 10%정도는 차이가 있어야 함을 알 수 있다

6-3

data exam.concent;

input group $ concentration @@;

cards;

A1 46.8 A1 58.0 A1 51.4 A1 61.0 A1 45.8

A2 51.2 A2 62.4 A2 30.6 A2 46.0 A2 48.8

A3 50.2 A3 38.2 A3 46.8 A3 26.7 A3 22.7

A4 21.4 A4 22.1 A4 28.2 A4 42.7 A4 25.2

;

run;

proc anova data=exam.concent;

class group;

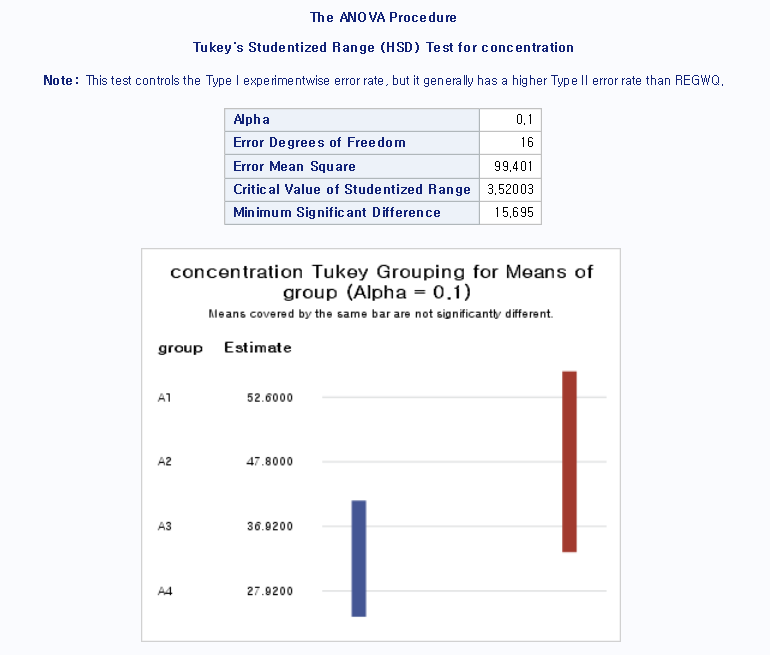
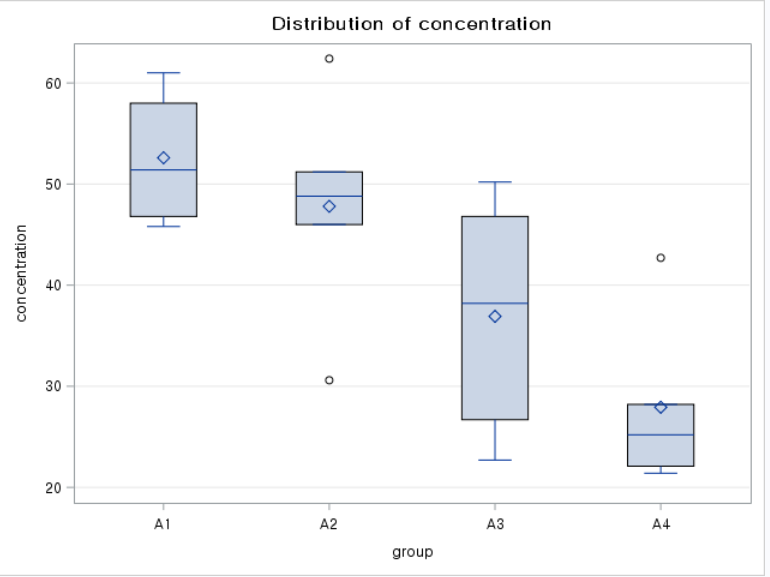
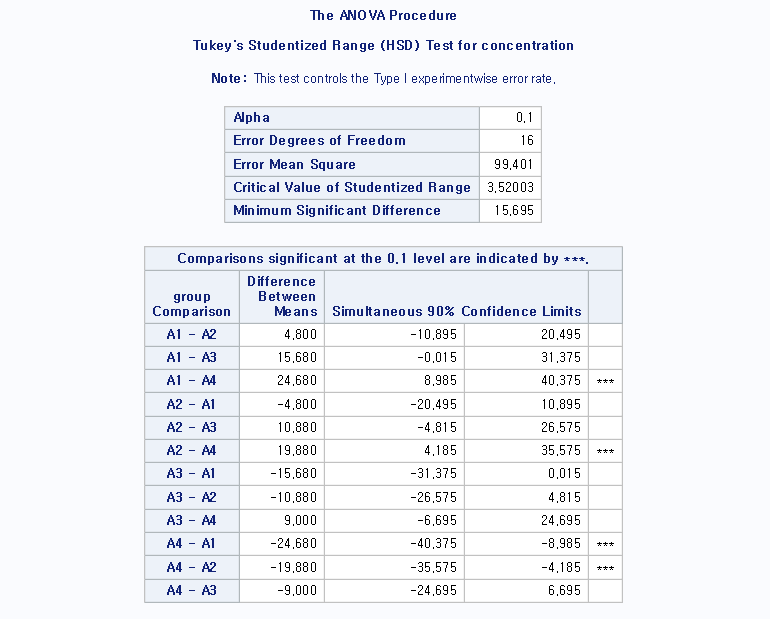
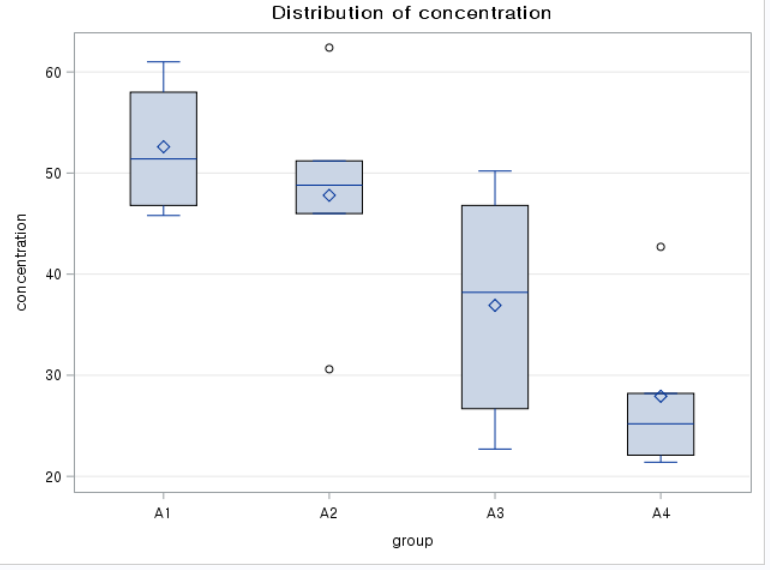
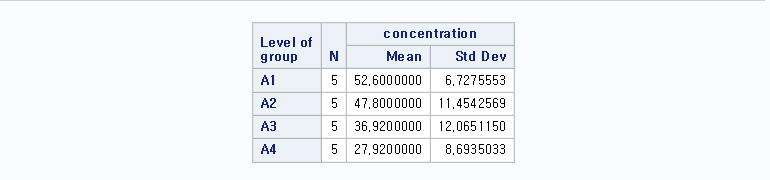
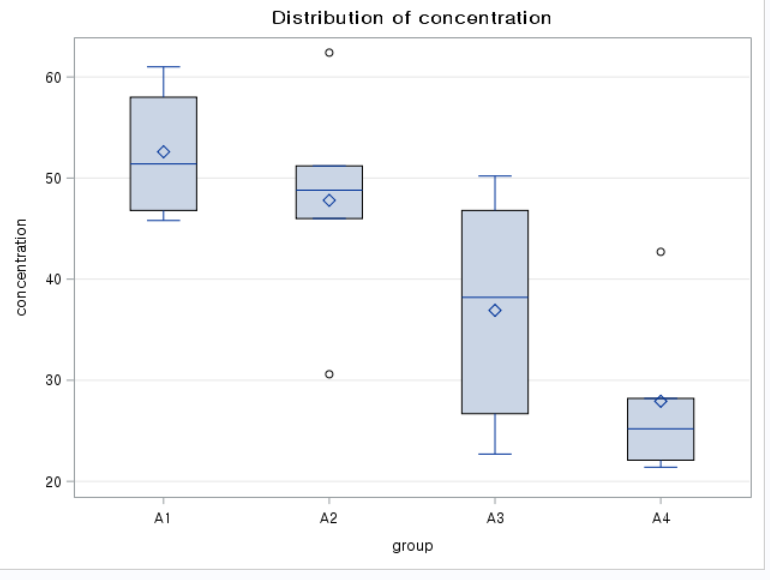
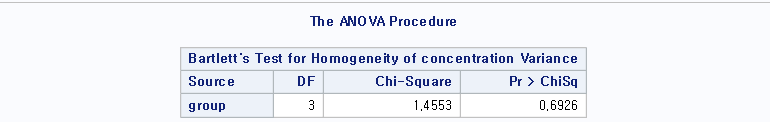
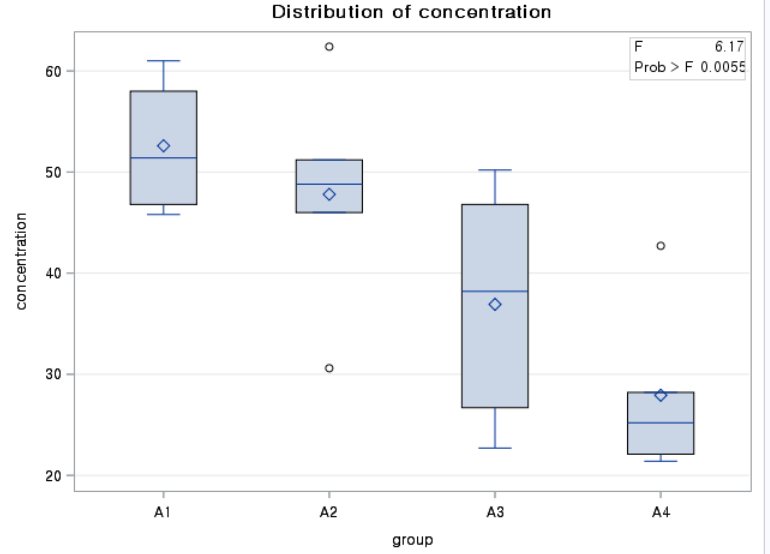
model concentration = group;

means group / hovtest = bartleft;

means group / tukey cldiff alpha=0.10;

means group / tukey lines alpha=0.10;

run;



해설

결과를 살펴보면 가공시의 처리액의 농도에 따른 차이가 유의하다는 것을 알 수 있다

즉 가공시의 처리액의 농도에 따라 평균 직물의 인장강도가 차이가 있다고 할 수 있다 하지만 다중비교 부분을 살펴보면 농도의 차이가 최소 1%은 되어야 인장강도에 차이가 있다고 할 수 있다는 것을 알 수 있다

6-5

data exam.b\_count;

input color $ count @@;

cards;

A 45 A 59 A 48 A 46 A 38 A 47

B 21 B 12 B 14 B 17 B 13 B 17

C 37 C 32 C 15 C 25 C 39 C 41

D 16 D 11 D 20 D 21 D 14 D 7

;

run;

proc anova data=exam.b\_count;

class color;

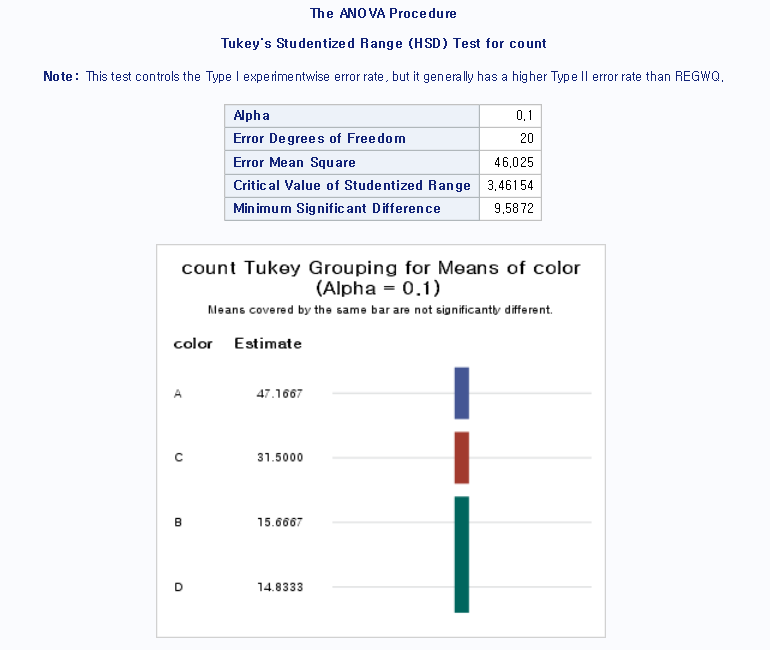
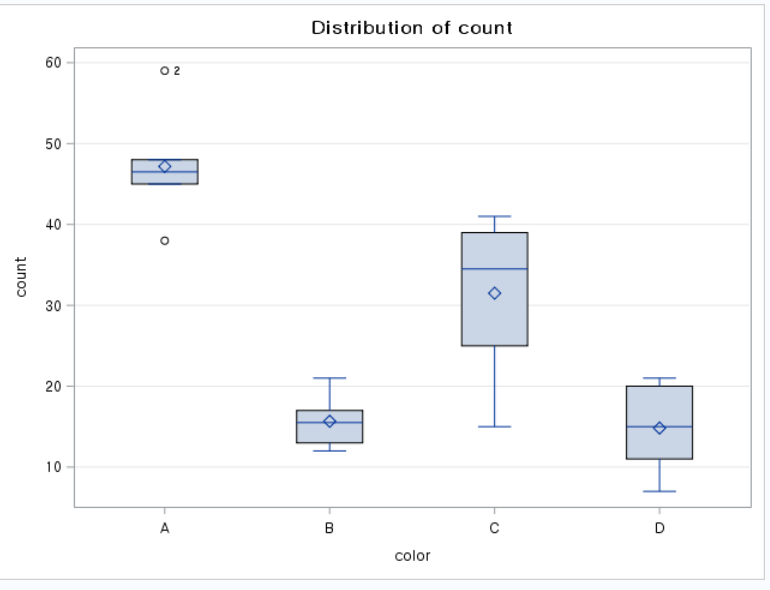
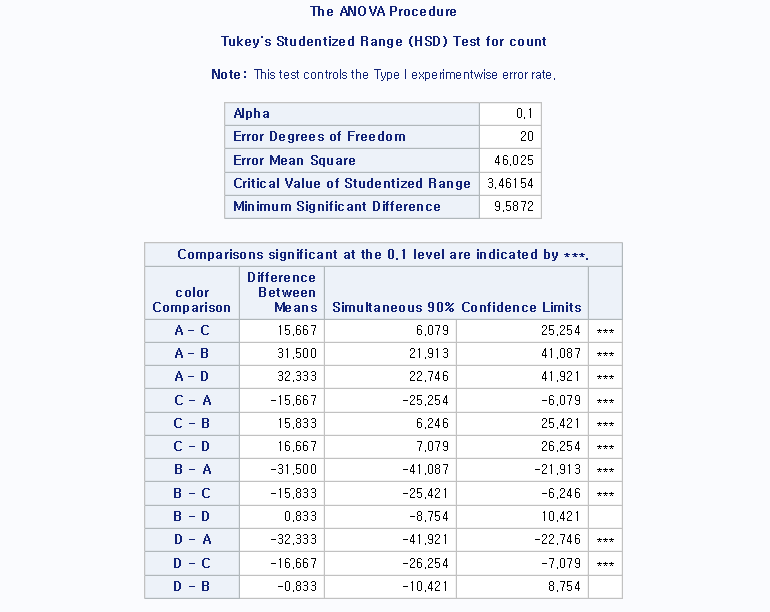
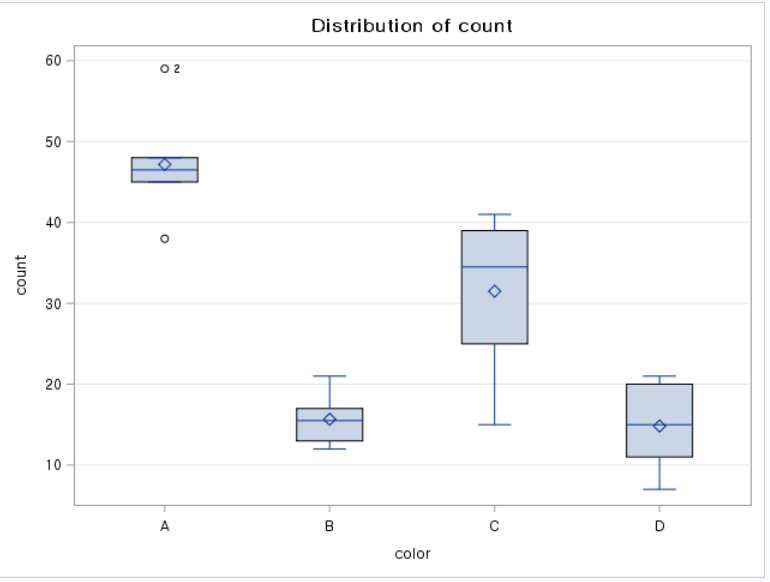
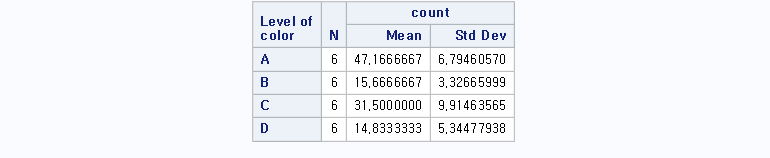
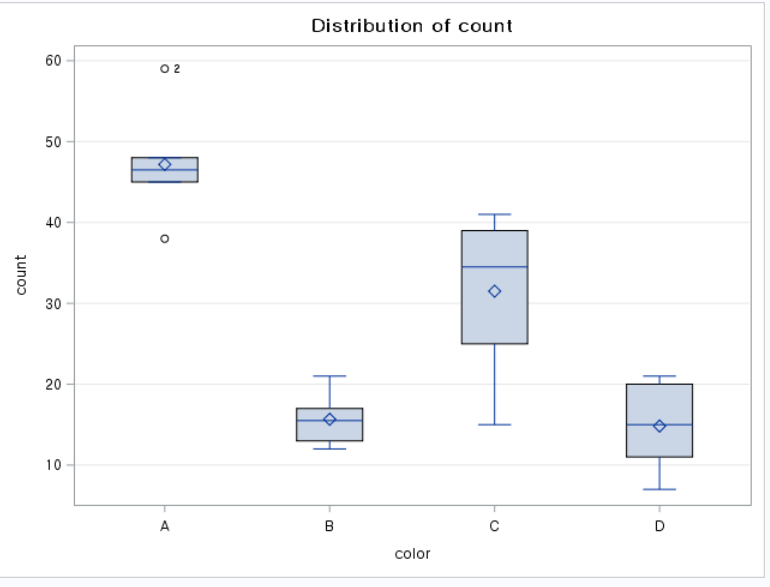
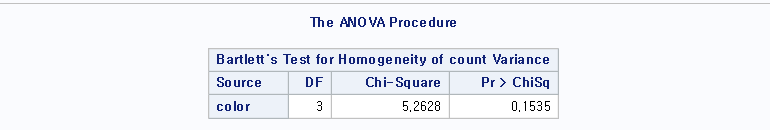
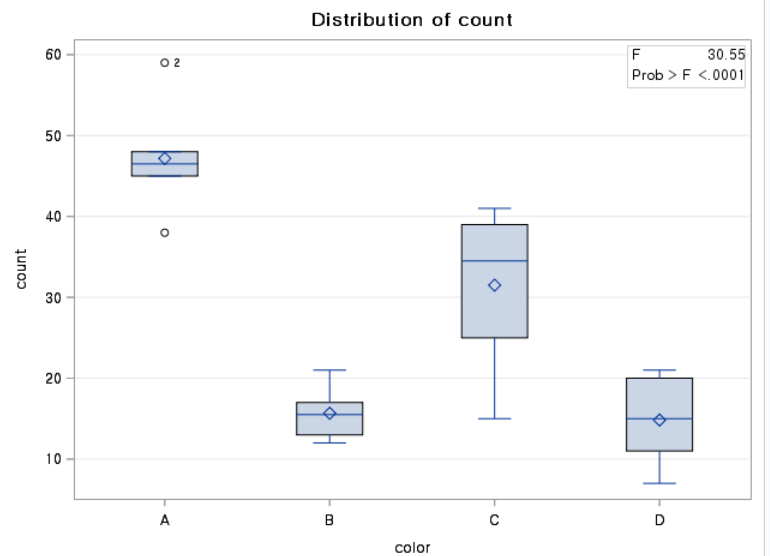
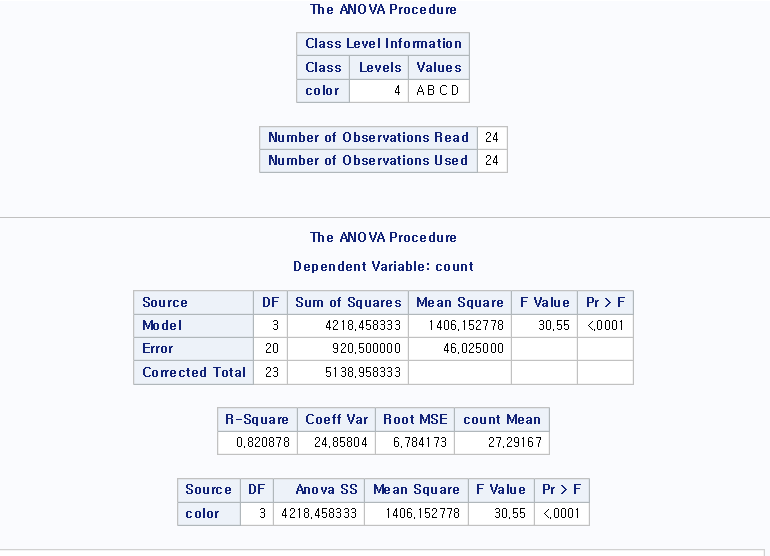
model count = color;

means color / hovtest = bartleft;

means color / tukey cldiff alpha=0.10;

means color / tukey lines alpha=0.10;

run;



해설

결과를 살펴보면 색상에 대한 차이가 유의하다는 것을 알 수 있다 즉 딱정벌레가 선호하는 색에 대한 평균이 차이가 있다는 것이다. 하지만 다중비교 부분을 살펴보면 흰색과 파란색에 대한 선호도는 차이가 없다는 것을 알 수 있다

6-7

data exam.norm;

input group $ value @@;

cards;

A 31 A 20 A 18 A 17 A 9 A 8 A 10 A 7

B 18 B 17 B 14 B 11 B 10 B 7 B 5 B 6

;

run;

proc ttest data=exam.norm cochran;

class group;

var value;

run;

proc anova data=exam.norm;

class group;

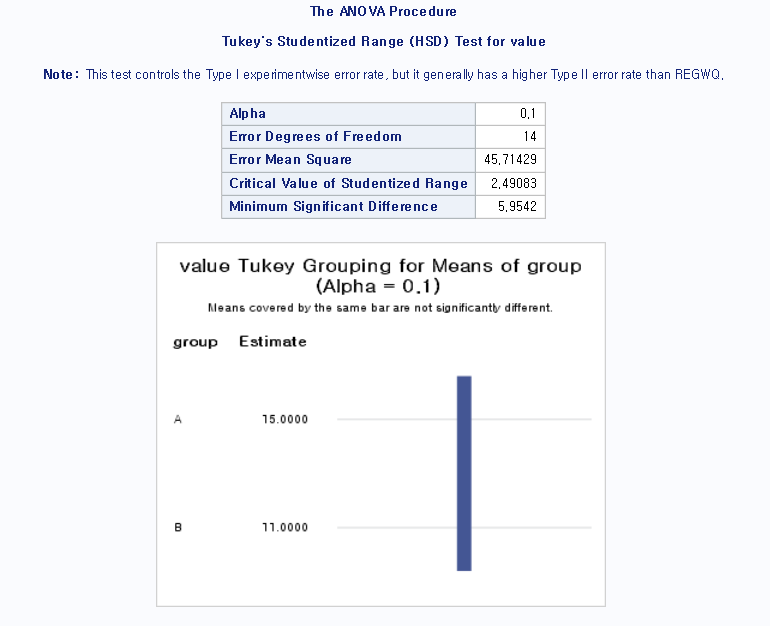
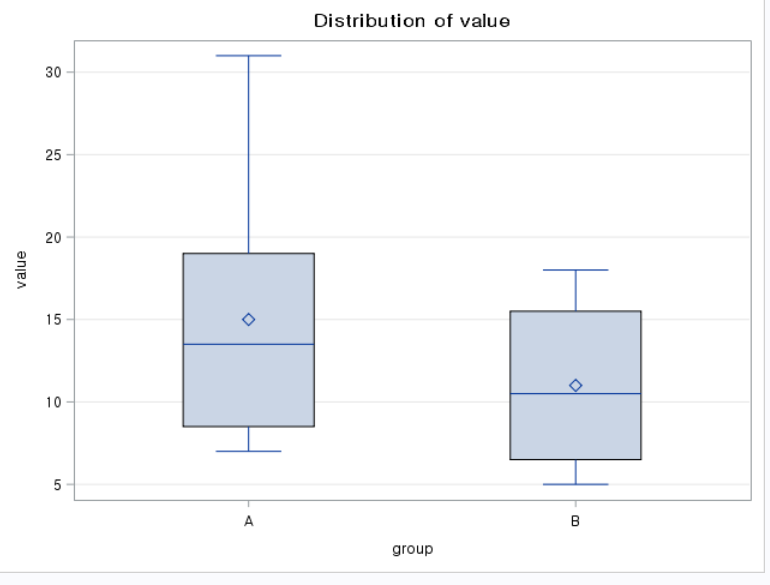
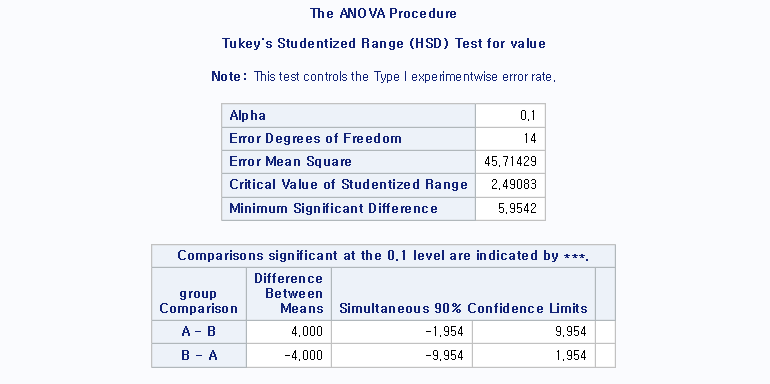
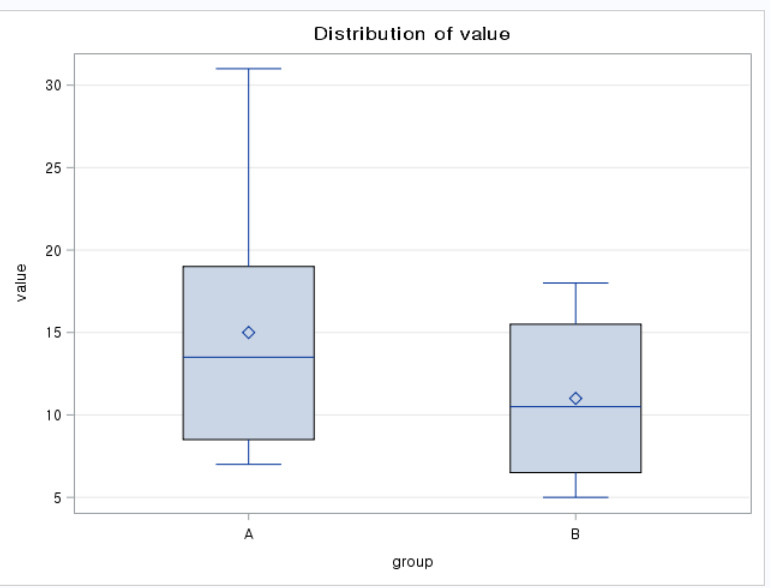
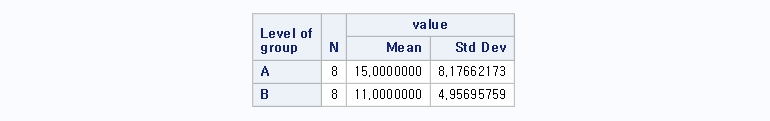
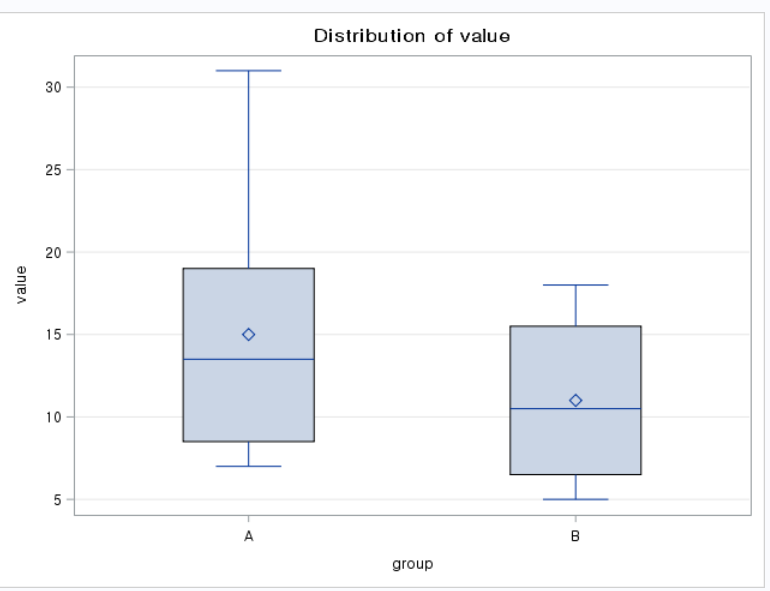
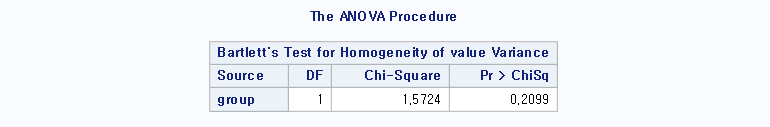
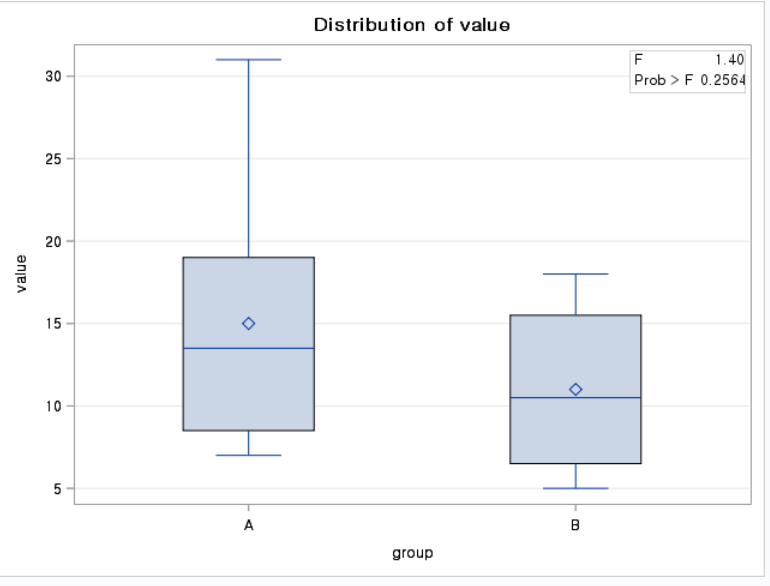
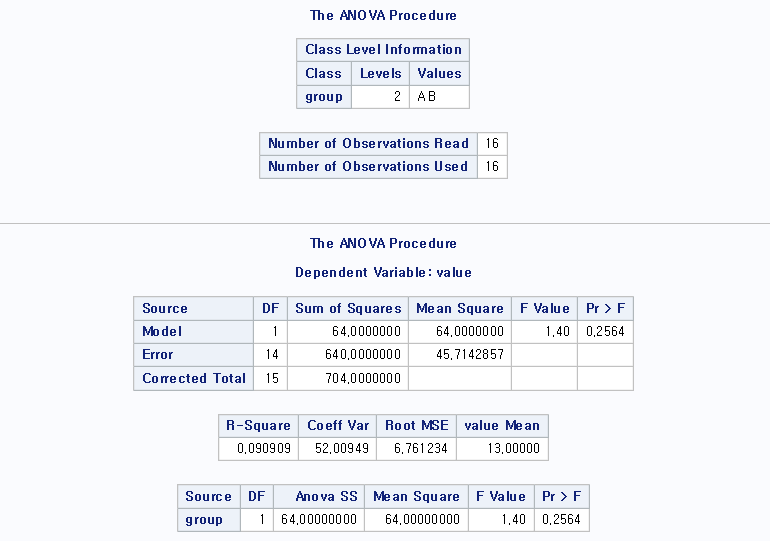
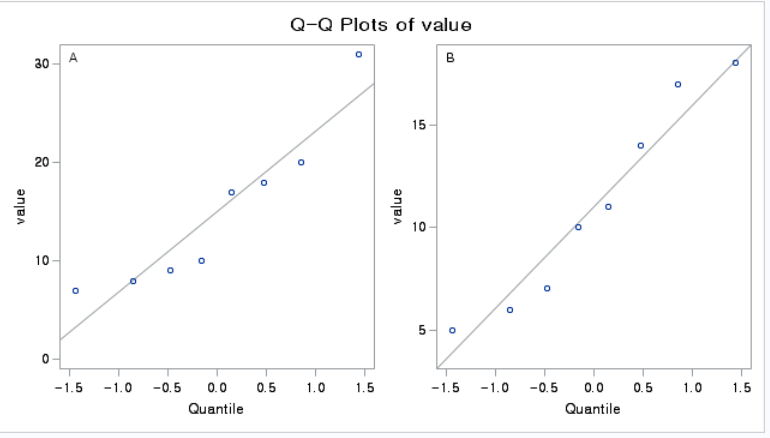
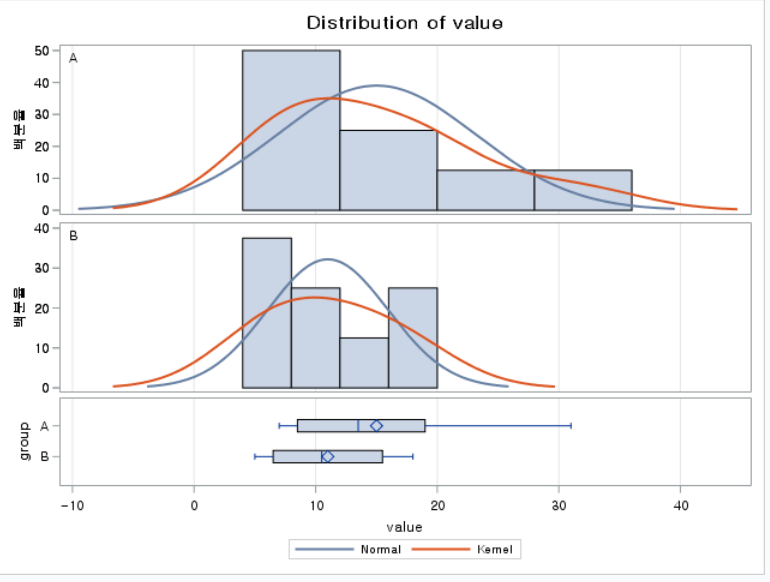
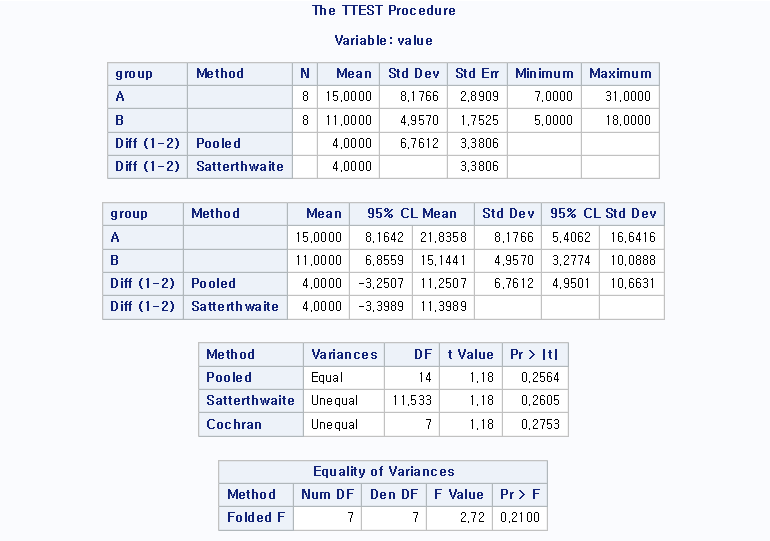
model value = group;

means group / hovtest = bartleft;

means group / tukey cldiff alpha = 0.10;

means group / tukey lines alpha = 0.10;

run;



해설

결과를 살펴보면 다 처리의 차의 평균에 대한 유의확률이 0.2564로 처리별로 통계적으로 유의한수준의 차이가 있다고 할 수 없다는 것을 알 수 있다

6-9

data exam.cholest;

do lab = 'A1', 'A2', 'A3', 'A4';

do food = 1 to 3 by 1;

input cholesterol @@;

output;

end;

end;

cards;

3.4 2.6 2.8

3.0 2.7 3.1

3.3 3.0 3.4

3.5 3.1 3.7

;

run;

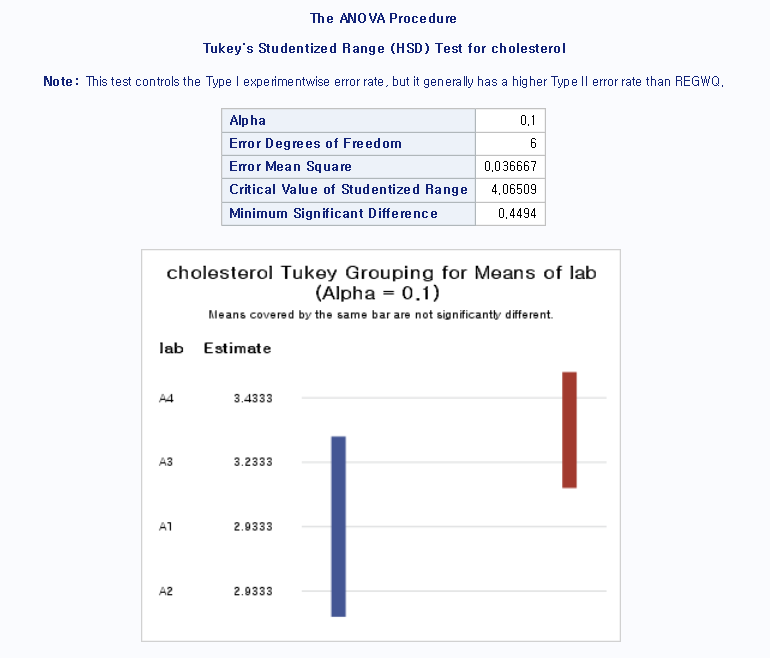
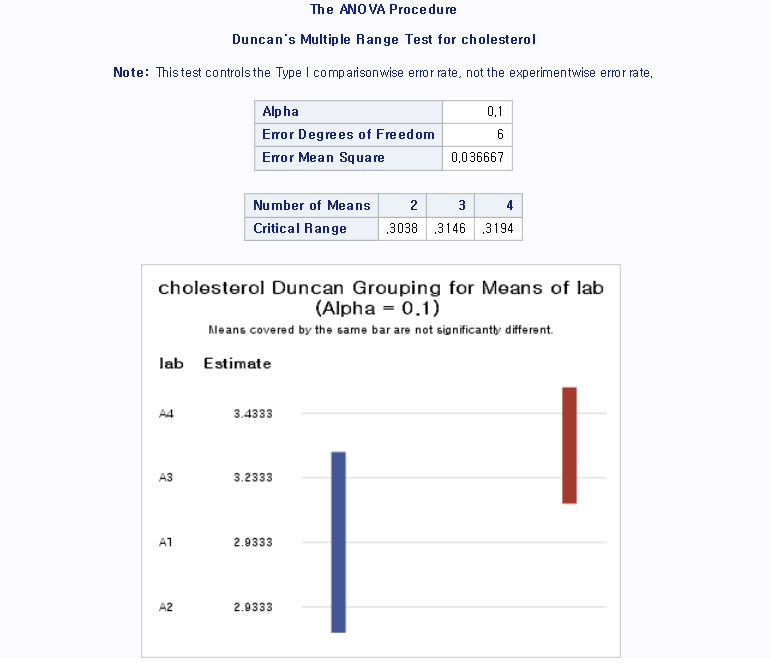
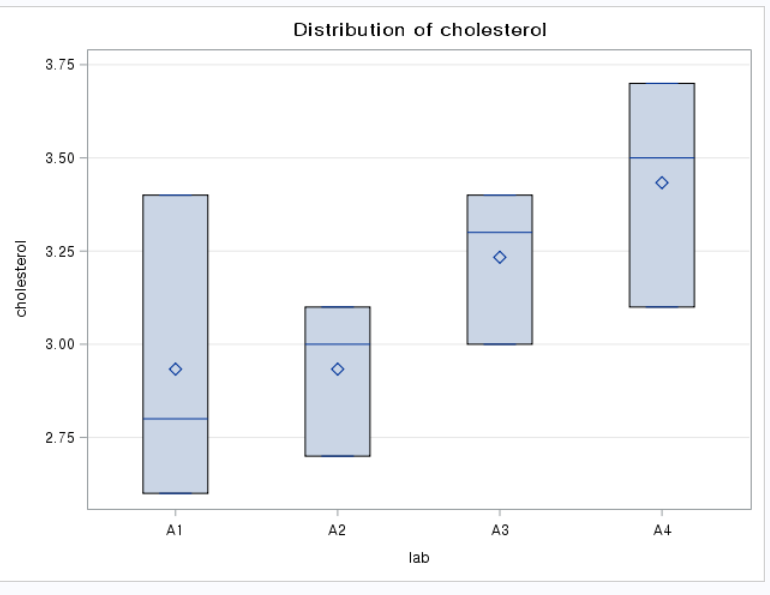
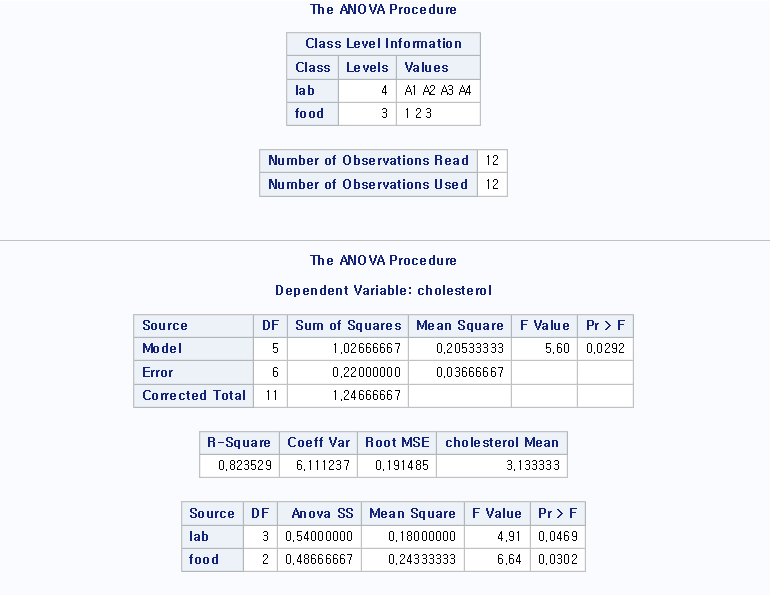
proc anova data =exam.cholest;

class lab food;

model cholesterol = lab food;

means lab / ducan tukey alpha=0.10;

run;



해설

결과를 살펴보면 food와 lab모두 통계적으로 유의한 차이를 보이고 있다는 것을 알 수 있다

또한 다중비교 부분을 살펴보면 실험실 C,D의 콜레스트롤 수치가 평균적으로 높다는 것을 알 수 있다

6-11

data exam.weight;

do factor = 'A1', 'A2', 'A3';

do protein = 'B1', 'B2';

do rep = 1,2,3,4,5,6,7,8,9,10;

input weight @@;

output;

end;

end;

end;

cards;

73 102 118 104 81 107 100 87 117 111

90 76 90 64 86 51 72 90 95 78

98 74 56 111 95 88 82 77 86 92

107 95 97 80 98 74 74 67 89 58

74 79 96 98 102 102 108 91 120 105

49 72 73 86 81 97 106 70 61 82

;

run;

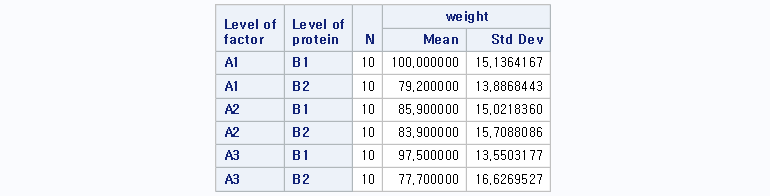
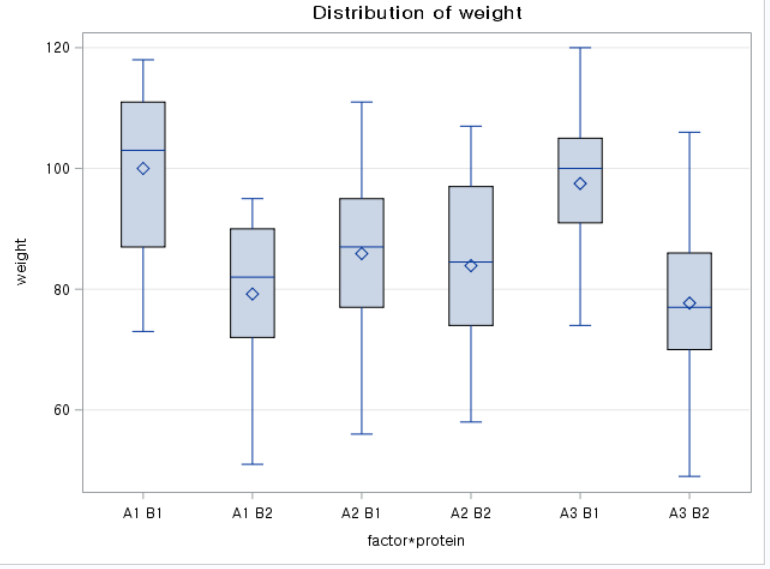
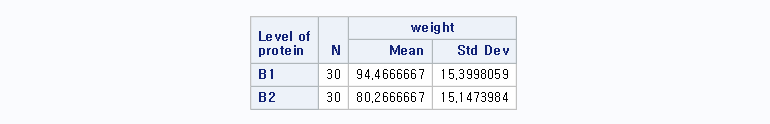
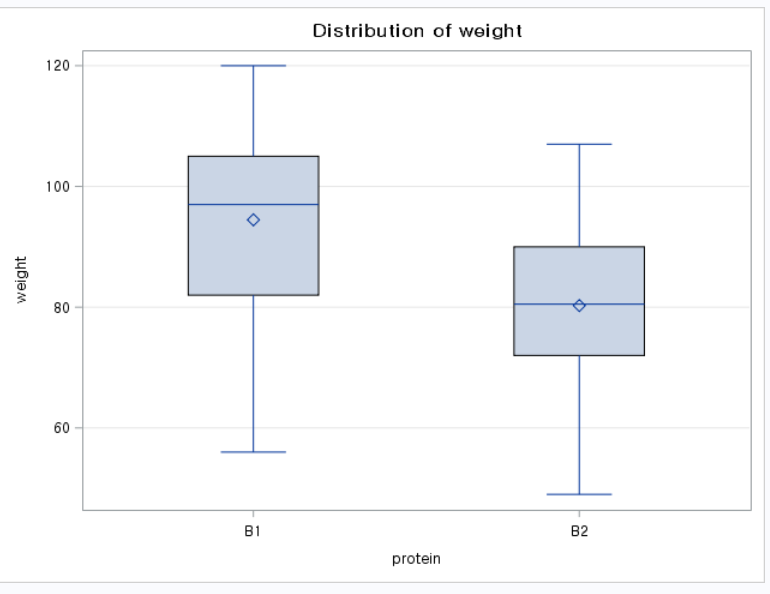
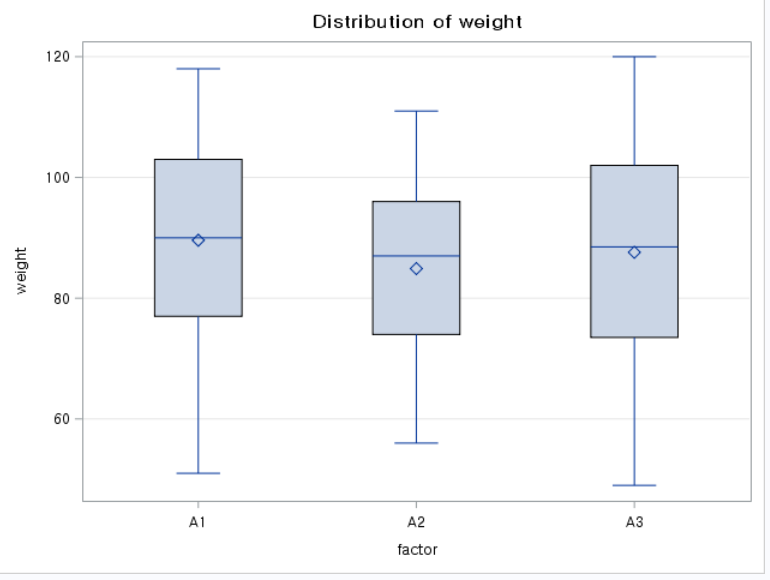
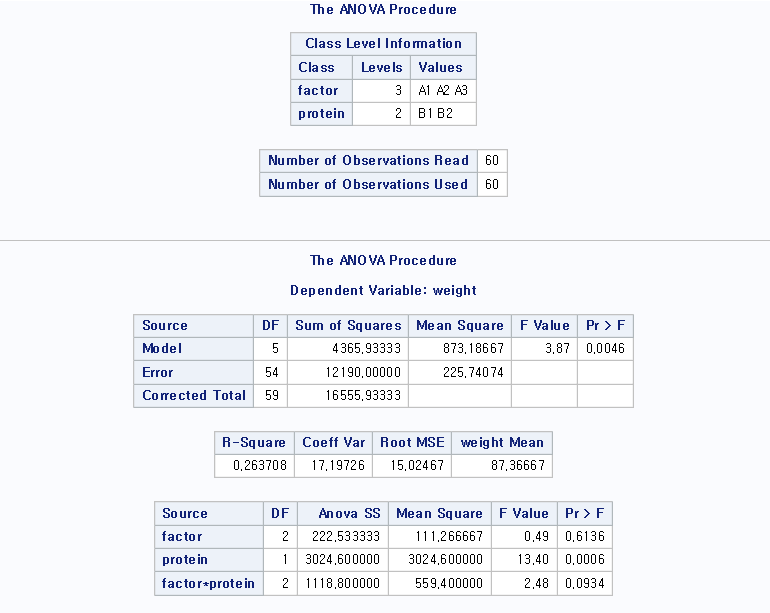
proc anova data=exam.weight;

class factor protein;

model weight = factor protein factor\*protein;

means factor protein factor\*protein;

run;



해설

결과를 살펴보면 단백질의 요소에 대한 유의학률과 단백질의 수준과 단백질의 요소의 상호작용에 대한 유의확률이 유의할 만큼 작지 않다는 것을 알 수 있다 즉 쥐의 체중증가에 영향을 미치는 요소는 단백질의 높고 낮음이다

예 제

<6.1>

data exam.harvest;

input fertile $ yield @@;

cards;

F1 148 F1 76 F1 134 F1 98

F2 166 F2 153 F2 255

F3 264 F3 214 F3 327 F3 304

F4 335 F4 436 F4 423 F4 380 F4 465

;

RUN;

proc anova data=exam.harvest;

class fertile;

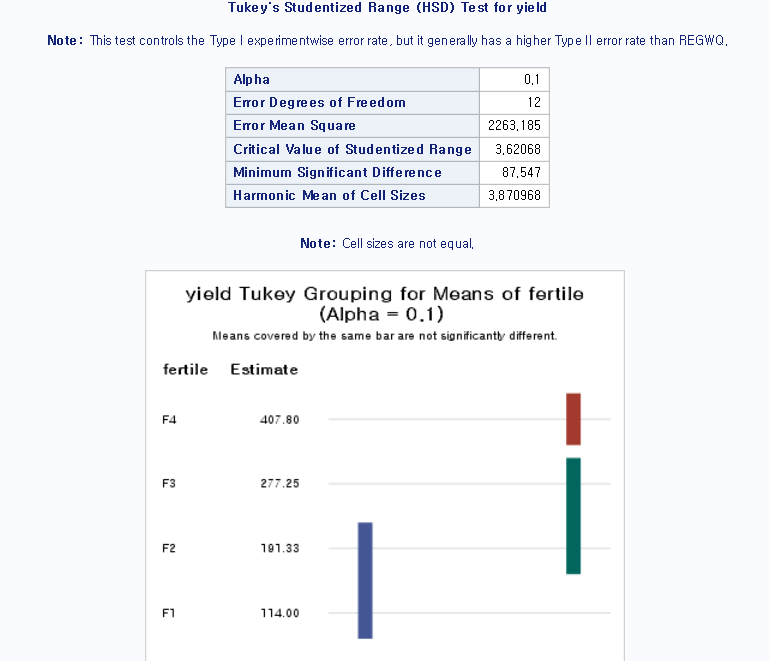
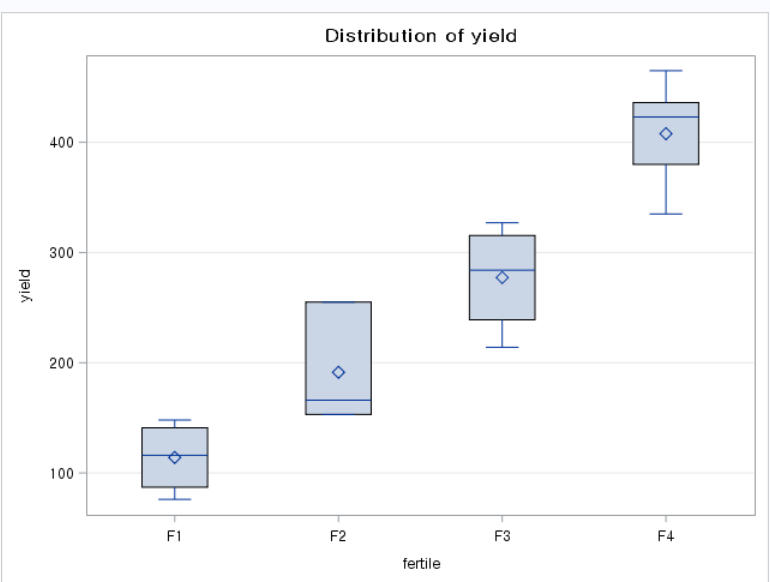
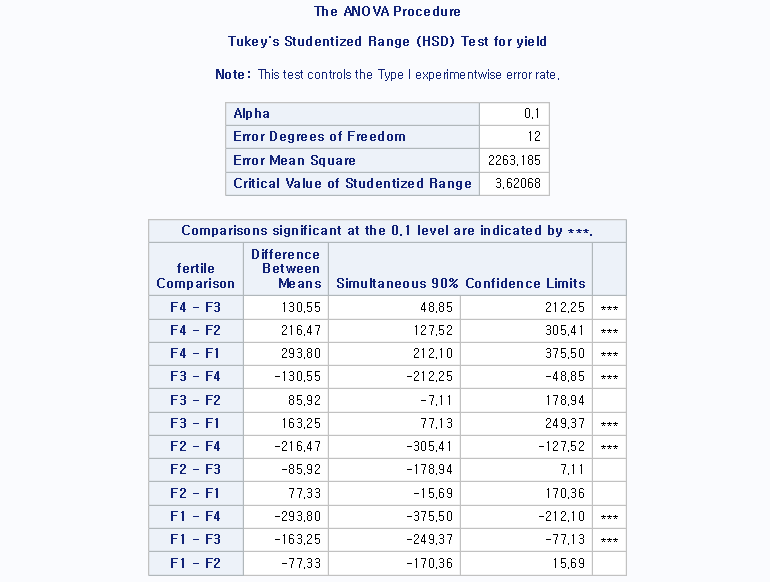
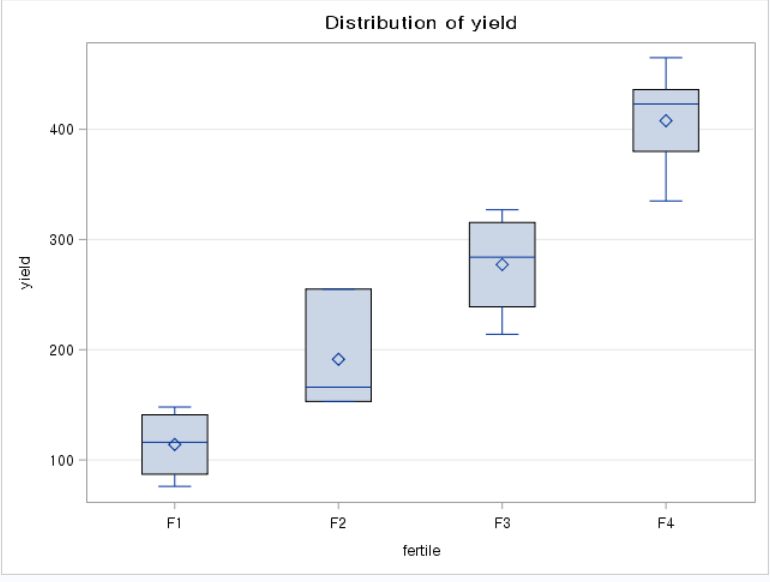
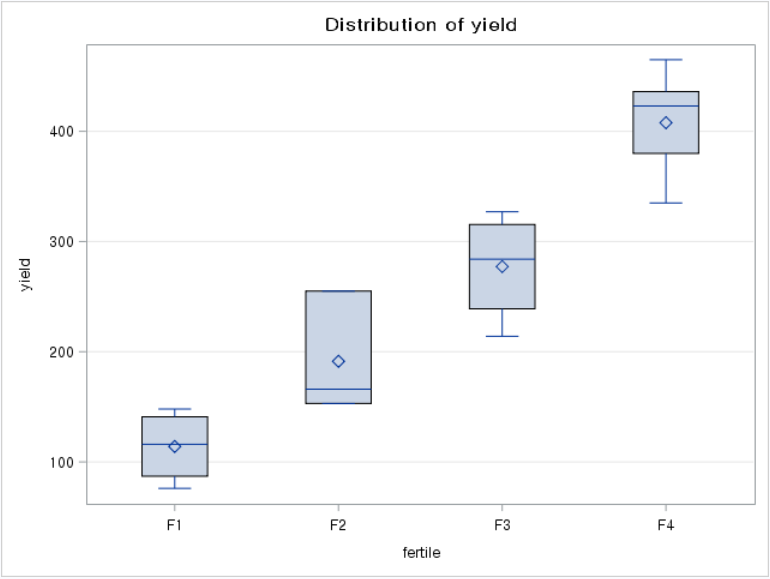
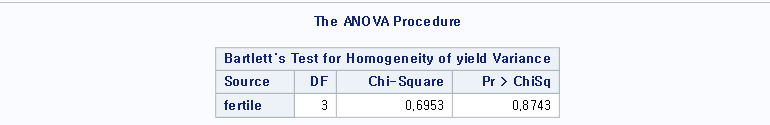
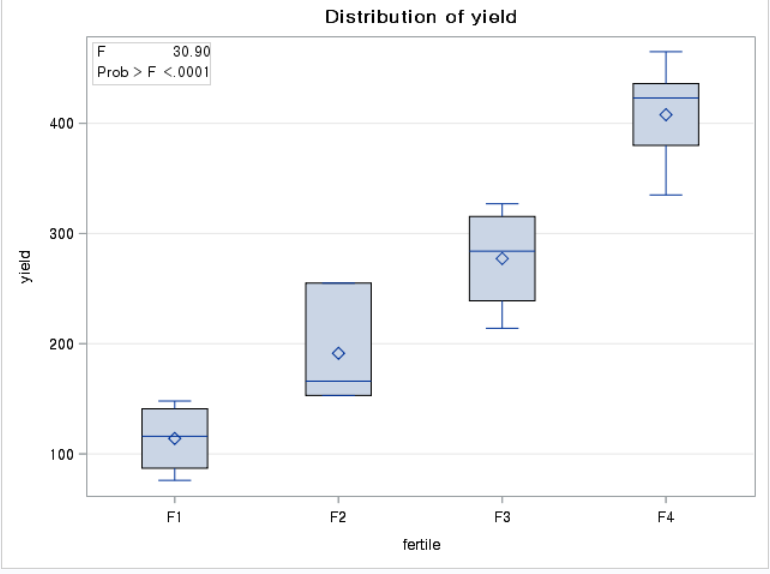
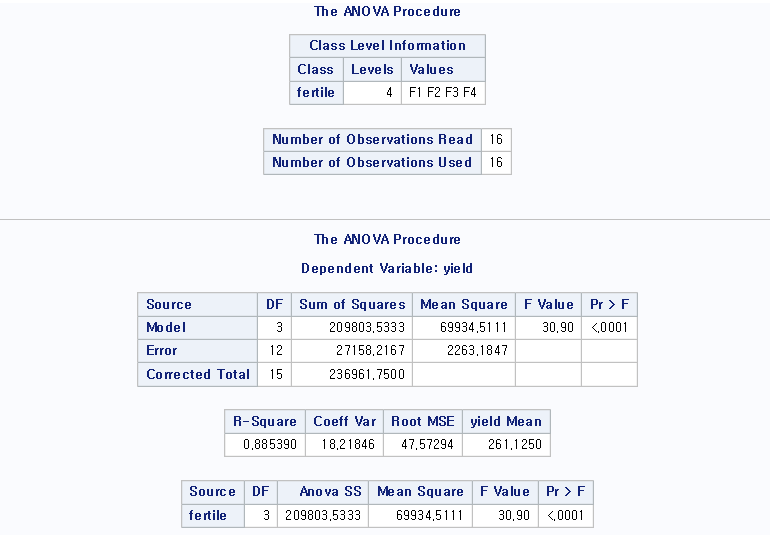
model yield=fertile;

means fertile / hovtest=BARTLEFT;

means fertile / tukey cldiff alpha=0.10;

means fertile / tukey lines alpha=0.10;

run;



해설

결과를 살펴보면 우선 F통계량에 대한 유의수준이 충분히 작기 때문에 H0:비료의 종류에 따라 수확량의 포평균에 차이가 없다를 기각 할 수 있다 한편 각 처리수준별 표본평균 및 표준편차를 확일 할 수 있으며 분산의 동일성 검정도 이루어지고 있다. 지금과 같이 귀무가설이 기각된다면 다중비교를 실시해야 한다 결과를 살펴보면 변수 2,3 그리고 변수 1,2에 대한 평균에 대한 차이가 없다는 것을 알 수 있다.

<6.2>

data exam.prefer;

do product= 'A1', 'A2', 'A3', 'A4';

do customer =1 to 5 by 1;

input prefer @@;

output;

end;

end;

cards;

5 7 9 10 8

2 3 4 5 2

4 7 6 5 7

6 4 2 2 1

;

run;

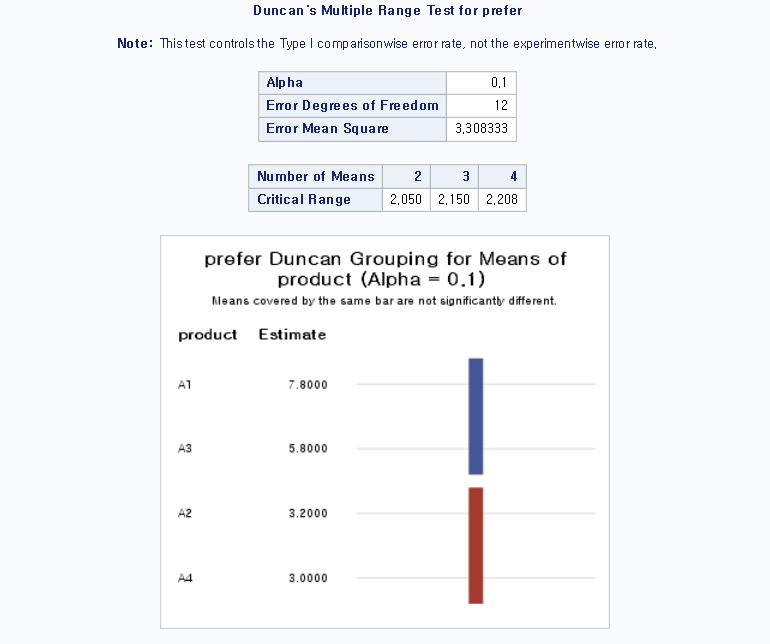
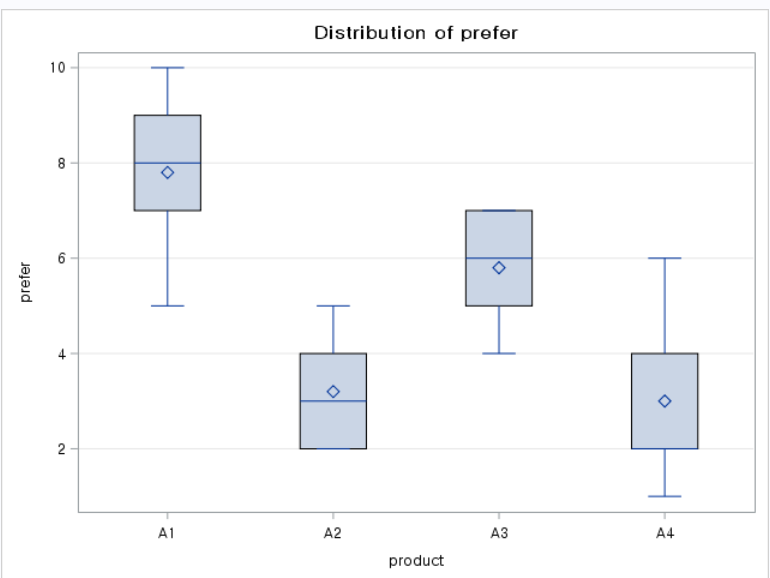
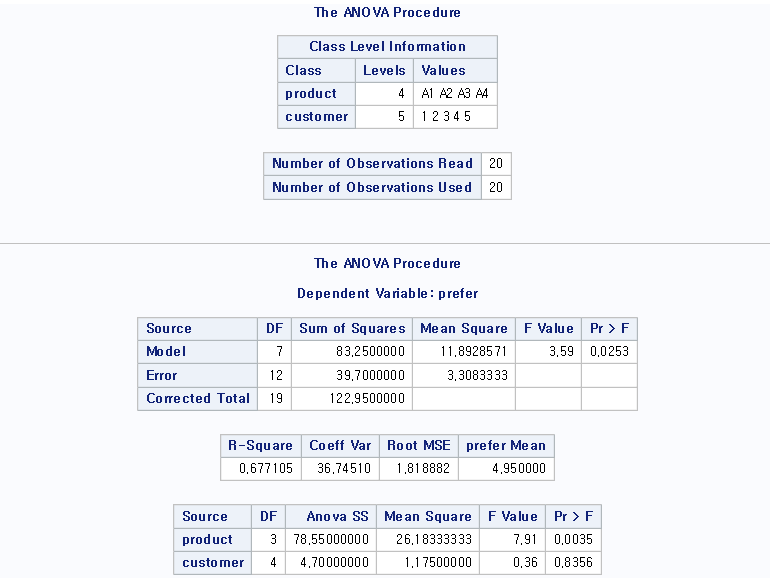
proc anova data=exam.prefer;

class product customer;

model prefer = product customer;

means product / DUCAN tukey alpha=0.10;

run;



해설

결과를 살펴보면 변수 product에 대한 F값의 유의확률이 충분히 작아서 유의함을 알 수 있다

반면 변수 cutomer에 대한 유의확률은 0.8356으로 유의한 차이를 보이고 있지 않다는 것을 알 수 있다 또한 다중 비교를 살펴보면 제품 A1과 A3의 경우가 A2와 A4에 비해서 선호도의 평균이 높다는 것을 알 수 있다

<6.3>

data exam.sales;

do city='large', 'middle', 'small';

do design='A', 'B', 'C';

do rep=1,2,3;

input sales @@;

output;

end;

end;

end;

cards;

23 20 21 22 19 20 19 18 21

22 20 19 24 25 22 20 19 22

18 18 16 21 23 20 20 22 24

;

run;

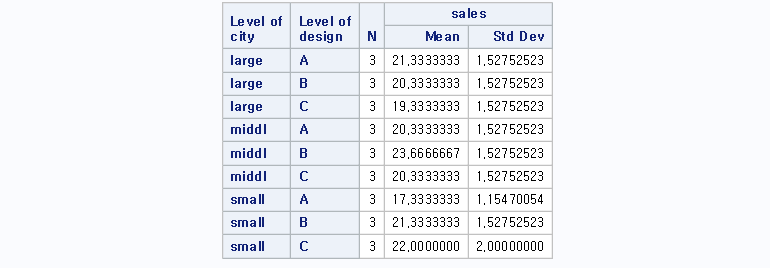
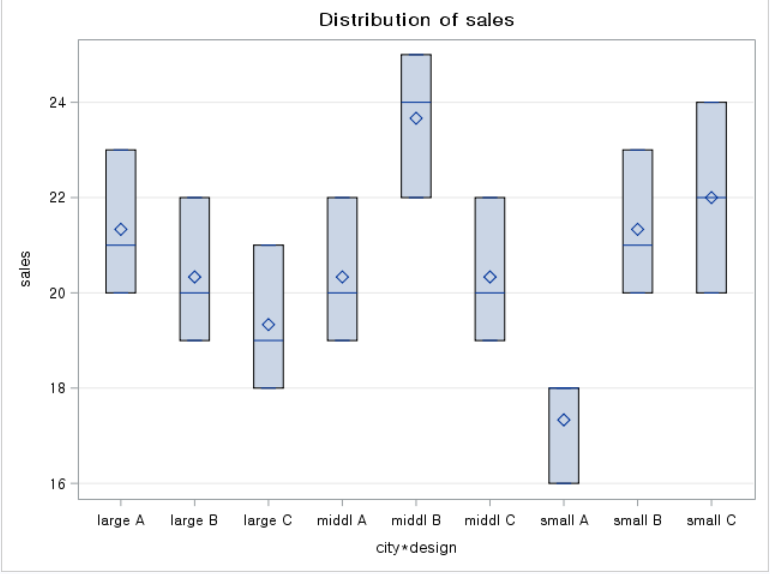
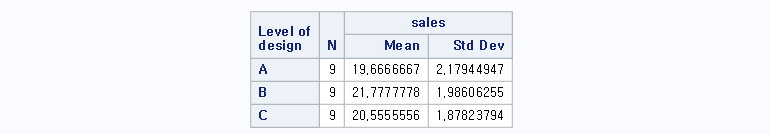
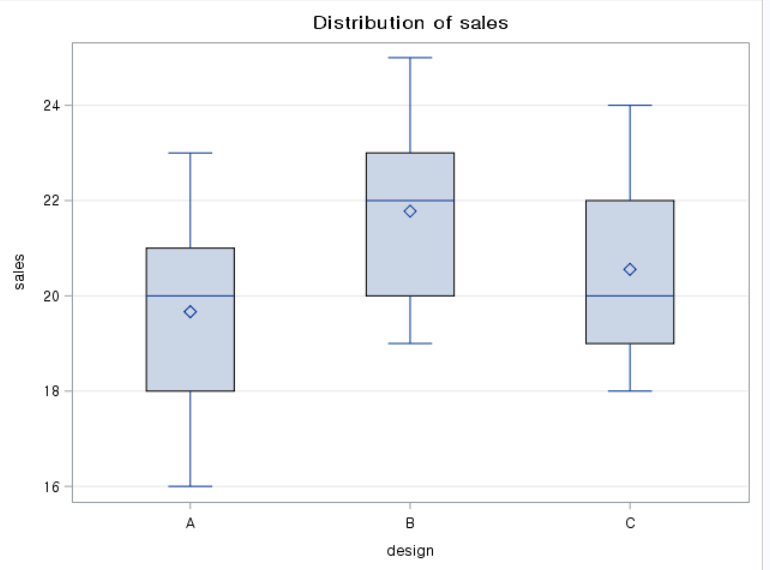
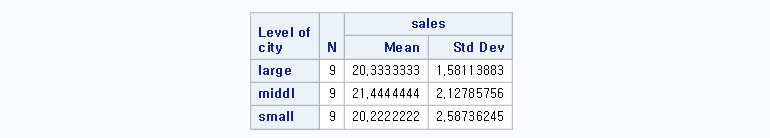
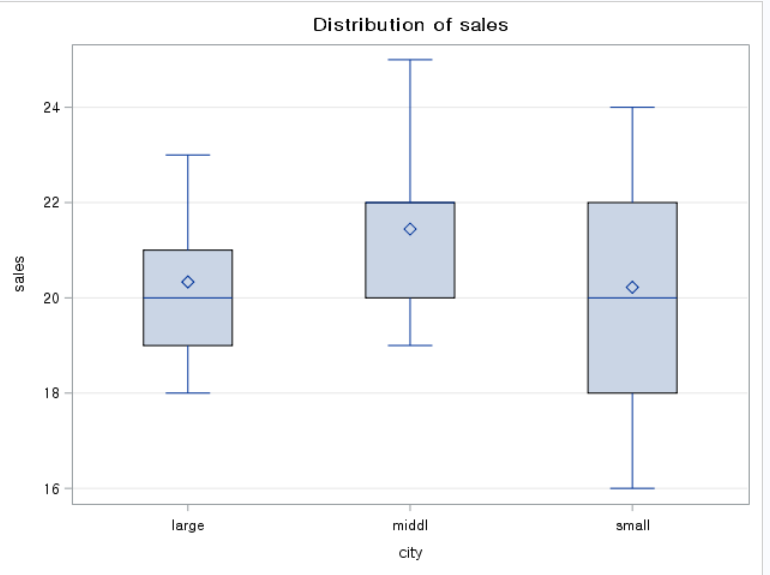
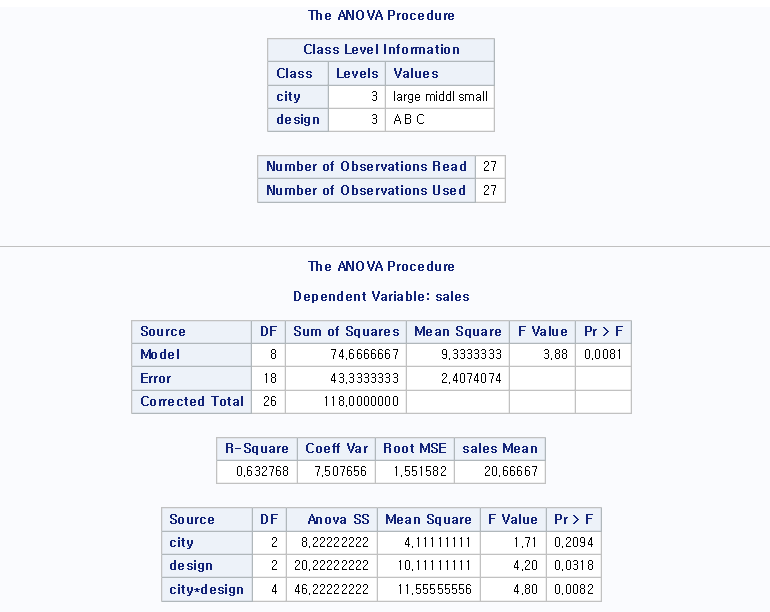
proc anova data=exam.sales;

class city design;

model sales= city design city\*design;

means city design city\*design;

run;



해설

반복이 있는 이원분산분석에서는 상호작용의 효과도 살펴봐야 하는데 결과를 보면 design\*city 에 대응하는 유의확률이 충분히 작으므로 상호작용의 효과가 통계적으로 유의하다고 할 수 있다 상호작용이 유의하다면 상호작용과 관련된 변수들은 유의확률이 낮더라도 함부로 제거하지 않는것이 좋다

<6.4>

proc summary data=exam.sales nway;

class city design;

var sales;

output out=meanout mean(sales)=mean;

run;

sumbol1 interpol=join width=1 value=dot cv=black height=2;

symbol2 interpol=join width=1 value=circle cv=red height=2;

symbol3 interpol=join width=1 value=square cv=blue height=2;

proc gplot data=meanout;

plot mean\*city=design;

run;

<6.5>

data exam.ancova;

input group $ age score @@;

cards;

A 31 30 A 28 0 A 25 10 A 34 40 A 39 55

A 26 20 A 30 65 A 26 5 A 31 40 A 23 0

B 36 65 B 33 50 B 31 90 B 29 25 B 41 99

B 36 60 B 32 25 B 32 80 B 27 5 B 32 99

;

run;

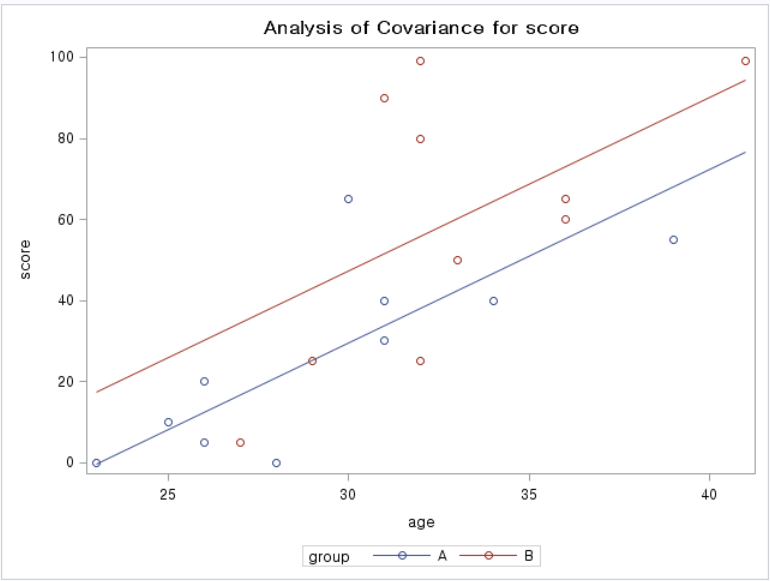
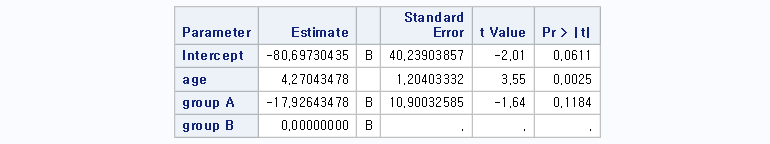
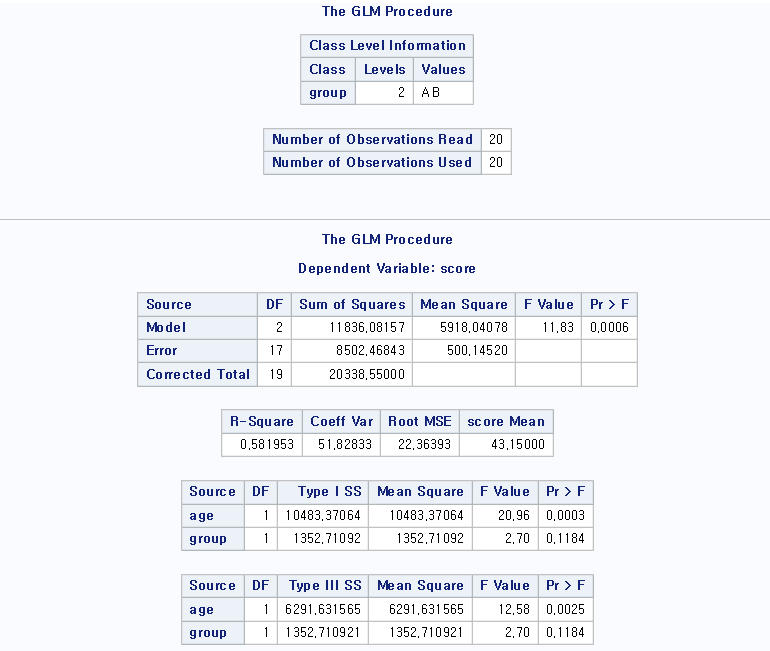
proc glm data=exam.ancova;

class group;

model score = age group

/ solutions;

run;



해설

결과를 살펴보면 age의 F값에 대한 유의확률은 충분히 작기 때문에 유의하나 group의 F값에 대한 유의확률은 0.118로 유의하지 않다 즉 나이가 통제된다면 두 집단에 대한 심리검사 점수의 평균 차이는 유의하지 않다고 할 수 있다

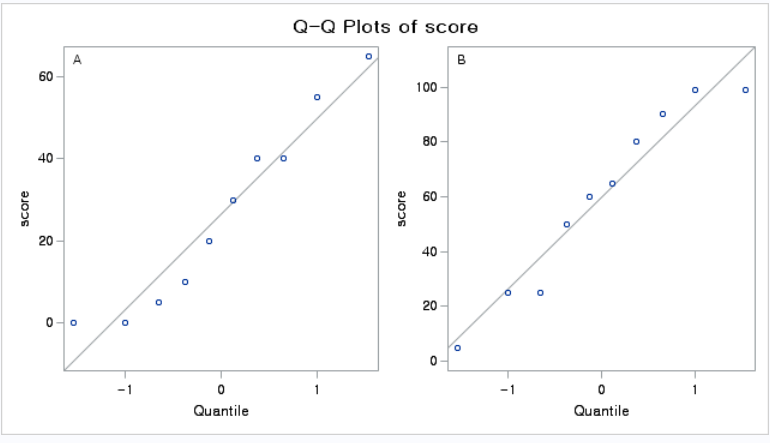
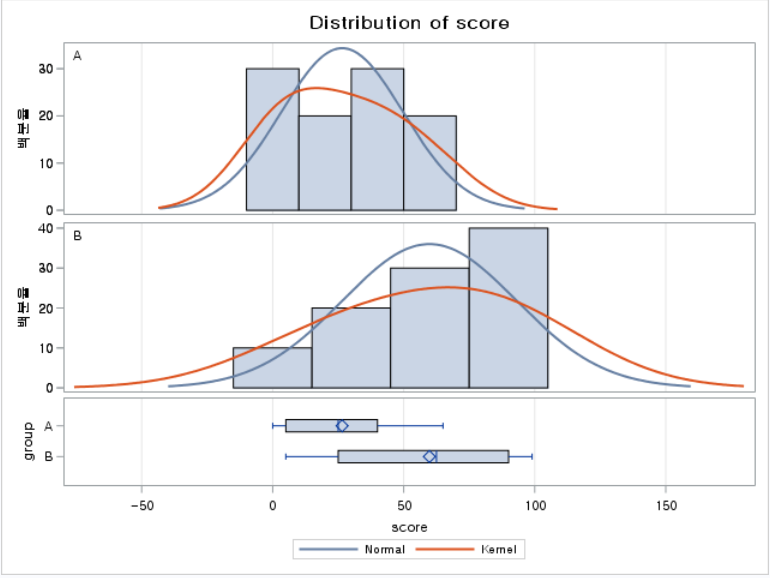
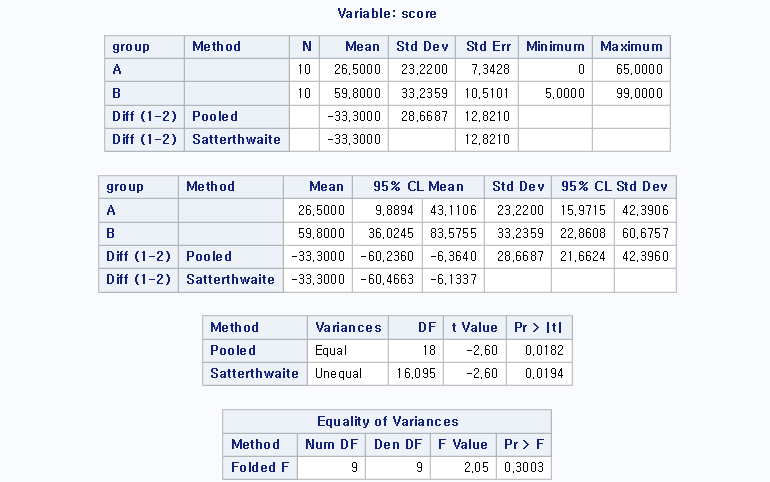
<6.6>

proc ttest data=exam.ancova;

class group;

var score;

run;



여기서 살펴 볼 수 있듯이 공변량이 존재하는데도 이를 고려하지 않고 분석을 수행한다면 제대로된 분석을 수행 할 수 없음을 알 수 있다 (group의 유의확률이 0.018로 유의)

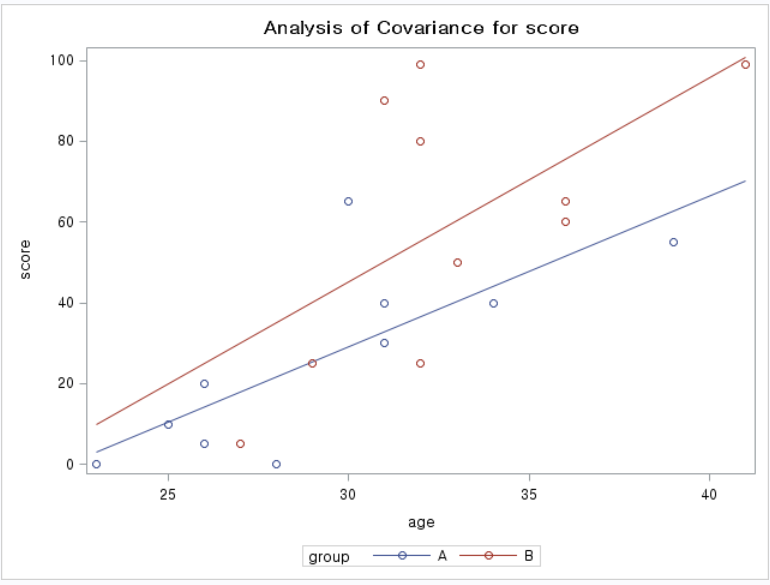
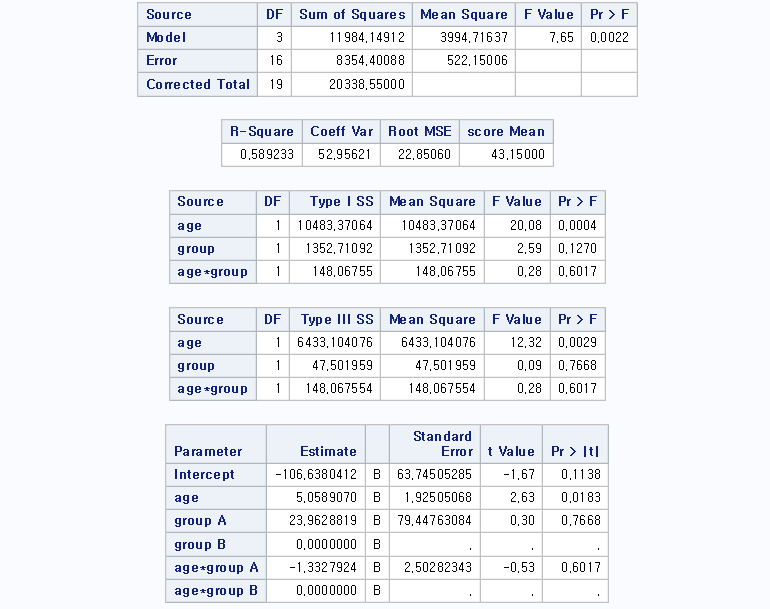
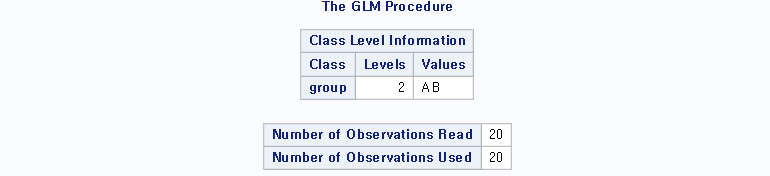
<6.7>

proc glm data=exam.ancova;

class group;

model score = age group age\*group / solutions;

run;



해설

결과를 살펴보면 age\*group의 유의확률이 0.6017로 상호작용의 효과가 유의하지 않으므로 기울기의 동일성을 가정할 수 있다