

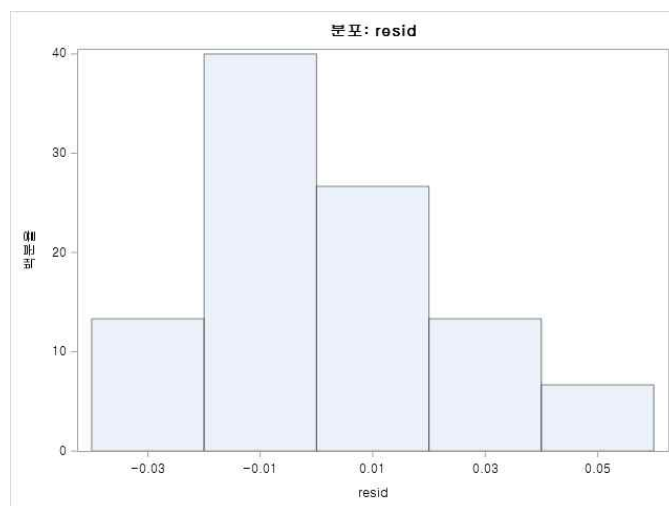
Problem 1.

1) 잔차의 히스토그램을 그려라.

```

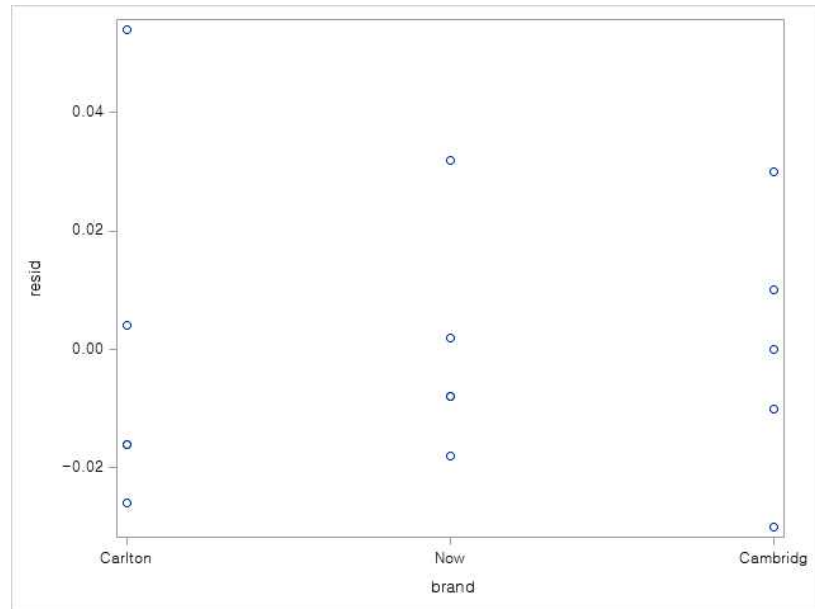
1  /* 1번 */
2  /* 데이터 생성 */
3  data dataset1;
4  input brand $ tar;
5  cards;
6  Carlton 0.16
7  Carlton 0.14
8  Carlton 0.21
9  Carlton 0.14
10 Carlton 0.13
11 Now 0.19
12 Now 0.20
13 Now 0.23
14 Now 0.18
15 Now 0.19
16 Cambridge 0.21
17 Cambridge 0.17
18 Cambridge 0.19
19 Cambridge 0.23
20 Cambridge 0.20
21 ;
22 run;
23
24 proc print data = dataset1;
25 run;
26
27 /* 잔차 생성 */
28 proc glm data = dataset1;
29 class brand;
30 model tar = brand;
31 output out = dataset2 p = pred r = resid;
32
33 proc print data = dataset2;
34 run;
35
36 /* 1) 잔차 히스토그램 */
37 proc univariate data = dataset2 normal noprint; /* normal: 정규성 검정을 시행 */
38 /* noprint: PROC UNIVARIATE 문이 생성하는 모든 기술 통계 테이블을 억제한다 */
39 histogram resid;
40 run;

```



2) 잔차의 산점도 (x 축은 담배브랜드, y 축은 잔차)

```
42 /* 2) 잔차 산점도 */
43 proc sgplot data = dataset2;
44     scatter x = brand y = resid;
45 run;
```



3) 바틀렛의 검정을 이용한 등분산성을 검정하라

```
47 /* 3) 등분산 검정(bartlett test) */
48 proc glm data = dataset2;
49     class brand;
50     model tar = brand;
51     means brand / hovtest = bartlett;
52 /* means brand: brand별로 tar의 산술평균과 표준편차 계산 */
53 run;
```

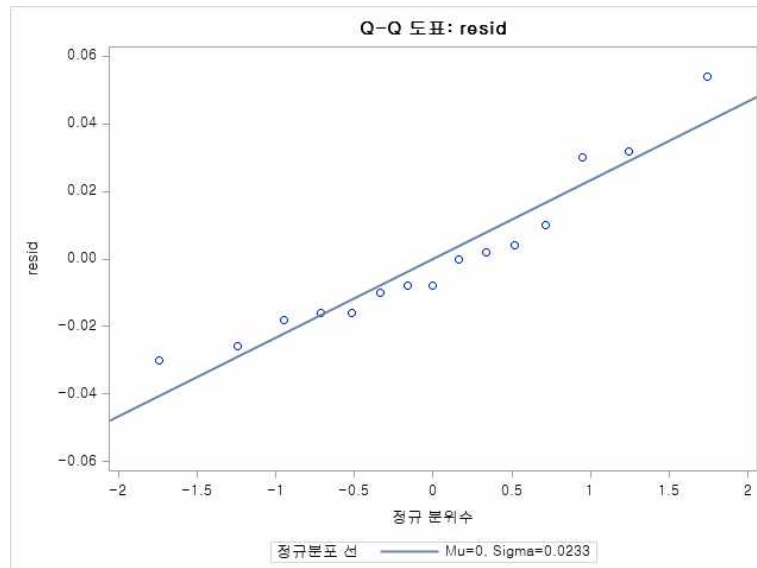
H_0 : 그룹 간 분산이 같다. H_1 : 그룹 간 분산이 다르다.

| The GLM Procedure | | | |
|---|----|------------|------------|
| Bartlett's Test for Homogeneity of tar Variance | | | |
| Source | DF | Chi-Square | Pr > ChiSq |
| brand | 2 | 1.0352 | 0.5959 |

유의수준 0.05 하에서 귀무가설 채택
(즉, 그룹 간 분산이 같다고 할 수 있다)

4) 잔차를 qq-plot을 이용하여 그리고 검정하라.

```
55 /* 4) Q-Q plot */
56 proc univariate data = dataset2 normal noprint; /* normal: 정규성 검정을 시행 */
57 /* noprint: PROC UNIVARIATE 문이 생성하는 모든 기술 통계 테이블을 억제한다 */
58 qqplot resid / normal(mu = est sigma = est);
59 /* est는 resid 데이터로부터 계산한 평균과 표준편차를 따르는 정규분포선을 그림 */
60 run;
```



```
62 /* 4) 정규성 검정 */
63 /* H_0: 잔차가 정규분포를 따른다 H_1: 잔차가 정규분포를 따르지 않는다 */
64 proc univariate data = dataset2 normal;
65 class brand;
66 var resid;
67 run;
```

H_0 : 잔차가 정규분포를 따른다. H_1 : 잔차가 정규분포를 따르지 않는다.

| 정규성 검정 | | | | |
|--------------------|------|----------|-----------|---------|
| 검정 | 통계량 | p 값 | | |
| Shapiro-Wilk | W | 0.999293 | Pr < W | 0.9998 |
| Kolmogorov-Smirnov | D | 0.12736 | Pr > D | >0.1500 |
| Cramer-von Mises | W-Sq | 0.01837 | Pr > W-Sq | >0.2500 |
| Anderson-Darling | A-Sq | 0.139422 | Pr > A-Sq | >0.2500 |

모두 유의수준 0.05 하에서 귀무가설 채택
(즉, 잔차가 정규분포를 따른다고 할 수 있다)

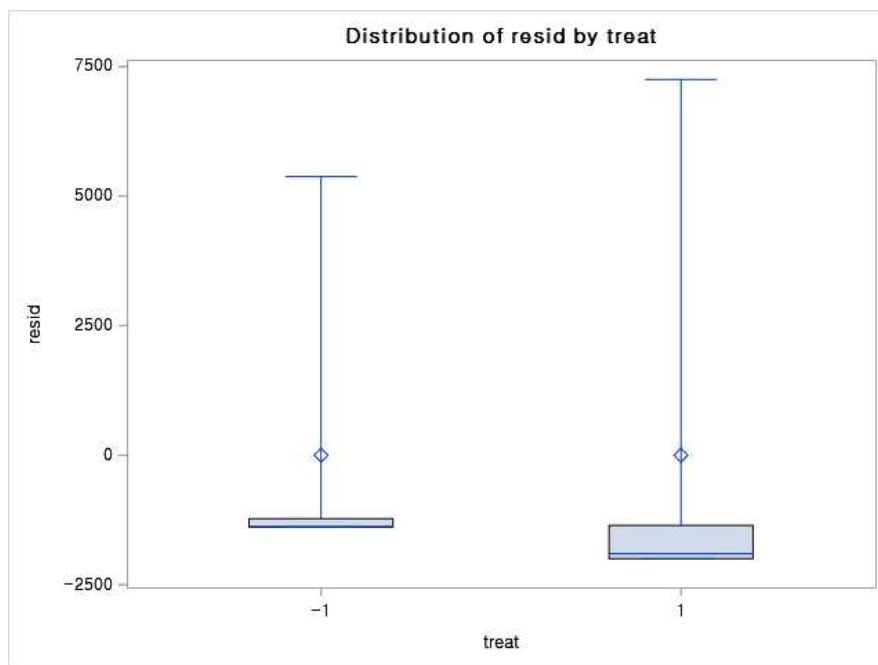
Problem 2.

```
69 /* 2번 */
70 /* 데이터 생성 */
71 data a;
72   do treat = -1 to 1 by 2;
73     do n = 1 to 5;
74       y = exp(4+3*rannor(111))+treat;
75       output;
76     end;
77   end;
78 run;
79
80 proc print data = a;
81 run;
82
83 /* 잔차 생성 */
84 proc glm data = a;
85 class treat;
86 model y = treat;
87 output out = b p = pred r = resid;
88
89 proc print data = b;
90 run;
```

a)

- 상자 그림

```
92 /* a) 상자그림 */
93 proc boxplot data = b;
94   plot resid * treat;
95 run;
```



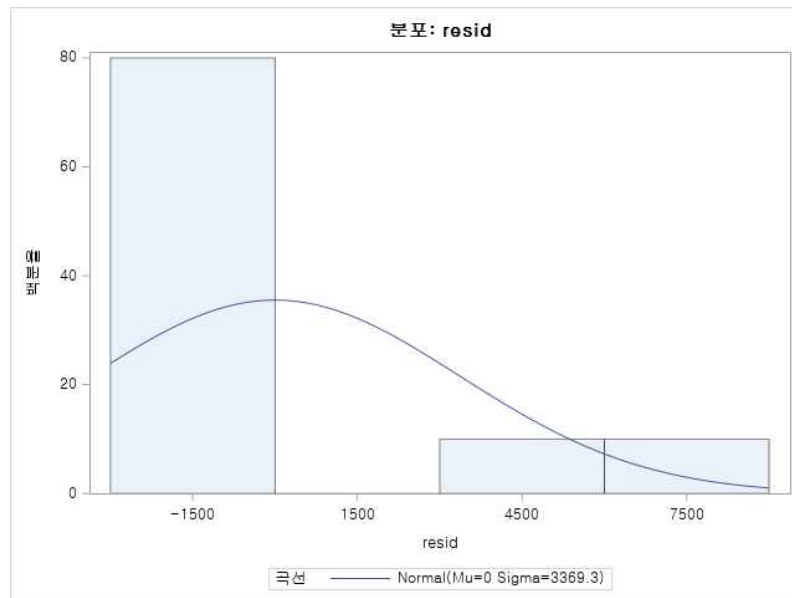
(잔차가 정규성을 띄기 힘든 것으로 보임)

- 히스토그램

```

97 /* a) 히스토그램 */
98 proc univariate data = b noprint;
99 histogram resid / normal(mu = est sigma = est);
100 run;

```



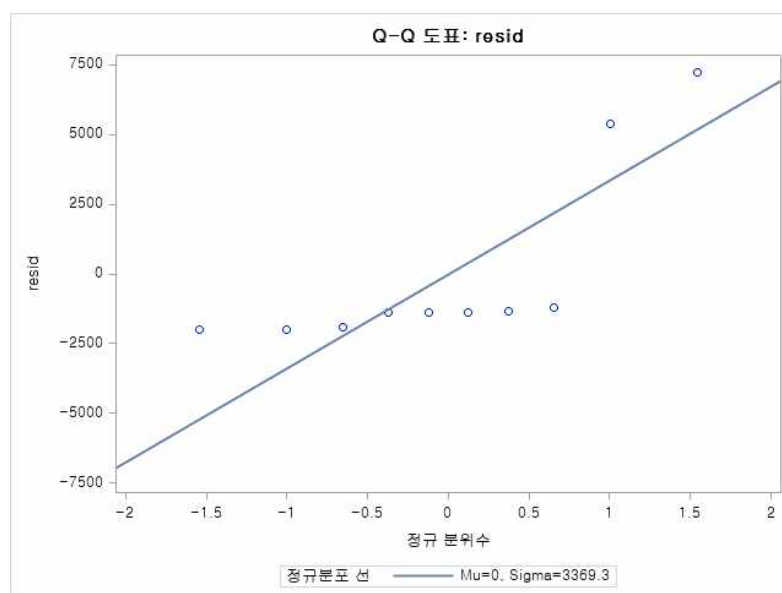
(정규분포 그래프에서 많이 벗어나는 것으로 보임)

- Q-Q Plot

```

102 /* a) Q-Q plot */
103 proc univariate data = b noprint;
104 qqplot resid / normal(mu = est sigma = est);
105 run;

```



(점들이 직선 위를 많이 벗어나는 것으로 보임)

- 샤피로-윌크, 앤더슨-달링 검정

```
107 /* a) 샤피로-윌크 검정/앤더슨-달링 검정 */
108 proc univariate data = b normal;
109 class treat;
110 var resid;
111 run;
```

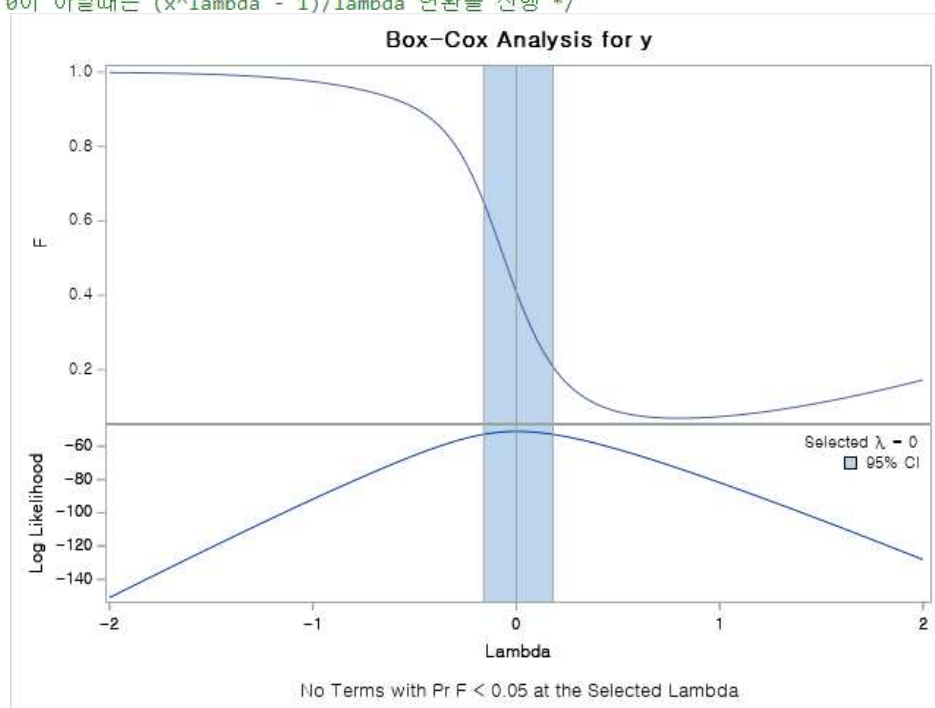
H_0 : 잔차가 정규분포를 따른다. H_1 : 잔차가 정규분포를 따르지 않는다.

| 정규성 검정 | | | | |
|--------------------|------|----------|-----------|---------|
| 검정 | 통계량 | | p 값 | |
| Shapiro-Wilk | W | 0.569184 | Pr < W | 0.0002 |
| Kolmogorov-Smirnov | D | 0.457948 | Pr > D | <0.0100 |
| Cramer-von Mises | W-Sq | 0.229744 | Pr > W-Sq | <0.0050 |
| Anderson-Darling | A-Sq | 1.150506 | Pr > A-Sq | <0.0050 |

모든 검정이 유의수준 0.05 하에서 귀무가설 기각
(즉, 잔차가 정규분포를 따르지 않는다고 할 수 있다)

b) Box-Cox 변환

```
113 /* b) Box-cox 변환 */
114 proc transreg details data = b;
115 model boxcox(y / convenient lambda = -2 to 2 by 0.01) = class(treat);
116 run;
117 /* 해석: 로그-가능도함수를 최대로 하는 lambda-hat이 0이 되어 로그변환을 추천 */
118 /* 0이 아닐때는 (x^lambda - 1)/lambda 변환을 진행 */
```



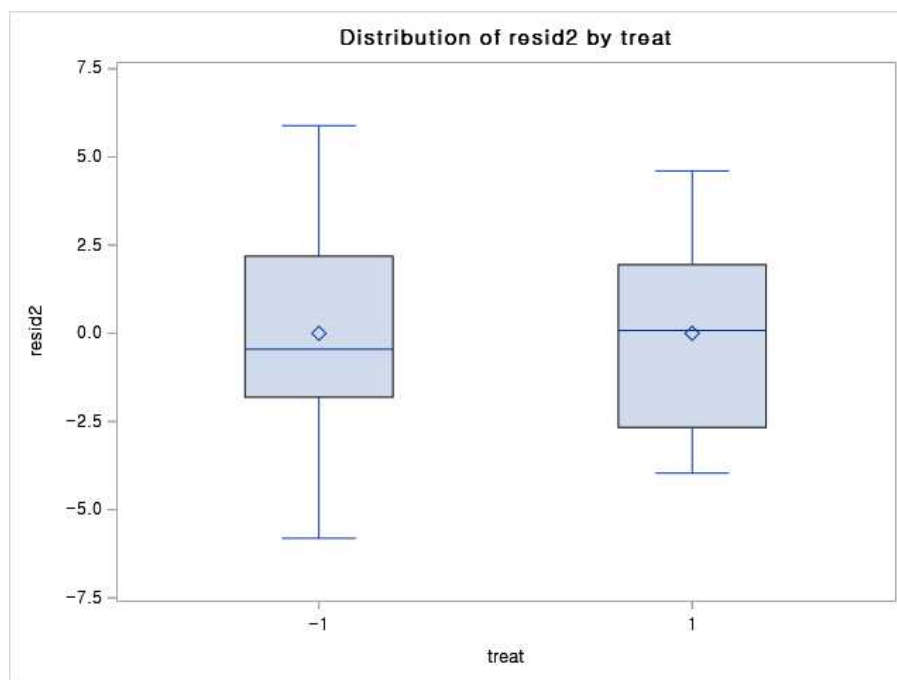
로그 가능도함수를 최대로 하는 $\hat{\lambda}$ 이 0이 되어 로그변환을 추천

c)

```
120 /* c) 데이터 변환 (로그 변환) */
121 data b;
122     set a;
123     logy = log(y);
124 run;
125
126 proc print data = b;
127 run;
128
129 /* c) 변환된 데이터의 잔차 생성 */
130 proc glm data = b;
131     class treat;
132     model logy = treat;
133     output out = c p = pred2 r = resid2;
134 run;
135
136 proc print data = c;
137 run;
```

- 상자 그림

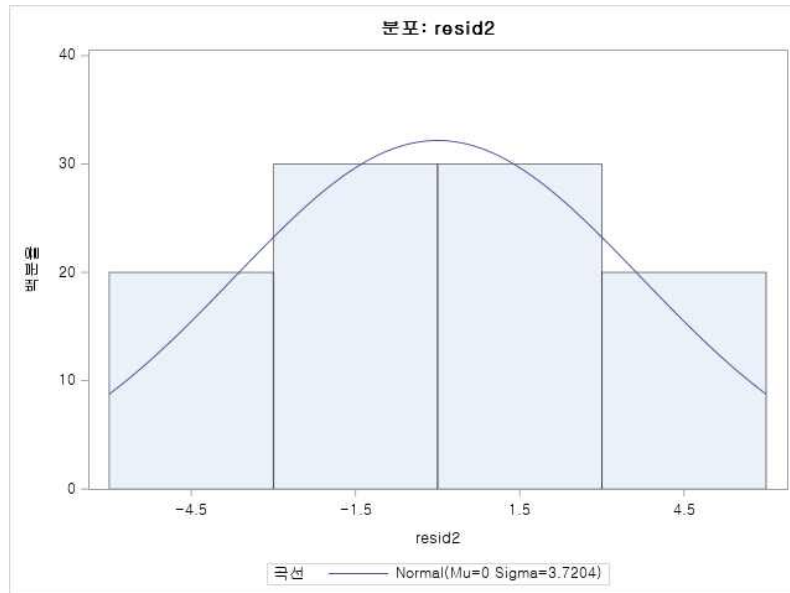
```
138 /* c) 상자그림 */
139 proc boxplot data = c;
140     plot resid2 * treat;
141 run;
```



(잔차가 정규성을 띄는 것으로 보임)

- 히스토그램

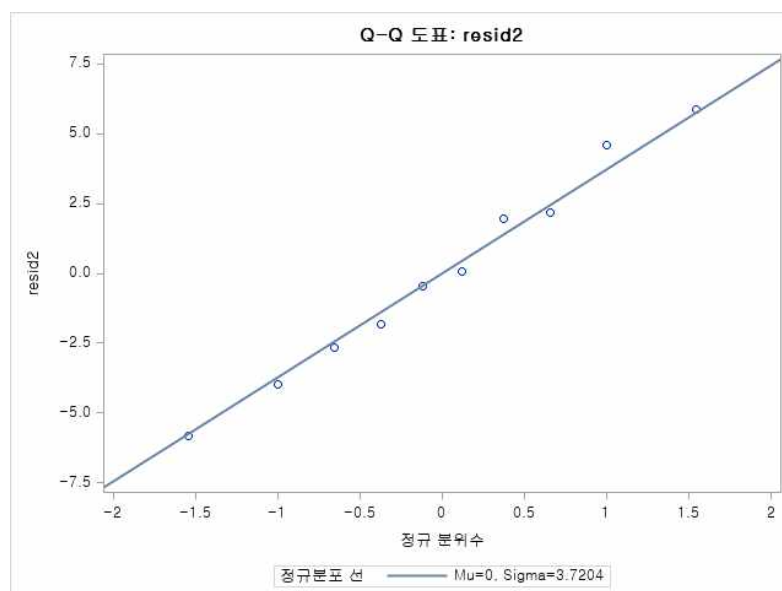
```
143 /* c) 히스토그램 */
144 proc univariate data = c noprint;
145 histogram resid2 / normal(mu = est sigma = est);
146 run;
```



(정규분포 그래프에 근접하는 것으로 보임)

- Q-Q Plot

```
148 /* c) Q-Q plot */
149 proc univariate data = c noprint;
150 qqplot resid2 / normal(mu = est sigma = est);
151 run;
```



(점들이 직선 위에 고르게 분포하는 것으로 보임)

- 샤피로-윌크, 앤더슨-달링 검정

```
153 /* c) 샤피로-윌크 검정/앤더슨-달링 검정 */
154 proc univariate data = c normal;
155 class treat;
156 var resid2;
157 run;
```

H_0 : 잔차가 정규분포를 따른다. H_1 : 잔차가 정규분포를 따르지 않는다.

| 정규성 검정 | | | | |
|--------------------|------|----------|-----------|---------|
| 검정 | 통계량 | | p 값 | |
| Shapiro-Wilk | W | 0.998055 | Pr < W | 0.9980 |
| Kolmogorov-Smirnov | D | 0.140946 | Pr > D | >0.1500 |
| Cramer-von Mises | W-8q | 0.020152 | Pr > W-8q | >0.2500 |
| Anderson-Darling | A-8q | 0.148315 | Pr > A-8q | >0.2500 |

모든 검정이 유의수준 0.05 하에서 귀무가설 채택
(즉, 잔차가 정규분포를 따른다고 할 수 있다)

Problem 3.

```
159 /* 3번 */
160 /* 데이터 생성 */
161 data a;
162 input Machine $ Employee $ Defects;
163 cards;
164 1 1 20
165 1 1 18
166 1 1 14
167 1 2 19
168 1 2 20
169 1 2 20
170 2 1 14
171 2 1 18
172 2 1 14
173 2 2 12
174 2 2 12
175 2 2 9
176 3 1 13
177 3 1 16
178 3 1 13
179 3 2 9
180 3 2 4
181 3 2 4
182 ;
183 run;
```

모형식: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$, ($i = 1, 2, 3$, $j = 1, 2$, $k = 1, 2, 3$)

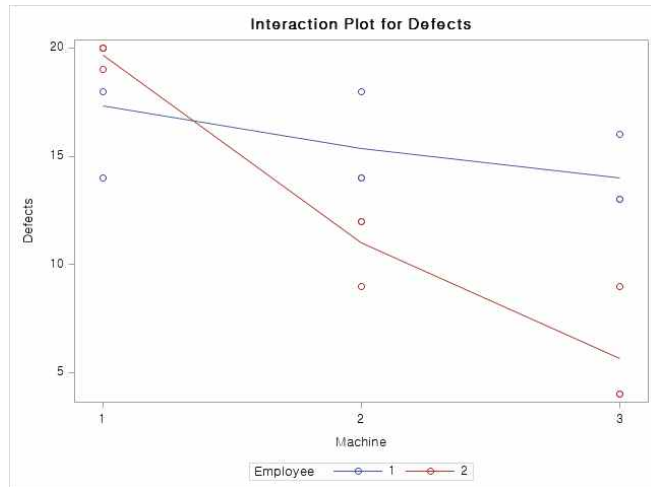
```
188 /* 반복수가 3인 이원배치분산분석 */
189 proc glm data = a;
190 class Machine Employee;
191 model Defects = Machine | Employee;
192 lsmeans Machine * Employee / slice = Machine;
193 run;
```

| The GLM Procedure | | | | | |
|-----------------------------|----|----------------|-------------|---------|--------|
| Dependent Variable: Defects | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 5 | 359.8333333 | 73.9666667 | 15.13 | <.0001 |
| Error | 12 | 58.6666667 | 4.8888889 | | |
| Corrected Total | 17 | 428.5000000 | | | |

| R-Square | Coeff Var | Root MSE | Defects Mean |
|----------|-----------|----------|--------------|
| 0.883088 | 15.98373 | 2.211083 | 13.83333 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|------------------|----|-------------|-------------|---------|--------|
| Machine | 2 | 229.3333333 | 114.6666667 | 23.45 | <.0001 |
| Employee | 1 | 53.3888889 | 53.3888889 | 10.92 | 0.0063 |
| Machine*Employee | 2 | 87.1111111 | 43.5555556 | 8.91 | 0.0042 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|------------------|----|-------------|-------------|---------|--------|
| Machine | 2 | 229.3333333 | 114.6666667 | 23.45 | <.0001 |
| Employee | 1 | 53.3888889 | 53.3888889 | 10.92 | 0.0063 |
| Machine*Employee | 2 | 87.1111111 | 43.5555556 | 8.91 | 0.0042 |



| Source | d.f. | SS | MS | F_0 | p-value |
|-----------|------|---------|---------|-------|---------|
| Machine | 2 | 229.333 | 114.666 | 23.45 | <.0001 |
| Employee | 1 | 53.388 | 53.388 | 10.92 | 0.0063 |
| Mach*Empl | 2 | 87.111 | 43.555 | 8.91 | 0.0042 |
| Error | 12 | 58.666 | 4.888 | | |
| Total | 17 | 428.5 | | | |

H_0 : 기계에 따른 불량품 개수에 차이가 없다. H_1 : 기계에 따른 불량품 개수에 차이가 있다.
 유의수준 0.05 하에서 귀무가설 기각 (즉, 기계에 따른 불량품 개수에 차이가 있다)

H_0 : 작업자에 따른 불량품 개수에 차이가 없다. H_1 : 작업자에 따른 불량품 개수에 차이가 있다.
 유의수준 0.05 하에서 귀무가설 기각 (즉, 작업자에 따른 불량품 개수에 차이가 있다)

H_0 : 기계와 작업자의 교호작용 효과는 없다. H_1 : 기계와 작업자의 교호작용 효과는 있다.
 유의수준 0.05 하에서 귀무가설 기각 (즉, 기계와 작업자의 교호작용 효과는 있다)

| The GLM Procedure | | | | | |
|---|----|----------------|-------------|---------|--------|
| Least Squares Means | | | | | |
| Machine*Employee Effect Sliced by Machine for Defects | | | | | |
| Machine | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| 1 | 1 | 8.1666667 | 8.1666667 | 1.67 | 0.2205 |
| 2 | 1 | 28.1666667 | 28.1666667 | 5.76 | 0.0335 |
| 3 | 1 | 104.1666667 | 104.1666667 | 21.31 | 0.0006 |

H_0 : 1번 기계는 불량품 개수 차이에 영향을 주지 않는다. H_1 : 1번 기계는 불량품 개수 차이에 영향을 준다.

H_0 : 2번 기계는 불량품 개수 차이에 영향을 주지 않는다. H_1 : 2번 기계는 불량품 개수 차이에 영향을 준다.

H_0 : 3번 기계는 불량품 개수 차이에 영향을 주지 않는다. H_1 : 3번 기계는 불량품 개수 차이에 영향을 준다.

2번과 3번 기계가 유의수준 0.05 하에서 귀무가설을 기각하여 불량품 개수 차이에 영향을 주는 기계로 판단 된다.

Problem 4.

```

195 /* 4번 */
196 /* 데이터 생성 */
197 data tool;
198   input A $ B $ C $ lifetime;
199   cards;
200   1 1 1 22
201   1 1 1 31
202   1 1 1 25
203   2 1 1 32
204   2 1 1 43
205   2 1 1 29
206   1 2 1 35
207   1 2 1 34
208   1 2 1 50
209   2 2 1 55
210   2 2 1 47
211   2 2 1 46
212   1 1 2 44
213   1 1 2 45
214   1 1 2 38
215   2 1 2 40
216   2 1 2 37
217   2 1 2 36
218   1 2 2 60
219   1 2 2 50
220   1 2 2 54
221   2 2 2 39
222   2 2 2 41
223   2 2 2 47
224 ;
225 run;
226
227 proc print data = tool;
228 run;

```

1) 세 요인을 모두 고정효과로 간주하고 3 요인 상호작용은 무시한 모형식을 작성하여라.

$$\begin{aligned}
 \text{모형식: } y_{ijkl} &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijkl} \\
 (i &= 1, 2, j = 1, 2, k = 1, 2, l = 1, 2, 3)
 \end{aligned}$$

2) 위 1)번의 분산분석표를 작성하고 (유의수준 0.05) 각 요인의 효과에 대하여 논하라.

```

230 proc glm data = tool;
231   class A B C;
232   model lifetime = A | B | C;
233 run;

```

| The GLM Procedure | | | | | |
|------------------------------|----|----------------|-------------|---------|--------|
| Dependent Variable: lifetime | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 7 | 1612.666667 | 230.380952 | 7.64 | 0.0004 |
| Error | 16 | 482.666667 | 30.166667 | | |
| Corrected Total | 23 | 2095.333333 | | | |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| A | 1 | 0.666667 | 0.666667 | 0.02 | 0.8837 |
| B | 1 | 770.666667 | 770.666667 | 25.55 | 0.0001 |
| A*B | 1 | 16.666667 | 16.666667 | 0.55 | 0.4681 |
| C | 1 | 280.166667 | 280.166667 | 9.29 | 0.0077 |
| A*C | 1 | 468.166667 | 468.166667 | 15.52 | 0.0012 |
| B*C | 1 | 48.166667 | 48.166667 | 1.60 | 0.2245 |
| A*B*C | 1 | 28.166667 | 28.166667 | 0.93 | 0.3483 |

| Source | d.f. | SS | MS | F_0 | p-value |
|--------|------|----------|---------|-------|---------|
| A | 1 | 0.666 | 0.666 | 0.02 | 0.8891 |
| B | 1 | 770.666 | 770.666 | 25.64 | 0.0001 |
| A*B | 1 | 16.666 | 16.666 | 0.55 | 0.4684 |
| C | 1 | 280.166 | 280.166 | 9.32 | 0.0071 |
| A*C | 1 | 468.166 | 468.166 | 15.58 | 0.0010 |
| B*C | 1 | 48.166 | 48.166 | 1.60 | 0.2229 |
| Error | 17 | 510.832 | 30.049 | | |
| Total | 23 | 2095.333 | | | |

H_0 : A에 따른 유효기간에 차이가 없다. H_1 : A에 따른 유효기간에 차이가 있다.

유의수준 0.05 하에서 귀무가설 채택 (즉, 절삭 속도에 따른 유효기간에 차이가 없다)

H_0 : B에 따른 유효기간에 차이가 없다. H_1 : B에 따른 유효기간에 차이가 있다.

유의수준 0.05 하에서 귀무가설 기각 (즉, 도구의 모양에 따른 유효기간에 차이가 있다)

H_0 : C에 따른 유효기간에 차이가 없다. H_1 : C에 따른 유효기간에 차이가 있다.

유의수준 0.05 하에서 귀무가설 기각 (즉, 절삭 각도에 따른 유효기간에 차이가 있다)

H_0 : A와 B의 교호작용 효과는 없다. H_1 : A와 B의 교호작용 효과는 있다.

유의수준 0.05 하에서 귀무가설 채택 (즉, A와 B의 교호작용 효과는 없다)

H_0 : A와 C의 교호작용 효과는 없다. H_1 : A와 C의 교호작용 효과는 있다.

유의수준 0.05 하에서 귀무가설 기각 (즉, A와 C의 교호작용 효과는 있다)

H_0 : B와 C의 교호작용 효과는 없다. H_1 : B와 C의 교호작용 효과는 있다.

유의수준 0.05 하에서 귀무가설 채택 (즉, B와 C의 교호작용 효과는 없다)