

확률과통계 12주차 과제 모범답안

1. 브랜드 선호도(Y)는 수분 함유량(X1)과 당도(X2)와 관련이 있다고 한다. 다중회귀분석을 통해 수분 함유량과 당도의 값에 대하여 브랜드 선호도를 예측하기 위하여 실험을 실시하고 자료를 수집하였다.

Y	X1	X2
64	4	2
73	4	4
61	4	2
76	4	4
72	6	2
80	6	4
71	6	2
83	6	4
83	8	2
89	8	4
86	8	2
93	8	4
88	10	2
95	10	4
94	10	2
100	10	4

- (1) 산점도 매트릭스와 상관계수 매트릭스를 구하라. (5 점)

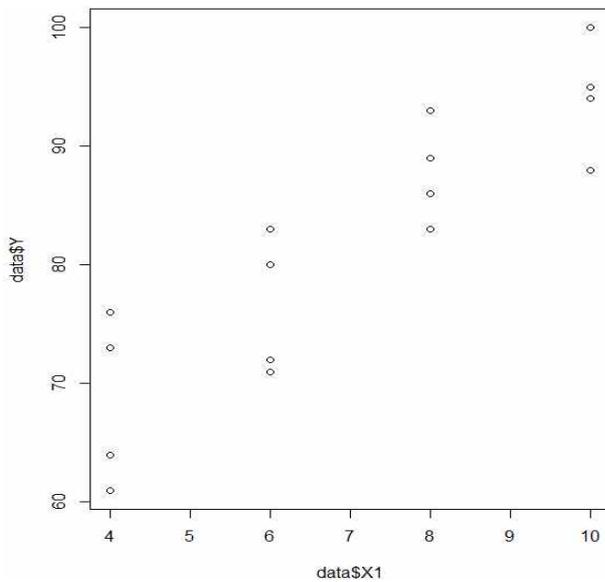


그림 1. X_1 산점도 매트릭스

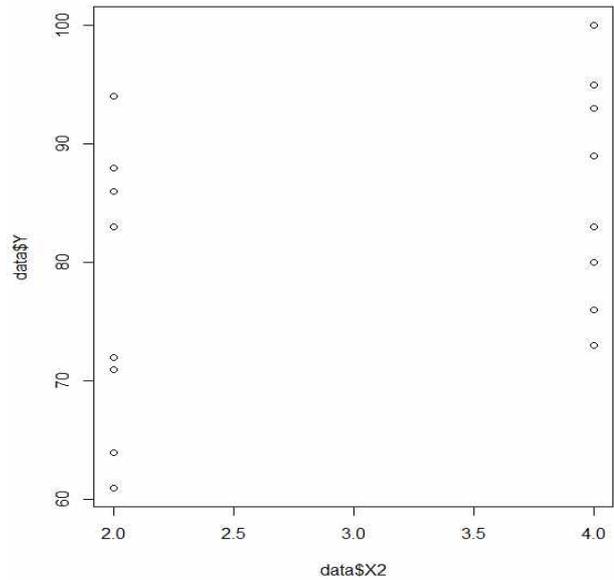


그림 2. X_2 산점도 매트릭스

표 1. 상관계수 매트릭스

	Y	X_1	X_2
Y	1.0000000	0.8923929	0.3945807
X_1	0.8923929	1.0000000	0.0000000
X_2	0.3945807	0.0000000	1.0000000

(2) 다중회귀분석(추정, 검정, 회귀진단 등 포함)을 실시하고, 최종 모형에 대해 설명하라 (15점)

```

Residuals:
    Min       1Q   Median       3Q      Max
-4.400 -1.762  0.025  1.587  4.200

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.6500    2.9961  12.566 1.20e-08 ***
x1           4.4250    0.3011  14.695 1.78e-09 ***
x2           4.3750    0.6733   6.498 2.01e-05 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.693 on 13 degrees of freedom
Multiple R-squared:  0.9521,    Adjusted R-squared:  0.9447
F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09

```

다중회귀 분석 결과 추정된 모델은 다음과 같다.

$$\text{추정: } \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} = 37.65 + 4.425X_{1i} + 4.375X_{2i}$$

검정 1: $H_0: \beta_0 = 0$
 $H_1: \beta_0 \neq 0$

β_0 의 P -value = $1.20e-08$ 이고,
 $1.20e-08 < 0.05$ 이므로 귀무가설 기각

검정 2: $H_0: \beta_1 = 0$
 $H_1: \beta_1 \neq 0$

β_1 의 P -value = $1.78e-09$ 이고,
 $1.78e-09 < 0.05$ 이므로 귀무가설 기각

검정 3: $H_0: \beta_2 = 0$
 $H_1: \beta_2 \neq 0$

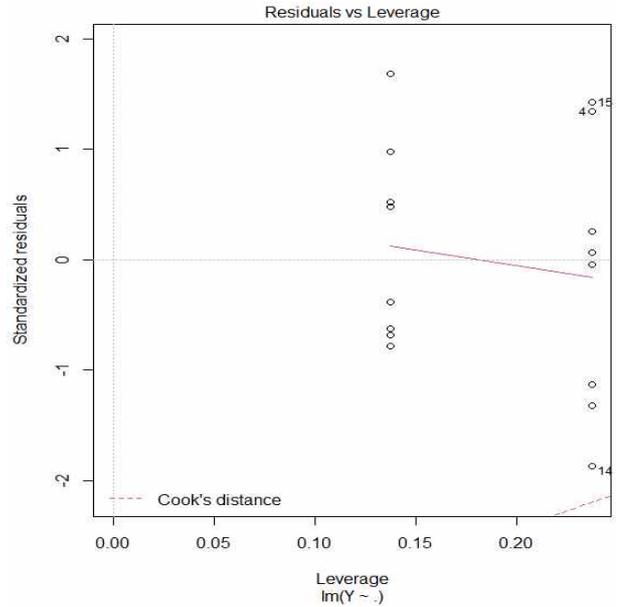
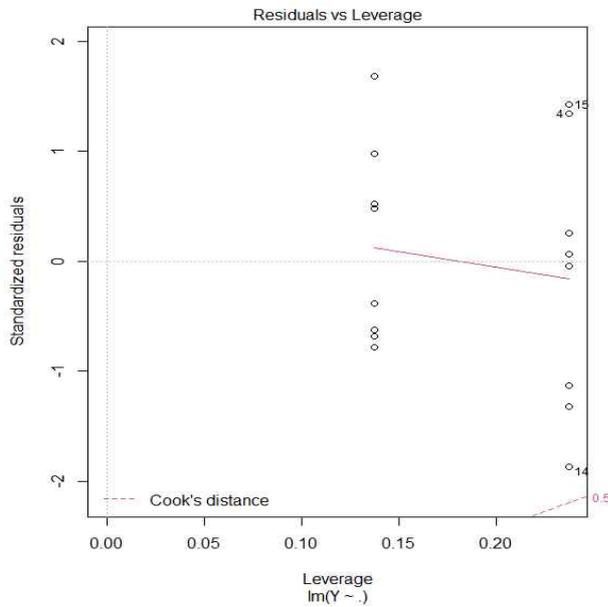
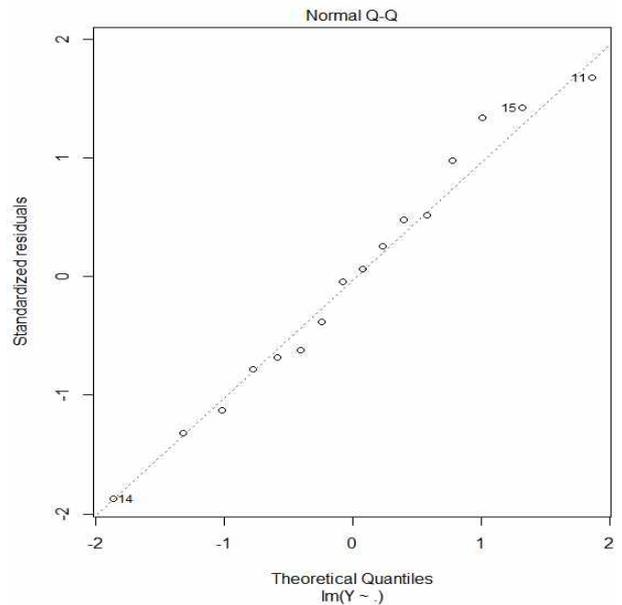
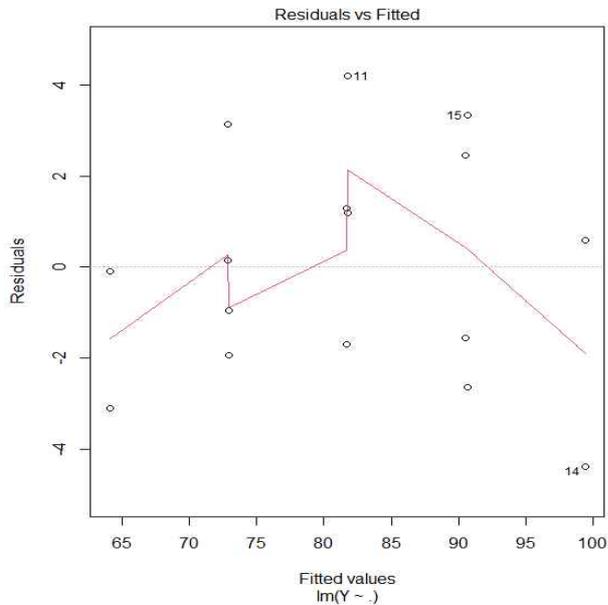
β_2 의 P -value = $2.01e-005$ 이고,
 $2.01e-05 < 0.05$ 이므로 귀무가설 기각

검정 결과 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ 모두 유의하다고 할 수 있다.

최종모형을 해석하면 X_1 이 1 증가할 때 마다 Y 값에 4.425 만큼 영향을 준다고 할 수 있고

X_2 이 1 증가할 때 마다 Y 값에 4.375 만큼 영향을 준다고 할 수 있다.

회귀진단:



shapiro-wilk normality test

```
data: model$residuals  
w = 0.97585, p-value = 0.9222
```

회귀진단 결과 대체로 선형성, 정규성, 독립성, 등분산성을 만족한다.