

초보 개발자의 성장기

대학교 2-2/회귀

5.4 연습문제

Launa 2024. 11. 28. 19:11



표 4.4 데이터에 대한 선형회귀 모형을 적합하고 다음에 답하시오

<표 4.4>

환자번호	콜레스트롤(mg/100ml)	몸무게(kg)	나이(year)
1	354	84	46
2	190	73	20
3	405	65	52
4	263	73	30
5	451	76	57
6	302	69	25
7	288	63	28
8	385	72	36
9	402	79	57
10	365	75	44
11	209	27	24
12	290	89	31
13	346	65	52
14	254	57	23
15	395	59	60
16	434	69	48

17	220	60	34
18	374	79	51
19	308	75	50
20	220	82	34
21	311	59	46
22	181	67	23
23	274	85	37
24	303	55	40
25	244	63	30

(a) 가정에 대한 검토로 잔차를 이용하여 등분산성을 점검하시오

```
> cholesterol = data.frame(
+   환자번호 = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
25),
+   콜레스테롤 = c(354, 190, 405, 263, 451, 302, 288, 385, 402, 365, 209, 290, 346, 254, 395,
434, 220, 374, 308, 220, 311, 181, 274, 303, 244),
+   몸무게 = c(84, 73, 65, 70, 76, 69, 63, 72, 79, 75, 27, 89, 65, 57, 59, 69, 60, 79, 75, 82, 59,
67, 85, 55, 63),
+   나이 = c(46, 20, 52, 30, 57, 25, 28, 36, 57, 44, 24, 31, 52, 23, 60, 48, 34, 51, 50, 34, 46, 23,
37, 40, 30)
+ )
> print(cholesterol)
```

환자번호 콜레스테롤 몸무게 나이

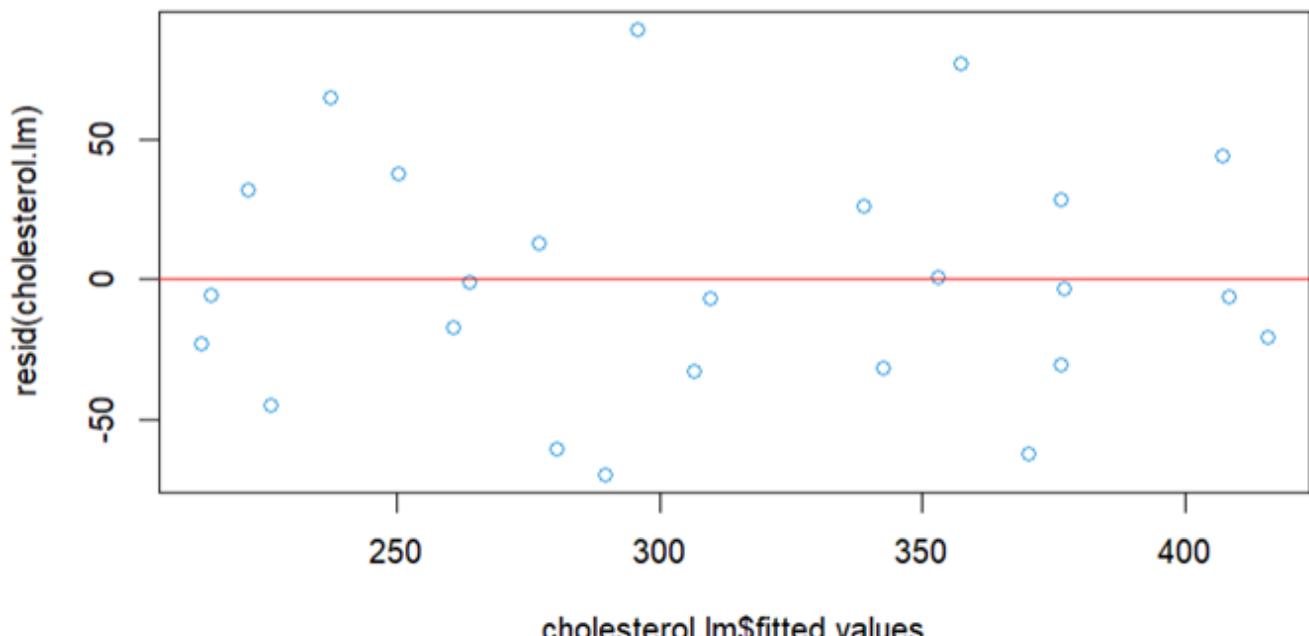
1	1	354	84	46
2	2	190	73	20
3	3	405	65	52
4	4	263	70	30
5	5	451	76	57
6	6	302	69	25
7	7	288	63	28
8	8	385	72	36
9	9	402	79	57

10	10	365	75	44
11	11	209	27	24
12	12	290	89	31
13	13	346	65	52
14	14	254	57	23
15	15	395	59	60
16	16	434	69	48
17	17	220	60	34
18	18	374	79	51
19	19	308	75	50
20	20	220	82	34
21	21	311	59	46
22	22	181	67	23
23	23	274	85	37
24	24	303	55	40
25	25	244	63	30

```
> cholesterol.lm = lm(콜레스테롤 ~ 몸무게 + 나이, data = cholesterol)
```

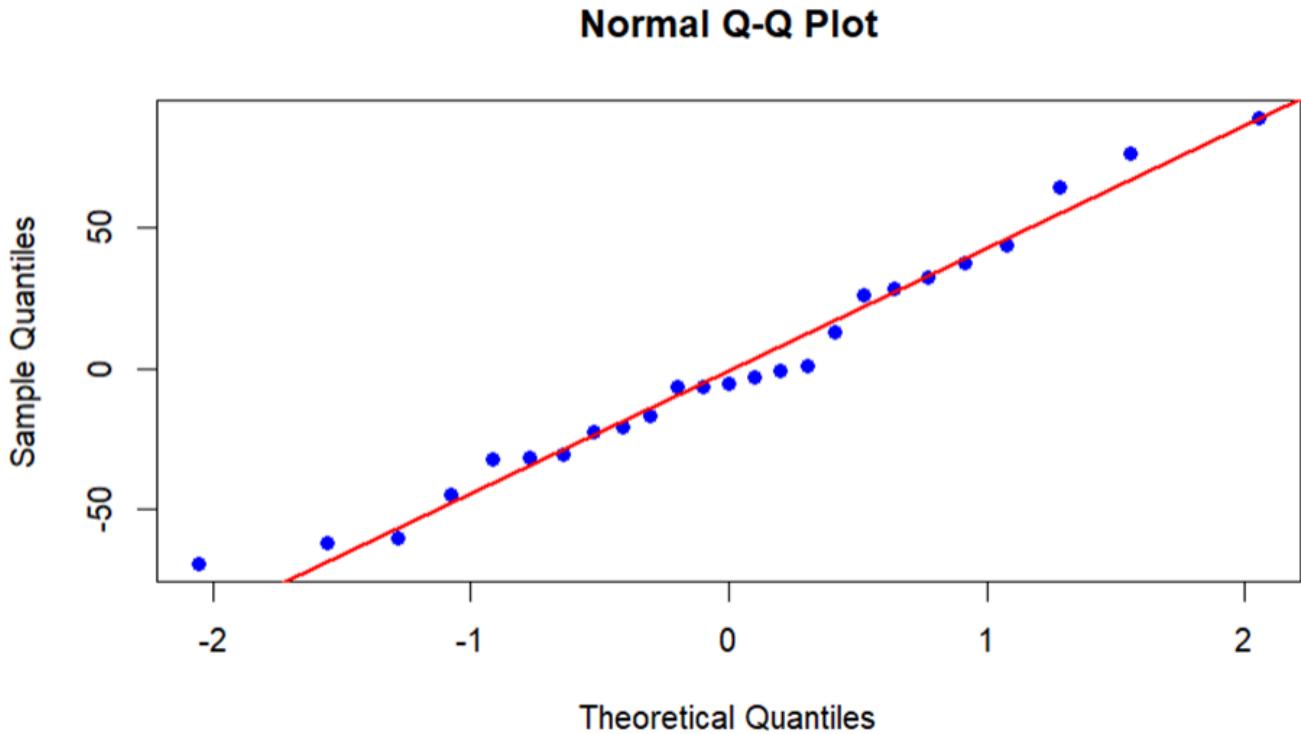
```
> plot(cholesterol.lm$fitted.values, resid(cholesterol.lm), col=4)
```

```
> abline(h = 0, col = "red")
```



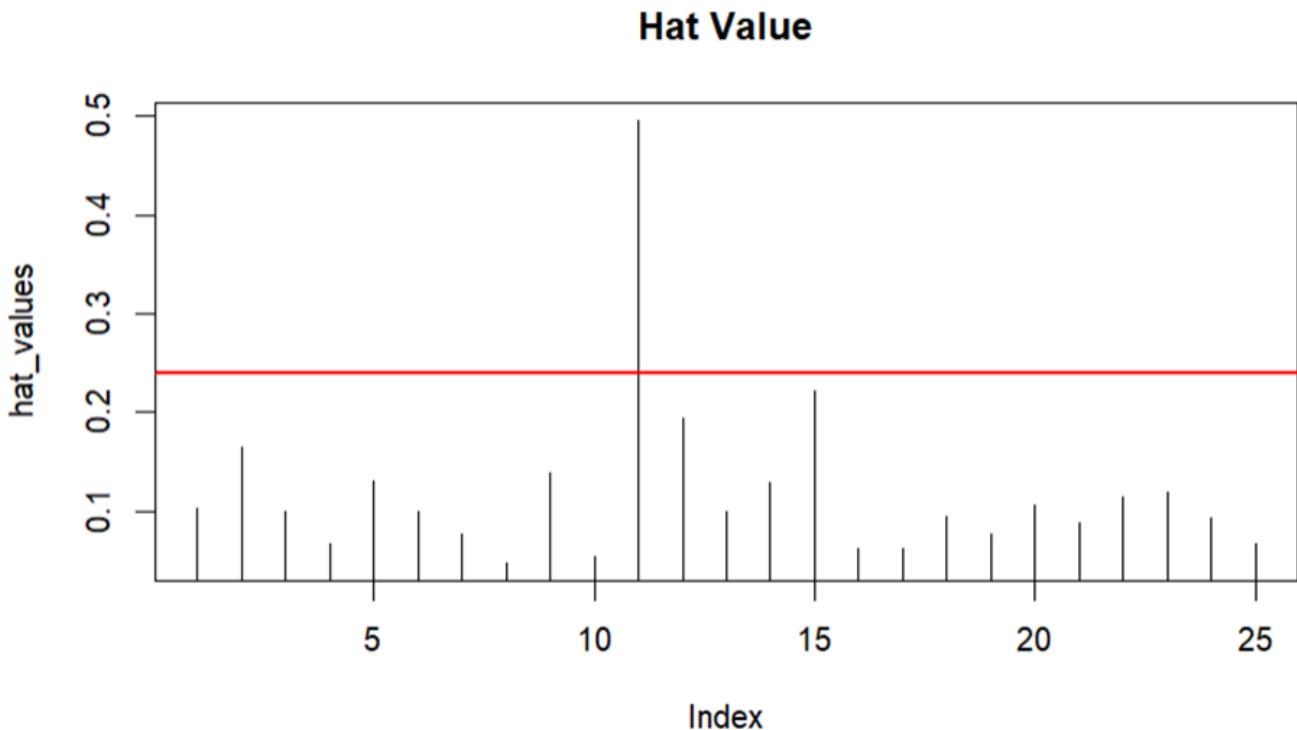
(b) 가정에 대한 검토로 잔차를 이용하여 정규성을 점검하시오

```
> residuals=residuals(cholesterol.lm)
> qqnorm(residuals,pch = 19, col = "blue")
> qqline(residuals, col = "red", lwd = 2)
```



(c) 지렛대점이 있는지 점검하시오

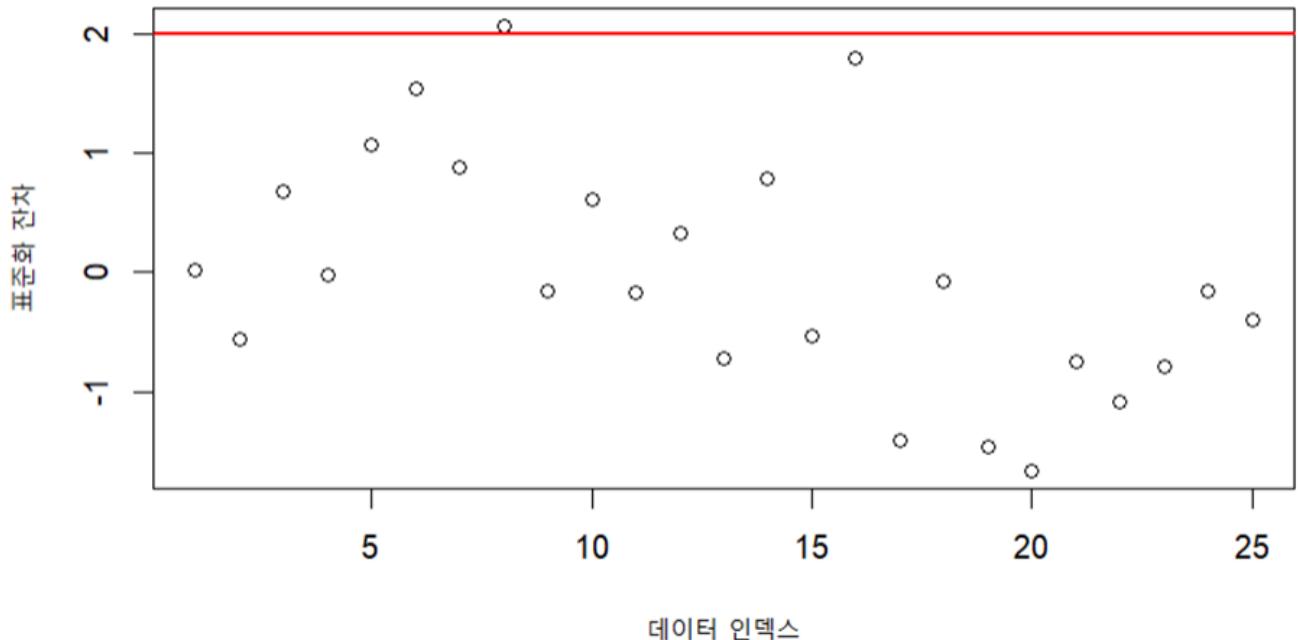
```
> hat_values = hatvalues(cholesterol.lm)
> n=nrow(cholesterol)
> p=length(coef(cholesterol.lm))-1
> threshold = 2 * (p + 1) / n
> plot(hat_values, type = "h", main = "Hat Value")
> abline(h = threshold, col = "red", lwd = 2)
```



(d) 이상점이 있는지 검토하시오

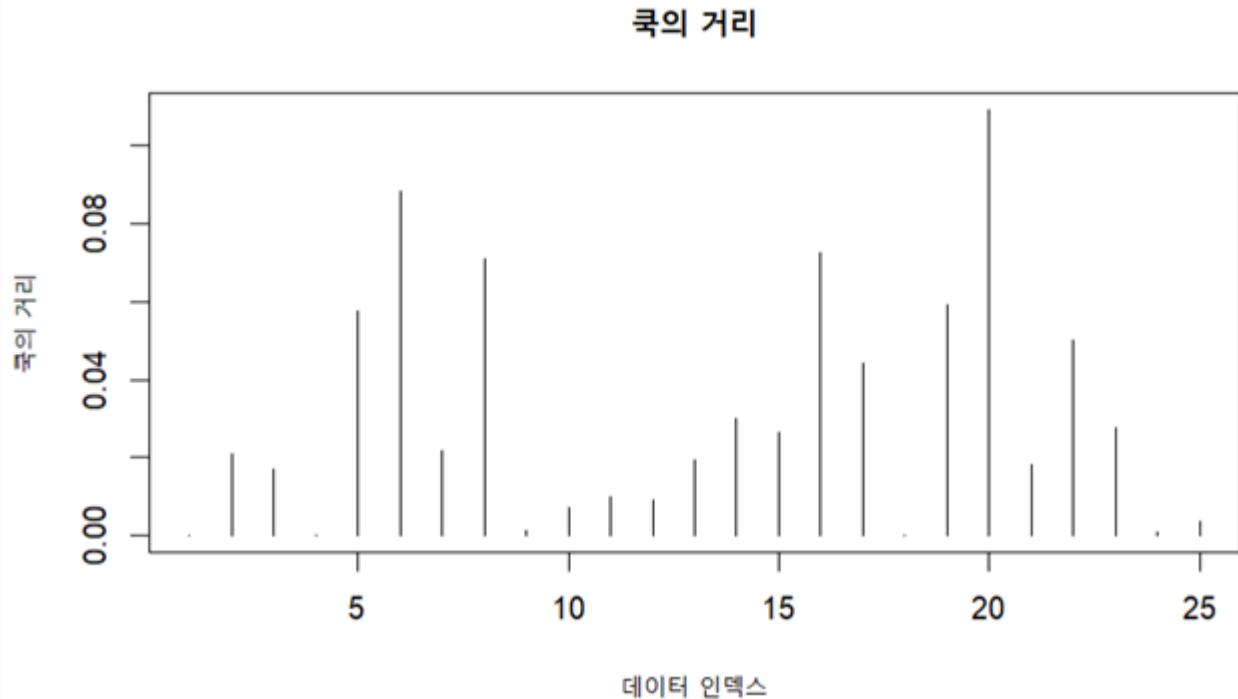
```
> standardized_residuals = rstandard(cholesterol.lm)
> outliers_residuals = which(abs(standardized_residuals) > 2)
> outliers_residuals
8
> plot(standardized_residuals, main = "표준화 잔차", ylab = "표준화 잔차", xlab = "데이터 인덱스")
> abline(h = c(-2, 2), col = "red", lwd = 2)
```

표준화 잔차



(e) 영향점이 있는지 검토하시오

```
> cooks_distance = cooks.distance(cholesterol.lm)
> influential_points = which(cooks_distance > 0.5)
> outliers_cooks
named integer(0)
> plot(cooks_distance, type = "h", main = "쿡의 거리", xlab = "데이터 인덱스", ylab = "쿡의 거리")
> abline(h = 0.5, col = "red", lty = 2)
```



(f) 설명변수들과 반응변수 간의 회귀식이 적합한지 진단하시오

> `summary(swiss.lm)`

Call:

`lm(formula = Fertility ~ ., data = swiss)`

Residuals:

Min	1Q	Median	3Q	Max
-15.2743	-5.2617	0.5032	4.1198	15.3213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	66.91518	10.70604	6.250	1.91e-07
Agriculture	-0.17211	0.07030	-2.448	0.01873
Examination	-0.25801	0.25388	-1.016	0.31546
Education	-0.87094	0.18303	-4.758	2.43e-05
Catholic	0.10412	0.03526	2.953	0.00519
Infant.Mortality	1.07705	0.38172	2.822	0.00734

(Intercept) ***

Agriculture *

Examination

Education ***

Catholic **

Infant.Mortality **

Signif. codes:

0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.165 on 41 degrees of freedom

Multiple R-squared: 0.7067, Adjusted R-squared: 0.671

F-statistic: 19.76 on 5 and 41 DF, p-value: 5.594e-10

공감

초보 개발자의 성장기

Launa 님의 블로그입니다.

댓글 0

Launa

내용을 입력하세요.

등록