

## 초보 개발자의 성장기

대학교 2-2/회귀

### 3.9 연습문제

Launa 2024. 10. 16. 16:09

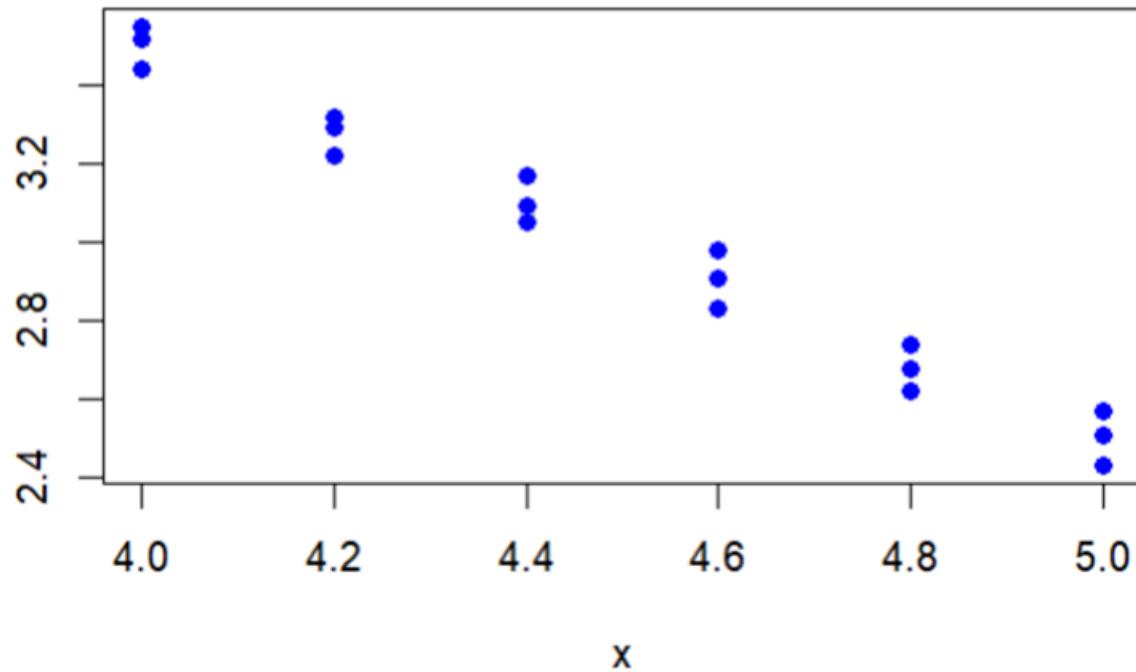


화학 오염 물질에 대한 메기의 민감성을 알아보기 위해 실험한 결과이다  
2리터의 물에 오염물질의 양을 달리하여 생존시간을 측정하여 데이터를 받았다

(A) 오염물질양(X)와 생존시간(Y) 간의 산점도를 그리시오

>

```
y=c(2.51,2.57,2.43,2.62,2.74,2.68,2.83,2.91,2.98,3.17,3.05,3.09,3.32,3.22,3.29,3.44,3.52,3.55)  
> x=c(5.0,5.0,5.0,4.8,4.8,4.8,4.6,4.6,4.6,4.4,4.4,4.4,4.4,4.2,4.2,4.2,4.0,4.0,4.0)  
> plot(x,y,pch = 19, col = "blue")
```



(B) 오염물질양(X)과 생존시간(Y)간의 상관계수를 구하시오

```
> cor(x, y)
```

```
[1] -0.9882052
```

상관계수는 cor()함수로 구한다

(C) 단순회귀모형  $Y = \beta_0 + \beta_1 X + \varepsilon$ ,  $\varepsilon \sim \text{iid } N(0, \sigma^2)$  을 적합하시오

```
> catfish.lm=lm(y ~ x)
```

```
> summary(catfish.lm)
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.076127	-0.052294	0.004254	0.039254	0.084254

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
----------	------------	---------	----------

```
(Intercept) 7.48698 0.17450 42.91 < 2e-16 ***

```

```
x -0.99810 0.03867 -25.81 1.82e-14 ***
```

```
---
```

```
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 0.05603 on 16 degrees of freedom

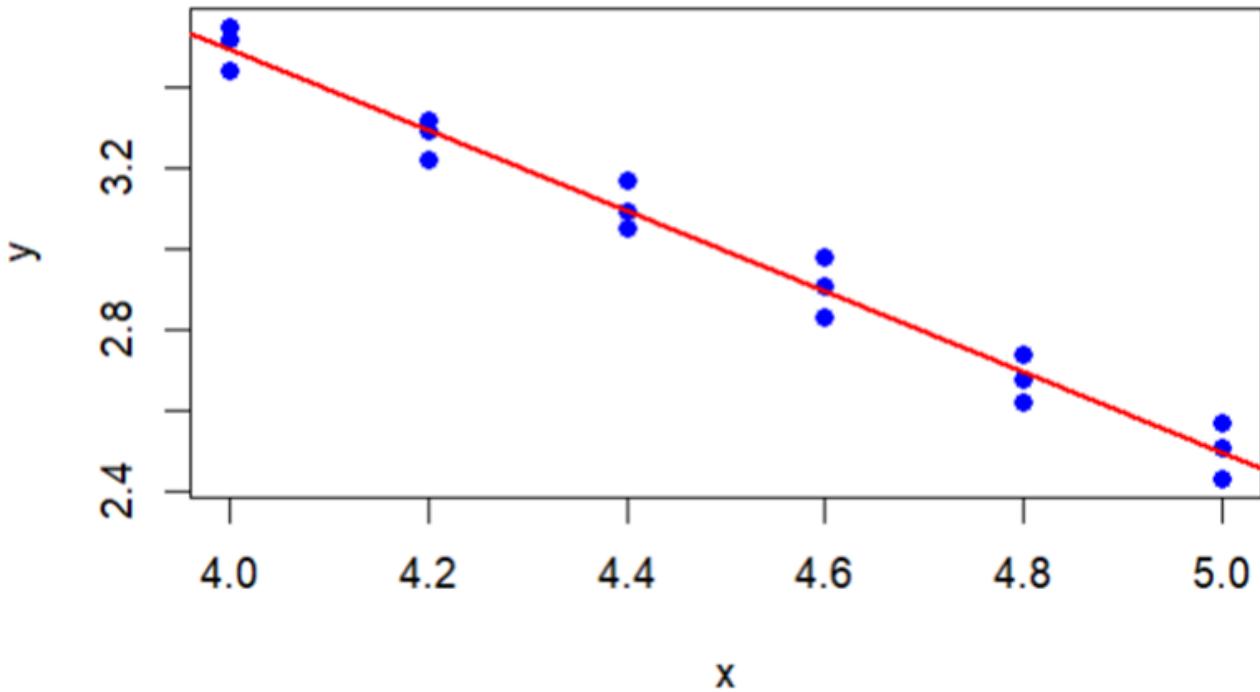
Multiple R-squared: 0.9765, Adjusted R-squared: 0.9751

F-statistic: 666.3 on 1 and 16 DF, p-value: 1.815e-14

lm을 사용해 단순회귀 모형을 쉽게 적합할수 있다

(d) (c)에서 구한 회귀식을 산점도와 같은 평면에 그리시오

```
> abline(catfish.lm, col = "red", lwd = 2)
```



(e) 기울기가 유의한지 유의수준  $\alpha = 0.05$ 에서 검정하시오

$H_0: \beta_1 = 0$  에 대해  $H_1 = \beta_1 \neq 0$

```
> tb1 = abs(summary(catfish.lm)$coefficients[2, "t value"])
> df = summary(catfish.lm)$df[2]
> ptb1=2 * pt(tb1, df, lower.tail = FALSE)
> ptb1
[1] 1.815094e-14
```

t통계량 계산

잔차의 자유도 확인

`df[1]`: 회귀 자유도 (회귀 모형에서 추정된 계수의 수)

`df[2]`: 잔차 자유도 (모형을 적합하고 남은 데이터 포인트의 자유도)

p value 계산

## p-값 계산 방법

1. 단측 검정:

- 만약 귀무가설이  $H_0 : \beta_1 \leq 0$  또는  $H_0 : \beta_1 \geq 0$ 와 같은 단측 검정이라면:

$$p = pt(t\text{-값}, df, lower.tail = FALSE)$$

- 여기서 `lower.tail = FALSE` 는 t-값이 양수일 때 사용하며, 음수일 때는 `lower.tail = TRUE`로 설정합니다.

2. 양측 검정:

- 귀무가설이  $H_0 : \beta_1 = 0$ 와 같은 양측 검정인 경우:

$$p = 2 \times pt(abs(t\text{-값}), df, lower.tail = FALSE)$$

- `abs(t-값)`은 t-값의 절대값을 사용하며, 양측 검정을 위해 p-값을 두 배로 계산합니다.

(f) 기울기에 대한 95% 신뢰구간을 구하시오

```
> confint(catfish.lm, level = 0.95)
```

2.5 % 97.5 %

(Intercept) 7.117056 7.8569127

x -1.080066 -0.9161247

`confint(catfish.lm, level = 0.95)` : 선형 회귀 모형의 각 계수에 대한 신뢰구간을 계산

(Intercept):  $\beta_0$

$x : \beta_1$

(g) 결정계수를 구하고 해석하시오

```
> r_squared = summary(catfish.lm)$r.squared
```

```
> r_squared
```

```
[1] 0.9765494
```

summary(catfish.lm)\$r.squared : 결정계수 구하기

(h) 잔차 e를 구하고 잔차제곱합을 구하시오

```
> residuals(catfish.lm)
```

1	2	3	4	5
---	---	---	---	---

0.013492063	0.073492063	-0.066507937	-0.076126984	0.043873016
-------------	-------------	--------------	--------------	-------------

6	7	8	9	10
---	---	---	---	----

-0.016126984	-0.065746032	0.014253968	0.084253968	0.074634921
--------------	--------------	-------------	-------------	-------------

11	12	13	14	15
----	----	----	----	----

-0.045365079	-0.005365079	0.025015873	-0.074984127	-0.004984127
--------------	--------------	-------------	--------------	--------------

16	17	18		
----	----	----	--	--

-0.054603175	0.025396825	0.055396825		
--------------	-------------	-------------	--	--

```
> deviance(catfish.lm)
```

```
[1] 0.05023683
```

residuals(catfish.lm): 잔차 구하기

deviance(catfish.lm): 잔차제곱합 구하기

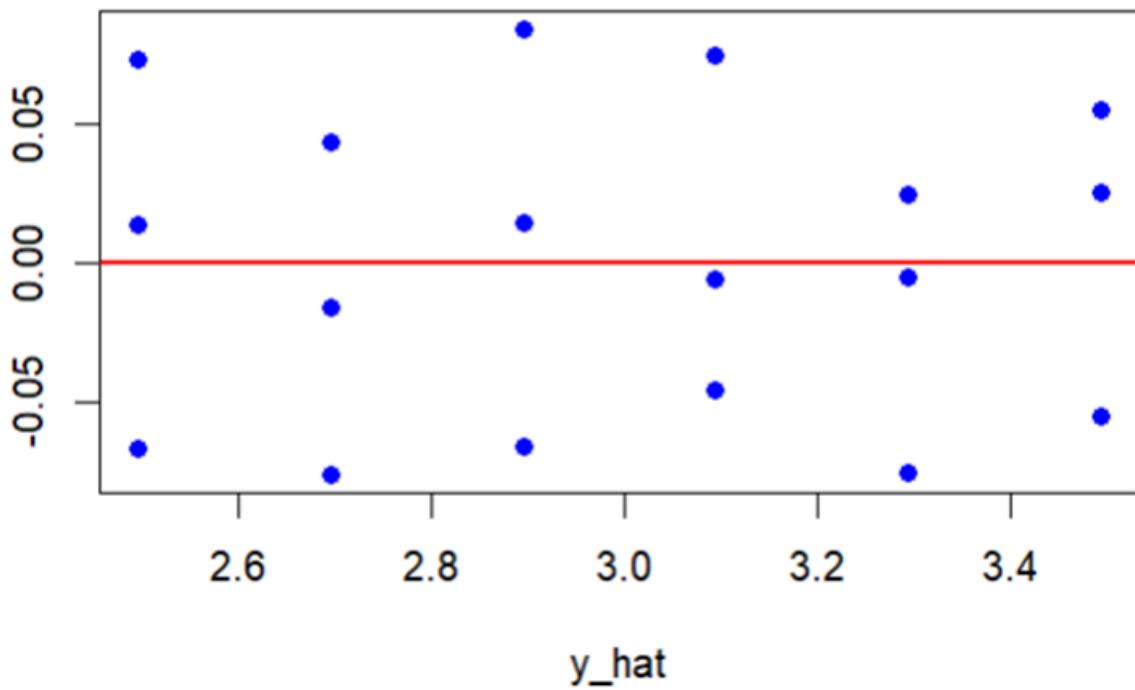
(i) 잔차그림 ( $\hat{Y}$ , e) 을 그리고 오차의 독립성에 대해 설명하시오

```
> y_hat=fitted(catfish.lm)
```

```
> residuals=residuals(catfish.lm)
```

```
> plot(y_hat, residuals,pch = 19, col = "blue")
```

```
> abline(h = 0, col = "red", lwd = 2)
```



`y_hat=fitted(catfish.lm)` : 예측값  $\hat{Y}$  구하기

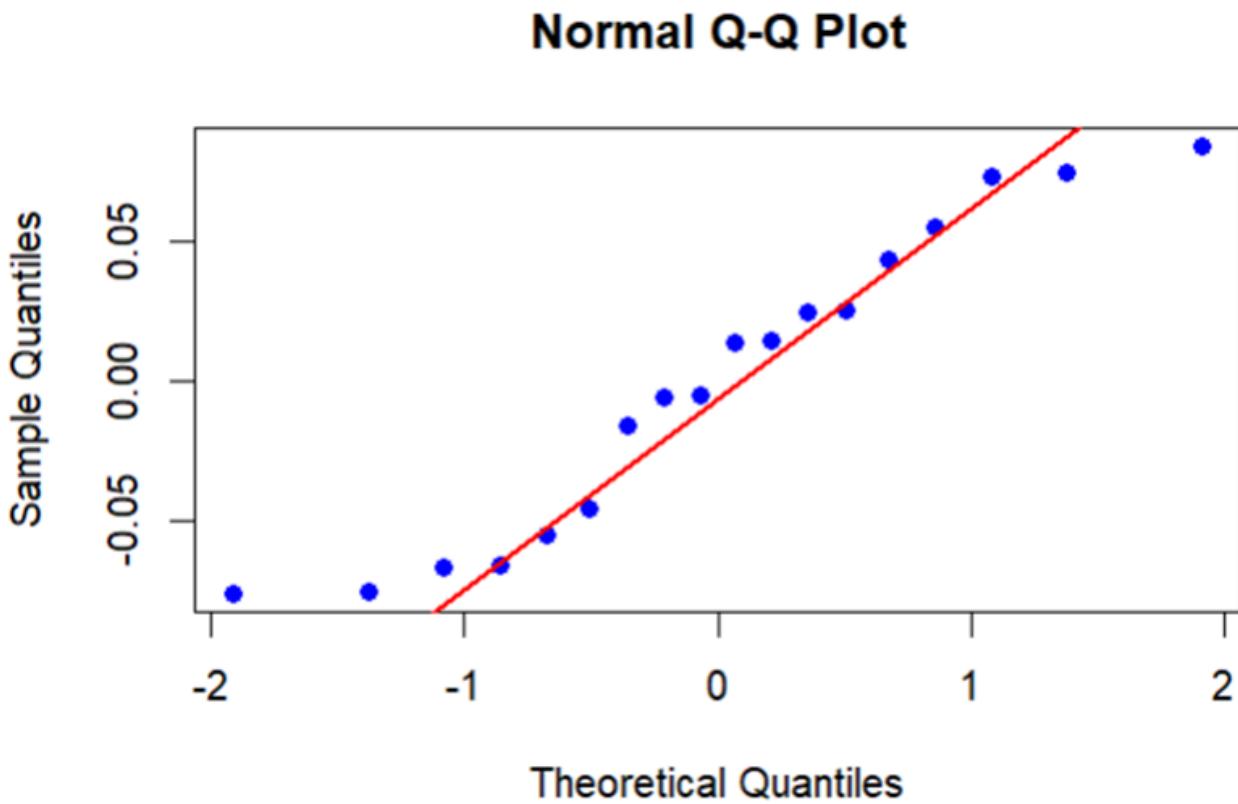
`residuals = residuals(catfish.lm)`: 잔차( $e$ ) 구하기

`abline(h = 0, col = "red", lwd = 2)`: 기준선 추가

(j) 잔차에 대한 정규 Q-Q그림을 그리고 설명하시오

```
> qqnorm(residuals, pch = 19, col = "blue")
```

```
> qqline(residuals, col = "red", lwd = 2)
```



`qqnorm(residuals, pch = 19, col = "blue")` : 정규 Q-Q 그림을 생성

`qqnorm()` 함수: 잔차의 정규 Q-Q 그림을 생성

Q-Q 그림은 데이터를 정규분포와 비교하기 위해 사용

residuals는 잔차 값을 나타냄

`qqline(residuals, col = "red", lwd = 2)`: 기준선 추가

(k) 오염물질 양이 5.5 일때 생존시간을 예측하시오

```
> new_data = data.frame(x = 5.5)
> predicted_y = predict(catfish.lm, newdata = new_data)
> predicted_y
```

1

1.99746

3.10

`new_data = data.frame(x = 5.5)`: 변수가 5.5인 데이터프레임 생성

`predict(catfish.lm, newdata = new_data)`: 데이터 프레임 대입

공감

**대학교 2-2>회귀** 카테고리의 다른 글[5.4 연습문제 \(0\)](#)

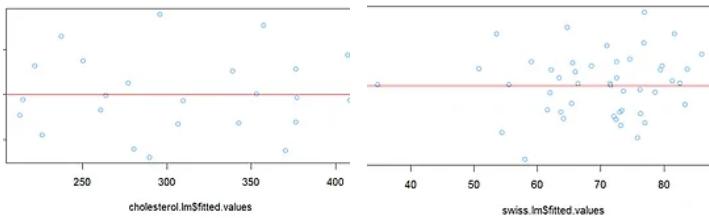
2024.11.28

[5.3 연습문제 \(0\)](#)

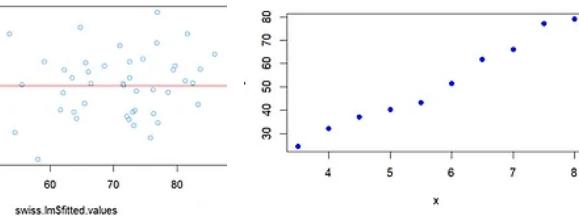
2024.11.28

[3.10 연습문제 \(2\)](#)

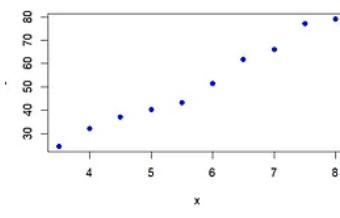
2024.11.08

**'대학교 2-2/회귀' Related Articles**

5.4 연습문제



5.3 연습문제



3.10 연습문제

**초보 개발자의 성장기**

Launa 님의 블로그입니다.

댓글 0

Launa

내용을 입력하세요.

등록

DESIGN BY TISTORY 관리자