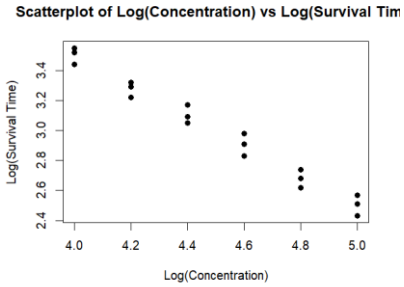
<회귀분석 3장 연습문제 과제_20221612김서윤>

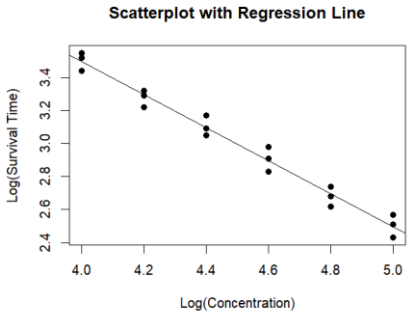
3.9)

```
> # 3.9 메기 민감성 데이터 정의
> 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, 4.2, 4.2, 4.2, 4.0, 4.0, 4.0)
> 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, 3.4
4, 3.52, 3.55)
>
> # (a) 오염물질양(X)과 생존시간(Y) 간의 산점도 그리기
> plot(X, Y, main="Scatterplot of Log(Concentration) vs Log(Survival Time)", xlab="Log(Concentratio
n)", ylab="Log(Survival Time)", pch=19)
>
```



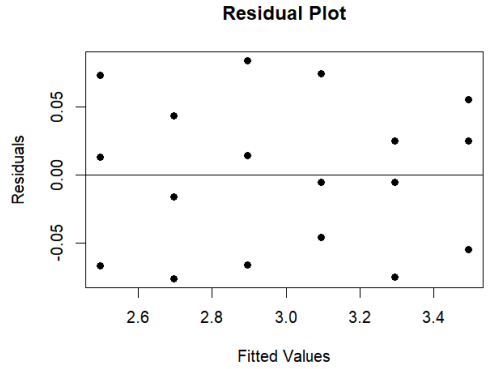

```
> # (b) 오염물질양(X)과 생존시간(Y) 간의 상관계수 구하기
> correlation <- cor(X, Y)
> correlation
[1] -0.9882052
>
> # (c) 단순회귀모형 적합하기
> model <- lm(Y ~ X)
> summary(model)

Call:
lm(formula = Y ~ X)

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

Coefficients:
            Estimate Std. Error t value
(Intercept)  7.48698    0.17450   42.91
X           -0.99810    0.03867  -25.81
            Pr(>|t|)
(Intercept)  < 2e-16 ***
X           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
```

```
> # (d) 산점도와 회귀식 그리기
> plot(X, Y, main="Scatterplot with Regression Line", xlab="Log(Concentration)", ylab="Log(Survival Time)", pch=19)
> abline(model)  # 회귀선 추가
```

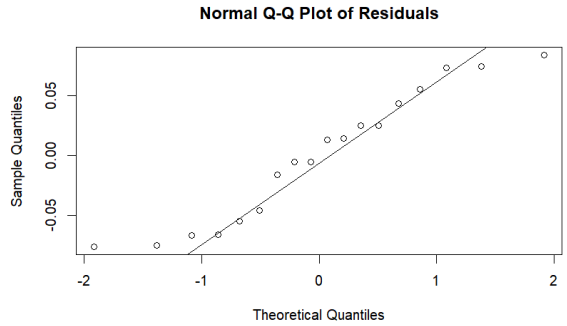



```
> # (e) 기울기가 유의한지 검정하기 (유의수준 α=0.05)
> summary(model)  # 회귀모형의 요약 정보에서 p-value 확인

Call:
lm(formula = Y ~ X)

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

Coefficients:
            Estimate Std. Error t value
(Intercept)  7.48698    0.17450   42.91
X           -0.99810    0.03867  -25.81
            Pr(>|t|)
(Intercept)  < 2e-16 ***
X           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

>
> # (f) 기울기에 대한 95% 신뢰구간 구하기
> confint(model, level=0.95)
                2.5 %     97.5 %
(Intercept)  7.117056  7.8569127
X           -1.080066 -0.9161247
>
> # (g) 결정계수 구하고 해석하기
> R_squared <- summary(model)$r.squared
> R_squared
[1] 0.9765494
```

```
> # (h) 잔차 e를 구하고 잔차제곱합 구하기
> residuals <- residuals(model)
> residual_sum_of_squares <- sum(residuals^2)
> residual_sum_of_squares
[1] 0.05023683
>
> # (i) 잔차그림 그리고 오차의 독립성 설명하기
> plot(fitted(model), residuals, main="Residual Plot", xlab="Fitted Values", ylab="Residuals", pch=19)
> abline(h=0)  # 잔차의 기준선
>
```



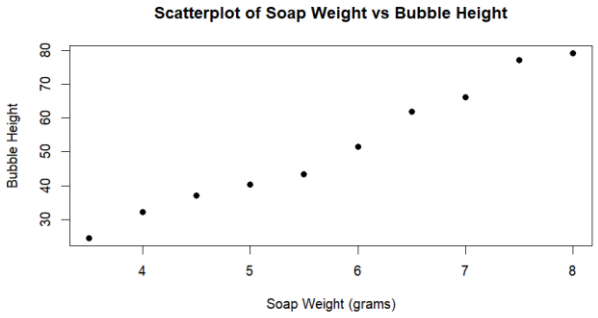

```
> # (j) 잔차에 대한 정규 Q-Q 그림 그리기
> qqnorm(residuals, main="Normal Q-Q Plot of Residuals")
> qqline(residuals)
```




```
> # (k) 오염물질양이 5.5일 때 생존시간 예측하기
> new_data <- data.frame(X=5.5)
> predicted_Y <- predict(model, newdata=new_data)
> predicted_Y
      1
1.99746
```

3.10)

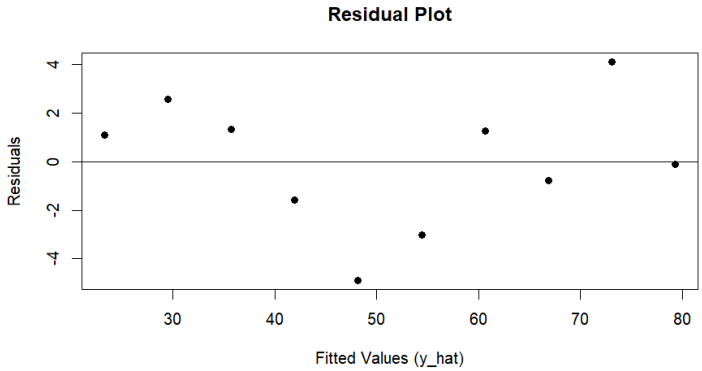
```
> # 3.10 비누 데이터 정의
> X <- c(3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0)  # 비누 중량
> Y <- c(24.4, 32.1, 37.1, 40.4, 43.3, 51.4, 61.9, 66.1, 77.2, 79.2)  # 비누거품 높이
>
> # (a) (X, Y) 산점도를 그리시오
> plot(X, Y, main="Scatterplot of Soap Weight vs Bubble Height", xlab="Soap Weight (grams)",
ylab="Bubble Height", pch=19)
```



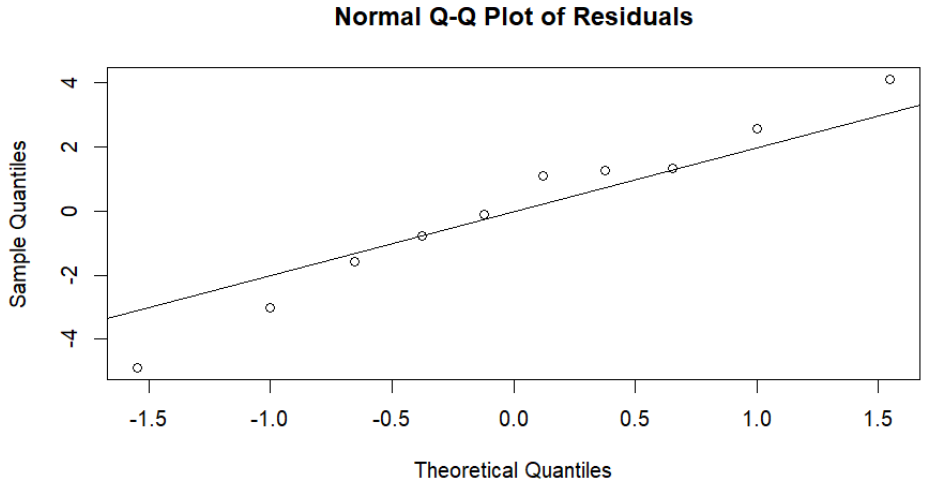

```
> # (b) 통계적 모형 Y = B1 * X + e 를 적합하시오
> model <- lm(Y ~ X)
> summary(model)

Call:
lm(formula = Y ~ X)

Residuals:
   Min     1Q Median     3Q    Max
-4.899 -1.374  0.490  1.322  4.116

Coefficients:
            Estimate Std. Error t value
(Intercept) -20.2339     3.6995  -5.469
X            12.4424     0.6242  19.933
            Pr(>|t|)
(Intercept) 0.000595 ***
X           4.18e-08 ***
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
  0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.835 on 8 degrees of freedom
Multiple R-squared:  0.9803,    Adjusted R-squared:  0.9778
F-statistic: 397.3 on 1 and 8 DF,  p-value: 4.183e-08
```

```
> # (c) Ho : B1 = 0에 대해 H1 : B1 != 0 유의수준 5%에서 검정하시오
> summary(model)  # 회귀모형의 요약 정보에서 p-value 확인

Call:
lm(formula = Y ~ X)

Residuals:
   Min     1Q Median     3Q    Max
-4.899 -1.374  0.490  1.322  4.116

Coefficients:
            Estimate Std. Error t value
(Intercept) -20.2339     3.6995  -5.469
X            12.4424     0.6242  19.933
            Pr(>|t|)
(Intercept) 0.000595 ***
X           4.18e-08 ***
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
  0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.835 on 8 degrees of freedom
Multiple R-squared:  0.9803,    Adjusted R-squared:  0.9778
F-statistic: 397.3 on 1 and 8 DF,  p-value: 4.183e-08

>
> # (d) 결정계수를 구하고 해석하시오
> R_squared <- summary(model)$r.squared
> R_squared
[1] 0.9802624
```

```
> # (e) (b)번의 모형을 이용하여 각 X에 대응하는 y의 hat을 구하시오
> y_hat <- fitted(model)
> y_hat
       1        2        3        4
23.31455 29.53576 35.75697 41.97818
       5        6        7        8
48.19939 54.42061 60.64182 66.86303
       9       10
73.08424 79.30545
>
> # (f) 각 Y에 대응하는 잔차를 구하시오
> residuals <- residuals(model)
> residuals
         1          2          3
 1.0854545  2.5642424  1.3430303
         4          5          6
-1.5781818 -4.8993939 -3.0206061
         7          8          9
 1.2581818 -0.7630303  4.1157576
        10
-0.1054545
>
> # (g) (y의 hat, e) 잔차그림을 그리고 독립성과 등분산성에 대해 설명하시오
> plot(y_hat, residuals, main="Residual Plot", xlab="Fitted Values (y_hat)", ylab="Residual
s", pch=19)
> abline(h=0)  # 잔차의 기준선
```




```
> # (h) 잔차가 정규분포를 따른다고 할 수 있는지 Q-Q 그림을 그리고 설명하시오
> qqnorm(residuals, main="Normal Q-Q Plot of Residuals")
> qqline(residuals)
```




```
> # (i) x = 5.3에서의 y의 hat을 구하시오
> new_data <- data.frame(X=5.3)
> predicted_Y <- predict(model, newdata=new_data)
> predicted_Y
       1
45.71091
```