Bonus Activity

Code

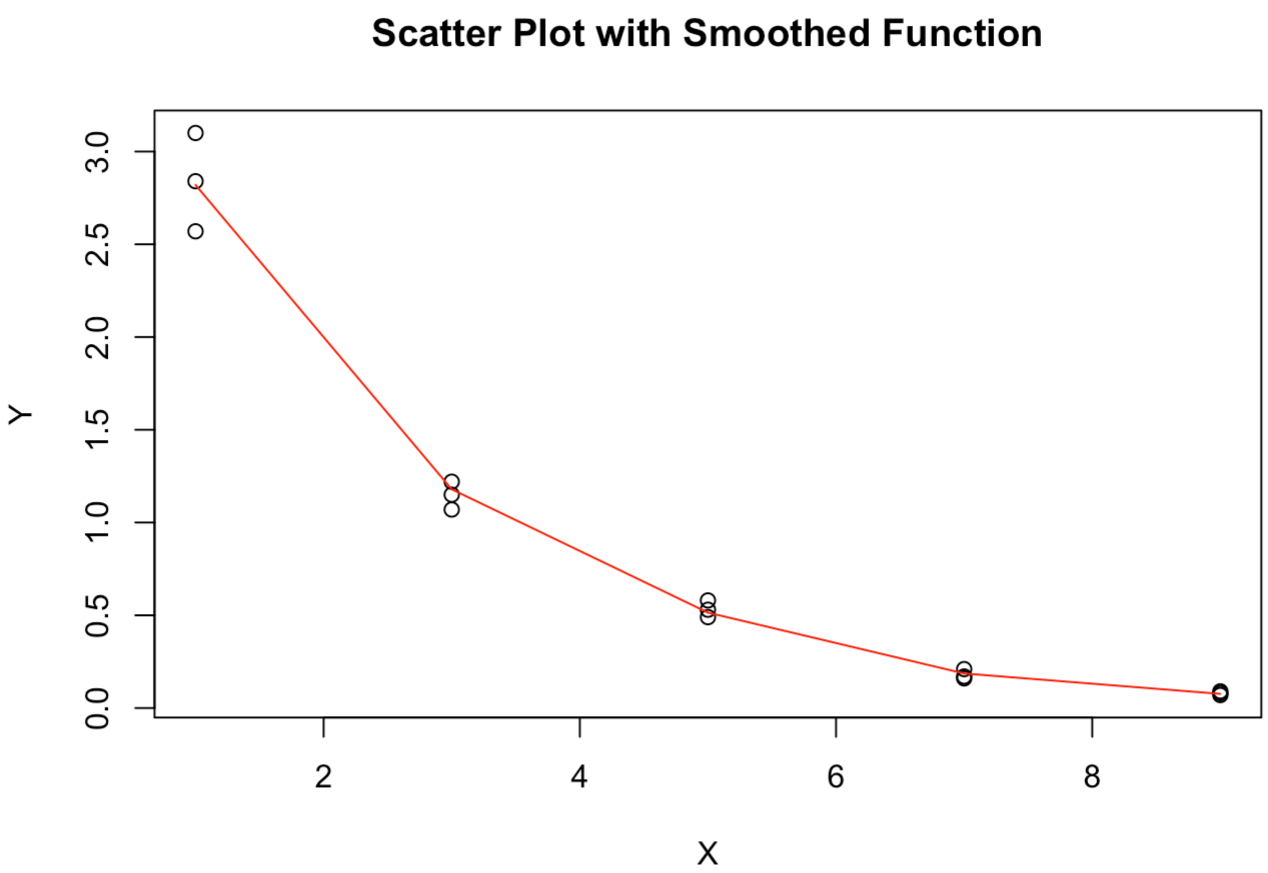
Hide

```
library(ggplot2)
data <- read.table('CH03PR15.txt')
colnames(data) <- c('Y', 'X')
```

(A)Plot the data. Include a smoothed function on the plot. What are the noteworthy features of this plot?

Hide

```
plot(data$X, data$Y, xlab = "X", ylab = "Y", main = "Scatter Plot with Smoothed Funct
ion")
lines(smooth.spline(data$X, data$Y), col = "red")
```



If X increased, Y would decrease.

(B)Fit a straight line model (simple linear regression) to the data. Give the results about the t-test to determine whether or not a linear relationship between the response and independent variables.

Hide

```
model <- lm(Y ~ X, data = data)
summary(model)
```

```
Call:
lm(formula = Y ~ X, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5333 -0.4043 -0.1373  0.4157  0.8487

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.5753     0.2487  10.354 1.20e-07 ***
X            -0.3240     0.0433  -7.483 4.61e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4743 on 13 degrees of freedom
Multiple R-squared:  0.8116,    Adjusted R-squared:  0.7971
F-statistic: 55.99 on 1 and 13 DF,  p-value: 4.611e-06
```

Since the p-value of the coefficient of X is very small, there are significant linear relationship between X and Y.

(C)Perform an ANOVA F-test of whether or not a linear relationship between the response and independent variables.

Hide

```
anova(model)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
X	1	12.597120	12.5971200	55.99384	4.611199e-06
Residuals	13	2.924653	0.2249733	NA	NA

2 rows

The F-statistic is 55.994, and the p-value is very small. Therefore, there are significant linear relationship between X and Y.

(D)Compute the sample Pearson correlation coefficient between the response and independent variables. Read the section “Inference on Correlation Coefficients” on p.83-87 of the textbook. Test whether or not the population correlation coefficient is significant different from zero. State the hypotheses, test statistics, decision rule, and conclusions.

Hide

```
#sample_Pearson_correlation
cor(data$X, data$Y)
```

```
[1] -0.9008759
```

Hide

```
cor.test(data$X, data$Y)
```

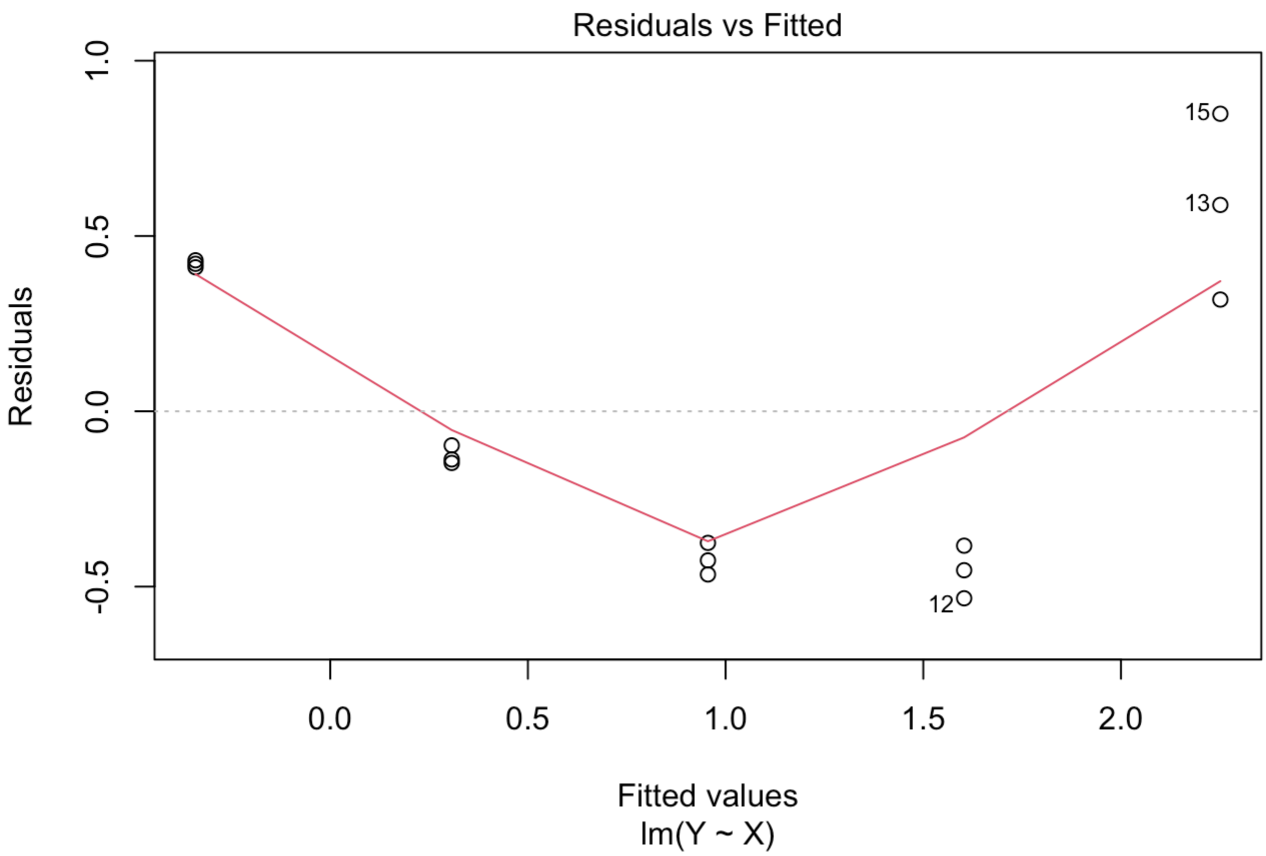
```
    Pearson's product-moment correlation

data:  data$X and data$Y
t = -7.4829, df = 13, p-value = 4.611e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9669196 -0.7216386
sample estimates:
       cor
-0.9008759
```

The coefficient of correlation is -0.9008759, stating that there is strong “negative” relationship between X and Y. And the p-value is small enough, we can conclude that there’re relationship between X and Y.

(E)Refer to parts (b), (c), and (d). Compare the results. Do they yield the same conclusions? **Ans:The results from (b), (c), (d) are all the same. Therefore, there are relationships between X and Y.** (F)Graph the residuals versus fits and prepare a normal probability plot of the residuals.

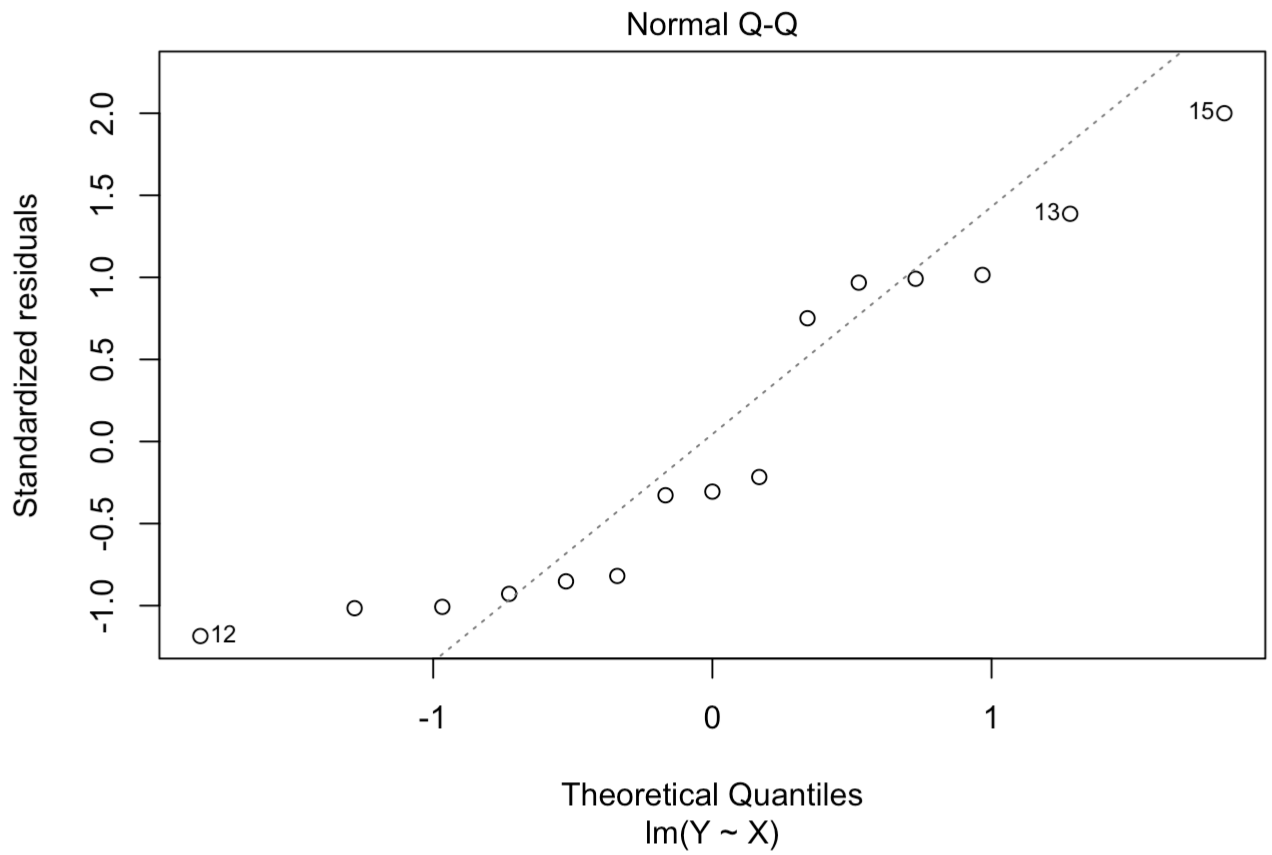
Hide

```
#residuals versus fits and prepare a normal probability plot
plot(model, which = 1)
```



Hide

```
plot(model, which = 2)
```



(G)Fit a simple linear regression model in which the response variable is the nature log of the response variable.

Hide

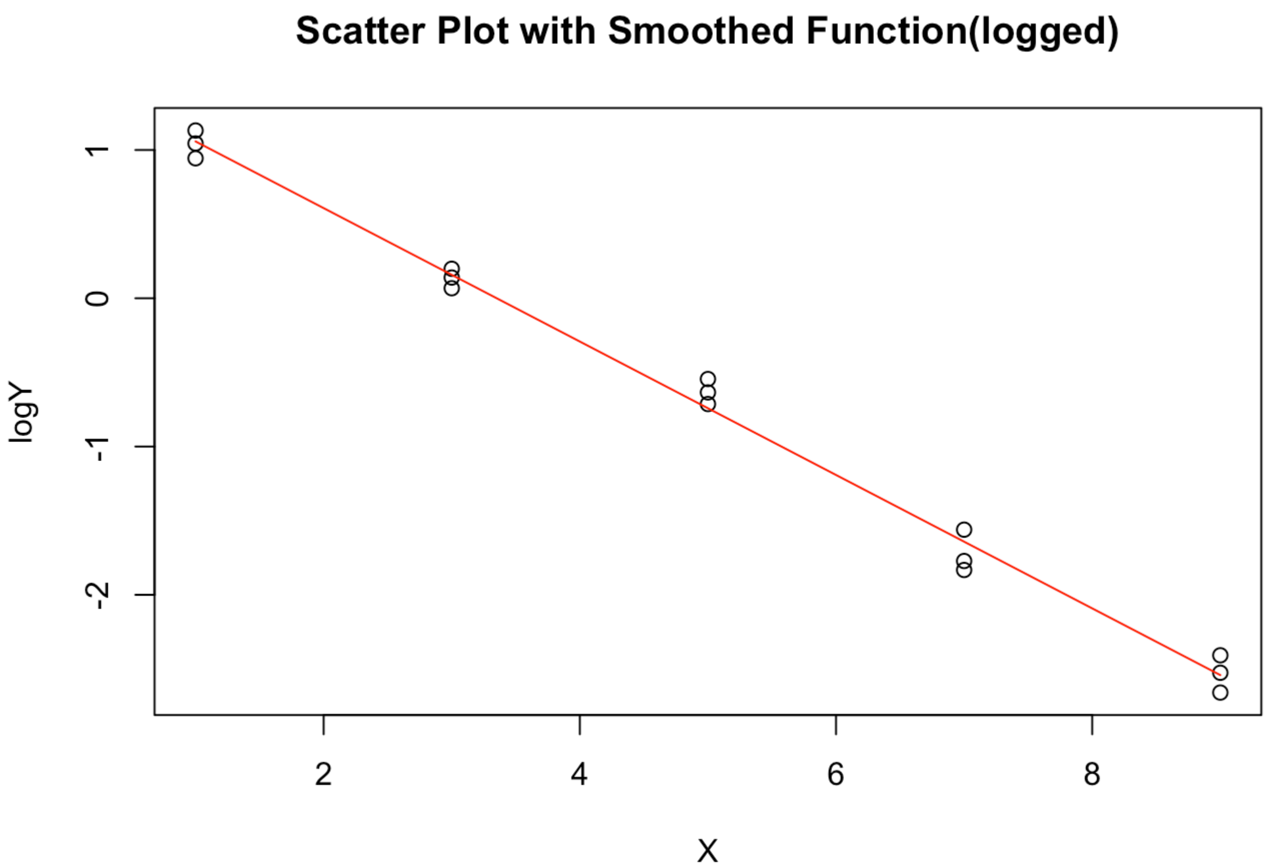
```
#logarithm
data$logY <- log(data$Y)
logged <- lm(logY ~ X, data = data)
summary(logged)
```

```
Call:
lm(formula = logY ~ X, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.19102 -0.10228  0.01569  0.07716  0.19699 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.50792    0.06028   25.01 2.22e-12 ***
X           -0.44993    0.01049  -42.88 2.19e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.115 on 13 degrees of freedom
Multiple R-squared:  0.993, Adjusted R-squared:  0.9924 
F-statistic:  1838 on 1 and 13 DF,  p-value: 2.188e-15
```

(H)Refer to part (g). Repeat the part (a) and part (f) using the transformed data. What is the result then?

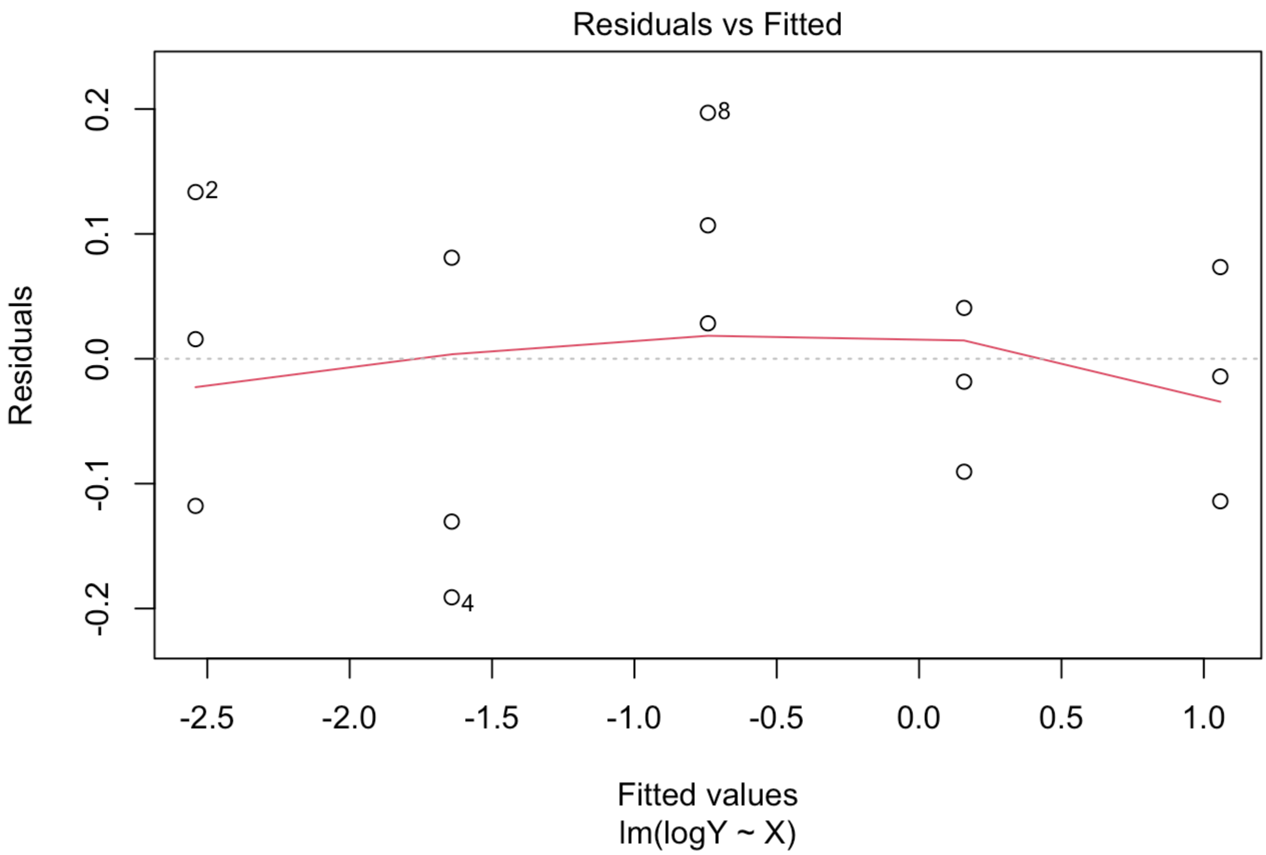
Hide

```
plot(data$X, data$logY, xlab = "X", ylab = "logY", main = "Scatter Plot with Smoothed
Function(logged)")
lines(smooth.spline(data$X, data$logY), col = "red")
```



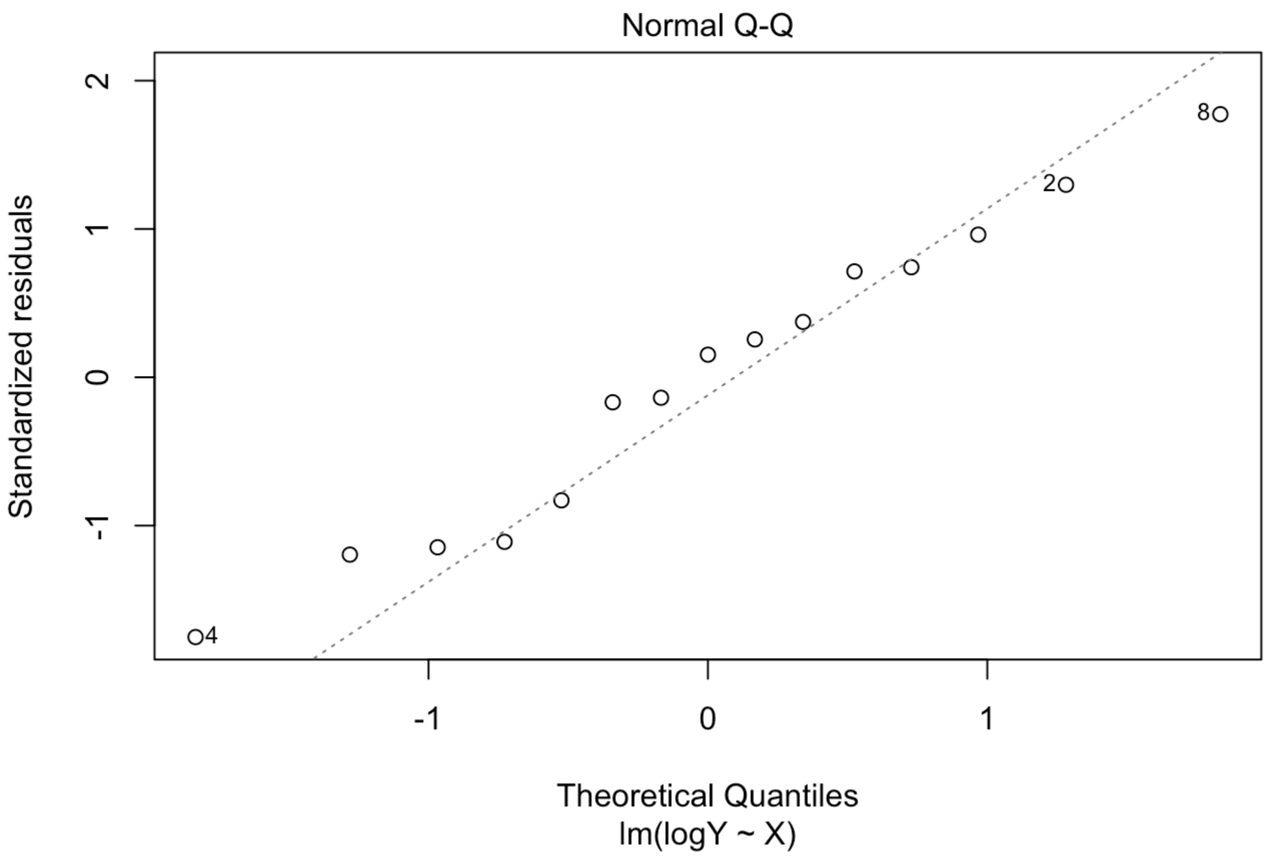
Hide

```
plot(logged, which = 1)
```



Hide

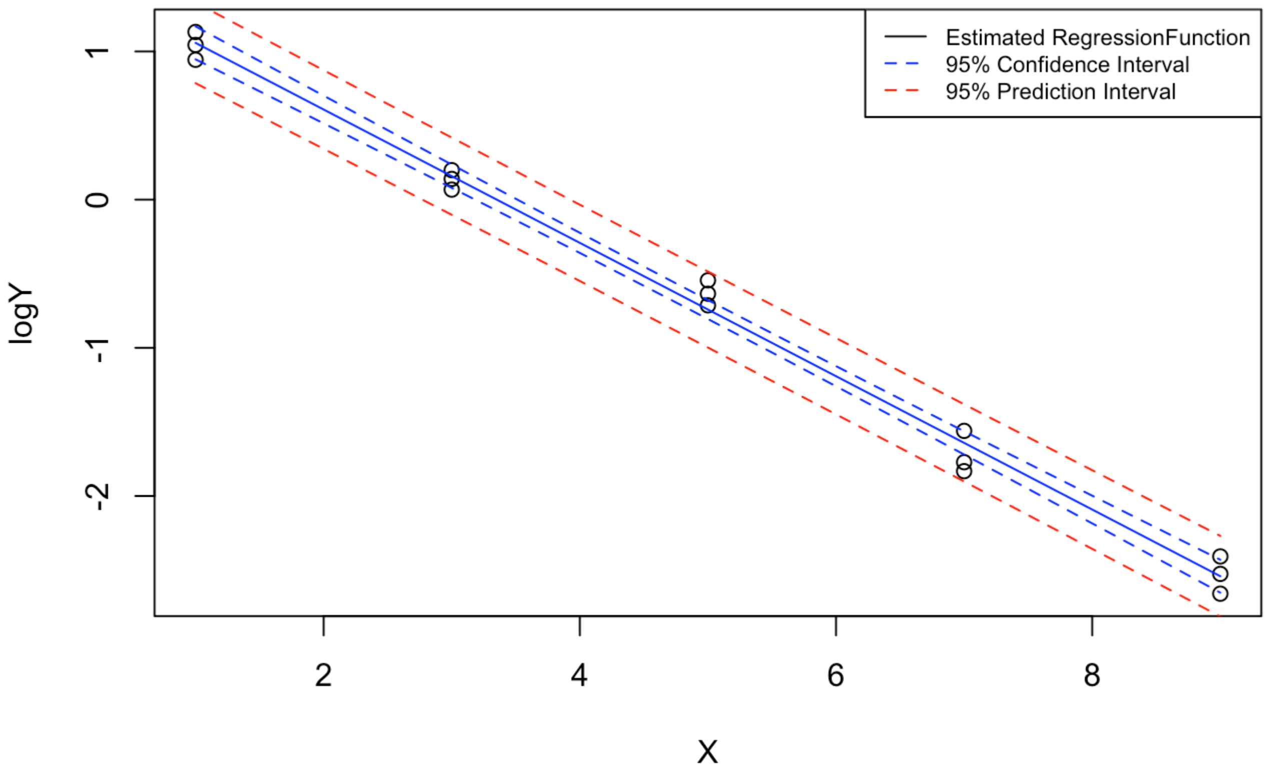
```
plot(logged, which = 2)
```



(I)Refer to part (g). Plot the estimated regression function, the 95% confidence intervals for the mean response, and 95% prediction intervals for a range of X values that span the data.

Hide

```
newdata <- data.frame(X = seq(min(data$X), max(data$X), length.out = 100))
conf_interval <- predict(logged, newdata, interval = "confidence")
pred_interval <- predict(logged, newdata, interval = "prediction")
plot(data$X, data$logY, xlab = "X", ylab = "logY")
lines(newdata$X, conf_interval[,1], col = "blue")
lines(newdata$X, conf_interval[,2], col = "blue", lty = 2)
lines(newdata$X, conf_interval[,3], col = "blue", lty = 2)
lines(newdata$X, pred_interval[,2], col = "red", lty = 2)
lines(newdata$X, pred_interval[,3], col = "red", lty = 2)
legend("topright", legend = c("Estimated RegressionFunction", "95% Confidence Interva
l","95% Prediction Interval"),
col = c("black", "blue","red"), lty = c(1,2,2), cex = 0.7)
```



(J)Refer to part (g). Express the estimated regression function in the original units.

Hide

```
#Express the estimated regression function in the original units
exp(coef(logged)[1] + coef(logged)[2] * newdata$X)
```

```
  [1] 2.88057675 2.77772679 2.67854905 2.58291242 2.49069045 2.40176124
  [7] 2.31600722 2.23331501 2.15357530 2.07668266 2.00253545 1.93103564
 [13] 1.86208870 1.79560349 1.73149211 1.66966981 1.61005484 1.55256841
 [19] 1.49713451 1.44367986 1.39213379 1.34242815 1.29449723 1.24827767
 [25] 1.20370836 1.16073039 1.11928692 1.07932319 1.04078634 1.00362543
 [31] 0.96779135 0.93323670 0.89991582 0.86778465 0.83680071 0.80692304
 [37] 0.77811214 0.75032993 0.72353966 0.69770594 0.67279460 0.64877271
 [43] 0.62560851 0.60327138 0.58173179 0.56096127 0.54093235 0.52161855
 [49] 0.50299435 0.48503511 0.46771711 0.45101744 0.43491402 0.41938557
 [55] 0.40441156 0.38997219 0.37604837 0.36262170 0.34967442 0.33718942
 [61] 0.32515019 0.31354082 0.30234596 0.29155081 0.28114109 0.27110305
 [67] 0.26142341 0.25208938 0.24308862 0.23440923 0.22603973 0.21796906
 [73] 0.21018655 0.20268192 0.19544523 0.18846693 0.18173779 0.17524890
 [79] 0.16899171 0.16295792 0.15713956 0.15152895 0.14611866 0.14090155
 [85] 0.13587071 0.13101949 0.12634149 0.12183051 0.11748060 0.11328600
 [91] 0.10924116 0.10534074 0.10157959 0.09795273 0.09445536 0.09108286
 [97] 0.08783078 0.08469482 0.08167082 0.07875479
```