STA 506 2.0 Linear Regression Analysis

Mid Evaluation Assignment

1 November 2020

Total marks: 40

Question

Any sound that is loud enough and lasts long enough can damage hearing and lead to hearing loss. A study was conducted to assess the effects of noise on hearing loss. The independent variable (X) was maximal noise level. Each individual was exposed regularly (5 hours a day, for a period of 5 years) to an environment with a specific noise level. The noise level is measured in units called decibels. The dependent variable was hearing loss. Hearing loss was measured as a percentage, compared with normal hearing. The data are given below.

```
hearing <- data.frame(NoiseLevel = c(62, 65, 70, 72, 81, 86, 90, 111, 123,
                                     128, 142, 106, 100, 100, 112, 142),
Loss = c(12, 15, 10, 18, 16, 23, 20, 34, 38, 48, 50, 32, 35, 36, 30, 58))
hearing
```

```
   NoiseLevel Loss
1          62   12
2          65   15
3          70   10
4          72   18
5          81   16
6          86   23
7          90   20
8         111   34
9         123   38
10        128   48
11        142   50
12        106   32
13        100   35
14        100   36
15        112   30
16        142   58
```

Q1 [4 marks]

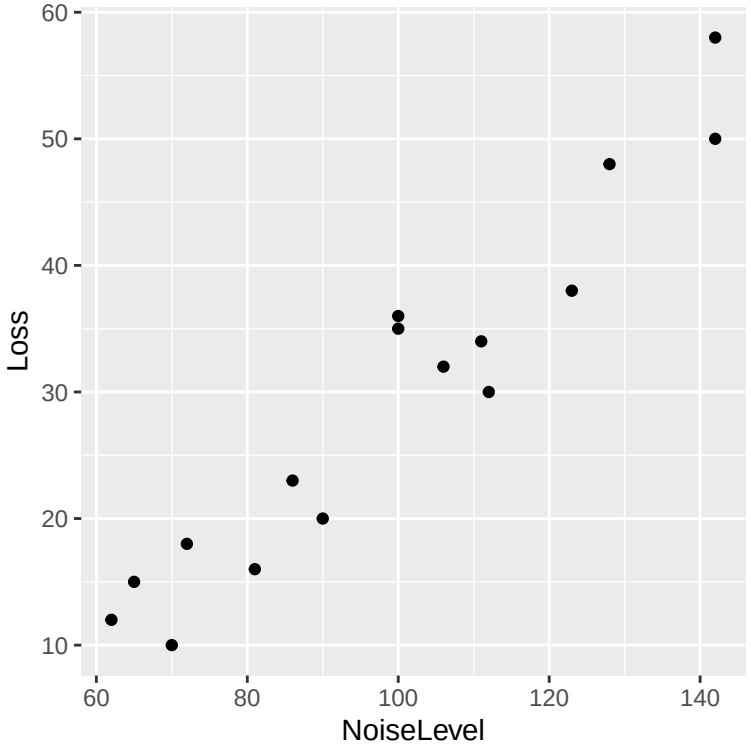

Figure 1: Relationship between noise level and hearing loss

The Pearson's correlation coefficient between X and Y is

```
[1] 0.9584264
```

Comment on the results given above.

Q2 [5 marks]

Write the regression model that you would be fitted for the above variables. Define every term and state any assumptions regarding the model.

Q3

A researcher fitted a simple linear regression model. The output is shown below

```
Call:
lm(formula = Loss ~ NoiseLevel, data = hearing)

Coefficients:
(Intercept)   NoiseLevel
   -22.8699       0.5289
```

```
Call:
lm(formula = Loss ~ NoiseLevel, data = hearing)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3646 -4.0149 -0.4025  3.2532  5.9820

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -22.86986    4.31407  -5.301 0.000112 ***
NoiseLevel    0.52888    0.04208  12.568 5.16e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.239 on 14 degrees of freedom
Multiple R-squared:  0.9186,    Adjusted R-squared:  0.9128
F-statistic:   158 on 1 and 14 DF,  p-value: 5.155e-09
```

The results of residual analysis

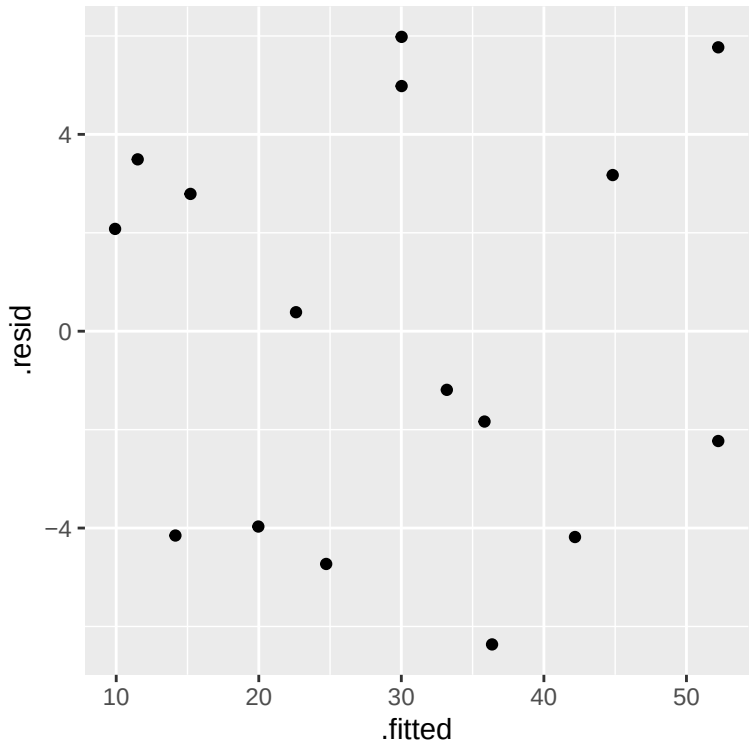

Figure 2: Residuals vs Fitted Values

```
shapiro.test(hearing_fit_resid$.resid)
```

```
    Shapiro-Wilk normality test

data:  hearing_fit_resid$.resid
W = 0.92954, p-value = 0.2397
```

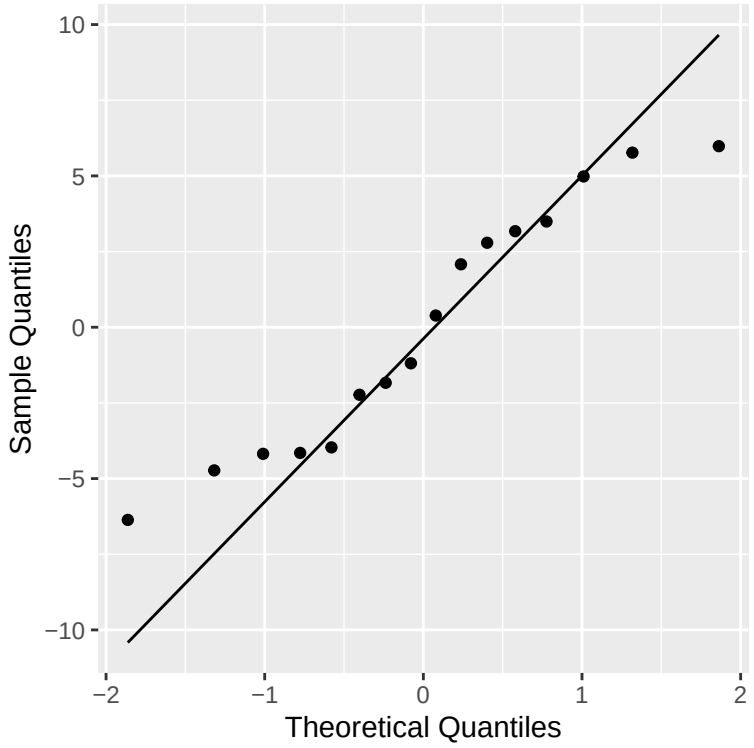

Figure 3: Normality test of residuals

Q3 - i [3 marks]

Write the fitted regression model.

Q3 - ii [5 marks]

Check the validity of the assumptions. You should clearly specify the reasons referring to suitable outputs.

Q3 - iii [5 marks]

Test the significance of the model. You should clearly write the hypothesis, decision and conclusions.

Q3 - iv [5 marks]

What proportion of the variation in the response is explained by the model fitted?

Q3 - v [4 marks]

Interpret the value $\hat{\beta}_1 = 0.5289$.

Q3 - vi [3 marks]

If an individual's environment noise level is 100, give a point estimate for its predicted hearing loss value.

Q4 [6 marks]

Another researcher fitted a regression model without an intercept. The results are given below. Which model would you conclude is superior (model with intercept or model without intercept)? Give reasons for your answer.

```
Call:
lm(formula = Loss ~ NoiseLevel - 1, data = hearing)

Coefficients:
NoiseLevel
    0.3126
```

```
Call:
lm(formula = Loss ~ NoiseLevel - 1, data = hearing)

Residuals:
    Min      1Q  Median      3Q     Max
-11.884  -5.836  -2.512   3.987  13.607

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
NoiseLevel  0.31263    0.01732   18.05 1.38e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.101 on 15 degrees of freedom
Multiple R-squared:  0.956, Adjusted R-squared:  0.9531
F-statistic: 325.9 on 1 and 15 DF,  p-value: 1.385e-11
```

The results of residual analysis

```
shapiro.test(hearing_fit_resid_nointercept$.resid)
```

```
    Shapiro-Wilk normality test

data:  hearing_fit_resid_nointercept$.resid
W = 0.96337, p-value = 0.7231
```

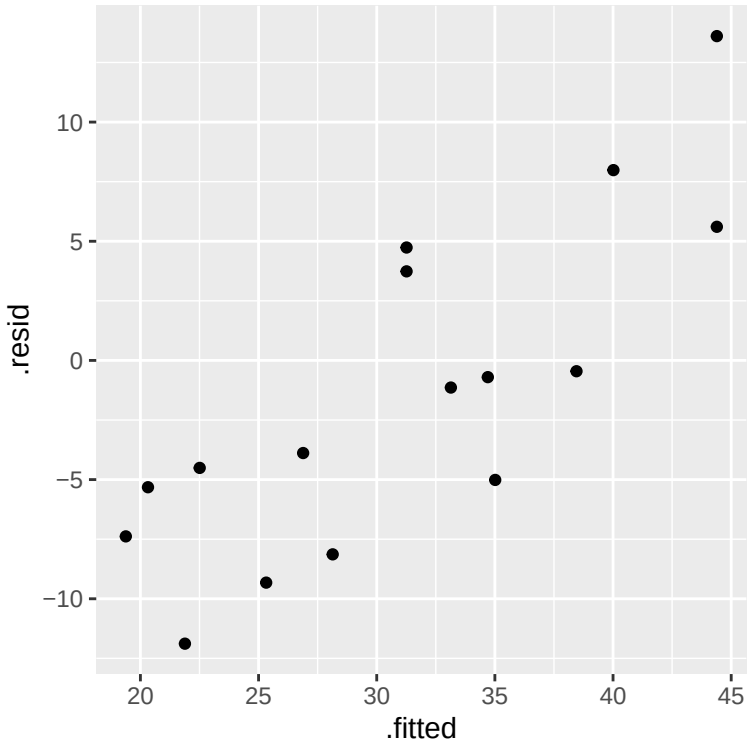

Figure 4: Residuals vs Fitted Values (Model without intercept)

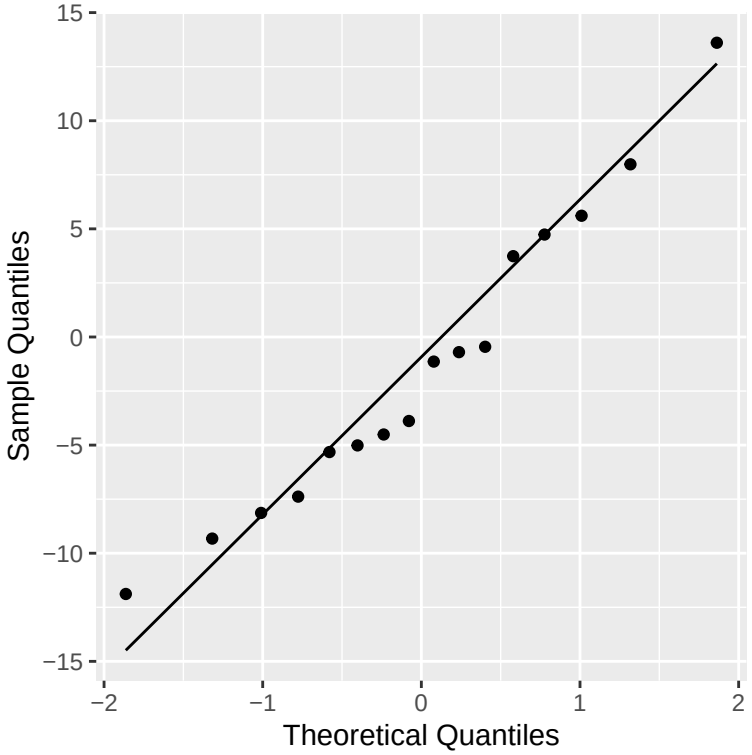

Figure 5: Normality test of residuals (Model without intercept)