Untitled

2024-12-01

1.) (IN R) Consider the text file called muscle which consists of paired data of individual's muscle mass and age. We want to examine this relationship.

```
setwd("~/Desktop/Personal_save/Stat_405_Module_14/Lab_14.3")
muscle <- read.table(file="muscle.txt",header=TRUE)
```

  a. Begin by looking at a scatterplot and finding the correlation. Does the association look linear? How strong is it?

```
muscle
```

```
##      mass age
## 1      82  71
## 2      91  64
## 3     100  43
## 4      68  67
## 5      87  56
## 6      73  73
## 7      78  68
## 8      80  56
## 9      65  76
## 10     84  65
## 11    116  45
## 12     76  58
## 13     97  45
## 14    100  53
## 15    105  49
## 16     77  78
```

```
plot(x=muscle$age,y=muscle$mass)
```
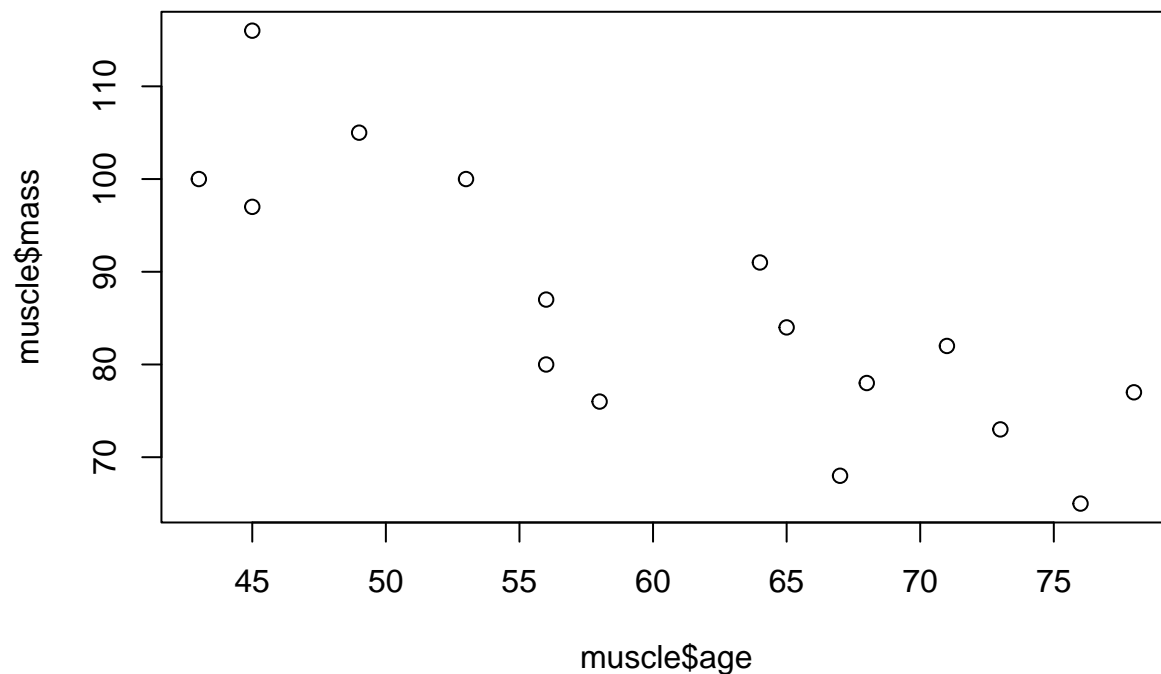
```r
cor(x=muscle$age,y=muscle$mass)
```

```
## [1] -0.8238943
```

The association looks linear, and very strong.

    b. Produce the necessary diagnostic plots to prepare to fit a regression model.

fitted values means predicted response values.

- Residuals vs. Predicted Response
- Residuals vs. Predictor
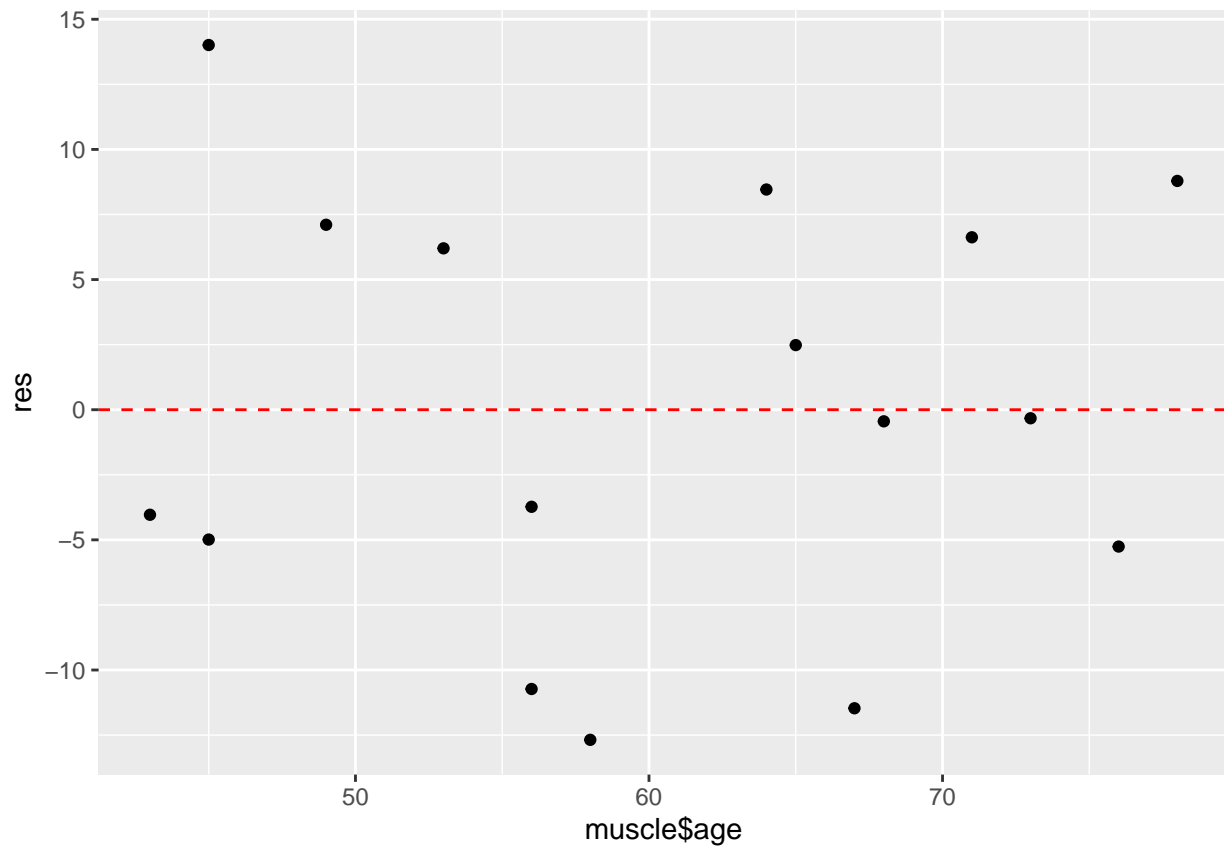- Studentized Residuals vs. Predictor
- QQ plot of Residuals

```r
library(ggplot2)
model <- lm(muscle$mass ~ muscle$age)

res <- resid(model)
fit <- fitted(model)

#Residuals vs. Predicted Response
ggplot(mapping = aes(x = fit, y = res)) +
    geom_point() +
    geom_hline(yintercept = 0, colour = "red", linetype = "dashed")
```
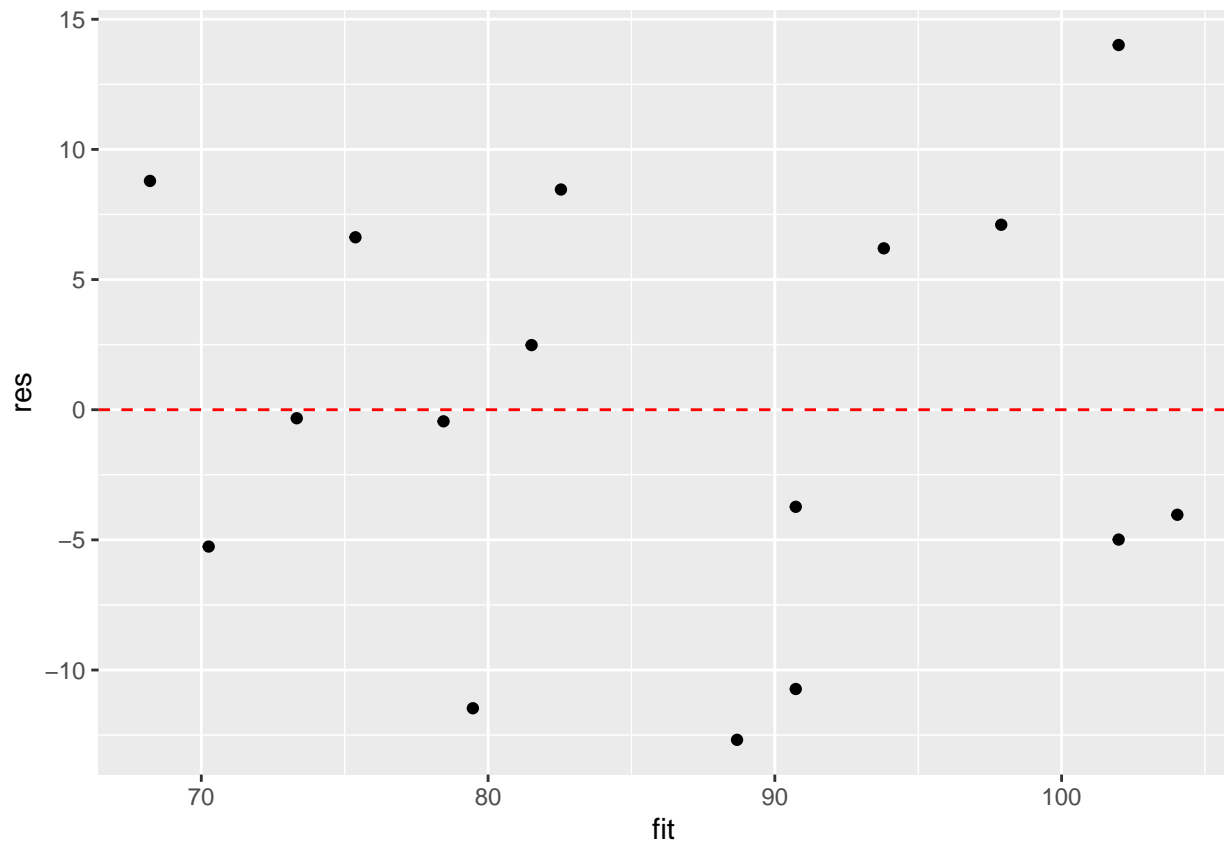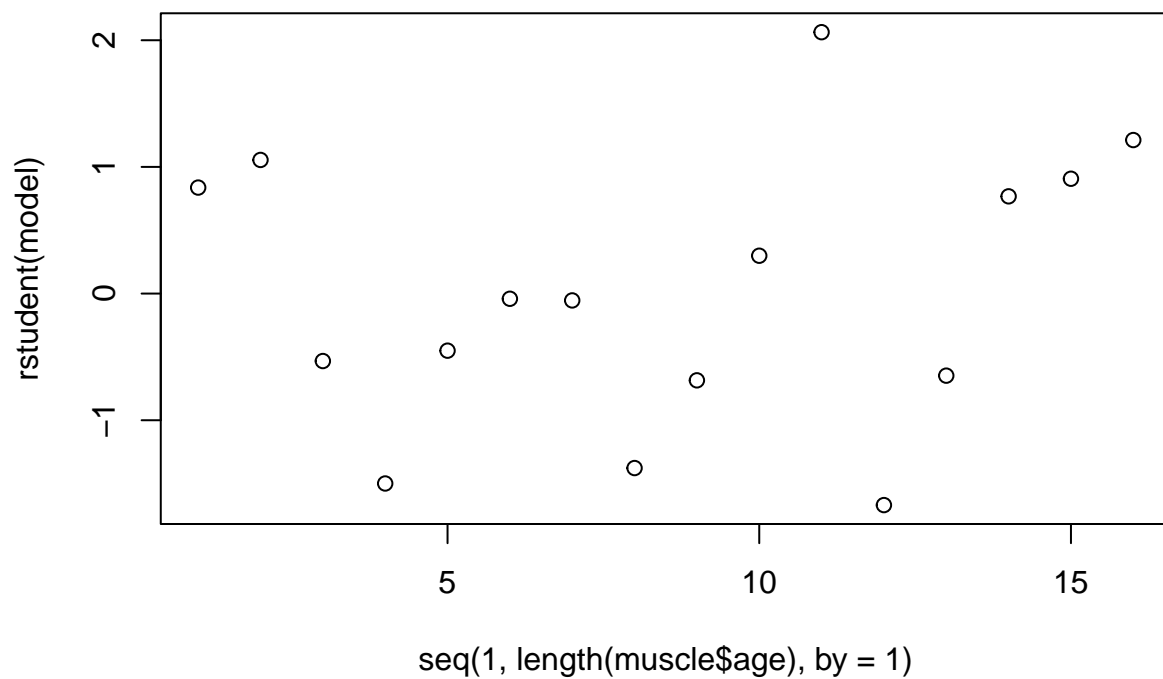
```
#Residuals vs. Predictor
ggplot(mapping = aes(x = muscle$age, y = res)) +
    geom_point() +
    geom_hline(yintercept = 0, colour = "red", linetype = "dashed")
```
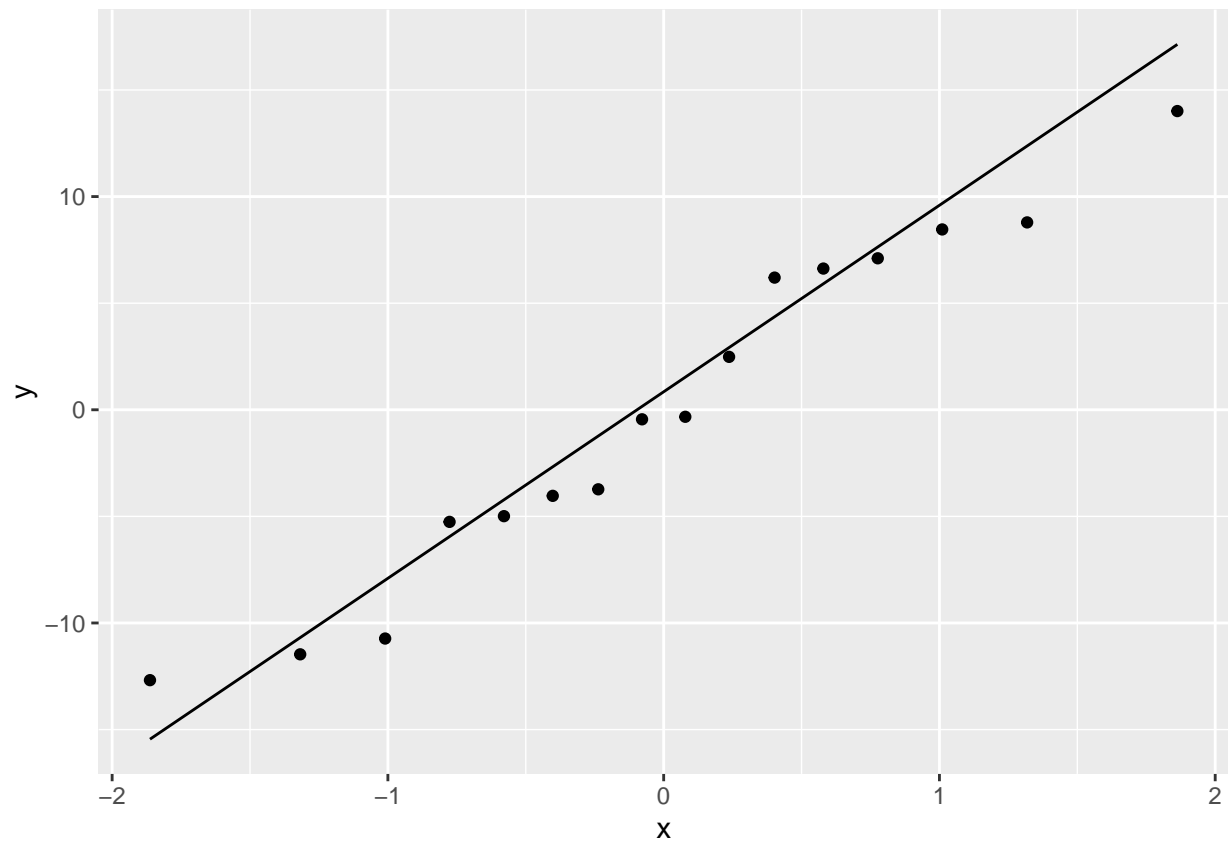
```
#Studentized Residuals vs. Predictor observation number
plot(rstudent(model) ~ seq(1,length(muscle$age),by=1))
```
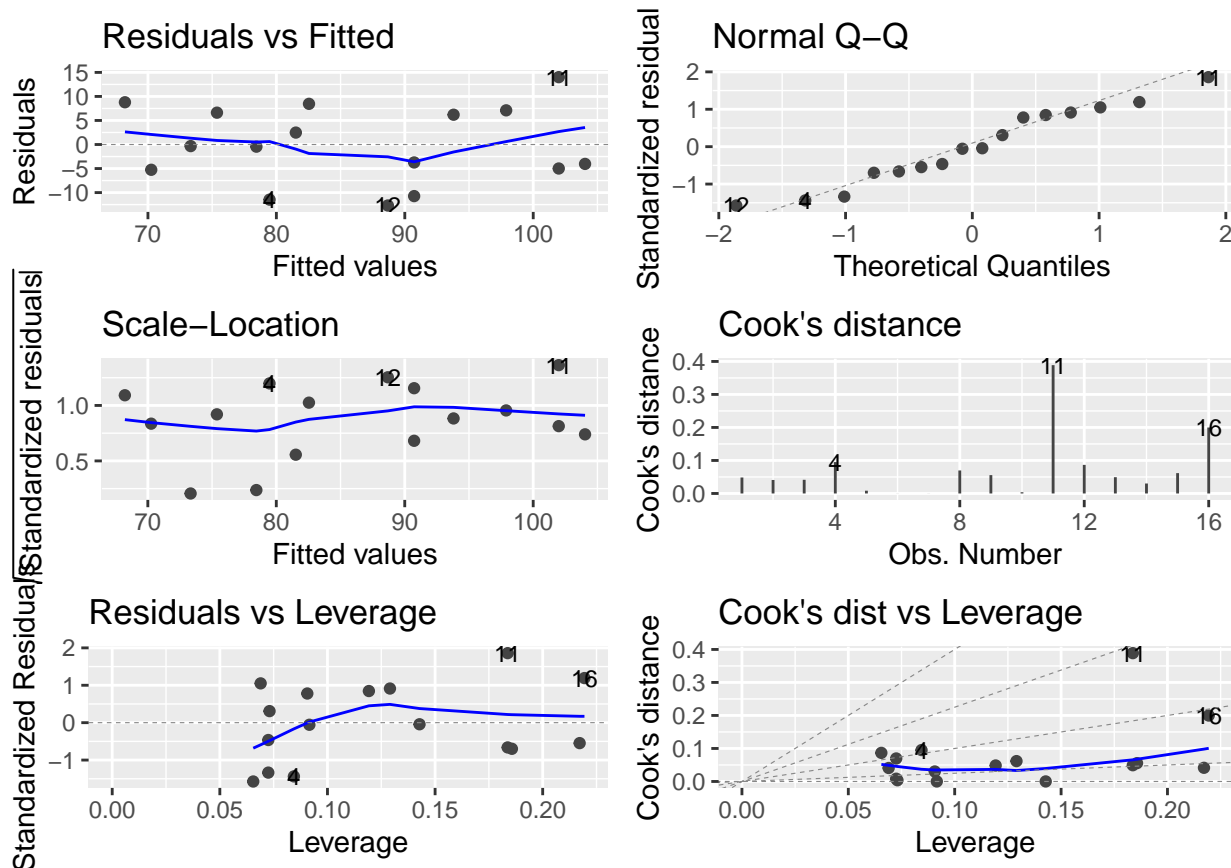


```
#QQ plot of Residuals
ggplot(mapping = aes(sample = res)) +
    stat_qq() +
```

4

```r
stat_qq_line()
```



```r
library(ggfortify)
autoplot(model, which = 1:6, ncol = 2, label.size = 3)
```

c. Using this information, would you recommend going ahead and fitting a simple linear regression model to this data? (Do not fit the model yet, just answer yes or no and provide evidence)

Yes I would recommend fitting a linear model, everything appears random with no significant patters. Observation 11 is particularly influential in the leverage plot and has a relatively high studentised residual but not enough to warrant removal.

c. Write a code to fit a simple linear regression model.

```
model <- lm(mass ~ age, data = muscle)
summary(model)
```

```
##
## Call:
## lm(formula = mass ~ age, data = muscle)
##
## Residuals:
##      Min      1Q   Median       3Q      Max
## -12.6825  -5.0563  -0.3876   6.7444  14.0108
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 148.0507    11.5629  12.804 4.05e-09 ***
## age          -1.0236     0.1882  -5.439 8.72e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.344 on 14 degrees of freedom
```

```
## Multiple R-squared:  0.6788, Adjusted R-squared:  0.6559
## F-statistic: 29.59 on 1 and 14 DF,  p-value: 8.721e-05
```

d. What does the slope mean?

The meaning of the slope is that as age increases by one year the model predicts that muscle mass will decrease by about one unit of measurement.

e. Predict the muscle mass for a 66-year-old person. (Hint: use the predict( ) function. Further detail of the predict function for linear regression model can be found in the R documentation. help(predict.lm))

```
predicted_mass <- predict(model, newdata = data.frame(age = 66))
predicted_mass
```

```
##        1
## 80.49378
```

f. Calculate a 98% CI for the muscle mass for a 40-year-old. (Hint: use the predict( ) function.)

```
prediction_40 <- predict(model, newdata = data.frame(age = 40),
                         interval = "confidence",
                         level = 0.98)

prediction_40
```

```
##       fit      lwr      upr
## 1 107.1071 95.62437 118.5898
```