Stat 415 Regression: Homework 2 (Residuals)

Frankie Tyndall

2022-07-09

## Study of Muscle Mass

A person’s muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each lO-year age group, beginning with age 40 and ending with age 79 The results follow; X is age, and Y is a measure of muscle mass.

x = age

y = measure of muscle mass

## Muscle Mass Dataset: Use and show R code to read in the Muscel Mass Data Excel File

MuscelMassData

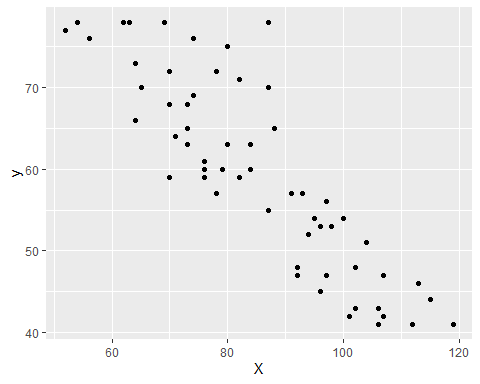
## # A tibble: 62 × 2  
## X y  
## <dbl> <dbl>  
## 1 106 43  
## 2 106 41  
## 3 97 47  
## 4 113 46  
## 5 96 45  
## 6 119 41  
## 7 92 47  
## 8 112 41  
## 9 92 48  
## 10 102 48  
## # … with 52 more rows

## Hypothesis for F-Test

H0:β1=0 # null hypothesis: linear relationship does not exist HA:β1≠0 # alternative hypothesis: linear relationship does exist

## Use and show R code to generate a scatter plot and comment on general trends that are evident involving the dependent variable and the independent variable.

qplot(x = X, y = y, data = MuscelMassData, geom = "point")

 Based on the scatter plot, the bivariate data with age and measure muscle mass has a strong negative correlation. This indicates that as age increases, the muscle mass of an individual decreases. Based on the scatterplot, it looks like one outlier is present.

## Use and show R code to produce a regression model. Interpret the slope of your model.

lm(y~X, data = MuscelMassData)

##   
## Call:  
## lm(formula = y ~ X, data = MuscelMassData)  
##   
## Coefficients:  
## (Intercept) X   
## 113.3524 -0.6275

Slope = -0.6275 y intercept = 113.3524

y(hat) = 113.3524 - 0.6275X

The slope indicates that for every increase by 1 year in age (X), the measure of muscle mass was found to have decreased by -0.6275 years on average.

## Use and show R code to produce all of the residuals

model\_muscle <- lm(y~X, data = MuscelMassData)  
residual\_model <- resid(model\_muscle)  
residual\_model

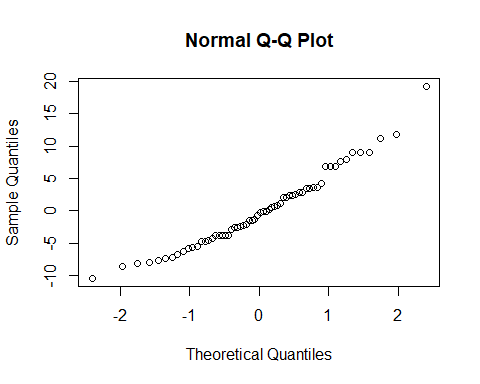
## 1 2 3 4 5 6   
## -3.8361564 -5.8361564 -5.4837652 3.5564282 -8.1112773 2.3215007   
## 7 8 9 10 11 12   
## -8.6213257 -2.0710839 -7.6213257 -1.3462048 -4.2086444 0.7913556   
## 13 14 15 16 17 18   
## -6.3462048 2.8114524 -7.9737169 -3.7588861 0.7511622 3.5162348   
## 19 20 21 22 23 24   
## -2.8964466 -7.4064949 0.2612106 1.1437468 -2.3663015 -0.1112773   
## 25 26 27 28 29 30   
## 3.3987710 -0.6414224 -10.4265917 2.9088194 -6.6615191 2.0061864   
## 31 32 33 34 35 36   
## 0.4559446 -4.5440554 -5.6615191 -0.1514708 2.3585776 -4.7990796   
## 37 38 39 40 41 42   
## -7.1916642 6.8686260 -3.7789828 6.8686260 -3.7789828 6.8686260   
## 43 44 45 46 47 48   
## -2.5440554 2.0834567 -4.6615191 11.2411139 -1.4265917 7.9458963   
## 49 50 51 52 53 54   
## -1.4667851 3.5533116 7.5935051 -2.5641521 -0.1916642 9.0834567   
## 55 56 57 58 59 60   
## 19.2411139 4.1808237 9.1035534 11.8485292 -3.7218093 -2.2117609   
## 61 62   
## 2.5734083 9.0834567

## Use and show R code to produce a qqplot for the residuals. Interpret the qqplot by commenting on normality for the residuals.

model\_muscle <- lm(y~X, data = MuscelMassData)  
residual\_model <- resid(model\_muscle)  
residual\_model

## 1 2 3 4 5 6   
## -3.8361564 -5.8361564 -5.4837652 3.5564282 -8.1112773 2.3215007   
## 7 8 9 10 11 12   
## -8.6213257 -2.0710839 -7.6213257 -1.3462048 -4.2086444 0.7913556   
## 13 14 15 16 17 18   
## -6.3462048 2.8114524 -7.9737169 -3.7588861 0.7511622 3.5162348   
## 19 20 21 22 23 24   
## -2.8964466 -7.4064949 0.2612106 1.1437468 -2.3663015 -0.1112773   
## 25 26 27 28 29 30   
## 3.3987710 -0.6414224 -10.4265917 2.9088194 -6.6615191 2.0061864   
## 31 32 33 34 35 36   
## 0.4559446 -4.5440554 -5.6615191 -0.1514708 2.3585776 -4.7990796   
## 37 38 39 40 41 42   
## -7.1916642 6.8686260 -3.7789828 6.8686260 -3.7789828 6.8686260   
## 43 44 45 46 47 48   
## -2.5440554 2.0834567 -4.6615191 11.2411139 -1.4265917 7.9458963   
## 49 50 51 52 53 54   
## -1.4667851 3.5533116 7.5935051 -2.5641521 -0.1916642 9.0834567   
## 55 56 57 58 59 60   
## 19.2411139 4.1808237 9.1035534 11.8485292 -3.7218093 -2.2117609   
## 61 62   
## 2.5734083 9.0834567

qqnorm(residual\_model)

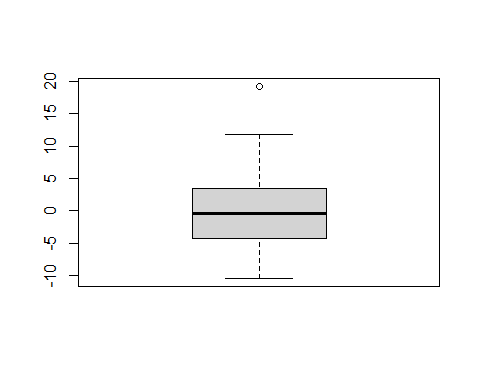
 The residuals are normally distributed as the data appears to roughly be a straight line.

## Use and show R code to produce a graph/plot that will indicate the existence of any outliers for your residuals.

model\_muscle <- lm(y~X, data = MuscelMassData)  
residual\_model <- resid(model\_muscle)  
residual\_model

## 1 2 3 4 5 6   
## -3.8361564 -5.8361564 -5.4837652 3.5564282 -8.1112773 2.3215007   
## 7 8 9 10 11 12   
## -8.6213257 -2.0710839 -7.6213257 -1.3462048 -4.2086444 0.7913556   
## 13 14 15 16 17 18   
## -6.3462048 2.8114524 -7.9737169 -3.7588861 0.7511622 3.5162348   
## 19 20 21 22 23 24   
## -2.8964466 -7.4064949 0.2612106 1.1437468 -2.3663015 -0.1112773   
## 25 26 27 28 29 30   
## 3.3987710 -0.6414224 -10.4265917 2.9088194 -6.6615191 2.0061864   
## 31 32 33 34 35 36   
## 0.4559446 -4.5440554 -5.6615191 -0.1514708 2.3585776 -4.7990796   
## 37 38 39 40 41 42   
## -7.1916642 6.8686260 -3.7789828 6.8686260 -3.7789828 6.8686260   
## 43 44 45 46 47 48   
## -2.5440554 2.0834567 -4.6615191 11.2411139 -1.4265917 7.9458963   
## 49 50 51 52 53 54   
## -1.4667851 3.5533116 7.5935051 -2.5641521 -0.1916642 9.0834567   
## 55 56 57 58 59 60   
## 19.2411139 4.1808237 9.1035534 11.8485292 -3.7218093 -2.2117609   
## 61 62   
## 2.5734083 9.0834567

boxplot(residual\_model)

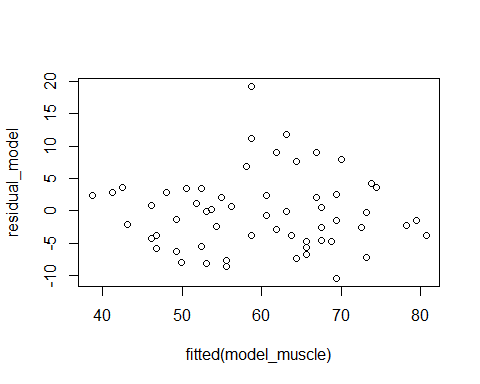
 Based on the boxplot there does suggest the existence of an outlier in the data. There is one outlier that stand out from the basic pattern of residuals.

## Now use and show R code to produce a residual plot of residuals plotted against the fitted values. Examine the residual plot and comment on the appropriateness of linearity. That is, explain why your residual plot does or does not suggest a linear relationship for the bivariate data.

model\_muscle <- lm(y~X, data = MuscelMassData)  
residual\_model <- resid(model\_muscle)  
residual\_model

## 1 2 3 4 5 6   
## -3.8361564 -5.8361564 -5.4837652 3.5564282 -8.1112773 2.3215007   
## 7 8 9 10 11 12   
## -8.6213257 -2.0710839 -7.6213257 -1.3462048 -4.2086444 0.7913556   
## 13 14 15 16 17 18   
## -6.3462048 2.8114524 -7.9737169 -3.7588861 0.7511622 3.5162348   
## 19 20 21 22 23 24   
## -2.8964466 -7.4064949 0.2612106 1.1437468 -2.3663015 -0.1112773   
## 25 26 27 28 29 30   
## 3.3987710 -0.6414224 -10.4265917 2.9088194 -6.6615191 2.0061864   
## 31 32 33 34 35 36   
## 0.4559446 -4.5440554 -5.6615191 -0.1514708 2.3585776 -4.7990796   
## 37 38 39 40 41 42   
## -7.1916642 6.8686260 -3.7789828 6.8686260 -3.7789828 6.8686260   
## 43 44 45 46 47 48   
## -2.5440554 2.0834567 -4.6615191 11.2411139 -1.4265917 7.9458963   
## 49 50 51 52 53 54   
## -1.4667851 3.5533116 7.5935051 -2.5641521 -0.1916642 9.0834567   
## 55 56 57 58 59 60   
## 19.2411139 4.1808237 9.1035534 11.8485292 -3.7218093 -2.2117609   
## 61 62   
## 2.5734083 9.0834567

plot(fitted(model\_muscle), residual\_model)

 This residual plot does suggest a linear relationship between age and muscle mass because the residuals “bounce randomly” around the 0 line with no distinct pattern shown.