Chad Huntebrinker’s Homework 7

Chad Huntebrinker

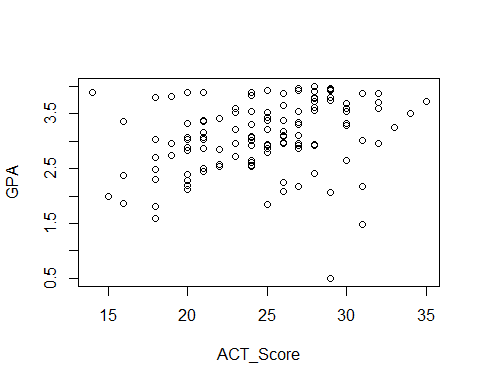
2024-10-28

Problem 2.23

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Grade\_Point\_Average\_Data.xlsx")  
  
# GPA ~ ACT  
# Fit the model and get the summary of the model  
model\_1 <- lm(GPA~ACT\_Score,data=excel\_data)  
sum\_of\_model\_1 <- summary(model\_1)  
sum\_of\_model\_1

##   
## Call:  
## lm(formula = GPA ~ ACT\_Score, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.74004 -0.33827 0.04062 0.44064 1.22737   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.11405 0.32089 6.588 1.3e-09 \*\*\*  
## ACT\_Score 0.03883 0.01277 3.040 0.00292 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6231 on 118 degrees of freedom  
## Multiple R-squared: 0.07262, Adjusted R-squared: 0.06476   
## F-statistic: 9.24 on 1 and 118 DF, p-value: 0.002917

# Plot the graph just to see  
plot(GPA~ACT\_Score,data=excel\_data)



#Problem 2.23a  
anova\_table <- anova(model\_1)  
  
#Problem 2.23b  
#MSR  
anova\_table$`Mean Sq`[1]

## [1] 3.587846

#MSE  
anova\_table$`Mean Sq`[2]

## [1] 0.3882848

#They will estimate the same quantity when the null hypothesis is true (H0: B1 = 0).  
  
#Problem 2.23c  
#H0: B1 = 0: F\* <= qf(1 - 0.01, 1, 118)  
#Ha: B1 != 0: F\* > qf(1 - 0.01, 1, 118)  
#Check to see if Ha is true:  
anova\_table$`F value`[1]

## [1] 9.240243

qf(1 - 0.01, 1, 118)

## [1] 6.854641

anova\_table$`F value`[1] > qf(1 - 0.01, 1, 118)

## [1] TRUE

#Conclusion: B1 != 0, there is a linear association  
  
#Problem2.23d  
model\_1\_reduced <- lm(GPA~1,data=excel\_data)  
anova\_table2 <- anova(model\_1\_reduced,model\_1)  
  
#Absolute magnitude of reduction  
anova\_table2$`Sum of Sq`[2]

## [1] 3.587846

#Relative reduction  
anova\_table2$`Sum of Sq`[2] / anova\_table2$RSS[1]

## [1] 0.07262044

#It's also known as R^2 or coefficient of determination  
summary(model\_1)$r.square

## [1] 0.07262044

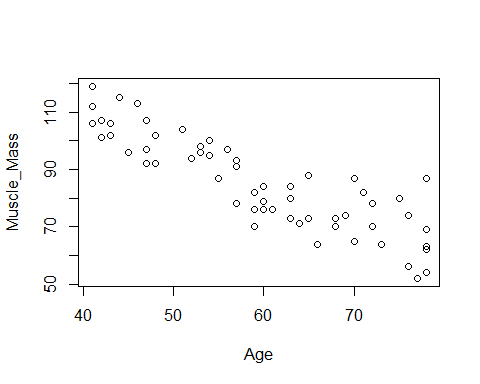
#Problem 2.23e  
sqrt(anova\_table2$`Sum of Sq`[2] / anova\_table2$RSS[1])

## [1] 0.2694818

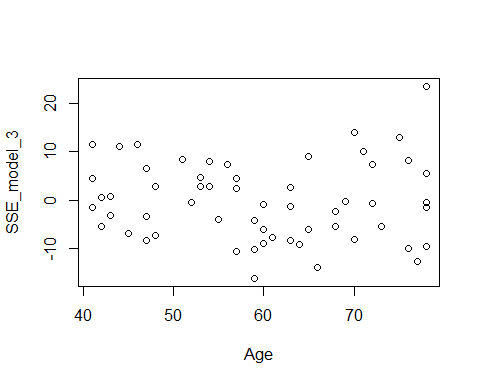
#r = +0.2694818  
  
#Problem 2.23f  
#In this case, r has a more clear-cut operational interpration. This is due to the answer  
#of r (+0.2694818) showing that there is a slight positive linear relationship between  
#the two variables. And we can see that is the case on the graph too (positive linear relationship  
#between points until we get to around 27).

Problem 2.29

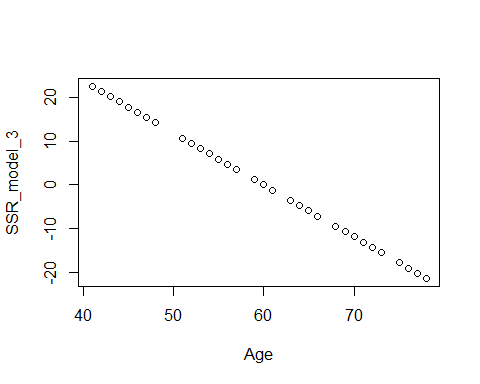
#Chad Huntebrinker  
  
#Load the library and data in to R  
library(readxl)  
excel\_data <- read\_excel("Muscle\_Mass\_Data.xlsx")  
  
#Muscle\_Mass ~ Age  
#Fit the model  
model\_3 <- lm(Muscle\_Mass~Age,data=excel\_data)  
sum\_of\_model\_3 <- summary(model\_3)  
  
#Plot to see the graph  
plot(Muscle\_Mass~Age, data = excel\_data)



#Problem 2.29a  
SSE\_model\_3 <- excel\_data$Muscle\_Mass - predict(model\_3, data.frame(Age=excel\_data$Age))  
  
  
SSR\_model\_3 <- predict(model\_3, data.frame(Age=excel\_data$Age)) - mean(excel\_data$Muscle\_Mass)  
  
plot(SSE\_model\_3 ~ Age, data = excel\_data)



plot(SSR\_model\_3 ~ Age, data = excel\_data)



#It looks like the SSR plays a larger component of SSTO.  
#That means that R^2 will be closer to 1, leading to a greater degree in linearity  
  
  
#Problem 2.29b  
anova\_table <- anova(model\_3)  
  
#Problem 2.29c  
#H0: B1 = 0: F\* <= qf(1 - 0.05, 1, 58)  
#Ha: B1 != 0: F\* > qf(1 - 0.05, 1, 58)  
#Check to see if Ha is true:  
anova\_table$`F value`[1]

## [1] 174.062

qf(1 - 0.05, 1, 58)

## [1] 4.006873

anova\_table$`F value`[1] > qf(1 - 0.05, 1, 58)

## [1] TRUE

#Conclusion: B1 != 0, there is a linear association  
  
#Problem 2.29d  
model\_3\_reduced <- lm(Muscle\_Mass~1,data=excel\_data)  
anova\_table2 <- anova(model\_3\_reduced,model\_3)  
  
1 - anova\_table2$`Sum of Sq`[2] / anova\_table2$RSS[1]

## [1] 0.2499332

#0.2499332 or 24.99%  
#It is relatively small  
  
#Problem 2.29e  
#R^2 = 0.7500668  
anova\_table2$`Sum of Sq`[2] / anova\_table2$RSS[1]

## [1] 0.7500668

#r = -0.866064  
sqrt(anova\_table2$`Sum of Sq`[2] / anova\_table2$RSS[1])

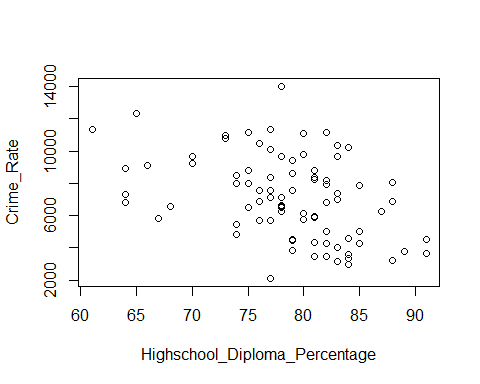
## [1] 0.866064

Problem 2.32

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Crime\_Rate\_Data.xlsx")  
  
# Fit the model and get the summary of the model  
model\_4 <- lm(Crime\_Rate~Highschool\_Diploma\_Percentage,data=excel\_data)  
sum\_of\_model\_4 <- summary(model\_4)  
sum\_of\_model\_4

##   
## Call:  
## lm(formula = Crime\_Rate ~ Highschool\_Diploma\_Percentage, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5278.3 -1757.5 -210.5 1575.3 6803.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20517.60 3277.64 6.260 1.67e-08 \*\*\*  
## Highschool\_Diploma\_Percentage -170.58 41.57 -4.103 9.57e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2356 on 82 degrees of freedom  
## Multiple R-squared: 0.1703, Adjusted R-squared: 0.1602   
## F-statistic: 16.83 on 1 and 82 DF, p-value: 9.571e-05

# Plot the graph just to see  
plot(Crime\_Rate~Highschool\_Diploma\_Percentage,data=excel\_data)



#Problem 2.32a  
model\_4\_reduced <- lm(Crime\_Rate~1,data=excel\_data)  
summary(model\_4)

##   
## Call:  
## lm(formula = Crime\_Rate ~ Highschool\_Diploma\_Percentage, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5278.3 -1757.5 -210.5 1575.3 6803.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20517.60 3277.64 6.260 1.67e-08 \*\*\*  
## Highschool\_Diploma\_Percentage -170.58 41.57 -4.103 9.57e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2356 on 82 degrees of freedom  
## Multiple R-squared: 0.1703, Adjusted R-squared: 0.1602   
## F-statistic: 16.83 on 1 and 82 DF, p-value: 9.571e-05

summary(model\_4\_reduced)

##   
## Call:  
## lm(formula = Crime\_Rate ~ 1, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5006.2 -2090.7 -180.7 1728.3 6904.8   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7111.2 280.5 25.35 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2571 on 83 degrees of freedom

#Problem 2.32b  
#1)  
SSE\_F <- deviance(model\_4)  
  
#2)  
SSE\_R <- deviance(model\_4\_reduced)  
  
#3)  
dfF = 82  
  
#4)  
dfR = 83  
  
#5)  
F\_stat <- ((SSE\_R - SSE\_F) / (dfR - dfF)) / (SSE\_F / dfF)  
  
#6)  
#H0: B1 = 0: F\* <= qf(1 - 0.01, dfR - dfF, dfF)  
#Ha: B1 != 0: F\* > qf(1 - 0.01, dfR - dfF, dfF)  
#Check to see if Ha is true:  
F\_stat

## [1] 16.83376

qf(1 - 0.01, dfR - dfF, dfF)

## [1] 6.95442

F\_stat > qf(1 - 0.01, dfR - dfF, dfF)

## [1] TRUE

#Conclusion: B1 != 0, there is a linear association  
  
#Problem 2.32c  
#t-test for with alpha = 0.01 that B1 != 0  
abs(sum\_of\_model\_4$coefficients[2,3]) > qt(1 - 0.01/2, 82)

## [1] TRUE

abs(sum\_of\_model\_4$coefficients[2,3])

## [1] 4.102897

sqrt(F\_stat)

## [1] 4.102897

#Yes, they are

Problem 2.64

#Chad Huntebrinker  
  
#Load the library and data in to R  
library(readxl)  
excel\_data <- read\_excel("SENIC\_Data.xlsx")  
  
#Length\_of\_Stay ~ Infection\_Risk  
#Fit the model  
model\_5 <- lm(Length\_of\_Stay~Infection\_Risk,data=excel\_data)  
sum\_of\_model\_5 <- summary(model\_5)  
  
#Length\_of\_Stay ~ Available\_facilities\_and\_services  
model\_6 <- lm(Length\_of\_Stay~Available\_facilities\_and\_services,data=excel\_data)  
sum\_of\_model\_6 <- summary(model\_6)  
  
#Length\_of\_Stay ~ Routine\_Chest\_X-ray  
model\_7 <- lm(Length\_of\_Stay~`Routine\_Chest\_X-ray`,data=excel\_data)  
sum\_of\_model\_7 <- summary(model\_7)  
  
sum\_of\_model\_5$r.square

## [1] 0.2845623

sum\_of\_model\_6$r.square

## [1] 0.1264072

sum\_of\_model\_7$r.square

## [1] 0.1462924

#The largest reduction in variability with the average length of stay is Infection Risk