Chad Huntebrinker’s Homework 5

Chad Huntebrinker

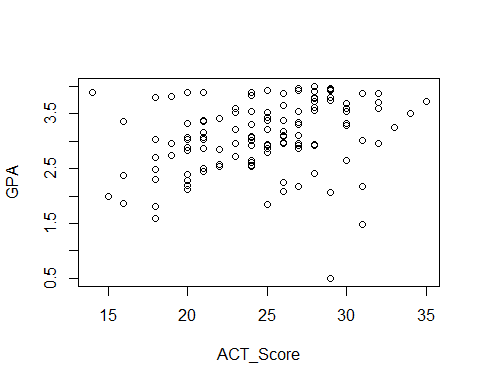
2024-10-01

Question 3.3:

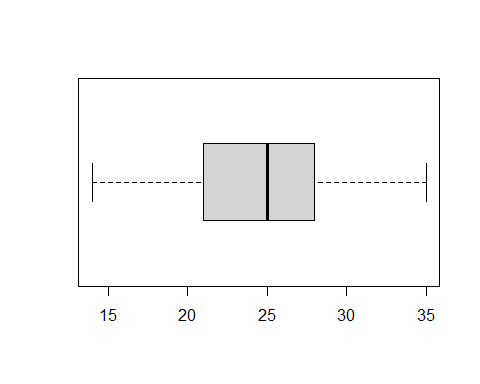
#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Grade\_Point\_Average\_Data.xlsx")  
  
# GPA ~ ACT  
# Fit the model and get the summary of the model  
model\_1 <- lm(GPA~ACT\_Score,data=excel\_data)  
sum\_of\_model\_1 <- summary(model\_1)  
sum\_of\_model\_1

##   
## 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

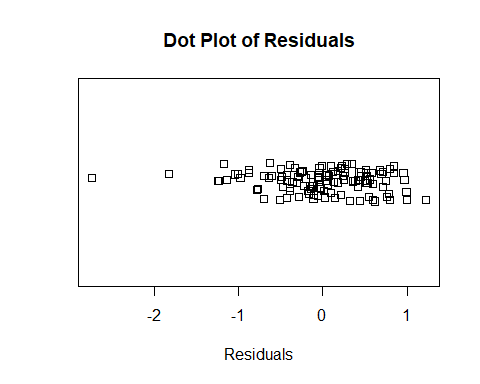
# Plot the graph just to see  
plot(GPA~ACT\_Score,data=excel\_data)



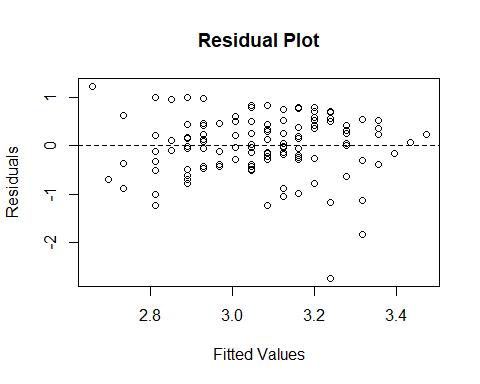
#Problem 3.3a - Create boxplot  
boxplot(excel\_data$ACT\_Score, horizontal = TRUE)



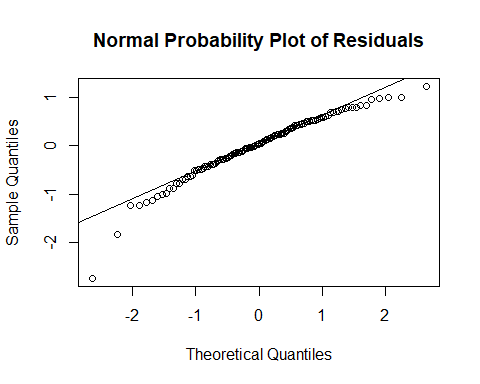
#Problem 3.3b - Create dot plot of residuals  
residuals\_model\_1 <- resid(model\_1)  
  
stripchart(residuals\_model\_1, main = "Dot Plot of Residuals", xlab = "Residuals", method = "jitter")



#Problem 3.3c - Create plot of residuals against the fitted values  
fitted\_values <- fitted(model\_1)  
plot(residuals\_model\_1 ~ fitted\_values, main = "Residual Plot", xlab = "Fitted Values", ylab = "Residuals")  
abline(h = 0, lty = 2)



#Problem 3.3d - Create Normal Plot  
q <- qqnorm(residuals\_model\_1, main = "Normal Probability Plot of Residuals")  
qqline(residuals\_model\_1)



cor.test(q$x, q$y, method=c("pearson")) # Table B6 in book

##   
## Pearson's product-moment correlation  
##   
## data: q$x and q$y  
## t = 47.128, df = 118, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9634938 0.9821475  
## sample estimates:  
## cor   
## 0.9744497

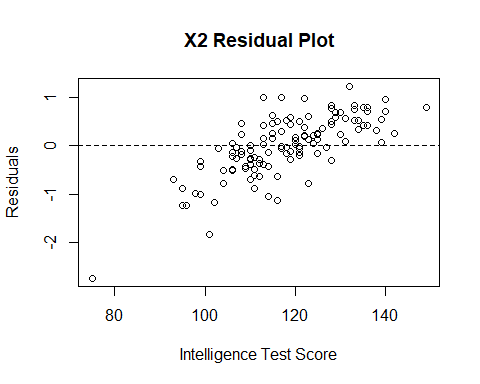
#Problem 3.3e - calculate t sample test statistic and the Brown-Forsythe test  
group\_1 <- subset(excel\_data, excel\_data$ACT\_Score < 26)  
group\_2 <- subset(excel\_data, excel\_data$ACT\_Score >= 26)  
  
model\_group1 <- lm(GPA~ACT\_Score,data=group\_1)  
model\_group2 <- lm(GPA~ACT\_Score,data=group\_2)  
  
median\_1 <- median(resid(model\_group1))  
median\_2 <- median(resid(model\_group2))  
  
abs\_dev\_group1 <- abs(resid(model\_group1) - median\_1)  
abs\_dev\_group2 <- abs(resid(model\_group2) - median\_2)  
  
  
s <- sqrt((sum((abs\_dev\_group1 - mean(abs\_dev\_group1))^2) + sum((abs\_dev\_group2 - mean(abs\_dev\_group1))^2))   
 / (nrow(excel\_data) - 2))  
  
two\_sample\_t <- ((mean(abs\_dev\_group1) - mean(abs\_dev\_group2)) / (s \* (sqrt(1/nrow(group\_1) + 1/nrow(group\_2)))))  
print(two\_sample\_t)

## [1] -0.9533587

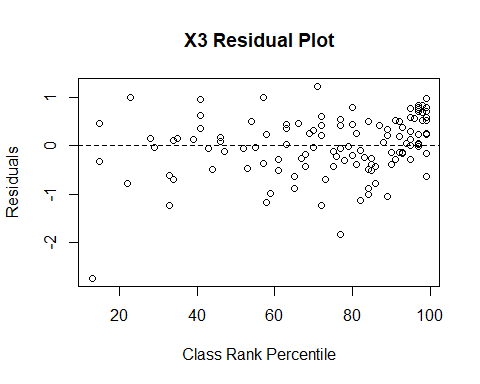
#Is H0 true? We find it is not  
abs(two\_sample\_t) > qt(0.995, 118)

## [1] FALSE

#Problem 3.3f - Plot x2 and x3 with the residuals  
plot(residuals\_model\_1 ~ excel\_data$Intelligence\_Test\_Score,   
 main = "X2 Residual Plot", xlab = "Intelligence Test Score", ylab = "Residuals")  
abline(h = 0, lty = 2)



plot(residuals\_model\_1 ~ excel\_data$Class\_Rank\_Percentile,   
 main = "X3 Residual Plot", xlab = "Class Rank Percentile", ylab = "Residuals")  
abline(h = 0, lty = 2)

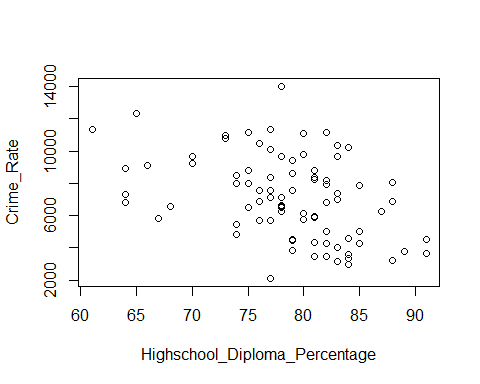


Question 3.8:

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Crime\_Rate\_Data.xlsx")  
  
# Fit the model and get the summary of the model  
model\_1 <- lm(Crime\_Rate~Highschool\_Diploma\_Percentage,data=excel\_data)  
sum\_of\_model\_1 <- summary(model\_1)  
sum\_of\_model\_1

##   
## 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

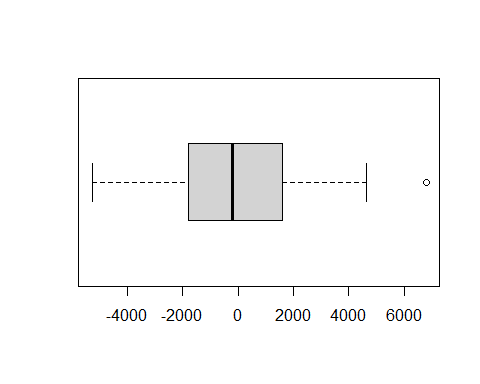
# Plot the graph just to see  
plot(Crime\_Rate~Highschool\_Diploma\_Percentage,data=excel\_data)



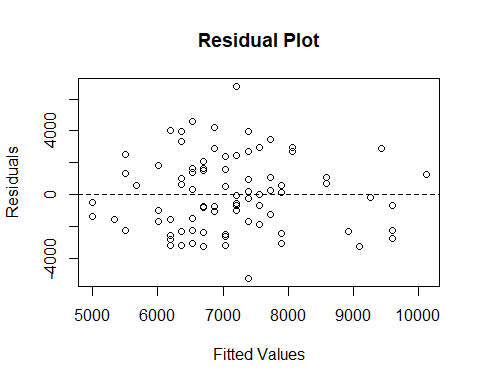
#Problem 3.4a - Create stem and leaf plot  
stem(excel\_data$Highschool\_Diploma\_Percentage)

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 6 | 1444  
## 6 | 5678  
## 7 | 00334444  
## 7 | 5555666677777778888888999999  
## 8 | 000011111112222222233333344444  
## 8 | 55578889  
## 9 | 11

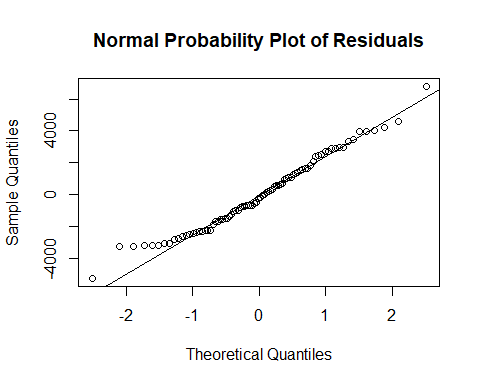
#Problem 3.4b - Create boxplot of residuals  
residuals\_model\_1 <- resid(model\_1)  
  
boxplot(residuals\_model\_1, horizontal = TRUE)



#Problem 3.4c - Create a residual plot with the fitted values  
fitted\_values <- fitted(model\_1)  
plot(residuals\_model\_1 ~ fitted\_values, main = "Residual Plot", xlab = "Fitted Values", ylab = "Residuals")  
abline(h = 0, lty = 2)



#Problem 3.4d - Get the correlation coefficient and the normal probability plot  
q <- qqnorm(residuals\_model\_1, main = "Normal Probability Plot of Residuals")  
qqline(residuals\_model\_1)



cor.test(q$x, q$y, method=c("pearson")) # Table B6 in book

##   
## Pearson's product-moment correlation  
##   
## data: q$x and q$y  
## t = 59.465, df = 82, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9824370 0.9926123  
## sample estimates:  
## cor   
## 0.988603

#Problem 3.4e - Do the Brown-Forsythe test  
group\_1 <- subset(excel\_data, excel\_data$Highschool\_Diploma\_Percentage < 69)  
group\_2 <- subset(excel\_data, excel\_data$Highschool\_Diploma\_Percentage >= 69)  
  
model\_group1 <- lm(Crime\_Rate~Highschool\_Diploma\_Percentage,data=group\_1)  
model\_group2 <- lm(Crime\_Rate~Highschool\_Diploma\_Percentage,data=group\_2)  
  
median\_1 <- median(resid(model\_group1))  
median\_2 <- median(resid(model\_group2))  
  
abs\_dev\_group1 <- abs(resid(model\_group1) - median\_1)  
abs\_dev\_group2 <- abs(resid(model\_group2) - median\_2)  
  
  
s <- sqrt((sum((abs\_dev\_group1 - mean(abs\_dev\_group1))^2) + sum((abs\_dev\_group2 - mean(abs\_dev\_group1))^2))   
 / (nrow(excel\_data) - 2))  
  
two\_sample\_t <- ((mean(abs\_dev\_group1) - mean(abs\_dev\_group2)) / (s \* (sqrt(1/nrow(group\_1) + 1/nrow(group\_2)))))  
print(two\_sample\_t)

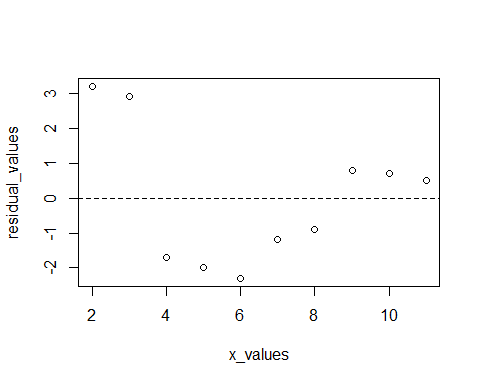
## [1] -0.9155

#Is H0 true? We find it is not.  
abs(two\_sample\_t) > qt(0.975, 82)

## [1] FALSE

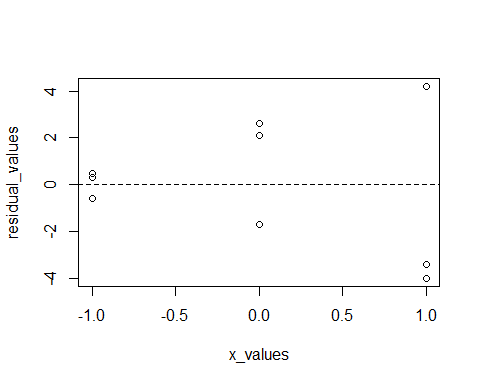
Question 3.9:

#Chad Huntebrinker  
  
#Problem 3.9  
x\_values <- c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)  
residual\_values <- c(3.2, 2.9, -1.7, -2.0, -2.3, -1.2, -0.9, 0.8, 0.7, 0.5)  
  
plot(residual\_values ~ x\_values)  
abline(h = 0, lty = 2)



Question 3.11:

#Chad Huntebrinker  
  
#Problem 3.11a  
  
x\_values <- c(-1, 0, 1, -1, 0, 1, -1, 0, 1)  
residual\_values <- c(0.5, 2.1, -3.4, 0.3, -1.7, 4.2, -0.6, 2.6, -4.0)  
  
plot(residual\_values ~ x\_values)  
abline(h = 0, lty = 2)



#They have a megaphone type  
  
#Problem 3.11b  
m.sig <- lm(residual\_values^2~x\_values)  
anova(m.sig)

## Analysis of Variance Table  
##   
## Response: residual\_values^2  
## Df Sum Sq Mean Sq F value Pr(>F)   
## x\_values 1 330.04 330.04 51.348 0.0001828 \*\*\*  
## Residuals 7 44.99 6.43   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

csq <- (anova(m.sig)$"Sum Sq"[1]/2)/((anova(m.sig)$"Sum Sq"[2]/9)^2)  
csq > qchisq(0.95,1)

## [1] TRUE

#Ha is true