MA 4710 Homework 2

Benjamin Hendrick

February 8, 2016

Problem 2.13

Load the data.

```
CH01PR19 <- read.table("/var/folders/6j/_n_8wtg912559sqd775wfctw0000gn/T//Rtmp8EbqFk/datacdf786779fb",
                       quote="\"", comment.char="")
```

Part A

Obtain the 95% confidence interval using the `confint` function in R.

```
model <- lm(V1 ~ V2, data = CH01PR19)
confint(model)
```

```
##                  2.5 %     97.5 %
## (Intercept) 1.47859015 2.74950842
## V2          0.01353307 0.06412118
```

Based on the output above, the 95% confidence interval is (0.01353307, 0.06412118). Because zero is not in the confidence interval, we can assume that slope is not zero.

Part B

```
predict(model, CH01PR19, interval="predict")[120,]
```

```
##      fit      lwr      upr
## 3.201209 1.959355 4.443063
```

The prediction interval for a score of 28 is (1.959355, 4.443063). The prediction interval contains the fitted score (3.201209). This means that we can be 95% sure that the predicted GPA, with an ACT score of 28, will be between 1.959355 and 4.443063.

Part C

The prediction interval is larger than the confidence interval. The prediction interval should be larger because it takes into account the variation between Y and X.

Part D

```r
yh <- 2.11405 + (28)*0.03883
w2 <- qf(.95,2,118)
yh + sqrt(w2)*yh
```

```
## [1] 8.813226
```

```r
yh - sqrt(w2)*yh
```

```
## [1] -2.410646
```

The boundary values of the 95% confidence interval are (-2.410646, 8.813226). This interval is larger than the interval in part B. This interval shouldn't be much larger than the prediction interval.

Problem 2.23

Use the same data and model as in Problem 2.13

Part A

```r
anova(model)
```

```
## Analysis of Variance Table
##
## Response: V1
##            Df Sum Sq Mean Sq F value   Pr(>F)
## V2          1  3.588  3.5878  9.2402 0.002917 **
## Residuals 118 45.818  0.3883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The code above provides the following ANOVA table.

	df	SS	MS	F
Model	1	3.588	3.588	9.2402
Error	118	45.818	0.3883	
Total	119	49.507		

Part B

MSE estimates the variance. MSR estimates the variance plus the residual error. MSR and MSE predict the same values if the residual error is zero.

Part C

State the hypothesis:

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$

Reject the null hypothsis if the F value from the ANOVA table is greater than the critical value. Find the critical value.

$$F_{0.01}(1, 118) = 6.854641$$

We reject the null hypothesis becasue the test statistic (9.2402) is greater than the critical value (6.854641).

Problem 2.26

Read the data.

```
CH01PR22 <- read.table("/var/folders/6j/_n_8wtg912559sqd775wfctw0000gn/T//Rtmp8EbqFk/datacdf58d91d97",
                       quote="\"", comment.char="")
```

Part A

```
model <- lm(V1~V2, data = CH01PR22)
anova(model)
```

```
## Analysis of Variance Table
##
## Response: V1
##           Df Sum Sq Mean Sq F value    Pr(>F)
## V2         1 5297.5  5297.5  506.51 2.159e-12 ***
## Residuals 14  146.4    10.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The code above provides the following ANOVA table.

	df	SS	MS	F
Model	1	5297.5	5297.5	506.51
Error	14	146.4	10.5	
Total	15	5,443.9		

Part B

State the hypothesis:

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$

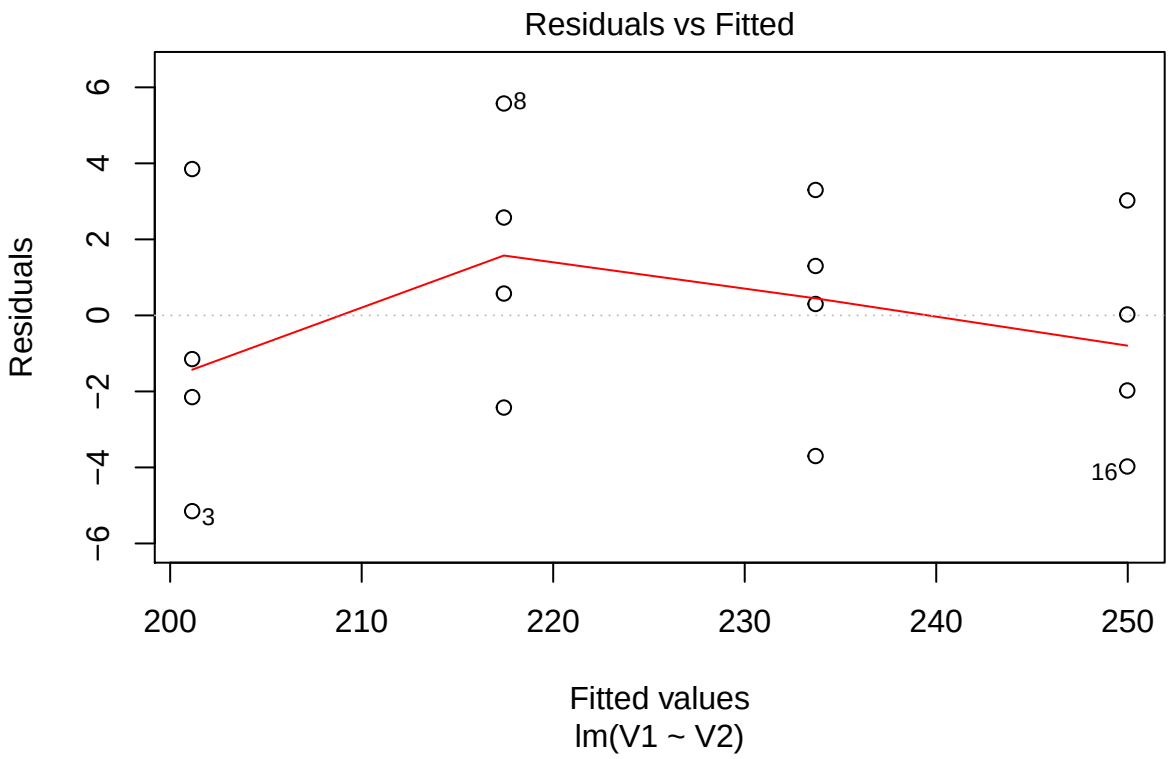
Reject the null hypothsis if the F value from the ANOVA table is greater than the critical value. Find the critical value.

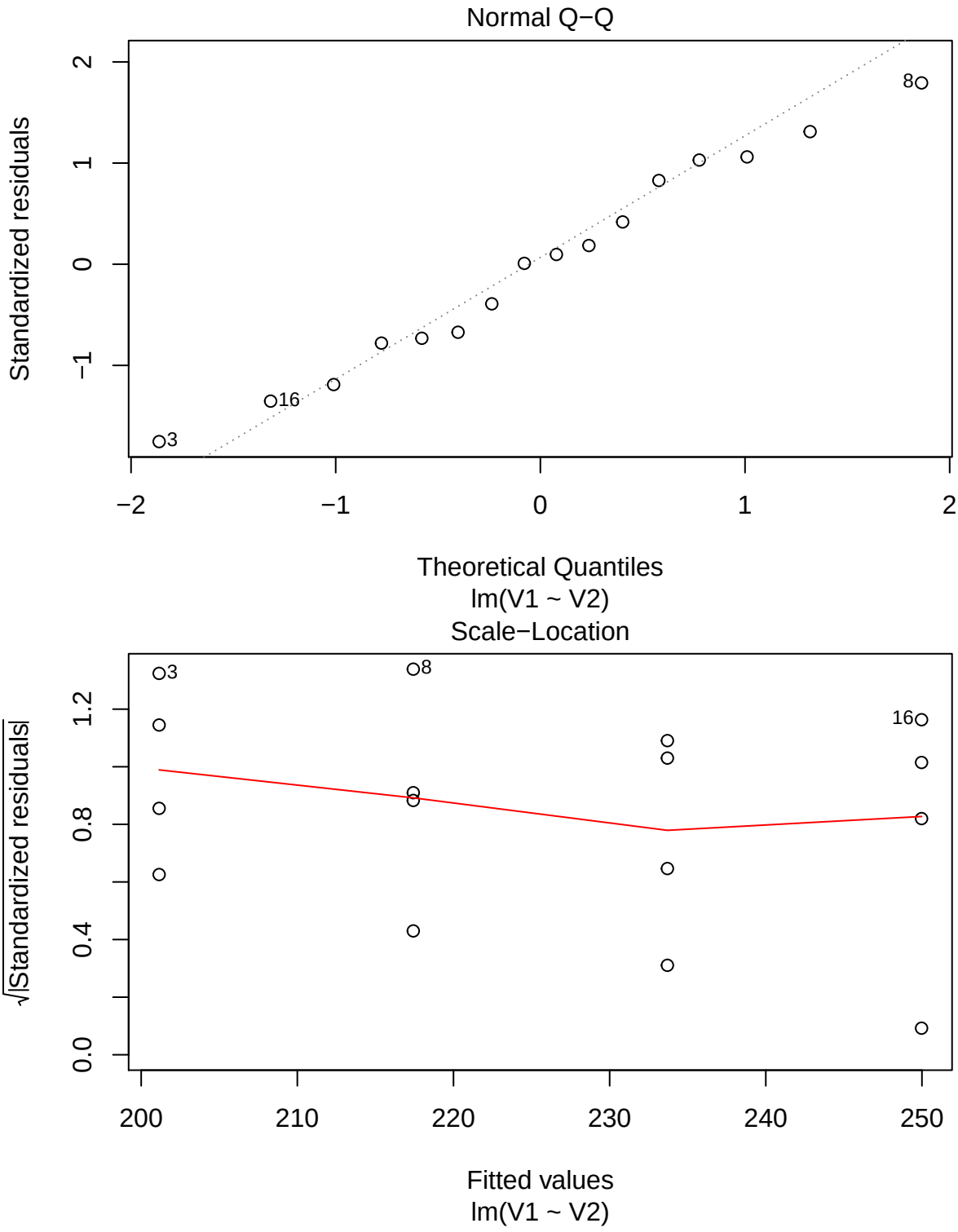
$$F_{0.01}(1, 14) = 8.861593$$

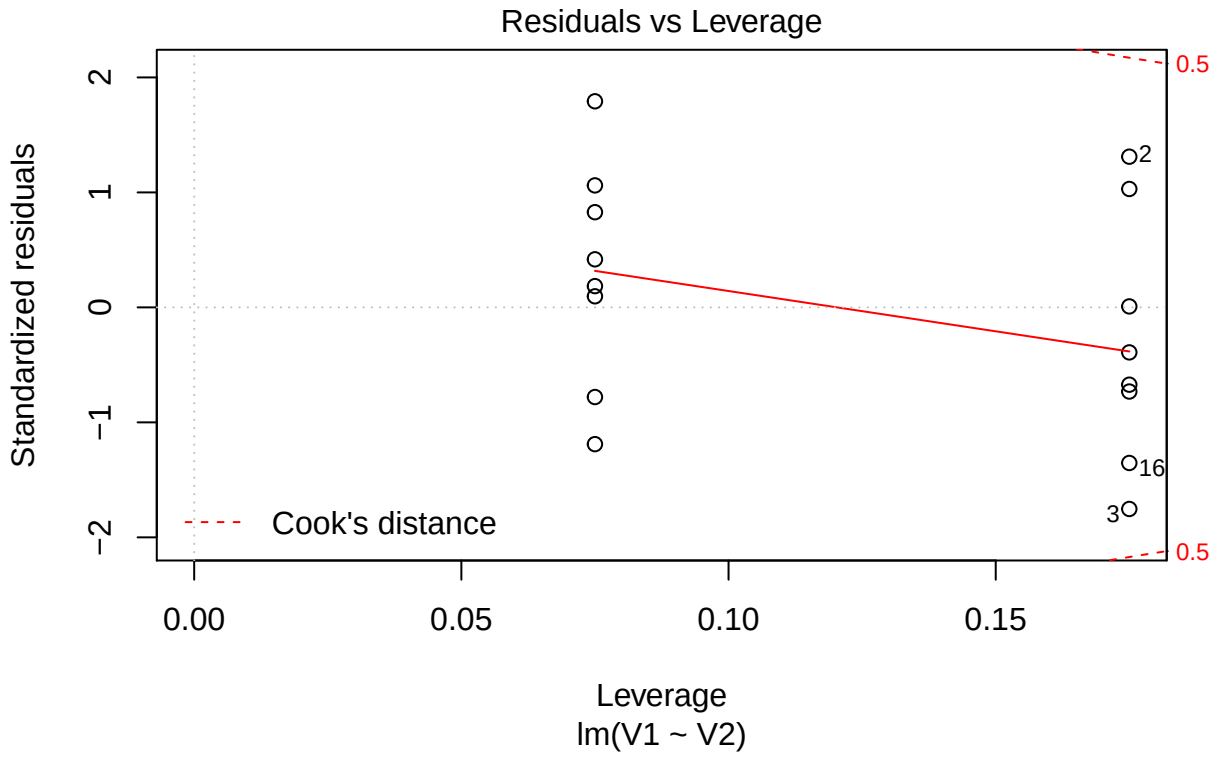
We reject the null hypothesis becasue the test statistic (506.51) is greater than the critical value (8.861593).

Part C

```
plot(model)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(V1 ~ V2)
Scale−Location
√|Standardized residuals|
Fitted values
lm(V1 ~ V2)
3
8
16



The two plots of interest are the residual plot (first plot) and the QQ plot (second plot). Based on the plots, the SSR appears to be a larger component of the SSTO. This implies that the magnitude of R^2 will be large.

Part C

```
summary(model)
```

```
##
## Call:
## lm(formula = V1 ~ V2, data = CH01PR22)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1500 -2.2188  0.1625  2.6875  5.5750
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 168.60000    2.65702   63.45  < 2e-16 ***
## V2            2.03438    0.09039   22.51 2.16e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.234 on 14 degrees of freedom
## Multiple R-squared:  0.9731, Adjusted R-squared:  0.9712
## F-statistic: 506.5 on 1 and 14 DF,  p-value: 2.159e-12
```

From the code above, $R^2 = 0.9731$ and $r = \sqrt{0.9731} = 0.9865$.