Problem Set 5

QTM 200: Applied Regression Analysis

Due: March 4, 2020

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on the course GitHub page in .pdf form.

- This problem set is due at the beginning of class on Wednesday, March 4, 2020. No late assignments will be accepted.

- Total available points for this homework is 100.

Using the `teengamb` dataset, fit a model with `gamble` as the response and the other variables as predictors.

```
gamble <- (data=teengamb)
# run regression on gamble with specified predictors
model1 <- lm(gamble ~ sex + status + income + verbal, gamble)
```

Answer the following questions:

(a) Check the constant variance assumption for the errors by plotting the residuals versus the fitted values.

```
res <- model1$residuals
fitted <- model1$fitted.values
plot(res ~ fitted, data = gamble)
abline(h=0)
```

Figure 1: Constant Variance
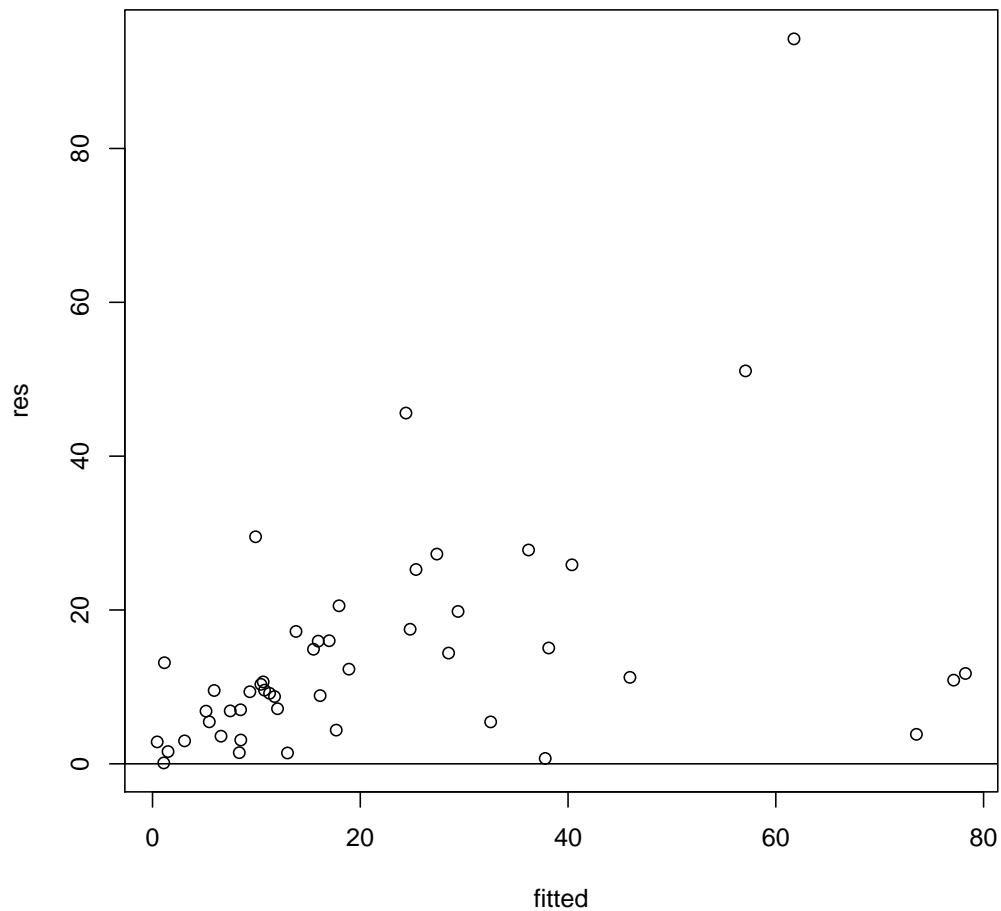
(b) Check the normality assumption with a Q-Q plot of the studentized residuals.
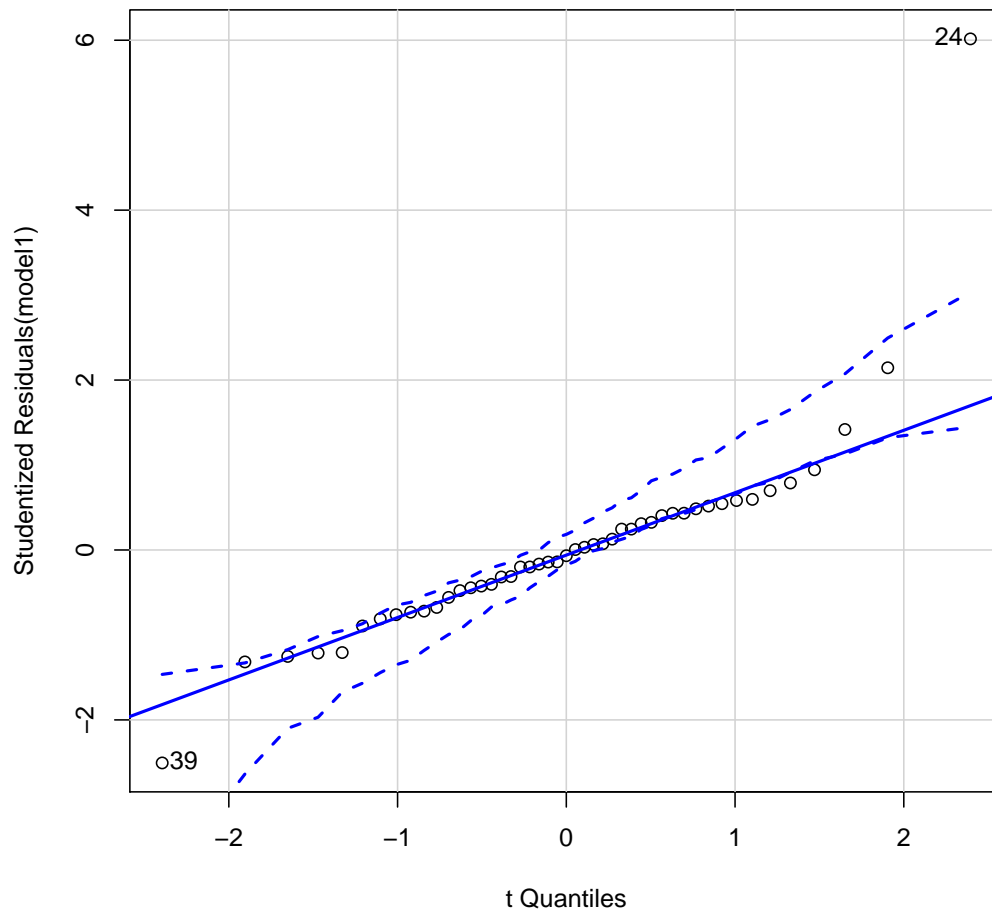
```
1  qqplot ( model1 )
```

See Figure 2 on page 3...

(c) Check for large leverage points by plotting the h values.

```
1  hatvalues ( model1 )
2  str ( model1 )  #n=47
3  plot ( hatvalues ( model1 ) )
4  abline ( h=2*3/47 ,  lty=2)
5  abline ( h=3*3/47 ,  lty=3)
```

The graph is on page 4...

Figure 2: Normality

(d) Check for outliers by running an `outlierTest`.

```
outlierTest(model1)
```

```
    rstudent unadjusted p-value Bonferroni p
24 6.016116         4.1041e-07    1.9289e-05
```

(e) Check for influential points by creating a "Bubble plot" with the hat-values and studentized residuals.

```
influencePlot(model1, sub="Circlee size is proportial to Cook's Distance")
```
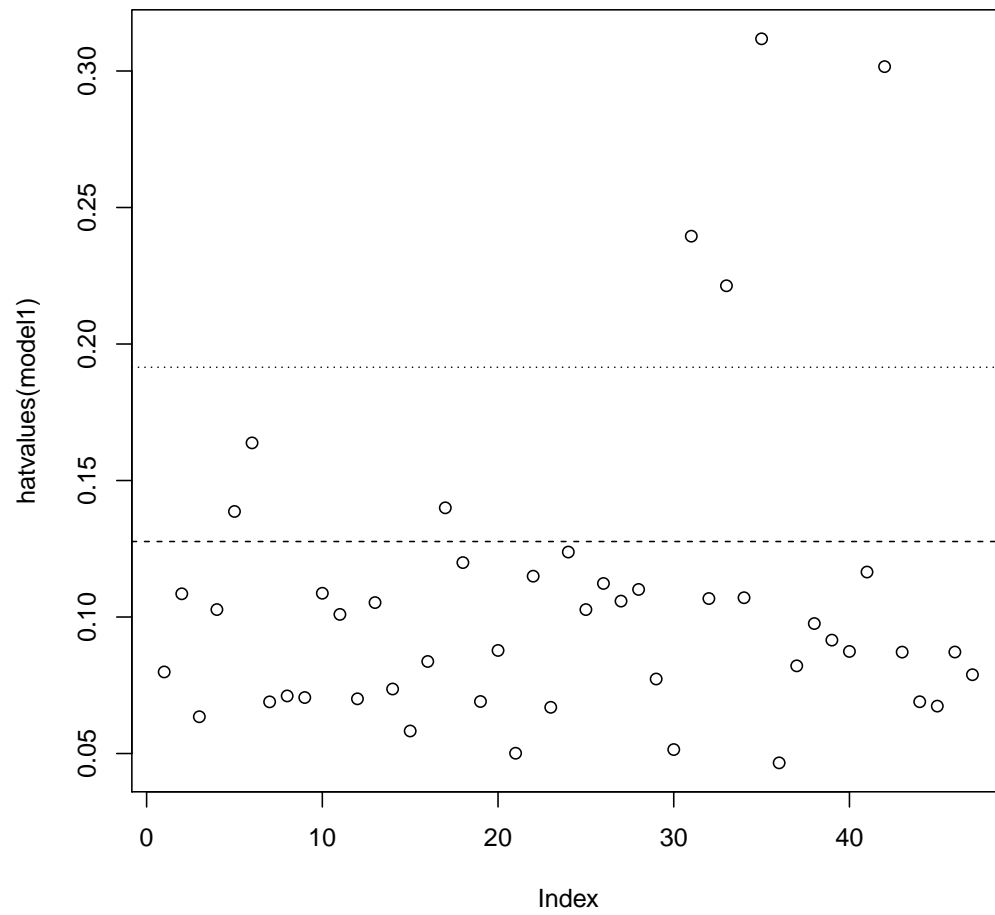
See the graph at the very bottom...

Figure 3: Hatvalues

Figure 4: Influential Points