week6Homework

M. Onimus

10/17/2020

Contents

Read in Data **1**

Question 1 **1**
Question/Answer 1a . . . 1
Question/Answer 1b . . . 1
Question/Answer 1c . . . 1
Question/Answer 1d . . . 2
Question/Answer 1e . . . 2
Question/Answer 1h . . . 4

Question 2 **4**
Question/Answer 2a . . . 4
Question/Answer 2b . . . 5
Question/Answer 2c . . . 5
Question/Answer 2d . . . 5
Question/Answer 2e . . . 7

Read in Data

Question 1

David is hoping to come up with a linear model that can be used to predict an individuals BMI using a subset of the PHMCcommunity health survey database.

Next, we wish to control for subject sex and food stamp receipt when looking at BMI and age.

Question/Answer 1a

What is the partial correlation between age and BMI, controlling for sex and food stamp receipt?

The partial correlation between age and BMI controlling for Sex and Food Stamps is 0.0197.

Question/Answer 1b

Write out the regression equation for the prediction of BMI using subject age, sex, and food stamp receipt.

$BMI = \beta_0 + \beta_1 age + \beta_2 sex + \beta_3 foodStamp + \epsilon_i$

Question/Answer 1c

What are the null and alternative hypotheses for the overall F test for your model?

$H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ $H_A : \beta_1 \neq 0$ and/or $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$

Question/Answer 1d

What are the results of inference about the overall model? For the three predictors?

```
##                 Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept) 33.782551975 0.452647203  74.633294 0.000000e+00
## RESPAGE      0.007700271 0.003973602   1.937857 5.266972e-02
## SEX01       -0.698436171 0.129064085  -5.411546 6.397380e-08
## GETSTAMP    -2.732417373 0.180969559 -15.098768 6.251975e-51
```

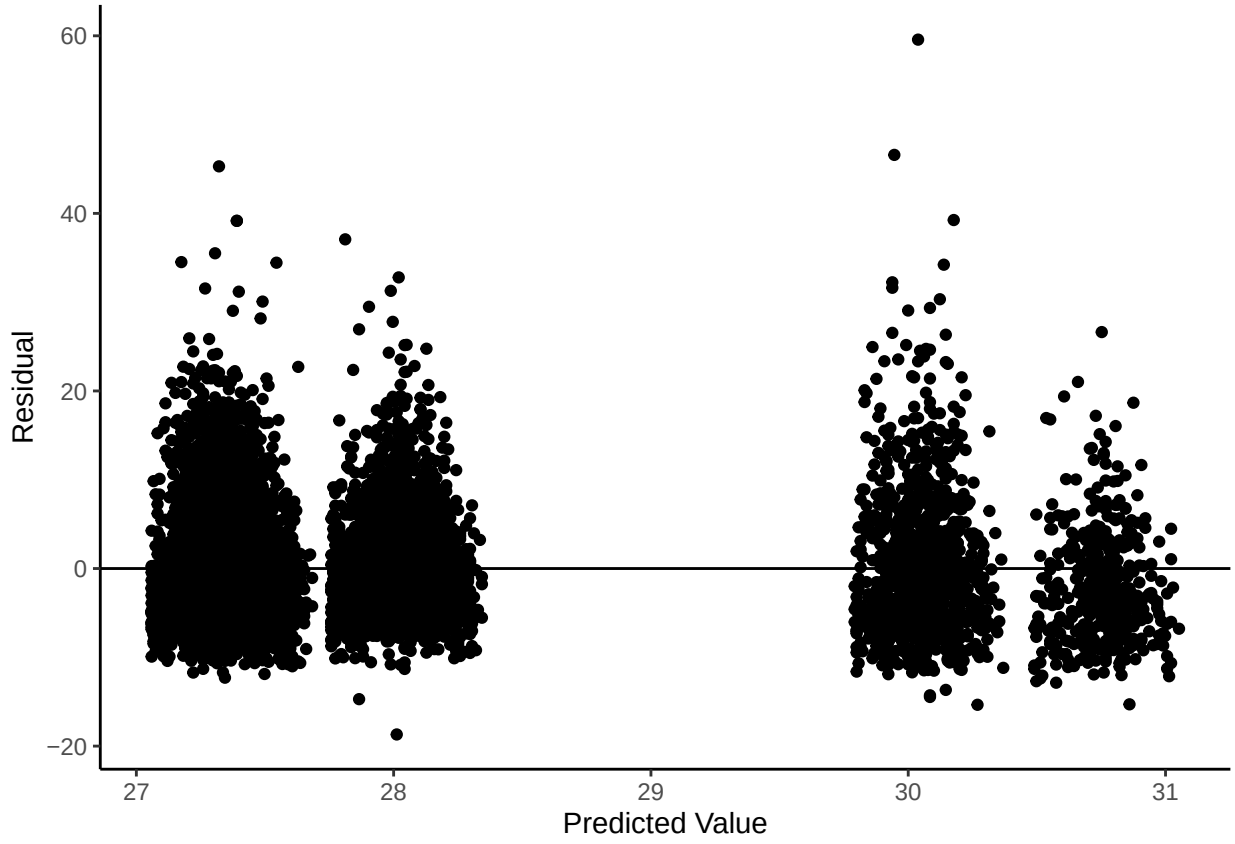
RESPAGE: $t_{3,9687} = 1.938$ $p = 0.053$

SEX01: $t_{3,9687} =$ -5.412 $p = 0$

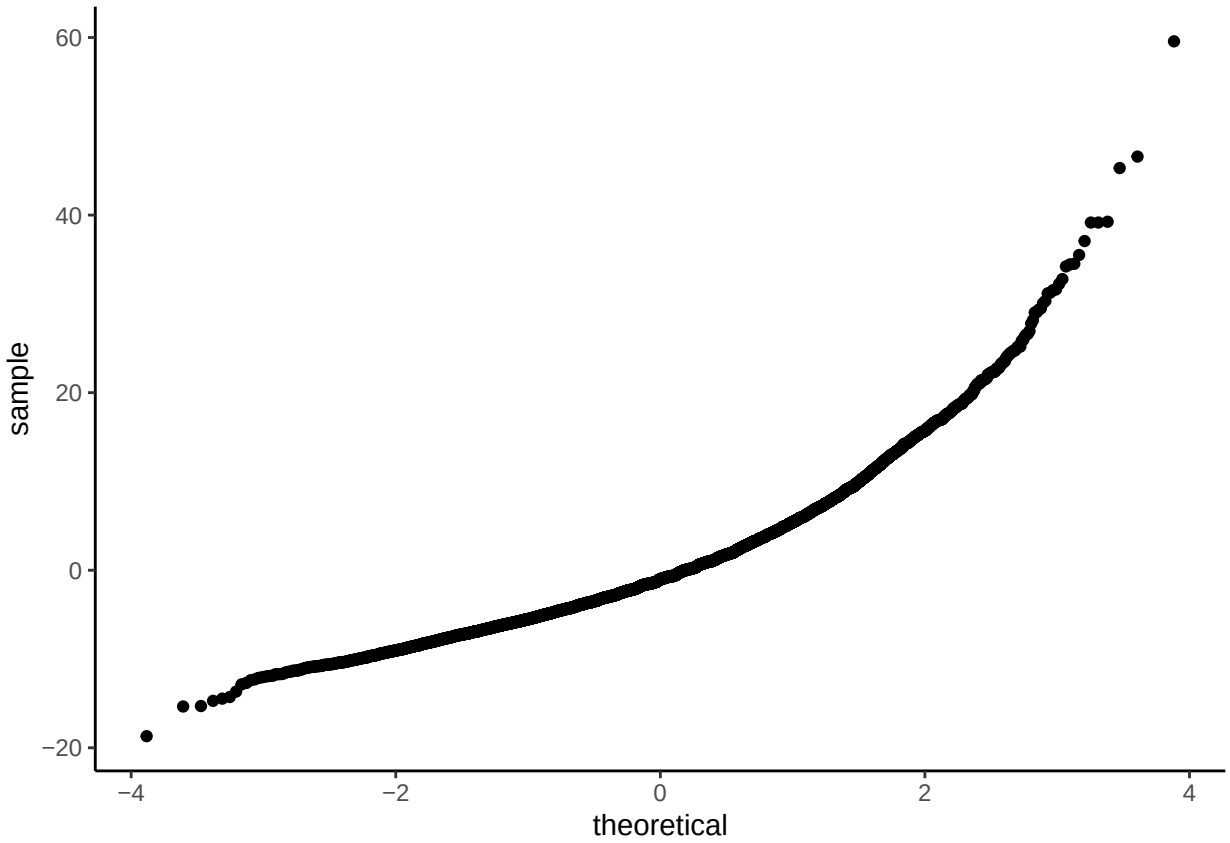
GETSTAMP: $t_{3,9687} =$ -15.099 $p = 0$

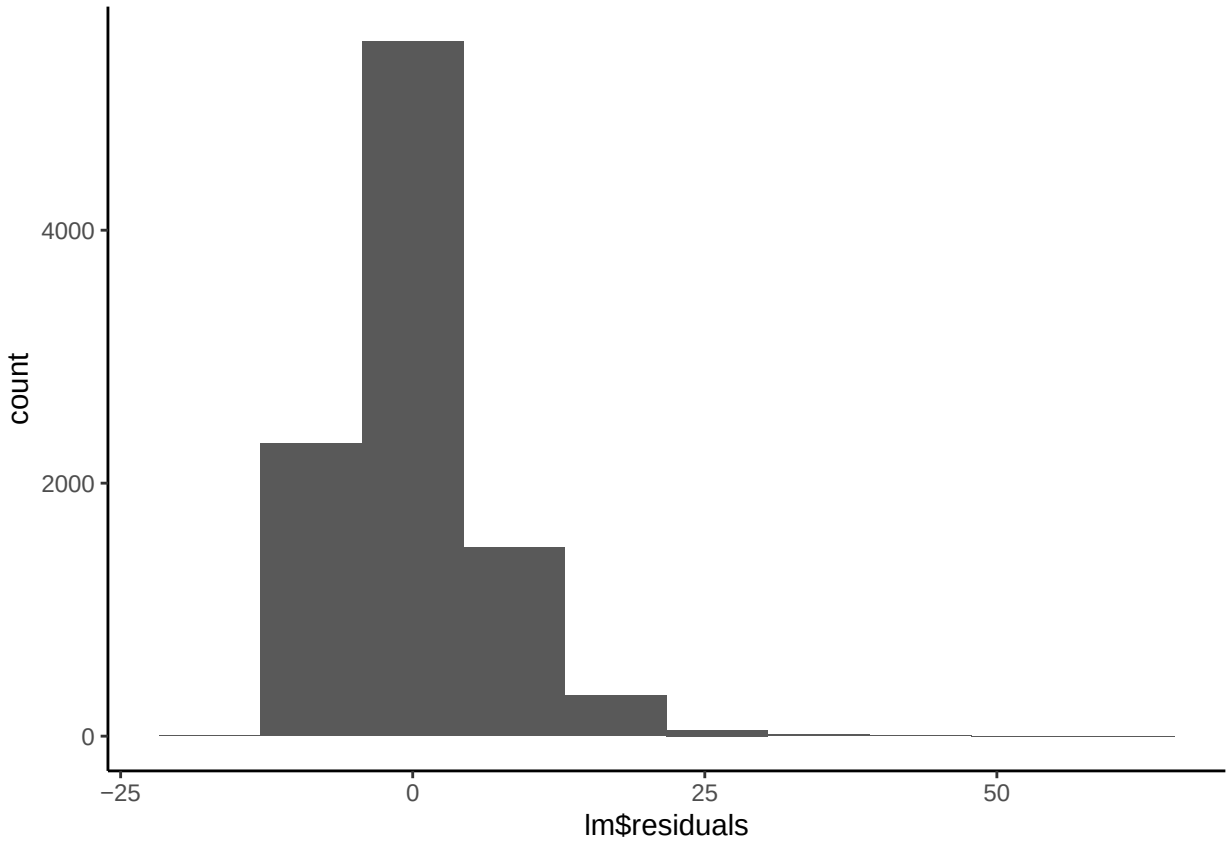
Question/Answer 1e

How do the assumptions of linear regression hold up in this case? Use the appropriate plots to support your discussion.



The residual/predicted plot above shows a slightly skewed distrubtion with a higher density of points below 0.


60
40
20
0
−20
sample
−4
−2
0
2
4
theoretical



The qqplot is not linear at all, showing extensive curving. The histogram shows a pretty skewed distrubtion. This is expected after seeing the residuals plot and qq plot.

Question/Answer 1h

Write a few sentences describing the results.

There maybe a signficant linear association between the 3 variables of interest:

RESPAGE: $t_{3,9687} = 1.938$ $p = 0.053$

SEX01: $t_{3,9687} =$ -5.412 $p = 0$

GETSTAMP: $t_{3,9687} =$ -15.099 $p = 0$

but the resulting R-square (0.025) is relatively weak. Additionally, there may be significant issues with using the model for prediction based upon the residuals and qq plot as well as the histogram.

Question 2

As it is known that there are between-sex differences in the aging process (e.g. rate of osteoporosis), it is of interest to include a sex-by-age interaction term in the model.

Question/Answer 2a

Practically speaking, what does having the interaction in the regression model do?

An interaction between two variables signify a relationship between the variables and the response is interdependent. Practiaclly speaking, this interaction means that both predictors are needed to actually model and predict the outcome variable. These interactions can greatly influence a model and leads to the main effects being inconsequential.

Question/Answer 2b

Write out the regression equation for the prediction of BMI using subject age, sex, their interaction, and food stamp receipt.

$BMI = \beta_0 + \beta_1 age + \beta_2 sex + \beta_3 age * sex + \beta_4 foodStamp + \epsilon_i$

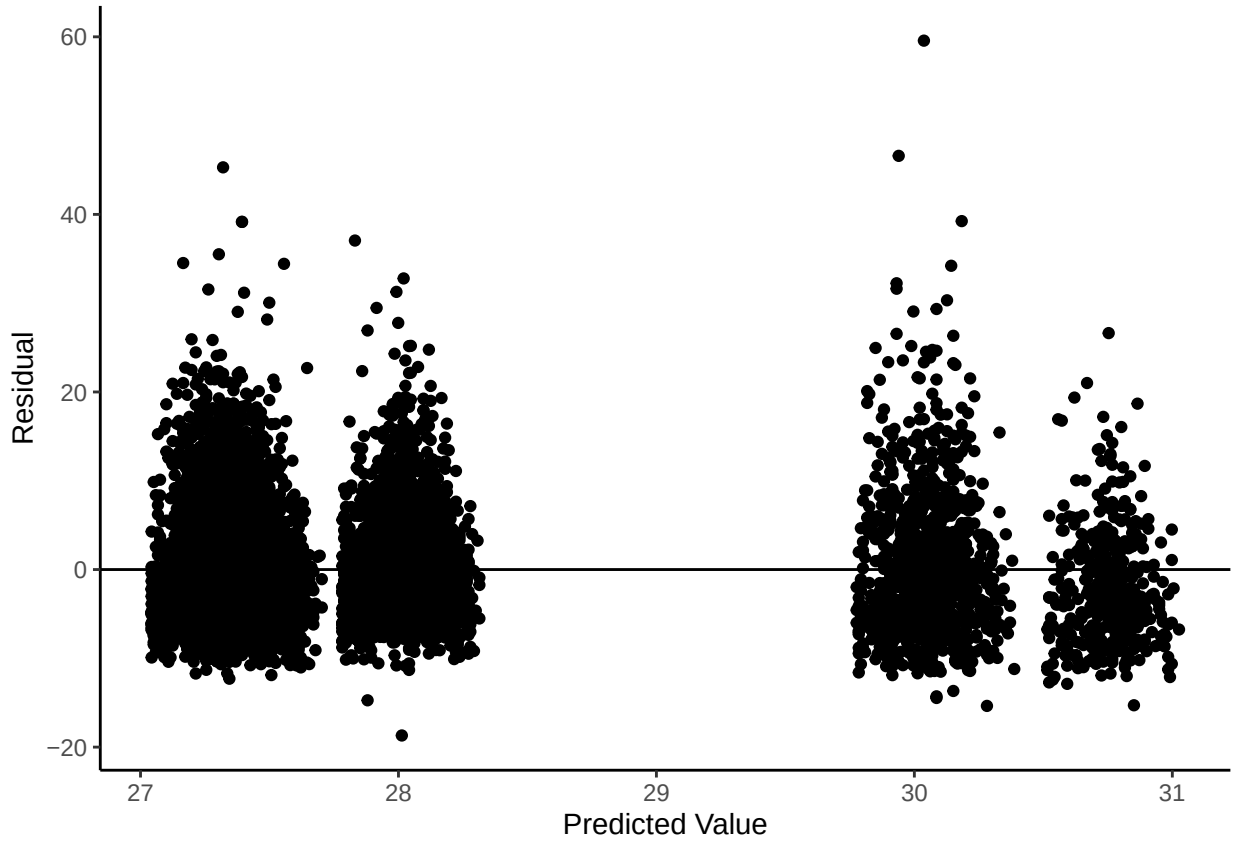
Question/Answer 2c

What are the results of inference about the interaction between age and sex on BMI, controlling for food stamp receipt?

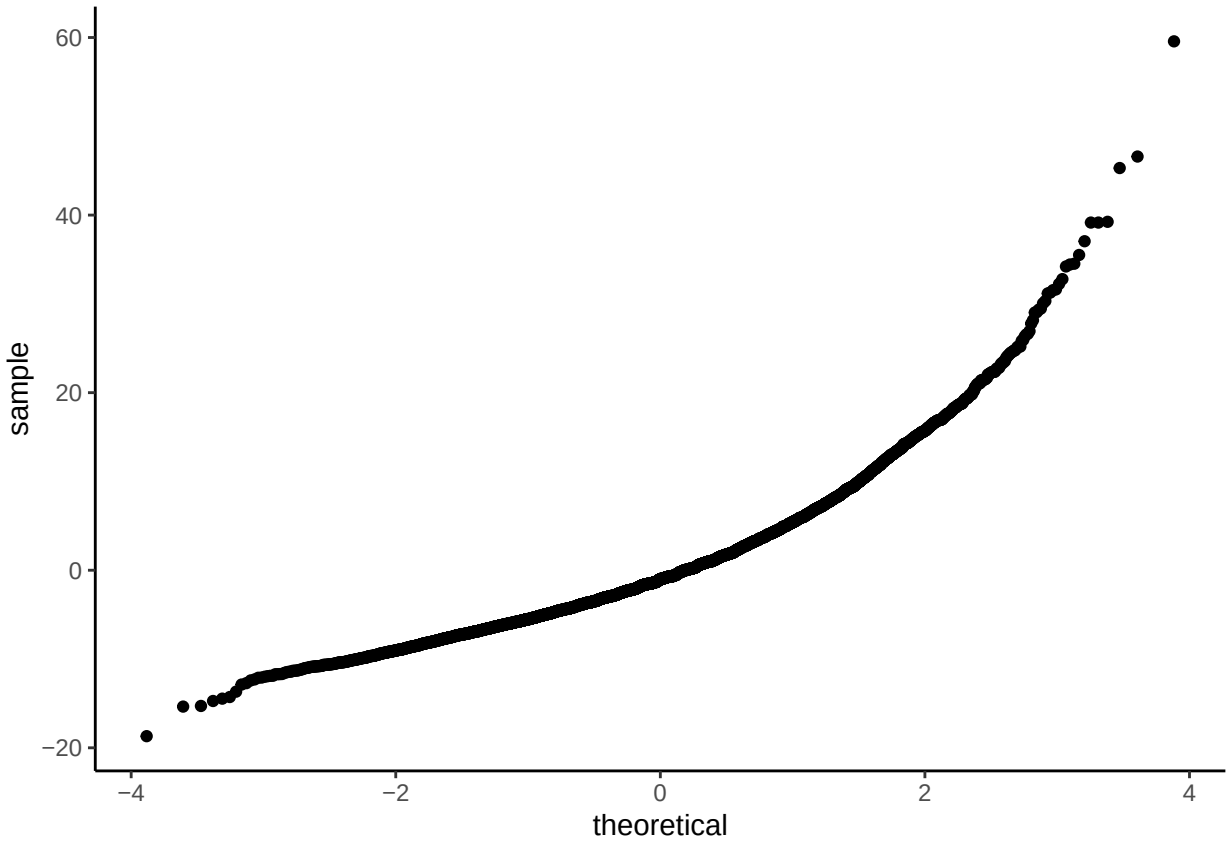
$t_{4,9686} =$ **0.14,** $p =$ **0.8876**

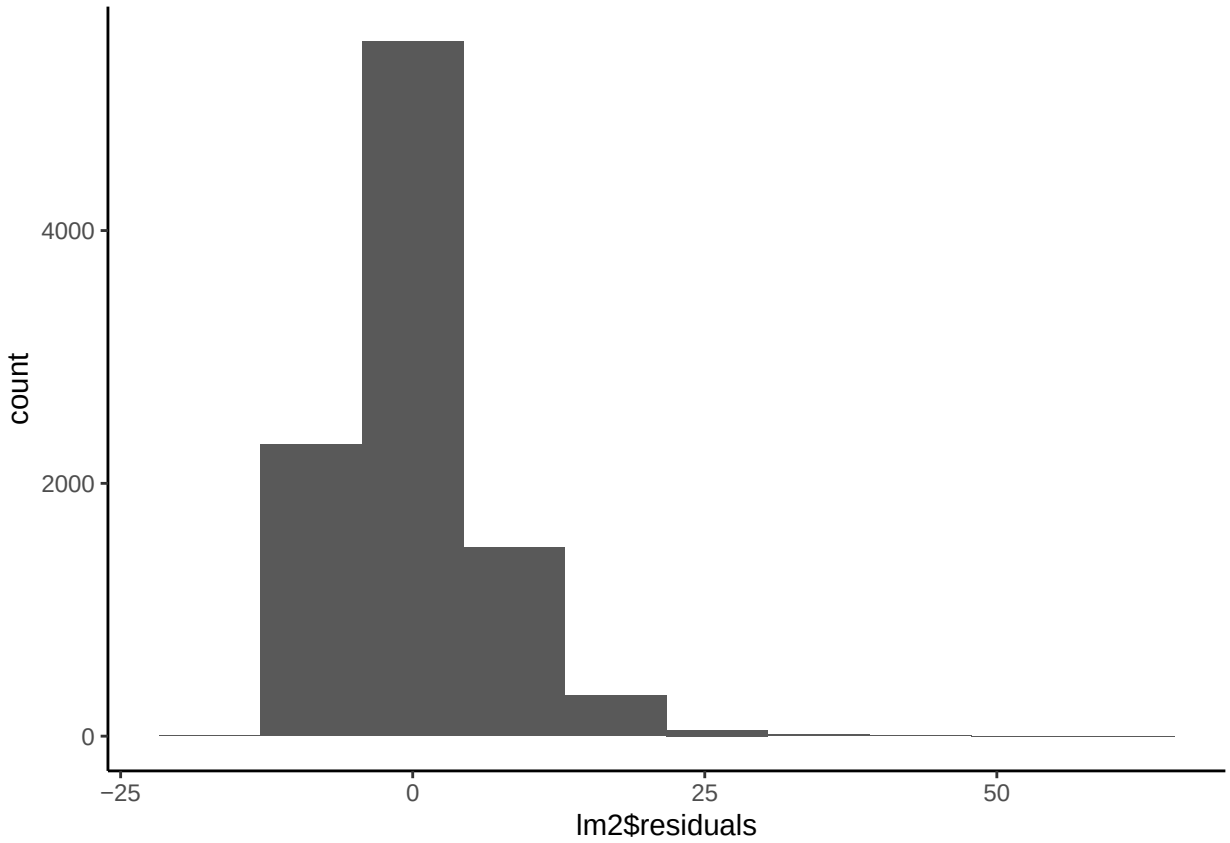
Question/Answer 2d

How do the assumptions of linear regression hold up in this case? Use the appropriate plots to support your discussion.



The residual/predicted plot above shows a slightly skewed distrubtion with a higher density of points below 0.


60
40
20
0
−20
sample
−4
−2
0
2
4
theoretical



The qqplot is not linear at all, showing extensive curving. The histogram shows a pretty skewed distrubtion. This is expected after seeing the residuals plot and qq plot.

Question/Answer 2e

Write a few sentences describing the results.

We did not observe a signifcant interaction between age and sex interaction term on BMI ($t_{4,9686} = 0.14$, $p = 0.8876$). The residual and qq plot as well as the histogram show non-normal trends, this model should be used with extreme caution and my not accurately predict BMI.