**LINEAR REGRESSION**: Homework 

*Professor Jingchen Liu*

Fan Yang

UNI: fy2232

# Problem 1 (2.2)

No, this conclusion does not imply that *X* and *Y* have no linear association. This result only tells us that *X* and *Y* are negative correlated, which means when *X* grows, value of *Y* decreases.

# Problem 2 (2.23)

## (a)



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DF | SS | MS | F |
| Regression | 1 | 3.588 | 3.5878 | 9.239763 |
| Error | 118 | 45.818 | 0.3883 |  |
| Total | 119 | 49.406 |  |  |

## (b)



So  is estimated by *MSR*



So  is estimated by *MSE*

When , *MSR* and *MSE* estimate the same quantity.

## (c)





if , then conclude 

else , reject 

While , so we can reject  and conclude 

## (d)

SSR = 3.588 is the absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model.

The relative reduction is . This is the same as coefficient of determination.

## (e)





## (f)

I think  has the more clear-cut operational interpretation. Because  equals to

Explained variation divided by Total variation, which represents the percentage of variation can be explained by our linear model.

# Problem 3 (2.26)

## (a)



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DF | SS | MS | F |
| Regression | 1 | 5297.5 | 5297.5 | 506.51 |
| Error | 14 | 146.4 | 10.5 |  |
| Total | 15 | 5443.9 |  |  |

## (b)



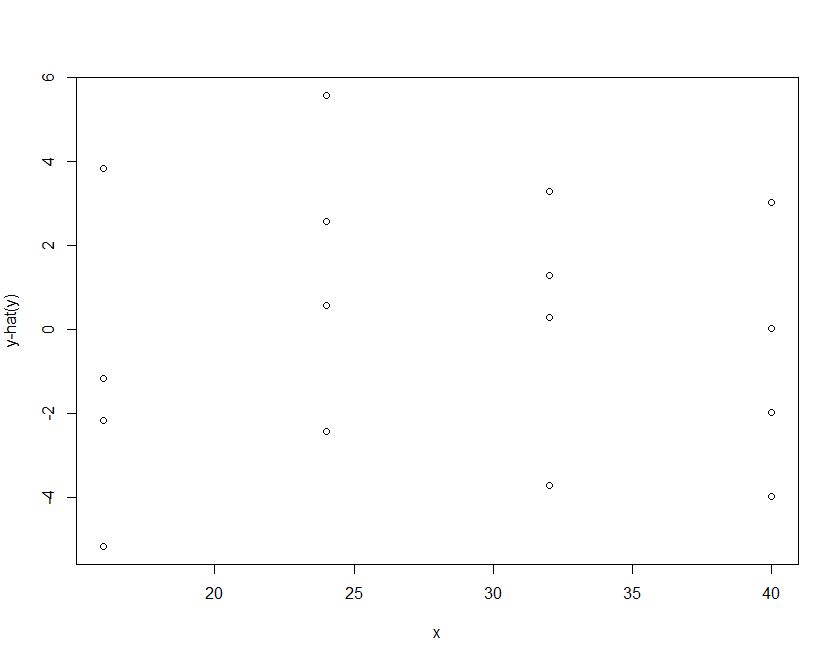


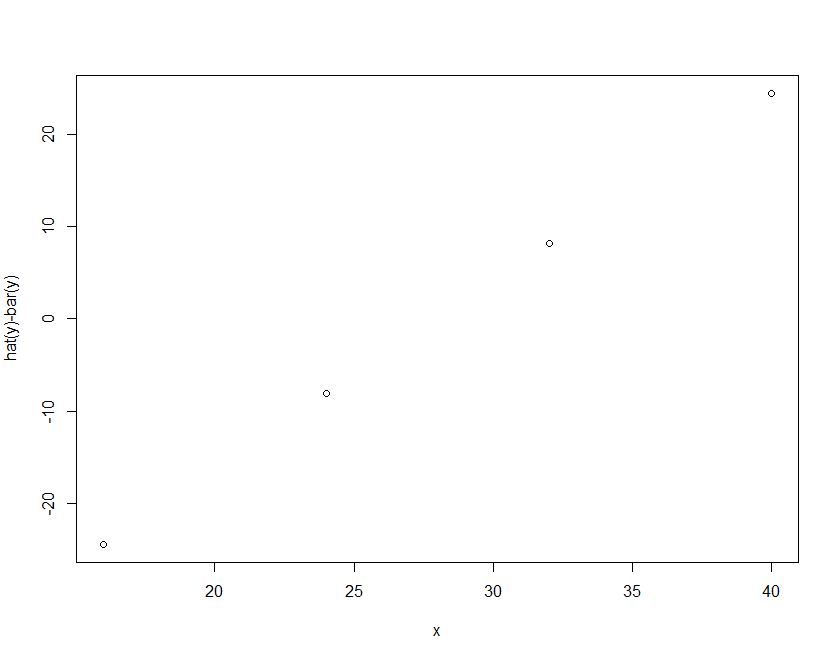
if , then conclude 

else , reject 

While , so we can reject  and conclude 

## (c)





From the graph we can say that SSR appear to be the larger component of SST.

While  so  is large.

# Problem 4 (2.56)

## (a)







## (b)



, 

, 

It would be worse to observe . Because it will bring more variation to our model. What’s more, these observations do not change the range of predictive variables.

But when we want to estimate the mean response for *X* = 8, adding these observations to our model will help us improve our preciseness. Because these observation points lie near *X* = 8, which provides more information around *X* = 8.

# Problem 5 (2.61)



As for this fraction, the denominator and numerator each has same expression for X and Y, which means X and Y have symmetric expressions. Therefore, the ratio is the same whether  is regressed on  or  is regressed on .

# Problem 6 (2.66)

## (a)





So ****

When 

The 95 percent confidence interval is



which is





## (b)



## (c)

The mean and standard deviation of the 200 estimates are:



As for theoretical results,





The mean of the 200 estimates is 3.967998, which is very close to theoretical results 4.

And the theoretical expectation of  should be 0.3952847. Compared with the standard deviation 0.3660616, we find the result differs from the theoretical expectation a little, but the difference is just about 0.3

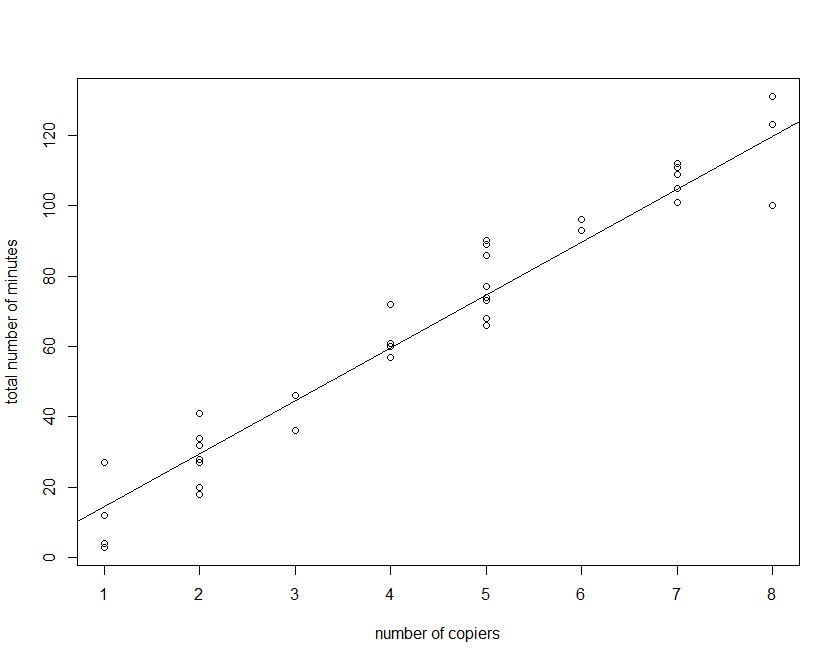
## (d)



From the results above, we find 100% of the 200 confidence intervals include. This result is consistent with theoretical expectations.

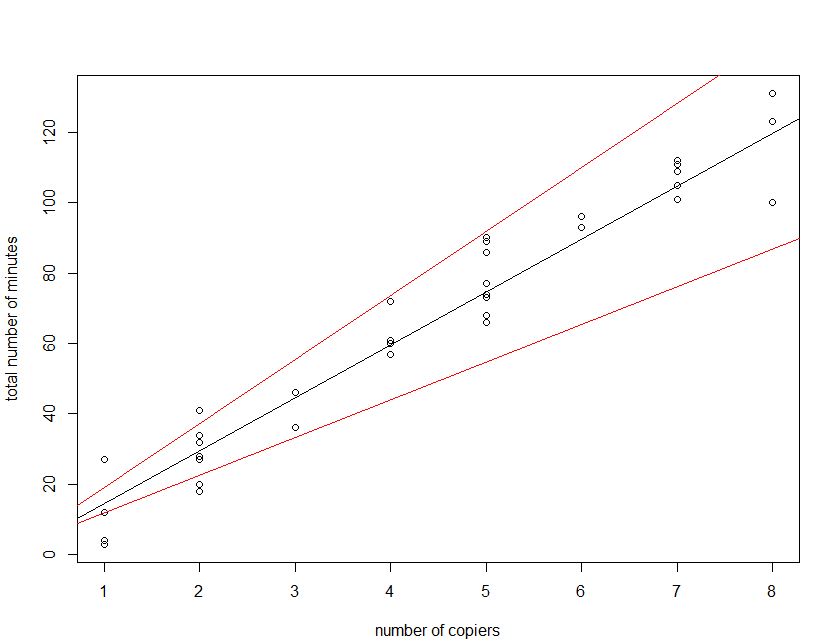
# Problem 7 (2.68)

## (a)



## (b)





The fitted regression line entirely lies between the confidence band. Except for some few points, most points also lie between the bands and near the fitted regression line. So the true regression relation has been precisely estimated.