Untitled

Edwin Chau

10/12/2019

Question 11 Part a

The coefficient estimate is 1.9939, the standard error is 0.1065, the t-statistic is 18.73 and the resulting p value is about 0. The model estimated that y increases by 1.9939 times for every increase in x. The standard estimate for the increase in y is about 0.1065, and the extremely small p value is supportive of x's significance as a predictor of y.

Question 11 Part b

When regressing y on x instead, the coefficient estimate is 0.3911, standard error for the estimate is 0.02689, and the t value and p value are the same. This means that for every increase in y, x increases by 0.3911 times. Furthermore, the similarly large t value and miniscule p value show that y is a very significant predictor of x.

Question 11 Part c

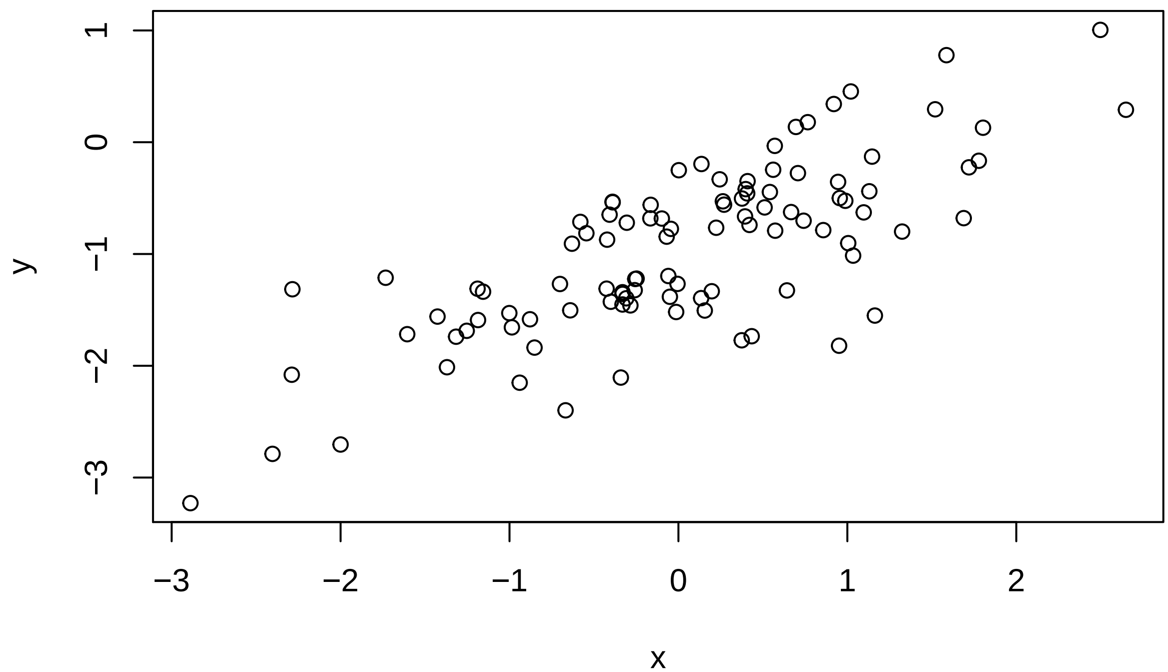
At first glance, the two regression models seem wrong because while regression y on x yielded a very similar formula to our original calculation of y using x, regressing x on y did not match. Instead of a slope of around 0.5 the slope was instead 0.3991. However, this could be due to the random nature of our x, and by switching the y and x axes we do not get a very symmetrical plot. By regressing x on y, the SSE effectively minimizes the horizontal distances instead of the vertical distances, and the non-symmetric data points will cause the two SSE minimizations to be different.

One thing of note is that the standard error of our y predictor is smaller than that of our x predictor, meaning our y coefficient estimate is more precise than our x coefficient estimate.

Question 13 Part c

The vector y is 100 entries long, with the intercept being -1 and the slope being 0.5.

Question 13 Part d

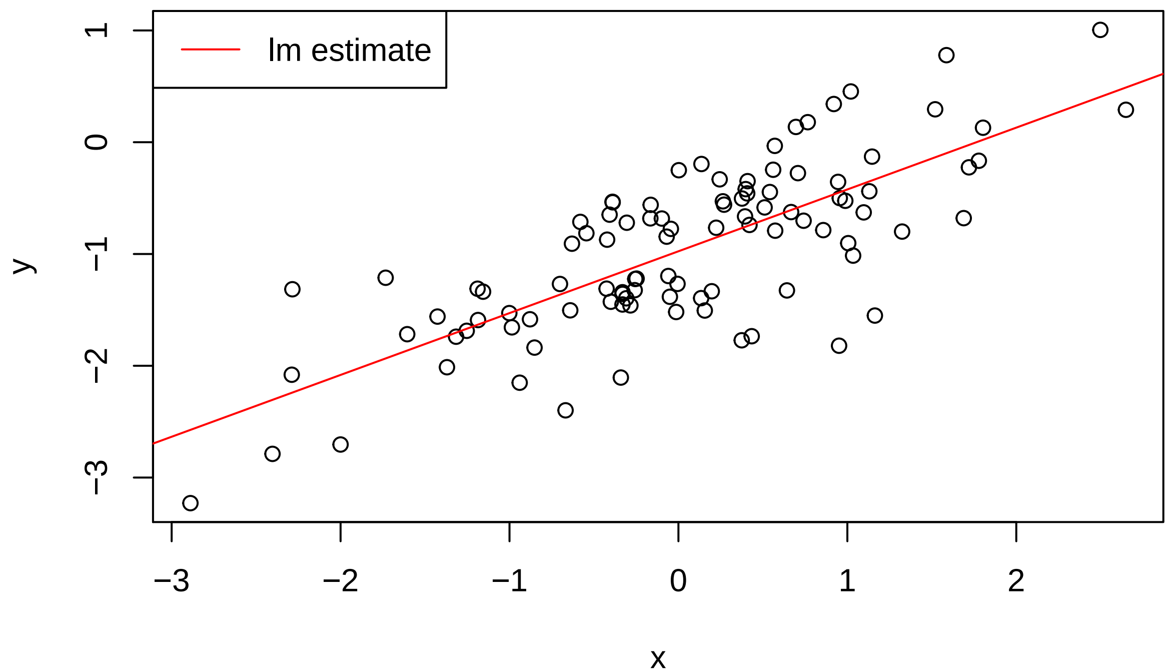


The relationship between x and y is moderately positively linear, as we introduced an error term to our linear equation.

Question 13 Part e

The linear regression model estimated the intercept as -0.97578 and the slope as 0.55311. These values are very close to the original -1 and 0.5, respectively.

Question 13 Part f

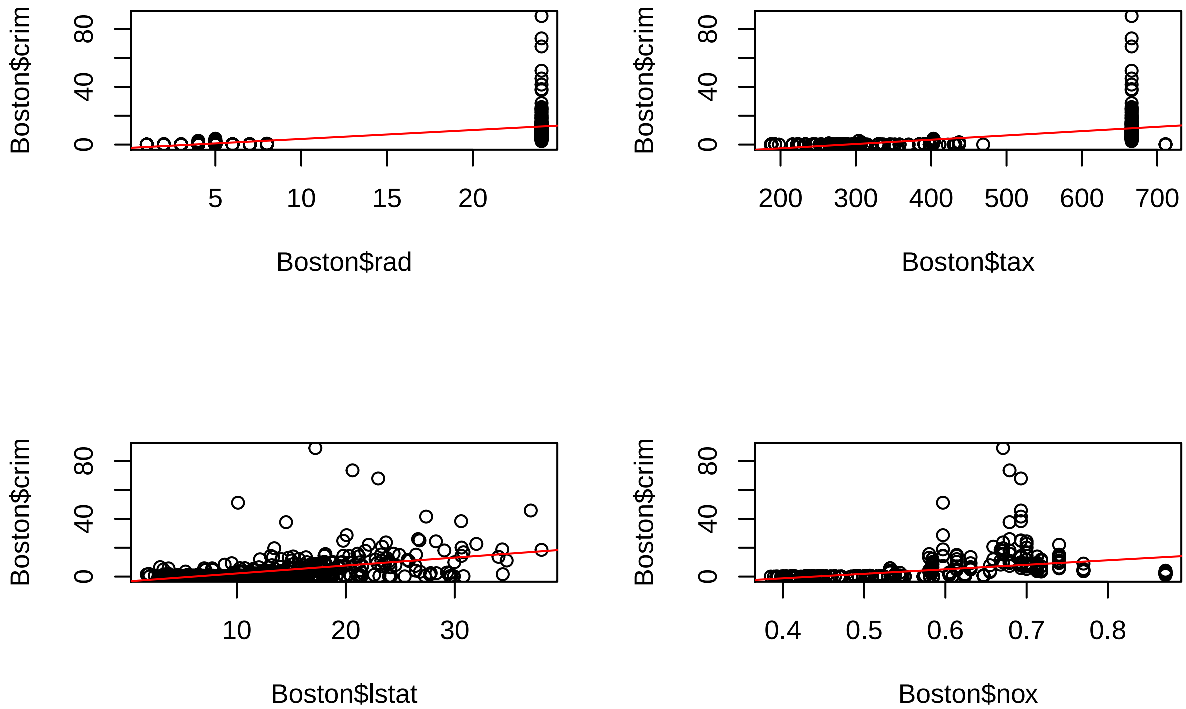


Question 13 Part g

There is no evidence that adding a quadratic term improves the model. The first clue is that the p value for the quadratic term is nonsignificant, which means it does not have a significant impact on the model estimations. The R-squared also dropped compared to the linear model, which is also indicative of a less accurate prediction.

Question 15 Part a

The predictors rad, tax, and lstat had R-square over 0.2 while nox was the only nonsignificant predictor. All other predictors had an R-square below 0.2. Looking the plots, even the highest R-square predictors are a seemingly poor choice, as a linear function does not seem to be the ideal model in these cases.



Question 15 Part b

In the multilinear regression model only a few predictors were significant for an alpha of 0.05, namely zn, dis, rad, black, and medv. For these predictors, we can reject the null hypothesis. All other predictors did not contribute to our model's prediction in a meaningful way.