MA575 C4 Team2

Yuhao Bai

Kiran Galani

Yiqin(Sabrina) Shi

Sixuan Wo

Kenji Lee

https://github.com/EricKenjiLee/LinearModelsLab.git

<http://jakevdp.github.io/blog/2015/07/23/learning-seattles-work-habits-from-bicycle-counts/>

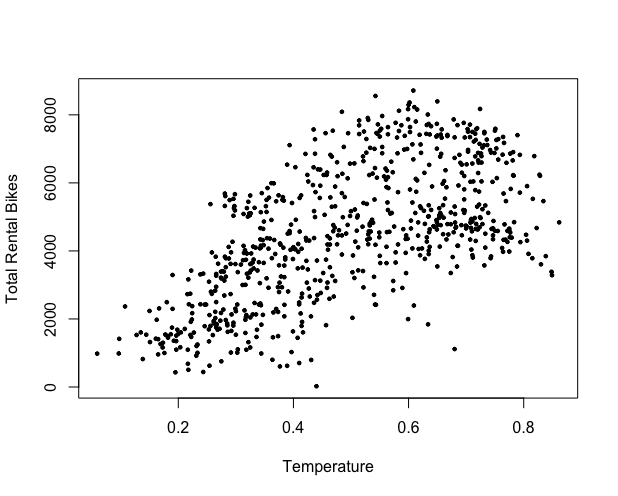
Biking Sharing Ordinary Least Squares

1.

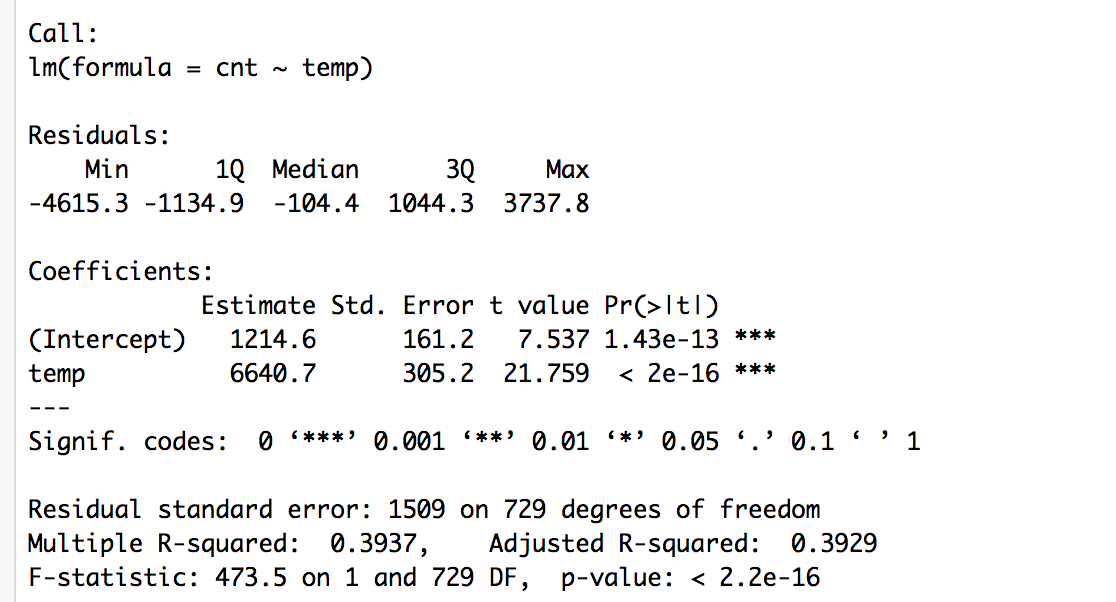
The response variable we choose is the count of total rental bikes (‘cnt’) which includes both casual and registered riders. The covariate variable we choose was the temperature (‘tmp’) which corresponds to the normalized temperature in Celsius (all values divided by 41 which was the max temperature).

We hypothesize that the relationship between the response variable and the covariate variable is linear based on our initial data exploration. When the temperature is low or very high, people will not choose to ride bikes and instead, they will take other forms of transportation. So, we predict that changes in temperature will influence the count of total rented bikes. Also, since the covariate variable is continuous, we think there won’t be a dramatic increase or decrease and overall the change will be linear. Therefore, we chose these two variables as our response and covariate variables.

2.



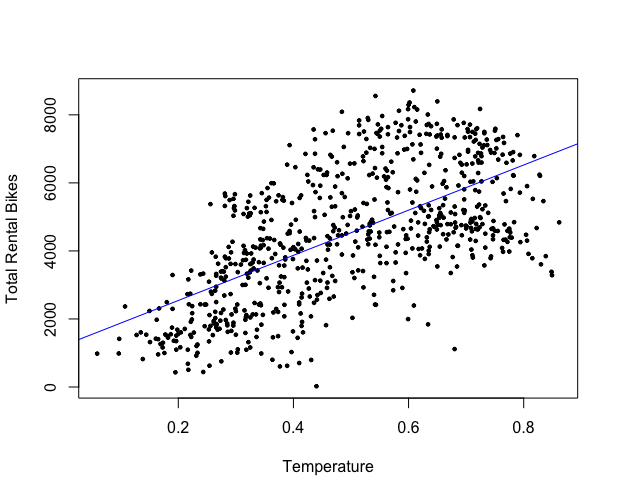
3 and 4.



According to the R summary, we know that the estimate for is 1214.6 and the estimate for is 6640.7. This means that when the normalized temperature is equal to 0 celcius, the number of total rented bikes would be 1214.6. For every unit increase in the normalized temperature, there is an increase of 6640.7 in the number of total rental bikes which corresponds to the slope .

The t-value for the intercept is 7.537. Because we are conducting the t-test at the 95% confidence level, we would reject the null hypothesis whenever the t-value exceeds 1.96. For the next estimated parameter, , the t-value was 21.759. As we are again conducting the test at the 95% confidence level, we would reject the null hypothesis when the t-value exceeds 1.96. In this experiment, we reject the null hypotheses for both parameters as the p-value for the intercept and linear slope are both less than . The p-values were vanishingly small and at the numerical limit with and correspondingly at 1.43e-13 and 2e-16. When rejecting the null hypothesis , we are stating that we have sufficient evidence to show that there is a significant linear relationship between the response and the covariate. This is the same as stating there is a less than 5% chance to observe t-values at least as extreme as shown given that the null hypothesis is true.

5.



It is a good model as it fits the overall trend and the R-squared score is 0.3937 (adjusted R-squared is 0.3929). However, this line does not explain all the data completely as it seems a quadratic fit might be better; this is more in line with our theory that ridership lowers for both low *and* very high temperatures. There weren’t any obvious outliers in the data either to compensate for.