

1. Download the bike share files. Each row in the dataset represents features measured from a single day. We will use columns 3:13 to predict one of the last 3 columns (casual daily users, registered daily users, total users). You can choose which response variable you like from the last 3 columns.
2. Import the data into matlab and separate the explanatory variables ( $x$ 's) from the response variable ( $y$ ). You should ignore the other possible response variables from the last 3 columns. Feel free to call the two variable arrays anything you like.
3. Use the `fitlm` function on the data to produce a linear regression.
4. Use the regression to predict the number of users (whichever of the 3 you chose) for a day with made up  $x$ -values for a bad weather day and made up  $x$ -values for a nice day.
5. Make a residual plot for your model. Does it seem like model assumptions are being met?
6. Make a 95% confidence interval for each of the betas. Taking the intervals into account along with the  $p$ -values in the last column, which variables do you think are predicting your  $y$  well?
7. Make pairwise plots for each of the  $x$ -variables. Do any of them seem to be correlated? If so, you will need to discard one.
8. Using all the information you have so far, select a subset of uncorrelated explanatory variables ( $x$ 's) that you think are good predictors for your response variable. Repeat the `fitlm` with your subset of variables to produce a second linear regression. How have the fit values like RSE, RSS,  $R^2$ ,  $F$ -test  $p$ -value changed? Has the regression improved from the first one? How are the individual beta  $p$ -values different?
9. Make a residual plot with the new regression. Does it seem like assumptions are being met?
10. Repeat the prediction in (4.). What do you think of the results?