

## **Multiple Linear Regression**

- 1) Which variables have the most explanatory power? Which have the least?
- 2) Remove some of the outlier countries, how does this effect your model?
- 3) Log-scale each of the variables, how does this change your model? Does it improve the models predictive power? How can you tell?
- 4) Can you think of any other modeling techniques (from class) that could be used instead of linear regression? Try using one of these and explain your results, with diagrams and if possible, a visualization as well as descriptive statistics
- 5) Think about how this model might be improved by adding more data. Then add this data to the model and test your hypothesis. What did you find. Provide descriptive statistics and visualizations as well as a few paragraphs explaining how you chose what data you did and why.
- 6) Using the model and data discussed in class predict how many cases a set of "new countries" would have (data to be provided in a separate csv file) Provide visualizations and a few paragraphs explaining your results.
- 7) Try other models discussed from class. What do these models predict and how do they differ from the linear regression model?
- 8) Now remove the variables with the least explanatory power. Does your linear regression improve compared to the other models? Does it do worse? Why? Please provide visuals and a few paragraphs of explanation
- 9) Now add in the extra data you found. Does your linear regression improved compared to the other models? Does it do worse? Why? Please provide visuals and a few paragraphs of explanation

<http://www.internetworldstats.com/>

<http://data.worldbank.org/indicator/IT.NET.USER.P2/countries>

--sources of internet usage

<http://www.internetlivestats.com/internet-users/>

--number of connected devices

10) download (or scrape) data from the above websites.

11) How much explanatory power does the model gain by adding the amount of internet penetration in a given country? How much does adding the total number of connected devices add?

12) Can you give an explanation of why or why not this does not add to the model's explanatory power? Is there another variable you might take away that is related to these variables?