Applying Linear Regression Models: Takeaways

Syntax

- Assessing the correlation of all the continuous variables in a `pandas` DataFrame:

```
df.corr()
```

- Using the logarithm function (base 2):

```
log2(y)
```

- The Box-Cox Transformation for giving the outcome a more bell-like distribution:

```
boxcox(y, lmbda = 0)
```

Concepts

- Feature selection is the process of choosing predictors/features for a model to improve its predictive ability.
  - One way to do so is to choose predictors that have high correlation with the outcome, or have different distributions of the outcome by categories.
  - Another way is to make more complex predictors, such as polynomial terms, splines, or interaction terms.
- We may want to transform the outcome so that extreme values are attenuated and the outcome has a more balanced distribution.
  - One way to do this is through a logarithmic transformation.
  - Another way to do this is through the Box-Cox transformation.