

We are going to build up to a k-fold cross validation comparison between two simple models. We will build two models to predict the variable `cnt` from the `bike` dataset found on my website.

1. Install the library `caret`.
2. Read in the `bike` dataset from my website. Read the help file associated with this dataset.
3. Write yourself a Mean Squared Error function. This function should have signature `MSE(y, yhat)` and should return a single number:

$$\text{MSE} = N^{-1} \sum_{n=1}^N (y_n - \hat{y}_n)^2.$$

4. Calculate the mean of the variable `cnt`. This is your first predictive model; a model that always predicts the mean. Let's refer to this value as `yhat`.
5. Call your function `MSE` on `cnt`, `y`, and your first model's predicted value, `yhat`.
6. Calculate ANOVA on an appropriately recognized categorical variable of your choice. Without defining "better", better choices should lead to smaller MSE.
7. Use the function `predict` to calculate `yhat` from this ANOVA model.
8. Call your function `MSE` on `cnt`, `y`, and your ANOVA model's predicted value, `yhat`.
9. Which MSE is smaller?
10. Are these MSEs reasonable quantifications of your model's ability to predict future (not your current) data? Why or why not?