# K-Folds 3

*Grant Esparza*

*November 30, 2018*

## Read data

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
bikes <- read.csv("https://roualdes.us/data/bike.csv")
```

## Create Mean Squared Error function

```
MSE <- function(y, yhat) {
  mean((y - yhat)^2)
}
```

## Call create folds

```
folds <- createFolds(bikes$cnt)
```

## Create two vectors to hold results

```
mse_mr01 <- rep(NA, 10)
mse_mr02 <- rep(NA, 10)
```

## Use a loop to do k-folds calculation

```
for(fold in 1:length(folds)) {
  training <- bikes[-folds[[fold]],]
  testing <- bikes[folds[[fold]],]
  mod01 <- lm(cnt ~ as.factor(season) + temp + as.factor(season):temp, data = bikes)
  mod02 <- lm(cnt ~ workingday + as.factor(season) + as.factor(season):workingday, data = bikes)
  mse_mr01[fold] <- MSE(testing$cnt, predict(mod01, newdata=testing))
  mse_mr02[fold] <- MSE(testing$cnt, predict(mod02, newdata=testing))
}
```

## Compare the mean of the Mean Squared Errors

```
mean(mse_mr01)
```

```
## [1] 1956550
```

```
mean(mse_mr02)
```

```
## [1] 2425596
```

### Take Away

Here we can see that the first model has a smaller MSE. The MSE allows us to compare two models but does little to tell us how good they actually are on their own. Therefore I would be cautious when using the MSE to declare a model as a good one. Here we can only say which one is worse.