MSE (training set):

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$

MSE (test set):

denote a test set obs. by (x_0, y_0)

~~If we had a large # of test set obs, could calculate~~

~~Average $\{y_0 - \hat{f}(x_0)\}^2$~~

~~average over many test set observations~~

Expected test MSE (at a particular value x_0)

$$E\left[\{y_0 - \hat{f}(x_0)\}^2\right] = Var\{\hat{f}(x_0)\} + \left[Bias\{\hat{f}(x_0)\}\right]^2 + Var(\varepsilon)$$

Expected test MSE:

Expected / Average squared error across
- all possible training sets / samples from the population
- all possible values of y_0 that could "go with" x_0

how much could $\hat{f}(x_0)$ change if we used a different training set?
↳ more flexible method → higher variance

on average, how different is $\hat{f}(x_0)$ from the "true" $f(x_0)$?
↳ more flexible method → lower bias

not in our control

Want a method w/ low MSE: low variance & low bias

More flexible methods → higher variance, lower bias

Less flexible methods → lower variance, higher bias

Where should we fall? Depends on how easy it is to find signal vs. noise

- sample size (larger n → more flexible method OK)
- # of predictors (larger p → more structured / less flexible method required)
- complexity of "true" f (more complex → more flexible method required)
- ~~depends on many...~~
- larger $Var(\varepsilon)$ (more structured / less flexible method required)