Let $\hat{\Theta}$ be an estimator for Θ.
Suppose $E(\hat{\Theta}) = \theta^*$ $\quad \theta - \theta^* =$ bias of $\hat{\Theta}$

$$MSE(\hat{\Theta}) = E[\hat{\Theta} - \Theta]^2 = E[\hat{\Theta} - \theta^* + \theta^* - \theta]^2$$

$$= E[\hat{\Theta} - \Theta]^2 + [\Theta^* - \Theta]^2 - 2(\theta^* - \theta)\underbrace{E[\hat{\Theta} - \theta^*]}_{\rightarrow 0}$$

$$= E[\hat{\Theta} - \theta]^2 + (\text{bias}(\hat{\Theta}))^2$$

$$= Var(\hat{\Theta}) + [\text{bias}(\hat{\Theta})]^2$$

If $\hat{\Theta}$ is an unbiased estimator then bias = 0.

$$MSE(\hat{\Theta}) = Var(\hat{\Theta})$$

---

unbiasedness is a "nice" property for an estimator

even better: unbiased + minimum variance among unbiased estimators

However, some bias & smaller variance than an unbiased estimator is also attractive.

Example: Ridge estimator for the "right" λ.