450A Political Methodology I
TA Section

Stanford University

September 30, 2016

Road Map

The purpose of TA section is to 1) review concepts from lecture, 2) extend proofs not covered in lecture, 3) provide guidance for `coding` homework.

Today we will cover

1. Introductions and housekeeping
2. Terminology
3. Properties of Estimators
4. Variance Proof
5. Mean Square Error Proof
6. Simulation
7. **R** examples

Who are you?

Who are we?

Plot Challenge!

Housekeeping

TA Information

Who?	Will Marble	Vincent Bauer
When?	Friday, 1:30-3:30	Tuesday, 10-noon
Where?	Encina 313	Encina 313

Housekeeping

TA Information		
Who?	Will Marble	Vincent Bauer
When?	Friday, 1:30-3:30	Tuesday, 10-noon
Where?	Encina 313	Encina 313

Students with documented disabilities

Students with documented disabilities can request academic accommodations through the Office of Accessible Education. Please let us know as soon as possible if you need accommodations for this class.

Terminology

Population, parameter, sample, and statistic.

Terminology

Population, parameter, sample, and statistic.

The population is the entire group that we are interested in, i.e. everyone in the US.

Terminology

Population, parameter, sample, and statistic.

The population is the entire group that we are interested in, i.e. everyone in the US. The parameter is the true information we want to know about the population, i.e. average support for Donald Trump.

Terminology

Population, parameter, sample, and statistic.

The population is the entire group that we are interested in, i.e. everyone in the US. The parameter is the true information we want to know about the population, i.e. average support for Donald Trump. If we had access to the entire population, we could just ask everyone their opinion on Trump and we would have no need for statistics.

Terminology

Population, parameter, sample, and statistic.

But we do not have enough time to interview everyone, and it turns out we can get a very close answer from just a small group of people, a sample.

Terminology

Population, parameter, sample, and statistic.

But we do not have enough time to interview everyone, and it turns out we can get a very close answer from just a small group of people, a sample. Each time we collect a sample, we estimate a new statistic for that sample, i.e. a particular guess about average support.

Terminology

Population, parameter, sample, and statistic.

But we do not have enough time to interview everyone, and it turns out we can get a very close answer from just a small group of people, a sample. Each time we collect a sample, we estimate a new statistic for that sample, i.e. a particular guess about average support. This guess will change slightly each time even if support stays the same because of sampling variability.

Terminology

- **Estimand:**
- **Estimator:**
- **Estimate:**

Terminology

- **Estimand:** The quantity in the population that we want to learn something about.
- **Estimator:**
- **Estimate:**

Terminology

- **Estimand:** The quantity in the population that we want to learn something about.
- **Estimator:** The function for estimating the estimand, we apply it to a sample from the population.
- **Estimate:**

Terminology

- **Estimand:** The quantity in the population that we want to learn something about.
- **Estimator:** The function for estimating the estimand, we apply it to a sample from the population.
- **Estimate:** The guess for the estimand produced by the estimator from a particular sample.

Terminology

Q: How is the expected value different from the mean value?

$$E[X] = \sum_{i=1}^{n} x_i p_i$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Terminology

Q: How is the expected value different from the mean value?

$$E[X] = \sum_{i=1}^{n} x_i p_i$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

The mean value is the actually observed average value of a particular sample.

Terminology

Q: How is the expected value different from the mean value?

$$E[X] = \sum_{i=1}^{n} x_i p_i$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

The mean value is the actually observed average value of a particular sample.
The expected value is the long-run average value you would expect to achieve from many iterations of the function.

Properties of Estimators

You are presented with a new estimator

$$\left(\frac{1}{n+5}\sum_{i=1}^{n} y_i\right) + \frac{3}{n}$$

and want to learn whether it is a good estimator to use, i.e. is it unbiased and/or consistent?

1. To learn whether an estimator is unbiased, we need to know its expectation.
2. To learn whether an estimator is consistent, we need to know the limit of its exectation and variance as n $\rightarrow \infty$.

Unbiasedness

Unbiasedness: Does $E[f(x)] = \mu$?

$$f(x) = \left(\frac{1}{n+5} \sum_{i=1}^{n} y_i \right) + \frac{3}{n}$$

Unbiasedness

Unbiasedness: Does $E[f(x)] = \mu$?

$$f(x) = \left(\frac{1}{n+5}\sum_{i=1}^{n} y_i\right) + \frac{3}{n}$$

$$E[f(x)] = E\left[(\frac{1}{n+5}\sum_{i=1}^{n} y_i) + \frac{3}{n}\right]$$

$$= E\left[(\frac{1}{n+5}\sum_{i=1}^{n} y_i)\right] + E\left[\frac{3}{n}\right]$$

Unbiasedness

Unbiasedness: Does $E[f(x)] = \mu$?

$$
\begin{aligned}
f(x) &= \left(\frac{1}{n+5} \sum_{i=1}^{n} y_i \right) + \frac{3}{n} \\
E[f(x)] &= E\left[\left(\frac{1}{n+5} \sum_{i=1}^{n} y_i \right) + \frac{3}{n} \right] \\
&= E\left[\left(\frac{1}{n+5} \sum_{i=1}^{n} y_i \right) \right] + E\left[\frac{3}{n} \right] \\
&= \frac{1}{n+5} * E\left[\sum_{i=1}^{n} y_i \right] + \frac{3}{n}
\end{aligned}
$$

Unbiasedness

Unbiasedness: Does $E[f(x)] = \mu$?

$$
\begin{aligned}
f(x) &= \left(\frac{1}{n+5}\sum_{i=1}^{n} y_i\right) + \frac{3}{n} \\
E[f(x)] &= E\left[(\frac{1}{n+5}\sum_{i=1}^{n} y_i) + \frac{3}{n}\right] \\
&= E\left[(\frac{1}{n+5}\sum_{i=1}^{n} y_i)\right] + E\left[\frac{3}{n}\right] \\
&= \frac{1}{n+5} * E\left[\sum_{i=1}^{n} y_i\right] + \frac{3}{n} \\
&= \frac{1}{n+5} * n\mu + \frac{3}{n} = \frac{n}{n+5}\mu + \frac{3}{n}
\end{aligned}
$$

Consistency

$$\text{Consistency} \begin{cases} (1) \lim_{n\to\infty} E[f(x)] = \mu \\ (2) \lim_{n\to\infty} Var[f(x)] = 0 \end{cases}$$

Consistency

$$\text{Consistency} \begin{cases} (1) \lim_{n\to\infty} E[f(x)] = \mu \\ (2) \lim_{n\to\infty} Var[f(x)] = 0 \end{cases}$$

(1)

$$\begin{aligned} \lim_{n\to\infty} E[f(x)] &= \lim_{n\to\infty} \frac{n}{n+5}\mu + \frac{3}{n} \\ &= \mu(!) \end{aligned}$$

Consistency

$$\text{Consistency} \begin{cases} (1) \lim_{n\to\infty} E[f(x)] = \mu \\ (2) \lim_{n\to\infty} Var[f(x)] = 0 \end{cases}$$

(1)

$$\begin{aligned} \lim_{n\to\infty} E[f(x)] &= \lim_{n\to\infty} \frac{n}{n+5}\mu + \frac{3}{n} \\ &= \mu(!) \end{aligned}$$

(2)

$$\begin{aligned} Var[f(x)] &= Var\left[(\frac{1}{n+5}\sum_{i=1}^{n} y_i) + \frac{3}{n}\right] \\ &= Var\left[\underbrace{\frac{1}{n+5}}_{constant}\sum_{i=1}^{n} y_i\right] + \underbrace{Var\left[\frac{3}{n}\right]}_{constant} \end{aligned}$$

Consistency

$$= \frac{1}{(n+5)^2} * Var\left[\sum_{i=1}^{n} y_i\right] + 0$$

$$= \frac{1}{(n+5)^2} * Var\left[y_1 + y_2 + ...y_n\right]$$

Consistency

$$= \frac{1}{(n+5)^2} * Var\left[\sum_{i=1}^{n} y_i\right] + 0$$

$$= \frac{1}{(n+5)^2} * Var\left[y_1 + y_2 + ...y_n\right]$$

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ... + Var[y_n] + Cov[y_1, y_2] +$$

$$Cov[y_2, y_1] + Cov[y_1, y_3] + ... + Cov[y_{n-1}, y_n] + Cov[y_n, y_{n-1}])$$

Consistency

$$= \frac{1}{(n+5)^2} * Var\left[\sum_{i=1}^{n} y_i\right] + 0$$

$$= \frac{1}{(n+5)^2} * Var\left[y_1 + y_2 + ...y_n\right]$$

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ... + Var[y_n] + Cov[y_1, y_2] +$$

$$Cov[y_2, y_1] + Cov[y_1, y_3] + ... + Cov[y_{n-1}, y_n] + Cov[y_n, y_{n-1}])$$

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ...Var[y_n] + \underbrace{2\sum_{i<j}^{n} Cov[y_i, y_j]}_{\text{zero by iid assumption}})$$

You will not receive full points if you do not show this step

Consistency

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ...Var[y_n])$$

Consistency

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ...Var[y_n])$$

$$= \frac{n\theta^2}{(n+5)^2}$$

Consistency

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ... Var[y_n])$$

$$= \frac{n\theta^2}{(n+5)^2}$$

Remember, we wanted to know whether $\lim_{n \to \infty} Var[f(x)] = 0$

Consistency

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ...Var[y_n])$$

$$= \frac{n\theta^2}{(n+5)^2}$$

Remember, we wanted to know whether $\lim_{n\to\infty} Var[f(x)] = 0$

$$\lim_{n\to\infty} Var[f(x)] = \lim_{n\to\infty} \frac{n\sigma^2}{(n+5)^2}$$

Consistency

$$= \frac{1}{(n+5)^2} * (Var[y_1] + Var[y_2] + ...Var[y_n])$$

$$= \frac{n\theta^2}{(n+5)^2}$$

Remember, we wanted to know whether $\lim_{n\to\infty} Var[f(x)] = 0$

$$\lim_{n\to\infty} Var[f(x)] = \lim_{n\to\infty} \frac{n\sigma^2}{(n+5)^2}$$

$$\approx \lim_{n\to\infty} \frac{1}{n}\sigma^2 = 0!$$

Summary

$$\text{Estimator: } \frac{1}{n+5}\sum_{i=1}^{n} y_i + \frac{3}{n}$$

Property	Criteria	Result
Unbiased	$E[f(x)] = \mu$	$\frac{n}{n+5}\mu + \frac{3}{n}$
Consistent	$\begin{cases} (1) \lim_{n\to\infty} E[f(x)] = \mu \\ (2) \lim_{n\to\infty} Var[f(x)] = 0 \end{cases}$	$\begin{cases} (1) = \mu \\ (2) = 0 \end{cases}$

Variance

Why does $Var[aX] = a^2 Var[X]$?

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$Var[aX] = E[(aX - E[aX])^2]$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$\begin{aligned} Var[aX] &= E[(aX - E[aX])^2] \\ &= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \end{aligned}$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$
\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\}
\end{aligned}
$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$
\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\} \\
&= E[a^2X^2] - E\{2aX\ \underbrace{E[aX]}_{constant}\} + E[aX]E[aX]
\end{aligned}
$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\} \\
&= E[a^2X^2] - E\{2aX\ \underbrace{E[aX]}_{constant}\} + E[aX]E[aX] \\
&= E[a^2X^2] - 2E[aX]E[aX] + E[aX]E[aX]
\end{aligned}$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$
\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\} \\
&= E[a^2X^2] - E\{2aX\ \underbrace{E[aX]}_{constant}\} + E[aX]E[aX] \\
&= E[a^2X^2] - 2E[aX]E[aX] + E[aX]E[aX] \\
&= E[a^2X^2] - E[aX]E[aX]
\end{aligned}
$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$
\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\} \\
&= E[a^2X^2] - E\{2aX\ \underbrace{E[aX]}_{constant}\} + E[aX]E[aX] \\
&= E[a^2X^2] - 2E[aX]E[aX] + E[aX]E[aX] \\
&= E[a^2X^2] - E[aX]E[aX] \\
&= a^2E[X^2] - a^2E[X]E[X]
\end{aligned}
$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$
\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\} \\
&= E[a^2X^2] - E\{2aX\underbrace{E[aX]}_{constant}\} + E[aX]E[aX] \\
&= E[a^2X^2] - 2E[aX]E[aX] + E[aX]E[aX] \\
&= E[a^2X^2] - E[aX]E[aX] \\
&= a^2E[X^2] - a^2E[X]E[X] \\
&= a^2(E[X^2] - E[X]^2) = a^2 Var[X]
\end{aligned}
$$

Variance

Why does $Var[aX] = a^2 Var[X]$?

$$
\begin{aligned}
Var[aX] &= E[(aX - E[aX])^2] \\
&= E\{a^2X^2 - 2aXE[aX] + E[aX]^2\} \\
&= E[a^2X^2] - E\{2aXE[aX]\} + E\{\underbrace{E[aX]E[aX]}_{constant}\} \\
&= E[a^2X^2] - E\{2aX\underbrace{E[aX]}_{constant}\} + E[aX]E[aX] \\
&= E[a^2X^2] - 2E[aX]E[aX] + E[aX]E[aX] \\
&= E[a^2X^2] - E[aX]E[aX] \\
&= a^2E[X^2] - a^2E[X]E[X] \\
&= a^2(E[X^2] - E[X]^2) = a^2 Var[X]
\end{aligned}
$$

$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abCov[X, Y]$

Mean Squared Error Criterion

When would we want to use a criterion such as MSE?

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

Sometimes we may want to :

1. Accept a little bias for an estimator that has less variance (any given estimate closer to the true parameter value)
2. Decide between several estimators with varying degrees of bias and variance.

Especially useful for selecting and weighting models in machine learning (450C).

Mean Squared Error statistic

Q: Why do we care about MSE, why not MAE?

$$MAE[\hat{\theta}] = E[|\hat{\theta} - \theta|]$$

1. Differentiable $\rightarrow$ analytic solutions
2. Higher weight to larger errors
3. Path dependency, other mathematics depend on MSE (i.e. OLS).
4. More efficient estimate of population parameter (Fischer 1920) *simulation

Q: Why does MSE $=$ Bias^2 $+$ Var?

Two different ways to derive the result, both equally valid but differ in complexity.

Deriving MSE: The Trick Way

We want to show that

$$MSE \equiv E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + Bias[\hat{\theta}]^2$$

Deriving MSE: The Trick Way

We want to show that

$$MSE \equiv E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + Bias[\hat{\theta}]^2$$

Start with the definition of variance:

$$Var[X] = E[X^2] - E[X]^2$$

Deriving MSE: The Trick Way

We want to show that

$$MSE \equiv E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + Bias[\hat{\theta}]^2$$

Start with the definition of variance:

$$Var[X] = E[X^2] - E[X]^2$$
$$X = (\hat{\theta} - \theta)$$

Deriving MSE: The Trick Way

We want to show that

$$MSE \equiv E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + Bias[\hat{\theta}]^2$$

Start with the definition of variance:

$$Var[X] = E[X^2] - E[X]^2$$
$$X = (\hat{\theta} - \theta)$$
$$Var[\hat{\theta} - \theta] = E[(\hat{\theta} - \theta)^2] - E[\hat{\theta} - \theta]^2$$

Deriving MSE: The Trick Way

We want to show that

$$MSE \equiv E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + Bias[\hat{\theta}]^2$$

Start with the definition of variance:

$$Var[X] = E[X^2] - E[X]^2$$
$$X = (\hat{\theta} - \theta)$$
$$Var[\hat{\theta} - \theta] = E[(\hat{\theta} - \theta)^2] - E[\hat{\theta} - \theta]^2$$
$$\underbrace{Var[\hat{\theta} - \theta]}_{Var[\hat{\theta}]} = \underbrace{E[(\hat{\theta} - \theta)^2]}_{MSE} - \underbrace{E[\hat{\theta} - \theta]^2}_{Bias^2}$$

Deriving MSE: The Algebra Way

We can also get the same result with much more algebra. It is useful to go through to force us to think about what terms are random variables and what terms are constants.

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

Deriving MSE: The Algebra Way

We can also get the same result with much more algebra. It is useful to go through to force us to think about what terms are random variables and what terms are constants.

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

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

Deriving MSE: The Algebra Way

We can also get the same result with much more algebra. It is useful to go through to force us to think about what terms are random variables and what terms are constants.

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

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

Deriving MSE: The Algebra Way

We can also get the same result with much more algebra. It is useful to go through to force us to think about what terms are random variables and what terms are constants.

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

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2 + 2\{(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)\} + (E[\hat{\theta}] - \theta)^2\right]$$

Deriving MSE: The Algebra Way

We can also get the same result with much more algebra. It is useful to go through to force us to think about what terms are random variables and what terms are constants.

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

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2 + 2\{(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)\} + (E[\hat{\theta}] - \theta)^2\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2E\left[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)\right] + E\left[(E[\hat{\theta}] - \theta)^2\right]$$

Deriving MSE: The Algebra Way

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2E\left[(\hat{\theta} - E[\hat{\theta}])\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\right] + E\left[\underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}\right]$$

Deriving MSE: The Algebra Way

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2E\left[(\hat{\theta} - E[\hat{\theta}])\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\right] + E\left[\underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\underbrace{E\left[\hat{\theta} - E[\hat{\theta}]\right]}_{=E[\hat{\theta}]-E[\hat{\theta}]=0} + \underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}$$

Deriving MSE: The Algebra Way

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2E\left[(\hat{\theta} - E[\hat{\theta}])\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\right] + E\left[\underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\underbrace{E\left[\hat{\theta} - E[\hat{\theta}]\right]}_{=E[\hat{\theta}]-E[\hat{\theta}]=0} + \underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + \underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}$$

Deriving MSE: The Algebra Way

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2E\left[(\hat{\theta} - E[\hat{\theta}])\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\right] + E\left[\underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}\right]$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + 2\underbrace{(E[\hat{\theta}] - \theta)}_{constant}\underbrace{E\left[\hat{\theta} - E[\hat{\theta}]\right]}_{=E[\hat{\theta}]-E[\hat{\theta}]=0} + \underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}$$

$$= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + \underbrace{(E[\hat{\theta}] - \theta)^2}_{constant}$$

$$= Var[\hat{\theta}] + Bias[\hat{\theta}]^2$$

Simulation: the Process

Here is the general process for a simulation

1. Identify (and look at) the population of interest, Y.
2. Specify your estimator, $\hat{\theta}(s)$, and code a function for it.
3. Pick how many simulations you want to run, m
4. Pick a sample size, n, for every simulation
5. Create an empty container vector of length m to hold estimates
6. Draw m separate samples, each of size n, from the population (with replacement)
7. For each sample, apply the estimator, $\hat{\theta}(s)$, to the sample which generates an estimate $\hat{\theta}_j$ for each sample.
8. Store each sample estimate $\hat{\theta}_j$ to your container vector. This is the empirical/simulated sampling distribution of $\hat{\theta}(s)$ for sample size n.
9. Analyze your sampling distribution: mean, variance, histogram, density plot, quantiles, etc

Simulation: in R

There are two general methods to complete steps 6 through 8

1. loops: iterate through steps 6 through 8 m number of independent times. This way is easier to conceptualize and program (at first) but more computationally expensive and easier to make mistakes.
2. apply: draw an n by m matrix from the population and then apply the function to the columns simultaneously. Much cleaner to code and much more efficient.