Lecture 12: Sampling Distributions & Standard Errors

Chapter 4.1

Goals for Today

Chapter 4: Arguably the most important chapter as it goes to the heart of statistical inference. Three important definitions:

1. point estimate
2. sampling distribution
3. standard error

Point Estimates

Behavior of Point Estimates

Ex: Say we draw samples of size $n = 100$ from a large population with $\mu = 5$ and $\sigma = 2$.

Behavior of Point Estimates

Ex: Say we draw samples of size $n = 100$ from a large population with $\mu = 5$ and $\sigma = 2$.

Two Important Questions:

Behavior of Point Estimates

Ex: Say we draw samples of size $n = 100$ from a large population with $\mu = 5$ and $\sigma = 2$.

Two Important Questions:

1. Is $\overline{x}$ going to be exactly 5?

Behavior of Point Estimates

Ex: Say we draw samples of size $n = 100$ from a large population with $\mu = 5$ and $\sigma = 2$.

Two Important Questions:

1. Is $\overline{x}$ going to be exactly 5?
2. Say we get $\overline{x} = 5.025$. If we repeat this procedure, will we get $\overline{x} = 5.025$ again?

Behavior of Point Estimates

Ex: Say we draw samples of size $n = 100$ from a large population with $\mu = 5$ and $\sigma = 2$.

Two Important Questions:

1. Is $\overline{x}$ going to be exactly 5?
2. Say we get $\overline{x} = 5.025$. If we repeat this procedure, will we get $\overline{x} = 5.025$ again?

We need to characterize the random error.

Behavior of Point Estimates

Let's repeat this procedure, say, 1000 times:

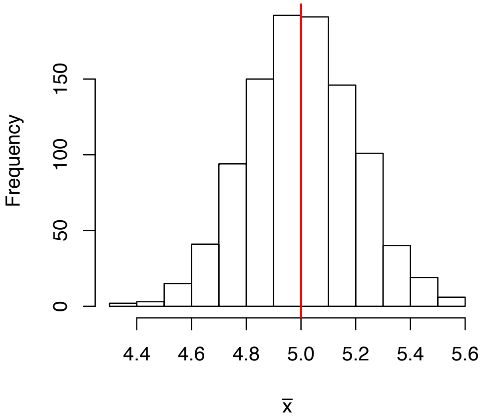
Behavior of Point Estimates

Let's repeat this procedure, say, 1000 times:

1st time	We get $\overline{x} = 4.831$
2nd time	We get $\overline{x} = 5.104$
3rd time	We get $\overline{x} = 4.965$
...	
1000th time	We get $\overline{x} = 4.957$

Sampling Distribution

This histogram is the 1000 instances of $\overline{x}$ i.e. the sampling distribution of $\overline{x}$:



Sampling Distributions

Sampling Distributions

We can define the sampling distributions for any point estimate, not just $\overline{x}$:

Sampling Distributions

We can define the sampling distributions for any point estimate, not just $\overline{x}$:

- s
- the sample median
- etc.

Sampling Distributions

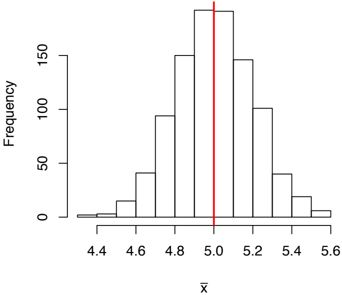
We can define the sampling distributions for any point estimate, not just $\overline{x}$:

- s
- the sample median
- etc.

We will only focus on sample means, including the sample proportion $\widehat{p}$.

Measure of Spread

What about spread? $[4.6, 5.4]$ contains roughly 95% of the data.

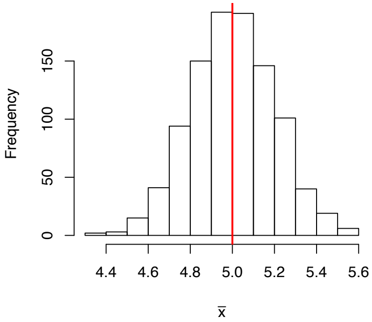


Standard Errors

Standard Error of $\overline{x}$

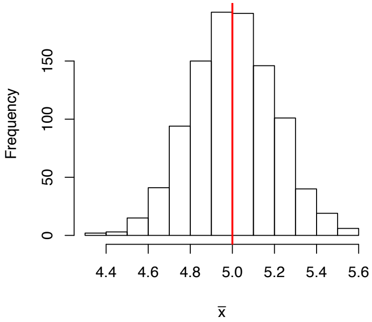
Back to Histogram

Samples were of size $n = 100$ with $\sigma = 2$. We estimated that the SD was 0.2.



Back to Histogram

Samples were of size $n = 100$ with $\sigma = 2$. We estimated that the SD was 0.2.



Using the formula:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \frac{2}{10} = 0.2$$

Standard Error of the Sample Mean $\overline{x}$

Compare 1000 instances of $\overline{x}$ when

Standard Error of the Sample Mean $\overline{x}$

Compare 1000 instances of $\overline{x}$ when

- $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$

Standard Error of the Sample Mean $\overline{x}$

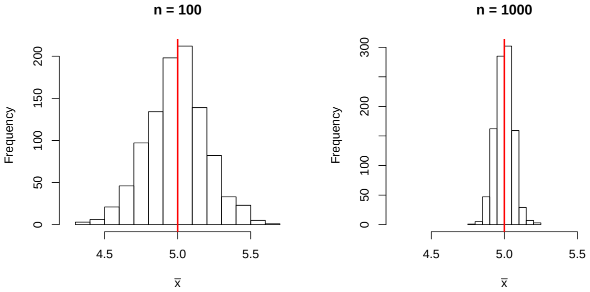
Compare 1000 instances of $\overline{x}$ when

- $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$
- $n = 1000$. $SE = \frac{2}{\sqrt{1000}} = 0.0632$. Smaller!

Standard Error of the Sample Mean $\overline{x}$

Compare 1000 instances of $\overline{x}$ when

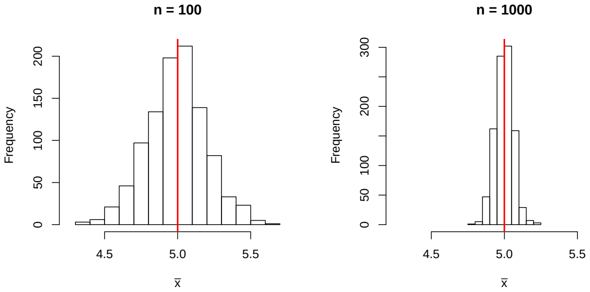
- $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$
- $n = 1000$. $SE = \frac{2}{\sqrt{1000}} = 0.0632$. Smaller!



Standard Error of the Sample Mean $\overline{x}$

Compare 1000 instances of $\overline{x}$ when

- $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$
- $n = 1000$. $SE = \frac{2}{\sqrt{1000}} = 0.0632$. Smaller!



Both are "accurate", but estimates on the right are "more precise."

Repeated Sampling

Standard Error of the Sample Mean

Population Distribution vs Sampling Distribution

Recap

- Point estimates are based on a sample $x_1, \ldots, x_n$ and are used to estimate population parameters.

Recap

- Point estimates are based on a sample $x_1, \ldots, x_n$ and are used to estimate population parameters.
- The sampling distribution characterizes the (random) behavior of point estimates.

Recap

- Point estimates are based on a sample $x_1, \ldots, x_n$ and are used to estimate population parameters.
- The sampling distribution characterizes the (random) behavior of point estimates.
- The standard deviation of a sampling distribution is the standard error: it quantifies the uncertainty/variability of point estimates.

Next Time

- Confidence Intervals
- When quoting survey results, what does: "the results of this survey are estimated to be accurate within 3.1 percentage points, 19 times out of 20" mean?
- Big One: Central Limit Theorem