

## Today's key words

- point estimate
- standard error
- sampling distribution

# Point estimates

- sample proportion
  - Each measurement is a 0 or 1.
  - 0 usually means “no” or “false” or “fail”.
  - 1 usually means “yes” or “true” or “success”.
  - The proportion is the average of the 0s and 1s.
- sample mean
  - Each measurement is a weight, height, mass, volume, count, etc...
  - The sample mean is the average of the measurements.

## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no.

## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no. What is your best guess for the proportion of BHCC students who like coconut water?

## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no. What is your best guess for the proportion of BHCC students who like coconut water?

$$\frac{5}{12} = 0.417$$

*Our point estimate is about 42% of BHCC students like coconut water.*

## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no. What is your best guess for the proportion of BHCC students who like coconut water?

$$\frac{5}{12} = 0.417$$

*Our point estimate is about 42% of BHCC students like coconut water.*

Would you be surprised if you asked 12 more random BHCC students and 7 said yes?

## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no. What is your best guess for the proportion of BHCC students who like coconut water?

$$\frac{5}{12} = 0.417$$

*Our point estimate is about 42% of BHCC students like coconut water.*

Would you be surprised if you asked 12 more random BHCC students and 7 said yes?

*No, there is natural fluctuation...*

## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no. What is your best guess for the proportion of BHCC students who like coconut water?

$$\frac{5}{12} = 0.417$$

*Our point estimate is about 42% of BHCC students like coconut water.*

Would you be surprised if you asked 12 more random BHCC students and 7 said yes?

*No, there is natural fluctuation...*

Based on all 24 students, what is the point estimate of the population proportion?



## Example of sample proportion

You ask 12 random BHCC students if they like coconut water. Only 5 say yes; the rest say no. What is your best guess for the proportion of BHCC students who like coconut water?

$$\frac{5}{12} = 0.417$$

*Our point estimate is about 42% of BHCC students like coconut water.*

Would you be surprised if you asked 12 more random BHCC students and 7 said yes?

*No, there is natural fluctuation...*

Based on all 24 students, what is the point estimate of the population proportion?

$$\frac{5 + 7}{24} = 0.5$$

*Our point estimate is about 50% of BHCC students like coconut water.*

Consider the probability distribution (infinite population) of rolling a fair 4-sided die.

$x$	1	2	3	4
$P(x)$	0.25	0.25	0.25	0.25

What is the expected value when rolling a 4-sided die?

Consider the probability distribution (infinite population) of rolling a fair 4-sided die.

$x$	1	2	3	4
$P(x)$	0.25	0.25	0.25	0.25

What is the expected value when rolling a 4-sided die?

$$\mu = (1)(0.25) + (2)(0.25) + (3)(0.25) + (4)(0.25) = \boxed{2.5}$$

Let's sample from this population.

Consider the probability distribution (infinite population) of rolling a fair 4-sided die.

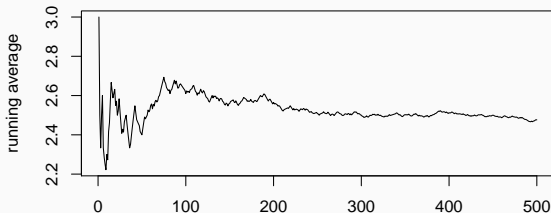
$x$	1	2	3	4
$P(x)$	0.25	0.25	0.25	0.25

What is the expected value when rolling a 4-sided die?

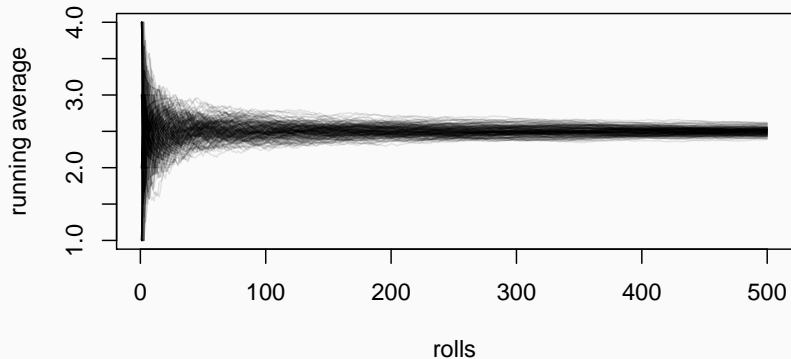
$$\mu = (1)(0.25) + (2)(0.25) + (3)(0.25) + (4)(0.25) = \boxed{2.5}$$

Let's sample from this population.

**The point estimate approaches the expected value.**



## Overlay of many Running Averages

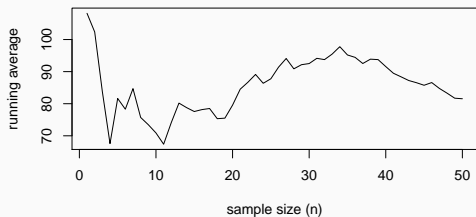


Notice the uncertainty gets smaller with larger sample size. However, there are diminishing returns...

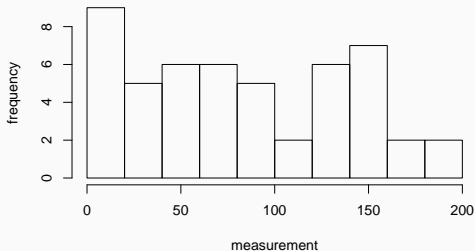
The accuracy improves drastically from  $n = 1$  to  $n = 100$ , but not nearly as drastically from  $n = 401$  to  $n = 500$ .

Now, imagine we sample from a new population/distribution, but we don't know the population parameters. What can we conclude?

**How accurate is our point estimate?**

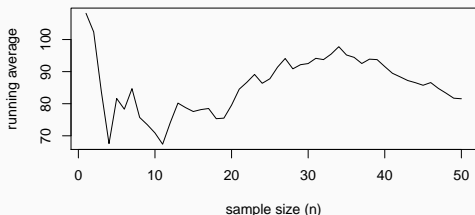


**Histogram of sample from unknown population**



Now, imagine we sample from a new population/distribution, but we don't know the population parameters. What can we conclude?

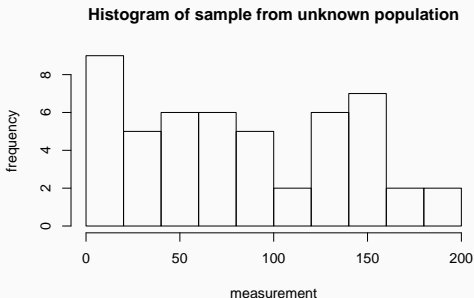
**How accurate is our point estimate?**



*Well, our point estimate is*

$$\mu \approx \bar{x} = 81.5$$

*However, we want to also describe our uncertainty. To me, based on the previous slide, I'd guess the uncertainty is about  $\pm 1/10$  of the range?*



$$\mu = 81.5 \pm 20$$

## Standard error

Standard error quantifies our uncertainty of a point estimate.

$$SE = \frac{\sigma}{\sqrt{n}}$$



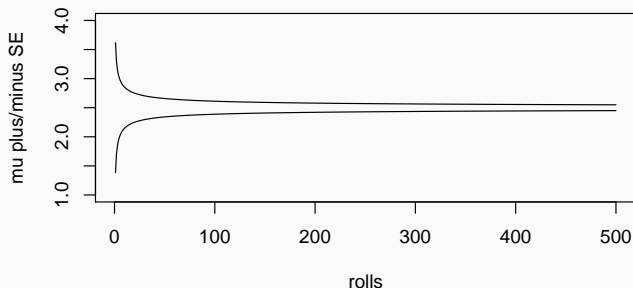
## Standard error

Standard error quantifies our uncertainty of a point estimate.

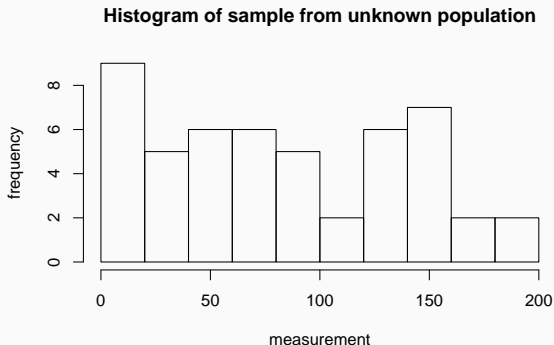
$$SE = \frac{\sigma}{\sqrt{n}}$$

Remember the 4-sided die. That distribution has  $\sigma = 1.118$ . We can plot  $\mu \pm SE$  as a function of  $n$ .

**The standard error decreases with more rolls.**

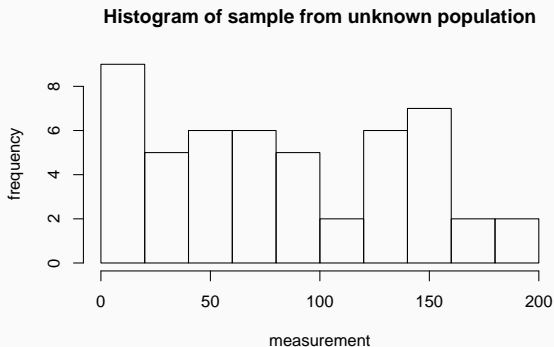


Let's return to our sample of 50 measurements from an unknown population.



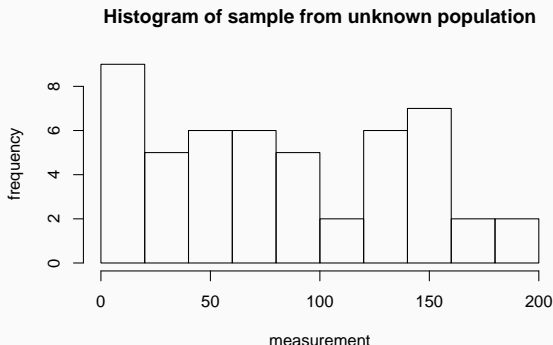
We do not know  $\sigma$ .

Let's return to our sample of 50 measurements from an unknown population.



We do not know  $\sigma$ . We can estimate  $\sigma$  from  $s$ .

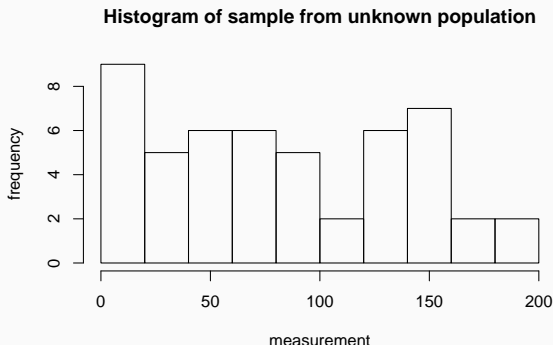
Let's return to our sample of 50 measurements from an unknown population.



We do not know  $\sigma$ . We can estimate  $\sigma$  from  $s$ . I calculated  $s = 55.75$ . Calculate the standard error.

$$SE \approx \frac{55.75}{\sqrt{50}} = 7.88$$

Let's return to our sample of 50 measurements from an unknown population.



We do not know  $\sigma$ . We can estimate  $\sigma$  from  $s$ . I calculated  $s = 55.75$ . Calculate the standard error.

$$SE \approx \frac{55.75}{\sqrt{50}} = 7.88$$

## Sampling Distributions

Let  $X_i$  be the  $i$ th draw from a population. Let  $n$  represent the number of draws. Let  $Y$  be the average of those draws.

$$Y = \frac{\sum_{i=1}^n X_i}{n}$$

By using the rules of Ch 2.4 we can show

$$\mu_Y = \mu_X$$

$$\sigma_Y = \frac{\sigma_X}{\sqrt{n}}$$

We say  $Y$  is determined by a sampling distribution. That sampling distribution has the same mean as the population, but it has a smaller standard deviation (and its SD shrinks as  $n$  increases). The SD of  $Y$  is the SE.

$$SE = \sigma_Y$$

# Sampling Distributions

Let  $X_i$  be the  $i$ th draw from a population. Let  $n$  represent the number of draws. Let  $\bar{X}$  be the average of those draws.

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

By using the rules of Ch 2.4 we can show

$$E(\bar{X}) = E(X)$$

$$SD(\bar{X}) = \frac{SD(X)}{\sqrt{n}}$$

The book also uses  $SD_{\bar{x}}$  to represent standard error.

$$SE = SD(\bar{X}) = SD_{\bar{x}}$$

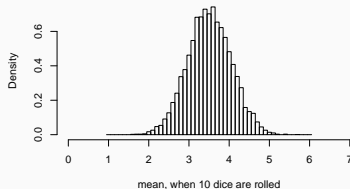
# Sampling Simulations

Let's roll 6-sided dice (on a computer to save time).

10000 repetitions of rolling 1 dice



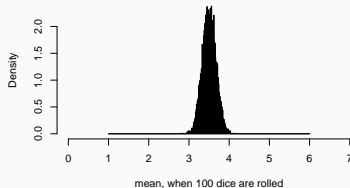
10000 repetitions of rolling 10 dice



10000 repetitions of rolling 2 dice



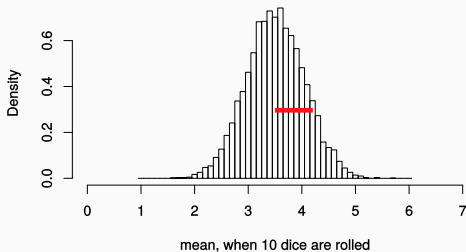
10000 repetitions of rolling 100 dice





# Practice

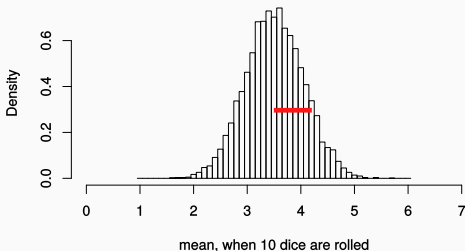
10000 repetitions of rolling 10 dice



Estimate the standard error when rolling 10 dice at a time.

## Practice

10000 repetitions of rolling 10 dice

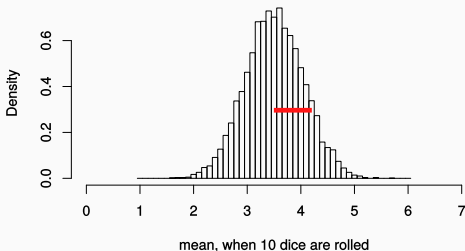


Estimate the standard error when rolling 10 dice at a time.

*Notice the sampling distribution looks nearly normal. I estimate that the standard deviation looks to be about **0.7**?*

# Practice

10000 repetitions of rolling 10 dice



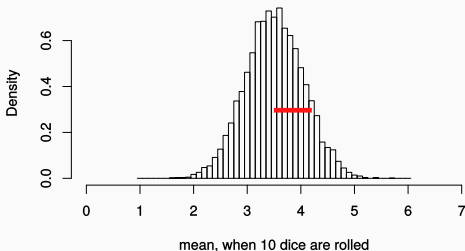
Estimate the standard error when rolling 10 dice at a time.

*Notice the sampling distribution looks nearly normal. I estimate that the standard deviation looks to be about **0.7**?*

I calculated that for rolling a single die,  $\sigma = 1.71$ . Calculate the standard error when rolling 10 dice.

## Practice

10000 repetitions of rolling 10 dice



Estimate the standard error when rolling 10 dice at a time.

*Notice the sampling distribution looks nearly normal. I estimate that the standard deviation looks to be about **0.7**?*

I calculated that for rolling a single die,  $\sigma = 1.71$ . Calculate the standard error when rolling 10 dice.

$$SE = \frac{1.71}{\sqrt{10}} = 0.54$$