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. What is your best guess for the proportion of BHCC students who like coconut water?

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

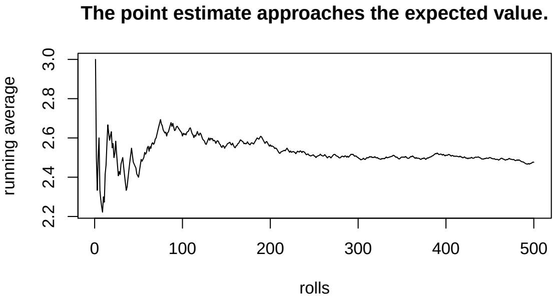
Based on all 24 students, what is the point estimate of the population proportion?

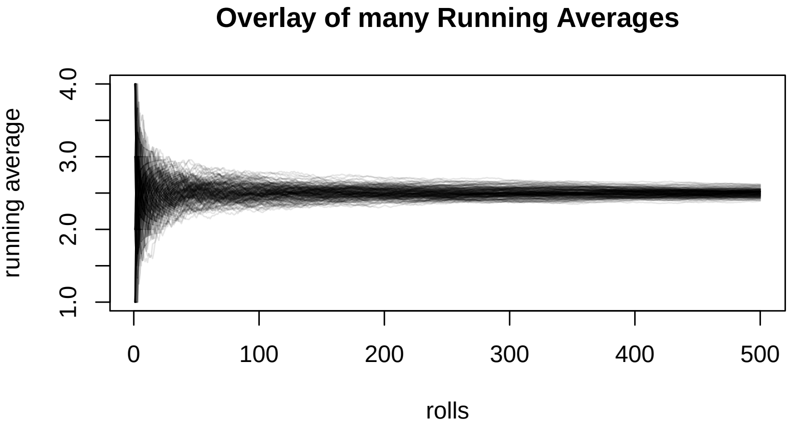
Consider the probability distribution (infinite population) of rolling a fair 4-sided die.

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?

Let's sample from this population.

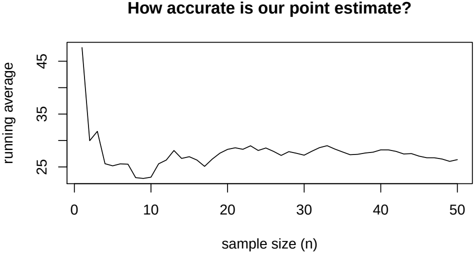


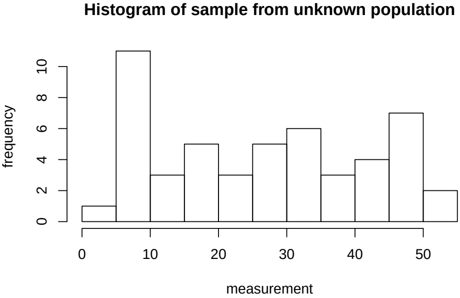


Notice the uncertainty gets smaller with larger sample size. However, there are diminishing returns...

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?



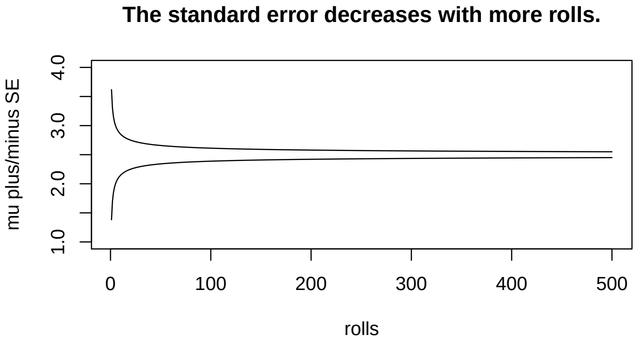


Standard error

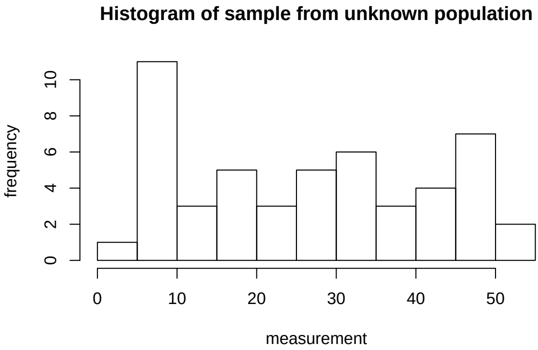
Standard error quantifies our uncertainty of a point estimate.

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

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



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



We do not know σ. We can estimate σ from s. I calculated $s = 15.5$.

$$SE \approx \frac{15.5}{\sqrt{50}} = 2.19$$

So we think our estimate $\mu \approx \bar{x} = 26.4$ has an "uncertainty" of 2.19.
But we need to define SE better...

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 ith 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

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

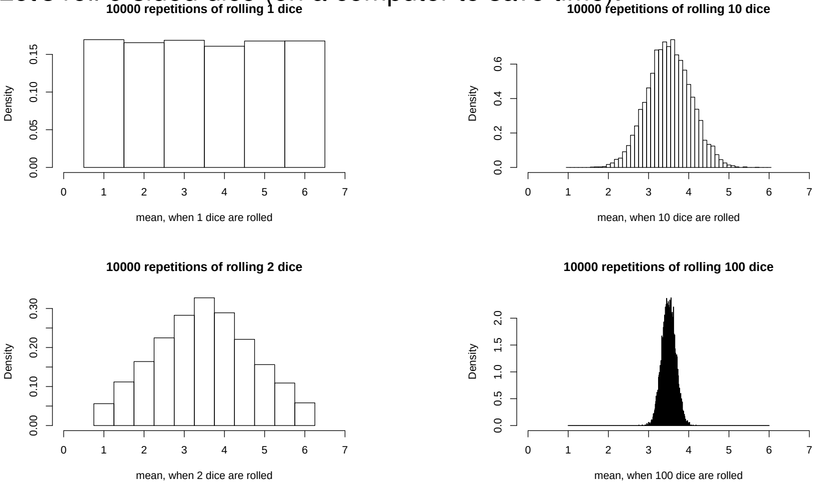
$$SD(\bar{X}) = \frac{SD(X)}{\sqrt{n}}$$

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

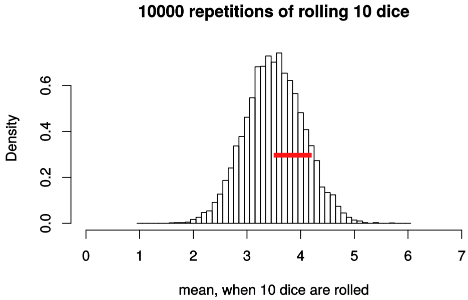
$$SE = SD(\bar{X}) = SD_{\bar{x}}$$

Sampling Simulations

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



Practice



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

I calculated that for rolling a single die, $\sigma = 2.92$. Calculate the standard error when rolling 10 dice.