81010

@msgunz3a

Chapter 7
Sampling Distributions

7.1 What is a sampling distribution?

Outcome: I will distinguish between a parameter and a statistic, use a sampling distribution of a statistic to evaluate a claim about a parameter, and distinguish among the distribution of a population and a sample.

X = # of hits out of 500 times at bat

Binomial

$n = 500$

$p = .26$

$.300(500)$

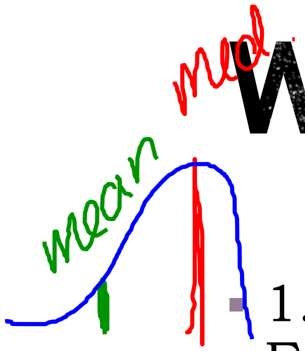
150

$P(X \geq 150) = .0246$

$\leq$

$1 - P(X \leq 149)$

Warm Up! (Recall Chapter 1)

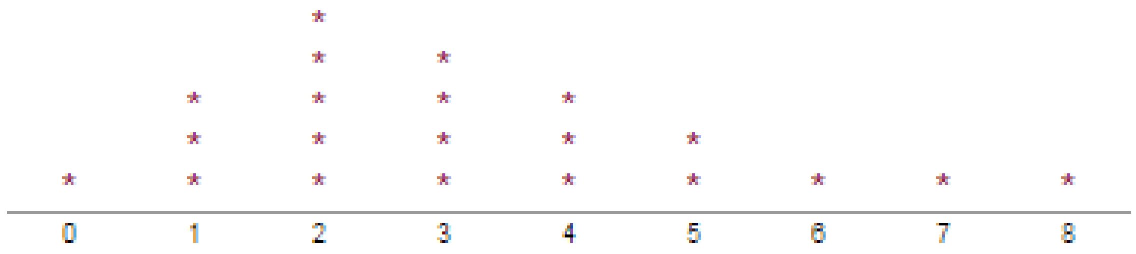


- 1. When a distribution is skewed to the left, is the mean or median larger? Explain.
- 2. The dotplot below shows the number of televisions owned by each family on a city block.

1.5(IQR)

$Q_3 = 4.5$

$Q_1 = 2$



Which of the following statements are true?

(A) The distribution is right-skewed with no outliers.

(B) The distribution is right-skewed with one outlier.

(C) The distribution is left-skewed with no outliers.

(D) The distribution is left-skewed with one outlier.

(E) The distribution is symmetric.

Parameter vs. Statistic

- On the AP Exam, many students lose points because they cannot distinguish between the two.

- **Parameter:** a number that describes some characteristic of the <u>population</u>
  - Think! **P** for **Population**
  - We use p to represent a population proportion
  - We use μ to represent population mean

- **Statistic**: a number that describes some characteristic of a <u>sample</u>
  - Think! **S** for **Sample**
  - We use $\hat{p}$ (p-hat) to represent a sample proportion (used to estimate the unknown parameter p)
  - We use $\bar{x}$ to represent sample mean

Example

Identify the population, the parameter, the sample, and the statistic in each of the following settings.

a.) The Gallup Poll asked a random sample of 515 U.S. adults whether or not they believe in ghosts. Of the respondents, 160 said "Yes".

Pop: All U.S. adults

Sam: 515 U.S. adults

p: Prop. of all U.S. Adults who said yes.

$\hat{p}$: $160/515 \approx 1/3$

b.) During the winter months, the temperatures outside of the Starnes' cabin in Colorado can stay well below freezing (32°F) for weeks at a time. To prevent the pipes from freezing, Mrs. Starnes sets the thermostat at 50°F. She wants to know how low the temperature actually gets in the cabin. A digital thermometer records the indoor temperature at 20 randomly chosen times during a given day. The minimum reading is 38°F.

Pop: temp. at all points throughout the day

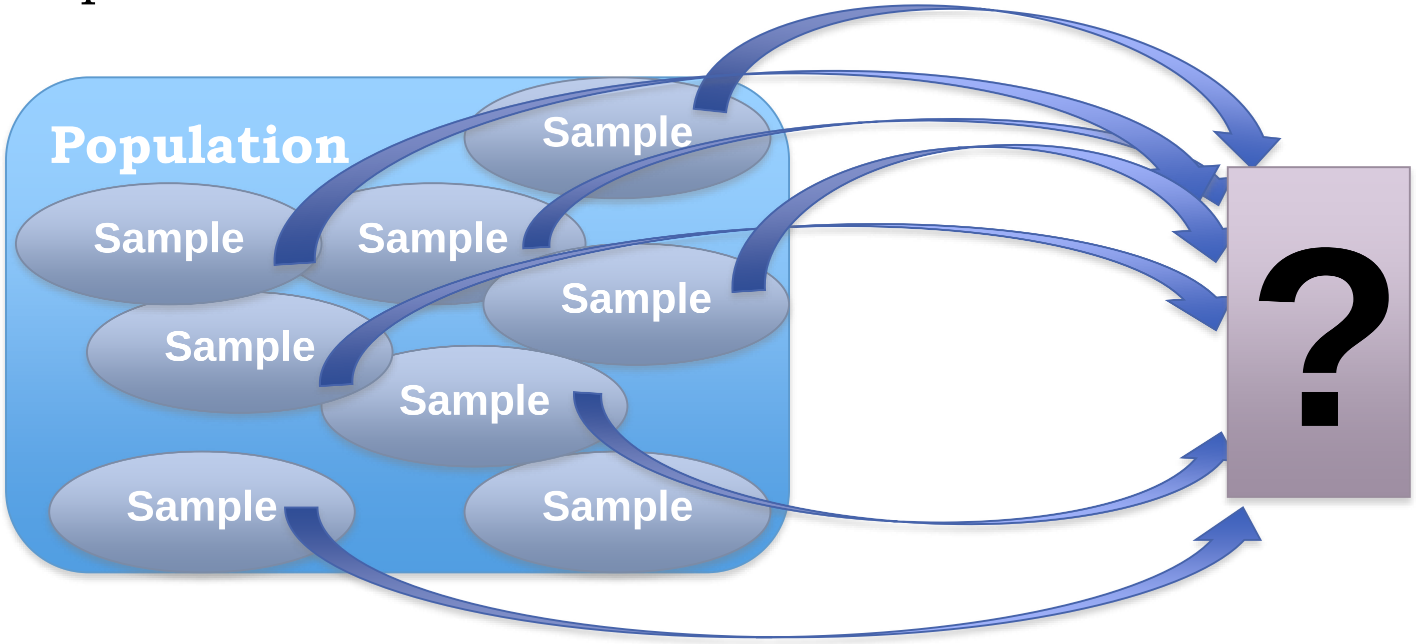
Sam: 20 chosen times

p: true minimum temp.

$\hat{p}$: 38°F

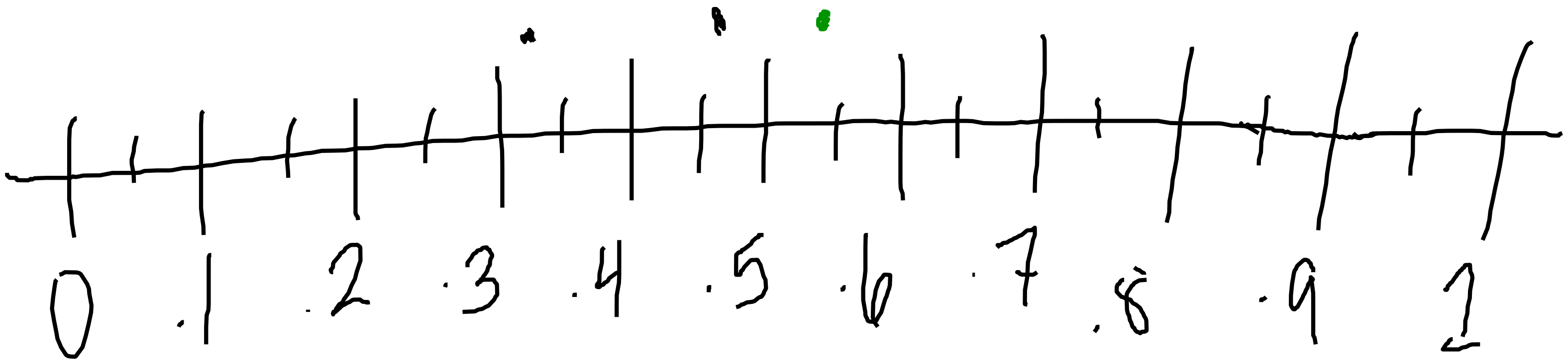
Sampling Variability

- How can $\bar{x}$ be an accurate representation of μ? After all, different random samples would produce different values of $\bar{x}$.
- This basic fact is called **sampling variability**: The value of a statistic varies in repeated random sampling.
- To make sense of sampling variability, we ask, "What would happen if we took many samples?"



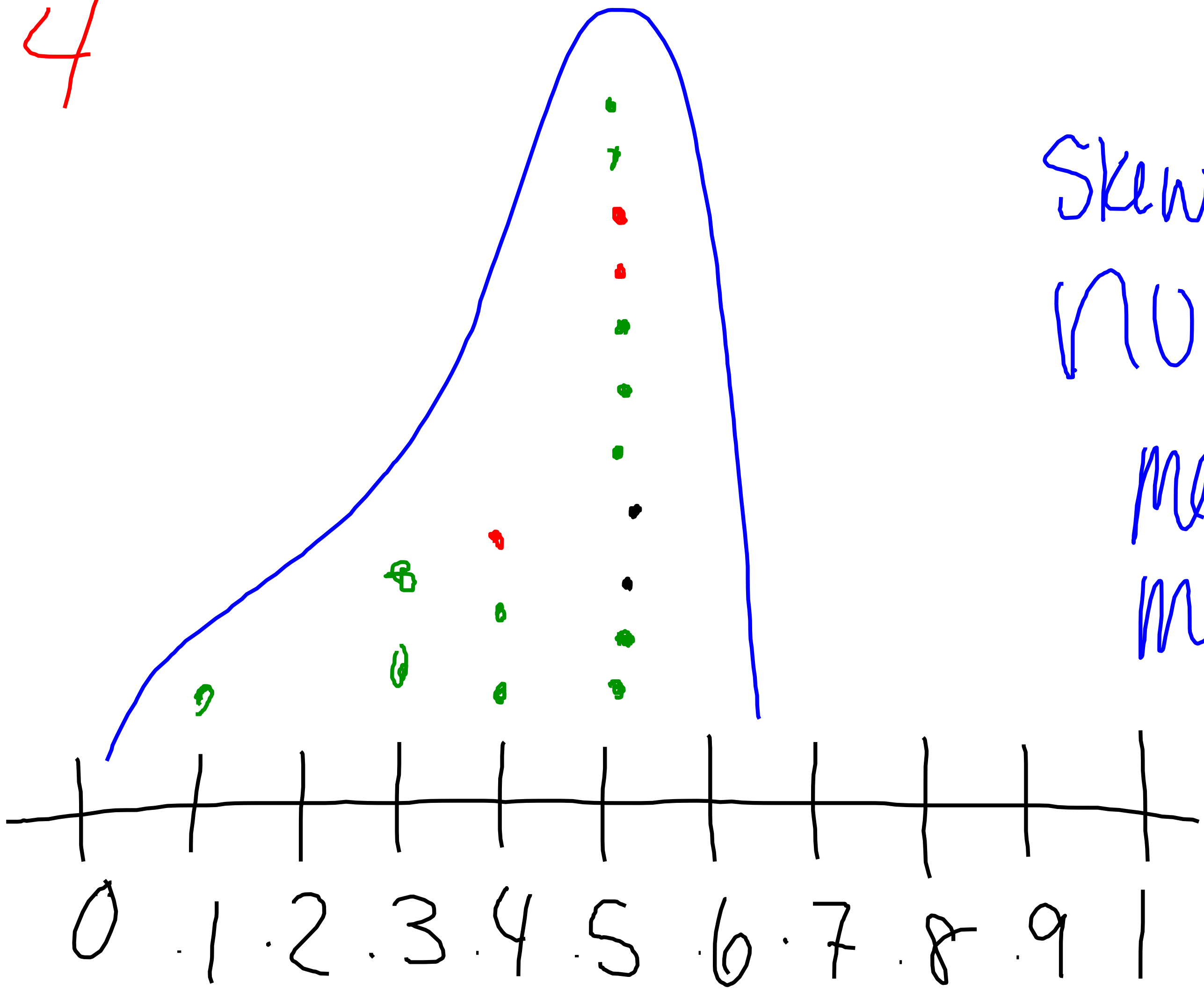
Reaching For Dice

- Each cluster will take a random sample of 15 dice from the bowl at the front of the room and note the sample proportion $\hat{p}$ of **green dice.**
- Each cluster will share the $\hat{p}$ value and the teacher will plot it on the class' dotplot.
- Repeat the process one more time (The more data we have, the better!)
- Describe what you see: shape, center, spread, and any outliers or unusual features.


0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1

SOCS

$$\frac{80}{200} = .4$$



Skewed left

no visible outliers

mean: .3-.4

mod: .5

Spread:

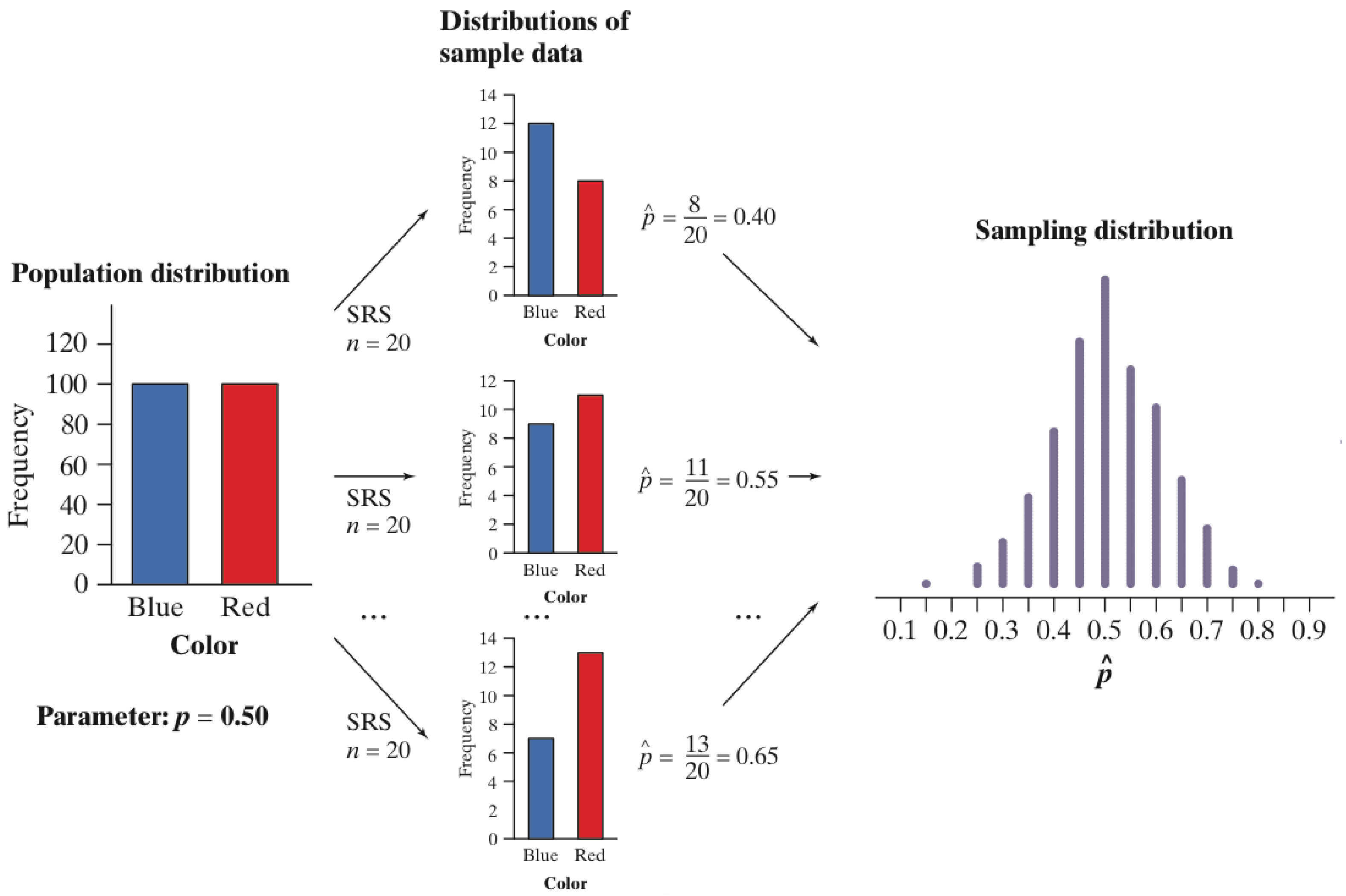
Sampling Distribution

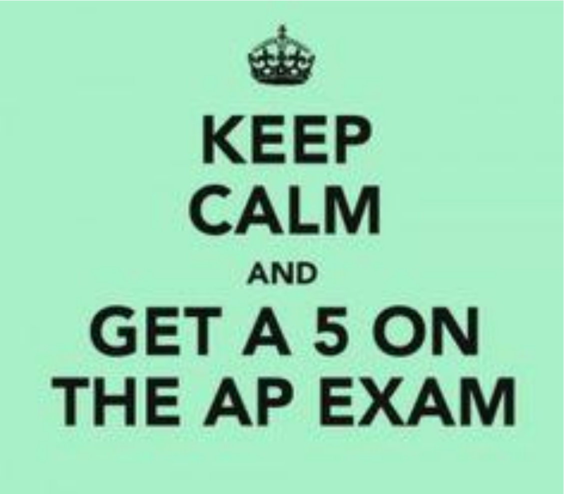
- If we took every one of the possible samples of size n from a population, calculated the sample proportion $\hat{p}$ for each, and graphed all of those values, we'd have a sampling distribution.

- **Sampling Distribution:** the distribution of values taken by the statistic in all possible samples of the same size n from the same population.

- In practice, it is usually difficult to take all possible samples of size n to obtain the actual sampling distribution of a statistic. Instead, we can use simulations to imitate the process of taking many, many samples.

Sampling Distribution vs. Population Distribution





- **Terminology matters.** Don't say "sample distribution" when you mean sampling distribution. You will lose credit on free response questions for misusing statistical terms.

Biased and Unbiased Estimators

- **CENTER**
- A statistic used to estimate a parameter is an **unbiased estimator** if the mean of its sampling distribution is equal to the value of the parameter being estimated.

  - BASICALLY: $p = \hat{p}$

  - Unbiased does not mean it is perfect. An unbiased estimator will almost always provide an estimate that is *not* equal to the value of the population parameter → It will most likely be VERY CLOSE.

Sampling Heights

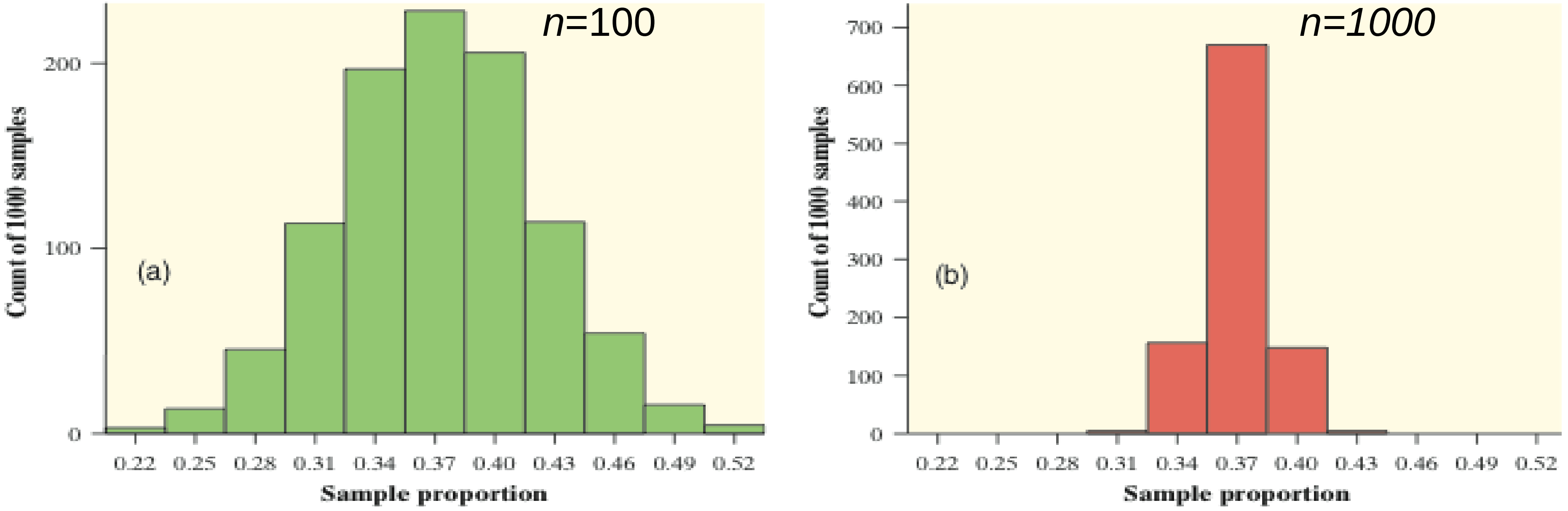
- Each student will write his or her height (In inches) neatly on a small piece of cardstock and then pass it forward.
- You will get the bag and take a sample of four cards and record the heights of the four students chosen. Then put the cards back and pass the bag to the next student.
- For your SRS, copy the table below and write in your data.

Height (in.)	Sample mean ($\overline{x}$)	Sample Range (max – min)

- Plot the values of you sample mean and sample range on the two class dot plots.
- Based on the two sampling distributions, which statistic appears to be an unbiased estimator?

Lower Variability is Better!

- **SPREAD**



- Larger samples have a clear advantage over smaller samples. They are much more likely to produce an estimate close to the true value of the parameter.

Variability of a Statistic

- Variability of a statistic is described by the spread of its sampling distribution.

- This spread is determined mainly by the size of the random sample.
  - Larger samples give smaller spreads.

Taking a larger sample will reduce variability of a statistic but it will not eliminate bias.

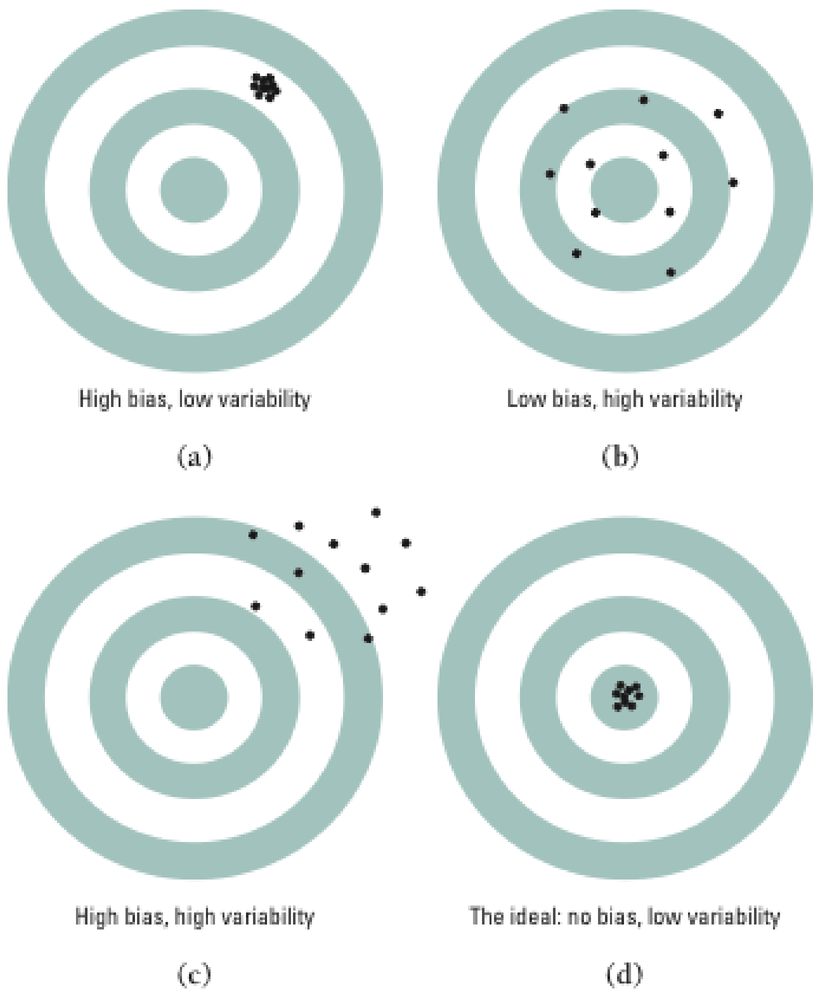
Bias, Variability, and Shape

- We can think of the true value of the population parameter as the bull's-eye on a target and of the sample statistic as an arrow fired at the target.

- Both bias and variability describes what happened when we take many shots at the target.

Bias means that our aim is off and we consistently miss the bull's-eye in the same direction. Our sample values do not center on the population value.

High **variability** means that repeated shots are widely scattered on the target. Repeated samples do not give very similar results.



Exit Ticket! FRQ for a Formative Grade

- Drop in the Green, Yellow or Red Folder when you finish.

- HOMEWORK: P. 436-437/ #6, 8, 10, 12, 14.