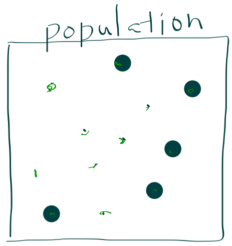
Sometimes important ideas seem simple and they get over looked

for that reason we'll start today by looking back at 3 big ideas.

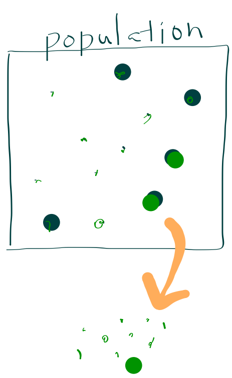
Big Idea #1



A parameter is a number that describes some characteristic of a population

[like the average cell phone screen time per week for all teachers]

μ



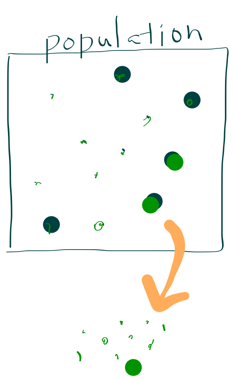
A parameter is a number that describes some characteristic of a population

[like the average cell phone screen time per week for all teachers]

μ

A statistic estimates the value of a parameter using a sample from the population $\overline{x}$

Big Idea #2



A parameter is a number that describes some characteristic of a population

[like the average cell phone screen time per week for all teachers]

μ

A statistic estimates the value of a parameter using a sample from the population $\bar{x}$

Statistics of samples vary

the mean number of cellphone screen time changes from sample to sample.

Statistics of samples vary

the mean number of cellphone screen time changes from sample to sample.

↓

which is why statistics have distributions (parameters do not)

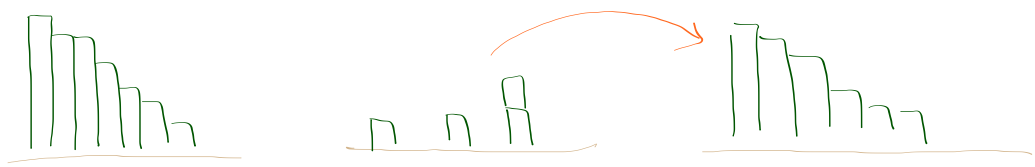
Big Idea #3

there is a difference between

Population distributions

distributions of a sample

Sampling distributions of a statistic

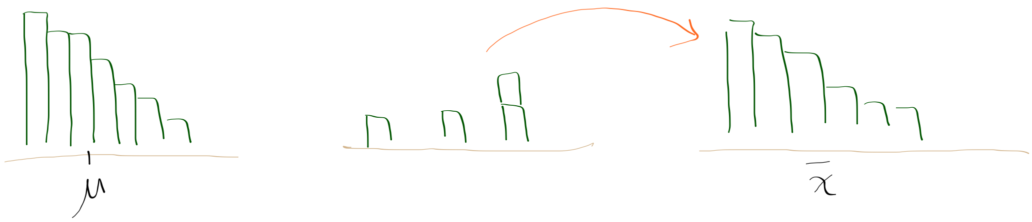


one of today's Big ideas

Population distributions

distributions of a sample

Sampling distributions of a statistic



if the mean of a sampling distribution is the same as the mean of the parameter being estimated, the statistic ($\bar{x}$) is called an unbiased estimator.

Use a sampling Distribution of a statistic to evaluate a claim about a parameter.

Determine if a statistic is an unbiased estimator of a population parameter.

Describe the relationship between sample size and the variability of a statistic

Use a sampling Distribution of a statistic to evaluate a claim about a parameter.

from an actual population at SHS

Determine if a statistic is an unbiased estimator of a population parameter.

Describe the relationship between sample size and the variability of a statistic

Lesson 7.1 (Day 2): What was the real Final Exam average?

Today, we will be taking a **sample** from a **population**. We will use the average from the **sample** to estimate the average for a real **population**. Yesterday we looked at a very small class of 4 students as the population. Today we will look at larger population, the combined final exam scores from three Algebra 2 classes at Sheldon High School. (actual scores)

Take a random sample of 5 students and record their scores. Then find the mean. Repeat this for a total of 4 times.

Scores:______________ Mean:______ Scores:______________ Mean:______

Scores:______________ Mean:______ Scores:______________ Mean:______

1. Write each mean on a different sticker and put the stickers in the appropriate location on the poster at the front of the room. Copy down the dotplot that is created on the poster.

1. Write each mean on a different sticker and put the stickers in the appropriate location on the poster at the front of the room. Copy down the dotplot that is created on the poster.

2. What does each dot on the poster represent?

3. What do you think the true final exam average is?

4. A **sampling distribution** shows the means calculated from all of the possible samples of size 5 from the population. Is the above dotplot a sampling distribution? Explain.

5. Suppose we took a random sample of 5 final exam scores from Churchill High School and got a mean of 68. Is this convincing evidence that Churchill students did worse than students at our school?

2. What does each dot on the poster represent?

One mean from a random sample of 5

3. What do you think the true final exam average is?

4. A **sampling distribution** shows the means calculated from all of the possible samples of size 5 from the population. Is the above dotplot a sampling distribution? Explain.

5. Suppose we took a random sample of 5 final exam scores from Churchill High School and got a mean of 68. Is this convincing evidence that Churchill students did worse than students at our school?

2. What does each dot on the poster represent?

One mean from a random sample of 5

3. What do you think the true final exam average is? $\mu =$

4. A **sampling distribution** shows the means calculated from all of the possible samples of size 5 from the population. Is the above dotplot a sampling distribution? Explain.

No. we didn't take all possible samples.

(That would be ${}_{74}C_5 = 16,108,764$ samples)

5. Suppose we took a random sample of 5 final exam scores from Churchill High School and got a mean of 35. Is this convincing evidence that Churchill students did worse than students at our school?

how many of our samples of 5 are 35 or lower?

$\frac{1}{52} \approx 1.9\%$ So, we do have ~~ea~~ evidence that CHS students did worse

Pick 3
Grading

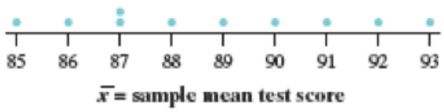
Pick Three Grading- Mrs. Perry Ameter, the teacher, has an interesting approach to assigning grades in her statistics class. Of the 5 tests students take throughout the semester, Mrs. Ameter selects a random sample of 3, finds the average score of these tests, and records this average as the student's final grade. Joe's test scores are as follows: **93, 87, 96, 78, 90**.

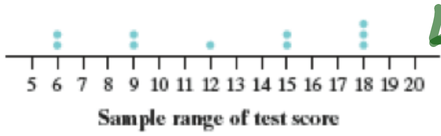
(a) List all 10 possible samples of size 3. (this is actually done for you on the right!)

93, 87, 96
93, 87, 78
93, 87, 90
93, 96, 78
93, 96, 90
93, 78, 90
87, 96, 78
87, 96, 90
87, 78, 90
96, 78, 90

(b) Calculate the mean of each sample and display the **sampling distribution of the sample mean** using a dotplot. (already done on the right).

And while you are at it, calculate the range of each sample and display the **sampling distribution of the sample range** using dot plot. (Has Mr. Cedarlund gone mad?.... All of this has also done for you below as well).





Sample	Sample mean	Sample range
93, 87, 96	92	9
93, 87, 78	86	15
93, 87, 90	90	6
93, 96, 78	89	18
93, 96, 90	93	6
93, 78, 90	87	15
87, 96, 78	87	18
87, 96, 90	91	9
87, 78, 90	85	12
96, 78, 90	88	18

c) Is the sample mean an **unbiased estimator** of the population mean? Explain your answer.

?

Population

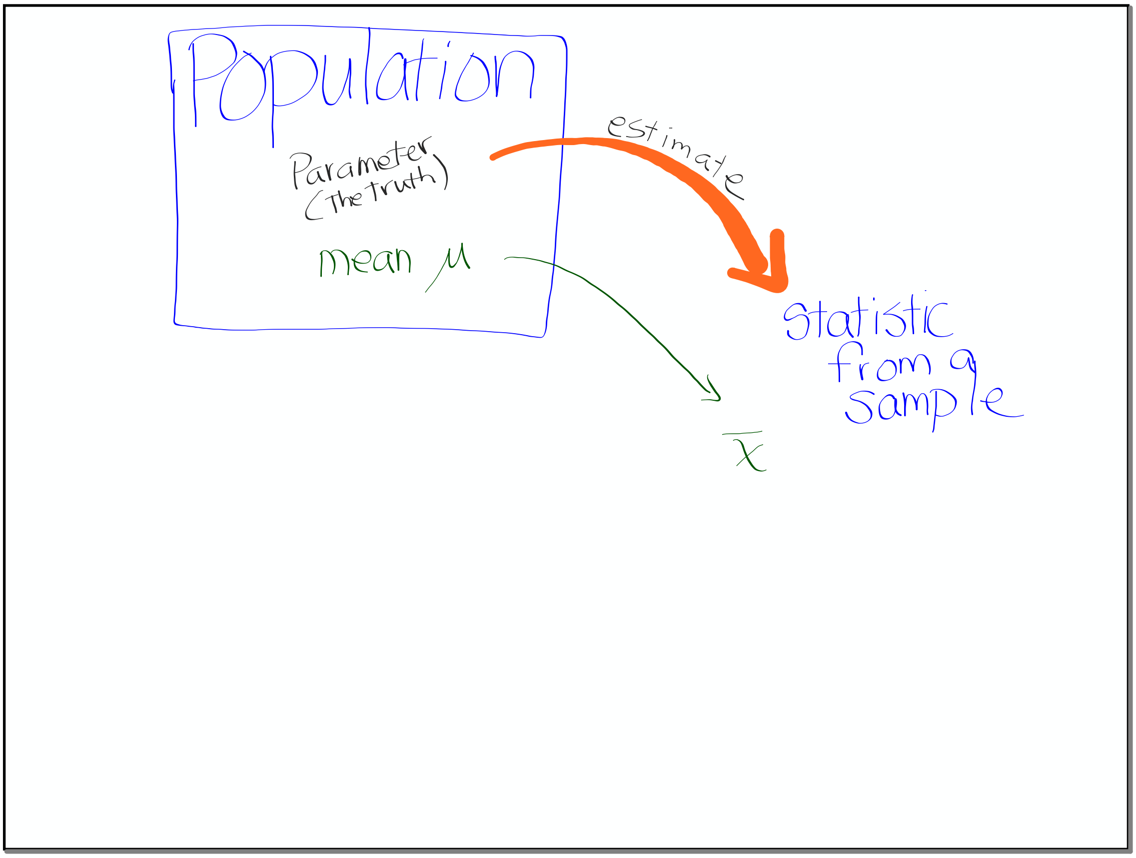
Parameter
(The truth)

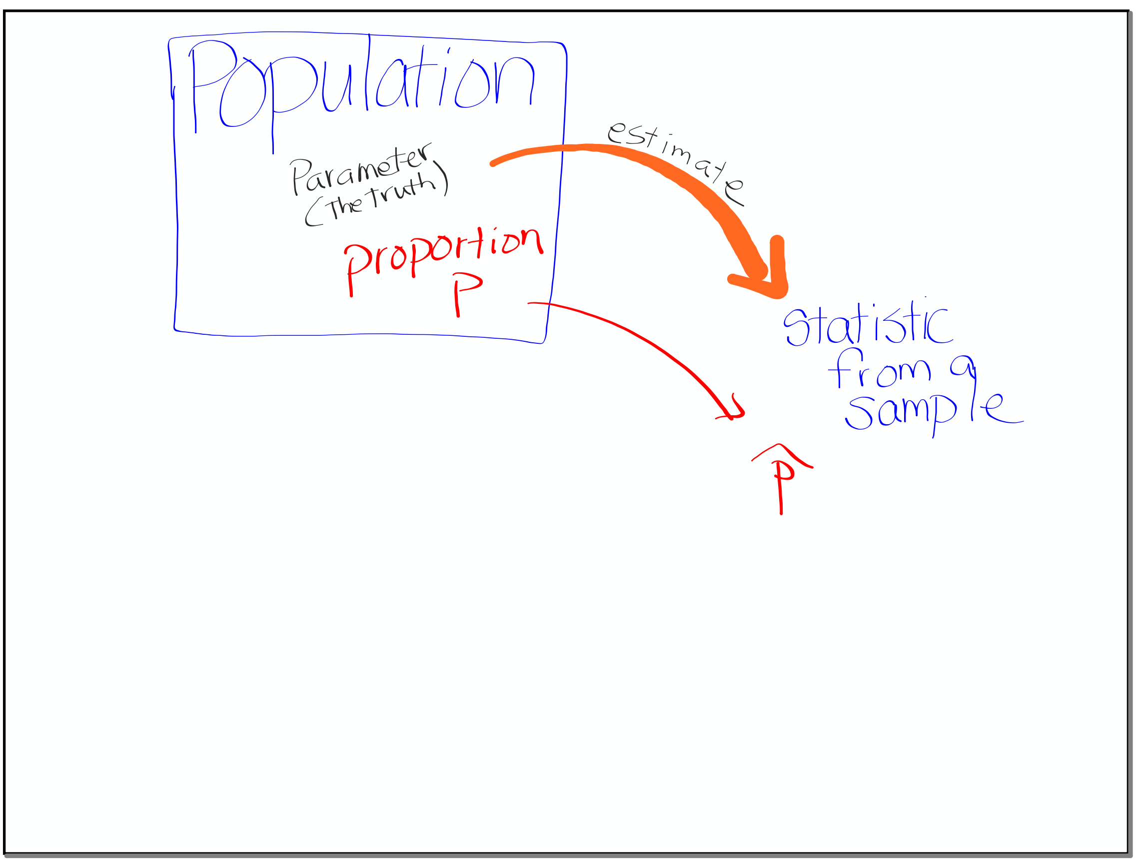
Population

Parameter
(The truth)

estimate

Statistic
from a
sample


Population
Parameter
(The truth)
mean μ
estimate
Statistic from a sample
$\bar{x}$


Population
Parameter
(The truth)
proportion
p
estimate
Statistic from a sample
$\hat{p}$

Population

Parameter
(The truth)

MAXIMUM

estimate

Statistic
from a
sample

maximum

Many things are estimators
some tend to over or underestimate

c) Is the sample mean an **unbiased estimator** of the population mean? Explain your answer.

mean of pop. $\mu = \dfrac{93+87+96+78+90}{5} = 88.8$

mean of sample distribut $\bar{x} = \dfrac{92+86+90+89+93+87+87+91+85+88}{10}$

$= 88.8$

...ics class. Of the 5 tests students take throughout ...
ple of 3, finds the average score of these tests, and
oe's test scores are as follows: **93, 87, 96, 78, 90.**

Sample	Sample mean	Sample rang
93, 87, 96	92	9
93, 87, 78	86	15
93, 87, 90	90	6
93, 96, 78	89	18
93, 96, 90	93	6
93, 78, 90	87	15
87, 96, 78	87	18
87, 96, 90	91	9
87, 78, 90	85	12
96, 78, 90	88	18

$96 - 78$
$= 18$

c) Is the sample mean an **unbiased estimator** of the population mean? Explain your answer.

mean of pop. $\mu = \frac{93+87+96+78+90}{5} = 88.8$

mean of sample distribut $\bar{x} = \frac{92+86+90+89+93+87+87+91+85+88}{10}$

$= 88.8$

Because they are the same, the sample mean is an unbiased estimator of the population mean.

Is the **range** an unbiased estimator of the **population range**?

NO, the population range is 18 but the sample range was 12.6.

Example

If we say $\hat{p}$ is an unbiased estimator of P,

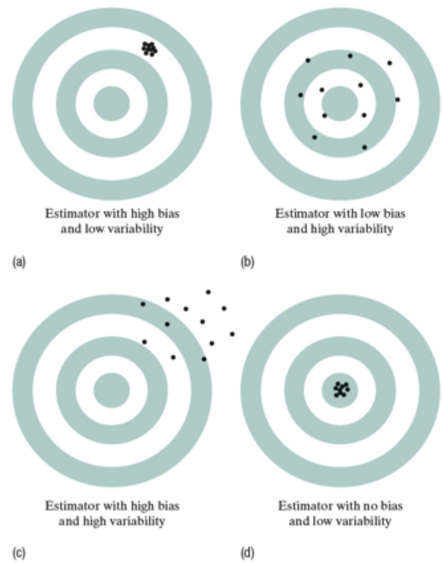
we assume that the value of $\hat{p}$ came from an SRS, not a convenience sample or voluntary response sample.

[likewise, No problems w/ Response bias, or NON-response]

To decide which estimator to use when there are several reasonable choices, consider both bias and variability.

Bias means that our sample statistics do not center on the population parameter. In other words, our estimates are not *accurate*.

High variability means that repeated samples do not give very similar results. In other words, our estimates are not very *precise*.



AP® Exam Tip

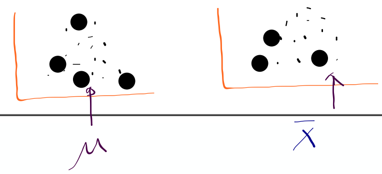
Make sure to understand the difference between **accuracy** and **precision** when writing responses on the AP® Statistics Exam. Many students use "**accurate**" when they really mean "**precise**."

For example, a response that says "increasing the sample size will make an estimate more **accurate**" is incorrect. It should say that increasing the sample size will make an estimate more **precise**. If you can't remember which term to use, don't use either of them. Instead, explain what you mean without using statistical vocabulary.

Biased and Unbiased Estimators

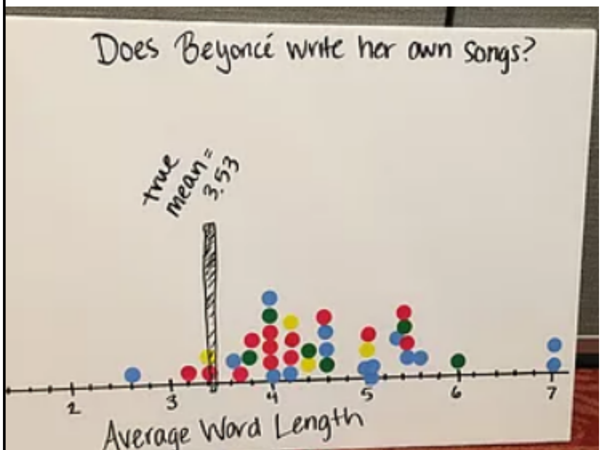
Important ideas:

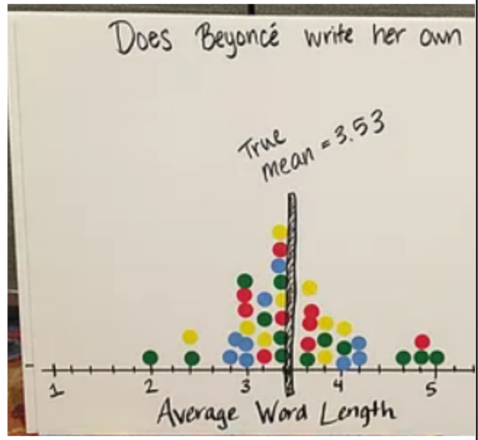
A statistic is an unbiased estimator if the mean of the sampling distribution is equal to the true parameter.



Remember Beyonce

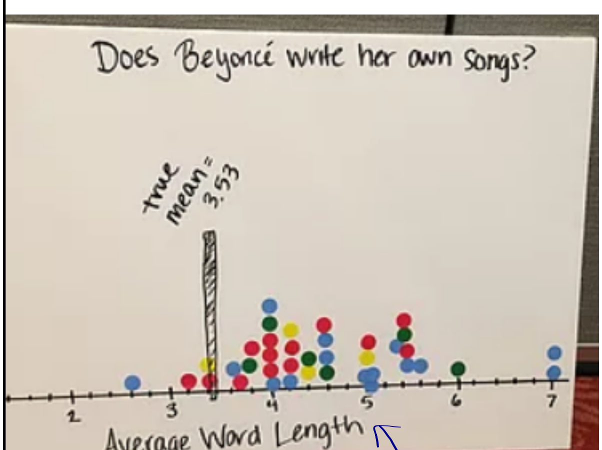
We were trying to estimate the true average word length from Crazy in Love so we could evaluate the claim that she did not write the lyrics.

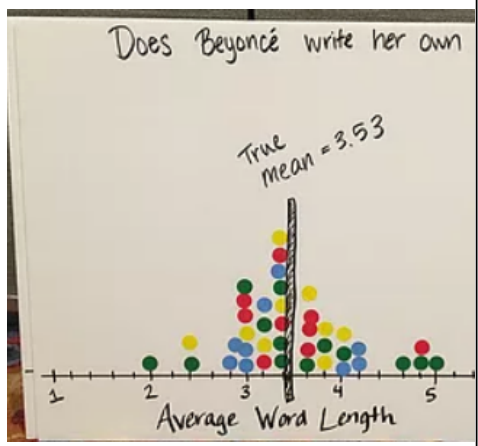




First sampling method had you quickly circle 5 words and find the mean.

SRS





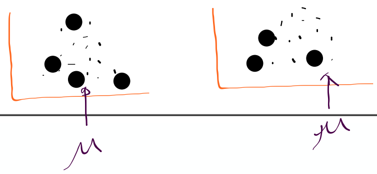
the mean of the sampling distribution was much greater than the true mean of the population

So the first mean μ was a biased estimator.

Biased and Unbiased Estimators

Important ideas:

A statistic is an unbiased estimator if the mean of the sampling distribution is equal to the true parameter.



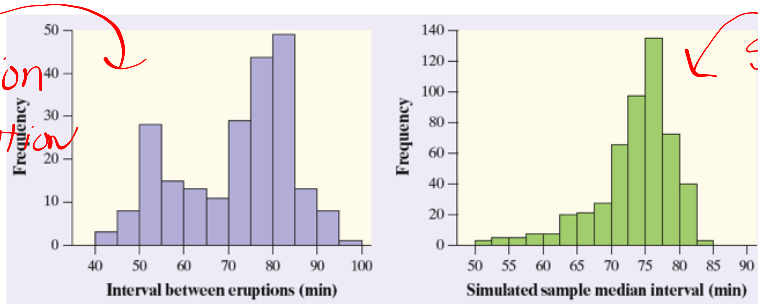
Increasing sample size decreases variability of the statistic.

Old Faithful

The histogram on the left below shows the interval (in minutes) between eruptions of the Old Faithful geyser for all 222 recorded eruptions during a particular month. For this population, the median is 75 minutes. We used technology to take 500 SRSs of size 10 from the population. The 500 values of the sample median are displayed in the histogram on the right. The mean of these 500 values is 73.5.

population distribution

Approx. Sampl. distribution

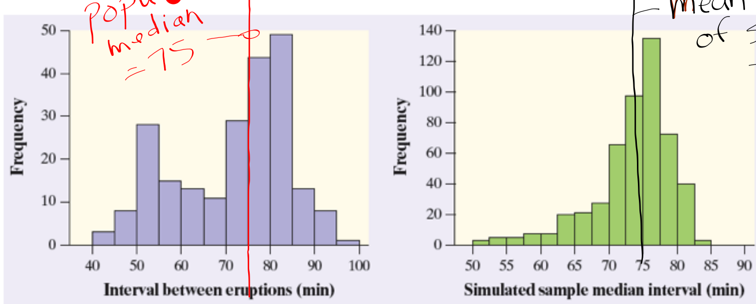


1. Is the sample median an unbiased estimator of the population median? Justify your answer.

The histogram on the left below shows the interval (in minutes) between eruptions of the Old Faithful geyser for all 222 recorded eruptions during a particular month. For this population, the median is 75 minutes. We used technology to take 500 SRSs of size 10 from the population. The 500 values of the sample median are displayed in the histogram on the right. The mean of these 500 values is 73.5.

Popul. median = 75

mean of 500 values of sample median 73.5



1. Is the sample median an unbiased estimator of the population median? Justify your answer.

No, the mean of the sample medians (73.5) is not the same as the true pop. median (75).

2. Suppose we had taken samples of size 20 instead of size 10. Would the variability of the sampling distribution of the sample median be larger, smaller, or about the same? Justify your answer.

3. Describe the shape of the sampling distribution of the sample median.

2. Suppose we had taken samples of size 20 instead of size 10. Would the variability of the sampling distribution of the sample median be larger, smaller, or about the same? Justify your answer.

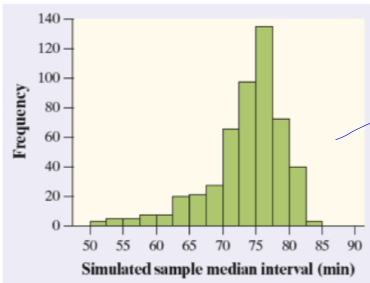
The variability would decrease because sample size was increased.

3. Describe the shape of the sampling distribution of the sample median.

2. Suppose we had taken samples of size 20 instead of size 10. Would the variability of the sampling distribution of the sample median be larger, smaller, or about the same? Justify your answer.

The variability would decrease because sample size was increased.

3. Describe the shape of the sampling distribution of the sample median.



skewed left with a single peak between 75 and 77.5

B.B.

7.1.....11, 13, 15, 19, 21, 25, 26-30

and study pp. 447-453