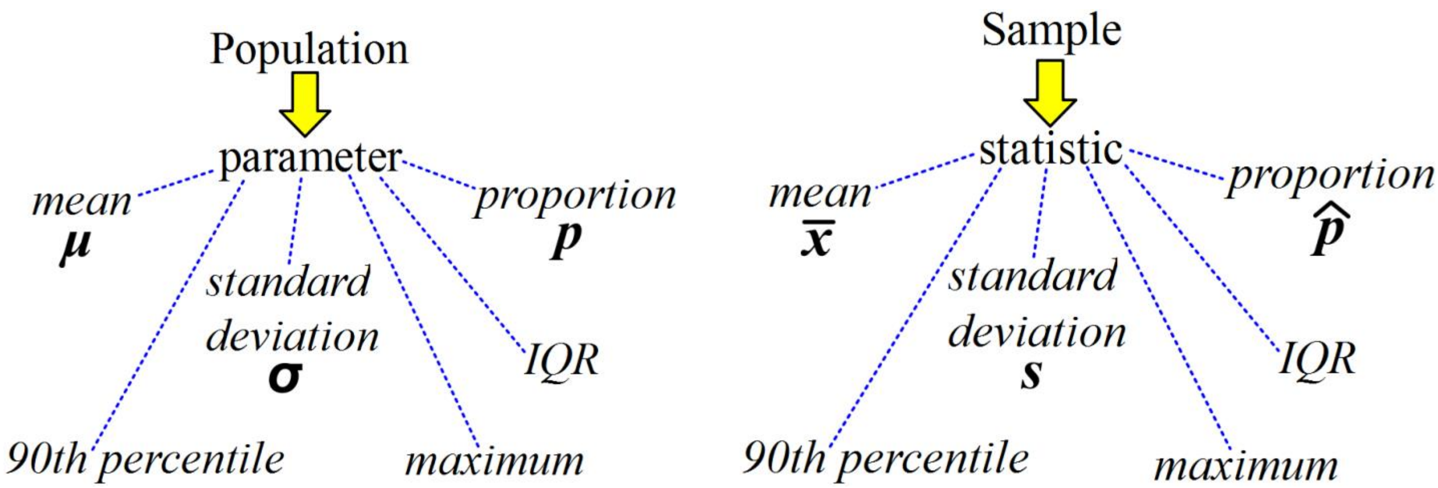
7.1 – Population vs. Sample & Parameter vs. Statistic

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

A statistic is a number that describes some characteristic of a sample.



Ex #1: You want to know the mean income of the subscribers to a particular magazine. You draw a random sample of 100 subscribers and determine that their mean income is $27,500.

- **What is the population?** ____
- **What is the population parameter of interest?** ______________________________
- **What is the sample?** ____
- **What is the sample statistic?** ____

Ex #2: You want to know how many students at IRHS consume alcohol. You survey a random sample of 200 IRHS students and conclude that 65% do not consume alcohol.

- **What is the population?** ____
- **What is the population parameter of interest?** ______________________________
- **What is the sample?** ____
- **What is the sample statistic?** ____

7.1– Sampling Distribution Introduction *(video 1)*

- ***Inference***: ____

When you can't access the whole population, you should take a _______________ (_____) from the population.

Since we don't know much of anything about our population (its SOCS), then we need a distribution that we can rely on…introducing the…_____________________ _____________________________.

Sampling Distribution: ____

____

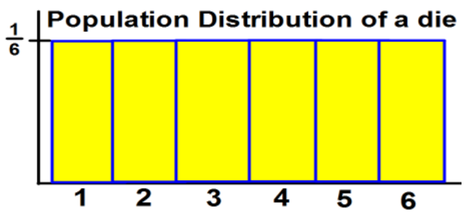
____

We do not need "many, many" samples thankfully. The **MAIN** thing we need is just _______ , _______ sample.

Is sampling distribution gives ***valuable information*** **(_________)** that we can use to determine whether a claim about the population is plausible or not plausible…

There may be only _______ population distribution, but there are ___________ different sampling distributions!!!

Let's explore sampling distributions and some awesome properties!!!

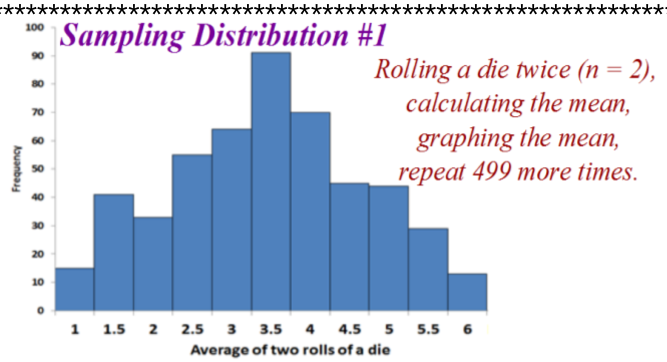


"S"hape: ____________________________________

"C"enter: ____________________________________

"S"pread: ____________________________________

****

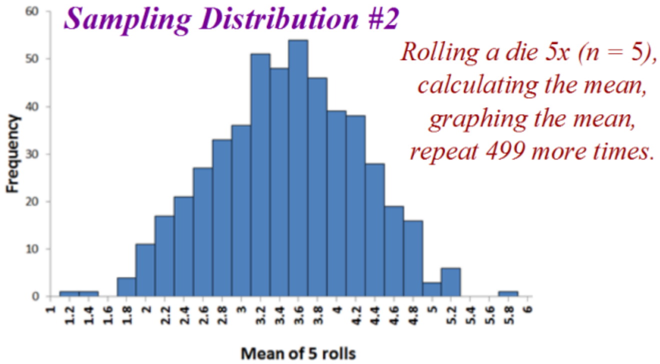


"S"hape: ____

"C"enter: ____

"S"pread: ____

(two more sampling distributions on the next page!)

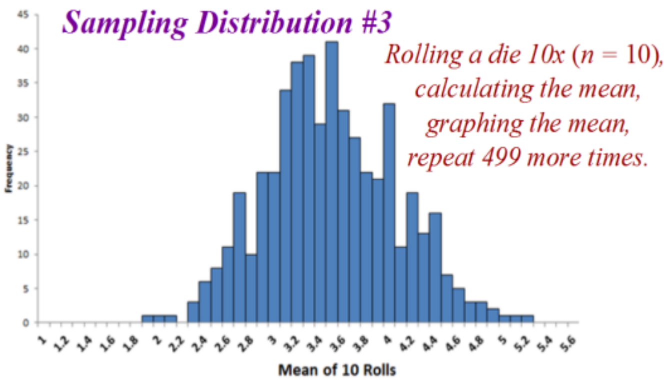


"S"hape: ______________________________________

"C"enter: ______________________________________

"S"pread: ______________________________________

****



"S"hape: ______________________________________

"C"enter: ______________________________________

"S"pread: ______________________________________

Summarize what you see with the population distribution vs. the various sampling distributions.

Shape: A sampling distribution will act more and more like an __________________ _______________ distribution as the sample size _____________, even when the population itself is not _________________ distributed.
Center: Sampling distributions have essentially the _____________ mean as the population from which it is drawn.
Spread: As you increase the sample size, the sampling distribution tends to be much less _____________________ or wild than the _______________________ from which it was drawn.
Outliers: Since large sample sizes tend to average out ____________ observations, sampling distributions typically do not have any outliers.

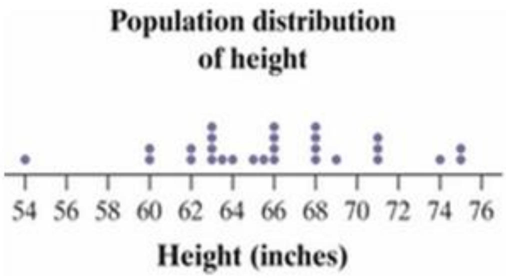
<u>7.1 – Bias and Variability</u> *(video 2)*

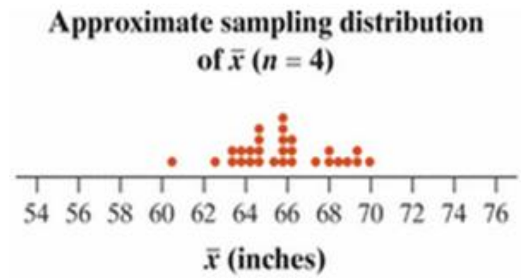
*Can **<u>ANY</u>** sample statistic be used to estimate its population parameter?* ______________________________

Some statistics produce too much ___________ or error to closely estimate the parameter of interest. This type of statistic is called a __________________ _____________________________.

But if a statistic ______________ estimates a parameter with very little bias or error, then that statistic is called an _____________________ ________________________________.

A statistic is ___________________ if the center of its sampling distribution is approximately the same as the population parameter.





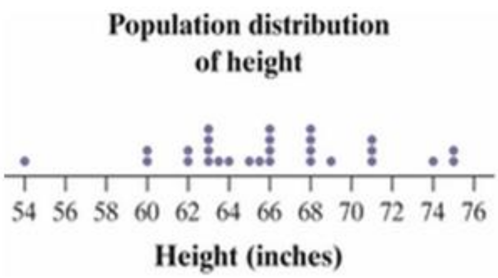
The center of this population distribution:

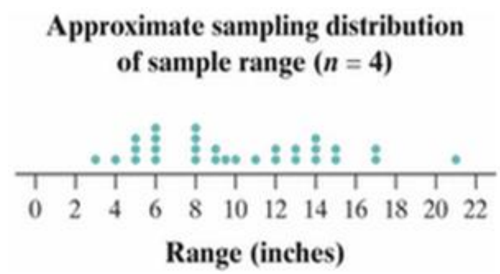
THE mean: _____________________

The center of this sampling distribution:

the mean of all the $\bar{x}$'s: _____________________

Is the mean an unbiased estimator??? **Conclusion…________**





The center of this population distribution:

THE range: _____________________

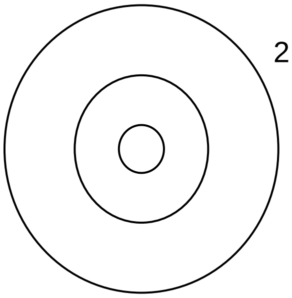
The center of this sampling distribution:

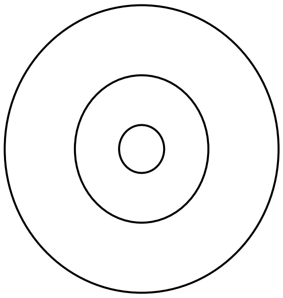
the mean of all the sample ranges:

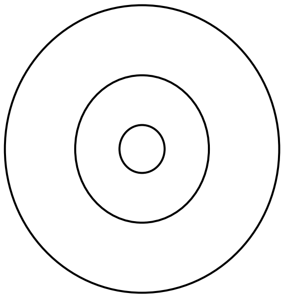
____

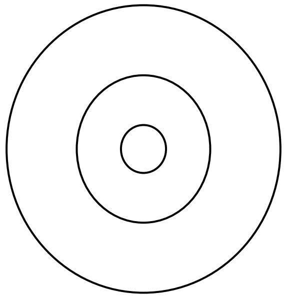
Is the range an unbiased estimator??? **Conclusion…________**

There are four situations regarding BIAS and VARIABILITY!

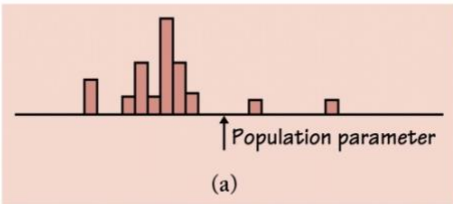
1. 
High Bias
Low Variability

2. 
Low Bias
High Variability

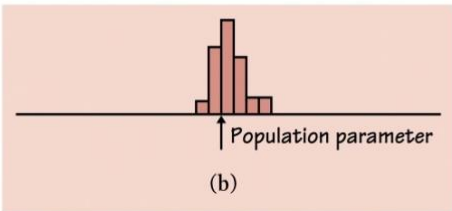
3. 
High Bias
High Variability

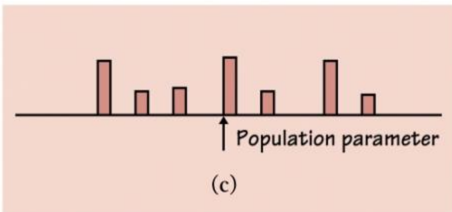
4. 
Low Bias
Low Variability

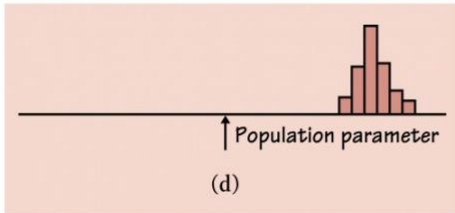
Match each histogram with one of the descriptions above.



Just because you are *precise* (low variability) does NOT mean you are *accurate* (low bias), too!







- **<u>Answer the last problems of the video here…</u>**

7.2 – Sampling Distributions of Sample Proportions *[VIDEO #3]*

Characteristics of the Sampling Distribution of Sample Proportions

1. **"S"hape** – The sampling distribution for sample proportions ($\hat{p}$) will be ______________ _____________ if the following condition is met: ________________________ and ____________________________.
   - The larger the sample size, *n,* the closer the shape is in becoming approx. normal.

2. **"C"enter –** The mean of all possible sample proportions (____) is equal to the population proportion, ___.

   ______________ = ______

3. **"S"pread –** The standard deviation of all possible sample proportions (_______) is

   ______________________________ **IF** the following condition is met!!!

   Is the population at least _______________ as large as the sample???

   This is referred to as the "independent condition" or the "10% condition".

Ex: One way of calculating the effect of undercoverage, nonresponse, and other sources of error in a sample survey is to compare the sample with known facts about the population. About 11% of Americans are teens. The proportion $\hat{p}$ of teens in a SRS of 1500 Americans should therefore be close to 11%. It is unlikely to be exactly 11% because of sampling variability. If a national sample contains only 9.2% teens, should we suspect that the sampling procedure is somehow under representing this group? Find the probability that a sample contains no more than 9.2% teens when the population actually consists of 11% teenagers.

1. Calculate the mean and standard deviation of the proportion $\hat{p}$ of the sample that are teens.

2. Calculate the probability that a sample contains no more than 9.2% teens when the population actually consists of 11% teenagers.

3. Interpret your results

<u>**7.3 – Sampling Distributions of Sample Means**</u> *[VIDEO #4]*

<u>Characteristics of the Sampling Distribution of Sample Means</u>

1. **"S"hape** – The sampling distribution for sample means (____) will be ______________ _____________ for <u>**ANY**</u> sample size, *n*, IF the population distribution is also _________________ _______________.

   What if the population is <u>**NOT**</u> approximately normal?!?!? → We will discuss that soon ☺

2. **"C"enter –** The mean of all possible sample means (____) is equal to the population mean, ____.

   ______________ = ______

3. **"S"pread –** The standard deviation of all possible sample means (_______) uses the formula

   ______________________________ <u>**IF**</u> the following condition is met!!!

   Is the population at least ________________ as large as the sample???

   This is referred to as the "independent condition" or the "10% condition".

<u>**Ex:**</u> A bottling company uses a filling machine to fill plastic bottles with soda. The bottles are supposed to contain 300 mL. In fact, the contents vary according to a normal distribution with mean, $\mu = 298$ mL, and standard deviation, $\sigma = 3$ mL.

a) What is the probability that the mean contents of six randomly selected bottles is less than 295 mL?

b) Would the probability that the mean contents of <u>**ten**</u> randomly selected bottles being less than 295 mL be less than or greater than your answer to part a)?

7.3 – The Central Limit Theorem (CLT) *[VIDEO #5]*

From Video #4…

Situation #1: When the Population Distribution is (Approximately) Normally Distributed

- …then we can assume the sampling distribution of the sample means is also **approx**. **normal**.

…………………………………………………………………………………………

Now for Video #5…

Situation #2: When the Population is Not Normally Distributed or not known altogether

- …then the sampling distribution CAN BE approximately normal **IF** we have a large ______________ ______________ ... all thanks to the ___________________ ________________ __________________ !!!
- How large a sample size *n* is needed for sampling distribution of sample means to be close to approximately normal depends on the shape of the ______________________ ______________________.
- More observations are required if the shape of the population distribution is far from ________________, but we can ***safely*** call the sampling distribution approx. normal when we reach a sample size of _____'ish.
- WARNING!!! The CLT is used with sampling distributions for sample ______________ ONLY!!!!

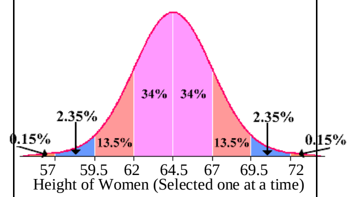
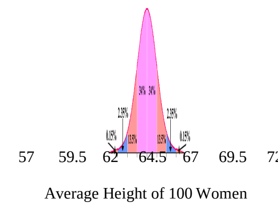
Ex: The number of lightning strikes on a square kilometer of open ground in a year has a mean of 6 and standard deviation of 2.4. The National Lightning Detection Network (NLDN) uses automatic sensors to watch for lightning in a random sample of 10 one-square-kilometer plots of land.

a) What are the mean and standard deviation of the sampling distribution of the sample mean number of strikes per square kilometer?

b) Explain why you cannot safely calculate the probability that the mean number of lightning strikes per square kilometer ($\bar{x}$) is less than 5 based on a sample size of 10.

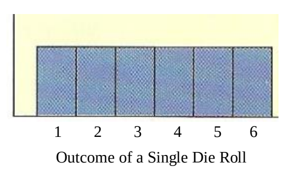
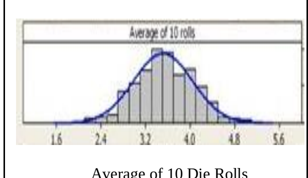
c) Suppose the NLDN takes a random sample of 50 square kilometers instead. Calculate the probability that the mean number of lightning strikes per square kilometer ($\bar{x}$) is less than 5.

The Central Limit Theorem (CLT) Summarized!

If the distribution of the Population is Normal… *(Example: Heights of Women)*					
	n	**Shape**	**Center**	**Spread**	**Picture**
Population Distribution	**1**	**Normal**	**μ = 64.5 in.**	**σ = 2.5 in**	
Sample Distribution	**n≥1**	**Also Normal!**	**Also 64.5!** $\mu_{\bar{x}}$ **= 64.5 in.**	**Much Less!** $\sigma_{\bar{x}} = \frac{2.5}{\sqrt{n}}$ **If n = 100** $\sigma_{\bar{x}} = 0.25$	

<u>Conclusions</u> ***Shape*: <u>Stays Normal!!!</u>** ***Center*: <u>Same!!!</u>** ***Spread*: <u>Smaller!!!</u>**

OR…

If the distribution of the Population is <u>NOT</u> Normal… *(Example: Rolling a Single Die)*					
	n	**Shape**	**Center**	**Spread**	**Picture**
Population Distribution	**1**	**Uniform**	**μ = 3.5**	**σ = 1.71**	
Sample Distribution	**n≥1**	**Becomes Normal!**	**Also 3.5!** $\mu_{\bar{x}}$ **= 3.5**	**Much Less!** $\sigma_{\bar{x}} = \frac{1.71}{\sqrt{n}}$	

<u>Conclusions</u> ***Shape*: <u>Becoming Normal!!!</u>** ***Center*: <u>Same!!!</u>** ***Spread*: <u>Smaller!!!</u>**

CLT & SOCS

The central limit theorem tells us that a sampling distribution always has significantly less wildness or variability, as measured by standard deviation, than the population it's drawn from. Additionally, the sampling distribution will look more and more like normal distribution as the sample size is increased, **even when the population itself is not normally distributed**!

Thanks to the central limit theorem, we can be sure that a mean or x-bar based on a reasonably large randomly chosen sample will be remarkably close to the true mean of the population. If we need more certainty we need only increase the sample size.

As the Sample Size "n" Increases:

Shape: becomes more and more approx. normal'ish

Center: stays the same as the population!

Spread: becomes less and less variable or spread out!

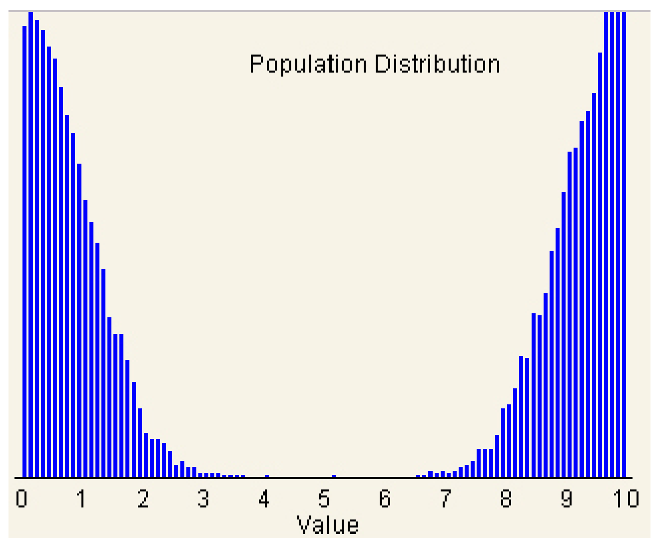


Fig. 2A) Histogram of Population - Bimodal Distribution: population = 16,000; mean = 5.002 std dev 4.242

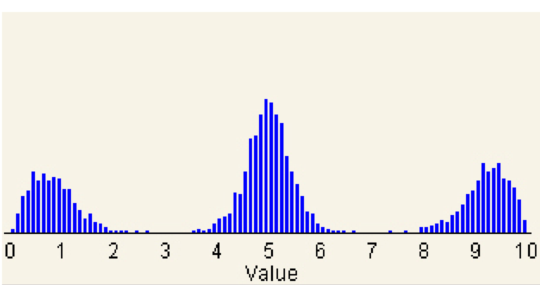


Fig. 2B) Sampling Distribution (from a bimodal population) n = 2: number of samples = 4000; mean = 4.977; std dev 3.017; std error = 2.999

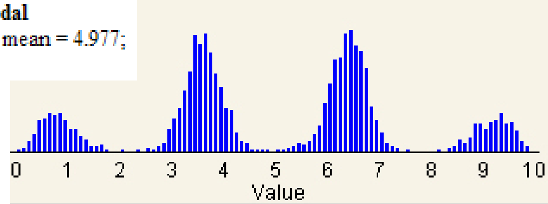


Fig. 2C) Sampling Distribution (from a bimodal population) n = 3: number of samples = 4000; mean = 4.946; std dev 2.425; std error = 2.449

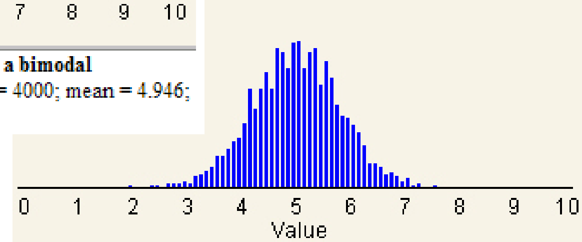


Fig. 2D) Sampling Distribution n = 30: number of samples = 4000; mean = 5.032; std dev 0.722; std error = 0.722

Preparing for Your Chapter 7 Test

- ❑ How to identify a parameter and a statistic from the context of the situation.
- ❑ How to find the mean of a sampling distribution (as long you have a SRS the mean of the sampling distribution should equal that of the population) - Know the proper notation.
- ❑ How to calculate the standard deviation of a sample mean and sample proportion (know the formulas and the proper notation)
- ❑ The exact definition of the important terms in the chapter such as: Sampling Distribution of a Statistic, Unbiased Estimator, Variability of a Statistic, etc.
- ❑ That the size of the **sample** is what impacts the spread (sampling variability) of the distribution. The **population** size does NOT affect spread (as long as the population is at least 10x the sample size).
- ❑ How to use and apply the Rule of Thumb #1 to sampling distributions
- ❑ How to use and apply the Rule of Thumb #2 to sampling distributions
- ❑ **How to calculate probabilities based on the normal approximations using either Table A or the calculator commands (normalcdf) – Look over HW problems. Use proper notation.**
- ❑ How to describe a sampling distribution. Address the following: shape, center, and spread. For example, "The distribution is normal with a mean of ____ and a standard deviation of ____".
- ❑ The law of large numbers ensures us that as the number of observations drawn increases, the mean ($\bar{x}$) of the observed values eventually approaches the mean μ of the population as closely as you specified and stays that close.
- ❑ The significance and use for the central limit theorem.
    - If the population is normally distributed, then the sampling distribution will also be normal regardless of the sample size.
    - If the population is NOT normally distributed, then the sampling distribution *becomes more and more normal* as the sample size increases. The larger the sample size, the more normally distributed we can assume the data to be.
- ❑ How to identify high/low bias and variability.