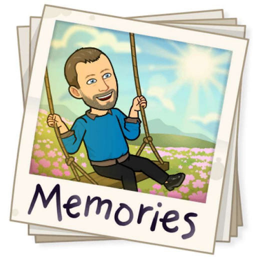
Introduction to Statistics I

Instructor: Jodin Morey   moreyj@lemoyne.edu

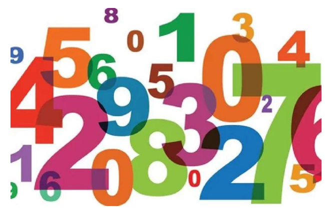
Previous Lecture

- How to apply steps of significance test to concrete examples

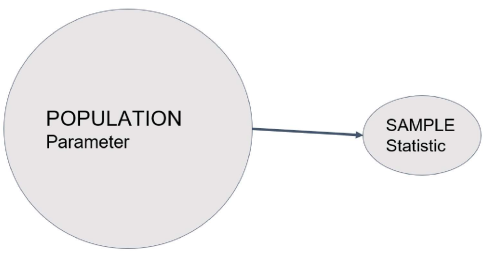


Topic 14: Sampling Dists for Means of Quantitative Vars

Previously we looked at sampling dist's and CLT with **categorical** vars.
How about **quantitative** ones?



Goal: How to use a quantitative statistic to estimate the parameter μ?



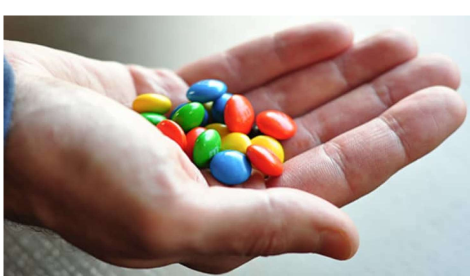
Recall:

	(Population) Parameter		(Sample) Statistic	
Proportion	π	"pi"	$\hat{p}$	"p-hat"
Mean	μ	"mu"	$\bar{x}$	"x-bar"
Standard Deviation	σ	"sigma"	s	

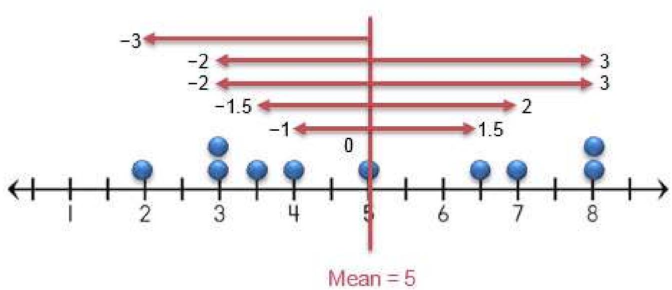
For our statistic, we'll focus on sample **means** ($\bar{x}$) (could also look at sample median, range, IQR, SD, etc).

We describe distr of the data **within** a sample by giving:

- **Shape** - normal, skew?
- **Center** - usually described by the mean $\bar{x}$, and
- **Spread** - usually described by SD (s) .

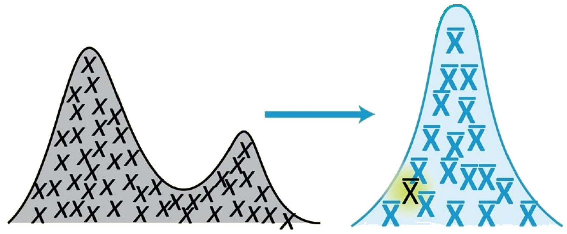


→



Calculating SD of a sample by finding distance of each M&M's diameter to

As we did with catagorical sample $\widehat{p}$'s, for quantitative vars, we ask how do the statistics $\bar{x}$ vary from **sample to sample**?



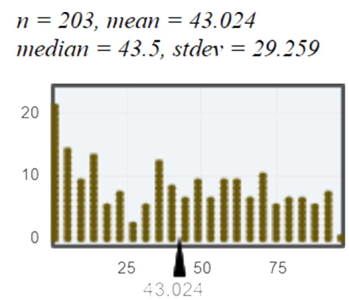
Distr of Pop. values. Distr of sample $\bar{x}$'s.

Example: Internet Access

Suppose we let our population be all 203 countries.

For each country, we can record the % of people w/internet access.

Parameters μ, σ from graph?



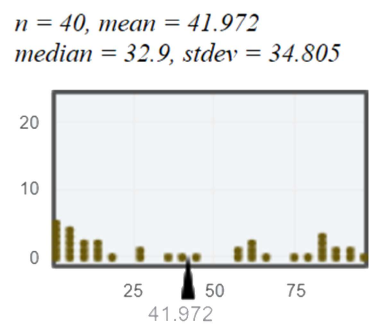
Population (each dot is a country)

$\mu = 43.02, \qquad \sigma = 29.26.$

Sampling: We take a sample of 40 countries.

What symbols do we use for mean/SD from this sample?

Are these parameters or statistics?



Sample

Statistics: $\bar{x} = 41.97, \; s = 34.81.$

Questions to Explore:

- This is just one sample. What happens when we sample again?
- How does the sample mean $\bar{x}$ relate to population mean μ?
- If we take samples repeatedly, does $\bar{x}$ vary around μ the same way (as when using catagorical vars) that $\widehat{p}$ did around π?
- And what role does the SD play in this relationship?

Let's experiment: On this applet choose "Percent w/Internet Access-2e,"
$n = 40$, then generate samples.

What does shape look like? Different from population distr above? SD?
What if $n = 200$?

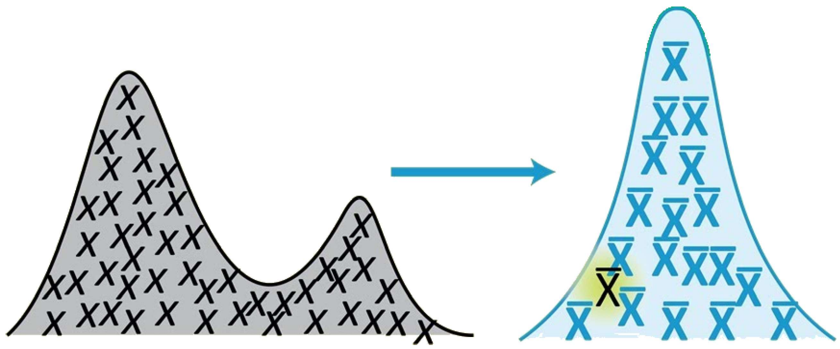
bit.ly/introstatsdata

Applet: Sampling Distribution for a Mean

Sampling Distribution

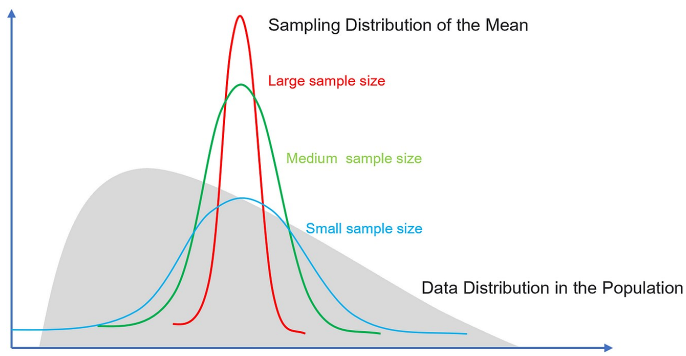
We take repeated samples of size n from a population with mean μ and SD of σ.

Each time, we calculate each sample's mean $\bar{x}$. What does the distr of these $\bar{x}$'s look like?



CLT: sample means → sample mean $\bar{x}$'s distr

- **Shape**: Approximately normal.
- **Center**: at the population mean μ.
- **Spread**: SD decreases as n increases.



Sampling distr is approx normal, even if population distr isn't!

Central Limit Theorem (for quantitative var)

Suppose samples of size n are taken from a population w/population mean μ and SD σ.

The sampling distr of sample means $\bar{x}$ will have:

- **Shape** - approximately normal (exactly so if population shape is normal).
- **Mean** - at μ.
- **SD** - $\frac{\sigma}{\sqrt{n}}$.

Two Tech Conds: CLT holds if:

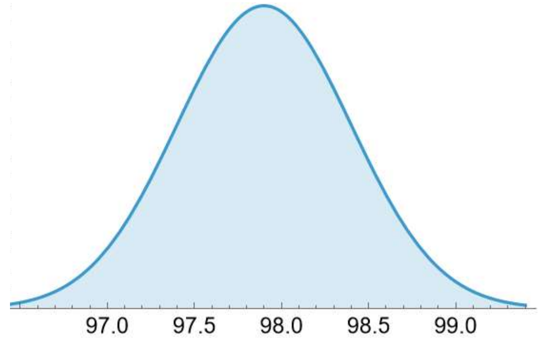
- Simple Random Samples (SRS) are taken, and
- Population is normal, **OR** as long as sample size is "big enough" ($n \geq 30$).

Body Temperatures

Body temperatures of healthy adults follow a Normal distr w/mean of $97.9^{\circ}F$ and SD of 0.5°. Are 97.9 & 0.5 parameters or statistics?

Symbols?

$\mu = 97.9, \qquad \sigma = 0.5$



What's the probability of finding a **single adult** w/temp above 98°?

$$z = \frac{\bar{x}-\mu}{\sigma} = \frac{98-97.9}{0.5} = 0.2.$$

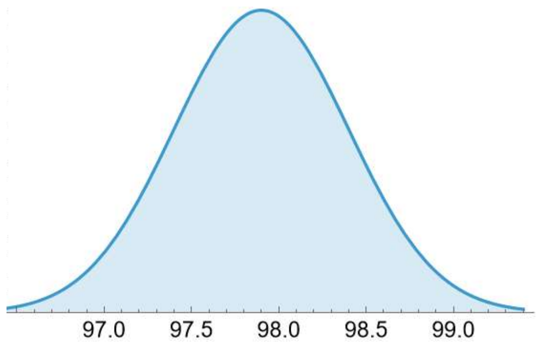
$P(x > 0.2) = 0.4207$ (from calculator.net: "z-score")

Instead, let's do many larger samples. What if we took SRS's w/$n = 130$? Does CLT apply?

Yes, "SRS" and "temps follow a Normal distr" (and $n = 130 \geq 30$).

What is the sampling distr: shape/center/spread?

Recall: $\sigma = 0.5$, $\mu = 97.9$ (from population)



Sample distr has:

- **Shape**: Normal.
- **Center**: At population mean $\mu = 97.9$.
- **Spread**: SD is $\frac{\sigma}{\sqrt{n}} = \frac{0.5}{\sqrt{130}} \approx 0.04385$.

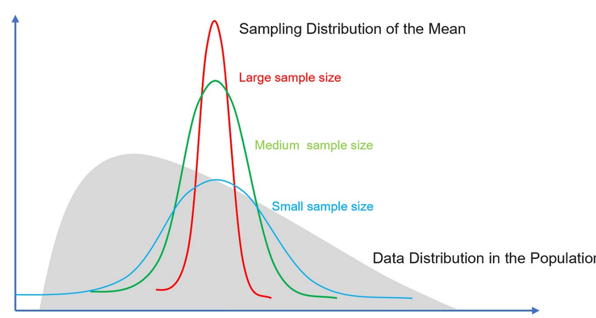
What's the prob of a SRS of 130 people having sample mean above 98°?

$$z = \frac{\bar{x}-\mu}{s} = \frac{98-97.9}{0.04385} \approx 2.281.$$

$P(x > 2.281) = 0.01127$ (from calculator.net: "z-score")

Probability of a **single adult** having body temp above 98° is 0.4207.

Probability of a **SRS of** 130 people having sample mean above 98° is 0.01127. (Woah!)



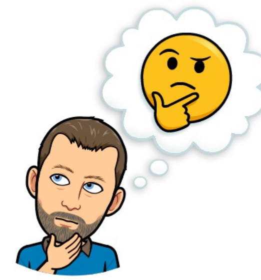
Statistic Distr Spread vs n

Increase sample size (n) $\Rightarrow$ spread of statistics ($\bar{x}$) decreases.

Thus, the $\bar{x}$'s are more likely to be near μ as n increases.

❗ Note that we are using the word "mean" in 3 different ways:

- **Population mean** μ
- **Sample mean** $\bar{x}$
- **Mean of the sample means**. Also known as the **mean of the sampling distr**.



CLT Comparison

	Categorical Var	**Quantitative Var**
Sample Statistic	$\hat{p}$ (proportion)	$\bar{x}$ (mean)
Tech Cond: **SRS and**:	When $n\pi \geq 10$ and $n(1-\pi) \geq 10$ or $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$	When population is normal or $n \geq 30$
Shape of Sampling Distr	Approximately Normal	Normal or Approximately Normal
Standard Deviation	$\sqrt{\frac{\pi(1-\pi)}{n}}$ or $se = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\frac{\sigma}{\sqrt{n}}$ or (see next lesson)
Center (same as pop)	π	μ

Activities: 14-1

What did we learn?

- Distributions of Quantitative Vars
- CLT: Shape/Center/SD, if pop. is normal or $n \geq 30$
- Symbols: π, μ, σ, $\hat{p}$, $\bar{x}$, s