WELCOME BACK!!!

Sampling Distributions
with proportions and means

AP Statistics
Chapter 18

SAMPLING* DISTRIBUTION

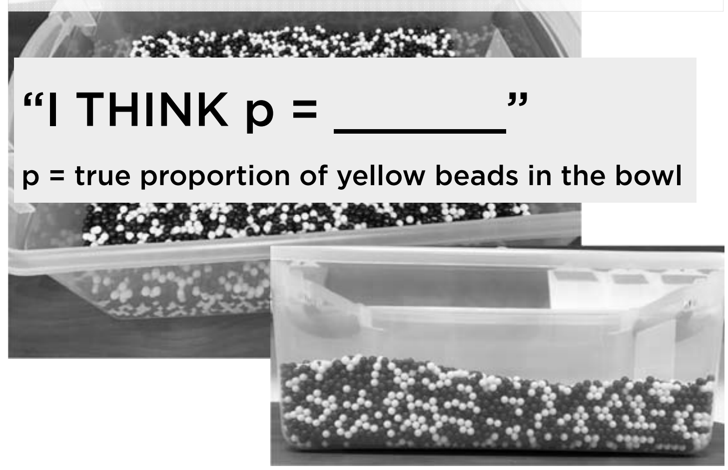
Collection of a LARGE NUMBER of sample statistics (means or proportions) of a given sample size.

**not "sample distribution"!*

PARAMETERS VS. STATISTICS
(population) (sample)

	MEANS *(numerical data)*	**PROPORTIONS** *(categorical data)*
Population **PARAMETER**		
Sample **STATISTIC**		

sampling bowl!



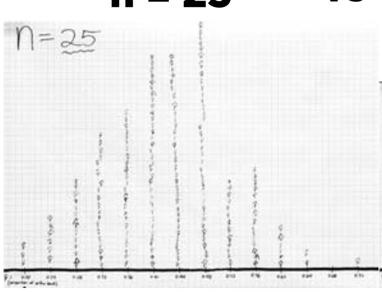
"I THINK p = ______"

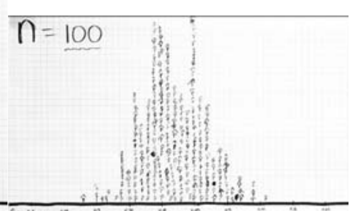
p = true proportion of yellow beads in the bowl

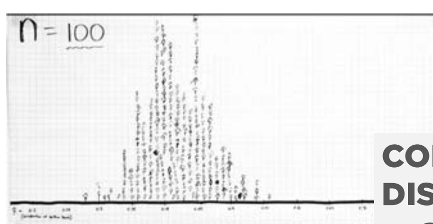
GO LOOK AT THE TWO GRAPHS…

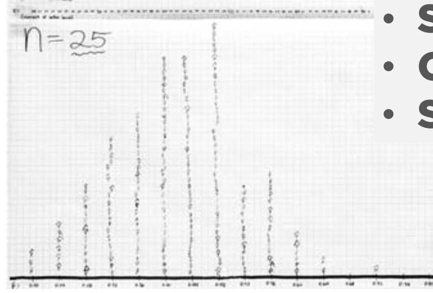
For which sample size are we **MORE** likely to get a sample proportion (p-hat) of **less than 30%?**

n = 25 **VS** **n = 100**









COMPARE THE DISTRIBUTIONS…

- **SHAPE?**
- **CENTER?**
- **SPREAD?**

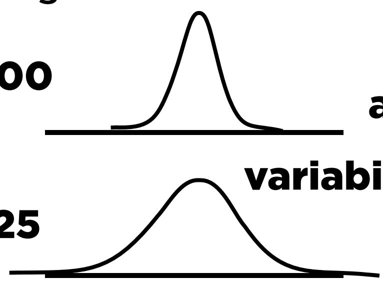
SAMPLE SIZE vs VARIABILITY (SPREAD)

What do we notice about the sampling models for...

n = 100

vs

n = 25



as n ↑,

variability ↓.

BIG PICTURE: **SAMPLE SIZE** affects the variability... (<u>not</u> center)

REMEMBER BIAS? (THIS WAS FROM THE FIRST WEEK OF SCHOOL...)

a) After 9/11, President Bush authorized government wiretaps on some phone calls in the U.S. without getting court warrants, saying this was necessary to reduce the threat of terrorism. Do you approve or disapprove of this?

53% of respondents approved

b) After 9/11, George W. Bush authorized government wiretaps on some phone calls in the U.S. without getting court warrants. Do you approve or disapprove of this?

46% of respondents approved

*THIS WAS JUST ONE FORM OF BIAS (RESPONSE BIAS)... THERE WERE OTHER FORMS OF BIAS AS WELL (VOLUNTARY, NON-RESPONSE, ETC)

UNBIASED VS BIASED ESTIMATORS

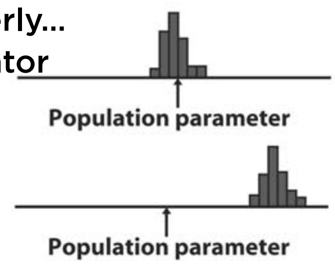
If we sample properly...

UNBIASED estimator

If there is BIAS in sampling method...

BIASED estimator

(sampling model does not properly represent the "truth")



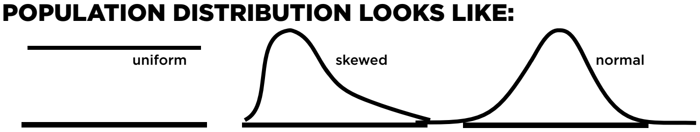
BIG PICTURE: **BIAS** has to do with **CENTER**... (not spread/variability)

NOW LET'S SHIFT GEARS AND TALK ABOUT SAMPLE MEANS...

Let's take random samples from a larger population.... and see what happens....

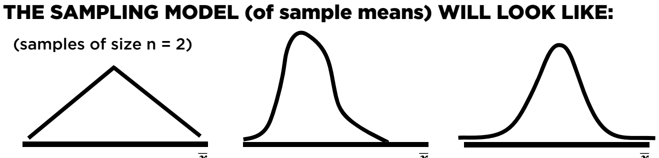
http://onlinestatbook.com/stat_sim/sampling_dist/index.html

POPULATION DISTRIBUTION LOOKS LIKE:

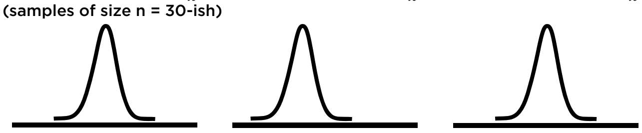


IF WE TAKE MAAAAANY SAMPLES OF THIS SIZE, THE SAMPLING MODEL (of sample means) WILL LOOK LIKE:

(samples of size n = 2)



(samples of size n = 30-ish)



CENTRAL LIMIT THEOREM (CLT)

When the sample size (n) is <u>LARGE ENOUGH</u>*, the distribution of sample <u>means</u> ($\bar{x}$) is **APPROXIMATELY NORMAL**, even when the population distribution is NOT.

*"Large enough" means "n" should be at least about 25 or 30… give or take depending on the type of distribution…

CONDITIONS *(for both means and proportions)*

I. RANDOM SAMPLE

of the population *(or at least a **<u>representative</u>** sample of the population… hopefully no reason to suspect bias???)*

II. 10% CONDITION

(ONLY check if sampling WITHOUT REPLACEMENT)

Sample size *(n)* should <u>not exceed</u> 10% of the population size.

*(this is so that sampling without replacement is **SIMILAR ENOUGH** to sampling WITH replacement)*

III. LARGE ENOUGH SAMPLE SIZE

(next slide…)

III. LARGE ENOUGH SAMPLE SIZE CONDITION

(different for means and proportions!)

SAMPLE PROPORTONS:

Both np and nq ≥ 10

Then the distribution of p-hat is approximately normal…

SAMPLE MEANS:

Two ways to show this*:

- Population is normally distributed
- Sample size > 30 *(by the Central Limit Theorem)*

Then the distribution of x-bar is approximately normal…

…for now. We will introduce a 3rd way to check this condition later…

DESCRIBING THE CENTER AND SPREAD FOR…

<u>PROPORTIONS</u>

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

These are found on the formula chart!

<u>MEANS</u>

$$\mu_{\bar{x}} = \mu$$

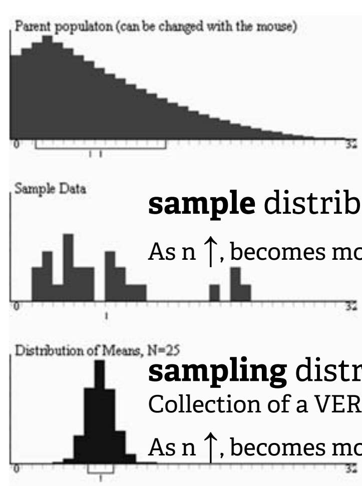
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

These are found on the formula chart!

sampling distribution

vs.

sample distribution



sample distribution (of ONE sample)

As n ↑, becomes more like population distribution

sampling distribution

Collection of a VERY LARGE # of sample means.

As n ↑, becomes more Normal.

Consider the following scenarios…

The mean height of females in the United States is about 64 inches (5' 4"), with a standard deviation of about 3 inches. Consider the likelihood of the following:

- One randomly selected female has a height of at least 5' 10".
- A random sample of FIVE females has a mean height of at least 5' 10".
- A random sample of 100 females has a mean height of at least 5' 10".

RANDOM **SAMPLE**

(of experiment units / subjects)

Allows us to GENERALIZE our results to a larger population

(very rare in an experiment)

vs

RANDOM **ASSIGNMENT**

(of subjects to treatments)

Allows us to draw CAUSAL conclusions (cause and effect)

ONTO THE ACTUAL PROBLEMS…

[POSTED IN A SEPARATE LINK ON THE WEBSITE!]