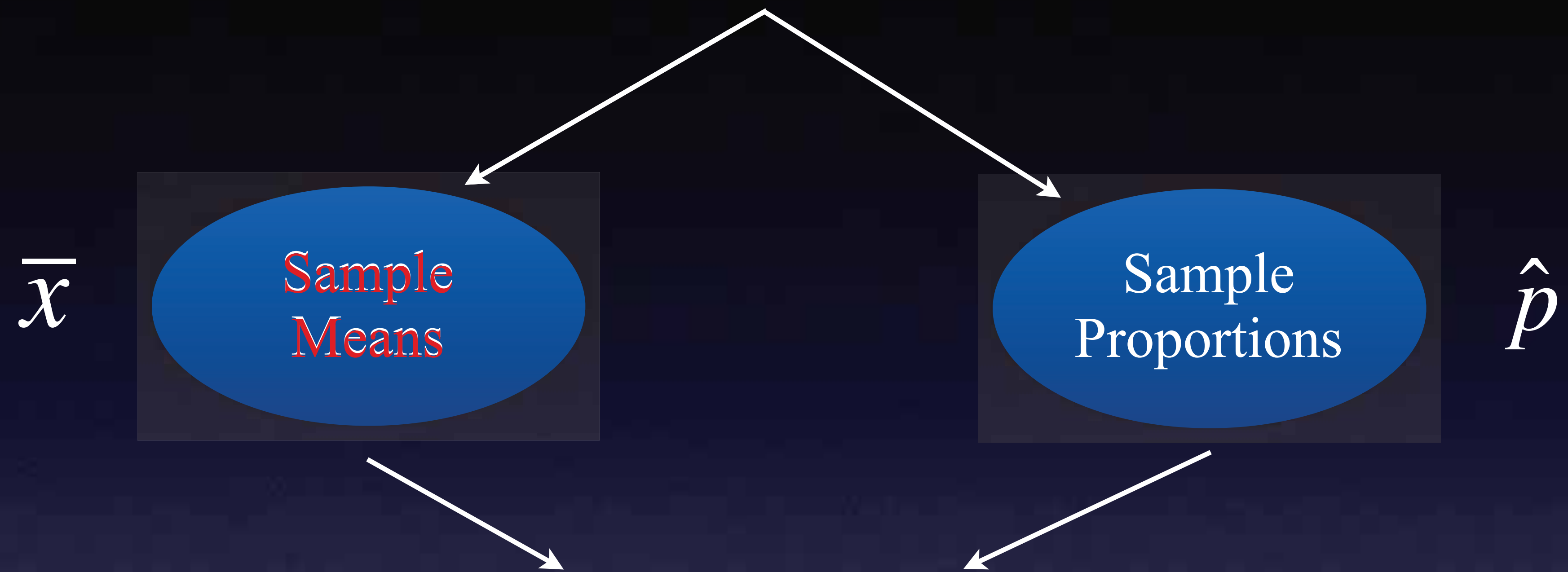
Sampling Distributions



- Mean
- Standard Deviation
- Normality - ?
- Calculate Probabilities

We will first focus on Sampling Distributions for Sample Means

This just means that it has a normal distribution with mean μ

$$\bar{x} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right)$$

and standard deviation $\frac{\sigma}{\sqrt{n}}$

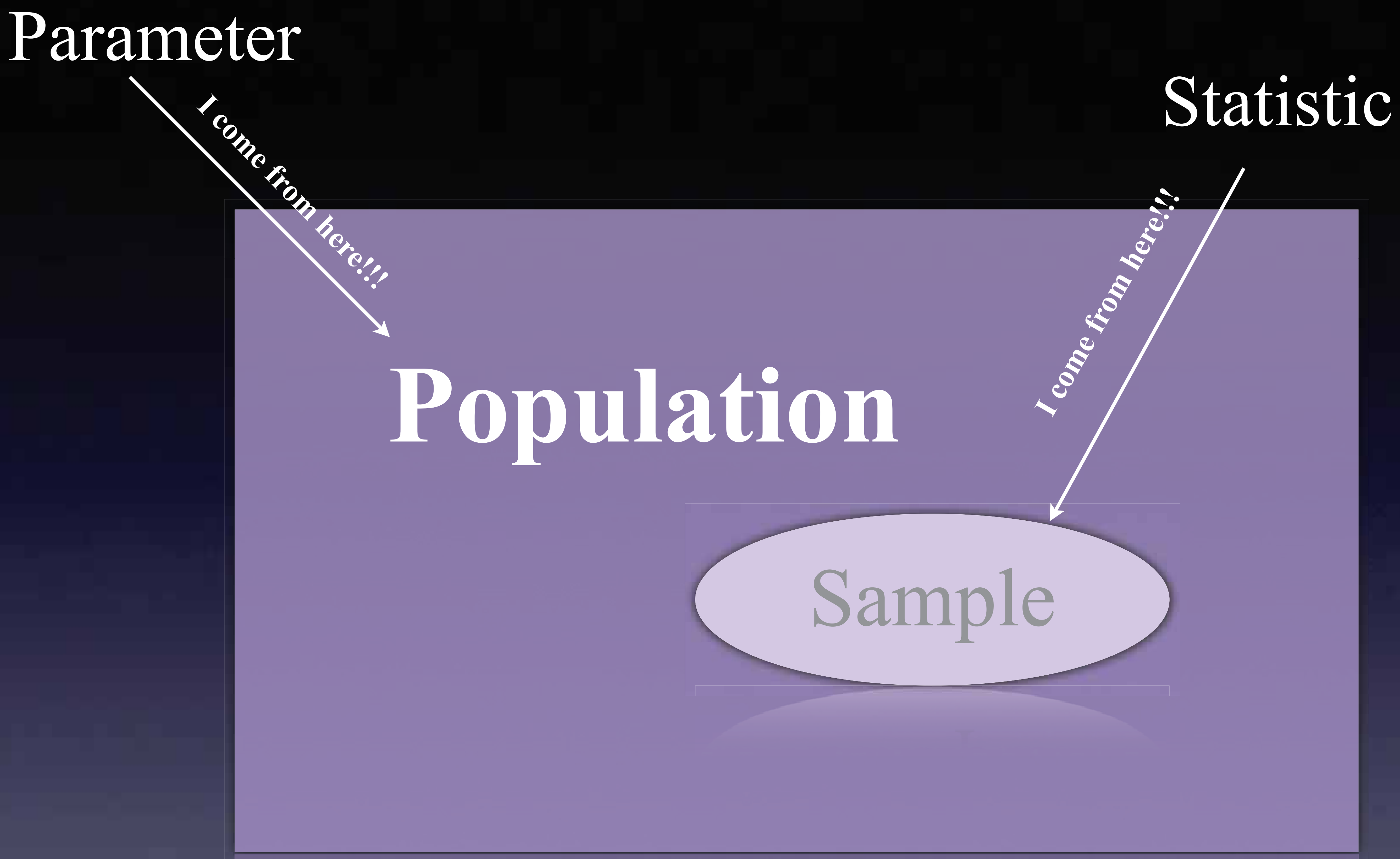
Parameter - Greek alphabet

$$\mu, \sigma, p$$

Statistic - Our alphabet

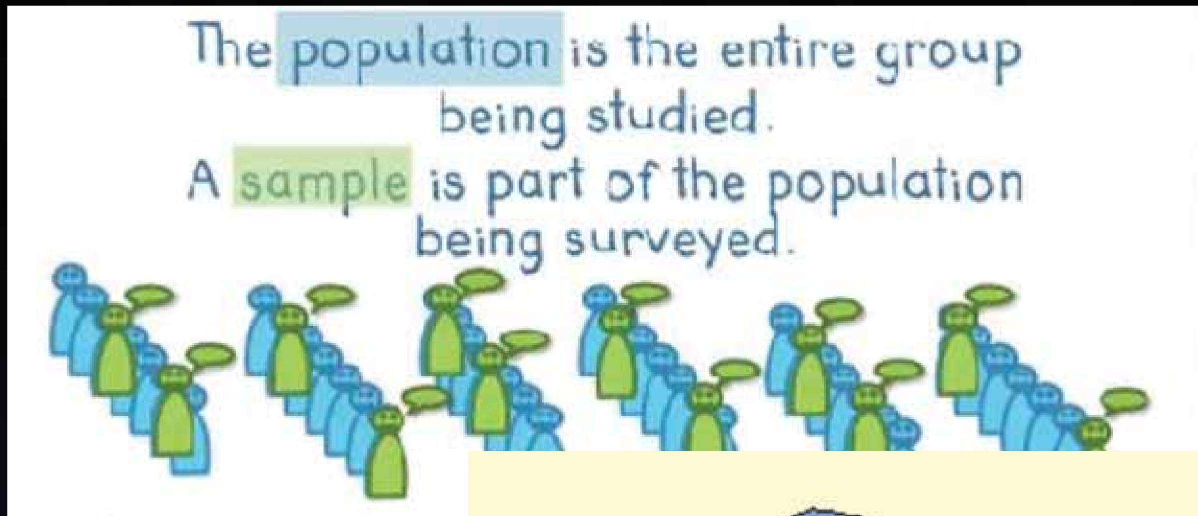
$$\bar{x}, s, \hat{p}$$

We will use statistics to approximate parameters.

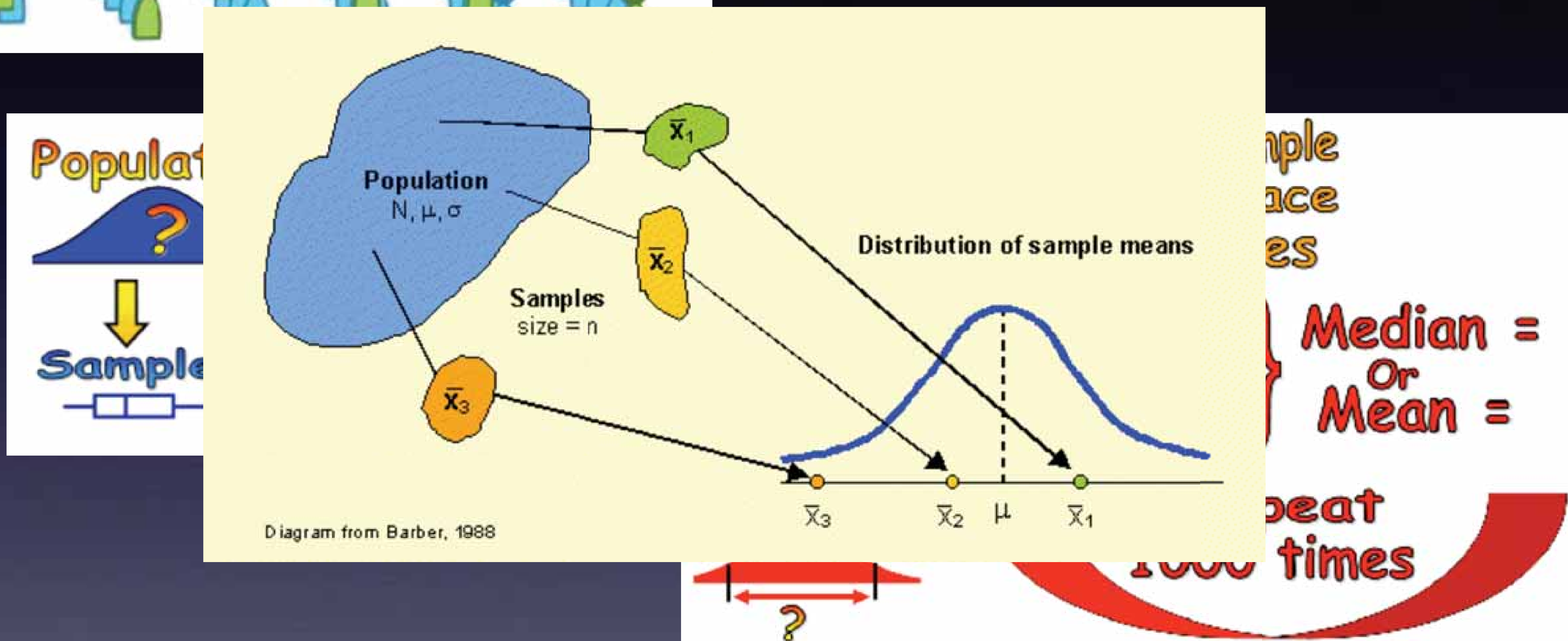


We won’t always have the same sample ... so we won’t always get the same mean for our statistics =

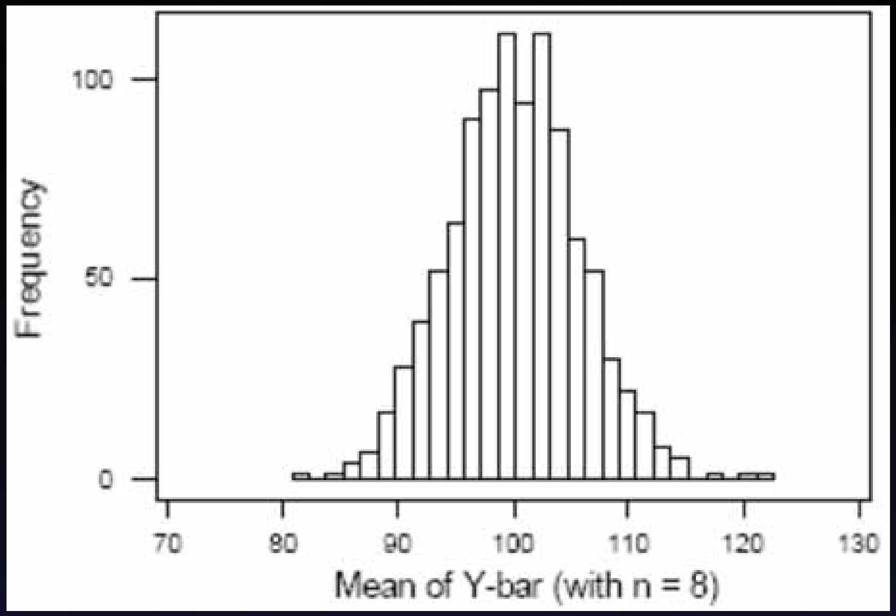
SAMPLING VARIABILITY



Recall that replication is important in experimentation

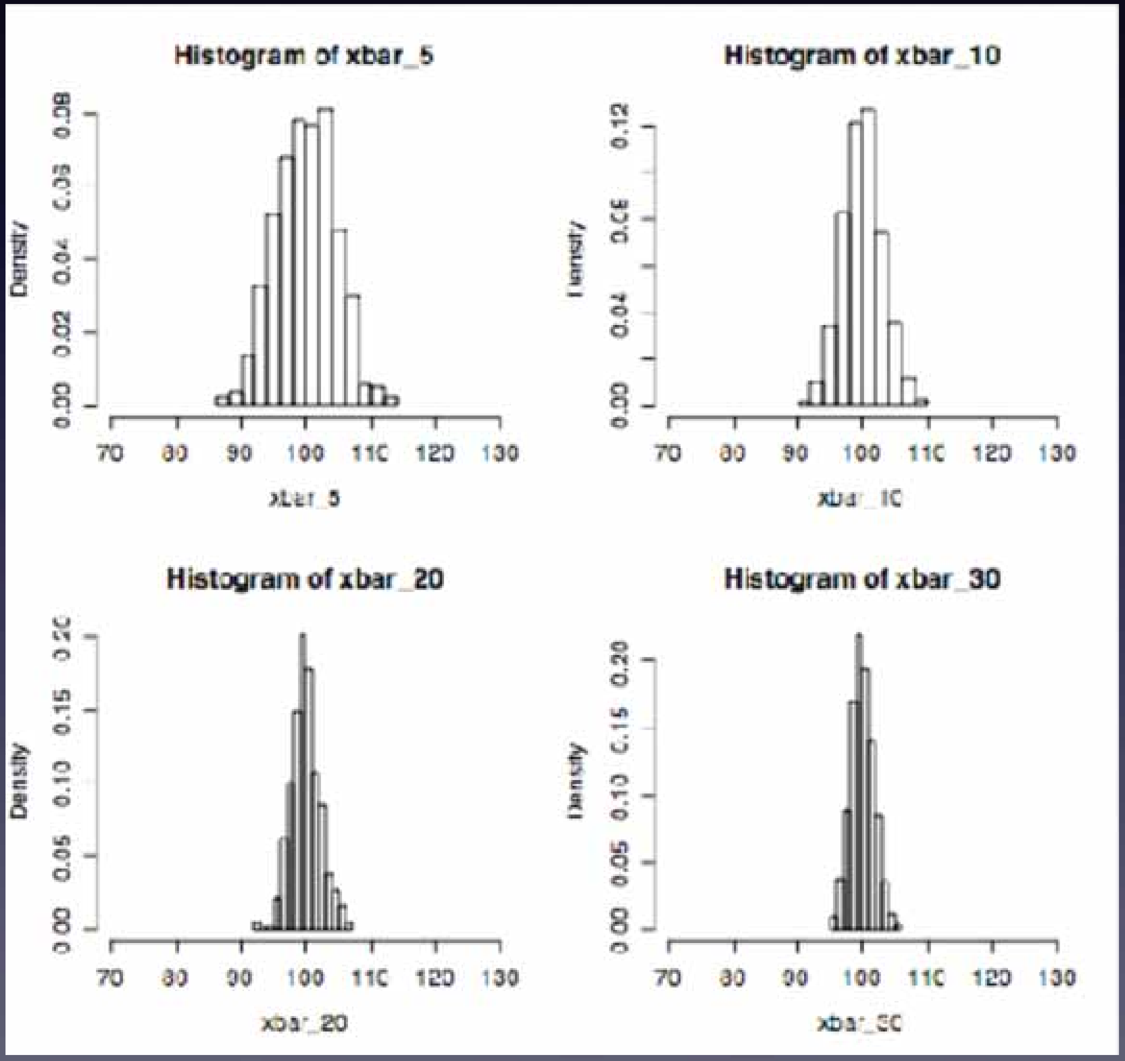


So what we are consolidating here are the means of all the samples to estimate the population mean



Population Distribution

$$x \sim N(100,\ 5)$$

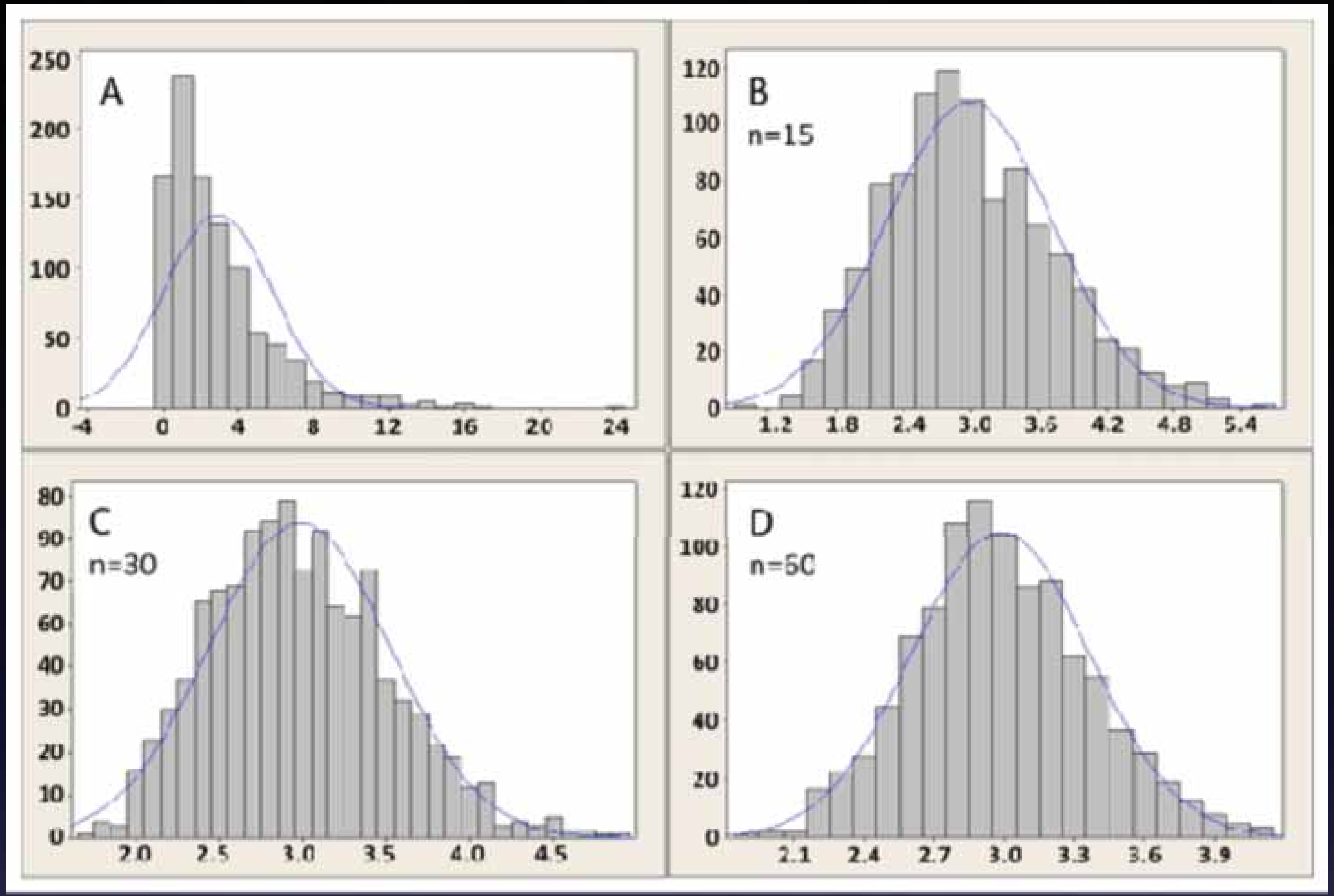


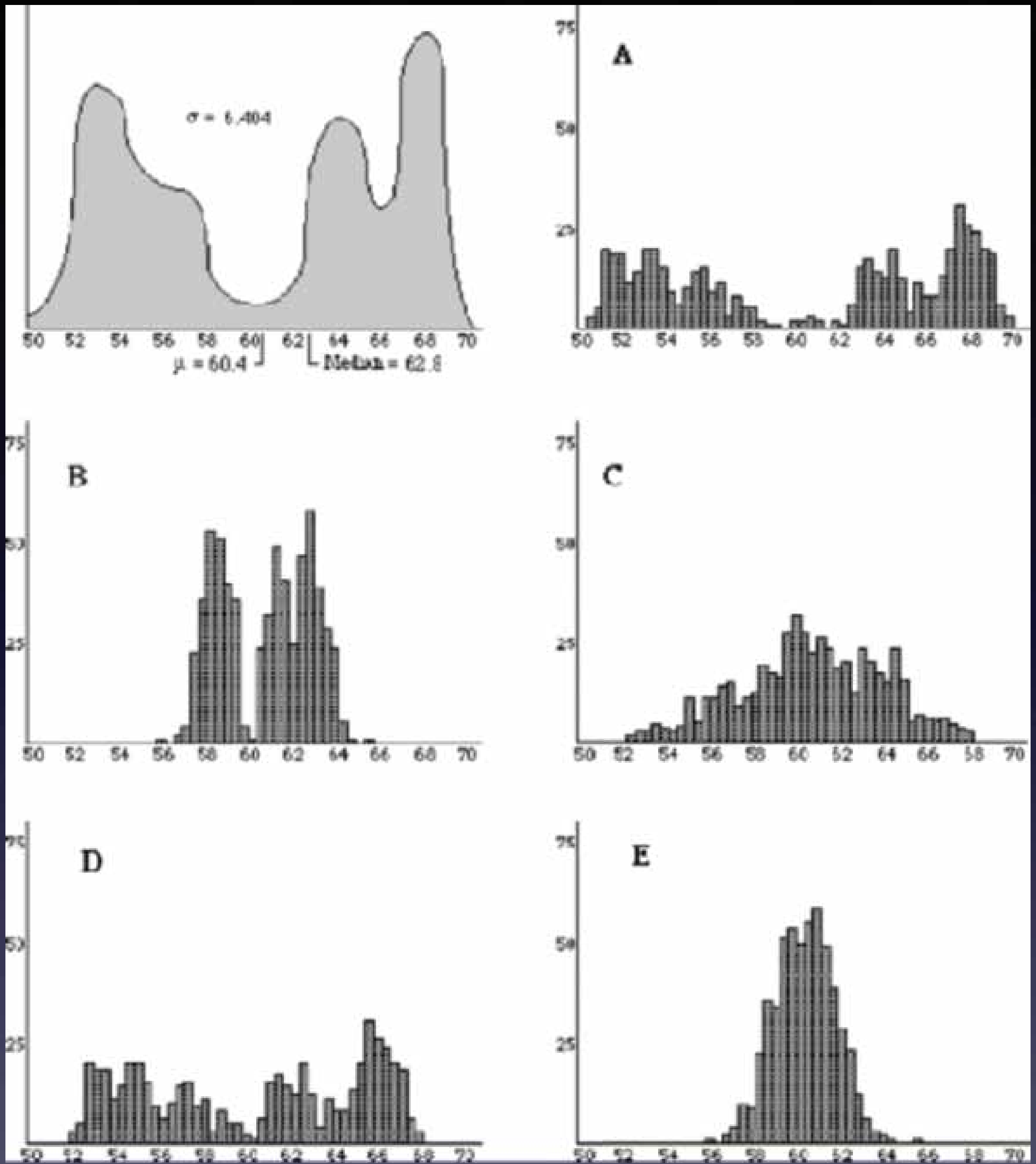
Sampling Distribution of the Sample Mean

$$\bar{x} \sim N\left(100,\ \frac{5}{\sqrt{n}}\right)$$

We can see here that as *n* increases, the standard deviation gets smaller because a larger *n* gets us closer to the actual mean

If the population distribution is ~*N*, then the sampling distribution of $\bar{x}$ will be ~*N*, no matter what the sample size.

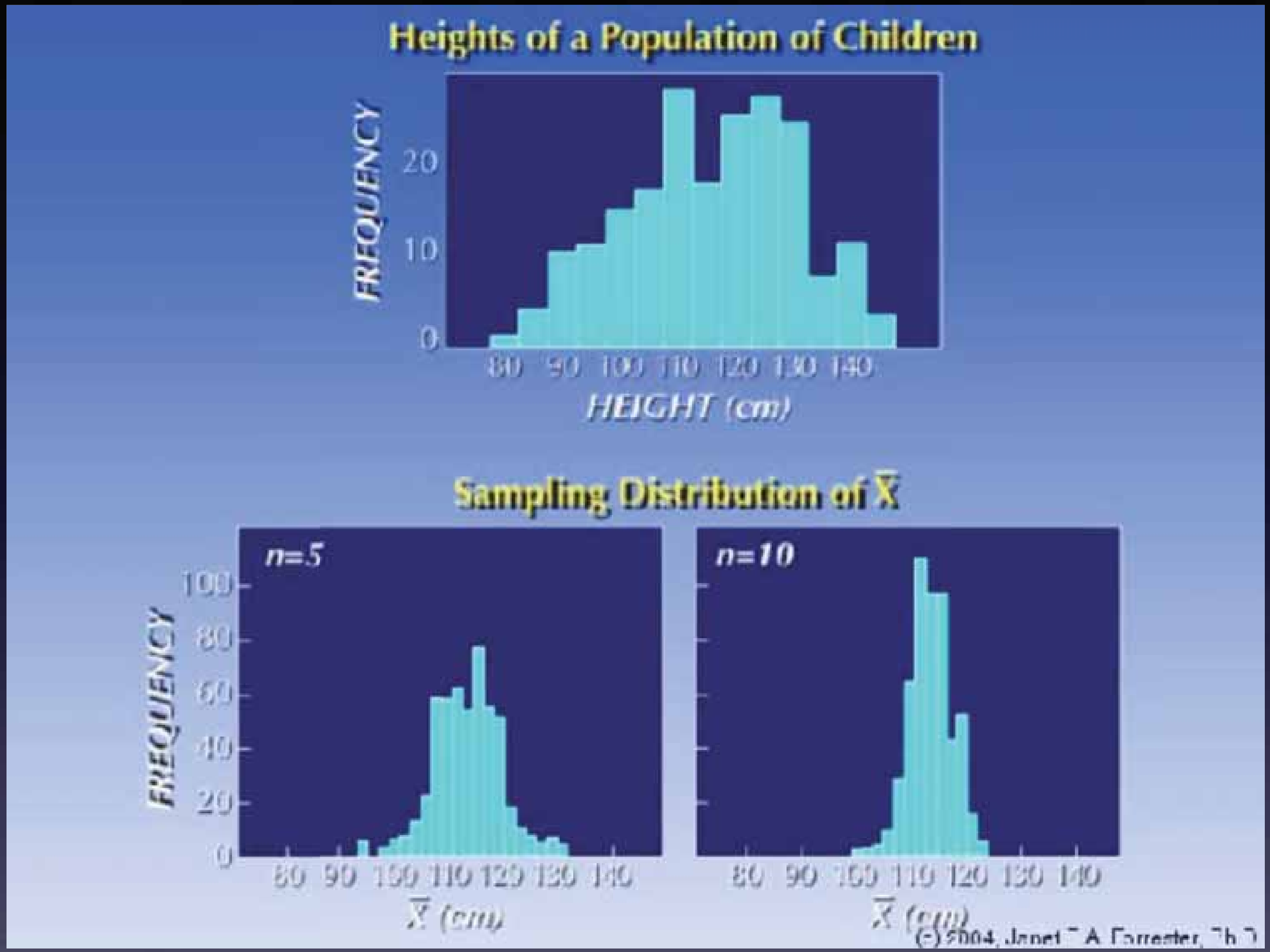




Notice that as the sample size increases - the centers stay the same, the spreads become smaller, and the shape of the sampling distribution gets closer to a normal curve.

Population Distribution

$x \sim SL(110,\ 10)$



$\bar{x} \sim SSL(110,\ 4.5)$ $\bar{x} \sim SSSL(110,\ 3.2)$

If the population distribution is not normal, then the sampling distribution of $\bar{x}$ will not necessarily be normal.

Sampling Distributions for Sample Means

- The **mean** of the sampling distribution is the same as the mean of the population:

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

- The **standard deviation** of the sampling distribution gets smaller according to this equation:

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

- **Normality** is attained in one of 2 ways:
  - The population distribution is stated to be normal.
  - The ***Central Limit Theorem*** states that as the sample size n increases, the sampling distribution becomes more normal (regardless of the shape of the population). In practice, if $n \geq 30$, we assume the distribution is approximately normal.

- Calculate **probabilities** using *normalcdf*

A Piece of Advice -

ALWAYS, I MEAN ALWAYS, CHECK YOUR ASSUMPTIONS!!!

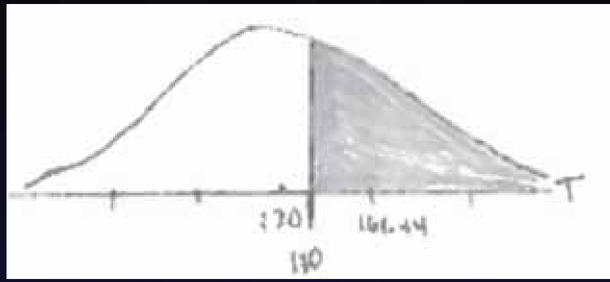
In "Mean Land" - (as opposed to Proportion Land which is next section)

- **Normality** is attained in one of 2 ways:
  - The initial distribution is stated to be normal.
  - The Central Limit Theorem states that as the sample size n increases, the sampling distribution becomes more normal (regardless of the shape of the population). In practice, if $n \geq 30$, we assume the distribution is approximately normal.

Free Response

Suppose that $T = 2X + 3Y$, where is X your score on the multiple-choice part of a test, Y is your score on the written part of a test, and T is your total score. In this case, your total score is calculated by doubling your multiple-choice score and tripling your written score. Suppose $X \sim N(30, 7)$ and $Y \sim N(20, 13)$

(a) What is the probability that T will be greater than 130?



$$\mu_T = 2(30) + 3(20) = 120$$

$$\sigma_T^2 = 2^2 \cdot 7^2 + 3^2 \cdot 13^2 = 1717$$

$$\sigma_T = \sqrt{1717} = 41.44$$

$$T \sim N(120, 41.44)$$

$$P(T > 130) = normalcdf(130,\ 1E99,\ 120,\ 41.44) = 0.405$$

(b) Suppose that we find three students' scores independently of one another. What is the probability that <u>at least one</u> of these measurements will be greater than 130?

success = measurement > 130

3 independent trials

$n = 3$

$X \sim B(3,\ 0.405)$

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) = 1 - P(X = 0)$$

$$= 1 - binompdf(3,\ 0.405,\ 0) = 0.789$$

(c) What is the probability that the <u>mean</u> of the three independent scores will be greater than 130?

$$\bar{T} \sim N\left(120, \frac{41.44}{\sqrt{3}} = 23.925\right)$$

$$P(\bar{T} > 130) = normalcdf(130,\ 1E99,\ 120,\ 23.925) = 0.338$$