

Sampling Distributions
Key to Statistical Inference



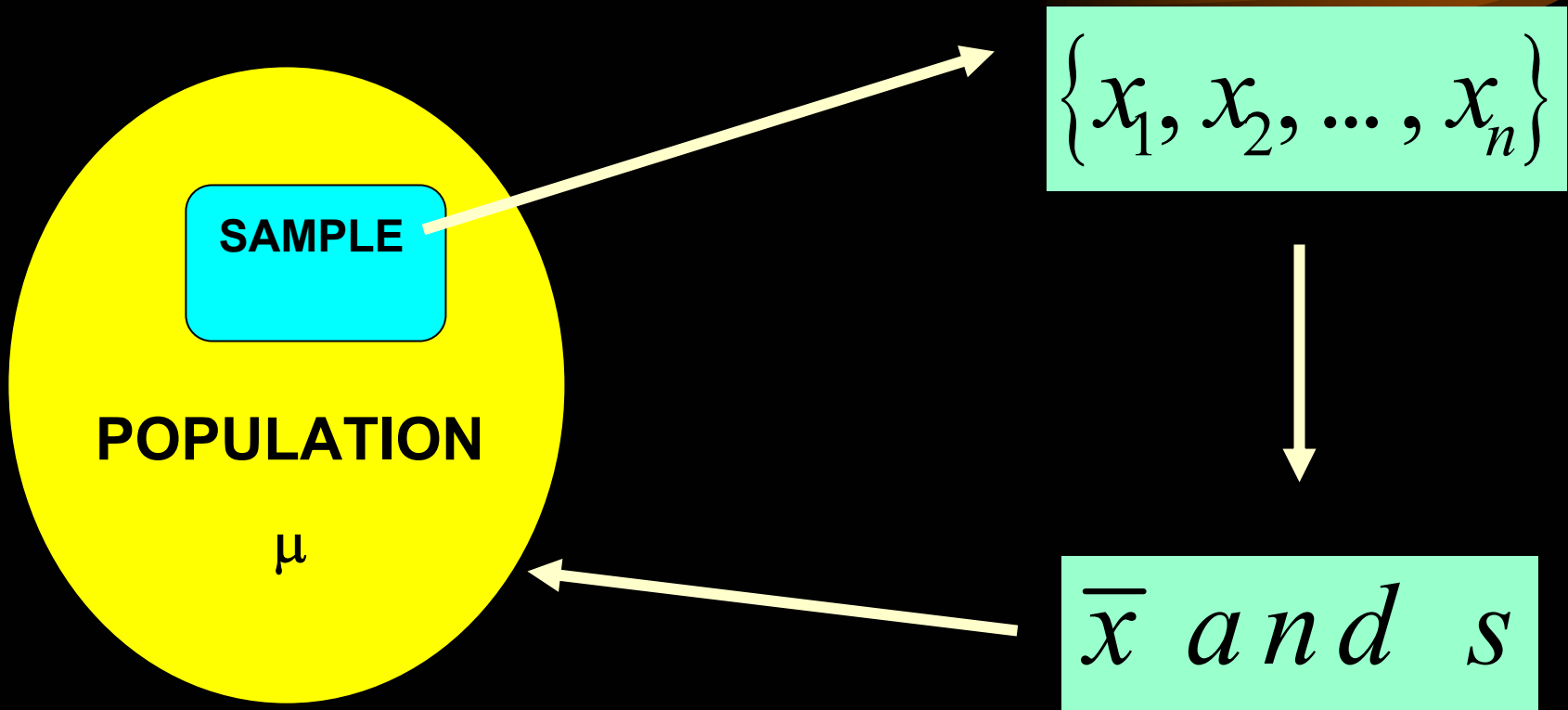
Jim Bohan

Manheim Township School District

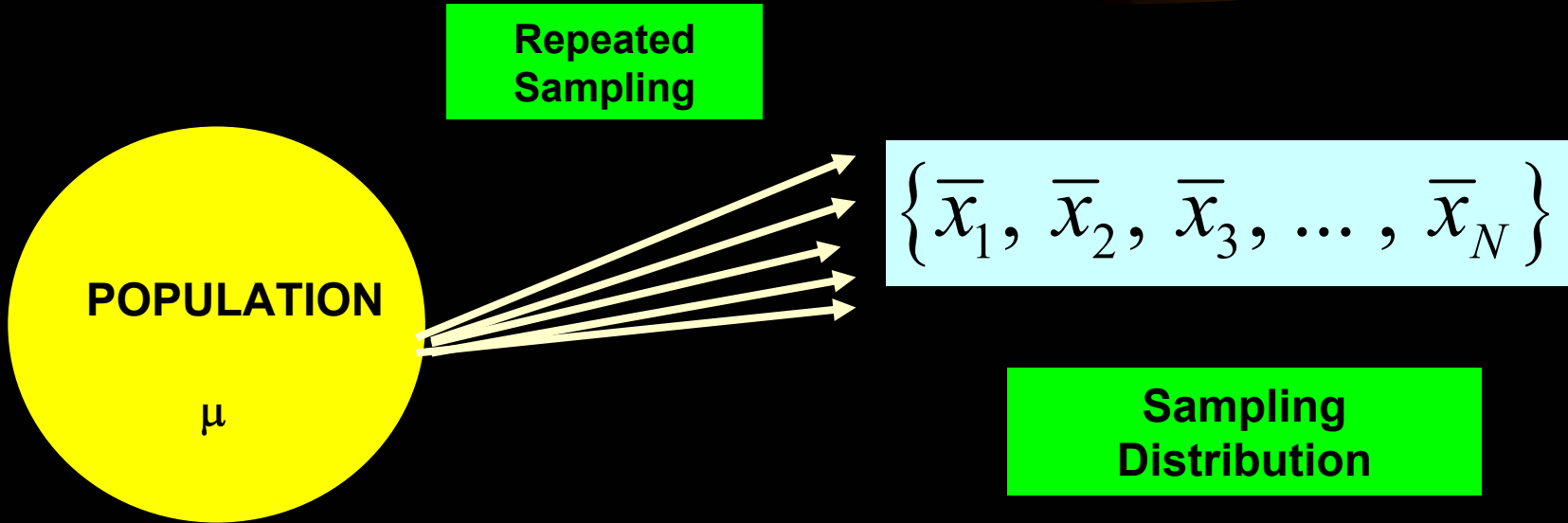
Lancaster Pennsylvania

jim_bohan@mtwp.k12.pa.us

The Sampling Process



The Sampling Distribution



And the Sampling Distribution is?

If the sampling fits the criteria for the Central Limit Theorem, we can recognize the Sampling Distribution. For example,

<p>SRS and Normal Population or Large Sample</p>	<p>$\{\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_N\}$</p> <p>→ $t_i = \frac{x_i - \bar{x}}{s} \Rightarrow \{t_1, t_2, \dots, t_n\}$</p> <p><i>approximately follows</i> <i>t - dist (df = n - 1)</i></p>
--	---

Elementary Sampling Distributions



- Normal (μ, σ)
- t-distribution (df)
- Chi-square (df)

If the Sampling Distribution is known...

Probability questions about sample statistics can be answered. For example,

A simple random sample of 50 is selected from a normal population with a mean of 50 and a standard deviation of 10. What is the probability that the sample mean will be greater than 53?

The Answer...

A simple random sample of 50 is selected from a normal population with a mean of 50 and a standard deviation of 10. What is the probability that the sample mean will be greater than 53?

Normal(μ, σ) \Rightarrow *Sampling Distribution Normal* $\left(\mu, \frac{\sigma}{\sqrt{n}} \right)$

Normal(50, 10) \Rightarrow *Sampling Distribution Normal* $\left(50, \frac{10}{\sqrt{50}} \right)$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{53 - 50}{\frac{10}{\sqrt{50}}} = 2.12 \Rightarrow p = .017$$

Two Deceptively Similar Problems

Census data states that the mean height of adult males in a city are normally distributed with a mean of 69.2 inches and a standard deviation of 4.3 inches. A sample of 35 is selected. What is the probability that one of these subjects has a height greater than 70 inches?

⇒ ***DATA
DISTRIBUTION***

Census data states that the mean height of adult males in a city are normally distributed with a mean of 69.2 inches and a standard deviation of 4.3 inches. A sample of 35 is selected. What is the probability that the mean of these subjects has a height greater than 70 inches?

⇒ ***SAMPLING
DISTRIBUTION***

The Key to Significance Testing

$$H_0 : \mu = 50$$

$$H_a : \mu > 50$$

$n = 40; \bar{x} = 55; s = 10 \Rightarrow$ Central Limit Theorem \Rightarrow

Sampling Distribution: approximate t ($df = 39$) \Rightarrow

p -value = $P(\text{sample of } 40 \text{ has mean of } 55 \text{ or more when } \mu = 50)$

$$\therefore t = \frac{55 - 50}{\frac{10}{\sqrt{40}}} = 3.162 \Rightarrow p\text{-value} = .0015$$

Another Example

$$H_0 : p = .42$$

$$H_a : p < .42$$

$$n = 30; \hat{p} = .40 \Rightarrow \text{Central Limit Theorem} \Rightarrow$$

$$\text{Sampling Distribution : Approx. Normal} \left(.42, \sqrt{\frac{(.42)(.58)}{30}} \right) \Rightarrow$$

$$p\text{-value} = P(\text{sample of 30 has } \hat{p} = .40 \text{ or less when } p = .42)$$

$$z = \frac{.40 - .42}{\sqrt{\frac{(.42)(.58)}{30}}} = -2.22 \Rightarrow p\text{-value} = .0132$$

Simulating Sampling Distributions

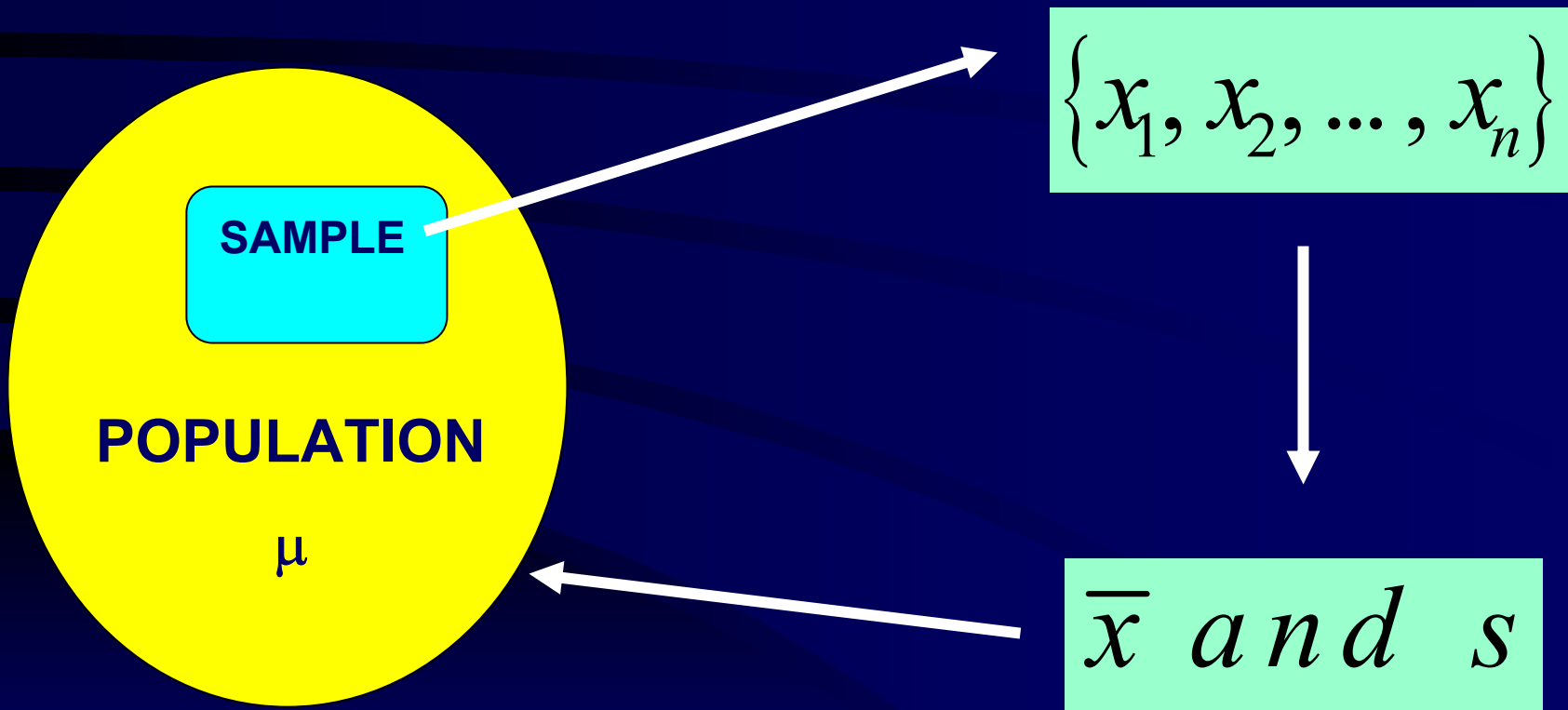
Jim Bohan

Manheim Township School District

Lancaster PA

jim_bohan@mtwp.k12.pa.us

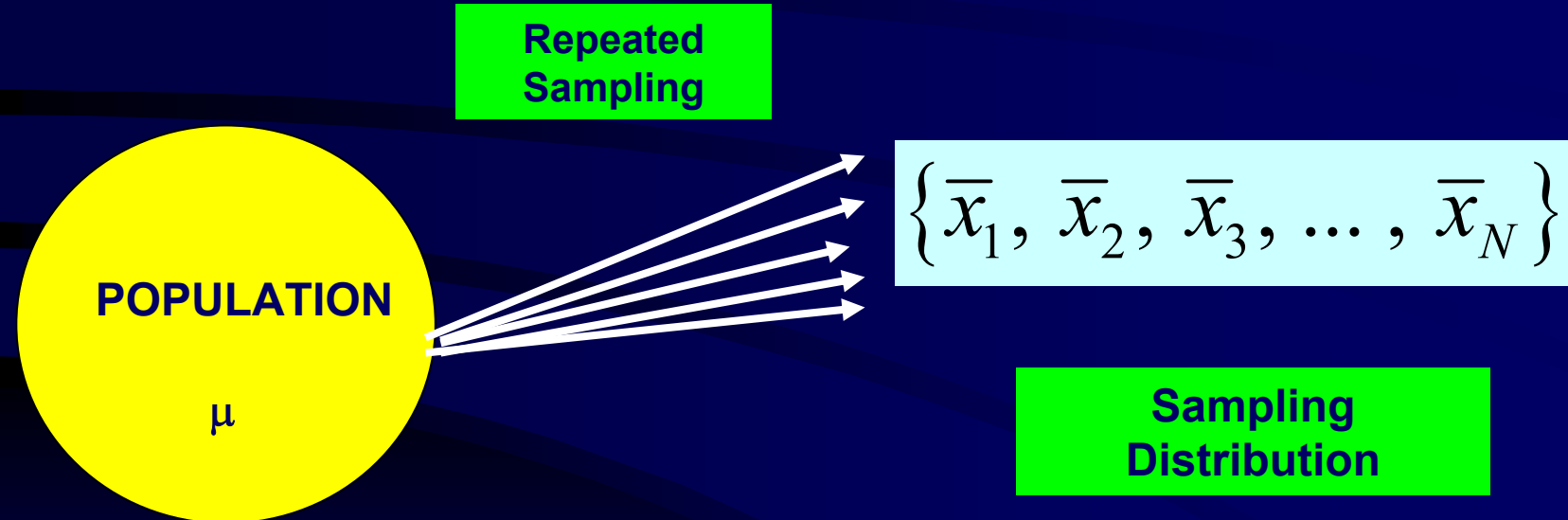
The Sampling Process



The Concept of a Sampling Distribution

- Repeated samples of the same size are selected from the same population.
- The same sample statistic is calculated from the data in EACH sample.
- The distribution of the sample statistics is the SAMPLING DISTRIBUTION of that sample statistic.

The Sampling Distribution



Using the TI-83/83+

- We will create a distribution of random integers from 1 to 100.
- Any population created using randInt(should be approximately uniform, that is, should have a histogram that is relatively rectangular.
- We will then create 50 samples of size 5 each from the population.

Instructions for the TI-83/83+

- Load L1, L2, ..., L5 with `randInt(1,100,50)`
- To check the shape of the population (uniform distribution), display the histogram of L1.
- Consider each row as a sample of size 5.
- Calculate the mean of each sample by loading $L6 = (L1+L2+L3+L4+L5)/5$

Analysis of the 50 “row” samples

- Display the histogram of the 50 sample means.
- Calculate summary statistics of the set of sample means.
- The mean of the uniform distribution of integers from 1 to 100 is 50. Compare this value to your mean of sample means.