



Computer  
Science

# **CSC196: Analyzing Data**

## **Sampling Distributions**

**Jason Pacheco and Cesim Erten**

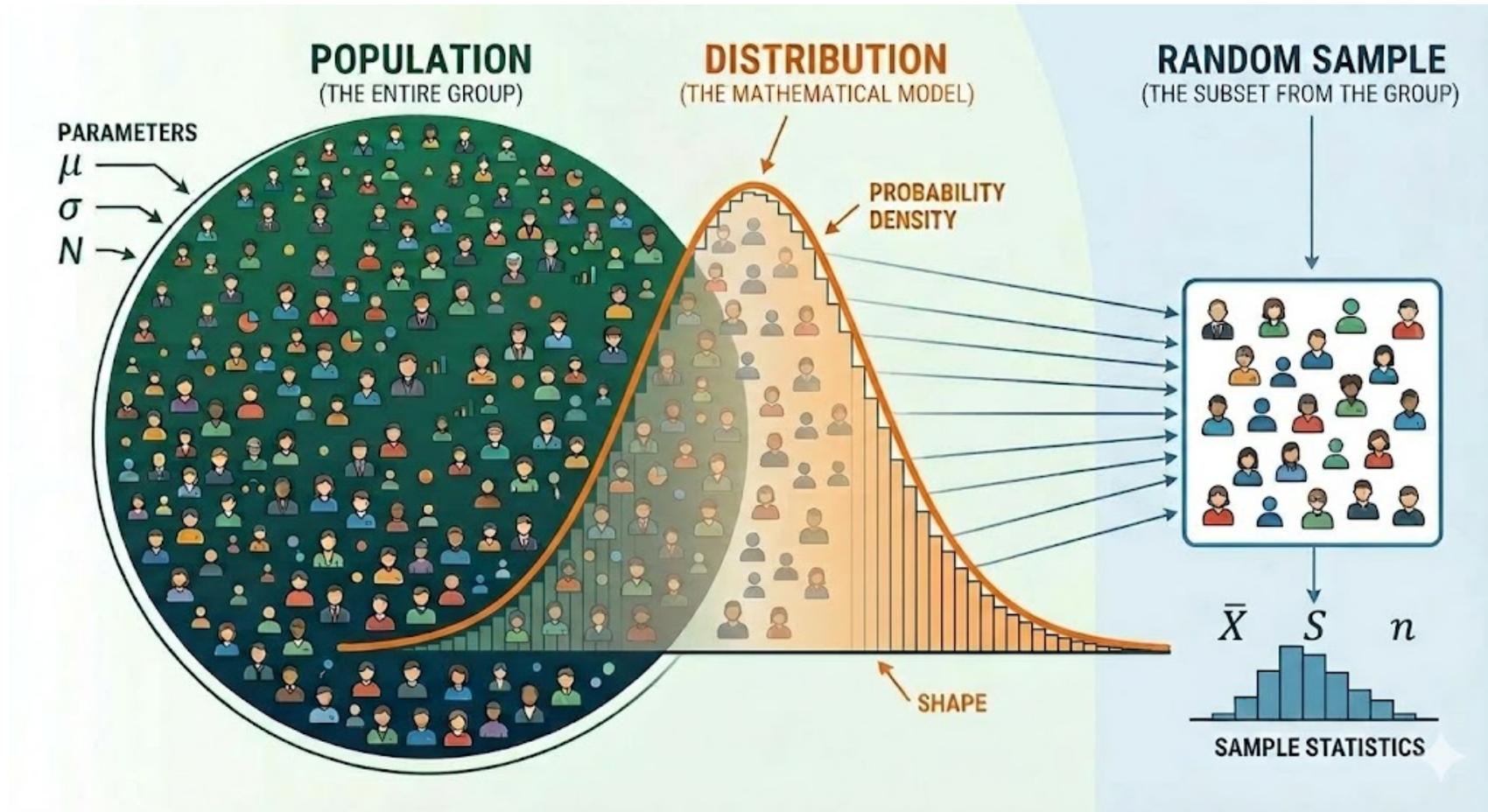
# QUIZ

Draw two normal distributions with the same mean  $\mu_1 = \mu_2$  and  $\sigma_1 < \sigma_2$  on the same plot.

# Outline

- Random Sampling
- Sampling Distributions
- Central Limit Theorem and Sampling Distribution of Means

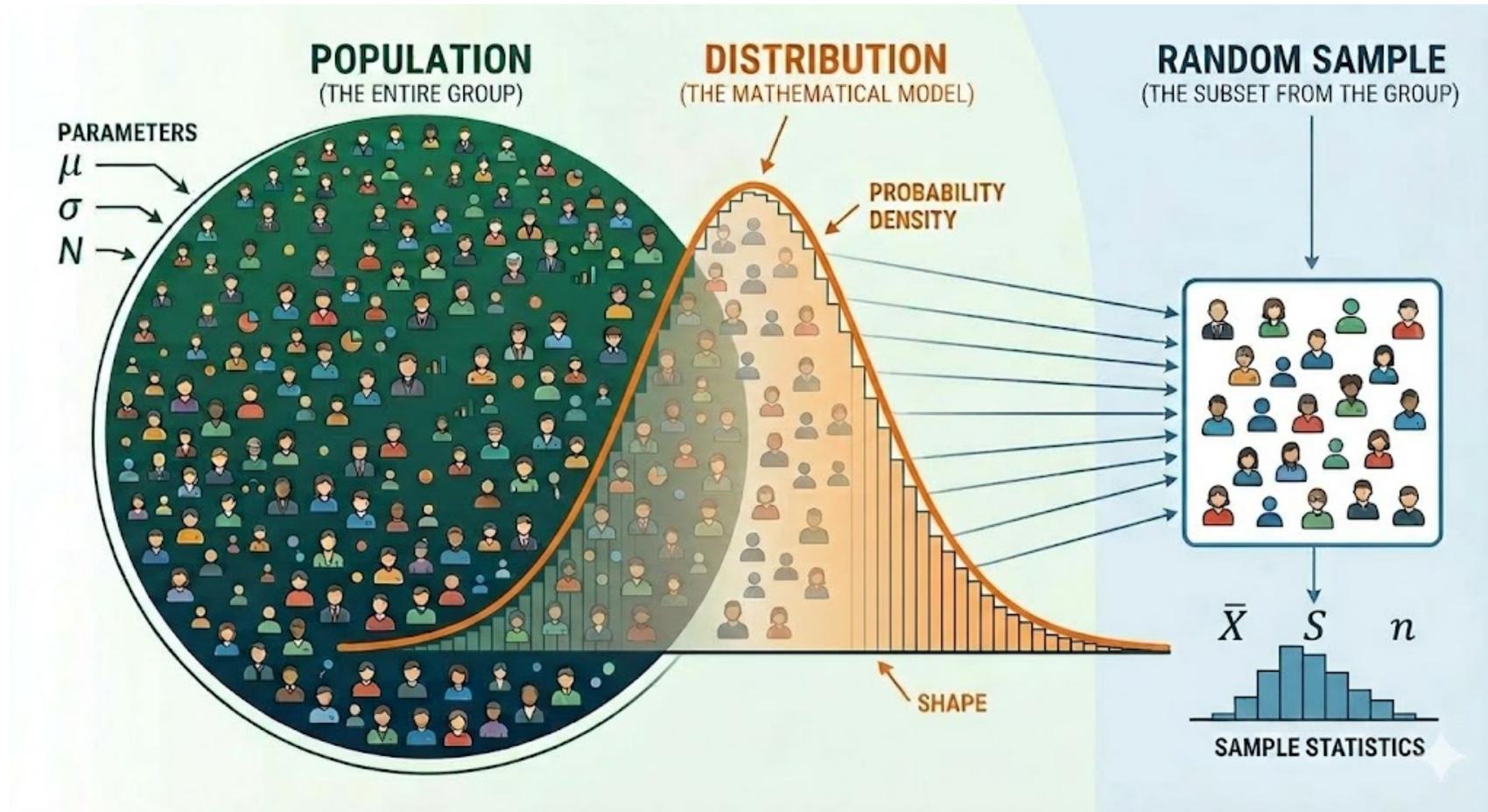
# Random Sampling



A **population** is the totality of the observations with which we are concerned.

“Normal population”  $\equiv$  A population whose observations are values of a random variable having a normal distribution

# Random Sampling



A **sample** is a subset of a population.

In a **random sample** observations are made independently and at random:

Independent random variables  $X_1, X_2, \dots, X_n$  each with the same probability distribution  $f(x)$ .

# Random Sampling

Any function of the random variables constituting a random sample is called a **statistic**. Let's review some statistics from initial lectures:

Let  $X_1, \dots, X_n$  iid (independently, identically distributed) random variables:

Sample mean:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Sample median:  $\tilde{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{otherwise} \end{cases}$

Sample mode: Value in the sample that occurs the most.

# Random Sampling

Any function of the random variables constituting a random sample is called a **statistic**. Let's review some statistics from initial lectures:

Let  $X_1, \dots, X_n$  iid (independently, identically distributed) random variables:

Sample variance: 
$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Sample standard deviation: 
$$S = \sqrt{S^2}$$

Sample range: 
$$X_{max} - X_{min}$$

# Sampling Distributions and Sampling Distribution of Means

Probability distribution of a statistic is called a **sampling distribution**.

Let  $X_1, \dots, X_n$  iid with mean  $\mu$  and variance  $\sigma^2$ .

Probability distribution of  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  is **sampling distribution of the mean**.

Note that  $\bar{X}$  is a random variable:

$$E(\bar{X}) =$$

$$Var(\bar{X}) =$$

# Examples Combining Sample Mean with Chebyshev's

**Review:**  $X$  a random variable  $\Rightarrow \forall t > 0, P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$

# Examples Combining Sample Mean with Chebyshev's

**Example:** Poll 100,000 people. Assume each person in the population votes for candidate  $C$  independently with probability  $p$ .

Bound probability that poll is off by  $\geq 1\%$  ?

Use Chebyshev's Theorem.

**Solution:**

First find maximum possible variance of a Bernoulli RV.