Module 03
Statistical Inference 1

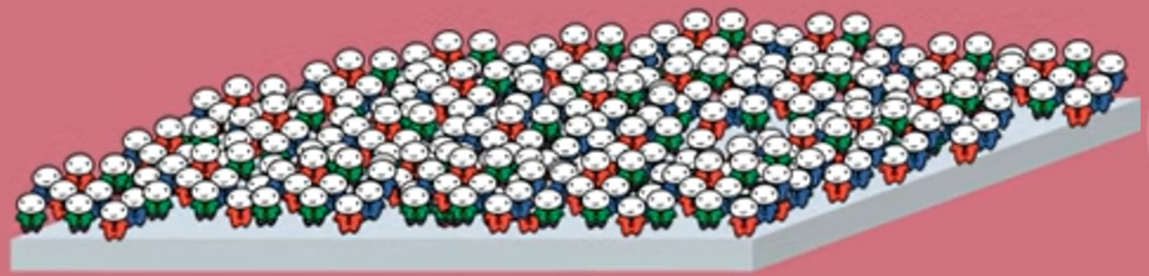
- Sampling distribution - by Shreya Sarjare
- Standard error - by Shreelekha Gundal
- Testing of hypotheses - by Shreevalli Gundal

SAMPLING DISTRIBUTION

SAMPLE DISTRIBUTION vs SAMPLING DISTRIBUTION

Sample

> A small portion of the population that we examine and draw conclusions from



Population = 10,000

μ = 5'4

$\bar{x}$ = 5'3

$\bar{x}$ = 5'7

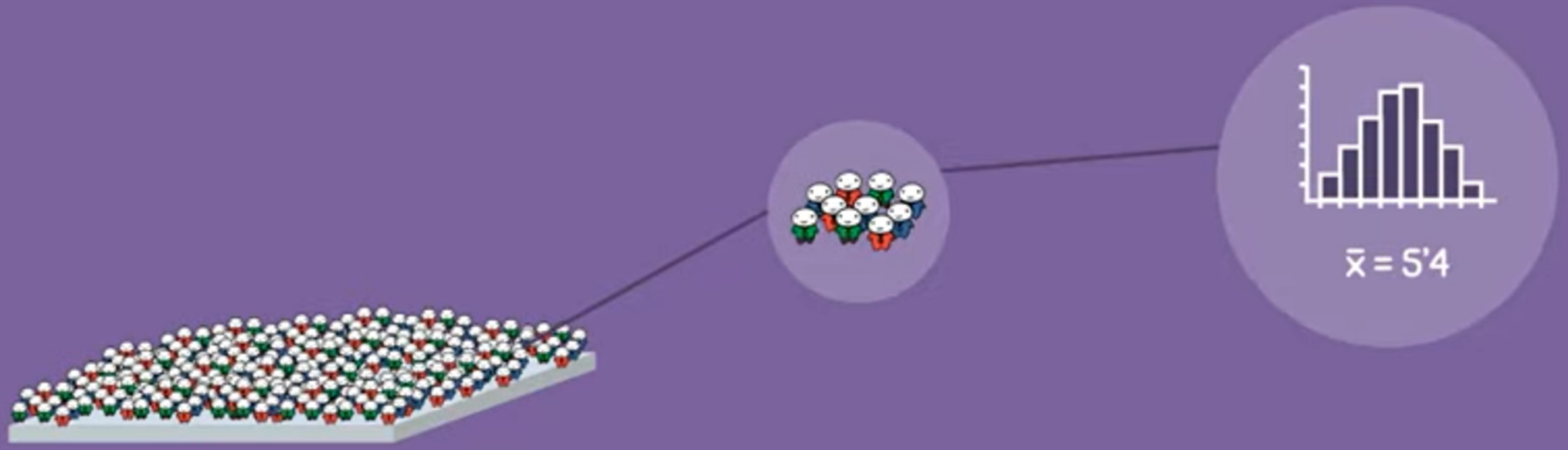
$\bar{x}$ = 5'4

Population = 10,000

μ = 5'4

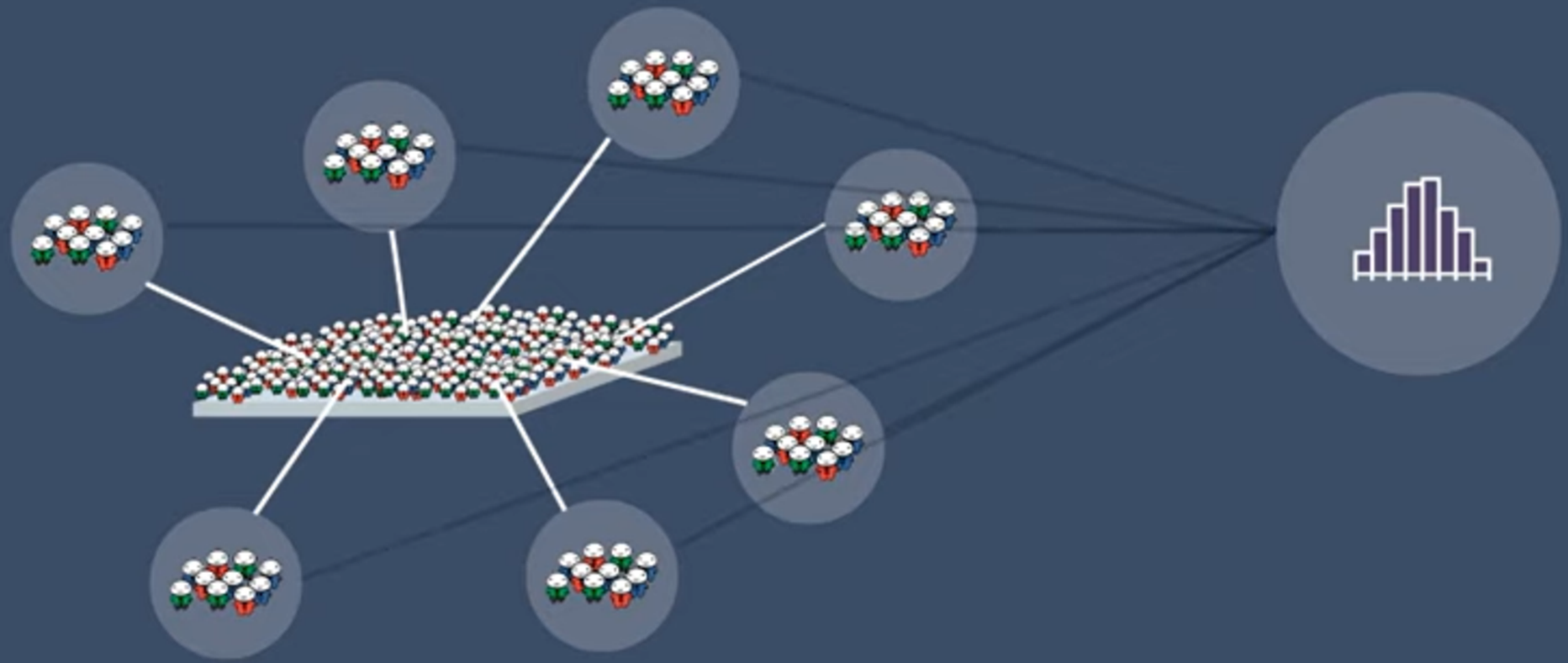
Sample Distribution

> Involves taking a singular sample from a population, and interpreting the data



Sampling Distribution

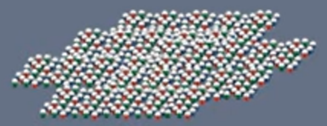
> A distribution of a statistic made from multiple simple random samples drawn from a specific population



Sampling Distribution of the Sample Mean

1

Population



Variable: Height

2

Random Sample

n = 5

104 cm 120 cm 110 cm

142 cm 169 cm

3

Calculate $\bar{x}$

$$= \frac{104 + 120 + 110 + 142 + 169}{5}$$

$$= 129 \text{ cm}$$

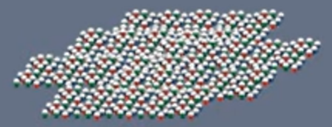
$$\bar{x} = 129 \text{ cm}$$

4

Frequency Distribution

1

Population



Variable: Height

2

Random Sample

n = 5

177 cm 105 cm 110 cm

115 cm 122 cm

3

Calculate $\bar{x}$

$$= \frac{177 + 105 + 110 + 115 + 122}{5}$$

$$= 125.8 \text{ cm}$$

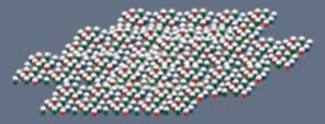
$$\bar{x} = 125.8 \text{ cm}$$

4

Frequency Distribution

Sampling Distribution of the Sample Mean

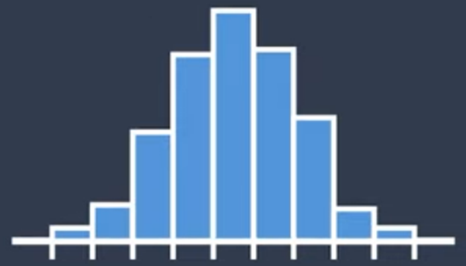
1 **Population**



2 **Random Sample**

3 **Calculate $\bar{x}$**

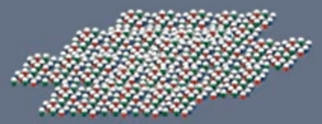
4 **Frequency Distribution**



Variable: Height

Sampling Distribution

1 **Population**



2 **Random Sample**

3 **Calculate $\bar{x}$**

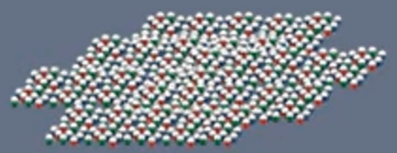
4 **Frequency Distribution**

Variable: Height

Sampling Distribution

Sampling Distribution of the Sample Mean

1. Population



Variable: Height

2. Random Sample

3. Calculate $\bar{x}$

4. Frequency Distribution

Normal Distribution

- **x -random variable**

Population Distribution

population distribution is a distribution that is created for measuring every single individual in the population

$X \sim N(\mu, \sigma)$

Mean

μ

Standard deviation

σ

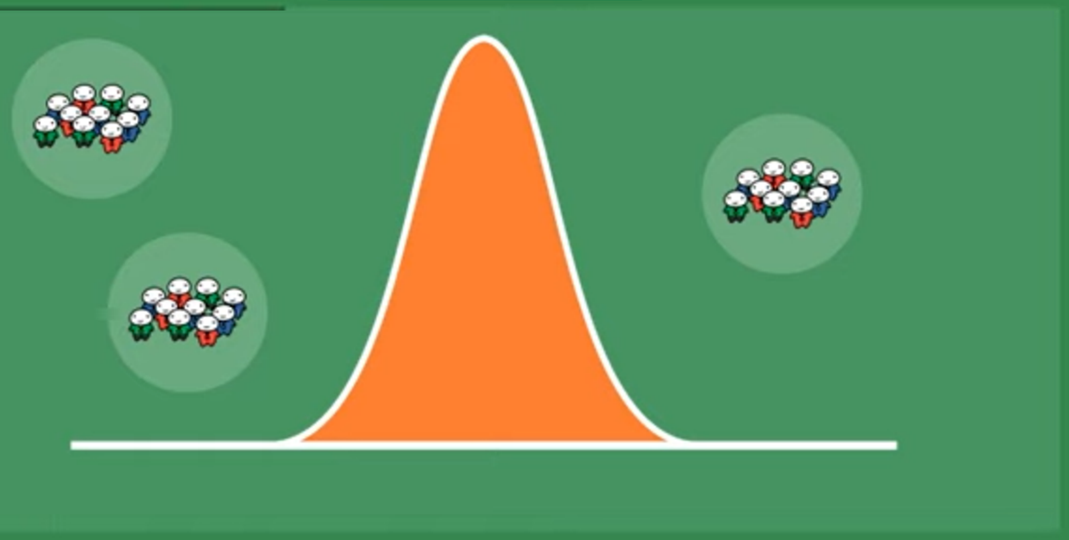
Standardization formula

$$z = \frac{x - \mu}{\sigma}$$

$\sigma_{\bar{x}} < \sigma$

Sampling Distribution

a sampling distribution involves repeatedly taking a sample and calculating a statistic for each individual sample and then combining that information to create a distribution



Mean

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

Standard deviation

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

Standardization formula

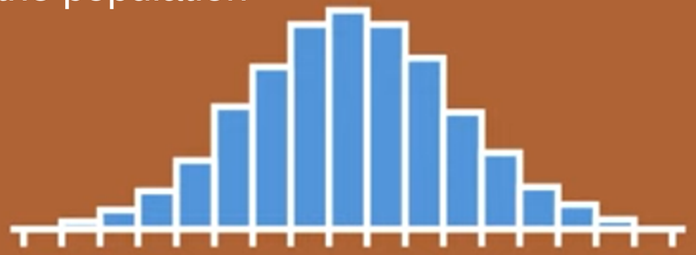
$$z = \frac{x - \mu}{\sigma / \sqrt{n}}$$

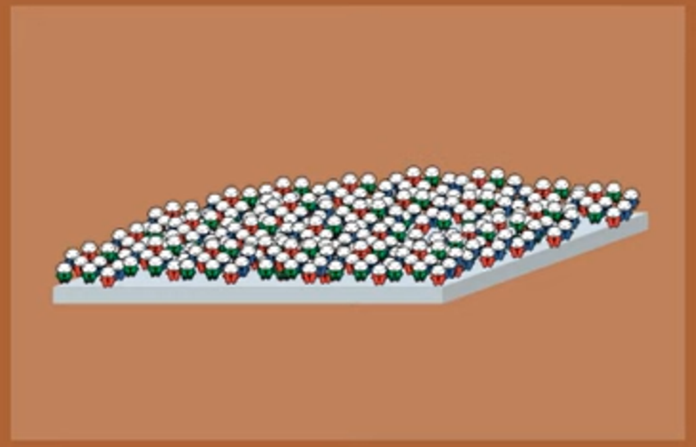
Sample Distribution

a sample distribution is a distribution that is created for measuring every single individual in this sample

Population Distribution

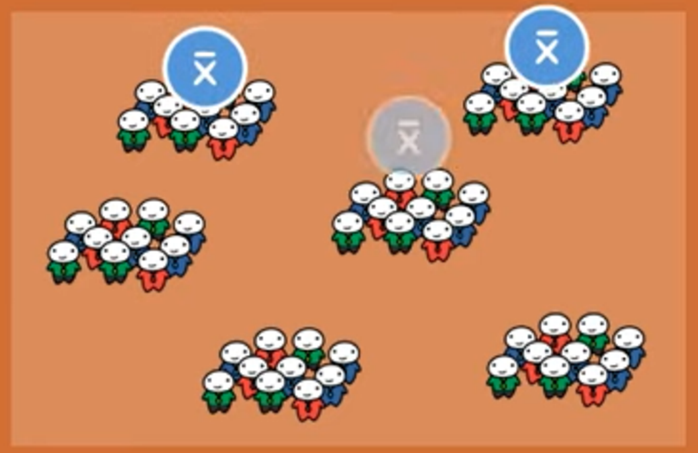
population distribution is a distribution that is created for measuring every single individual in the population





Sampling Distribution

a sampling distribution involves repeatedly taking a sample and calculating a statistic for each individual sample and then combining that information to create a distribution



Standard Error

In statistics, the standard error (SE) helps us understand how much the sample mean (average of a sample) is likely to differ from the true population mean. Essentially, SE tells us how precise our sample mean is in estimating the population mean.

Formula to calculate standard error:

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}$$

Where

- 'σ' is the population standard deviation, and
- 'n' is the sample size.

Formula to findout SE of mean :

When the population standard deviation σ is unknown, we estimate SE using the sample standard deviation 's:'

$$\mathrm{SE} = \frac{s}{\sqrt{n}}$$

Steps to calculate standard error:

1) Calculate the Sample Mean (x)
2) Calculate the Sample Standard Deviation (s)
3)Calculate the Standard Error (SE)

Question:

A teacher wants to estimate the average marks of all the students in her class. She takes a sample of 5 students and records the following marks: 70, 75, 80, 85, and 90. Calculate the sample mean, sample standard deviation, and the standard error of the mean.

solution:

Step 1: sample mean is the average of all marks

$$\bar{x} = \frac{70 + 75 + 80 + 85 + 90}{5} = \frac{400}{5} = 80$$

So, the sample mean is 80.

Step 2: Calculate the Sample Standard Deviation (s)

Find the deviations from the mean, square them, sum them up, divide by (n-1) and put square root.

1)

- $70 - 80 = -10 \rightarrow (-10)^2 = 100$
- $75 - 80 = -5 \rightarrow (-5)^2 = 25$
- $80 - 80 = 0 \rightarrow (0)^2 = 0$
- $85 - 80 = 5 \rightarrow (5)^2 = 25$
- $90 - 80 = 10 \rightarrow (10)^2 = 100$

2)

$$100 + 25 + 0 + 25 + 100 = 250$$

3)

$$\frac{250}{4} = 62.5$$

(n-1)=(5-1)=4

4)

$$s = \sqrt{62.5} \approx 7.91$$

Step 3: Calculate the Standard Error (SE)

$$\text{SE} = \frac{s}{\sqrt{n}} = \frac{7.91}{\sqrt{5}} = \frac{7.91}{2.236} \approx 3.54$$

So, the standard error is approximately 3.54

Testing of Hypothesis:

Hypothesis:

A Hypothesis is a proposed explanation or prediction that can be tested through experiments or observations. It serves as a starting point for research.

There are two types of hypotheses:

- Null Hypothesis (H_0)
- Alternative Hypothesis (H_1)

The Null Hypothesis

Definition

This is a statement of no effect, no difference, or no relationship. It is the hypothesis that we initially assume to be true.

It is often the hypothesis that we start with or assume to be true.

Symbol

It’s typically denoted as H_0.

Example

The average height of students in a class is 160 cm.

Purpose

The goal of hypothesis testing is usually to test whether there is enough evidence to reject the null hypothesis

Alternative Hypotheses

Definition

The alternative hypothesis is the opposite of the null hypothesis. It is a statement that there is an effect, there is a difference, or there is a relationship in the population. It represents what you aim to prove.

Example

The average height of students in a class is not 160 cm.

Purpose

The purpose of stating an alternative hypothesis is to compare it with the null hypothesis and determine if there is enough evidence to support the rejection of the null hypothesis in favor of the alternative hypothesis.

Symbol

The alternative hypothesis is typically denoted as H_1.

Example to make it clear

.

Let’s say you want to test if a new teaching method improves student performance

Null Hypothesis (H_0): “The new teaching method does not improve student performance.” This is the starting point. We assume that the new method has no effect.

Alternative Hypothesis (H_1): “The new teaching method does improve student performance.” This is what you’re trying to prove. You want to find out if the new method does make a difference.

The goal is to see if there is enough evidence to reject the null hypothesis and support the alternative hypothesis

Significance levels

Definition : The significance level is the probability of rejecting a null hypothesis when it is true.

Common values : 1%, 5%

How to use:Compare the p-value to the significance level.

When to accept the null hypothesis : If $p < \alpha$, accept the null hypothesis

When to reject the null hypothesis : If $p \geq \alpha$, reject the null hypothesis

Types of Errors in Hypothesis Testing

:

1 **Type I Error(false positive)**

Rejecting the Null hypothesis when it is true.

2 **Type II Error(false negative)**

Failing to reject the null hypothesis when it is false.(OR accepting the false Null hypothesis)

Type I Error

Definition

This happens when we *reject* the null hypothesis when it is actually *true.*

Type II Error

Definition

Failing to reject the null hypothesis when it is actually false.(OR we may *accept the false* null hypothesis)

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	**Type I error** False positive Probability = α	**Correct decision** True positive Probability = 1 - β
Not rejected	**Correct decision** True negative Probability = 1 - α	**Type II error** False negative Probability = β

Steps of testing the hypothesis

The process generally follows these steps:

- *Formulate the hypotheses:* Define the null and alternative hypotheses
- *Set the significance level (α):* This is the probability of rejecting the null hypothesis when it is actually true (commonly set at 0.05).
- *Choose the appropriate test:* Depending on the type of data and research question, you select a statistical test (e.g., t-test, chi-square test, ANOVA).
- *Collect and analyze data:* Obtain a sample and compute the test statistic.
- *Make a decision:* Compare the test statistic to a critical value or use a p-value to decide whether to reject or fail to reject the null hypothesis

Thank You !