



# STATISTICS FOR DATA SCIENCE

## HYPOTHESIS and INFERENCE

---

**Dr. Deepa Nair**  
Department of Science and Humanities

# STATISTICS FOR DATA SCIENCE

---

## UNIT-4      HYPOTHESIS and INFERENCE

### Session-1

### INTRODUCTION

**Dr. Deepa Nair**

Department of Science and Humanities

# STATISTICS FOR DATA SCIENCE

## Introduction

---



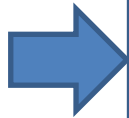
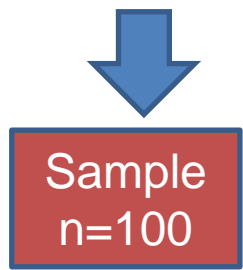
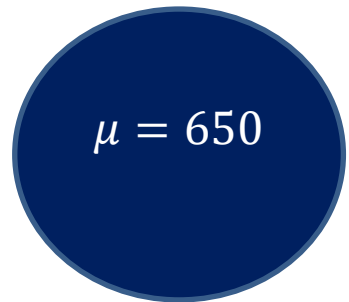
**Hypothesis Test:**

**Let us begin with an example**

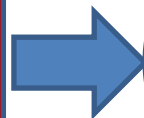
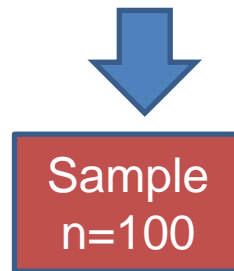
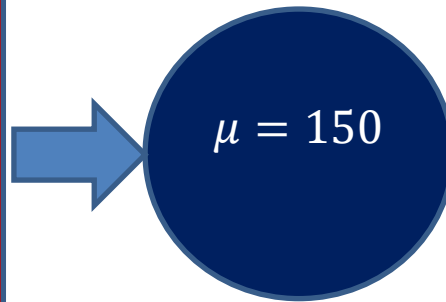
# STATISTICS FOR DATA SCIENCE

## Introduction

**bore well water  
(TDS)**



**water after  
filtration**



Is it plausible that this sample, with its mean of 150, could have come from a population whose mean is 650 or more?

# STATISTICS FOR DATA SCIENCE

## Introduction

---



- A hypothesis test produces a number between 0 and 1.
- That measures the degree of certainty we may have in the truth of a hypothesis about a quantity such as a population mean or proportion.

▪

# STATISTICS FOR DATA SCIENCE

## Introduction

---



- It turns out that hypothesis tests are closely related to confidence intervals.
- In general, whenever a confidence interval can be computed, a hypothesis test can also be performed, and vice versa.

# STATISTICS FOR DATA SCIENCE

## Introduction

---



- A Hypothesis is an assumption about population Parameter

# STATISTICS FOR DATA SCIENCE

## Introduction

---



### Types of Statistical Hypotheses:

There are two types of statistical hypotheses:

#### Null hypothesis:

The null hypothesis, denoted by  $H_0$ , is usually the hypothesis that sample observations result purely from chance



# STATISTICS FOR DATA SCIENCE

## Introduction

---



### Types of Statistical Hypotheses:

#### Alternative hypothesis:

The alternative hypothesis, denoted by  $H_1$  or  $H_a$ , is the hypothesis that sample observations are influenced by some non-random cause (investigator's belief).

# STATISTICS FOR DATA SCIENCE

## Introduction

---



### Example:

Do men and women having different average salaries after graduating University?

# STATISTICS FOR DATA SCIENCE

## Introduction

---



**Solution:**

$$H_0: \mu_M = \mu_F$$

$$H_1: \mu_M \neq \mu_F:$$

# STATISTICS FOR DATA SCIENCE

## Introduction

---



### Example:

A coin was flipped 50 times, resulting in 40 Heads and 10 Tails.

# STATISTICS FOR DATA SCIENCE

## Introduction

---



**Solution:**

$$H_0: p = 0.5$$

$$H_1: p \neq 0.5$$

# STATISTICS FOR DATA SCIENCE

## Introduction

---



### Example:

A new type of battery will be installed in heart pacemakers if it can be shown to have a mean lifetime is greater than eight years.

# STATISTICS FOR DATA SCIENCE

## Introduction

---



**Solution:**

$$H_0: \mu \leq 8$$

$$H_1: \mu > 8$$

# STATISTICS FOR DATA SCIENCE

## Introduction

---

- A parameter is the characteristic of the population  
Like its mean or variance
- The parameter must be identified before the analysis.

I Assume the average weight  
Of this class is 58 KG





# STATISTICS FOR DATA SCIENCE

## Introduction

---



- The best way to determine whether a statistical hypothesis is true would be to examine the entire population.
- Since that is often impractical, researchers typically examine a random sample from the population.
- If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.

# STATISTICS FOR DATA SCIENCE

## Introduction

---



In performing Hypothesis Test, we essentially put the null hypothesis on trial.

- We begin by assuming  $H_0$  is true.
- The random sample provides the evidence.
- The Hypothesis test involves measuring the strength of disagreement between the sample and  $H_0$ .

# STATISTICS FOR DATA SCIENCE

## Introduction

---



**Two methods :**

- 1) Traditional Method : Rejection region approach**
- 2) P-value approach (Used in book – Commonly used)**

# STATISTICS FOR DATA SCIENCE

## Introduction

---

### Hypothesis Testing



**Null Hypothesis:  $H_0$**



- State the Hypothesized Value of the parameter Before Sampling
- The assumption we wish To test.



**Alternate Hypothesis:  $H_1$**



- All [possible alternatives other than the null Hypothesis

# STATISTICS FOR DATA SCIENCE

## Introduction

### Steps:

Define  $H_0$  and  $H_1$ .



Assume  $H_0$  to be true



Compute a **test statistic**.



Compute the  $P$ -value of the test statistic.



State a conclusion about the strength of the evidence against  $H_0$ .



**Dr. Deepa Nair**

Department of Science and Humanities

---

**[deepanair@pes.edu](mailto:deepanair@pes.edu)**