Statistics

Lecture 01: |Introduction

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Intro

- The **collection and analysis** of data are **fundamental** to science and engineering
- A major difficulty with scientific data is that they are **subject to random variation**, or **uncertainty**.
  - When scientific measurements are repeated, they come out **somewhat differently each time**.
- This poses a problem: **How can one draw conclusions** from the results of an experiment when those results could have come out differently?

Intro

- To address this question, a **knowledge of statistics** is essential.
- **Statistics** is the field of study concerned with the collection, analysis, and interpretation of uncertain data.
- The methods of statistics allow scientists and engineers to **design** valid experiments and to **draw reliable conclusions** from the data they produce

Basic Idea

- The basic idea behind all statistical methods of data analysis is to make inferences about a population by studying a relatively small sample chosen from it
- Consider a machine that makes **steel rods** for use in optical storage devices. The specification for the diameter of the rods is **0.45 ± 0.02** cm. During the last hour, the machine has made **1000 rods**. The quality engineer **wants to know approximately how many of these rods meet the specification**. He does not have time to measure all 1000 rods. So he draws a **random sample of 50 rods**, measures them, and finds that **46 of them (92%) meet the diameter specification**

Basic Idea

- Now, it is unlikely that the sample of 50 rods represents the population of 1000 perfectly.
- The proportion of good rods in the population is likely to differ somewhat from the sample proportion of 92%.
- What the engineer needs to know is just how large that difference is likely to be. For example, is it plausible that the population percentage could be as high as 95%? 98%? As low as 90%? 85%?

Basic Idea

Here are some specific questions that the engineer might need to answer on the basis of these sample data:

- The engineer needs to compute a rough estimate of the likely size of the difference between the sample proportion and the population proportion. How large is a typical difference for this kind of sample?
  - Computation of a standard deviation

Basic Idea

Here are some specific questions that the engineer might need to answer on the basis of these sample data:

- Having observed that 92% of the sample rods were good, he will indicate the percentage of acceptable rods in the population as an interval of the form 92% ± x%, where x is a number calculated to provide reasonable certainty that the true population percentage is in the interval. How should x be calculated?
  - construction of a confidence interval

Basic Idea

Here are some specific questions that the engineer might need to answer on the basis of these sample data:

- The engineer wants to be fairly certain that the percentage of good rods is at least 90%; otherwise he will shut down the process for recalibration. How certain can he be that at least 90% of the 1000 rods are good?
  - hypothesis test

Main course concern

- The topics listed here concern methods of **drawing conclusions from data**.
- These methods form the field of **inferential statistics**
- We must first learn more about methods of **collecting data and of summarizing** clearly the basic information they contain.
- These are the topics of sampling and **descriptive statistics**

To conclude

- A numerical summary of a sample is called a **statistic**.
- A numerical summary of a population is called a **parameter**.
- **Statistics are often used to estimate parameter**s.