

Lecture 1

Data $\rightarrow$ Analysis $\rightarrow$ Utilizing Data

We begin by considering two primary modelling approaches:

- **Deterministic Models:** No uncertainty is involved. Examples include:

$$V = IR, \quad F = ma$$

- **Statistical Models:** Incorporate uncertainty. Examples include:

Relation between Temperature and Atmospheric Pressure

Relation between Weight and Height

Relation between Income and Expenditure

A statistical formulation can be expressed as:

$$w = f(u) + \epsilon$$

where ϵ represents **uncertainty**.

Remark 1. A *mathematical structure* is used to define a statistical model.

As behaviour changes, uncertainty tends to increase. The target in modelling is to:

1. Reduce error
2. Increase accuracy

1 Statistics

Statistics can be divided into three broad categories:

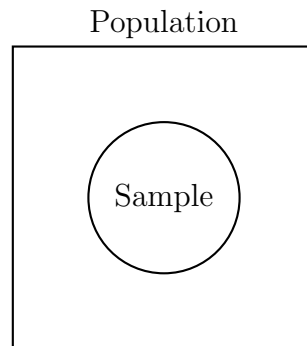
1.1 Descriptive Statistics

Provides a historical perspective of data. Examples include:

- Measures of central tendency: **Mean, Median, Mode**
- Measures of dispersion: **Range, Variance, Covariance**
- Data visualization: Bar Chart, Pie Chart, Box-plot, Scatter diagram

1.2 Predictive Statistics

- Analyzing historical data for prediction
- Using a **mathematical structure** for prediction to understand the behaviour of the entire population



After modelling, we test the model to ensure reliability.

1.3 Prescriptive / Inferential Statistics

Focused on:

- **Testing**
- **Validation**

2 Probability Theory

2.1 Understanding Probability

Probability is synonymous with **chance**.

Definition 1 (Empirical Definition). *Probability is defined as:*

$$\frac{\text{Favourable outcomes}}{\text{Total outcomes}}$$

This is the **classical** definition, not the rigorous mathematical one.

2.2 Mathematical Structure of Probability

A **Probability Function** can be illustrated using examples:

Example 1. *Tossing a coin $\Rightarrow P(\text{Head}) = \frac{1}{2}$ Throwing a die $\Rightarrow P(\text{Each face}) = \frac{1}{6}$*

2.3 Random Experiment

- Number of possible outcomes is fixed.
- In a **deterministic experiment**, the number of outcomes is fixed and known.
- For a random experiment, the set of outcomes is fixed, but the actual outcome is not fixed.
- Repetition under identical conditions is possible.

2.4 Sample Space

Definition 2. The **Sample Space** S is the set of all possible outcomes.

Example 2. 1. Weekday of your birthday: $S = \{\text{Mon}, \dots, \text{Sun}\}$

2. Height of a student in a class

3. Set of all even numbers

S may be finite, countably infinite, or uncountable. Example: $S = \{2, 4, 6, \dots\}$.

2.5 Events

Definition 3. An **Event** is any subset of the sample space S . The empty set $\emptyset$ is also an event.

2.6 Axiomatic Definition of Probability

A probability is a function:

$$P : F \rightarrow [0, 1]$$

satisfying:

1. $P(S) = 1$

2. If $A \cap B = \emptyset$, then:

$$P(A \cup B) = P(A) + P(B)$$

2.7 Power Set

The **Power Set** $\mathcal{P}(S)$ is the set of all subsets of S .

- Contains all subsets of S
- Contains $\emptyset$ and S
- If $A \in \mathcal{P}(S)$, then $A^c \in \mathcal{P}(S)$

If $|S| = \infty$, then:

$$|\mathcal{P}(S)| = 2^\infty = \infty$$