Introduction to Statistics

Understanding quantitative analysis, interpretation, and decision-making in modern business and academic contexts

Etymology: The Roots of Statistics

1

Latin

Statisticum

State affairs

2

Italian

Statistia

Political state

3

German

Statistik

Numerical data

4

Modern Meaning

Quantitative information of human interest

numerical / quantitative

What Statistics Means Today

Statistics represents **quantitative and numerical data** used for:

- Systematic calculation
- In-depth analysis
- Evidence-based decision-making

Two Fundamental Types of Data

Qualitative Data

Descriptive information

Categories and characteristics

Non-numerical attributes

Quantitative Data

Measurable information

Numerical values

Can be calculated

The Statistical Analysis Process

01

Collection of Data

Gathering information from reliable sources and establishing foundation

02

Classification & Tabulation

Organising raw data into structured, comparable formats

03

Analysis

Applying formulas, charts, and statistical methods to reveal patterns

04

Interpretation

Drawing conclusions and translating findings into actionable insights

Phase 1: Collection of Data

The **foundation** of all statistical analysis

- Identify reliable sources
- Ensure data accuracy
- Establish sampling methods
- Document collection process

Phase 2: Classification & Tabulation

Transform raw data into **organised, structured information**

- Group similar data points together
- Create meaningful categories
- Enable comparison and pattern recognition

Phase 3: Analysis of Data

Transform numbers into insights

- Apply statistical formulas
- Create visual representations
- Identify trends and patterns
- Test hypotheses

Phase 4: Interpretation of Data

> The final step: translating analytical results into **meaningful conclusions** and actionable recommendations

Why Statistics Matters

Business & Economics

Market analysis, forecasting, strategic planning

Medicine & Healthcare

Clinical trials, disease tracking, treatment efficacy

Weather & Climate

Prediction models, pattern analysis, forecasting

Finance & Banking

Risk assessment, investment decisions, credit evaluation

Sports Analytics

Performance tracking, strategy optimisation, predictions

Core Functions of Statistics

Presents Facts Clearly

Transforms complex information into understandable formats

Simplifies Complexity

Reduces large datasets into manageable summaries

Enables Comparison

Facilitates meaningful comparisons across groups and time

Supports Forecasting

Predicts future trends based on historical patterns

Statistics in Business & Economics

Essential for strategic decision-making

- Demand forecasting
- Market trend analysis
- Consumer behaviour patterns
- Competitive intelligence
- Resource allocation

Applications in Medicine & Weather

Medical Research

Clinical trial design

Disease prediction models

Treatment effectiveness analysis

Epidemiological studies

Weather Forecasting

Climate pattern analysis

Predictive modelling

Extreme event forecasting

Long-term climate projections

Finance & Banking Applications

Risk Analysis

Assessing investment risks and market volatility

Credit Evaluation

Determining creditworthiness and loan eligibility

Portfolio Management

Optimising investment portfolios for maximum returns

Understanding Statistical Limitations

- **No Individual Study**
  Focuses on aggregates, not individual cases
- **Ignores Qualitative Data**
  Cannot analyse non-numerical information
- **Average-Based Results**
  May obscure important variations and outliers
- **Potential for Bias**
  Results depend on collection and analysis methods

Population vs Sample

Population

The **entire dataset** of interest

Complete collection of all elements

Often impractical to study fully

N

Sample

A **representative subset** of the population

Selected for practical study

Used to make inferences about population

n

Variates and Attributes

Variate	Attribute
Numerical characteristic that can be measured	**Qualitative characteristic** that describes categories
Examples: height, weight, temperature, income	Examples: gender, colour, nationality, occupation
Varies quantitatively	Varies qualitatively

Types of Quantitative Variables

Discrete Variables #1

Countable distinct values

Whole numbers only

Examples:

- Number of family members
- Cars owned
- Students in class

Continuous Variables #2

Measurable values on a scale

Any value within range (fraction / integer)

Examples:

- Height and weight
- Temperature
- Time duration

Sample vs Population Measures

Category	Mean	Standard Deviation	Size
Population	μ (mu)	σ (sigma)	N
Sample	$\bar{x}$ (x-bar)	s	n

Greek letters denote **population parameters**; Roman letters denote sample statistics

Primary Data

Collected firsthand by the researcher

Advantages:

- Highly accurate and reliable
- Tailored to specific needs
- Up-to-date information

Disadvantages:

- Time-consuming
- Costly to collect

Secondary Data

Information **previously collected** by other agencies or researchers

Advantages:

- Quick and economical
- Readily accessible
- Covers broader scope

Disadvantages:

- May lack precision for specific needs
- Potential accuracy concerns
- Possible outdated information

Primary vs Secondary Data

1

Primary Data

Original collection

More reliable and specific

Higher cost and time investment

2

Secondary Data

Derived from existing sources

More accessible and economical

Lower precision for specific research

Classification of Data

Arranging data into homogeneous groups to reveal patterns

- Identifies relationships between variables
- Facilitates meaningful comparisons
- Reveals hidden trends and patterns

Four Bases of Classification

Qualitative

Based on attributes or qualities

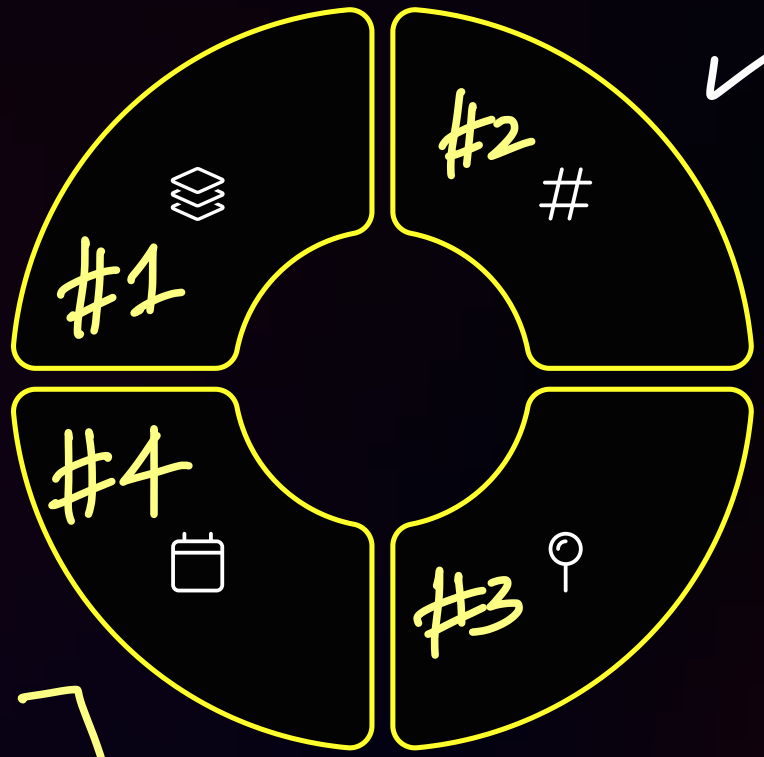
Example: gender, education level

Chronological

Based on time periods

Example: year, quarter

[Temporal Base]



Quantitative

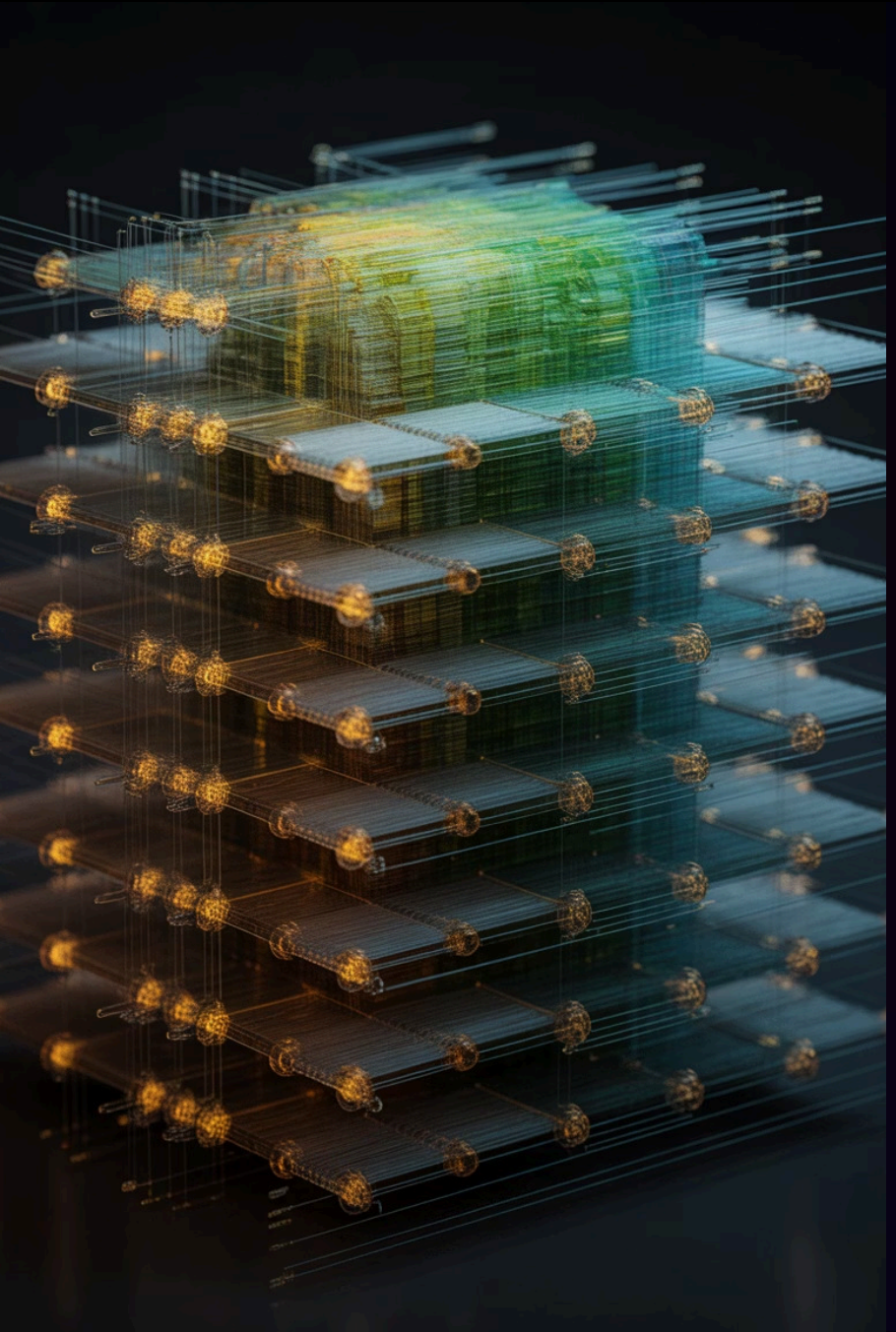
Based on numerical values

Example: age, income

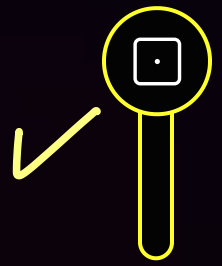
Geographical

Based on location or region

Example: country, city



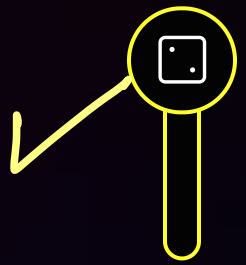
Types of Classification



One-Way Classification

Data classified by **single characteristic**

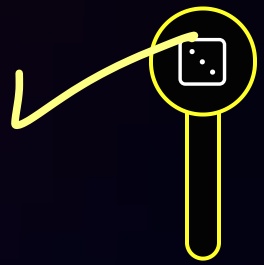
Example: Students by age



Two-Way Classification

Data classified by **two characteristics**

Example: Students by age and gender



Multi-Way Classification

Data classified by **more than two characteristics**

Example: Students by age, gender, and course

#9ml

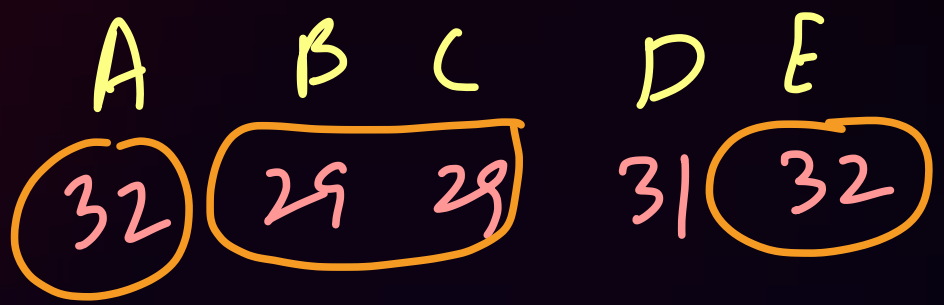
Frequency Distribution

Variable → discrete
→ Continuous

Shows **how often** each value occurs in a dataset

Used for:

- Discrete data (exact counts)
- Continuous data (grouped intervals)
- Identifying patterns and outliers
- Understanding data distribution



31	1
32	2
29	2

Class interval

LCL – UCL

Class Limits & Intervals

1

Lower Class Limit

Smallest value in the class

2

Upper Class Limit

Largest value in the class

3

Class Width

Upper limit − Lower limit

(41, 42)

60-69, 70-79

Class intervals → Exclusive
→ Inclusive

Mid-Value (Class Mark)

> The **midpoint** of a class interval

Formula

Class Mark = **(Lower Limit + Upper Limit) ÷ 2**

Purpose

Represents the central value of each class for calculations

0-9

0-10, 10-20, 20-30

LCL UCL

$\frac{10+0}{2} = 5$ $\frac{20+10}{2}$

15

NEW LCI = $5 - \frac{[10-9]}{2} = 4.5$

NEW UCI = OLD UCI + $\frac{10-9}{2}$

① = 9 + 0.5 = 9.5

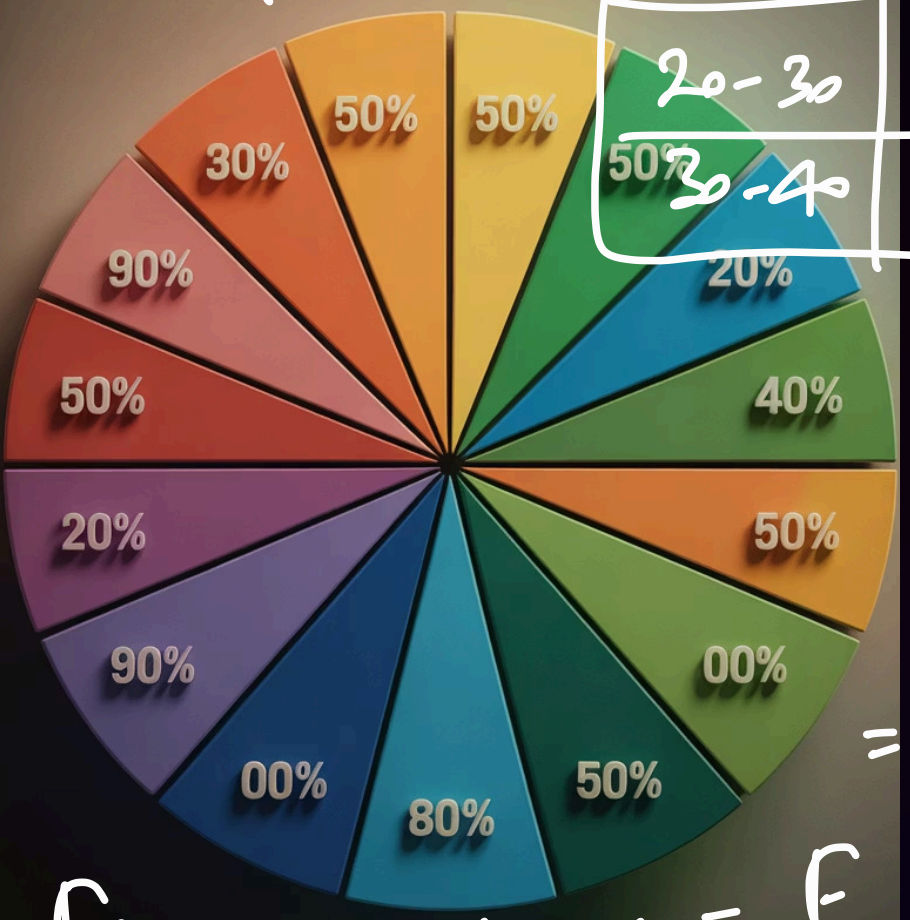
In → Ex

5-9, 10-14, 15-19

① ② [4.5-9.5]

C.W ⇒ 30-20
= 10

0-10	5
10-20	7
20-30	12
30-40	8



Frequency density = F / width = 12 / 10 = 1.2

Relative & Percentage Frequency

Relative Frequency

Formula: f / Σf

Proportion of observations in each class

Values between 0 and 1

f/Σf

20-30 → 12/32

Percentage Frequency

Formula: (f / Σf) × 100

Percentage of observations in each class

Values between 0% and 100%

12/32 × 100

Discrete vs Continuous Frequency Distributions

Discrete Distribution

Distinct, countable values

Each value listed separately

Example: Number of children per family

Continuous Distribution

Grouped into intervals or ranges

Data spans a continuum

Example: Height ranges (150–160 cm, 160–170 cm)

15-20 10
20-25 12
25-30 15
30-35

Both serve as the foundation for calculating further statistical measures like mean, median, and standard deviation