

Statistics



Data

Data Fundamentals	2
Data Types	2
Population vs. Sample Data	2
Visualizing Data	3

Descriptive Statistics

Data Normalization and Outliers

Probability Theory

Hypothesis Testing

T-Tests

Confidence Intervals

Correlation

Analysis of Variance

Regression

Statistical Power and Sample Sizes

Clustering and Dimension-Reduction

Signal Detection Theory

Data



Data Fundamentals

- **Data:** units qualitative or quantitative information about persons or objects collected via observation.
 - Note: data is different from information—information resolves uncertainty, while data has the potential to be transformed into information post-analysis.
 - Data as a general concept refers to the fact that some existing information or knowledge can be represented in a form suitable for processing.

Data Types

- Data types have two different general meanings:
 - **Data type (computer science):** involves the format of data storage and has implications on operations and storage space.
 - **Data type (statistics):** involves the category of data and has implications on the methods used for analysis.
- There are many data types, with more specific definitions than the following definitions, but for now these are frequently used and adequate for topics covered.

Relevant Statistical Data Types

Category	Type	Description	Example
Numerical	Interval	Degree of difference	Temperature
	Ratio	Interval + meaningful zero	Height
	Discrete	Count (integers)	Population
Categorical	Ordinal	Sortable, discrete	Educational level
	Nominal	Non-sortable, discrete	Movie genre

Population vs. Sample Data

- **Population data** μ : data from *all* members of a group.
- **Sample data** $\hat{\mu}$: data from a *subset* of members of a group (hopefully random).
- Statistical procedures generally are designed for sample or population data; wrong conclusions can be drawn if the distinction is not clear.
 - Note: most data are sample data in practice, as generalization of populations using sample data is usually the goal of statistics.

Anecdotes: a case study of a rare occurrence, or a sample size of only one; insights may be possible, but poor confidence in generalizations should be noted.

Visualizing Data

-

Descriptive Statistics



Data Normalization and Outliers



Probability Theory



Hypothesis Testing



T-Tests



Confidence Intervals



Correlation



Analysis of Variance



Regression



Statistical Power and Sample Sizes



Clustering and Dimension-Reduction



Signal Detection Theory

