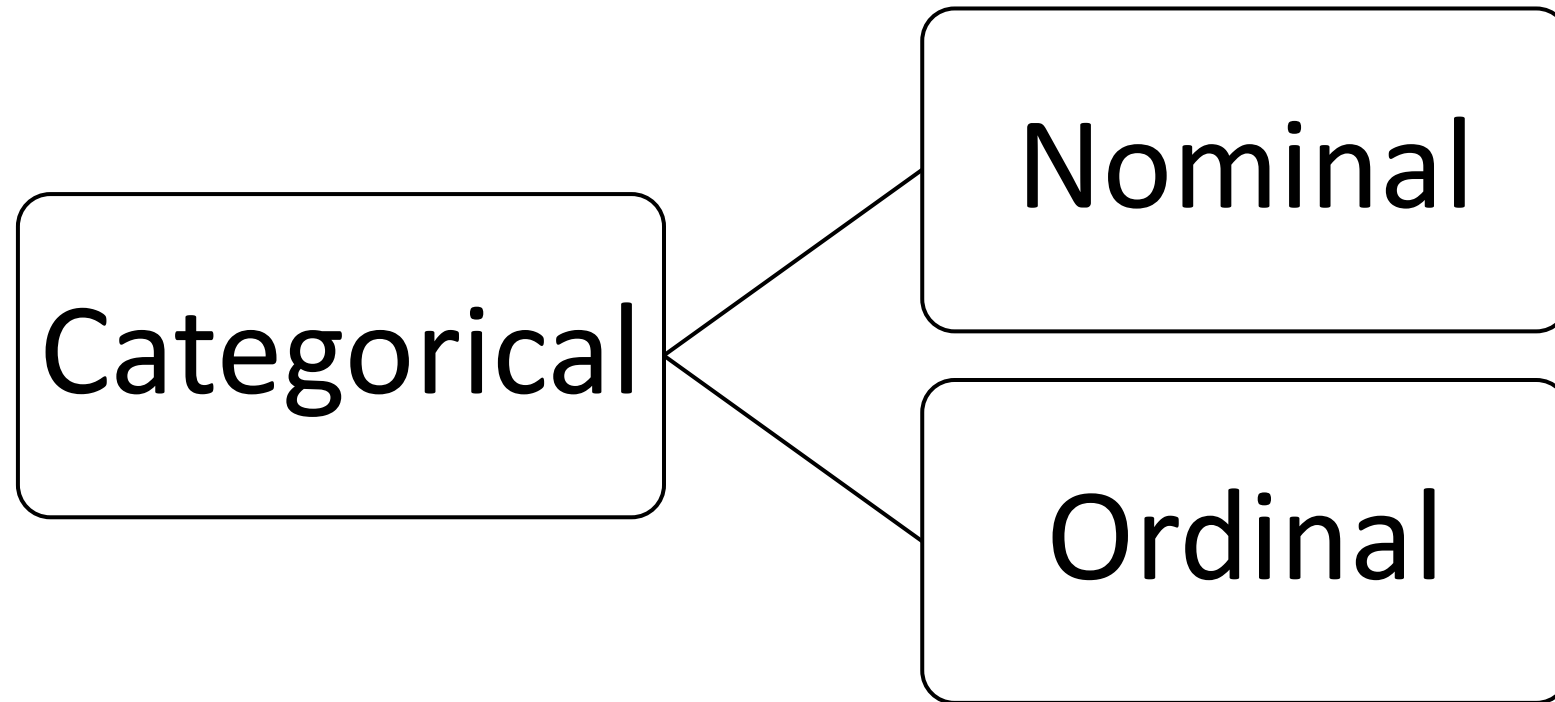
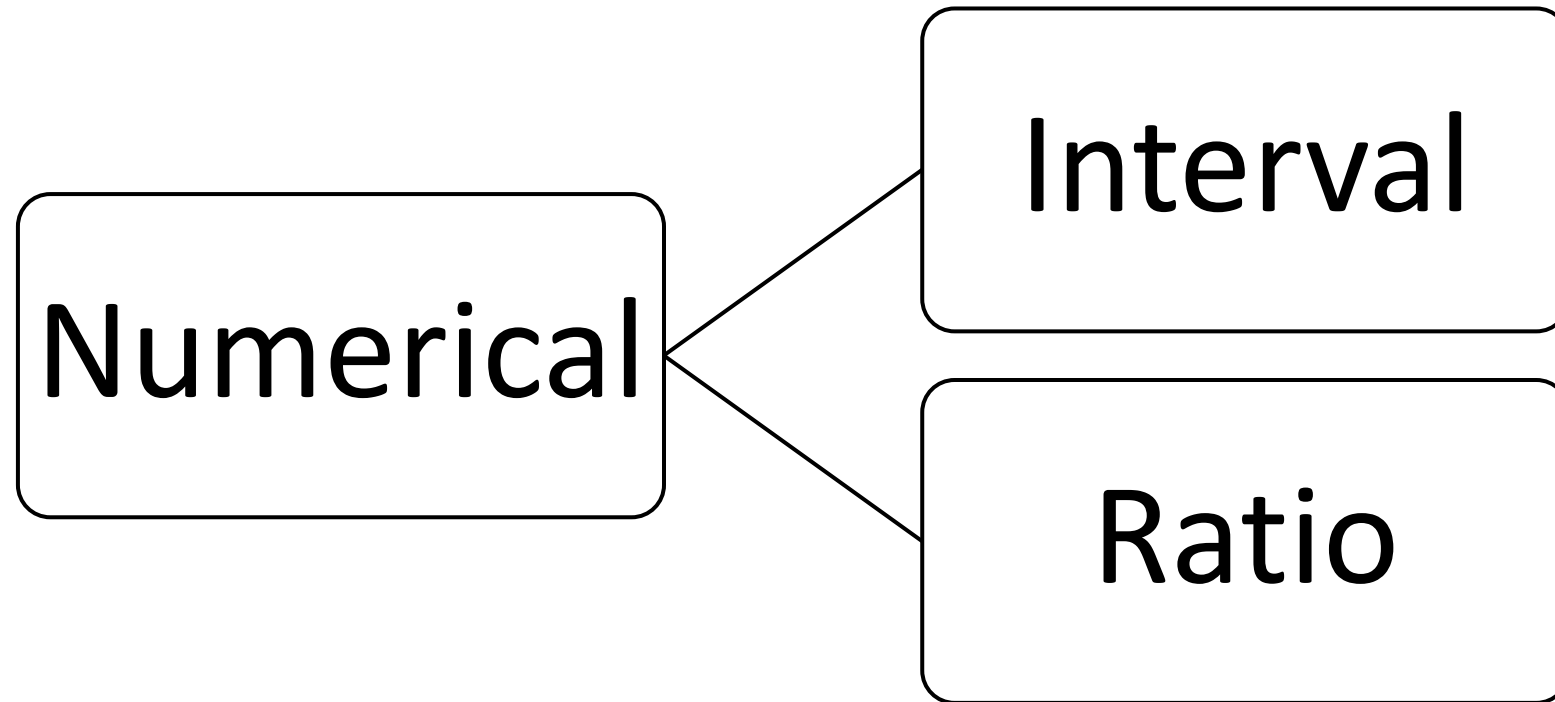


Data Types and More..

Classification of Variables



Classification of Variables



Categorical and Numerical Variables

Categorical variables produce responses that belong to groups or categories.

Eg. Gender, marital status, major, level of satisfaction (1: very satisfied, 2: satisfied, 3: neutral, 4: disappointed, 5: very disappointed)

Numerical variables produce responses that are either discrete or continuous numbers.

Eg. The number of units earned, number of stocks in a portfolio, temperature in F, height, weight, distance from campus

Measurement Levels

Nominal data are words that describe the categories or class of responses.

Ordinal data indicate the rank ordering of items of responses.

Interval data are obtained from numerical variables, whose differences are meaningful.

Ratio data are also numerical having meaningful absolute 0.

Nominal Data

- Nominal data refers to a type of categorical data where the values represent distinct categories without any intrinsic ordering or ranking.
- often used to label variables without implying any quantitative relationship.
- Each category is unique and doesn't carry any quantitative significance.

Key Characteristics of Nominal Data

- Categorical: That data can be divided into various categories or groups.
- No order or ranking: Unlike ordinal data, the categories cannot be logically ordered or ranked.
- Unique Labels: Each value is used to uniquely label a category or group.
- Mutually Exclusive: Each data point belongs to exactly one category.

Examples

- Gender: Categories like Male, Female, Non-binary.

Examples

- Gender: Categories like Male, Female, Non-binary.
- Colours: Categories like Red, Blue, Green etc.

Examples

- Gender: Categories like Male, Female, Non-binary.
- Colours: Categories like Red, Blue, Green etc.
- Country of origin: Categories like India, China etc.

Examples

- Gender: Categories like Male, Female, Non-binary.
- Colours: Categories like Red, Blue, Green etc.
- Country of origin: Categories like India, China etc.
- Marital status: Categories like Single, Married, Divorced etc.

Analysis Techniques for Nominal Data

- Frequency count: The number of times each category appears.
- Mode: The most frequent category.
- Chi-square test: To examine the relationship between two nominal variables.
- Bar charts/pie charts: To visually represent the distribution of categories.

Common Statistical Methods

- Cross tabulation: For analysing the relationship between two nominal variables.
- One-hot Encoding: A technique used in machine learning to convert nominal data into binary vectors.

Ordinal Data

- Ordinal data refers to a type of categorical data that represents variables with natural, meaningful order or ranking among the categories.

Ordinal Data

- Ordinal data refers to a type of categorical data that represents variables with natural, meaningful order or ranking among the categories.
- Intervals between the categories are not uniform or known.

Key Characteristics of Ordinal Data

- Categorical with Order: The data can be grouped into categories, and these categories have a logical order or ranking.
- Unknown Intervals: The exact differences between categories are not measurable or consistent.
- Relative Comparison: One can compare the order of the data points (e.g., greater than, less than), but cannot perform meaningful arithmetic operations like addition or subtraction.

Examples

- Socio-economic status: low, medium, high

Examples

- Socio-economic status: low, medium, high
- Education Level: High School, Bachelor's, Master's, Doctorate.

Examples

- Socio-economic status: low, medium, high
- Education Level: High School, Bachelor's, Master's, Doctorate.
- Customer Satisfaction: Very Unsatisfied, Unsatisfied, Neutral, Satisfied, Very Satisfied.

Examples

- Socio-economic status: low, medium, high
- Education Level: High School, Bachelor's, Master's, Doctorate.
- Customer Satisfaction: Very Unsatisfied, Unsatisfied, Neutral, Satisfied, Very Satisfied.
- Health Status: Poor, Fair, Good, Excellent.

Analysis Techniques for Ordinal Data

- Median and Mode: The median and mode can be used to summarize ordinal data.
- Rank Correlation: Tests like Spearman's rank correlation or Kendall's tau can measure the association between two ordinal variables.
- Non-Parametric Tests: Statistical methods like the Mann-Whitney U test, Wilcoxon signed-rank test, or Kruskal-Wallis test are appropriate for analyzing ordinal data.
- Cross-Tabulation and Chi-Square Test: To explore relationships between ordinal variables or between ordinal and nominal variables.

Visualization Techniques for Ordinal Data

- Bar Charts: To display the frequency of each category.
- Stacked Bar Charts: To compare ordinal data across groups.
- Box Plots: When numeric equivalents are assigned to ordinal data.

Practical Applications of Ordinal Data

- Surveys and Questionnaires: Likert scales (e.g., satisfaction ratings).
- Healthcare: Pain severity levels (mild, moderate, severe).
- Education: Academic grading systems (e.g., A, B, C, D, F).
- Market Research: Ranking preferences for products or services.

Interval Data

- Interval data is a type of quantitative data where the values are numeric, and the differences between values are meaningful and consistent.

Interval Data

- Interval data is a type of quantitative data where the values are numeric, and the differences between values are meaningful and consistent.
- However, interval data does not have a **true zero point**, meaning zero does not indicate the absence of the quantity being measured.

Key Characteristics of Interval Data

- Numeric: Data values are expressed in numbers.
- Equal Intervals: The differences between consecutive values are consistent and meaningful.
- No True Zero: A value of zero does not imply a complete absence of the measured property.
- Arithmetic Operations: Addition and subtraction are meaningful, but multiplication and division are not.

Examples

- Temperature: Measured in Celsius or Fahrenheit (e.g., 10°C, 20°C, 30°C).

Examples

- Temperature: Measured in Celsius or Fahrenheit (e.g., 10°C, 20°C, 30°C).
- Dates: Calendar years (e.g., 2000, 2010, 2020).

Examples

- Temperature: Measured in Celsius or Fahrenheit (e.g., 10°C, 20°C, 30°C).
- Dates: Calendar years (e.g., 2000, 2010, 2020).
- IQ Scores: Intelligence test scores.

Examples

- Temperature: Measured in Celsius or Fahrenheit (e.g., 10°C, 20°C, 30°C).
- Dates: Calendar years (e.g., 2000, 2010, 2020).
- IQ Scores: Intelligence test scores.
- Standardized Test Scores: SAT, GRE scores.

Analysis Techniques for Interval Data

➤ Descriptive statistics

- a) Mean, median, and mode to describe central tendency.
- b) Range, variance, and standard deviation to measure variability.

Analysis Techniques for Interval Data

➤ Descriptive statistics

- a) Mean, median, and mode to describe central tendency.
- b) Range, variance, and standard deviation to measure variability.

➤ Correlation and regression

- a) Pearson correlation for relationships between interval variables.
- b) Linear regression to model dependent and independent interval variables.

Analysis Techniques for Interval Data

➤ Descriptive statistics

- a) Mean, median, and mode to describe central tendency.
- b) Range, variance, and standard deviation to measure variability.

➤ Correlation and regression

- a) Pearson correlation for relationships between interval variables.
- b) Linear regression to model dependent and independent interval variables.

➤ Inferential statistics

- a) T-tests to compare means of two groups.
- b) ANOVA (Analysis of Variance) for more than two groups.

Visualization Techniques for Interval Data

- Histogram: To show the distribution of the data.
- Line Graphs: To observe trends over time or sequences.
- Scatterplots: To analyze relationships between two interval variables.

Practical Applications of Interval Data

- Temperature analyses: Comparing temperature variations without considering absolute absence of heat.
- Psychology and Education: Evaluating IQ or standardized test performance.
- Finance: Interest rate changes or year-over-year growth rates.

Ratio Data

- Ratio data is the most informative type of quantitative data, characterized by numeric values that have a true zero point.

Ratio Data

- Ratio data is the most informative type of quantitative data, characterized by numeric values that have a true zero point.
- One may perform mathematical operations, including multiplication and division.

Ratio Data

- Ratio data is the most informative type of quantitative data, characterized by numeric values that have a true zero point.
- One may perform mathematical operations, including multiplication and division.
- This type of data enables comparisons in terms of absolute magnitude (e.g., "twice as much").

Key Characteristics of Ratio Data

- Numeric: Data values are numbers, often representing measurements.
- Equal Intervals: The intervals between values are consistent and meaningful.
- True Zero: A value of zero indicates the complete absence of the measured property.
- Full Arithmetic Operations: Addition, subtraction, multiplication, and division are all valid.

Examples

- Height: Measured in centimeters, meters, or inches.

Examples

- Height: Measured in centimeters, meters, or inches.
- Weight: Measured in kilograms, pounds, or grams.

Examples

- Height: Measured in centimeters, meters, or inches.
- Weight: Measured in kilograms, pounds, or grams.
- Age: Measured in years or months.

Examples

- Height: Measured in centimeters, meters, or inches.
- Weight: Measured in kilograms, pounds, or grams.
- Age: Measured in years or months.
- Income: Measured in dollars, euros, or other currencies.

Examples

- Height: Measured in centimeters, meters, or inches.
- Weight: Measured in kilograms, pounds, or grams.
- Age: Measured in years or months.
- Income: Measured in dollars, euros, or other currencies.
- Distance: Measured in kilometers, miles, or meters.

Analysis Techniques for Ratio Data

➤ Descriptive statistics

- a) Mean, median, and mode to describe central tendency.
- b) Range, variance, and standard deviation to measure variability.

Analysis Techniques for Ratio Data

➤ Descriptive statistics

- a) Mean, median, and mode to describe central tendency.
- b) Range, variance, and standard deviation to measure variability.

➤ Inferential statistics

- a) T-tests to compare means of two groups.
- b) ANOVA (Analysis of Variance) for more than two groups.

Analysis Techniques for Ratio Data

➤ Descriptive statistics

- a) Mean, median, and mode to describe central tendency.
- b) Range, variance, and standard deviation to measure variability.

➤ Inferential statistics

- a) T-tests to compare means of two groups.
- b) ANOVA (Analysis of Variance) for more than two groups.

➤ Advanced analysis

- a) Proportions and percentages (e.g., part-to-whole comparisons).
- b) Geometric and harmonic means for specific contexts.

Visualization Techniques for Ratio Data

- Histogram: To show the distribution of the data.
- Box plots: For comparing distributions of multiple datasets.
- Scatterplots: To analyze relationships between ratio variables.

Practical Applications of Ratio Data

- Scientific Measurements: Measuring physical quantities like speed, mass, and volume.
- Finance and Economics: Income, profit, revenue, and expenditure analysis.
- Health and Medicine: BMI (Body Mass Index), blood pressure, or calorie intake.
- Engineering: Lengths, weights, and capacities in product design or manufacturing.

Comparison Between Data Types

Feature	Nominal Data	Ordinal Data	Interval Data	Ratio Data
Order	No	Yes	Yes	Yes
Equal Intervals	No	No	Yes	Yes
True Zero	No	No	No	Yes
Arithmetic	None	Limited (rank-based)	Add/Subtract	All operations