

Introduction to Descriptive Statistics

Introduction to Descriptive Statistics

Key concepts, tools, and practical steps for summarizing data

1. What Is Descriptive Statistics?

- **Definition** – A set of techniques used to **summarize, organize, and present** data in a meaningful way.
 - **Purpose** – Provide a quick overview of the main features of a dataset without drawing conclusions about a larger population.
 - **Contrast with Inferential Statistics** – Descriptive statistics describe the sample itself; inferential statistics use the sample to make predictions or generalizations about a population.
-

2. Types of Data

Data Type	Characteristics	Typical Measures
Nominal	Categories without order (e.g., gender, eye color)	Frequency, mode, proportion
Ordinal	Ordered categories (e.g., rating scales)	Median, percentiles, mode
Interval	Numerical values with equal intervals, no true zero (e.g., temperature in °C)	Mean, standard deviation, range
Ratio	Numerical values with a meaningful zero (e.g., height, weight)	All interval measures + coefficient of variation

3. Summarizing a Single Variable

3.1. Frequency Distribution

- **Absolute frequency (count)** – Number of observations in each category or class.
- **Relative frequency (proportion)** – Count divided by total observations.
- **Cumulative frequency** – Running total of frequencies; useful for percentiles.

Introduction to Descriptive Statistics

Example Table

Class / Category	Frequency (f)	Relative Frequency (f/N)	Cumulative Frequency
0-9	5	0.10	5
10-19	12	0.24	17
...	...	...	...

3.2. Measures of Central Tendency

Measure	Formula (for a sample)	When to Use
Mean	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	Symmetric, interval/ratio data, no extreme outliers
Median	Middle value after sorting (or average of two middle values if n is even)	Skewed distributions, ordinal data
Mode	Most frequently occurring value	Nominal data, multimodal distributions

3.3. Measures of Dispersion (Spread)

Measure	Formula	Interpretation
Range	$(\max(x) - \min(x))$	Total spread; sensitive to outliers
Interquartile Range (IQR)	$(Q_3 - Q_1)$	Spread of the middle 50%; robust to outliers
Variance	$(s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1})$	Average squared deviation from the mean
Standard Deviation (SD)	$(s = \sqrt{s^2})$	Typical distance of observations from the mean
Coefficient of Variation (CV)	$(\frac{s}{\bar{x}} \times 100\%)$	Relative variability; useful for comparing different units

3.4. Shape of Distribution

- **Skewness** – Direction and degree of asymmetry.
 - Positive (right-sk)