

# Descriptive Statistics Cheat Sheet

## 1 Classification & Tabulation of Univariate Data

- **Univariate Data:** Data with only one variable (e.g., height of students).
  - **Classification:** Organizing data into groups/classes.
  - **Types:**
    - **Qualitative** (categorical: color, gender)
    - **Quantitative** (numerical: height, weight)
  - **Frequency Distribution Table:**
    - **Class Interval (CI):** Range of values (e.g., 10-20, 21-30)
    - **Frequency (f):** Number of occurrences in each CI
  - 
  - 
  -
- 

## 2 Graphical Representation

- **Bar Chart:** Used for categorical data.
  - **Histogram:** Represents frequency distribution of continuous data.
  - **Pie Chart:** Shows proportions as slices of a circle.
  - **Frequency Polygon:** Line graph connecting midpoints of class intervals.
  - **Ogive (Cumulative Frequency Curve):**
    - **Less than ogive:** Plots cumulative frequency for “less than” class boundaries.
    - **More than ogive:** Plots cumulative frequency for “more than” class boundaries.
  -
- 

## 3 Frequency Curves

- **Bell-Shaped Curve (Normal Distribution)**
  - **Positively Skewed (Right-Skewed)**
  - **Negatively Skewed (Left-Skewed)**
  - **J-Shaped Curve**
  - **U-Shaped Curve**
- 

## 4 Descriptive Measures

### A. Measures of Central Tendency

- **Mean (  $\bar{x}$  or  $\bar{x}$ ):**
  - Formula:  $\bar{x} = \frac{\sum x_i}{n}$
  - For grouped data:  $\bar{x} = \frac{\sum fx}{\sum f}$
- 
- **Median (M):**
  - Middle value when arranged in order.
  - Formula (for grouped data):
$$M = L + \left( \frac{\frac{n}{2} - CF}{f} \right) \times h$$
    - LLL = Lower boundary of median class
    - nnn = Total frequency
    - CFCFCF = Cumulative frequency before median class
    - fff = Frequency of median class

- $h$  = Class width
- 
- 
- **Mode (Z):**
- Most frequently occurring value.
- Formula (for grouped data):  

$$Z = L + \left( \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$$

$$Z = L + (2f_1 - f_0 - f_2) \times h$$
- $f_1$  = Modal class frequency
- $f_0$  = Previous class frequency
- $f_2$  = Next class frequency
- 
- 

## B. Measures of Dispersion

- **Range** = Maximum - Minimum
- **Variance ( $\sigma^2$  or  $s^2$ ):**
- Population:  $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$
- Sample:  $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
- 
- **Standard Deviation ( $\sigma$  or  $s$ ):**
- $\sigma = \sqrt{\sigma^2}$
- $s = \sqrt{s^2}$
- **Coefficient of Variation (CV):**
- $CV = \frac{\sigma}{\mu} \times 100$

## 5 Bivariate Data

- **Definition:** Data with two variables (e.g., height & weight).
- **Scatter Plot:** Graphical representation of bivariate data.

### Summarization of Bivariate Data

- **Mean of X ( $\bar{x}$ ):**  $\bar{x} = \frac{\sum x_i}{n}$
- **Mean of Y ( $\bar{y}$ ):**  $\bar{y} = \frac{\sum y_i}{n}$

### Marginal & Conditional Frequency Distributions

- **Marginal Frequency Distribution:** Row or column sums in a contingency table.
- **Conditional Frequency Distribution:** Probability of one variable given a specific value of another.

## Key Formulas Summary

Measure	Formula
Mean	$\bar{x} = \frac{\sum x_i}{n}$
Median (Grouped)	$M = L + \left( \frac{\frac{n}{2} - CF}{f} \right) \times h$
Mode (Grouped)	$Z = L + \left( \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$

Measure	Formula
Variance	$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$
CV	$\frac{\sigma}{\mu} \times 100$