Measures of Central Tendency and Dispersion (Ungrouped and Grouped Data)

1 Introduction

In descriptive statistics, we summarize data using:

- **Measures of central tendency** (means): Arithmetic, Geometric, Harmonic.
- **Measures of dispersion** (variability): Range, Variance, Standard deviation, Mean absolute deviation, Quartiles, and Interquartile range.

2 Measures of Central Tendency

2.1 Arithmetic Mean (AM)

A) Ungrouped data

Given observations $x_1, x_2, \ldots, x_n$, the arithmetic mean is:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

B) Grouped data (discrete frequency table)

If values x_i occur with frequencies f_i, then:

$$\bar{x} = \frac{\sum_i f_i x_i}{\sum_i f_i}.$$

C) Grouped data (continuous class intervals)

For classes $[L_i, U_i)$ with frequency f_i, use the class midpoint

$$m_i = \frac{L_i + U_i}{2},$$

then

$$\bar{x} = \frac{\sum_i f_i m_i}{\sum_i f_i}.$$

2.2 Geometric Mean (GM)

A) Ungrouped data

For positive data ($x_i > 0$):

$$GM = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Using logarithms:

$$GM = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln x_i\right).$$

B) Grouped data

For $x_i > 0$ with frequencies f_i:

$$GM = \exp\left(\frac{\sum_i f_i \ln x_i}{\sum_i f_i}\right).$$

2.3 Harmonic Mean (HM)

A) Ungrouped data

For nonzero observations:

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}.$$

B) Grouped data

For values $x_i \neq 0$ with frequencies f_i:

$$HM = \frac{\sum_i f_i}{\sum_i \frac{f_i}{x_i}}.$$

2.4 Relationship between the three means

For positive data:

$$HM \leq GM \leq AM.$$

3 Measures of Dispersion

3.1 Range

Ungrouped data

$$\text{Range} = x_{\max} - x_{\min}.$$

This measure is simple but highly sensitive to outliers.

3.2 Variance and Standard Deviation

A) Ungrouped data (population variance)

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

If μ is unknown and the sample mean $\bar{x}$ is used, a common population-style formula is:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Shortcut (using $\bar{x}$):

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2.$$

Sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

B) Grouped data

Using representative values x_i (or midpoints m_i for classes):

$$\sigma^2 = \frac{\sum_i f_i(x_i - \bar{x})^2}{\sum_i f_i}.$$

Shortcut:

$$\sigma^2 = \frac{\sum_i f_i x_i^2}{\sum_i f_i} - \bar{x}^2.$$

Standard deviation

$$\sigma = \sqrt{\sigma^2} \qquad \text{or} \qquad s = \sqrt{s^2}.$$

3.3 Mean Absolute Deviation (MAD)

Ungrouped data

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|.$$

Grouped data

$$MAD = \frac{\sum_i f_i|x_i - \bar{x}|}{\sum_i f_i}.$$

3.4 Coefficient of Variation (CV)

A relative measure of dispersion:

$$CV = \frac{\sigma}{\bar{x}} \times 100\%.$$

4 Quartiles and Interquartile Range (IQR)

4.1 Quartiles

Quartiles divide ordered data into four equal parts:

- Q_1: first quartile (25% below)
- Q_2: second quartile (median, 50% below)
- Q_3: third quartile (75% below)

4.2 A) Ungrouped data

Step 1: Sort data in ascending order.

Step 2: Compute quartile positions:

$$\text{Position}(Q_1) = \frac{n+1}{4}, \qquad \text{Position}(Q_2) = \frac{n+1}{2}, \qquad \text{Position}(Q_3) = \frac{3(n+1)}{4}.$$

Step 3: If a position is not an integer, interpolate between the surrounding observations.

4.3 B) Grouped data (class intervals)

Let $N = \sum_i f_i$ be the total frequency and CF be cumulative frequency.

Quartile positions

$$Q_1 \text{ is at } \frac{N}{4}, \qquad Q_2 \text{ is at } \frac{N}{2}, \qquad Q_3 \text{ is at } \frac{3N}{4}.$$

Quartile formula (grouped)

If the quartile lies in a class with:

- L = lower class boundary,
- CF_{before} = cumulative frequency before that class,
- f = frequency of the quartile class,
- h = class width,

then:

$$Q_k = L + \left(\frac{\frac{kN}{4} - CF_{\text{before}}}{f} \right) h, \qquad k = 1, 2, 3.$$

4.4 Interquartile Range (IQR)

$$IQR = Q_3 - Q_1.$$

It measures the spread of the middle 50% of the data and is resistant to outliers.

4.5 Semi-Interquartile Range

$$SIQR = \frac{Q_3 - Q_1}{2}.$$

4.6 Outlier detection using IQR

Define fences:

$$\text{Lower fence} = Q_1 - 1.5(IQR), \qquad \text{Upper fence} = Q_3 + 1.5(IQR).$$

Observations outside these fences are often considered **outliers**.

5 Summary Table

Measure	**Formula**
Arithmetic Mean (ungrouped)	$\bar{x} = \frac{1}{n}\sum x_i$
Arithmetic Mean (grouped)	$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$
Geometric Mean	$GM = \left(\prod x_i\right)^{1/n} = \exp\left(\frac{1}{n}\sum \ln x_i\right)$
Geometric Mean (grouped)	$GM = \exp\left(\frac{\sum f_i \ln x_i}{\sum f_i}\right)$
Harmonic Mean	$HM = \frac{n}{\sum(1/x_i)}$
Harmonic Mean (grouped)	$HM = \frac{\sum f_i}{\sum(f_i/x_i)}$
Range	$x_{\text{max}} - x_{\text{min}}$
Variance (grouped)	$\sigma^2 = \frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i}$
Standard deviation	$\sigma = \sqrt{\sigma^2}$
Quartiles (ungrouped positions)	$Q_1 : \frac{n+1}{4},\ Q_2 : \frac{n+1}{2},\ Q_3 : \frac{3(n+1)}{4}$
Quartiles (grouped formula)	$Q_k = L + \left(\frac{\frac{kN}{4} - CF_{\text{before}}}{f}\right) h$
Interquartile Range	$IQR = Q_3 - Q_1$