# Measure of central tendency and its types

Measure of central tendency and its types:

- Geometric and Harmonic mean for ungroup and group data. Its uses and applications

Measure of dispersion and its types

- Range for absolute and relative measure (ungroup and group)

# Geometric Mean

- The geometric mean is useful in finding the average change of percentages, ratios,indexes, or growth rates over time. It has a wide application in business and economics because we are often interested in finding the percentage changes in sales, salaries, or economic figures, such as the Gross Domestic Product, which compound or build on each other.

# Defination

- The geometric mean of a set of n positive numbers is defined as the nth root of the product of n values. The formula for the geometric mean is written:
- For ungroup

$$G.M = \sqrt[n]{x_1.x_2.x_3....x_n}$$

$$G.M = anti\log[\frac{1}{n}\{\sum \log x_i\}]$$

- Further study See book Chapter 3 page 77

- For group

$$G.M = anti\log\left[\frac{1}{\sum f}\left\{\sum f\log x_i\right\}\right]$$

# For Ungroup data

| Percentage rise in population | Population at the end of year $x_i$ | $\log x_i$ |
| --- | --- | --- |
| 15 | 115 | 2.0607 |
| 25 | 125 | 2.0969 |
| 5 | 95 | 1.9777 |
| | | 6.1353 |

$$G.M = \text{Antilog} \frac{\sum_{n}^{n} \log x_i}{n}$$

$$= \text{Antilog} \frac{(6.1353)}{3}$$

$$= \text{Antilog}\ (2.0451)$$

$$= 110.9$$

# For Group Data

| Marks | Mid point ($x_i$) | $f_i$ | $\log x_i$ | $f_i \log x_i$ |
|---|---|---|---|---|
| 4-8 | 6 | 6 | 0.7782 | 4.6692 |
| 8-12 | 10 | 10 | 1.0000 | 10.0000 |
| 12-16 | 14 | 18 | 1.1461 | 20.6298 |
| 16-20 | 15 | 30 | 1.2553 | 37.6590 |
| 20-24 | 22 | 15 | 1.3424 | 20.1360 |
| 24-28 | 26 | 12 | 1.4150 | 16.800 |
| 28-32 | 30 | 10 | 1.4771 | 14.7710 |
| 32-36 | 34 | 6 | 1.5315 | 9.1890 |
| 36-40 | 38 | 2 | 1.5798 | 3.1596 |
| Total | | N = 109 | | 137.1936 |

$$\text{G.M.} = \text{Antilog}\left[\frac{\sum_{i=1}^{n} f_i \log x_i}{N}\right]$$

$$= \text{Antilog}\left[\frac{137.1936}{109}\right] = \text{Antilog}\,[1.2587]$$

$$\text{G.M.} = 18.14$$

# Harmonic Mean

$H.M = \frac{1}{x}$

- Harmonic mean is the reciprocal of arithmetic mean and reciprocal of its values.
- Formula
- for ungroup $$H.M = \frac{n}{\sum \frac{1}{x}}$$
- for group $$H.M = \frac{\sum f}{\sum (\frac{f}{x})}$$

# Example for Ungroup data

| Truck Number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Km. driven | 40 | 50 | 60 | 75 |

| $x$ | $1/x$ |
|---|---|
| 40 | 0.02500 |
| 50 | 0.02000 |
| 60 | 0.01677 |
| 75 | 0.01333 |
| | 0.07500 |

$$H.M = \frac{N}{\sum 1/x}$$

$$H.M = \frac{4}{0.07500}$$

$$H.M = 53.33\ Km.$$

# Example for Group data

| Class Interval | 11 -- 15 | 16 -- 20 | 21 -- 25 | 26 -- 30 | 31 -- 35 |
|---|---|---|---|---|---|
| Frequencies | 15 | 20 | 60 | 150 | 15 |

| *Class Interval* | $x$ | $f$ | $f/x$ |
|---|---|---|---|
| 11 – 15 | 13 | 15 | 1.153846154 |
| 16 – 20 | 18 | 20 | 1.111111111 |
| 21 – 25 | 23 | 60 | 2.608695652 |
| 26 – 30 | 28 | 150 | 5.357142857 |
| 31 – 35 | 33 | 15 | 0.454545455 |
| | | 260 | 10.68534 |

$x = Midpoint$

$$x = \frac{L.C.L + U.C.L}{2}$$

$$e.g\ x = \frac{11 + 15}{2} = 13$$

$$H.M = \frac{N}{\sum f/x}$$

$$H.M = \frac{260}{10.68534} = 24.3$$

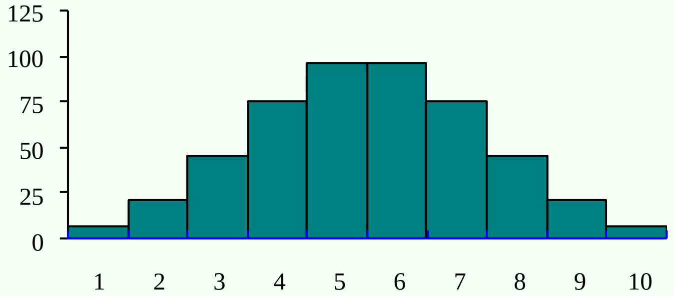
# Measures of Dispersion

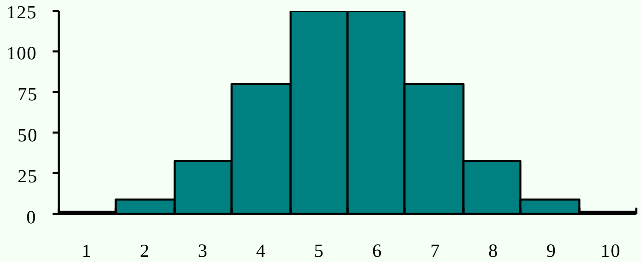
# Definition

- *Measures of dispersion* are descriptive statistics that describe how similar a set of scores are to each other
  - The more similar the scores are to each other, the lower the measure of dispersion will be
  - The less similar the scores are to each other, the higher the measure of dispersion will be
  - In general, the more spread out a distribution is, the larger the measure of dispersion will be

# Measures of Dispersion

- Which of the distributions of scores has the larger dispersion?
- The upper distribution has more dispersion because the scores are more spread out
  - That is, they are less similar to each other



# Types of dispersion

- Absloute Measure of dispersion
- Relative measure of dispersion

# Absloute Measure of dispersion

- Range
- The semi-interquartile range (SIR) or Quartile Deviation(Q.D)
- Mean Deviation (M.D)
- Standard deviation

# Relative measure of dispersion

- Co-efficient of Range
- Co-efficient of Quartile Deviation(Q.D)
- Co-efficient of Mean Deviation (M.D)
- Co-efficient of Standard deviation
- Variance

# The Range

- The *range* is defined as the difference between the largest score in the set of data and the smallest score in the set of data, $X_L - X_S$
- What is the range of the following data:
  4 8 1 6 6 2 9 3 6 9
- The largest score ($X_L$) is 9; the smallest score ($X_S$) is 1; the range is $X_L - X_S = 9 - 1 = 8$

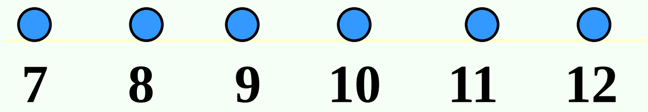
# The Range

- **Measure of Variation**
- **Difference Between Largest & Smallest Observations:**

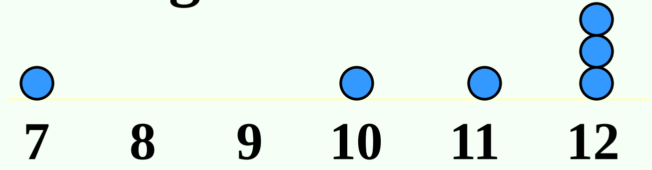
$$\text{Range} = x_{Largest} - x_{Smallest}$$

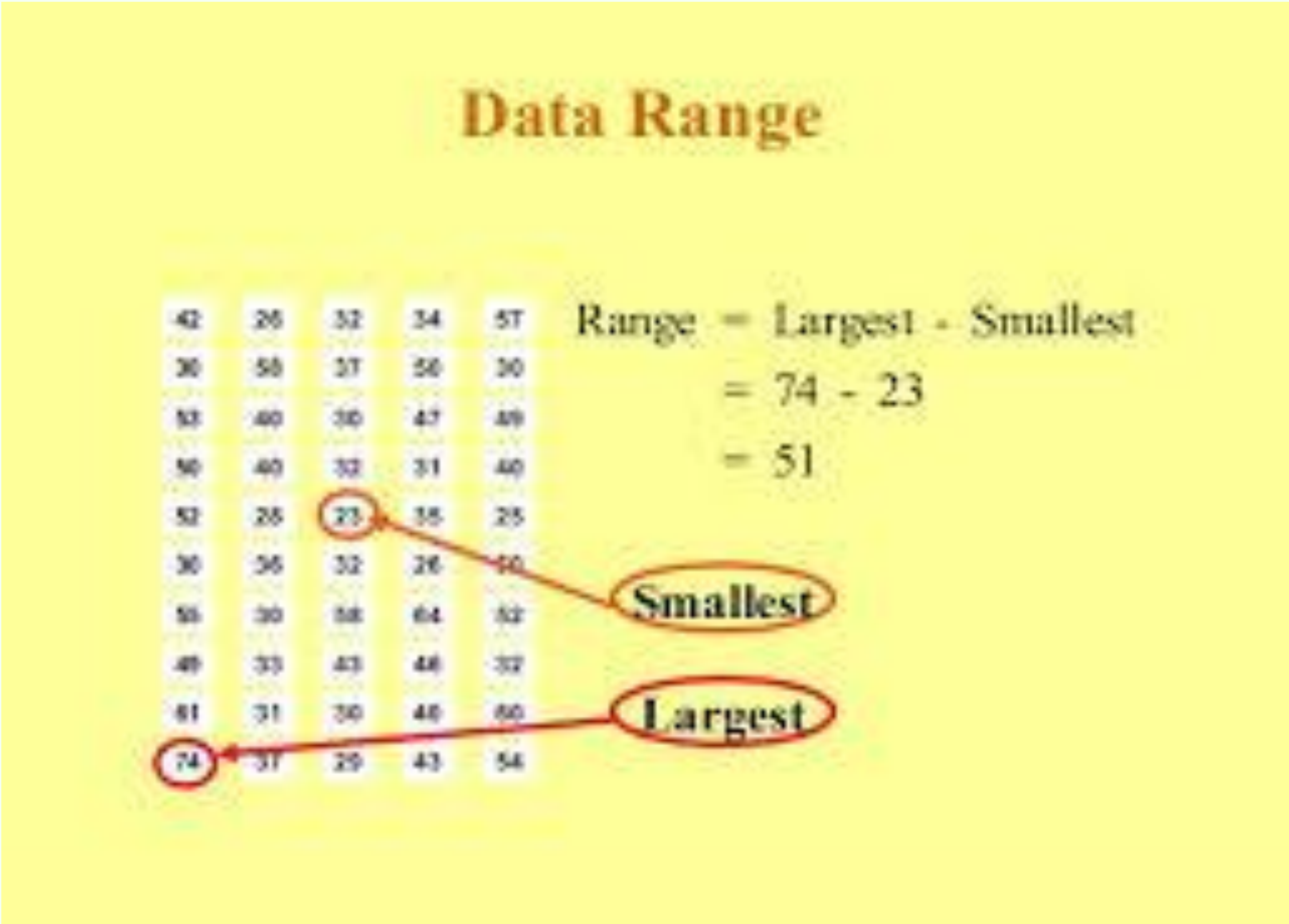
- **Ignores How Data Are Distributed:**

**Range = 12 - 7 = 5**




**Range = 12 - 7 = 5**




Data Range
Range = Largest - Smallest
= 74 - 23
= 51
Smallest
Largest

# Estimate Range from Grouped Data

| Employee $ contributed to profit sharing plan | | | Frequency f |
|---|---|---|---|
| $50 | Up To | $55 | 4 |
| 55 | Up To | $60 | 7 |
| 60 | Up To | 65 | 9 |
| 65 | Up To | 70 | 22 |
| 70 | Up To | 75 | 40 |
| 75 | Up To | 80 | 24 |
| 80 | Up To | 85 | 15 |
| 85 | Up To | 90 | 9 |
| | Total | | 130 |

| Estimated Range | = | Upper Limit of the Highest Class | - | Lower Limit of the Lowest Class | |
|---|---|---|---|---|---|
| Estimated Range | = | $90 | - | $50 | = $40 |

65

# When To Use the Range

- The range is used when
  - you have ordinal data or
  - you are presenting your results to people with little or no knowledge of statistics
- The range is rarely used in scientific work as it is fairly insensitive
  - It depends on only two scores in the set of data, $X_L$ and $X_S$
  - Two very different sets of data can have the same range:
    1 1 1 1 9 vs 1 3 5 7 9

# The Semi-Interquartile Range

- The *semi-interquartile range* (or *SIR*) is defined as the difference of the first and third quartiles divided by two
  - The first quartile is the $25^{th}$ percentile
  - The third quartile is the $75^{th}$ percentile
- $SIR = (Q_3 - Q_1) / 2$

# SIR Example

- What is the SIR for the data to the right?
- 25 % of the scores are below 5
  - 5 is the first quartile
- 25 % of the scores are above 25
  - 25 is the third quartile
- SIR = $(Q_3 - Q_1) / 2 = (25 - 5) / 2 = 10$

| |
|---|
| 2 |
| 4 |
| ← 5 = $25^{th}$ %tile |
| 6 |
| 8 |
| 10 |
| 12 |
| 14 |
| 20 |
| ← 25 = $75^{th}$ %tile |
| 30 |
| 60 |

# When To Use the SIR

- The SIR is often used with skewed data as it is insensitive to the extreme scores