CHAPTER 3

NUMERICAL DESCRIPTIVE MEASURES

Measures of Central Tendency

OBJECTIVE

Summarize data using measures of central tendency, such as mean, median, and mode.

- Population parameters/parameter
  A numerical measure calculated for a population data set.

  Example : μ, σ

- Sample statistic/ Statistic
  A summary measure calculated for a sample data

  Example : $\bar{x}$, s

MEAN

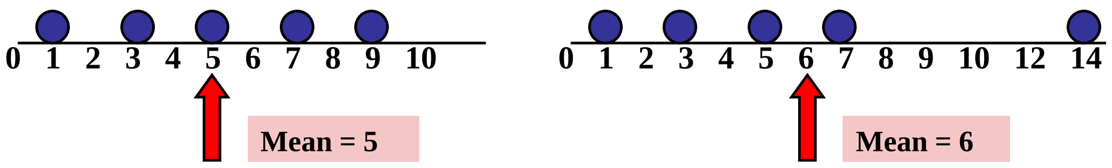
Use the mean to describe the middle of a set of data that *does not* have an outlier.

Advantages:

- Most popular measure in fields such as business, engineering and computer science.
- It is unique - there is only one answer.
- Useful when comparing sets of data.

Disadvantages:

- Affected by extreme values (outliers)



MEAN (Ungrouped Data)

The **arithmetic mean** (or simply the **mean**) of a list of numbers is the sum of all the members of the list divided by the number of items in the list (for ungrouped data)

$$Mean = \frac{Sum of\ all\ values}{Number of\ values}$$

The mean is the most commonly-used type of **average** and is often referred to simply as *the* average.

Calculating Mean

Mean for Ungrouped Data

$$\text{Mean for population data, } \mu = \frac{\sum X}{N}$$

$$\text{Mean for sample data, } \quad \bar{x} = \frac{\sum x}{n}$$

Mean for Grouped Data

$$\text{Mean for population data, } \mu / \bar{x} = \frac{\sum fX_m}{\sum f}$$

where Σx is the sum of all values, N is the population size, n is the sample size, μ is the population mean and is the sample mean.

MEDIAN (Ungrouped Data)

- The number separating the higher half of a sample or a population from the lower.
- The value of the *middle term* in a data set that has been ranked in increasing order.
- There is a unique median for each data set.
- It is not influenced by outliers and is therefore a valuable measure of central tendency when such values occur.

MEDIAN

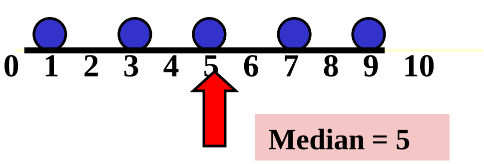
Use the median to describe the middle of a set of data that *does* have an outlier.

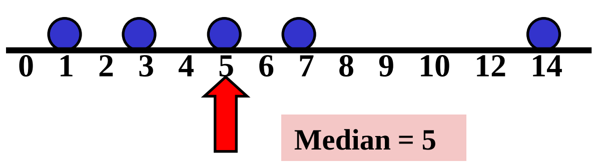
Advantages:

- Extreme values (outliers) do not affect the median as strongly as they do the mean.
- Useful when comparing sets of data.
- It is unique - there is only one answer.

Disadvantages:

- Not as popular as mean.





7

Calculating Median

The median for ungrouped data:

Steps : i - Rank the data set in increasing order

ii - Find the middle term.

If $\text{n} = \text{odd} \rightarrow \text{Median} = \text{middle term}$

$\text{n} = \text{even} \rightarrow \text{Median} = \text{average of 2 middle term}$

$$\text{Median} = \text{Value of the} \left(\frac{n+1}{2}\right)^{\text{th}} \text{term in a ranked data set}$$

The median for grouped data:

$$MD = L + \left(\frac{\frac{n}{2} - C_f}{f}\right)C$$

where L = Lower limit of MD class

C_f = Cum frequency before the MD class

f = Class frequency of the median class

C = Class size

8

MODE

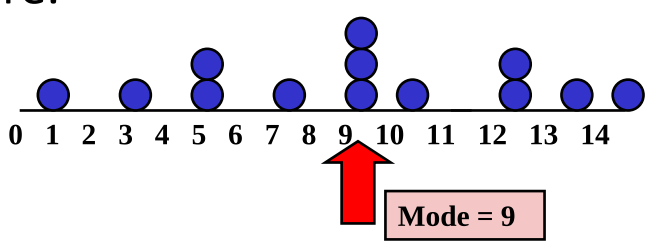
Use the mode when the data is non-numeric or when asked to choose the most popular item.

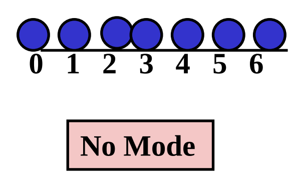
Advantages:

- Extreme values (outliers) do not affect the mode.

Disadvantages:

- Not as popular as mean and median.
- Not necessarily unique - may be more than one answer
- When there is more than one mode, it is difficult to interpret and/or compare.





9

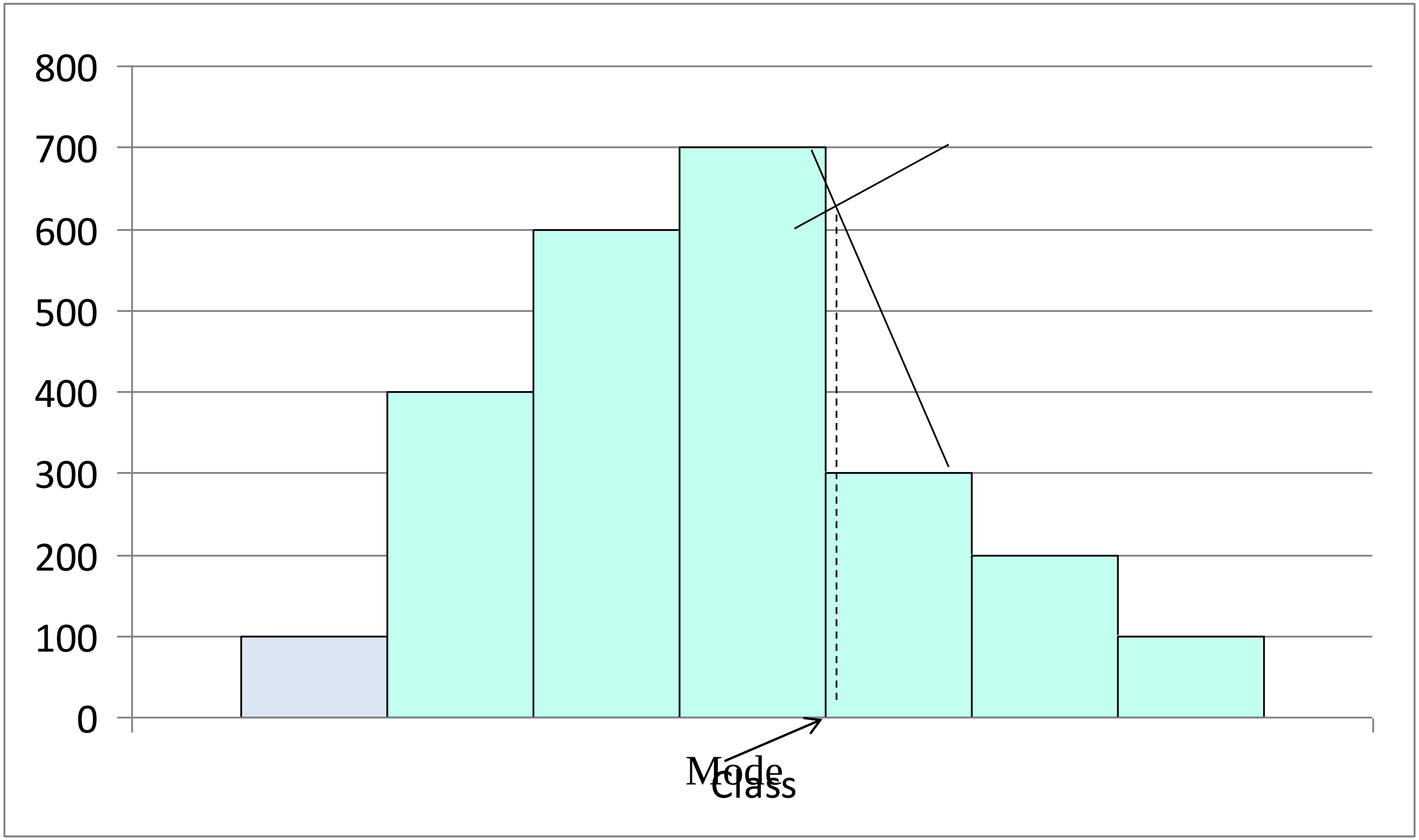
MODE (Ungrouped Data)

The value that occurs with the highest frequency in a data set.

A data set may have none or may have more than one mode, whereas it will have only one mean and only one median.

Unimodal	Only one mode
Bimodal	Two modes
Multimodal	More than two modes

Mode (Grouped Data)



MEASURES OF VARIATION (DISPERSION)

Objectives:

Describe data using measures of variation, such as the;

i. Range
ii. Variance
iii. Standard Deviation

12

Measures Of Variation For Ungrouped Data

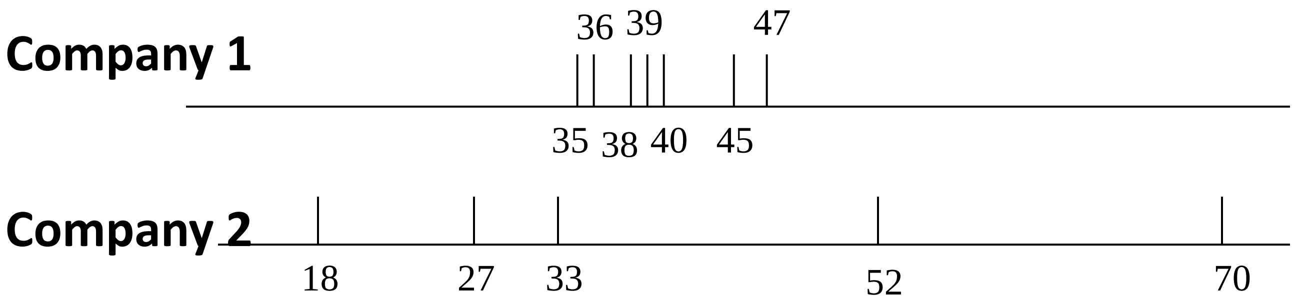
- The measures of central tendency do not reveal the whole picture of the distribution of a data set.
- Two data sets with the same mean may have completely different spreads.
- Consider the following two data sets on the ages of all workers in each of two small companies.

Company 1	47 38 35 40 36 45 39
Company 2	70 33 18 52 27

- Both Mean = 40, **BUT** the variation very different.

Measures of Variation For Ungrouped Data

- The ages of the workers in the second company have a much larger variation than the ages of the workers in the first company.



- The measures of dispersion gives the spread of a data set.

14

RANGE

- Range for Ungrouped Data :

$$\text{Range} = \text{Largest value} - \text{smallest value}$$

Example :

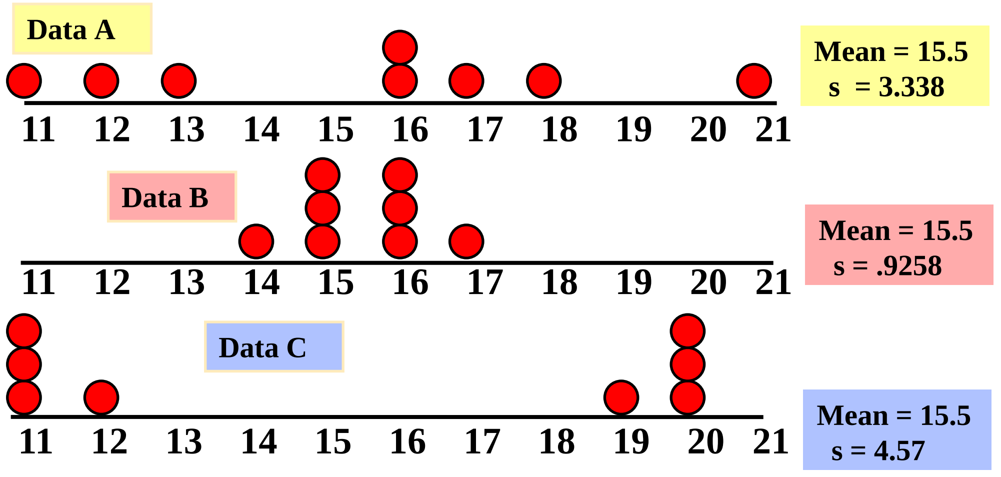
Find the range for the below data set:

53,182   49,651   69,903   267,277

- Disadvantage : - Influenced by outliers
  - Based on 2 values only

Variance and Standard Deviation

- The value of the standard deviation tells how closely the values of a data set are clustered around the mean.
- A lower value of the s.d. indicates that the values of that data set are spread over a relatively smaller range around the mean.
- A larger value of the s.d. indicates that the values of that data set are spread over a relatively larger range around the mean.



16

Variance and Standard Deviation
(Ungrouped Data)

	Population	Sample
Basic Formulas (Variance)	$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$	$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$
Short-Cut Formulas (Variance)	$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$	$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

Variance and Standard Deviation (Grouped Data)

	Population	Sample
Basic Formulas (Variance)	$\sigma^2 = \dfrac{\sum f(X_m - \mu)^2}{N}$	$s^2 = \dfrac{\sum f(X_m - \bar{x})^2}{n-1}$
Short-Cut Formulas (Variance)	$\sigma^2 = \dfrac{\sum X_m^2 f - \frac{(\sum X_m f)^2}{N}}{N}$	$s^2 = \dfrac{\sum X_m^2 f - \frac{(\sum X_m f)^2}{n}}{n-1}$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

Use of Standard Deviation

By using the mean and standard deviation, we can find the proportion or percentage of the total observations that fall within a given interval about the mean.

Example:

The heights of 30 bean plant are as follows:

Height (cm)	Frequency
3-5	1
6-8	2
9-11	11
12-14	10
15-17	5
18-20	1

Find the Variance and Standard Deviation.

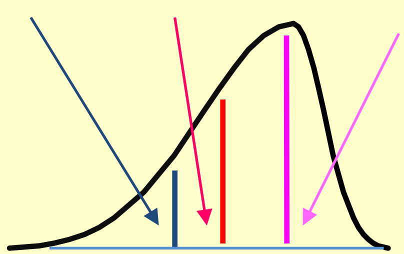
Height		Frequency	Xm	fXm	Xm-x'	(Xm-x')^2	f(Xm-x')^2
3	5	1	4	4	-7.9	62.41	62.41
6	8	2	7	14	-4.9	24.01	48.02
9	11	11	10	110	-1.9	3.61	39.71
12	14	10	13	130	1.1	1.21	12.1
15	17	5	16	80	4.1	16.81	84.05
18	20	1	19	19	7.1	50.41	50.41
Sum		30		357			296.7
Mean	**11.90**						
Variance	**10.23**						
Standard Deviation	**3.20**						

Shape of a Distribution

- Describes how data is distributed
- Measures of shape
  - Symmetric or skewed

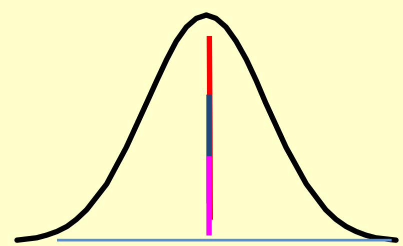
Left-Skewed

Mean < Median < Mode



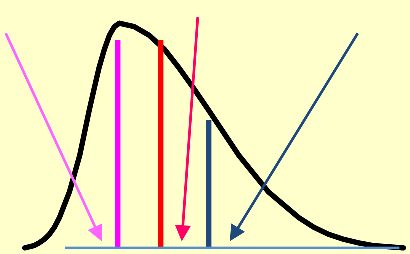
Symmetric

Mean = Median =Mode



Right-Skewed

Mode < Median < Mean



Skewness Measure

Measures of skewness indicate shape of distribution.

$$Measure\ of\ skewness = \frac{3\left(\bar{x} - Median\right)}{s}$$

A positive values indicates positive skewness and a negative value denotes negative skewness. A computed value greater than 1 (less than -1) denotes strong positive (negative) skewness.

Example:

The following data consisting of the number of aggressive acts per hour for 33 preschool children

1	**2**	**2**	**2**	**2**	**3**	**4**	**4**	**5**	**5**	**6**
7	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**10**	**10**	**10**
11	**11**	**12**	**12**	**12**	**12**	**12**	**13**	**13**	**14**	**15**

a) Compute Mean
b) Find the Median
c) Find the Mode
d) Given the Mean, Median and Mode, what is the skew of this distribution?