Probability & Statistical Modelling

AQ077-3-2-PSMOD and Version VD1

A · P · U

ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

Summary Measures of Statistics

Topic & Structure of The Lesson

- Introduction
- Measures of Location
- Measures of Position
- Measures of Variation
- Measures of Skewness

Learning Outcomes

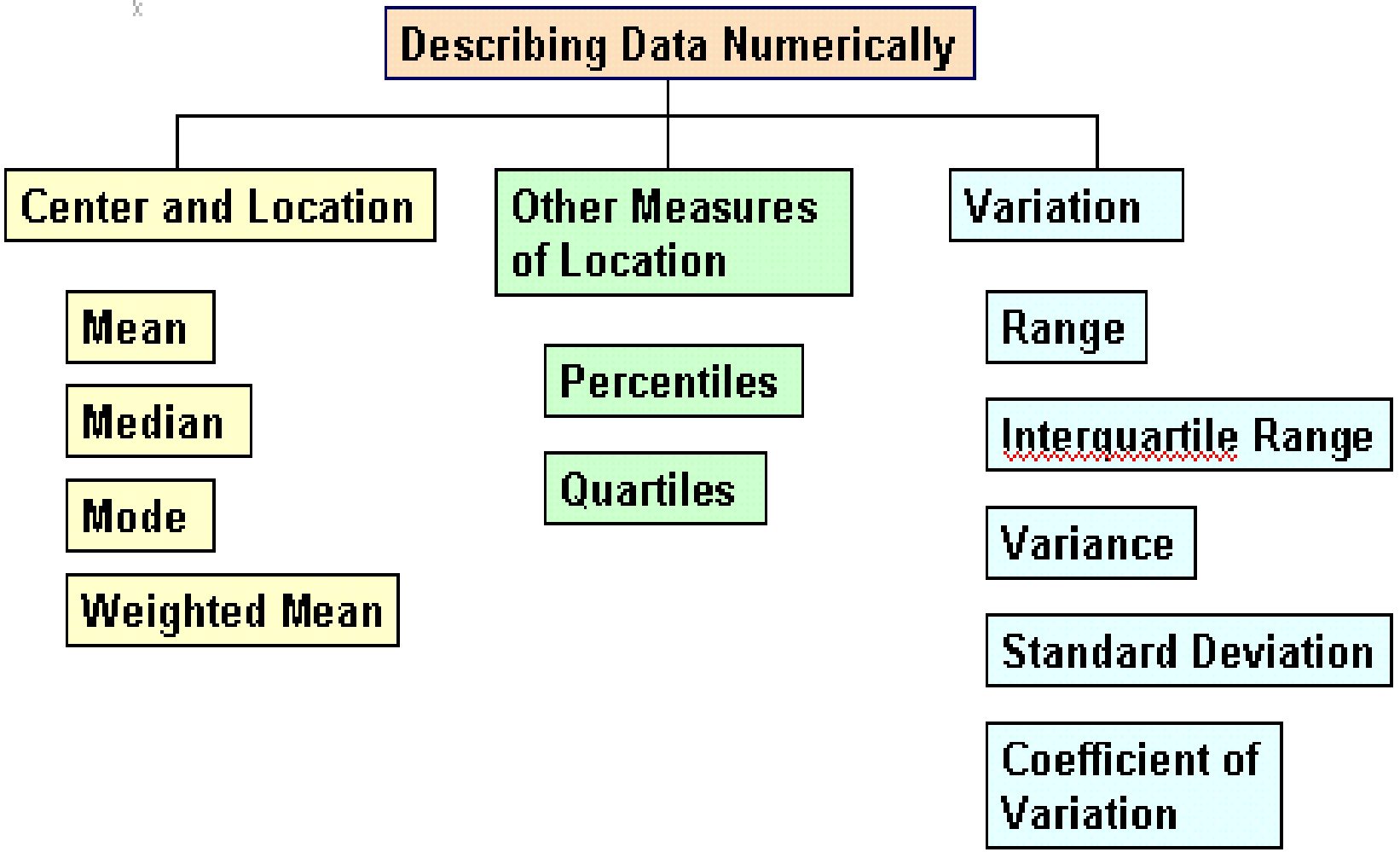
- **A the end of this topic, You should be able to:**
  - **Have refreshed your knowledge of statistics on**
    - **Mean**
    - **Standard Deviation and the variance**
  - **Interpret the output from a statistical package**

Key Terms You Must Be Able To Use

- **If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:**

  - **Measures of central tendency**
  - **Measures of location**
  - **Measures of variation**
  - **Measures of skewness**

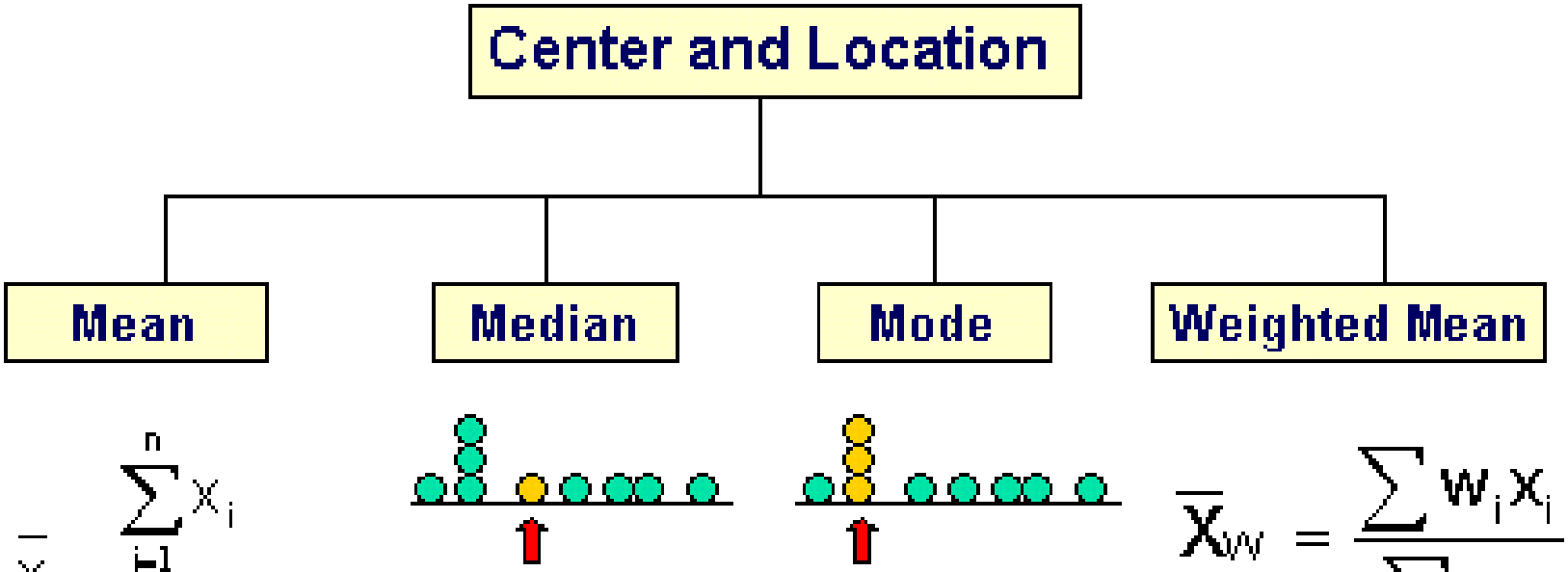
Summary Measures



Characteristics of Distribution

- Measures of central tendency – average (the typical size)
  - ✓ Mean
  - Median
  - Mode

Measures of Central Tendency



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

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

- **Mean**
  - It is the arithmetic average of data values and usually denoted by $\bar{x}$
  - For a set of values mean, $\bar{x} = \frac{\sum x}{n}$
  - For a simple frequency distribution, mean , $\bar{x} = \frac{\sum fx}{\sum f}$

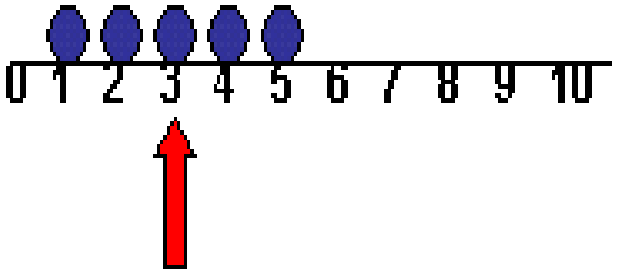
- For grouped frequency, it is impossible to find the total values of the items, which means, in effect that, it is impossible to calculate the mean exactly. However, it is possible to estimate it.

mean, $\bar{x} = \frac{\sum fx}{\sum f}$

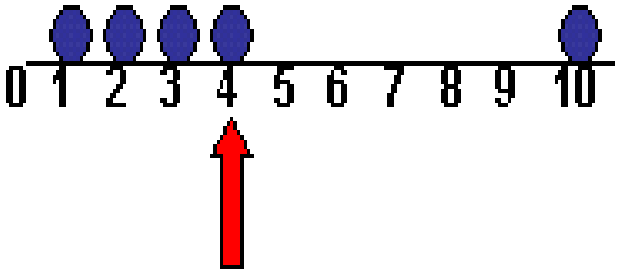
where x = class mid-point

- **Advantages**
  - it is widely understood
  - the value of every item is included in the computation of the mean.
  - it is well suited to further statistical analysis.
- **Disadvantages**
  - its value may not correspond to any actual value.

- **it might be distorted by extremely high or low values.**



$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$



$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

Quick Review Question

A · P · U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

Example:

- **Consider the following set of sample data and compute the mean.**

(a) 17, 15, 18, 20, 15, 15, 11, 15

(b)

x	1	2	3	4	5	6	7
f	3	4	7	12	12	7	5

(c)

Range	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
f	2	12	15	18	13	10

Consider the following data:

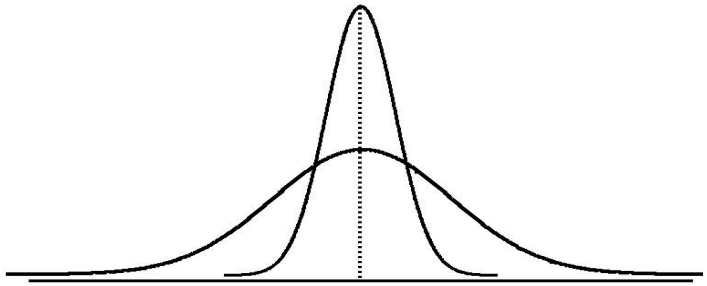
(a) 7, 7, 7, 7, 7

(b) 4, 6, 6.5, 7.2, 11.3

(c) -193, -46, 28, 69, 177

Find mean for each set of data above.

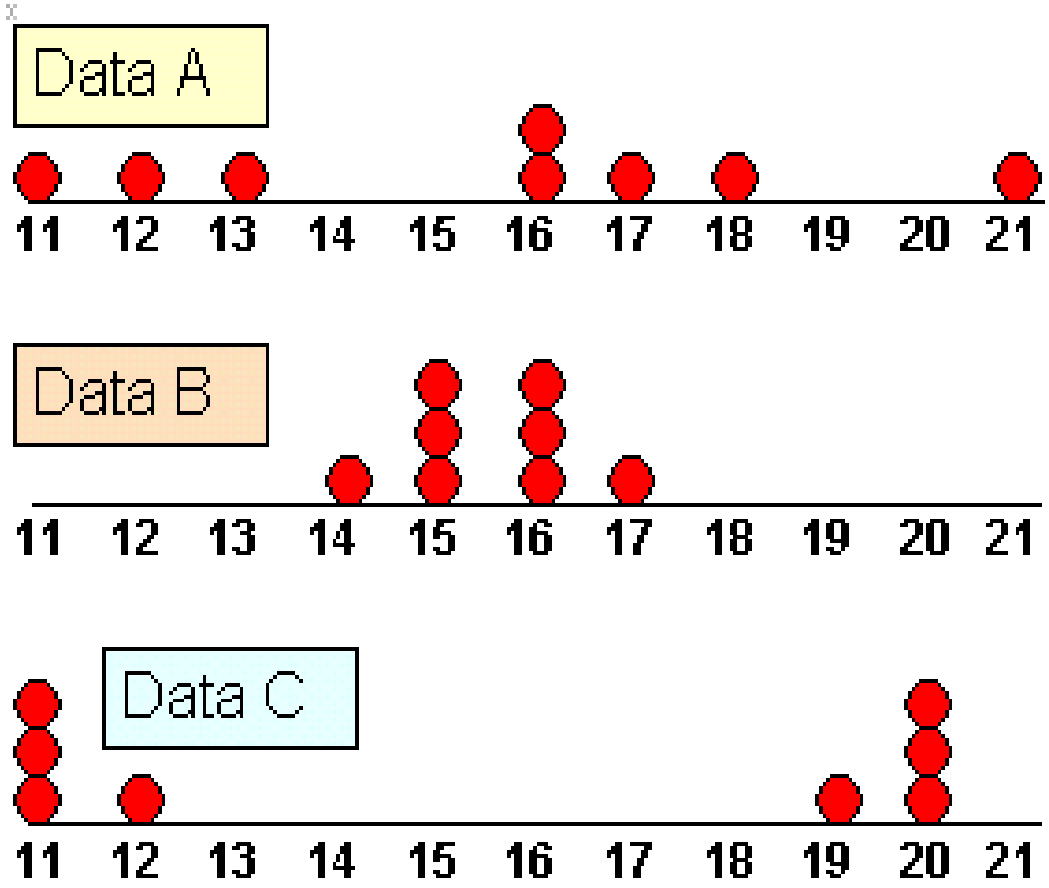
- **Measures of Dispersion (Variation)**
  - **A measure of the degree or dispersion of data**
    - **Range**
    - ✓ **Standard deviation**
    - ✓ **Variance**
    - **Quartile deviation**
    - **Mean deviation**



  - **Needed for two basic purposes.**
    - **To asses the reliability of the average of the data**
    - **To serve as a basis for control of the variability**

- **Standard deviation**
  - it is a measure of the extent to which data for a particular random variable (x) is spread about the mean.
  - The higher the standard deviation the greater the amount of scatter.

Comparing Standard Deviation



Mean = 15.5
s = 3.338

Mean = 15.5
s = .9258

Mean = 15.5
s = 4.57

- standard deviation for a set of values is calculated as follows:
  - For population

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

- **For sample**

$$s = \sqrt{\frac{\sum \left(x - \bar{x}\right)^2}{n - 1}}$$

- **Standard deviation for simple frequency distribution**
  - **For sample:** $$s = \sqrt{\frac{\sum f\left(x - \bar{x}\right)^2}{\sum f - 1}}$$
  - **For Population:** $$\sigma = \sqrt{\frac{\sum f\left(x - \mu\right)^2}{\sum f}}$$

where x = class mid-point

- **Standard deviation for grouped frequency distribution**
  - **For sample:**

$$s = \sqrt{\frac{\sum f\left(x - \bar{x}\right)^2}{\sum f - 1}}$$

  - **For Population:**

$$\sigma = \sqrt{\frac{\sum f\left(x - \mu\right)^2}{\sum f}}$$

where x = class mid-point

- **Another formulation which is convenient when a calculator is being used is as follows:**

$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

- **Advantages**
  - it takes all values into account, therefore it can be regarded as truly representative of the data.
  - it is suitable for further statistical analysis.
- **Disadvantages**
  - it is more difficult to understand than some other measures of dispersion.

Quick Review Question

A · P · U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

➢ **Example:**

(a) **Find the standard deviation for the following data:**
196, 198, 198, 199, 200, 200, 201, 201, 202, 205

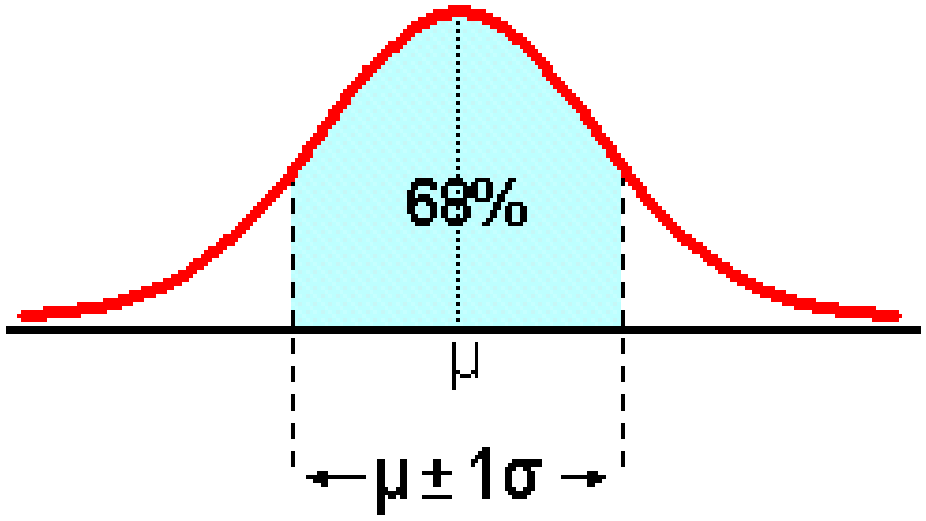
(b) **The time taken for 115 candidates to complete an intelligence test was recorded and summarise in the table below.**

Time (minutes)	0 - 1	1 - 2	2 - 3	3 - 5	5 - 10
Frequency	10	15	25	40	25

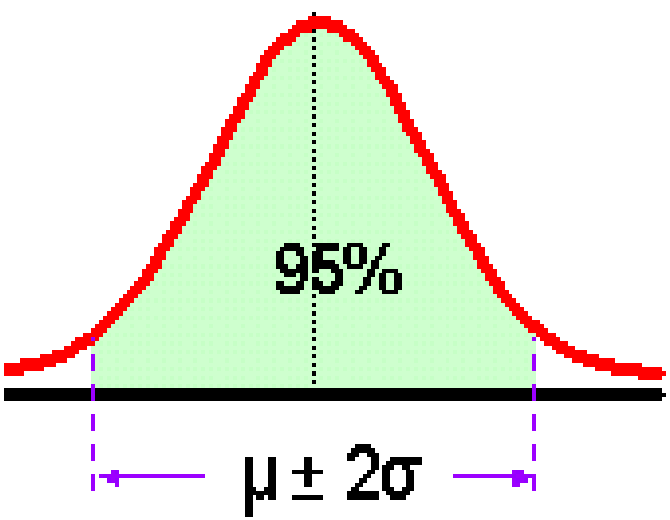
Calculate estimate of the mean and standard deviation of the time taken to complete the test.

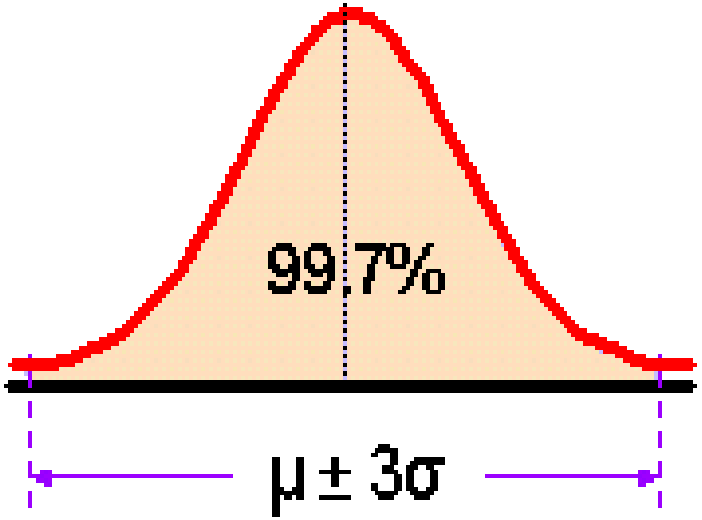
The Empirical Rule

- If the data distribution is bell-shaped, then the interval:
  - $\mu \pm 1\sigma$ contains about 68% of the values in the population or the sample



- $\mu \pm 2\sigma$ contains about 95% of the values in the population or the sample
- $\mu \pm 3\sigma$ contains about 99.7% of the values in the population or the sample





- **Variance**
  - it is a measure of the extent to which values in a data set vary around the mean.
  - if most values cluster closely around the mean, the variance will be a small figure.
  - if, however, most values are widely dispersed around the mean, the variance will be a large figure.
  - variance = $(\text{standard deviation})^2$

Proportion

- A proportion is a name we give to a statement that two ratios are equal. It can be written in two ways:
  - two equal fractions,

$$\frac{a}{b} = \frac{c}{d}$$

or,

- using a colon, **a:b = c:d**

EXCEL

=AVERAGE(number1, number2,...)

=MEDIAN(number1, number2, ...)

=MODE(number1, number2,...)

=QUARTILE(range, 1)

=QUARTILE(range, 3)

=STDEV(number1, number2,...)

Summary of Main Teaching Points

- **Measures of statistics:**
  - **Location**
  - **Position**
  - **Variation**
  - **Skewness**

Question and Answer Session

A · P · U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

Q & A

What we will cover next

- **Correlation & Regression Analysis**