LECTURE 1 OF 5

TOPIC	:	**6.0 DATA DESCRIPTION**
SUBTOPIC	:	**6.1 Introduction to Data**
LEARNING OUTCOMES	:	At the end of this lesson, students should be able to:

a) identify the discrete and continuous data
b) identify ungrouped and grouped data
c) construct and interpret stem-and leaf diagram

CONTENT

SET INDUCTION

These are Mathematics marks for 20 students who are taking mid-semester test

18 , 23, 24, 46, 34, 48, 56, 63, 23, 43, 65, 78, 84, 95, 98, 67, 73, 68, 58, 71

How can we interpret these marks?

INTRODUCTION

Statistics is a science that deals with collecting, organizing, summarizing, presenting and analyzing data in order to obtain useful information for decision making.

Statistics has a wide range of uses in the field of science, business, industry, economy, medicine, education, agriculture and so on.

For example:

a) In the field of science, statistical techniques are used to analyze data that is created from the experiment.

b) In the area of business, marketing surveys are carried out to determine the compatibility of the product with the economics and social demand.

c) In the formation of the national policy, data from the census is used in economic and social planning.

d) In the field of education, statistical techniques are used to analyze the progress of students in an examination.

Definitions

- **Population** is the collection of all elements whose characteristics are being studied.

- **Parameter** is a summary measure of a population (such as population means μ, variances σ^2, etc.

- **Sample** is a set of **measurements** that constitute part or all of a population, that is, a sample is a subset of a population.

- V**ariable** is any measured characteristic or attribute that differs for different subjects. For example, if the weight of 30 subjects were measured, then weight would be a variable.

  Quantitative variables are measured on an ordinal, interval, or ratio scale.

  Qualitative variables are measured on a nominal scale.

Example 1

Categorise the following information into qualitative or quantitative data.

a) Height of boys
b) Type of footwear
c) Age of lecturers
d) Colour of cars

Solution

a)
b)
c)
d)

Example 2

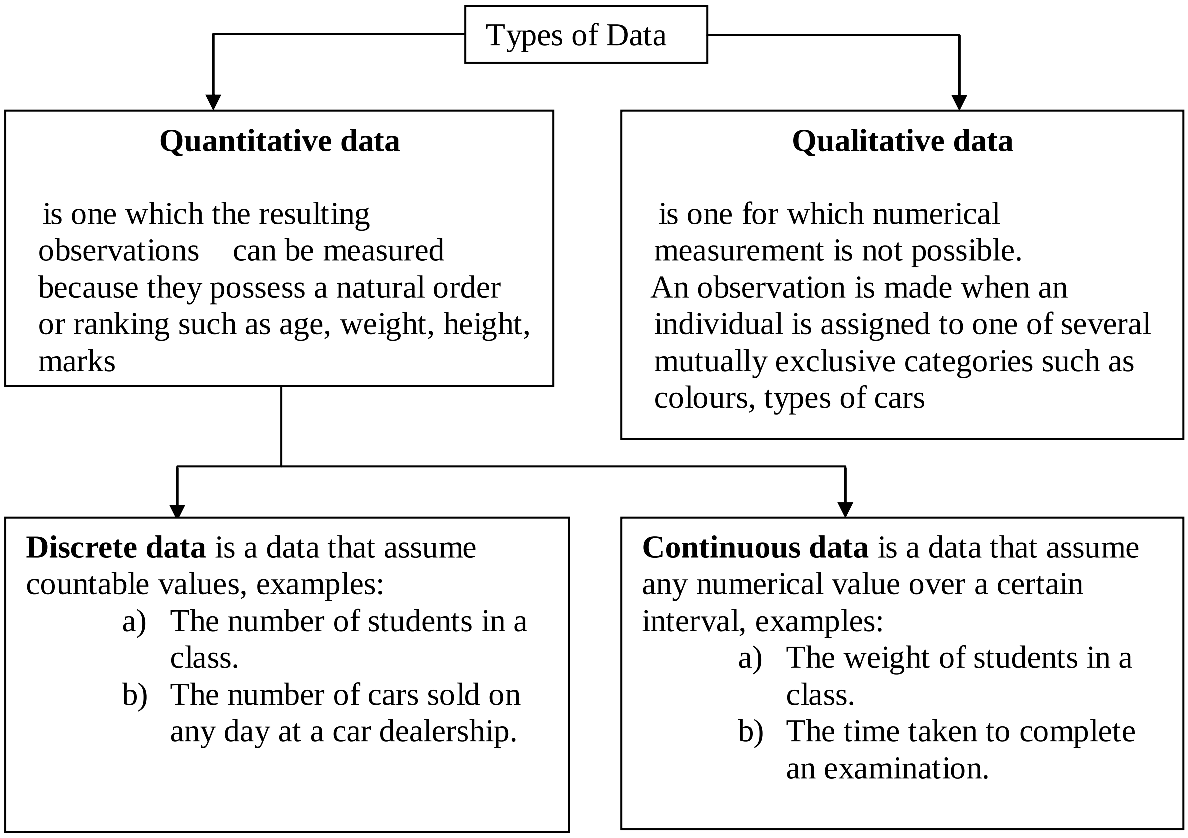
A survey was carried out on 20 boys and girls in a college to find out whether they like the subject Statistics or not. State

a) the population
b) the sample
c) the variable
d) the type of variable, qualitative or quantitative.

Solution

a) The population :
b) The sample :
c) The variable :
d)

- **Data** – collection of observations



Types of Data

Quantitative data

is one which the resulting observations can be measured because they possess a natural order or ranking such as age, weight, height, marks

Qualitative data

is one for which numerical measurement is not possible. An observation is made when an individual is assigned to one of several mutually exclusive categories such as colours, types of cars

Discrete data is a data that assume countable values, examples:

a) The number of students in a class.
b) The number of cars sold on any day at a car dealership.

Continuous data is a data that assume any numerical value over a certain interval, examples:

a) The weight of students in a class.
b) The time taken to complete an examination.

Example 3

Based on the following statements, determine either the data is discrete data or continuous data.

a) The time taken to travel from Ipoh to Kuala Lumpur
b) The number of pens sold by a stationary shop
c) The diameter of ten spheres
d) The number of customers in a cinema in one day
e) The weight of new born babies in a hospital

Solution

a) b) c)

d) e)

Raw data can be represented in **Ungrouped Data** and **Grouped Data**.

a) **Grouped** data is grouped in interval, are categorized into mutually exclusive intervals, can be presented in frequency distribution table, histogram, polygon, ogive.

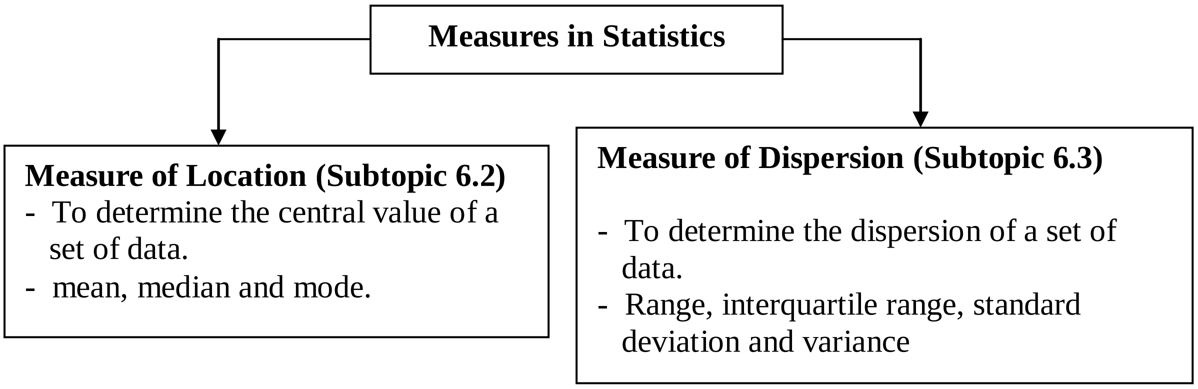
For example,

Height (cm)	150-155	155-160	160-165	165-170
Frequency	2	8	6	5

b) **Ungrouped** Data are listed as a sequence or in the form of a frequency table but without the use of intervals.

For example,

Number of children	0	1	2	3	4
Number of families	4	6	7	2	1



Measures in Statistics

Measure of Location (Subtopic 6.2)
- To determine the central value of a set of data.
- mean, median and mode.

Measure of Dispersion (Subtopic 6.3)

- To determine the dispersion of a set of data.
- Range, interquartile range, standard deviation and variance

Stem-and- leaf diagrams.

A stem and leaf diagram is used to present and display frequency of a group of data without losing information on individual observation. Each value in a stem-and-leaf is divided into two portions.

To construct a stem-and-leaf:

- Each score will be spilt into two parts.
- The first part (first digit) is called **stem**.
- The second part (second digit) is called **leaf**.
- Make sure the data is in ascending order.
- Advantage – the information on individual observations will not be lost.

Example 4

Construct a stem-and-leaf diagram for the data below:

12, 13, 21, 27, 33, 34, 35, 37, 40, 40, 41

Solution

The "stem" is the left-hand column which contains the tens digits. The "leaves" are the lists in the right-hand column, showing all the ones digits for each of the tens, twenties, thirties, and forties.

Stem	Leaf

Example 5

The heights of 15 Form 5 students correct to the nearest cm are given below.

172	**182**	**177**	**174**	**166**	**158**	**170**	**178**
163	**161**	**191**	**167**	**171**	**201**	**166**	

Construct a stem-and-leaf diagram to represent the data.

Solution

Rearrange the data in ascending order :

For this data, the first two digits are used as the stem, for example 15, 16, 17, 18, 19, 20. The last digit is the leaf.

Stem	Leaf

LECTURE 2 OF 5

TOPIC : **6.0 DATA DESCRIPTIVE**

SUBTOPIC : **6.2 Measures of Location**

LEARNING OUTCOMES : At the end of lesson students should be able to:

a) find and interpret the mean, mode, median and quartiles for ungrouped data.

b) construct and interpret box-and-whisker plots for ungrouped data.

CONTENT

The role of measures of location is to determine the central value of a set of data i.e mean, median and mode.

Mean, Median and Mode of Ungrouped Data

Mean of a set data $x_1, x_2, x_3, \ldots x_n$ is written as $\bar{x}$ and defined as

$$\bar{x} = \frac{\text{sum of all data}}{\text{number of data}}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$= \frac{\sum x}{n}$$

To key-in the data in calculator fx570ms:

$\boxed{\text{MODE}} \Rightarrow \boxed{\text{SD}}$

Key in data : x_1 $\boxed{\text{M+}}$ x_2 $\boxed{\text{M+}}$...$\boxed{\text{M+}}$$x_n$

$\Rightarrow \boxed{\text{AC}}$

$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\Rightarrow \sum x, \ \sum x^2$

$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\Rightarrow \bar{x}$, SD, Var

To key-in the data in calculator fx570es:

$\boxed{\text{MODE}} \Rightarrow \boxed{\text{STAT}} \Rightarrow \boxed{\text{1-VAR}}$

Key in data : $x_1$$\boxed{=}$ $x_2$$\boxed{=}$...$\boxed{=}$ x_n

$\Rightarrow$ AC

$\boxed{\text{SHIFT}} \Rightarrow \boxed{1} \Rightarrow \boxed{3} \Rightarrow \sum x, \ \sum x^2$

$\boxed{\text{SHIFT}} \Rightarrow \boxed{1} \Rightarrow \boxed{4} \Rightarrow \bar{x}$, SD, Var

Example 1

a) Find the mean of a set of numbers

3, 5, 7, 4, 5, 9, 6

b) Find the mean of a set of data

Number of Male Children	0	1	2	3	4	5
Frequency	2	5	7	3	2	1

Solution

a) $\bar{x} = \frac{\sum x}{n}$

b) $\bar{x} = \frac{\sum fx}{\sum f}$

Median is the middle value when a set of data is arranged in ascending order then choose the middle point.

For a set of data $x_1, x_2, x_3, \ldots x_n$ arranged in ascending order, there are two cases.

a) When the number of data (n) is **odd**, the median is the $\left(\frac{n+1}{2}\right)^{th}$ observation.

b) When the number of data (n) is **even**, the median is the mean of the $\frac{n}{2}$ th value and the $\left(\frac{n}{2}+1\right)$ th value.

Example 2

Find the median for the following sets of data

a) 21, 24, 17, 28, 36, 20, 32

b) 3.56, 2.71, 5.48, 8.61, 4.35, 6.22

Solution

a) Median = $\left(\frac{n+1}{2}\right)^{th}$ observation

Median = $\left(\frac{7+1}{2}\right)^{th}$ observation

= 4th observation

=

b) 2.71, 3.56, 4.35, 5.48, 6.22, 8.61

Median =

Mode of a set of data is the value that occurs most frequently.

Example 3

Find the mode for the following set of data

a) 5, 2, 3, 3, 5, 4, 28, 5

b) 2, 3, 5, 8, 10

c) 0.2, 0.4, 0.4, 0.4, 0.5, 0.7, 0.7, 0.7, 0.5

Solution

a) 2, 3, 3, 4, 5, 5, 5, 28

Mode =

b) 2, 3, 5, 8, 10

Mode =

c) 0.2, 0.4, 0.4, 0.4, 0.5, 0.5, 0.7, 0.7, 0.7

Mode =

NOTES

a) if mean = median = mode → The distribution is symmetrical

b) if mode < median < mean → The distribution is skewed to the right

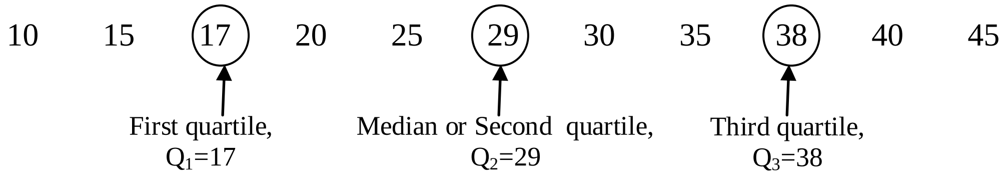
c) if mean < median < mode → The distribution is skewed to the left

Quartiles

- **First quartile (Q_1)** is a number such that 25% (quarter) of the total number of data has values less than Q_1.
- **Second quartile (Q_2)** also called **median** is a number such that 50% (half) of the total number of data has values less than Q_2.
- **Third quartile (Q_3)** is a number such that 75% (three quarter) of the total number of data has values less than Q_3.

Quartiles of ungrouped data

Quartiles are values which divide a set of data arranged in **ascending or descending order** into 4 equal parts as shown below.



10 15 17 20 25 29 30 35 38 40 45

First quartile, $Q_1=17$

Median or Second quartile, $Q_2=29$

Third quartile, $Q_3=38$

Example 4

Find the first quartile, median and third quartile for the following set of data.

a) 114, 120, 133, 138, 145, 148, 151

b) 23, 47, 32, 34, 42, 35, 44, 36, 52, 40, 42, 46

Solution

a) Arrange the data in ascending order, that is

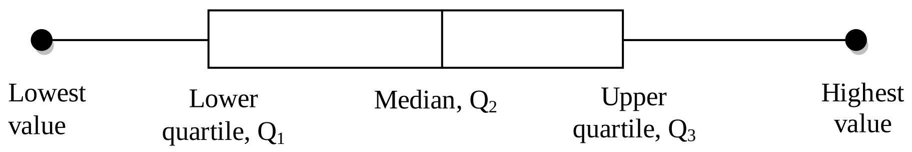
b) Arrange the data in ascending order, that is

Box-and-Whisker Plots

A Box-and-Whisker Plot gives a graphical description of data (a box and two whiskers) using 5 measures, namely the **median, first quartile, third quartile, smallest (min) and largest (max) values** in the data set.

It shows the *center, spread and skewness* of a set of data.

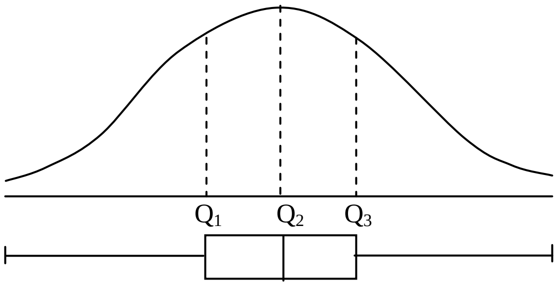
It can be represented horizontally.



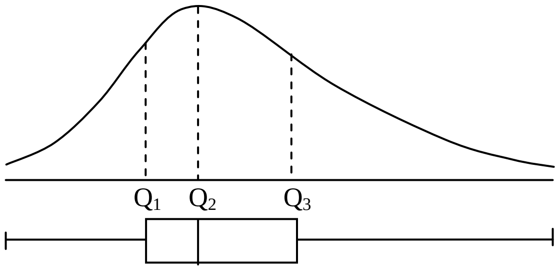
The box extends from Q_1 to Q_3 and encloses the middle 50% of the data. The whiskers extend from the box to the lowest and highest values and illustrate the range of the data.

Shape of data distribution – symmetry and skewness

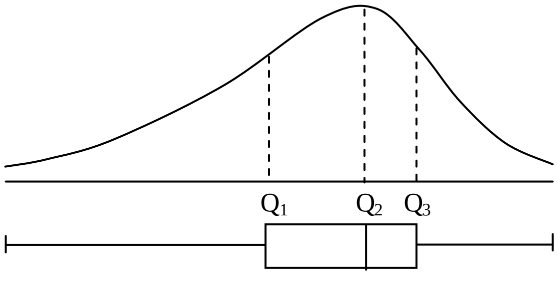
Symmetrical distribution – The ‘whiskers’ are the same length and the median is in the centre of the box.



Positively skewed distribution – the left ‘whiskers’ is shorter than the right ‘whiskers’ and the median is nearer to the Q_1.



Negatively skewed distribution – the left ‘whiskers’ is longer than the right ‘whiskers’ and the median is nearer to the Q_3.



To construct a Boxplot:

Step 1: Determine Q_1, Q_2, Q_3 and interquartile range (**IQR = Q_3- Q_1**) from the ranked data.

Step 2: Find the i) Lower fence = $Q_1 - 1.5$ IQR
ii) Upper fence = $Q_3 + 1.5$ IQR

Step 3: Determine the smallest and largest values within the two fences.

Step 4: Draw horizontal line and mark the expenses levels (all the values in the data set are covered). Check if there are values fall outside the two inner fences (outliers).

Example 5

The following data are the salaries (RM) for a sample of 15 households.

550 600 800 850 750 650 420 550
1200 680 900 1000 900 888 550

Construct a box-and-whiskers for the data and interpret it.

Solution

S1: Ranked data - 420 550 550 550 600 650 680 750 800 850 888
900 900 1000 1200

Q_1 = Q_2 = Q_3 = IQR =

S2 : Lower fence =
Upper fence =

S3: Smallest value Largest value

S4:

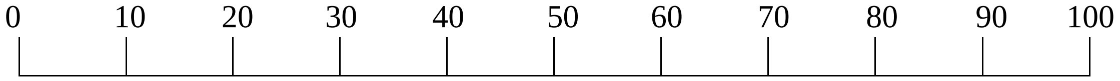
Example 6

The following data shows a summary of the marks for Mathematics and Biology for students in a class.

Subjects	**Minimum**	**Maximum**	**Median**	**First quartile**	**Third Quartile**
Mathematics	10	90	60	45	70
Biology	35	85	60	48	72

Draw two boxplots for this data and give comments regarding the distribution of marks for Mathematics and Biology.

Solution

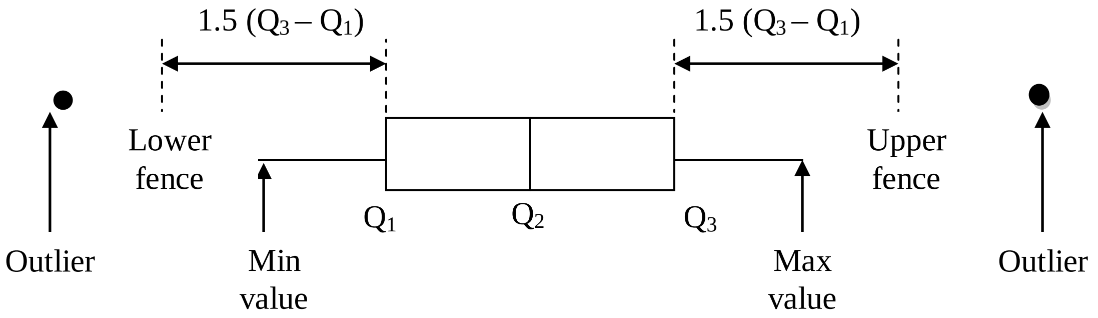


Comment:
The median marks for both subjects are the same but the marks for Mathematics show bigger range and it is skewed to the left. The Biology marks have a smaller range compared to Mathematics and the distribution of marks for Biology is almost symmetry.

Use of boxplots to identify 'outliers'

Sometimes, values which are unusually small or large occur in a set of data. The unusual values (outliers) occur probably because of an error in recording the data.

As a guide, points which are 1.5 times the interquartile range more than the third quartile or less than the first quartile are called 'outliers'.



Note: Upper fence = $Q_3 + 1.5\,(Q_3 - Q_1)$
Lower fence = $Q_1 - 1.5\,(Q_3 - Q_1)$

Example 7

The following stem-plot shows the maximum temperature for each day from 1st August to 23rd August in a town. Draw a boxplot and use the boxplot to identify the 'outliers'.

Stem	Leaf
5	1
5	9
6	2 3 3 4 4 4 4 4
6	5 7 8 8 8 9 9
7	0 2 2 3
7	6 7

Solution

Number of observations, n =

First quartile, Q_1 =

Median, Q_2 =

Third quartile, Q_3 =

Upper fence = $Q_3 + 1.5\ (Q_3 - Q_1)$
=

Lower fence = $Q_1 - 1.5\ (Q_3 - Q_1)$
=

Hence, the temperature that may be recorded wrongly is

For boxplot, the 'whisker' is drawn from $59^0 F$ to $77^0 F$.

LECTURE 3 OF 5

TOPIC : 6.0 DATA DESCRIPTION

SUBTOPIC : 6.2 Measures of Location

LEARNING OUTCOMES : At the end of this lesson, students will be able to:

a) find and interpret the mean, mode, median, quartiles and percentile for grouped data.

CONTENT

Mean, mode and median for Grouped Data.

Mean

If a set of grouped data given in frequency distribution, for example in the form of class intervals, the mean is defined as :

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_kx_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} f_ix_i}{\sum_{i=1}^{k} f_i} = \frac{\sum fx}{\sum f}$$

Where x_i is the **midpoint** of the i^{th} class and f_i is the corresponding **frequency**.

To key-in the data in calculator fx570ms:

$\boxed{\text{MODE}} \Rightarrow \boxed{\text{SD}}$

Key in data : x_1 $\boxed{\text{SHIFT}}$ $\boxed{,}$ f_1 $\boxed{\text{M+}}$

x_2 $\boxed{\text{SHIFT}}$ $\boxed{,}$ f_2 $\boxed{\text{M+}}$

x_n $\boxed{\text{SHIFT}}$ $\boxed{,}$ f_n $\boxed{\text{M+}}$

$\Rightarrow \boxed{\text{AC}}$

$\boxed{\text{SHIFT}} \Rightarrow \boxed{1} \quad \Rightarrow \sum x, \ \sum x^2$

$\boxed{\text{SHIFT}} \Rightarrow \boxed{2} \quad \Rightarrow \bar{x}$, SD, Var

To key-in the data in calculator fx570es:

$\boxed{\text{SHIFT}} \Rightarrow \boxed{\text{SET UP}}$ $\boxed{\blacktriangledown} \Rightarrow \boxed{\text{STAT}} \Rightarrow \boxed{\text{ON}}$

$\boxed{\text{MODE}} \Rightarrow \boxed{\text{STAT}} \Rightarrow \boxed{\text{1-VAR}}$

Key in data : $x_1 \boxed{=} x_2 \boxed{=} \ldots \boxed{=} x_n$

$f_1 \boxed{=} f_2 \boxed{=} \ldots \boxed{=} f_n$

$\Rightarrow$ AC

$\boxed{\text{SHIFT}} \Rightarrow \boxed{1} \Rightarrow \boxed{3} \Rightarrow \sum x, \ \sum x^2$

$\boxed{\text{SHIFT}} \Rightarrow \boxed{1} \Rightarrow \boxed{4} \Rightarrow \bar{x}$, SD, Var

Example 1

Calculate the mean of a life content of 40 batteries.

Time (years)	Number of batteries
1.5 - 1.9	2
2.0 - 2.4	1
2.5 - 2.9	4
3.0 - 3.4	15
3.5 - 3.9	10
4.0 - 4.4	5
4.5 - 4.9	3

Solution

Time (years)	Number of batteries	Midpoint (x_i)
1.5 - 1.9	2	
2.0 - 2.4	1	
2.5 - 2.9	4	
3.0 - 3.4	15	
3.5 - 3.9	10	
4.0 - 4.4	5	
4.5 - 4.9	3	

Mode

Mode can be calculated by using formulae:

$$\text{Mode} = L_B + \left(\frac{d_1}{d_1 + d_2}\right)C$$

Where;

L_B = lower class **boundary** of mode class

d_1 = the different between the mode class frequency and the **previous** class frequencies

d_2 = the different between mode class frequency and the class frequency **after** the mode class frequency.

C = class width

Example 2

Find the mode of frequency distribution given below :

Class Interval	Frequency
15 - 19	1
20 - 24	4
25 - 29	22
30 - 34	35
35 - 39	20
40 - 44	8

Solution

Class Interval	Frequency
15 – 19	1
20 – 24	4
25 – 29	22
30 – 34	35
35 – 39	20
40 – 44	8

Class boundary : 29.5-34.5

Mode Class

Mode =

Median

Median is the value for which 50% of the observations lie either side of it when arranged in ascending order.

The median class should be determined first before calculating the median.

The median lie at $\left(\frac{n}{2}\right)^{th}$ observations by referring to the cumulative frequency.

Median can be calculated by using formulae:

$$\text{Median} = L_k + \left(\frac{\frac{n}{2} - F_{k-1}}{f_k} \right) C$$

Where;

L_k = is the lower class boundary of median class

n = is the number of data or the sum of frequency

F_{k-1} = cumulative frequency before median class

C = class width

f_k = frequency of median class

Example 3

Find the cumulative frequency and calculate the median.

Class Interval	Frequency
1 - 5	1
6 - 10	3
11 - 15	5
16 - 20	7
21 - 25	13
26 - 30	9
31 - 35	7
36 - 40	3
41 - 45	2

Solution

Class Interval	Frequency	Cumulative frequency
1 - 5	1	
6 - 10	3	
11 - 15	5	
16 - 20	7	
21 - 25	13	
26 - 30	9	
31 - 35	7	
36 - 40	3	
41 - 45	2	

$Median = \left(\frac{50}{2}\right)^{th}$ observation = 25 th observation

Quartles for grouped data

$$Q_k = L_k + \left(\frac{\left(\frac{k}{4}\right)n - F_{k-1}}{f_k}\right)c_k \, ; \quad k = 1, 2, 3.$$

where L_k – lower boundary of the class where Q_k lies.

n – total number of observations.

F_{k-1} – cumulative frequency before the Q_k class.

f_k – frequency of the class where Q_k lies.

c – size of the class where Q_k lies.

Percentiles

For grouped data, the *k* th percentile,

$$\mathrm{P}_k = L_k + \left(\frac{\left(\frac{k}{100}\right)n - F_{k-1}}{f_k} \right) c_k \; ; k = 1, 2, 3, \ldots, 99.$$

where L_k – lower boundary of the class where P_k lies.

n – total number of observations.

F_{k-1} – cumulative frequency before the P_k class.

f_k – frequency of the class where P_k lies.

c – size of the class where P_k lies.

Note:

i. The 25 percentile is called the 1st quartile, Q_1.

ii. Median is the 50 percentile, also as the second quartile, Q_2.

iii. The 75 percentile is called the third quartile, Q_3.

iv. Interquartile range is the range between the 1st quartile and third quartile ($\mathrm{Q}_3 - \mathrm{Q}_1$).

Example 4

For the frequency distribution given below,

Class interval	Frequency
20 – 29	4
30 – 39	11
40 – 49	20
50 – 59	45
60 – 69	25
70 – 79	12
80 – 89	3

Find the

a) mean

b) mode

c) median

d) first quartile

e) third quartile

f) Interquartile range

g) 10th percentile.

h) the value of *m* if 30% of the data is greater than *m*.

Solution

Class interval	Class boundary	Midpoint, x	Frequency, f_i	Cumulative frequency, F
20 – 29				
30 – 39				
40 – 49				
50 – 59				
60 – 69				
70 – 79				
80 – 89				

LECTURE 4 OF 5

TOPIC : **6.0 DATA DESCRIPTION**

SUBTOPIC : **6.3 Measures of Dispersions**

LEARNING OUTCOMES : At the end of this lesson, students are able to

a) find and interpret the variance and standard deviation for ungrouped data.
b) find and interpret the variance and standard deviation for grouped data.

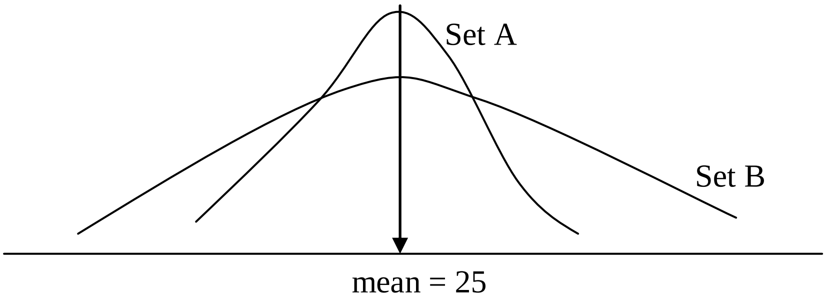
CONTENT

INTRODUCTION

Two sets of data might have the **same mean value** but may differ in their dispersion. For example, in the following data set,

Set A: 23, 24, 24, 25, 25, 25, 25, 26, 26, 27
Set B : 2, 6, 15, 22, 25, 25, 25, 30, 40, 60

have the same mean, that is, 25, but most of the numbers in the first set are around the mean value. On the other hand, the second set is more spread away from the mean.



The difference in the spread of data can be determined by the **measure of dispersion**.

Two common measures of dispersion are the variance and standard deviation.

A bigger value of variance or standard deviation means that the data are more spread out about the mean and less consistent.

Variance and standard deviation for ungrouped data

For ungrouped data

Variance, $s^2 = \dfrac{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}{n-1}$

Standard deviation, $s = \sqrt{\dfrac{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}{n-1}}$

If the variance increase by a constant k, new variance $= k^2 \times$ original variance.
If the standard deviation increase by a constant k, new variance $= k \times$ original standard deviation.

Example 1

Find the mean, variance and standard deviation for the data below.

2, 7, 10, 9, 2, 5, 16

Solution

Example 2

The following is the systolic blood pressure, in mm Hg, of 10 patients in a hospital.

165 135 151 155 158 146 149 124 162 173

a) Find the mean and the standard deviation of the systolic blood pressure of the 10 patients.
b) Find the number of patients whose systolic blood pressures exceed one standard deviation above or below the mean.

Solution

Example 3

The data below shows the marks obtained by Nik and Fizz in five tests:

Nik's marks	:	80, 80, 80, 80, 85
Fizz's marks	:	69, 78, 80, 80, 98

Find the mean and standard deviation for the above data. Which student shows a better overall performance?

Solution

Variance and standard deviation for grouped data

Variance, $s^2 = \dfrac{\sum fx^2 - \dfrac{\left(\sum fx\right)^2}{n}}{n-1}$

Standard deviation, $s = \sqrt{\dfrac{\sum fx^2 - \dfrac{\left(\sum fx\right)^2}{n}}{n-1}}$

Example 4

The frequency distribution table shows the masses of loaves of bread produced by a bakery.

Mass (g)	420 – 424	425 – 429	430 – 434	435 – 439	440 – 444
Frequency	16	24	25	18	17

a) Find the mean, variance and standard deviation.
b) The bakery allows only loaves of bread each with a mass of within one standard deviation from the mean to be sold in the market. Find the interval of mass of the loaves of bread allowed to be sold.

Solution

Mass (g)	***Midpoint, x***	***f***

a)

Example 5

The frequency distribution table shows the hourly wages of workers in a factory.

Wage (RM)	5 – 7	8 – 10	11 – 13	14 – 16	17 – 19
Frequency	9	16	11	8	6

a) Find the standard deviation, correct to three decimal places.
b) If the manager of the factory decides to increase the wage of each worker by 20%, find the new standard deviation.

Solution

Wage (RM)	***Midpoint, x***	***f***

LECTURE 5 OF 5

TOPIC : 6.**0** **DATA DESCRIPTION**

SUBTOPIC : **6.3** **Measures of Dispersion**

LEARNING OUTCOMES : At the end of this lesson, students will be able to :
(a) find and interpret the Pearson's coefficient of skewness.

SYMMETRY AND SKEWNESS

If mean = median = mode . The distribution is symmetrical.
If mode < median < mean . The distribution is skewed to the right.
If mean < median < mode . The distribution is skewed to the left.

Pearson's Coefficient of Skewness

The Pearson's coefficient of skewness is given by

$$\boxed{S_k = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}} \quad \text{or} \quad S_k = \frac{(\text{mean} - \text{mode})}{\text{standard deviation}}}$$

If $S_k < 0$ then the data distribution is skewed to the left / skewed negative .
If $S_k > 0$ then the data distribution is skewed to the right/ skewed positive.
If $S_k = 0$ then the data distribution is symmetrical.
(If $-0.1 < S_k < 0.1$, the term slightly is use)

Example 1

Given the following sorted data. Find the mean and median. Hence, calculate the Pearson's coefficient of skewness.

1.2, 1.5, 1.9, 2.4, 2.4, 2.5, 2.6, 3.0, 3.5, 3.8

Solution

Example 2

The frequency distribution of the age (in years) of 80 patients in a clinics is given in the table below.

Age	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40
Number of Patients	5	15	24	18	10	8

Find the mean and mode. Hence, calculate and interpret Pearson's coefficient of skewness given that the standard deviation is 6.798 years.

Solution

Age	x	f

Example 3

The time (in minutes) used by 120 students surfing the internet to perform a certain project is given in the following relative cumulative frequency table.

Time (x), in minutes	Relative cumulative frequency
$x \le 0$	0
$x \le 20$	$\frac{3}{40}$
$x \le 40$	$\frac{19}{60}$
$x \le 60$	$\frac{2}{3}$
$x \le 80$	$\frac{53}{60}$
$x \le 100$	1

Find

a) The median and mean.

b) Pearson’s skewness coefficient and comment on the value obtained.

Solution

Time (x), in minutes	x	Relative cumulative frequency	F	f