Topic 14: Statistical Data Handling

Notes:

1. The study of mathematical statistics includes 4 stages: **collection**, **organisation**, **presentation** and **interpretation of data**. When data is collected, organised in a meaningful manner and represented as visual statistical diagrams, we can look for trends and attempt to make predictions based on the data.

2. Gathering data requires an organised system of collecting the information. A simple way is to use tally marks to keep track of data collection. The use of tallies is most effective when the tally marks are recorded in an organised column chart as shown below. Grouping the tally marks in multiples of 5 allow for a faster evaluation of total items in the category. A title for the tally chart is required to provide the purpose of the data collection.

   Example of a tally chart:

   Pupils' Mode of Transport to School:

Transport	**Tally**	**Frequency**
Bus	//// //// ///	13
Car	//// //	7
Walking	////	5

Cycling	////	4

Pictogram:

3. A **pictogram** represents the quantity of data collected as pictures or icons. Each picture or icon is used to represent one or more units of data. A key or legend must be used to indicate the number represented by each picture or icon.

Bar Graph:

4. A **bar graph** or bar chart is a useful tool to organise discrete data. Bar graphs can be vertical or horizontal. The height or length of each bar represents the quantity of each category of items. The length of each bar is proportional to the amount that it represents. When constructing a bar chart it is important to choose a suitable scale to represent the frequency or quantity of the items.

   There are 3 main types of bar graphs:

   a) Vertical bar graph
   b) Horizontal bar graph
   c) Double bar graph

5. The **double bar graph** is used when we want to represent two sets of data on the same chart. We can put the bars side by side or we may put the bars of one set of data on top of the bars of the other set of data.

6. The choice of the forms of double bar graph, like **side by side** or **stacked**, depends on the main purpose of the graph, A side by side graph is more useful when we compare two sets of data, e.g. number of boys in a class compared to number of girls in a class, whereas the stacked chart emphasises the sum of two sets of data, e.g. total number of students in each class of the Sec 1 cohort.

Pie Chart:

7. A pie chart makes use of sectors in a circle. It displays each data as a percentage of the whole data set. The angle of a sector is proportional to the frequency or quantity of the data it represents, and it should have a label and its corresponding percentage or degree shown.

   The formula to determine the angle of a sector in pie chart is here:

   $\mathcal{A} = \frac{frequency\ of\ required\ item}{total\ frequency\ of\ all\ items}$ x 360°

   (Similarly, you can use an angle to find out the quantity for a category if you know the total)

Line Graph:

8. A **line graph** is formed by joining the points given by the data collected with straight lines. It is usually used to show the trend pattern over a period. This means that the horizontal axis is usually a time scale, e.g. minutes, days, months, years, weeks etc.

Appropriateness of Different Types of Statistical Diagrams:

9. The choice of an appropriate visual statistical diagram depends on the category and spread of the data collection as well as the type of analysis required. The strengths and weaknesses of the different types of statistical diagrams are shown in the following table:

Statistical Diagrams	**Strengths**	**Weaknesses**
Pictogram	• Easy to read • Visually appealing • Handles large data set easily with the uses of pictures and a legend	• Difficult to quantify a partial picture or icon • Time consuming to prepare as each picture or icon must be identical in shape or size • Only suitable for small quantity of categories; less than 5 • Low degree of accuracy in data presentation
Bar Graph	• Visually clear • 2 or 3 data sets can be easily compared	• Use only for discrete data

Pie Chart	• Visually appealing • Shows percentage of each category	• No exact numerical value • Difficult to compare between 2 data sets • Total quantity of the categories is unknown • Best for large number of categories like 3, 4, 5 or 6 • Use only for discrete data
Line Graph	• Shows a trend in data relationship • Allows projection of future values • Retains exact data values • Can be used for continuous data and hence estimation of values within the data set • Shows the minimum or	• Difficult to visualise the results in large data sets • Flat trend lines between 2 values may have no significant meaning

	maximum value and outliers	