Statistics Unit Notes

Table of Contents:

Statistical Terms:

Population:
A set of persons or objects considered in a statistical survey (all students in a school for example)

Sample:
A subset of a population (part of the school's population, i.e. all grade 7 students)

Census:
A statistical survey in which every element of a population is analyzed.

Survey:
Only a sample from the population is considered to gain insight into that population.

A sample is said to be representative of the population if it possesses the same characteristics of the population from which it was selected, and gives good insight to the population. If not representative, then it is said to be biased.

Study:
A statistical survey in which experts are consulted on areas that are targeting by the survey.

Tables of Values:
Used to summarize data collected during a statistical survey.

Size:
Total number of elements in the population or sample. $\underline{n}$
(sample size) = fraction
N (Population Size)

Data:
The facts or information collected for analysis

Variable:
The characteristic of each individual member of the population that is being studied. A variable can have different values.
i.e. the variable "number of siblings" has a value that can be 0, 1, 2, 3, 4, 5, etc.

Two Types of Variables:
1. **Quantitative:**
express a quantity and takes numerical values.
i.e. the variable number of siblings has a numerical value of 0, 1, 2, etc.
2. **Qualitative:**
express a quality and does not take numerical values. i.e. eye colour has values of blue, green, etc.

Frequency:
A value indicates the number of times that the value appears in the data.

Relative Frequency:
Is equal to the ratio of frequency to the total number of values. Often expressed as a percent.

Graphing Tips:
- **Always use a ruler.**
- **Label both axis and make a title.**
- **Use constant spacing and scale on each axis.**
- **Be neat.**

Types of Graphs:

Bar Graphs:
A bar graph allows for the illustration of a qualitative variable. We can construct a bar graph from a frequency table.

Circle Graphs (or pie chart):
Allows for the illustration of a qualitative variable. A circle graph represents each part of the whole (relative frequency). You will need to find the fraction value for each item and then turn it to a percent. Then find the percent of 360° (whole circle) to graph sectors for each item.

Broken-line Graphs:
Used to represent a variable that evolves in a continuous manner; variations in temperatures, the growth of a plant, the value of stock exchange. Composed of several consecutive segments connecting successive points.

Histogram:
Used to represent data that are grouped in classes. Bars touch as classes are continuous (one starts where the other

one ends). For example test scores grouped by intervals of percent, 10-20, 20-30, etc.

Measures of Central Tendency:

Sample data set of Grade 7 Math test scores (%):
$$72, 83, 64, 78, 91, 78, 64, 91, 78$$

Mean (Average):
You can find the average by finding the total of the data items and dividing by the number of items.
i.e. 72+83+64+78+91+78+64+91+78= 699 ÷ 9 = 77.67 or 78%

Median:
The median is the middle number in the data set. **The numbers must be in order from smallest to largest.

- If there is an odd number of data:

$$\frac{n+1}{2} \quad \text{(n=the number of data)}$$

  (9+1=10 ÷ 2= 5 so the 5th data is the median)
  i.e. $64, 64, 72, 78, 78, 78, 83, 91, 91$
  ↑

- If there is an even number of data:
  Get the median by taking the average of the two middle numbers. For example take the number of data for example (8 ÷ 2 = 4th data so take the average of the 4th and 5th data)
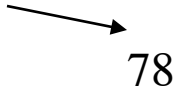 i.e. $64, 64, 72, 76, 78, 83, 91, 91$

$$76 + 78 = 154 \div 2 = 77$$

Mode:

The most frequently occurring data. The most popular data.

i.e. $64, 64, 72, 78, 78, 78, 83, 91, 91$

78

Range:
The measure of the dispersion of the data. The difference between the highest and lowest values. (maximum – minimum value)

i.e. $64, 64, 72, 78, 78, 78, 83, 91, 91$
so $91 - 64 = 27$ (range)