

Introduction to Statistics

In this unit we will learn:

- about the difference between a sample and the entire set of data
- about the different types of data
- different ways to measure how spread out the data is
- how to work with data that has been summarized

Quantitative vs Qualitative

Quantitative = Numerical data which can be counted or measured

Qualitative = Can be described in words or categories

Important Notes

It is challenging, but not impossible, to put qualitative data into quantitative terms.

Discrete vs Continuous

Discrete = Has an exact numerical value.

Continuous = Value that is measured, and its accuracy depends on the measuring tool used.

Example: Height in inches **Continuous**

Discrete

1. Decide which of these variables are continuous and which are discrete:

- (a) days in a month **Discrete**
- (b) shoe size **Discrete**
- (c) length of foot **Continuous**
- (d) weight of a gerbil **Continuous**
- (e) length of arm to the nearest centimetre **Discrete**
- (f) age in completed years. **Discrete**

Continuous

Discrete

Population vs Sample

Population = The entire group which we are interested in studying.

Sample = Subset of the population, or a group of data collected from the population.

Important Notes

1. The sample may not be representative of the population.
2. Even with a good sample, it is unlikely you will get the same statistics as you would for the entire population. (This can be shown mathematically!)

Random Samples can help us avoid problem #1 from the previous slide (samples not representative of the population).

Random Samples

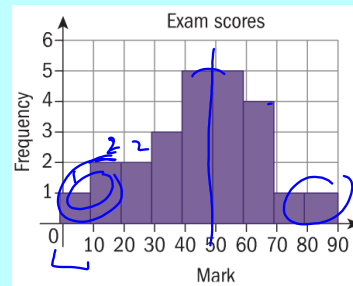
1. Each member of the population is equally likely to be selected
2. The probability of selecting any member of the population is independent of the selection of any other member.

2. For each of the following suggest why the sample taken is not random:

- (a) a questionnaire to find out attitudes towards a new shopping development taken on the street at 4 p.m. on a Saturday
- (b) interviewing students in a lesson about truancy rates
- (c) an internet survey of voting intentions
- (d) a cholesterol test from one person in each household in a street.

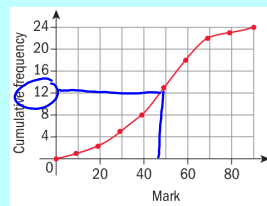
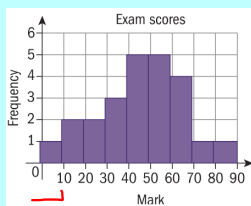
Histograms

Similar to bar charts/graphs, but with no gaps in between the bars. Also, the height of each bar is proportional to the frequency on that given interval.



Cumulative Frequency Diagrams

Diagrams which show the total/cumulative frequency of a data up to and including those at a given location.



Homework

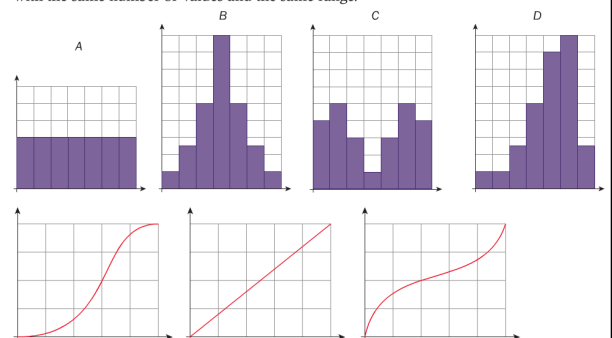
6A p.287 #2 (Histogram and CFD only) and #5

2. The table shows the age distribution of mathematics teachers who work at Caring High School.

- a. Is the data discrete or continuous?
- b. How many mathematics teachers work at Caring High School?
- c. Use your GDC to help you draw a fully labeled histogram to represent this data.

Age	Number of teachers
$20 \leq x < 30$	5
$30 \leq x < 40$	4
$40 \leq x < 50$	3
$50 \leq x < 60$	2
$60 \leq x < 70$	3

5. These histograms show four data sets A, B, C and D with the same number of values and the same range.



- a. Decide which data set A, B, C, or D goes with each of these cumulative frequency diagrams.
- b. Sketch a cumulative frequency diagram for the remaining data set.