Unit 14: Further Statistics

Sampling

When you are investigating a hypothesis, the population is the whole group that you are interested in.

A population is everything or everybody that could possibly be involved in an investigation

A population may divide into groups such as age range These groups are called strata.

In a stratified sample, the number of people taken from each group is proportional to the group size.

Stratified Sampling

The grouped frequency table shows information about the weights, in kilograms, of 20 students, chosen at random from Year 11.

Weight (w kg)	Frequency
$50 \le w < 60$	7
$60 \le w < 70$	8
$70 \le w < 80$	3
$80 \le w < 90$	2

There are 300 students in Year 11.

Work out an estimate for the number of students in Year 11 whose weight is between 50 kg and 60 kg.

$$\frac{7}{20} \times 300 = 105$$

Capture Recapture

Capture-Recapture is a technique that can be used to estimate the total population.

$$\frac{M}{N} = \frac{R}{T}$$

M = total marked
N = total population
R = number "recaptured"
T = total capture on 2nd visit

Clive wants to estimate the number of bees in a beehive.
Clive catches 50 bees from the beehive.

He marks each bee with a dye.
He then lets the bees go.

The next day, Clive catches 40 bees from the beehive.
8 of these bees have been marked with the dye.

(i) Work out an estimate for the number of bees in the beehive.

M = 50
T = 40
R = 8
N = ?

$$\frac{50}{n} = \frac{8}{40}$$

$$50 = \frac{8n}{40}$$

$$2000 = 8n$$

$$250 = n$$

A Cumulative Frequency table shows a running total of the frequencies

Mark	Frequency	Cumulative Frequency
1-10	3	3
11-20	6	3 + 6 = 9
21-30	11	9 + 11 = 20
31-40	13	20 + 13 = 33
41-50	18	33 + 18 = 51

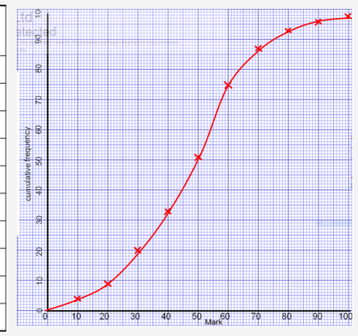
A cumulative frequency table shows how many data values Are less than or equal to the upper class boundary of each data class

The upper class boundary is the highest possible value in each class.

Cumulative Frequency

Drawing a Cumulative Frequency Graph

Mark	Frequency	Cumulative Frequency
1-10	3	3
11-20	6	9
21-30	11	20
31-40	13	33
41-50	18	51
51-60	24	75
61-70	12	87
71-80	6	93
81-90	3	96
91-100	2	98



Steps:

1) Start from 0
2) Plot using end points
3) Join using a smooth curve

Averages

1 1 3 5 7 9 11 14 16 19 19 20 21

n = 13

Median = $\frac{13+1}{2} = \frac{14}{2} = 7th\ value = 11$

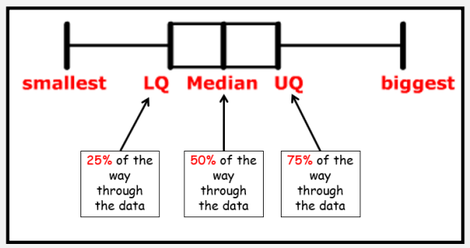
LQ = $\frac{13+1}{4} = \frac{14}{4} = 3.5th\ value = 4$

UQ = $3.5 \times 3 = 10.5th\ value = 19$

Median formula $= \frac{n+1}{2}$

Lower Quartile formula $= \frac{n+1}{4}$

Upper Quartile formula $= LQ \times 3$

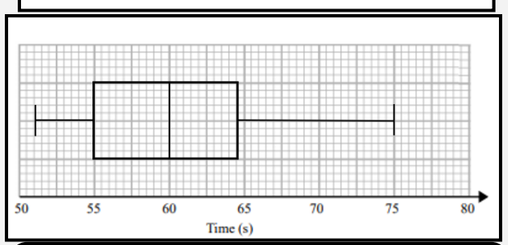


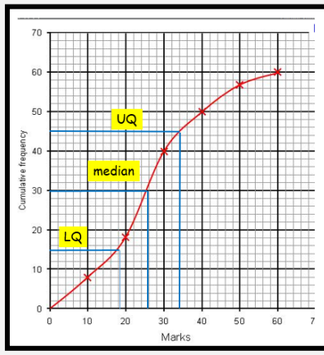
Drawing a box plot

The times, in seconds, of 15 students running a race are recorded below.

52 54 54 55 58 58 59 60 60 61 61 64 67 70 75

Draw a box plot for this information.





Median – Middle value found at total frequency ÷ 2

Lower quartile – Value which is a quarter of the way through the data found at total frequency ÷ 4

Upper quartile – Value which is three quarters of the way through the data Found at total frequency ÷ 4 x 3

Interquartile range = Upper quartile – Lower quartile

Median = 26

Lower Quartile = 18

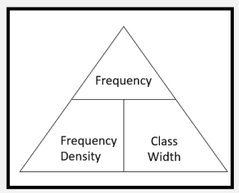
Upper Quartile = 34

IQR= 34 - 18 =16

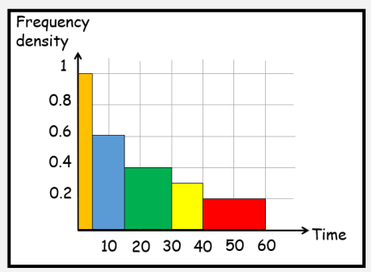
Box Plots + CF Diagrams

Drawing a histogram

Time, t, in minutes	Frequency density	Frequency
$0 \le t < 5$	1	5
$5 \le t < 15$	0.6	6
$15 \le t < 30$	0.4	6
$30 \le t < 40$	0.3	3
$40 \le t < 60$	0.2	4



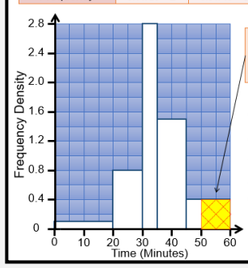
This is the formula triangle for a histogram



Interpreting a Histogram

The histogram shows the times a sample of students spent on the internet one evening. **(a)** Copy and complete the frequency table,
(b) Estimate how many students spent longer than 50 minutes on the internet.

Time, t, minutes	$0 \le t < 20$	$20 \le t < 30$	$30 \le t < 35$	$35 \le t < 45$	$45 \le t < 60$
Frequency	0.1 x 20 = **2**	0.8 x 10 = **8**	2.8 x 5 = **14**	1.5 x 10 = **15**	0.4 x 15 = **6**



50 Minutes or more

The area of the histogram represents area.

Area $50 \le t < 60$

= 0.4 x 10 = 4

4 students spent longer than 50 minutes on the internet.

Histograms