Statistics 1 Ch3 Representations of Data

BOX (AND WHISKER) PLOT

Min	Q1	Median	Q3	Max
3	10	12	17	27

Range

IQR

You can put the outlier boundary instead of the maximum (NOT BOTH)

The right box represents the same # of people as the left box (both 25%)
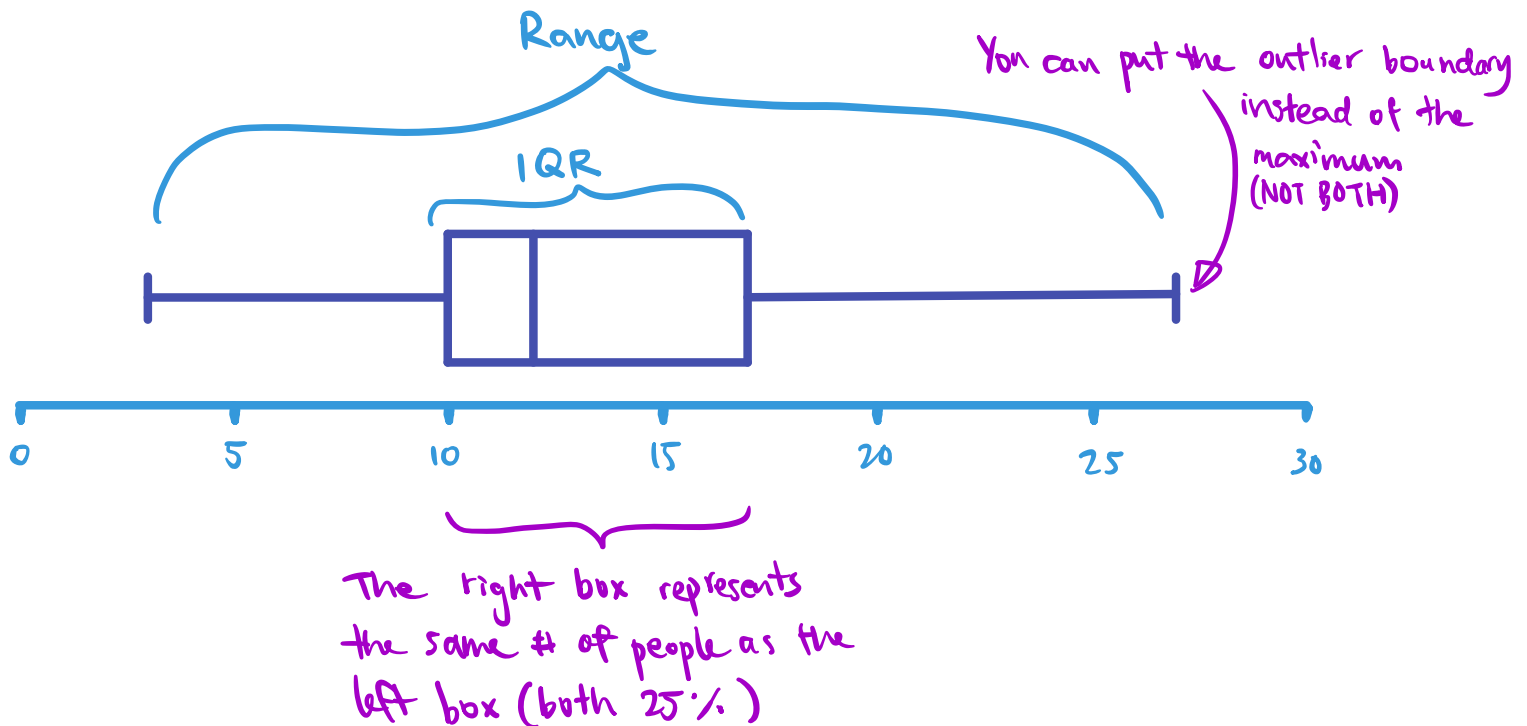
What is an outlier?

An extreme value

Usually 1.5 IQRs beyond Q1 or Q3

EXAMPLE

Diameter   2.2   2.5   2.9   2.7   2.8   3.0   3.1   3.1   3.2   4.0   4.7

$Q1 = 3^{rd}$ item = 2.7 $Q3 = 9^{th}$ item = 3.2

IQR = 3.2 - 2.7 = 0.5
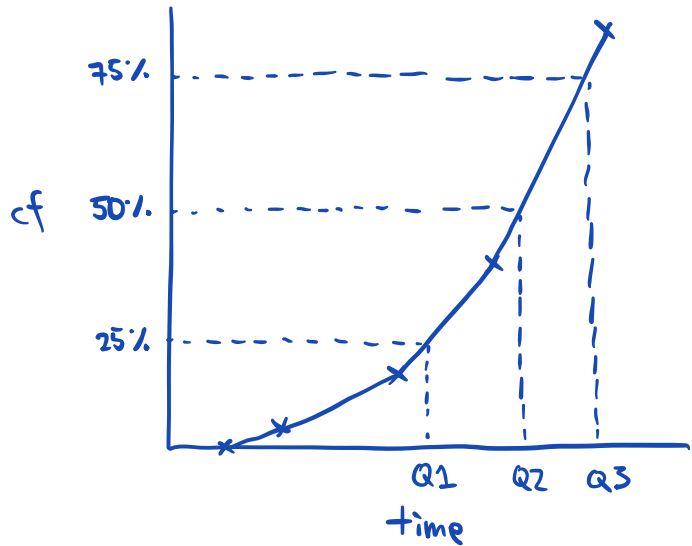
Lower outlier boundary = $2.7 - 1.5 \times 0.5 = 1.95$

Upper outlier boundary = $3.2 + 1.5 \times 0.5 = 3.95$

∴ Outliers are 4.0 and 4.7

αβγδεζηθικλμνξοπρστυφχψω
ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ

Cumulative frequency diagram

time (s)	frequency	cf
9.6-9.7	1	1
9.7-9.9	4	5
9.9-10.05	10	15
10.05-10.2	19	32

EX 3A (p.42)

② Male: IQR = 580-400 = 180g
    Lower boundary = 400-180 = 220g
    Upper boundary = 580+180 = 760g

Female: IQR = 340-260 = 80g
    Lower boundary = 260-80 = 180g
    Upper boundary = 340+80 = 420g

a) None.   b) 170g, 440g   c) 760g

④ $\Sigma x = 92$ $\Sigma x^2 = 1428$ $n = 9$

a) mean $= \frac{\Sigma x}{n} = \frac{92}{9} = 10.2$

$\sigma^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 = 158.7 - 104.5 = 54.2$
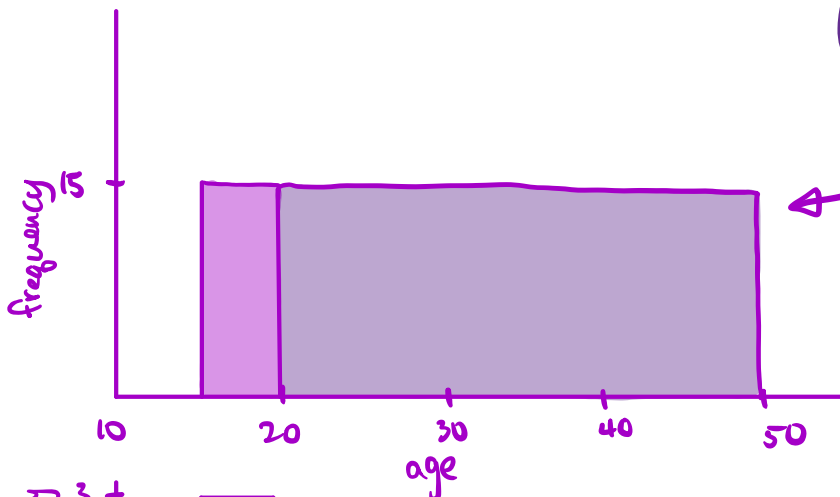
$\sigma = \sqrt{54.2} = 7.4$
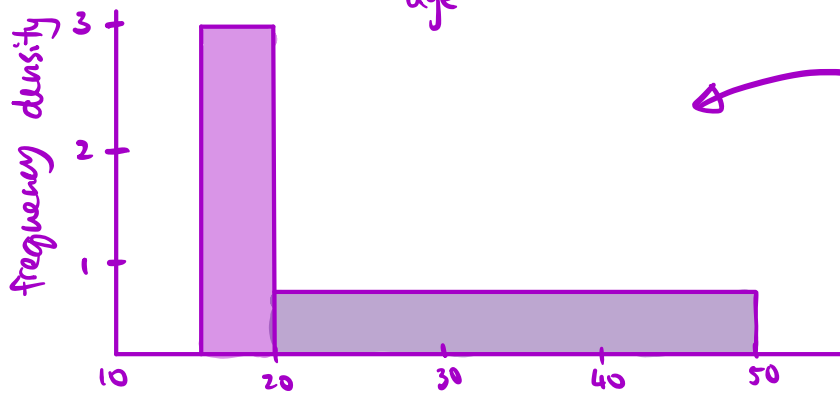
b) upper boundary $= 10.2 + 2(7.4) = 25.0$
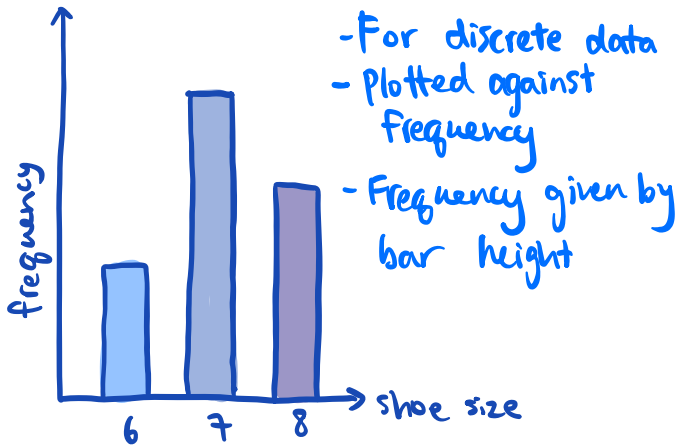∴ it is an outlier   ∵ 30 > 25

c) outliers can be a part of the data set.

age	frequency
$15 \le a < 20$	15
$20 \le a < 50$	15

Histograms

WRONG!

NICE!

BAR CHARTS

- For discrete data
- Plotted against Frequency
- Frequency given by bar height

frequency (y-axis)

shoe size (x-axis): 6 7 8

HISTOGRAM

- For continuous data
- Plotted against freq. density
- Freq. proportional to height × width

Frequency density (y-axis)

1.0 1.2 1.4 1.6 1.8
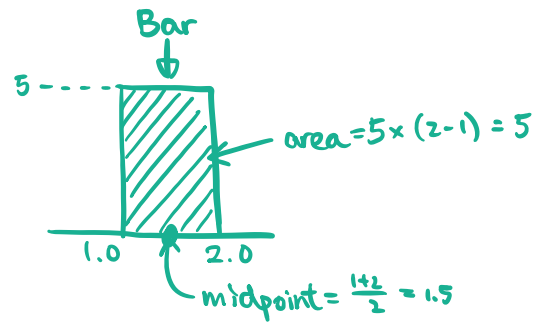
$$freq \propto area$$

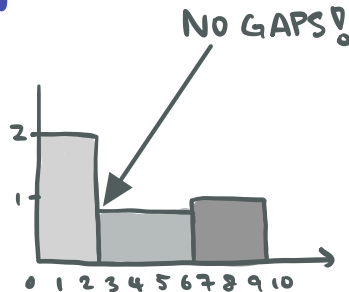$$freq = \underline{\underline{k}} \times area$$

Same for entire graph!
(can assume k=1 if no info given about frequency)

Finding the mean

$\bar{x}$ is $\frac{\Sigma fx}{\Sigma f}$ $\Sigma fx = \Sigma (k \times area \times bar\ midpoint)$

Bar

5 - - - -

area $= 5 \times (2-1) = 5$

1.0   2.0

midpoint $= \frac{1+2}{2} = 1.5$

Remember: NO GAPS!

Weight	Freq	FD.
1-2	4	4÷2=2
3-6	3	3÷4=0.75
7-9	3×1=3	1

NO GAPS!

0 1 2 3 4 5 6 7 8 9 10

this is actually 6.5 - 9.5!
∴ class width is 3, not 2!

① a)

frequency density — y-axis

Mass, m (pounds) — x-axis
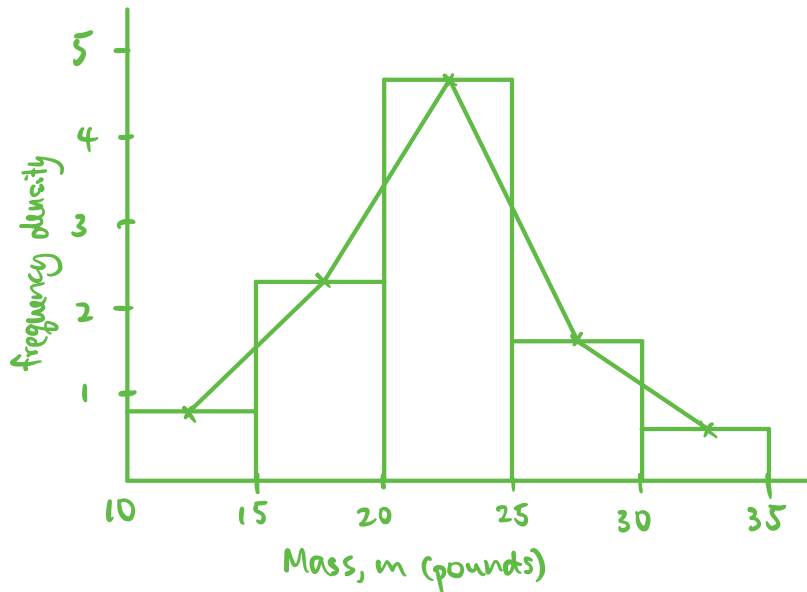
② a) time taken is continuous data

b) Class width = 20
   F.D. = 5   # of students = 20×5 = 100

c) 8.6×10 + 6×10 + 5×20 = 246

d) 246 + 14×5 + 12×5 + 3×30 = 466

③ a) distance is continuous data

b) 2×20 + 5×15 + 10×10 + 6×15 + 1×5 = 310

c) 5×5 + 10×5 = 25 + 50 = 75

d) 6×15 + 1×5 = 95

e) 2×20 + 5×5 = 40 + 25 = 65

④ a) Area of 28 ≤ m < 32 : 10×10 = 100 small squares = 32 lambs
                              25 small squares = 32/4 lambs = 8 lambs
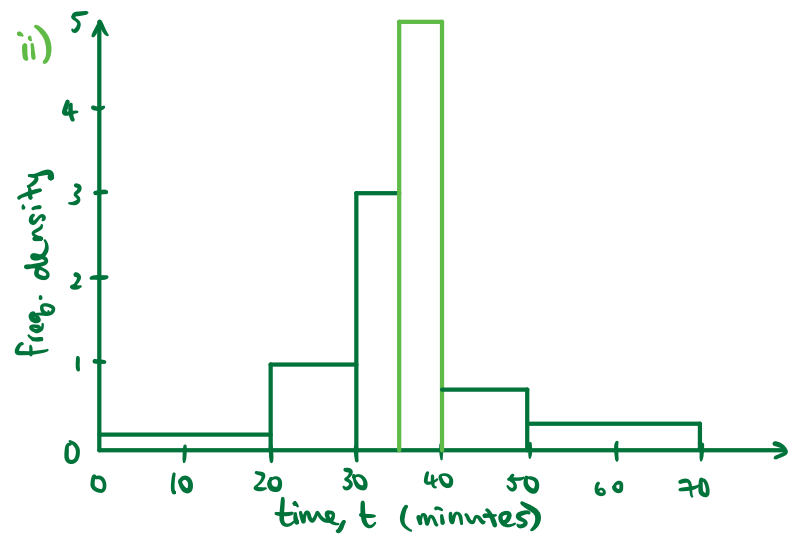
b) 20×5 = 100 small squares = 32 lambs

c) 10×10 + 20×5 + 40×5 + 10×10 + 5×5 = 525 small squares
   $32 \times \frac{525}{100} = 168$ lambs

d) 20×2.5 + 40×5 + 10×2.5 = 275 lambs

⑤ a) i)

time, t (min)	frequency
$0 \leq t < 20$	4
$20 \leq t < 30$	10
$30 \leq t < 35$	15
$35 \leq t < 40$	25
$40 \leq t < 50$	7
$50 \leq t < 70$	6

ii)

b) $1 \times 10 + 3 \times 5 + 5 \times 3 = 40$ people

⑥ a) $12.5 \leq y < 14.5$

b) Area = $4 \times 6 = 24$
Frequency = 24   ∴ k=1

i) 6 cm

ii) $18 \div 6 = 3$ cm