STATSTICS 1 CH2 : VARIABLES !

VARIABLES : STATISTICS VS ALGEBRA

vars in statistics

- represents the value of multiple objects (like a set)
- operations can be performed on th var (e.g. Σx = sum of x)
- $\bar{x}$ is the mean of x
- we can attatch a probability to it

Both

- represents values
- can be discrete or continuous
- can be part of calculations (eg. $2x$ = twice the value)

vars in algebra

- represents a single value (e.g. $x = 1$)
- operations such as Σx is meaningless as x is a single value
- we cannot attatch a probability to it

Measures of location	
Maximum Minimum Quartiles Deciles Percentiles Deciles	<u>Measures of central tendency</u> Mean Median Mode

Measures of spread	
Range	Standard deviation
Interquartile range	Variance

Diameter of coin (x cm)   2.2   2.5   2.6   2.65   2.9

$\bar{x}$ (ie the average) = 2.57

USING THE CLASSWIZ:

Menu → Statistics (6)

1-Variable (1) → Enter data points

$\boxed{AC}$ to execute calculation

OPTN → 1-variable calc (2)

frequency tables

# of children (x)	frequency (f)
0	4
1	3
2	9
3	2

SHIFT → SETUP

$\boxed{\downarrow}$ → STATISTICS (3)

Frequency ON/OFF

height of bear	frequency (f)
$0 \leq h < 0.5$	4
$0.5 \leq h < 1.2$	20
$1.2 \leq h < 1.5$	5
$1.5 \leq h < 2.5$	11

midpoint of range $= x$ $\bar{x} = \dfrac{\Sigma fx}{\Sigma x} = 1.17$ m

this mean is just an estimate

P.23 EX2A

5

Breakdowns	0	1	2	3	4	5
Frequency	8	11	12	3	1	1

a mode: 2

b median: 1

c mean: 1.47

d c because there are not a lot of outliers and the mean is very representative

7

# of Eggs	1	2	3
Frequency	7	p	2

if mean = 1.5, p = 2

P.24 EX2B

2 a $\bar{x} = 82.3$ dB

b because the data is grouped.

3 a $10 \leq t < 12$

b 11.4°c

4 $\bar{x}_A = 49.8$ Shop B is better at employing old workers

$\bar{x}_B = 51.0$

Combined Mean

Class A (20 pupils) mean: 62

Class B (30 pupils) mean: 75

$$\text{Mean} = \underbrace{\frac{(20 \times 62) + (30 \times 75)}{50}}_{\text{weighted average}} = 69.8$$

Median

Listed data

of data: n

if $n/2$ is decimal, round up

if $n/2$ is whole, halfway between $n/2$ and $(n/2)+1$

Grouped data

of data: n

find $n/2$ and then use linear interpolation

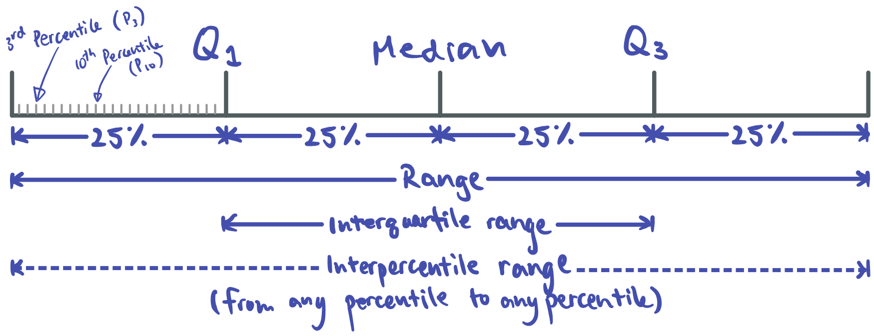
(DO NOT ROUND $n/2$)

Quartiles

Lower Quartile (LQ / Q1) just like median:

if $n/4$ is decimal, round up

if $n/4$ is whole, halfway between $n/4$ and $(n/4)+1$



Upper Quartile (UQ / Q3) just like median:

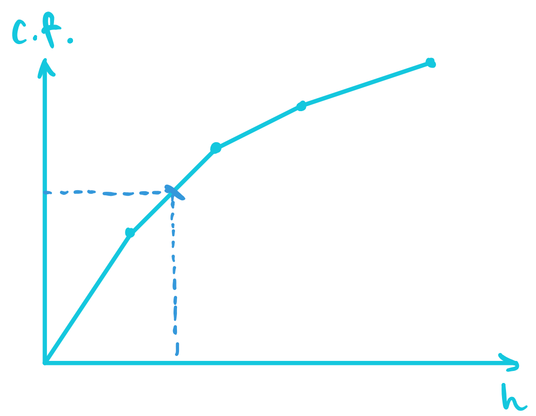
if $3n/4$ is decimal, round up

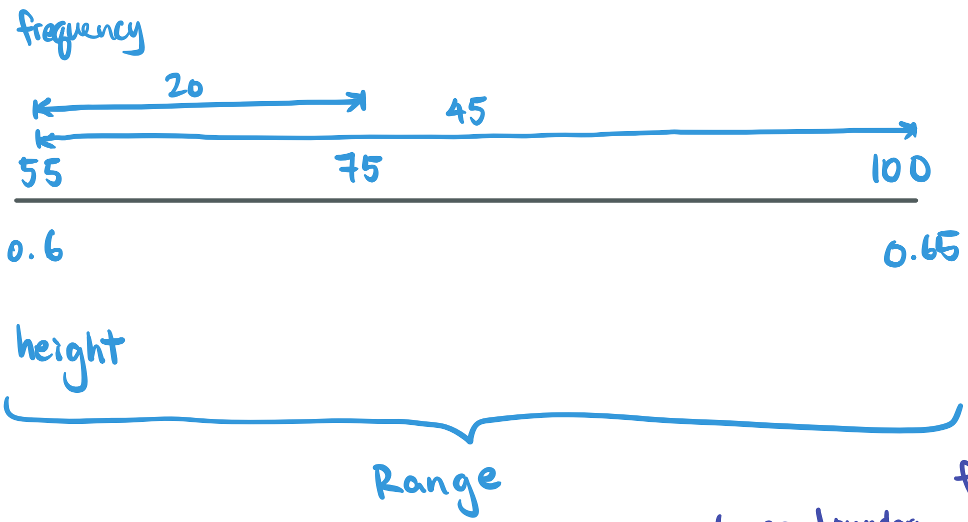
if $3n/4$ is whole, halfway between $3n/4$ and $(3n/4)+1$

Linear Interpolation

height of tree	f	c.f.
$0.55 \le h < 0.6$	55	55
$0.6 \le h < 0.65$	45	100
$0.65 \le h < 0.7$	30	130
$0.7 \le h < 0.75$	15	145
$0.75 \le h < 0.8$	5	150

Median position $= \frac{150}{2} = 75$





we are $\frac{20}{45}$ through the range

fraction

$$\text{Median} = \underbrace{0.6}_{\text{lower boundary}} + \left(\frac{20}{45} \times \overbrace{(0.65 - 0.6)}^{\text{class width}}\right)$$

$$= 0.622$$

⚠ e.g.

x	Freq
12 - 13	?
14 - 15	?
16 - 17	?

NOT 16–17

IT IS 15.5 - 17.5

Ex 2C (p. 27)

1 1009, 1013, 1014, 1017, 1017, 1017, 1018, 1019, 1021, 1022, 1024, 1024, 1025, 1027, 1029, 1031

a median = 1020

b Q1 : 1017

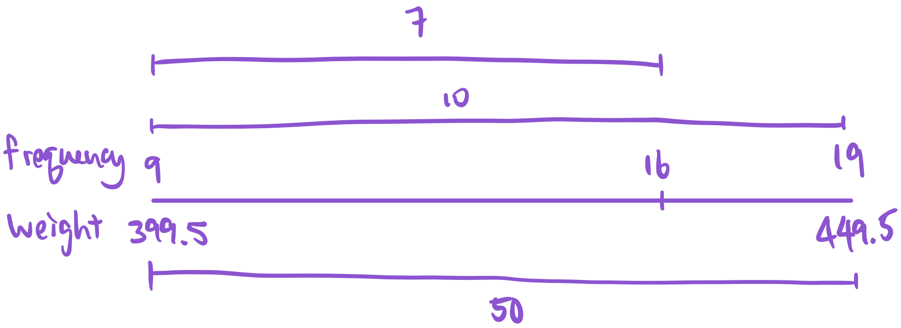
Q3 : 1024.5

2 $Q_1 = 37$

$Q_2 = 37$

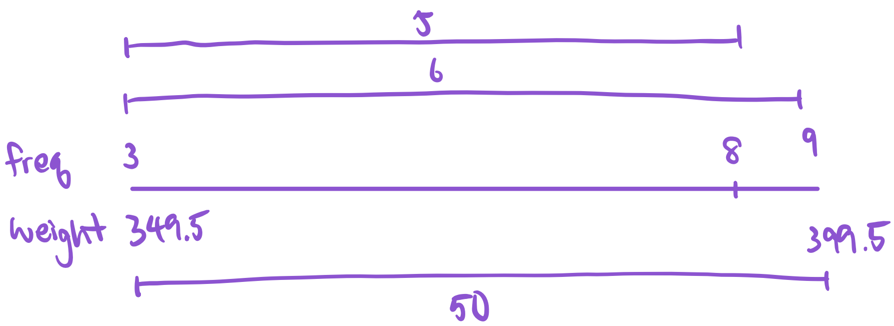
$Q_3 = 38$

4 a $\frac{3+6+10+7+5}{2} = 15.5$ index = 16



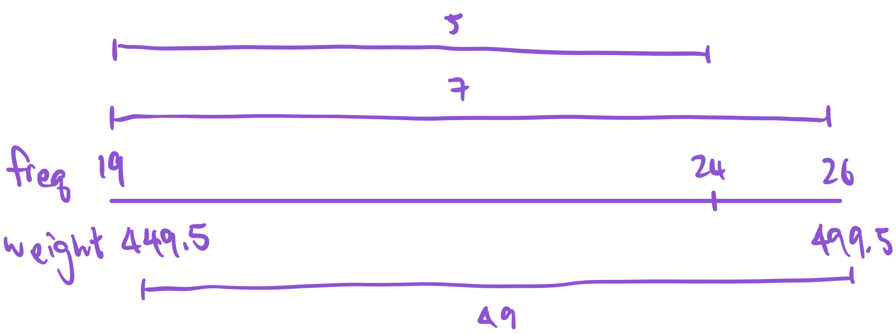
median = $400 + \left(\frac{7}{10} \times 50\right) = 435$

b $\frac{3+6+10+7+5}{4} = 7.75$ index = 8



$Q_1 = 350 + \left(\frac{5}{6} \times 50\right) = 391.7$

c $\frac{(3+6+10+7+5)3}{4} = 23.25$ index = 24



$Q_3 = 450 + \left(\frac{5}{7} \times 50\right) = 485.7$

d about 25% of the cows are heavier than 485.7 kg and 75% are lighter

5 a $\frac{6+10+18+13+2}{2} = 24.5$ index $= 26$

group: $40 \le t < 50$ range: 10

frequency: 18 index within group: 10

median $= 40 + \left(\frac{10}{18} \times 10\right) = 45.6$

b $\frac{(6+10+18+13+2)65}{100} = 31.85$ index $= 32$

group: $40 \le t < 50$ range: 10

frequency: 18 index within group: 16

$P_{65} = 40 + \left(\frac{16}{18} \times 10\right) = 48.9$

c $\frac{(6+10+18+13+2)90}{100} = 44.1$

group: $50 \le t < 60$ range: 10

frequency: 13 index within group: 10.1

$P_{90} = 50 + \left(\frac{10.1}{13} \times 10\right) = 57.8$

claim is invalid

VARIANCE & STANDARD DEVIATION

σ^2 = variance: $\frac{\Sigma(x-\bar{x})^2}{n}$ or $\frac{\Sigma x^2}{n} - \bar{x}^2$

the average squared distance from the mean

σ = standard deviation: $\sqrt{\sigma^2} = \sqrt{\text{variance}} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

the average distance from the mean

Grouped data?

Variance $= \frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2$

Ex 2E (p.32)

1 a mean $= \frac{\Sigma x}{n} = 3$

b $\sigma^2 = \frac{\Sigma x^2}{n} - \bar{x}^2 = \frac{78}{8} - 3^2 = 0.75$

c $\sigma = \sqrt{\sigma^2} = 0.866$

2 $\sigma^2 = \frac{\Sigma w^2}{n} - \left(\frac{\Sigma w}{n}\right)^2 = \frac{5905}{10} - \left(\frac{241}{10}\right)^2$

$= 590.5 - 580.81$

$= 9.69$

$\sigma = \sqrt{9.69} = 3.11$

3 a mean $= \frac{165+170+190+180+175+185+176+184}{8}$

$= 178.125$

b $\sigma^2 = \frac{\Sigma h^2}{n} - \bar{h}^2 = \frac{254307}{8} - (178.125)^2$

$= 59.86$

c $\sigma = 7.74$

4 mean $= \frac{50+86}{25} = 5.44$

$\sigma^2 = \frac{310+568}{25} - (5.44)^2$

$= 5.5264$

$\sigma = 2.35$

5 a $\bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{869}{84} \approx 10.35$

$\Sigma f x^2 = 9039$

$\sigma^2 = \frac{9039}{84} - \left(\frac{869}{84}\right)^2 = 0.583$

$\sigma = 0.764$

b $10.35 + 0.764 = £11.114$

20 students have $> £11.114$

IMAGINE...

After "stretching", everyone's height has tripled!

what happens to the variance? $\times 3^2 = \times 9$

What happens to σ? $\times 3$

proof: $\sigma^2 = \frac{\Sigma(3x)^2}{n} - \left(\frac{\Sigma(3x)}{n}\right)^2$

$$= \frac{\Sigma 9x^2}{n} - \left(\frac{3\Sigma x}{n}\right)^2 = \frac{9\Sigma x^2}{n} - 9\left(\frac{\Sigma x}{n}\right)^2$$

$$= 9\left(\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2\right)$$

CODING();

Coding	Effect on $\bar{x}$	Effect on σ
$y = x + 10$	$\bar{y} = \bar{x} + 10$	$\sigma_y = \sigma_x$
$y = 3x$	$\bar{y} = 3\bar{x}$	$\sigma_y = 3\sigma_x$
$y = 2x - 5$	$\bar{y} = 3\bar{x} - 5$	$\sigma_y = 2\sigma_x$
$y = ax + b$	$\bar{y} = a\bar{x} + b$	$\sigma_y = a\sigma_x$

⚠ when squaring ($y = x^2$), $\bar{y} = \bar{x}$ because $\Sigma x^2 = (\Sigma x)^2$

Application of coding:

cost x of diamond ring (£)

£1010 £1020 £1030 £1040 £1050

"code" the variable: $y = \frac{x - 1000}{10}$

£1 £2 £3 £4 £5

$\Downarrow$

$\sigma_y = 1.41$

$\sigma_x = 14.1$

Coding used to simplify the data

Ex 2F (p.34)

① $x: 110, 90, 50, 80, 30, 70, 60$

a) $y = \frac{x}{10}$ $11, 9, 5, 8, 3, 7, 6$

b) $\bar{y} = 7$

c) $\bar{y} = \frac{\bar{x}}{10}$

$\bar{x} = 10\bar{y} = 70$

② $x: 52, 73, 31, 73, 38, 80, 17, 24$

a) $y = \frac{x-3}{7}$

$7, 10, 4, 10, 5, 11, 2, 3$

b) $\bar{y} = 6.5$

c) $\bar{y} = \frac{\bar{x}-3}{7}$

$\bar{x} = 7\bar{y} + 3 = 48.5$

③ $y = \frac{x-65}{200}$ $\bar{y} = 1.5$

$\bar{y} = \frac{\bar{x}-65}{200}$

$\bar{x} = 200\bar{y} + 65 = 365$

④ $y = x - 40$ $\sigma_y = 2.34$

$\sigma_x = 2.34$

⑤ a) $\bar{y} = 25.1$ hours

b) $\bar{y} = \frac{\bar{x}-1}{20}$

$\bar{x} = 20\bar{y} + 1 = 503$ hours

c) $\sigma_y = 1.76$

$\sigma_y = \frac{\sigma_x}{20}$

$\sigma_x = 20\sigma_y = 35.2$

⑥ $y = \frac{i-90}{100}$, $\Sigma y = 131$, $\Sigma y^2 = 176.84$

$\sigma_y = \frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2$

$= 1.7684 - (1.31)^2$

$= 0.0523 = \frac{\sigma_i}{100}$

$\sigma_i = 5.23$