2. Summarizing Data

We describe centre with mean and median
" spread " standard deviation (SD)

Mean, median and SD all have same units as original data.

The **mean** is the average value

$$\text{mean} = \frac{\text{sum}}{n} \leftarrow \# \text{ of measurements}$$

Notation: μ for population mean
$\bar{x}$ sample mean

Data is assumed to be a population unless "sample" is indicated

Ex: Find the mean of the sample
11, 9, 17, 19, 4, 15

sample mean $\bar{x} = \frac{11+9+\ldots+15}{6} = 12.5$

Ex: A student has test marks 58, 63 and 71. What mark on 4th test gives an average of 70?

Let mark = x

$$\frac{58+63+71+x}{4} = 70$$

$$192 + x = 280$$

$$x = 88$$

Ex: If the two populations are combined into one, find the new mean

	Pop. Size	μ
Pop. 1	43	71
Pop. 2	26	68

For Pop. 1 $\quad \mu_1 = 71$

$$\frac{\text{sum}}{43} = 71$$

$$\text{Sum} = 3053$$

For Pop. 2 $\quad \text{Sum} = 26 \cdot 68 = 1768$

Overall pop. mean $\quad \mu = \frac{\text{sum}}{n} = \frac{(3053+1768)}{(43+26)} \approx 69.9$

Given

value	frequency
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$

$$\text{mean} = \frac{\sum xf}{\sum f}$$

Ex: Find the sample mean

Temp (°C)	Frequency
22	11
23	6
25	3

$$\bar{x} = \frac{22\cdot 11 + 23\cdot 6 + 25\cdot 3}{20}$$

$$= 22.75\,^\circ C$$

Given

value	relative frequency
x_1	r_1
x_2	r_2
$\vdots$	$\vdots$

mean $= \sum x r$

Ex: Find the sample mean

mass (g)	rel. freq.
82	0.55
86	0.4
88	0.05

$$\bar{x} = 82(0.55) + 86(0.4) + 88(0.05)$$
$$= 83.9 \text{ g}$$

Median: middle value when data is ordered
If # of measurements is even, average the middle two measurements

No notation for median

<u>Ex</u>: Find the median

a) 2, 9, 11, 5, 6

→ 2, 5, 6, 9, 11

median = 6

b) 2, 9, 11, 5, 6, 10

→ $2, 5, \underbrace{6, 9}, 10, 11$

median = $\frac{6+9}{2} = 7.5$

The median is in position $\frac{n+1}{2}$
(when data is ordered)

<u>Ex</u>: a) If $n = 161$ position = $\frac{162}{2} = 81$

Median is 81^{st} measurement

b) If $n = 200$ position = $\frac{201}{2} = 100.5$

Median is average of 100^{th} and 101^{st} measurements

Ex: Find the median

Temp (°C)	Freq.
22	11
23	6
25	3

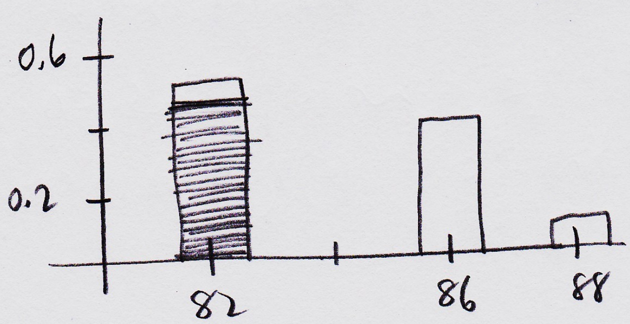
$\text{Position} = \frac{n+1}{2} = \frac{21}{2} = 10.5$

Median is average of 10^{th} and 11^{th}

Median = 22°C

Ex: Find the median

mass (g)	rel. freq.
82	0.55
86	0.4
88	0.05



Shade the first 50%

Median = 82g

Median is more representative than the mean when there are very high/low measurements

e.g. salaries in Canada

$$\boxed{\begin{array}{ll} \text{Population Variance} & \sigma^2 = \dfrac{\sum (x-\mu)^2}{n} \\ \text{Population SD} & \sigma = \sqrt{\sigma^2} \end{array}}$$

The more spread out the data, the bigger the SD is.

Ex: Find the pop. SD of 2, 5, 8, 9

x	$x-\mu$	$(x-\mu)^2$
2	-4	16
5	-1	1
8	2	4
9	3	9

$\mu = \dfrac{\text{sum}}{n}$

$= 6$

$\sigma^2 = \dfrac{\text{sum}}{n}$

$= \dfrac{30}{4}$

$= 7.5$

$\sigma = \sqrt{7.5} \simeq 2.7$

Sample Variance $s^2 = \dfrac{\sum (x-\bar{x})^2}{n-1}$

Sample SD $s = \sqrt{s^2}$

Later we'll use s^2 to estimate population data. Better approximation if we divide by n-1 instead of n.

Ex: Find the sample SD of 12, 13, 17

x	$x-\bar{x}$	$(x-\bar{x})^2$
12	-2	4
13	-1	1
17	3	9

$\bar{x} = \dfrac{\text{sum}}{n}$

$= 14$

$s^2 = \dfrac{\text{sum}}{n-1}$

$= \dfrac{14}{2}$

$= 7$

$s = \sqrt{7} \approx 2.6$

<u>Ex</u>: Which sample is more spread out?

a) 1, 4, 10

b) 31, 36, 38

$\boxed{\text{Which } s \text{ is larger?}}$

a) $s = \sqrt{21}$

b) $s = \sqrt{13}$

$\sqrt{21} > \sqrt{13}$

Sample a) is more spread out

Data is <u>accurate</u> : mean close to target value

<u>precise</u> : variance or SD is small

<u>Ex</u>: Two machines are filling 355 mL cans.

	Machine 1	Machine 2
$\bar{x}$	355.8	355.2
s^2	0.3	1.4

a) Which machine is more accurate?

Machine 2 : $\bar{x}$ is closer to 355

b) Which is more precise?

Machine 1: s^2 is smaller

Ex: Population = salaries at a small engineering firm. What happens to mean, median and SD in each situation?

a) each employee gets a \$2000 raise?

(55, 60, 80) → (57, 62, 82)

mean and median increase by \$2,000
SD does not change
(spread unchanged)

b) each salary is doubled?
mean, median and SD all double

(55, 60, 80) (110, 120, 160)

$\sigma \simeq 10.8$ $\sigma \simeq 21.6$

Notation

	Pop.	Sample
Mean	μ	$\bar{x}$
SD	σ	s
Variance	σ^2	s^2

On Sharp EL 531/520 :

Find $\mu, \bar{x}, \sigma, \sigma^2, s, s^2$ for 1, 4, 6, 8

$\boxed{\text{MODE}}$ $\boxed{1}$ $\boxed{0}$

1 $\boxed{\text{M+}}$
4 $\boxed{\text{M+}}$
6 $\boxed{\text{M+}}$
8 $\boxed{\text{M+}}$

$\boxed{\text{RCL}}$ $\boxed{\bar{x}}$ $\bar{x} = 4.75$

$\boxed{\text{ON/C}}$ to clear screen

$\boxed{\text{RCL}}$ $\boxed{\sigma x}$ $\sigma \approx 2.59$

$\boxed{x^2}$ $\boxed{=}$ $\sigma^2 = 6.6875$

$\boxed{\text{RCL}}$ $\boxed{sx}$ $s \approx 2.99$

$\boxed{x^2}$ $\boxed{=}$ $s^2 \approx 8.92$

Note : $\mu = \bar{x}$ always
$\mu = 4.75$

To clear data set
$\boxed{2^{nd}F}$ $\boxed{\text{ALPHA}}$ $\boxed{0}$ $\boxed{0}$

Find $\mu, \bar{x}, \sigma, \sigma^2, s, s^2$ for

X	Freq
1	4
3	1
5	5

$\boxed{\text{MODE}}\ \boxed{1}\ \boxed{0}$

1 $\boxed{\text{STO}}$ 4 $\boxed{\text{M+}}$

3 $\boxed{\text{STO}}$ 1 $\boxed{\text{M+}}$

5 $\boxed{\text{STO}}$ 5 $\boxed{\text{M+}}$

Recall values as above

$\mu = \bar{x} = 3.2$

$\sigma \simeq 1.9 \quad \sigma^2 = 3.56$

$s \simeq 1.99 \quad s^2 \simeq 3.96$

Can enter relative frequencies the same way as frequencies

$\boxed{\text{MODE}}\ \boxed{1}\ \boxed{0}$

1 $\boxed{\text{STO}}$ 0.4 $\boxed{\text{M+}}$

etc.