MEASURES OF SPREAD

A measure of the extent to which the values of a variable are spread out.

Range:
difference between the maximum value and the minimum value

Quartiles:
three values that divide a body of data into four equal parts

Interquartile Range (IQR):
difference between Q1 and Q3; 50% of the data lies in this range

Example:
Sondrine keeps track of the waiting time, in minutes, at two different fast-food restaurants. Find the quartile values—Q1, Q2, and Q3—and calculate the interquartile range (IQR) of both restaurants in order to compare the waiting times.

Burgers: 9.5, 0.9, 8.3, 9.1, 7.7, 5.2, 5.3, 7.3, 10.5, 3.6, 0.4, 3.7, 7.7, 6.5, 2.9, 8.9, 9.6, 8.9, 7.4, 7.1

'n Fries: 4.1, 3.3, 4.7, 5.8, 6.6, 4.4, 2.6, 7.3, 5.3, 5.2, 6.1, 3.2, 0.8, 5.0, 5.5, 6.5, 4.3, 4.4, 3.3, 3.5

Q1 4.45

	Burgers Time (min)		n Fries Time (min)		Burgers Time (min)		n Fries Time (min)
	0.4		0.8		7.4		4.7
	0.9		2.6		7.7		5.0
	2.9		3.2		7.7		5.2
	3.6		3.3		8.3		5.3
	3.7	3.4	3.3	Q3	8.9		5.5
	5.2	Q_1	3.5		8.9		5.8
	5.3		4.1		9.1		6.1
	6.5		4.3		9.5		6.5
	7.1		4.4		9.6		6.6
	7.3		4.4		10.5		7.3
Q2	7.35	Q_2	4.55				

5.65 Q_3

IQR $Q_3 - Q_1$ Burgers $= 8.9 - 4.45$
$= 4.45$

More consistent → nFries $= 5.65 - 3.4$
$= 2.25$

Fastest?
Compare Medians! (Q_2)

nFries is faster! (Q_2 is less)

Standard Deviation (σ):
mathematician's choice for measuring the spread of data; the measure of dispersion found by taking the square root of the variance

Standard Deviation (Ungrouped Data)

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Example:
Felix and Melanie are laying down patio stones. They record how many stones they put in place each hour.

	1st	2nd	3rd	4th	5th	6th
Felix	34	41	40	38	38	45
Melanie	51	28	36	44	41	46

(a) Which worker gets more done during the day?

Felix placed 236 stones
Melanie placed 246 stones

∴ Melanie got more work done

(b) Which worker is more consistent?

Felix ($\bar{x}$ = 39.3)			Melanie ($\bar{x}$ = 41)		
Stones Placed	$(x-\bar{x})$	$(x-\bar{x})^2$	Stones Placed	$(x-\bar{x})$	$(x-\bar{x})^2$
34	−5.3	28.09	51	10	100
41	1.7	2.89	28	−13	169
40	0.7	0.49	36	−5	25
38	−1.3	1.69	44	3	9
38	−1.3	1.69	41	0	0
45	5.7	32.49	46	5	25
$\Sigma(x-x)^2$		67.34	$\Sigma(x-x)^2$		328

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$= \sqrt{\frac{67.34}{6}}$$

$$\doteq 3.35$$

$$\sigma = \sqrt{\frac{328}{6}}$$

$$\doteq 7.39$$

∵ Felix's σ is lower (points are closer together)

∴ Felix is more consistent

Standard Deviation (Grouped Data)

$$\sigma = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{n}}$$

Example

A railway line gives out small bags of peanuts to its travellers, and each bag does not always contain the same number of peanuts. The following table represents a sample of 31 bags showing the number of peanuts per bag.

Number of Peanuts	28	29	30	31	32	33
Frequency	2	5	10	9	4	1

(a) Calculate the mean number of peanuts per bag.

$$\bar{x} = \frac{28\cdot 2 + 29\cdot 5 + 30\cdot 10 + 31\cdot 9 + 32\cdot 4 + 33\cdot 1}{2+5+10+9+4+1}$$

$$\doteq 30.4$$

(b) Calculate the standard deviation for this sample.

$$\sigma = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{2(28-30.4)^2 + 5(29-30.4)^2 + 10(30-30.4)^2 + 9(31-30.4)^2 + 4(32-30.4)^2 + 1(33-30.4)^2}{31}}$$

$$\doteq 1.2$$

The formula we use for standard deviation depends on whether the data is being considered a population of its own, or the data is a sample representing a larger population.

- If the data is being considered a population on its own, we divide by the number of data points, N.
- If the data is a sample from a larger population, we divide by one fewer than the number of data points in the sample, $n-1$.

The reason we divide by *n-1* instead of *n* when calculating the **sample standard deviation** relates to something called **Bessel's correction**. This correction compensates for the fact that a sample is typically smaller than the entire population, and thus provides a slightly biased estimate of the population's variability.

Population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Sample standard deviation:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

The steps in each formula are all the same except for one—we divide by one less than the number of data points when dealing with sample data.

Variance (σ^2):
measure of dispersion that is found by averaging the squares of deviation of each piece of data

Practice: #1, 3bc, 4, 6, 7, 9, 12