Measures of Variability

Measures of central tendency (averages) are of little use by themselves. They become useful when discussed in association with measures of variability. These summarise how much the data varies and include:

* the range;
* the variance;
* the standard deviation; and
* the mean absolute deviation.

Measures of Variability

The range of a data set is simply:

the largest value - the smallest value

Range

Example: Suppose two machines produce mobile phone covers which are on average 10cm long. 11 covers are selected from each machine.

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14
Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14

Which machine is better?

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14
Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14

Try calculating mean and range ...

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14
Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14

Try calculating mean and range ...

	Mean	Range
Machine A	10	8
Machine B	10	8

Doesn't really help - lets try variance

How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

2 5 10 11 12

Mean is ?

How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

2 5 10 11 12

Mean is 8

How can we work variance out?

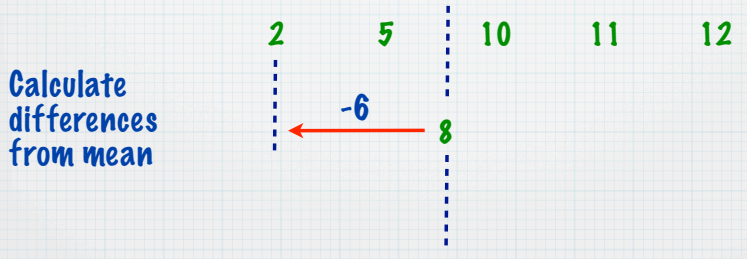
Lets try working out how far items are spread out from the mean (using a smaller example)

2 5 10 11 12

Mean is 8

How can we work variance out?

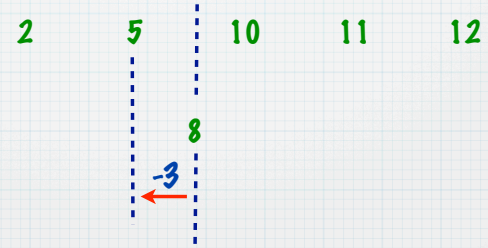
Lets try working out how far items are spread out from the mean (using a smaller example)



How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

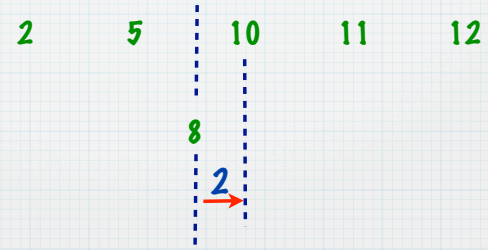
Calculate differences from mean



How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

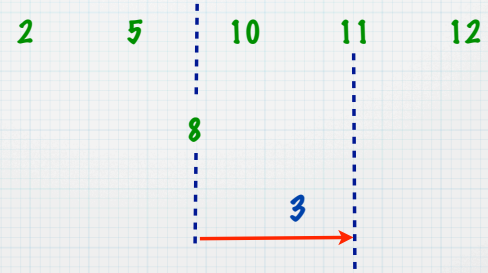
Calculate differences from mean



How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

Calculate differences from mean



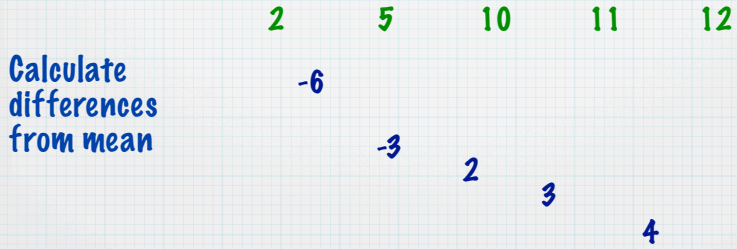
How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

Calculate differences from mean

How can we work variance out?

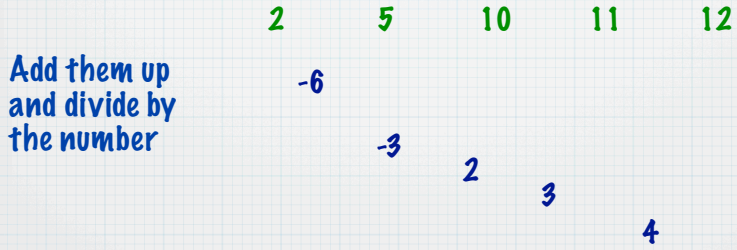
Lets try working out how far items are spread out from the mean (using a smaller example)



Calculate differences from mean

How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)



Add them up and divide by the number

How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

2	5	10	11	12
-6	-3	2	3	4

Add them up and divide by the number

How can we work variance out?

Lets try working out how far items are spread out from the mean (using a smaller example)

2	5	10	11	12
-6 + -3 + 2 + 3 + 4 =				0

Add them up and divide by the number of items

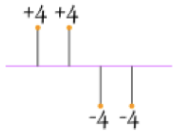
How can we work variance out?

Problem with this example? Sums to zero?

Try different examples - ages on each table

Trying out a few different ways to calculate

Could use absolute values

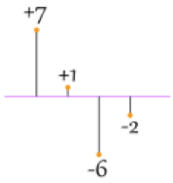


$$\frac{|4| + |4| + |-4| + |-4|}{4} = \frac{4 + 4 + 4 + 4}{4} = 4$$

Seems to work..

Trying out a few different ways to calculate

Could use absolute values



$$\frac{|7| + |1| + |-6| + |-2|}{4} = \frac{7 + 1 + 6 + 2}{4} = 4$$

Problem - values are spread out more but give same value

How can we work variance out?

Solution: square the difference before adding

2	5	10	11	12
-6	-3	2	3	4

Square difference

How can we work variance out?

Problem: sum of difference from mean is ALWAYS zero!

Solution: square the difference before adding

	2	5	10	11	12
	-6	-3	2	3	4
Square difference	36	9	4	9	16

How can we work variance out?

Problem: sum of difference from mean is ALWAYS zero!

Solution: square the difference before adding

	2	5	10	11	12
	-6	-3	2	3	4
	36	9	4	9	16
Add them up and divide by no of items	36 + 9 + 4 + 9 + 16 = 74 / 5				

How can we work variance out?

Problem: sum of difference from mean is ALWAYS zero!

Solution: square the difference before adding

	2	5	10	11	12
	-6	-3	2	3	4
	36	9	4	9	16
Add them up and divide by no of items	36 + 9 + 4 + 9 + 16 = 74 / 5				
	Variance = 14.8				

So - to calculate variance

Calculate mean

Find differences to mean

Square differences

Add them up and divide by the no of items

Try it on these two ...

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14
Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14

Interpreting the variance

Used in mathematics in many ways

Very useful when calculated to compare two sets

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14	4.36
Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14	10.18

So - machine A is better - less spread in results

The variance of a data set measures how much the data is spread around the mean. To work it out..

Calculate mean

Find differences to mean

Square differences

Add them up and divide by the no of items

The variance of a data set measures how much the data is spread around the mean. To work it out..

- Calculate mean
- Find differences to mean
- Square differences
- Add them up and divide by the no of items

Alternatively

The variance is the:

mean of the squared differences from the mean

Variance

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Variance

x_i

Variance

x_i

Calculate mean

Variance

$x_i - \mu$

Find differences to the mean

Variance

$(x_i - \mu)^2$

Square differences

Variance

$$\sum_{i=1}^{N} (x_i - \mu)^2$$

Add them up...

Variance

$$\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

... divide by the no of items

Variance

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Variance often represented by σ^2

Measures of Variability

Some useful facts about the variance.

Useful Fact 1:

The minimum value of the variance is zero. This implies no variance which itself implies that all the values are identical.

Useful Fact 2:

If we have two data sets whose values are of a similar magnitude, then the one with the larger variance, has the greater variability.

Measures of Variability

What is the unit of measure of variability?

If data items are cm cm^2

kg kg^2

£ $£^2$

So the square root of variance is often used

The standard deviation

Measures of Variability

The standard deviation of a data set is simply the

square root of the variance

Measures of Variability

The standard deviation of a data set is simply the

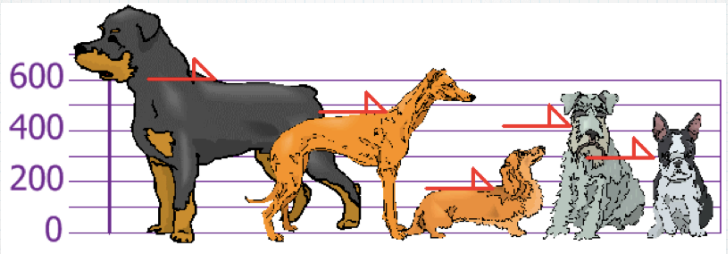
square root of the variance

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

Standard Deviation

Example: Dogs

Dog heights: 600, 470, 170, 430, 300 mm



Which dogs are too big / small ?

Standard Deviation

Example: Dogs

Dog heights: 600, 470, 170, 430, 300 mm

Calculate variance and standard deviation for these

Standard Deviation

Example: Dogs

Dog heights: 600, 470, 170, 430, 300 mm

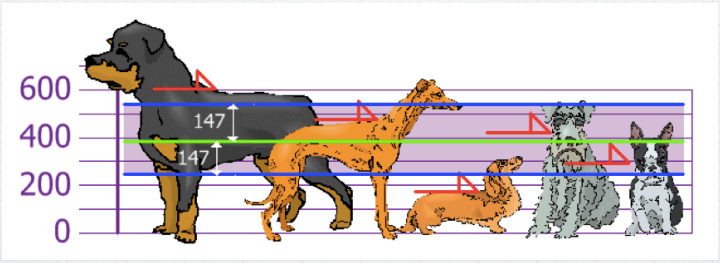
Variance 21704

SD 147.32mm

Standard Deviation

Lets say

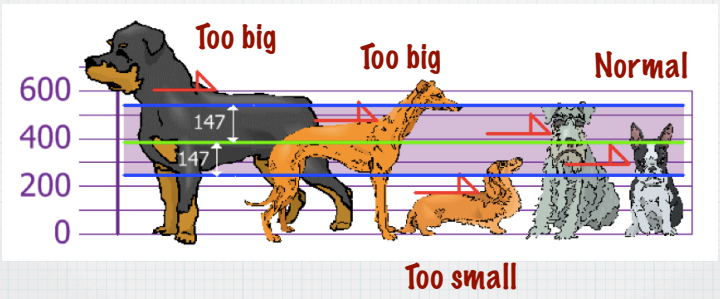
'Normal' dogs are within 1 standard deviation of the mean



Standard Deviation

Lets say

'Normal' dogs are within 1 standard deviation of the mean



Standard Deviation

We will frequently comment on items being within a certain number of standard deviations from the mean

We may set requirements that only particular numbers of SDs from the means are allowable