Standard Deviation (AS Ma)

Starter

1. (Review of last lesson) Four persons are chosen at random from a group of 10 persons consisting of 4 men and 6 women. Three of the women are sisters. Calculate the probabilities that the 4 persons chosen will:
   (a) consist of 4 women
   (b) consist of 2 women and 2 men
   (c) include the 3 sisters.

Notes

Standard deviation will be covered in greater depth in the maths course but without a general understanding of it, the next part of the further maths will be difficult.

Measures of central tendency vs. measures of spread (or dispersion)

Measures of central tendency are statistical values which represent the centre of the set of data. For example, mean, median and mode.

Measures of dispersion describe the spread of data around the central value. Such measures from GCSE, i.e. range and interquartile range, are not sophisticated enough so statisticians needed to devise a different measure of dispersion.

E.g. 1 Consider the data values 2, 3, 7, 4 and 9.
   (a) Calculate the mean for this data set.
   (b) Write down the deviation of each data value from the mean. If the value is below the mean, the deviation would be negative; if above the mean, positive.
   (c) Find the sum of the deviations. Comment on your answer

To overcome this problem, statisticians ***square each deviation*** and then sum them.

In our case: $(-3)^2 + (-2)^2 + 2^2 + (-1)^2 + 4^2 = 34$

This measure of deviation comes from 5 data values. If 10 data values returned the same value, they would not have the same spread — in fact, they would be much closer to the mean and the summary statistic must reflect that.

Therefore, ***divide by the number of values***.

In our case, $\frac{34}{5} = 6.8$ — this value is called the ***variance*** of the data set.

The problem is that the variance does not have the same units as the mean. Since the deviations are squared, the units of variance are the square of those of the mean. For example, if the mean is in metres, the variance would be in metres^2.

To solve this problem, simply square root the variance in order to get the ***standard deviation***, which is considered the ***most important measure of dispersion***. Standard deviation is given the symbol, σ (lower case "sigma") and variance is denoted σ^2.

N.B. The standard deviation is the ***positive*** square root of the variance.

For a data set $x_1, x_2, x_3, \ldots, x_n$ whose mean is μ:

Standard deviation: $$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

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

Standard deviation: ***square*** the deviations from the mean, ***divide*** by the number of values, ***square root*** the result.

Calculating standard deviation

While $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$ is the formula, there is a quicker way to calculate standard deviation.

Remember $\mu = \frac{\sum x_i}{n}$ (i.e. the sum of the values divided by the number of values)

There is no need to copy this.

Square the brackets: $$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n} = \frac{\sum (x_i^2 - 2x_i\mu + \mu^2)}{n}$$

Separating the terms: $$\sigma^2 = \frac{\sum x_i^2}{n} - \frac{\sum 2x_i\mu}{n} + \frac{\sum \mu^2}{n}$$

Since μ is a constant: $$\sigma^2 = \frac{\sum x_i^2}{n} - 2\mu\frac{\sum x_i}{n} + \frac{n\mu^2}{n}$$

N.B. $$\sum_{i=1}^{n} \mu^2 = \mu^2 \sum_{i=1}^{n} 1 = \mu^2(1 + 1 + 1 + \ldots + 1) = \mu^2 \times n$$

But $\frac{\sum x_i}{n} = \mu$: $$\sigma^2 = \frac{\sum x_i^2}{n} - 2\mu^2 + \mu^2$$

Please start copying again.

The final formula: $$\sigma^2 = \frac{\sum x_i^2}{n} - \mu^2$$

For standard deviation: $$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \mu^2}$$

E.g. 2 Find the standard deviation for these datasets.

(a) $24, 19, 17$

(b) $11, 14, 3, 8$

(c) $73, 65, 89, 68, 59, 82$

Working: (a) Mean $= 20$

Variance, $\sigma^2 = \frac{24^2 + 19^2 + 17^2}{3} - 20^2 = \frac{26}{3}$

Standard deviation $= \frac{\sqrt{78}}{3}$ (3 s.f.)

N.B. The AS maths course will also teach calculating standard deviation from a table of values.

Calculating the standard deviation using a Classwiz
Menu >> 6: Statistics >> 1: 1-Variable >> (Enter data) >> Press AC >> Press OPTN >> 2: 1-Variable Calc >> (Data displayed)

$\bar{x} \equiv$ mean
$\sigma^2 x \equiv$ variance
$\sigma x \equiv$ standard deviation

N.B. Always check your data after entering to make sure all values are correct.

Video: **Calculating standard deviation on the Classwiz**

E.g. 3 Use your calculator to find the mean and standard deviation for the following data:
(a) $165, 183, 167, 174, 192, 186, 195, 171$

(b) $6958, 7492, 5638, 4529, 3945, 4957, 4639$

Working: (a) Mean $= 179.125$
Standard deviation $= 10.7$ (3 s.f.)

Video: Standard deviation

Standard deviation (discrete) EQ

Solutions to Starter and E.g.s

Exercise
No exercise

Summary
The standard deviation is the ***positive*** square root of the variance.
Original formulae:

Standard deviation: $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$

Variance: $\sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$

Formulae for calculating:

Standard deviation (same units as mean): $\sigma = \sqrt{\dfrac{\sum x_i^2}{n} - \mu^2}$

Variance $\sigma^2 = \dfrac{\sum x_i^2}{n} - \mu^2$