

S4 Measures of Spread

Measuring Spread or Dispersion in Data

Consider the two sets of values below:

Set A: 4, 4, 5, 5, 5, 6, 6.

Set B: 1, 3, 4, 5, 6, 7, 9.

Both groups A and B have mean = median = 5 but the data sets are quite different.

The values in Set A are less spread out than those in Set B.

Range

To compare data sets it is also useful to look at the measure of spread.

The most basic measure of spread is the range, the distance from the smallest to the largest value.

Set A

$$\begin{aligned}\text{Range} &= \text{highest value} - \text{lowest value} \\ &= 6 - 4 \\ &= 2.\end{aligned}$$

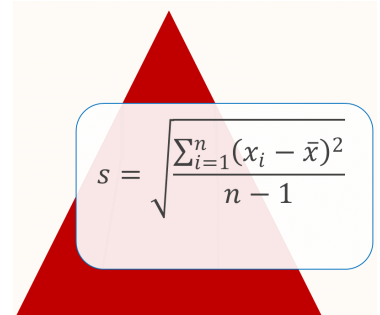
Set B

$$\begin{aligned}\text{Range} &= \text{highest value} - \text{lowest value} \\ &= 9 - 1 \\ &= 8.\end{aligned}$$

We can see that Set B has greater spread than Set A. But the problem with the range is that it uses only two of the values in the data set. One of these may be an odd or unusual value called an outlier.

Consider the two sets of values below:

Set Y: 1, 1, 2, 2, 2, 2, 2, 100.



Set Z: 1, 18, 23, 41, 59, 63, 87, 100.

The range for both is 99 because Set Y has one unusual value - the 100. Most of the data are ones or twos. The value of 100 is unusual and is an outlier.

Interquartile Range

The Interquartile Range (IQR) is the distance between the first quartile Q_1 and the third quartile Q_3 .

$$\text{IQR} = Q_3 - Q_1$$

The first and third quartiles are values that are $1/4$ and $3/4$ of the way through the ordered data. Q_1 is the median of the lower half of the data and Q_3 is the median of the upper half of the data.

(Note: there are other ways to find Q_1 and Q_3 so check with your teacher.)

Set Y: 1 $\underbrace{1 \downarrow 2}_{Q_1}$ 2 \downarrow 2 $\underbrace{2 \downarrow 2}_{Q_3}$ 100

The arrows indicate the location of the quartiles and

$$\begin{aligned} Q_1 &= \frac{1+2}{2} \\ &= 1.5 \\ Q_3 &= \frac{2+2}{2} \\ &= 2. \end{aligned}$$

Hence the interquartile range is $\text{IQR} = Q_3 - Q_1 = 2 - 1.5 = 0.5$.

Now consider the other set of data:

Set Z: 1 $\underbrace{18 \downarrow 23}_{Q_1}$ 41 \downarrow 59 $\underbrace{63 \downarrow 87}_{Q_3}$ 100

We have

$$\begin{aligned} Q_1 &= \frac{18+23}{2} \\ &= 20.5 \\ Q_3 &= \frac{63+87}{2} \\ &= 75. \end{aligned}$$

Hence the interquartile range is

$$\text{IQR} = Q_3 - Q_1 = 75 - 20.5 = 54.5$$

So for the data sets Y and Z the mean together with the IQR are better summaries of the data sets than the range.

Standard Deviation

A measure of dispersion or spread in a data set that takes into account all of the data is the standard deviation. It gives an indication of the typical or average distance of each score from the mean for the data. The standard deviation can be calculated using the formula

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

but it is much more convenient to use your calculator or the computer.

Some statistical tests make use of the variance which is the square of the standard deviation:

$$\text{Variance} = s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}.$$

Set A: 4 4 5 5 5 6 6

$$\bar{x} = 5$$

$$s = 0.82$$

Set B: 1 3 4 5 6 7 9

$$\bar{x} = 5$$

$$s = 2.65$$

We can interpret the standard deviation: the scores in set A are typically 0.82 away from the mean, but the scores in set B are typically 2.65 away from the mean. Though set A and B have the same centre, those in set B are clearly more dispersed or have greater spread than set A.

Exercises

1. Given the following scores: 12, 12, 13, 14, 14, 15, 15, 15, 16, find the standard deviation.

Answer: 1.414

2. A class of 22 students gained the following scores, out of 10, on a test : 5, 7, 8, 7, 6, 5, 6, 4, 7, 4, 8, 3, 7, 9, 4, 9, 7, 3, 6, 8, 7, 5. Find the (a) range (b) IQR (c) standard deviation.

Answers: (a) 6 (b) 2 (c) 1.807

3. Pistol Pete is the star full-forward for the local football team. Last season he played 20 games and kicked the following number of goals in each game: 5, 6, 6, 5, 7, 4, 3, 1, 3, 8, 7, 8, 6, 0, 5, 2, 7, 6, 5, 6.

(a) Find the mean and the standard deviation for the number of goals that Pete kicked per game. Hint: google an online standard deviation calculator or use statistics mode on your calculator.

(b) This season the mean number of goals Pete kicks per game is 5, with a standard deviation of 2.7. In which year was his performance more consistent?

Answers: (a) $\bar{x} = 5$, $s = 2.22$ (b) Last year: smaller standard deviation means less variation.