

## Lesson Plan on Standard Deviation

Time: 60 minutes

Class: S6

Objectives:

1. To introduce the necessity of standard deviation.
2. To understand the concept of standard deviation.
3. To apply the concept in daily life.

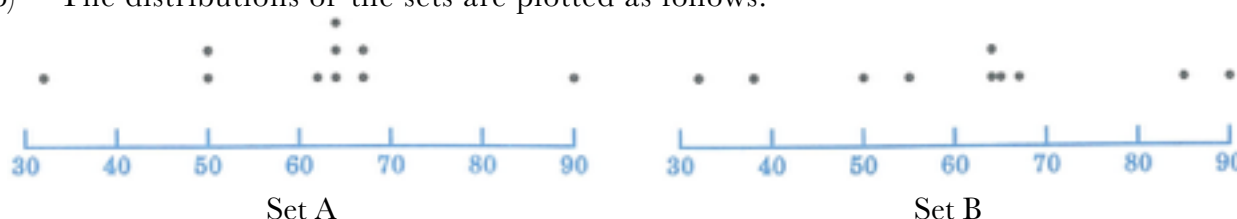
Time	Activity
5 min	T briefs Ss on the objectives and the planned progress of the lesson.
10 min	T gives an example of $\geq 2$ sets of data which have the same range and inter-quartile range but clearly different distributions. T guides the students to discuss the pros and cons of using inter-quartile range as a measure of dispersion.
10 min	T introduces the concept of standard deviation and how standard deviation will be defined in different situations. T guides the students to find the value of standard deviation by using a calculator.
15 min	Ss are asked to identify the suitable definition of standard deviation in different situations, and then find the corresponding standard deviation.
15 min	Ss are given information of samples collected with sampling error stated. Ss are asked to discuss its importance to the poll.  <a href="http://hkupop.hku.hk/english/report/mpCEnOCCw6/index.html">http://hkupop.hku.hk/english/report/mpCEnOCCw6/index.html</a>
5 min	T sums up the lesson

## Standard Deviation

1. Given the following sets of data:

<b>Set A</b>	32	50	50	62	64	64	64	67	67	90
<b>Set B</b>	32	38	50	55	64	64	65	67	85	90

- (a) (i) Find the range and inter-quartile range for each set of data.  
(ii) Do they have the same dispersion?  
(b) The distributions of the sets are plotted as follows:



Determine whether range or inter-quartile range indicates the dispersion appropriately.

Since both the range and the inter-quartile range involve only some but not all of the data, so they cannot precisely tell the spread of the entire set, and the effect of each datum upon the dispersion cannot be seen. Neither range nor inter-quartile range is a reliable measure of dispersion.

To overcome this limitation, every datum should be taken into account. To be precise, the distance of each datum from the centre should be calculated, and the average distance can be found. The greater the average distance of all data from the centre is, the more dispersed a set of data is.

As only mean takes all data into account, and to avoid negative numbers from involving in the calculation, every difference between a datum and the mean should be squared, taken average, and then taken the square root. This measure of dispersion is called standard deviation.

For a set of ungrouped data  $x_1, x_2, x_3, \dots, x_n$ , the formula is

$$\text{standard deviation } \sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n}},$$

where  $\mu$  is the mean. (In other words, we need to always find the mean before the standard deviation.)

Note:

- (a)  $\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n}$  is called the variance  $\sigma^2$ , which is also a useful measure.  
(b)  $\sigma = 0$  if and only if  $x_1 = \mu, x_2 = \mu, \dots, x_n = \mu$ , i.e.  $x_1 = x_2 = \dots = x_n = \mu$ , all data are equal.

2. Find the mean and standard deviation of the following sets of data.

- (a) 1, 2, 3  
(b) 1, 1, 2, 3, 3, 3, 4  
(c)

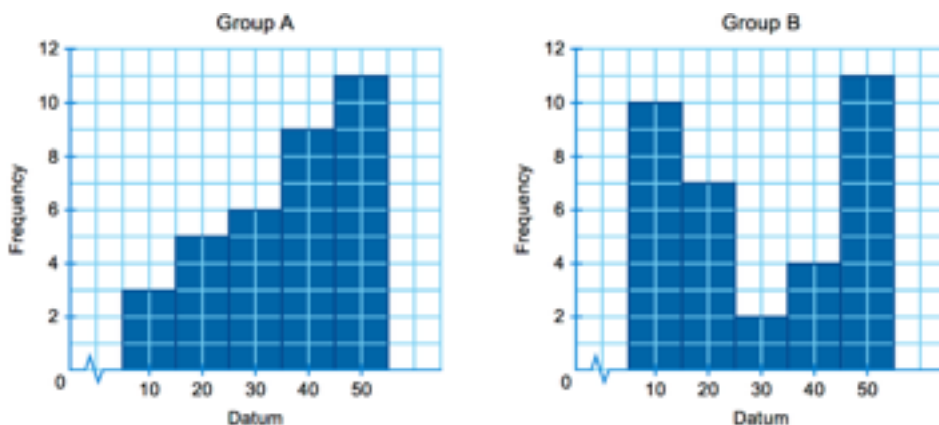
<b>Datum</b>	2	3	4	5	6
<b>Frequency</b>	4	6	7	3	1

3. The following table shows the speeds of vehicles recorded by the police.

Speed (km/h)	51 – 60	61 – 70	71 – 80	81 – 90	91 – 100	101 – 110
Class mark (km/h)						
Frequency	32	72	33	10	2	1

- Complete the above table
- Find the mean and standard deviation of speed.

4. The histogram below show two groups of data.



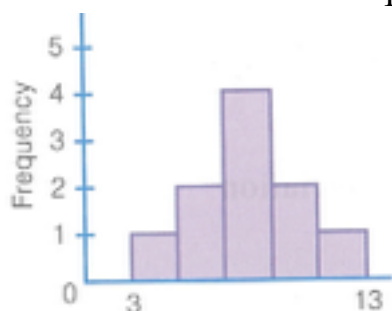
- Find the mean and the standard deviation of the data in each group.
- Which group has larger data in general? Explain your answer.
- Whose group has a less dispersed data? Explain your answer.

### Range vs Inter-quartile range vs Standard deviation

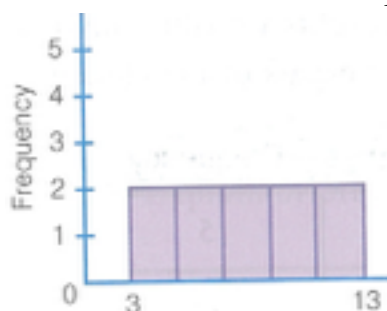
Measure	Advantage	Disadvantage
<b>Range</b>	Easy to calculate.	Affected by extreme values.
	Can be obtained from some graphs.	No further statistical applications.
<b>Inter-quartile range</b>	Easy to calculate.	Not showing the dispersion of the whole set of data.
	Can be obtained from some graphs.	No further statistical applications.
	Not affected by extreme values.	
<b>Standard deviation</b>	Not affected by extreme values.	Difficult to compute without using calculators with statistical function.
	Dispersion reflected accurately.	
	Used for further statistical applications.	

5. The figures show the histograms of three frequency distributions. Arrange them in descending order of magnitude of their standard deviations.

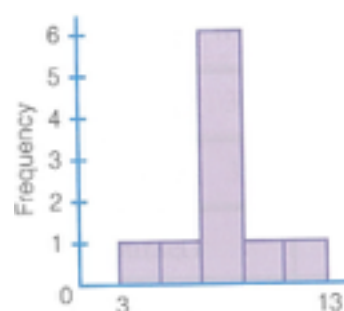
I.



II.



III.



- A. I, II, III  
 C. II, I, III
- B. I, III, II  
 D. III, I, II
6. If the range of 2, 3,  $y$ , 6, 4, 4, 5, 7 is 10, and  $y > 0$ , find the standard deviation.  
 A. 2.37  
 B. 2.51  
 C. 2.61  
 D. 2.91
7. The standard deviation of the five numbers  $10a + 1$ ,  $10a + 3$ ,  $10a + 5$ ,  $10a + 7$  and  $10a + 9$  is  
 A. 8  
 B.  $\frac{12}{5}$   
 C.  $\sqrt{10}$   
 D.  $2\sqrt{2}$
8. In the sixth round “Survey on CE Election and Occupy Central” sponsored by Ming Pao, the sampling error was less than 1.6%.  
 (a) Explain its significance.  
 (b) Why the sampling error was always the same throughout all the surveys conducted?