Statistics & Probability • Lesson 7

Measures of centre & spread

- Data can be summarised or described using measures of centre and measures of spread.

- Measures of centre include the mean, median and mode. These statistics describe a whole set of data using a single value that represents the centre or middle of a data set.

- The mean of a set of data is the average of the numbers. It is given by:

$$\bar{x} = \frac{sum\ of\ data\ values}{number\ of\ data\ values}$$

$\bar{x}$ (x bar) is the symbol used to represent mean.

- The median is the middle value of the data when the values are sorted in order from lowest to highest. If there is an even number of values in the data set, there will be two middle values and the median will be the average of these two numbers.

- The mode is the most common value and is the value that occurs most frequently. Multiple values can be the mode if they all share the highest frequency.

- Measures of spread include the range and interquartile range. They are used to describe the amount of spread in a set of data.

- The range of a set of data is the difference between the lowest and highest values.
$Range = highest\ number - lowest\ number$

- The range, mean and median can only be calculated for numerical data, but the mode can be found for either numerical or categorical data.

Let's try to answer the following question together before looking at the worked example.

Consider the ages (in years) of eight people who are surveyed in a coffee shop:

17, 28, 15, 47, 28, 65, 18, 54

a) Find the range of the values.

b) FInd the mean of this set of data.

c) Find the median of this set of data.

d) Find the mode of this set of data.

WORKED EXAMPLE 1

Consider the ages (in years) of eight people who are surveyed in a coffee shop:

17, 28, 15, 47, 28, 65, 18, 54

a) Find the range of the values.

b) FInd the mean of this set of data.

c) Find the median of this set of data.

d) Find the mode of this set of data.

- Firstly, it is always good practice to rearrange the values so that they are in ascending order.

  15, 17, 18, 28, 28, 47, 54, 65

- The range is the difference between the highest and lowest number.

  Range = highest number - lowest number

  = 65 - 15

  = 50

- To find the mean, we can use $\bar{x} = \frac{sum\ of\ data\ values}{number\ of\ data\ values}$.

  $\bar{x} = \frac{15+17+18+28+28+47+54+65}{8} = \frac{272}{8} = 34$

- The mode is the most common value. Here the most common value is 28 which occurs twice. Hence, mode = 12.

- The median is the middle value of the data set in ascending order.

  15, 17, 18, 28, 28, 47, 54, 65

  Now, there are two middle values (28 and 28), so the median is $\frac{28+28}{2} = 28$.

- The interquartile range (IQR) is the difference between the upper and lower quartile of a data set. The upper quartile is the median of the upper half of a data set while the lower quartile is the median of the lower half.

$IQR = upper\ quartile\ -\ lower\ quartile$

- As a summary statistic, the IQR has an advantage in that it is less influenced by extreme values (outliers).

- The general steps to find the IQR are:

    1) List data in ascending order.

    2) If there is an odd number of values, remove the middle value (which is the median).

    3) Split the data into two equal size groups.

    4) Find the lower quartile which is the median of the lower half.

    5) Find the upper quartile which is the median of the upper half.

    6) Subtract to determine IQR.
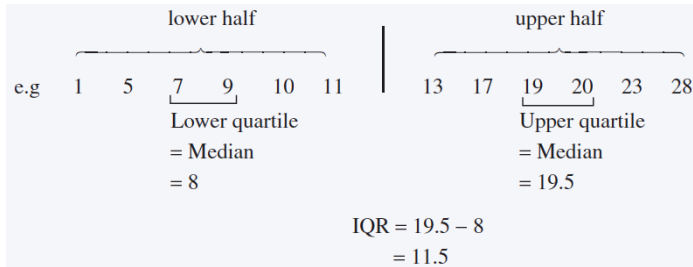
Let's try to answer the following question together before looking at the worked example.

Find the interquartile range of this set of data.

2, 7, 11, 8, 4, 8, 10, 2, 9, 5, 8

WORKED EXAMPLE 2

Find the interquartile range of this set of data.

2, 7, 11, 8, 4, 8, 10, 2, 9, 5, 8

- Firstly, sort the values so that they are in ascending order.

  2, 2, 4, 5, 7, 8, 8, 8, 9, 10, 11

- As there are an odd number of values, remove the middle value and divide so that there are two equal halves remaining.

  2, 2, 4, 5, 7, ~~8~~, 8, 8, 9, 10, 11

  Lower half   Upper half

  2, 2, 4, 5, 7   8, 8, 9, 10, 11

- Now, find the lower and upper quartile by determining the median of the lower and upper half respectively.

  Lower quartile = 4

  Upper quartile = 9

- Thus,

  IQR = upper quartile - lower quartile

  = 9 - 4

  = 6

Statistics & Probability • Revision 7

Level 1:

Q#	QUESTION	RESPONSE

1	State the range of the following sets of numbers.
	    a) 6, 7, 12, 8, 1
	    b) 0, 7, 3, 5, 1	

2	For the set of numbers 2, 5, 5, 7, 11, find the:
	    a) Total of the numbers when added
	    b) Mean
	    c) Median
	    d) Mode	

3	For each of the following sets of data, calculate the Range, Mean, Median and Mode for the following:
	    a) 2, 2, 10, 8, 13
	    b) 1, 22, 10, 20, 33, 10
	    c) 97, 31, 18, 54, 18, 63, 6	

4	Consider the set of numbers 2, 5, 8, 10, 11, 11, 14.
	    a) State the median of the lower half (2, 5, 8).
	    b) State the median of the upper half (11, 11, 14).
	    c) Hence state the interquartile range.	

5	Consider the set 9, 3, 10, 5, 5, 6, 4, 11.
	    a) Write this list in ascending order.
	    b) Calculate the lower quartile.
	    c) Calculate the upper quartile.
	    d) Calculate the IQR.	

Level 2:

Q#	QUESTION	RESPONSE
6	The number of aces that a tennis player serves per match is recorded over eight matches. 	

Match	1	2	3	4	5	6	7	8
# of Aces	11	19	11	16	18	22	23	13

a) What is the mean number of aces the player serves per match? Round your answer to 1 decimal place.
b) What is the median number of aces the player serves per match?
c) What is the range of this set of data?

Q#	QUESTION
7	James and Justin organise their maths test marks in a table.

Test	1	2	3	4	5	6	7	8	9	10
James	90	82	84	50	76	67	69	70	85	57
Justin	58	90	90	77	96	60	94	100	95	90

a) Which boy has the higher mean?
b) Which boy has the higher median?
c) Which boy has the smaller range?
d) Which boy do you think is better at tests? Explain why.

RESPONSE

Q#	QUESTION	RESPONSE
8	The set 3, 6, 10, 10 has one extra number added to it, and this causes the mean to be doubled. What is the number?	
9	A soccer goalkeeper recorded the number of saves he makes per game during a season in a table. | Number of saves	0	1	2	3	4	5
---	---	---	---	---	---	---
Number of games	4	3	0	1	2	2 | a) How many games did he play that season? b) What is the mean number of saves this goalkeeper made per game? c) What is the most common number of saves that the keeper had to make during a game?	
10	Give an example of a set of data with: a) A mean of 10 and a range of 2 b) A mean of 6, a median of 7 and a mode of 5	
11	Three whole numbers are chosen with a mean of 10 and a range of 2. What are the numbers?	
12	Explain why you can calculate the mode for numerical or categorical data but you can only calculate the mean, median and range from numerical data.	
13	Find the IQR of the following sets. a) 7, 9, 13, 16, 20, 28 b) 1.16, 2.97, 3.84, 3.94, 4.73, 6.14	

Q#	QUESTION	RESPONSE
14	Find the IQR of the following sets. a) 10, 11, 13, 22, 27, 30, 30 b) -19, -17, -6, 0, 3, 3, 18, 23, 26	
15	Find the IQR of the following sets. a) 6, 4, 6, 5, 14, 8, 10, 18, 16, 6 b) -4, -9, 17, 7, -8, -4, -16, 4, 2, 5	
16	Julia and Nadia compare the number of runs they score in cricket over a number of weeks. Julia: 17, 15, 8, 5, 12, 19, 15, 16 Nadia: 39, 6, 22, 19, 10, 22, 28, 30 a) Calculate Julia's range. b) Calculate Nadia's range. c) Who has the greater range? d) Which cricketer is more consistent, on the basis of their ranges only?	

Q#	QUESTION
17	Over 20 weeks, Joe and Ben tally their spelling test results.

Score	0	1	2	3	4	5	6	7	8	9	10
Joe		||		|		ЖΗ	|	||	|||	|	ЖΗ
Ben	|				|		||||		ЖΗ	||	ЖΗ-||

a) Find the range for:
  i) Joe
  ii) Ben

b) Find the IQR for:
  i) Joe
  ii) Ben

c) On the basis of range only, which student is more consistent?

d) On the basis of the IQR only, which student is more consistent?

RESPONSE

Q#	QUESTION	RESPONSE
18	Consider the set of numbers 2, 3, 3, 4, 5, 5, 5, 7, 9, 10. a) Calculate the range. b) Calculate the IQR. c) If the number 10 changed to 100, calculate the new range. d) Calculate the new IQR. e) Explain why the IQR is a better measure of spread if there are outliers in a data set.	
19	Two unknown numbers are chosen with mean = 10 and range = 4. What is the product of the two unknown numbers?	
20	For a set of 3 numbers, what effect is there on the range if: a) each number is increased by 10? b) each number is doubled?	