Data and Statistics - Lesson 2

5 - Figure Summaries, Boxplots, Interquartile Range and Semi - Interquartile Range

LI

- Know the meaning of Quartile.
- Work out a 5-Figure Summary of a data set.
- Draw a Boxplot.
- Calculate Interquartile and Semi-Interquartile Ranges.

SC

- Mean and Median.

Jan 10-12:11

A Quartile is one of 3 numbers that split up a list of numbers (from lowest to highest) into 4 equal groups

Start with a list of numbers (aka data set) *in order from lowest to highest*

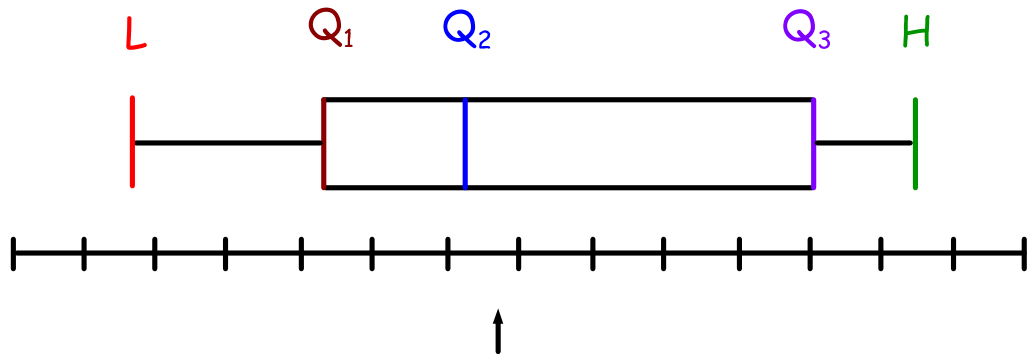
A 5-Figure Summary consists of 5 numbers :

- Smallest number in list : Lowest Value (L)
- Biggest number in list : Highest Value (H)
- Middle number of list : Median, aka Second Quartile (Q_2)
- Middle number of lower half of list : First Quartile (Q_1) aka Lower Quartile
- Middle number of upper half of list: Third Quartile (Q_3) aka Upper Quartile

The 5-figure Summary is then :

L, Q_1, Q_2, Q_3, H

A Boxplot is a diagram showing a 5 - figure summary :



(Even scale with numbers on it)

The Interquartile Range (IQR) is :

$$IQR = Q_3 - Q_1$$

The IQR measures the spread of data compared with the mean; it only deals with the middle 50 % of data.

The Semi - Interquartile Range (SIQR) is :

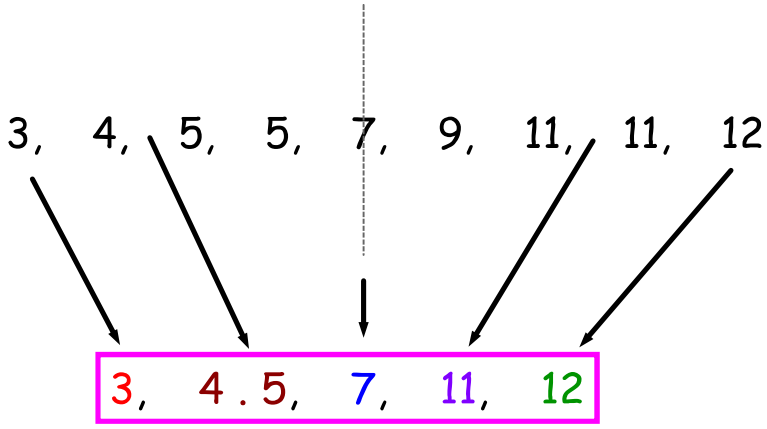
$$SIQR = \frac{Q_3 - Q_1}{2}$$

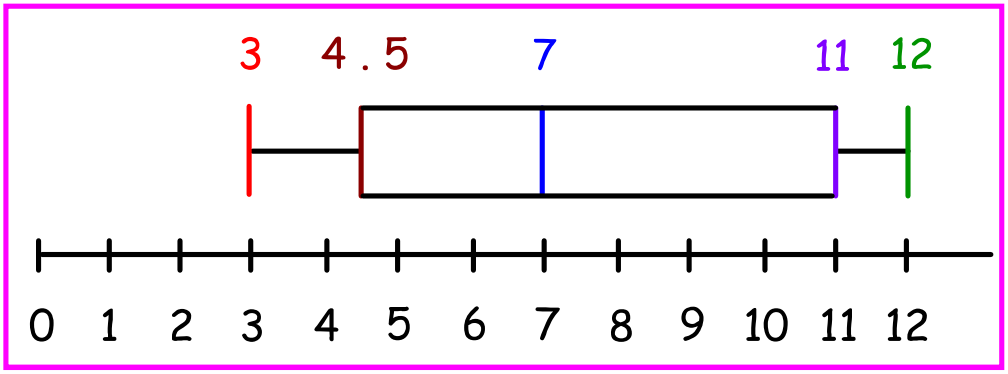
The SIQR measures the average spread from the centre of the data.

Example 1

Make a 5-figure summary, draw a boxplot and find the IQR for the following data set :

3, 4, 5, 5, 7, 9, 11, 11, 12





$IQR = Q_3 - Q_1$

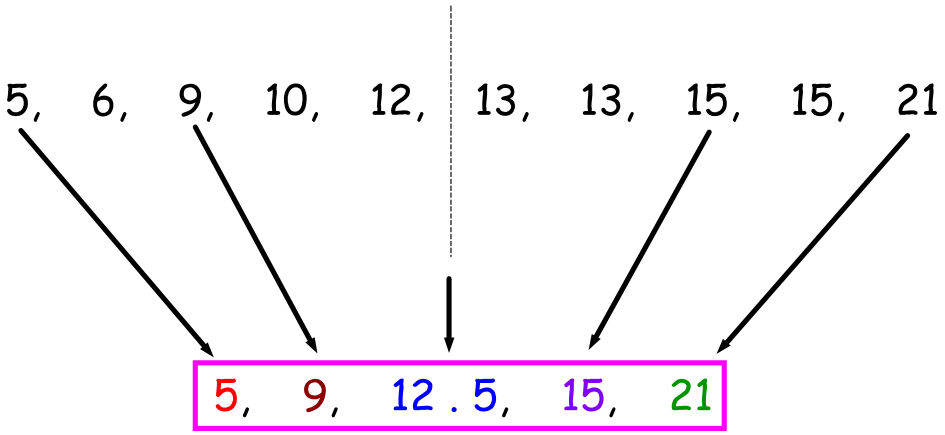
$IQR = 11 - 4.5$

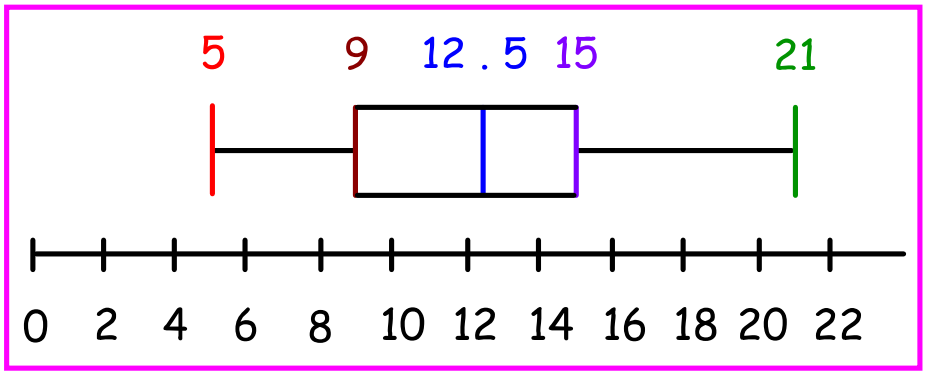
$IQR = 6.5$

Example 2

Make a 5-figure summary, draw a boxplot and find the SIQR for the following data set :

5, 6, 9, 10, 12, 13, 13, 15, 15, 21





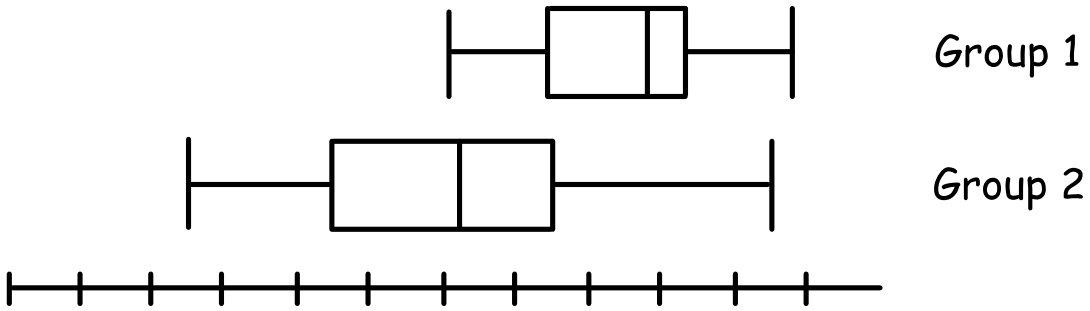
$$SIQR = \frac{Q_3 - Q_1}{2}$$

$$SIQR = \frac{15 - 9}{2}$$

$$SIQR = 3$$

Example 3

The heights of two different groups of people are shown :



Make two comments about the heights of these two groups.

Group 1 has, on average, the greater height (as the median is bigger than that for Group 2); the heights in Group 1 are more consistent (as the IQR is smaller than that for Group 2)

Questions

1 Find the median and quartiles of the following data sets:

a 1, 1, 3, 3, 3, 6, 7, 7, 8, 9, 9

b 23, 25, 29, 29, 32, 36, 41, 42, 42, 51, 52, 60

c 4.5, 3.2, 1.6, 8.9, 5.3, 2.7, 8.4, 7.2, 5.6

2 Biologists are studying the differences between hamsters and rats.

a The lengths (from nose to the tip of the tail) in cm of a sample of 10 Syrian hamsters are:

13 13 14 15 15 15 16 17 17 18

Calculate:

i the median **ii** the quartiles **iii** the interquartile range of the sample.

b The lengths (from nose to the tip of the tail) in cm of a sample of 10 brown rats are:

9 10 12 14 15 15 18 22 23 25

Calculate:

i the median **ii** the quartiles **iii** the interquartile range of the sample.

c Make two valid comparisons between the lengths of the hamsters and rats.

3 The price, in pence per litre, of diesel at 12 city garages is shown below:

136.9 145.4 139.2 148.2 138.1 142.5

137.5 143.6 138.2 140.8 139.8 146.7

a Calculate:

i the median **ii** the quartiles **iii** the interquartile range of these prices.

In 12 rural garages the petrol prices have a median of 146.3p and an interquartile range of 15.3p.

b How do the rural prices compare with the city prices?

Feb 4-12:58

4 Patients in Central Scotland waiting for hip replacements have to wait a number of days on a waiting list in order to see a consultant. The average number of days in 20 different hospitals is recorded in the stem and leaf diagram below.

Average number of days waiting to see a consultant prior to a hip replacement

0	2 5	
1	6	
2	0 2 2 6 6 7 7	Key: 2 | 6 = 26
3	1 2 4 4 8	
4	3 5	
5	1 2 8	

$n = 20$

a Find the median and quartiles.

b Calculate: **i** the range **ii** the interquartile range.

c Another hospital's data is included in the list. Its average waiting time is 76 days.

i Calculate the new range and interquartile range.

ii Which is most affected by the addition of one piece of data: the range or the interquartile range?

5 A popular city centre theatre has an audience capacity of 3039 people.

The audience figures for 12 performances of a musical were:

680 1427 2532 1793 2838 1982
717 1816 2997 2739 1407 1834

a Calculate:

i the median **ii** the quartiles **iii** the interquartile range of these figures.

b Draw a boxplot to represent this data.

During a Comedy Festival 12 different famous comedians were booked to perform at the same theatre on 12 separate performances. The audience figures were:

2782 2815 2976 3009 1897 2156
2419 2795 3030 2184 1982 2864

c Calculate:

i the median **ii** the quartiles **iii** the interquartile range of these figures.

d Draw a boxplot on the same scale as in **b** to represent this data.

e Based on these figures, which do you think is more popular, musicals or comedy? Give two reasons for your answer.

6 14 pupils sat a maths exam (marks out of 50). Their results were:

14 21 22 25 31 33 34 34 34 36 37 40 41 48

a Make a five-figure summary of the results.

b Draw a boxplot to illustrate the data.

Feb 25-21:39

Answers

1 **a** Q1 = 3, Q2 = 6, Q3 = 8

b Q1 = 29, Q2 = 38.5, Q3 = 46.5

c Q1 = 2 . 95, Q2 = 5.3, Q3 = 7 . 8

2 **a** **i** 15

ii 14, 17

iii 3

b **i** 15

ii 12, 22

iii 10

c Rat length is more inconsistent, average length is the same

3 **a** **i** 140.3

ii 138.15, 144.5

iii 6.35

b On average, city prices are cheaper; rural prices are less consistent.

4 **a** Q1 = 22, Q2 = 29, Q3 = 40.5

b **i** Range = 56

ii IQR = 18.5

c **i** Range = 74 , IQR = 22

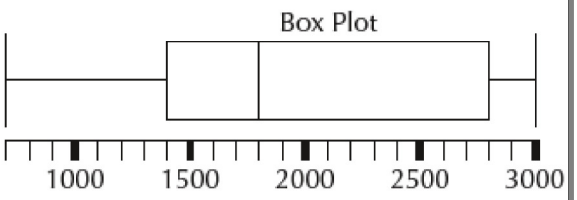
ii The range

5 **a** **i** 1825

ii 1417, 2635.5

iii 1218.5

b

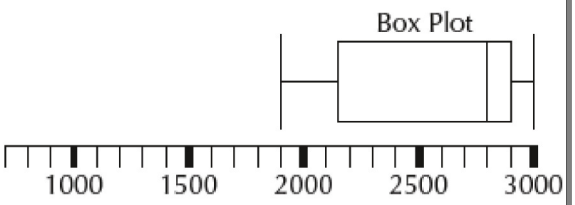


c **i** 2788.5

ii 2170, 2920

iii 750

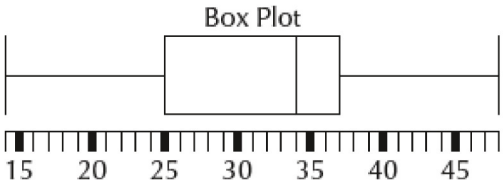
d



e Comedy is more popular. Less variation, higher average

6 **a** 14, 25, 34, 37, 48

b



Dec 18-14:20