Lesson 2

2.1 Box and Whisker Diagrams with Outliers

A Geography class of 30 pupils, when given a test, yields %age scores as follows;

57	74	63	58	57	12	37	45	28
42	51	48	35	64	55	08	59	46
54	50	56	63	60	37	47	65	67
58	53	36						

Using a computer spreadsheet (Excel, for example) the data is put into order.

08	12	28	35	36	37	37	42	45
46	47	48	50	51	53	54	55	56
57	57	58	58	59	60	63	63	64
65	67	74						

(i)   Which two, perhaps three, items of data are outliers ?

    ☞

(ii)   Determine the quartiles Q_1, Q_2 and Q_3
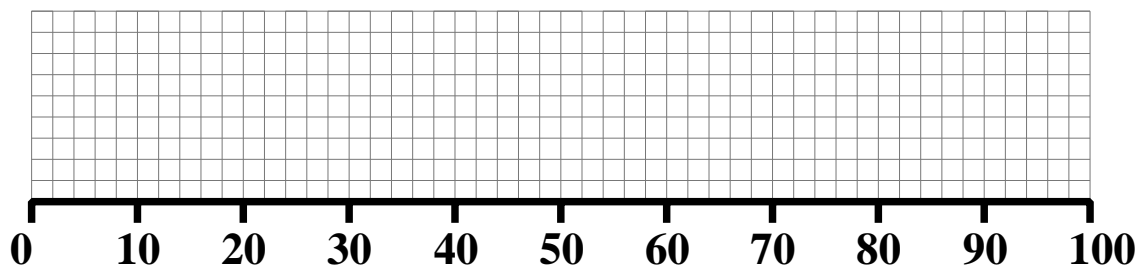
    ☞

(iii)   Calculate the IQR

    ☞

(iv)   An outlier is an item of data that is $1.5 \times$ IQR either above the Q_3 or below Q_1. Calculate these 'cut off' values.

☞

(v)   Hence produce a box plot of the data with any outliers clearly marked.

☞

(vi)   Is the skew positive, or negative or is the distribution symmetrical ?

☞

2.2 Exercise

Question 1

S1 Examination question from June 2005, Q4, (edited)
Aeroplanes fly from City *A* to City *B*. Over a long period of time the number of minutes delay in take-off from City *A* was recorded. the minimum delay was 5 minutes and the maximum delay was 63 minutes. A quarter of all delays were at most 12 minutes, half were at most 17 minutes and 75% were at most 28 minutes. Only one of the delays was longer than 45 minutes.

An outlier is an observation that falls either 1.5 × (interquartile range) above the upper quartile or 1.5 × (interquartile range) below the lower quartile.

(**a**)   Calculate the outlier 'cut-off' values.

[**2 marks**]

(**b**)   On graph paper draw a box plot to represent these data.

[**4 marks**]

(**c**)   Calculate

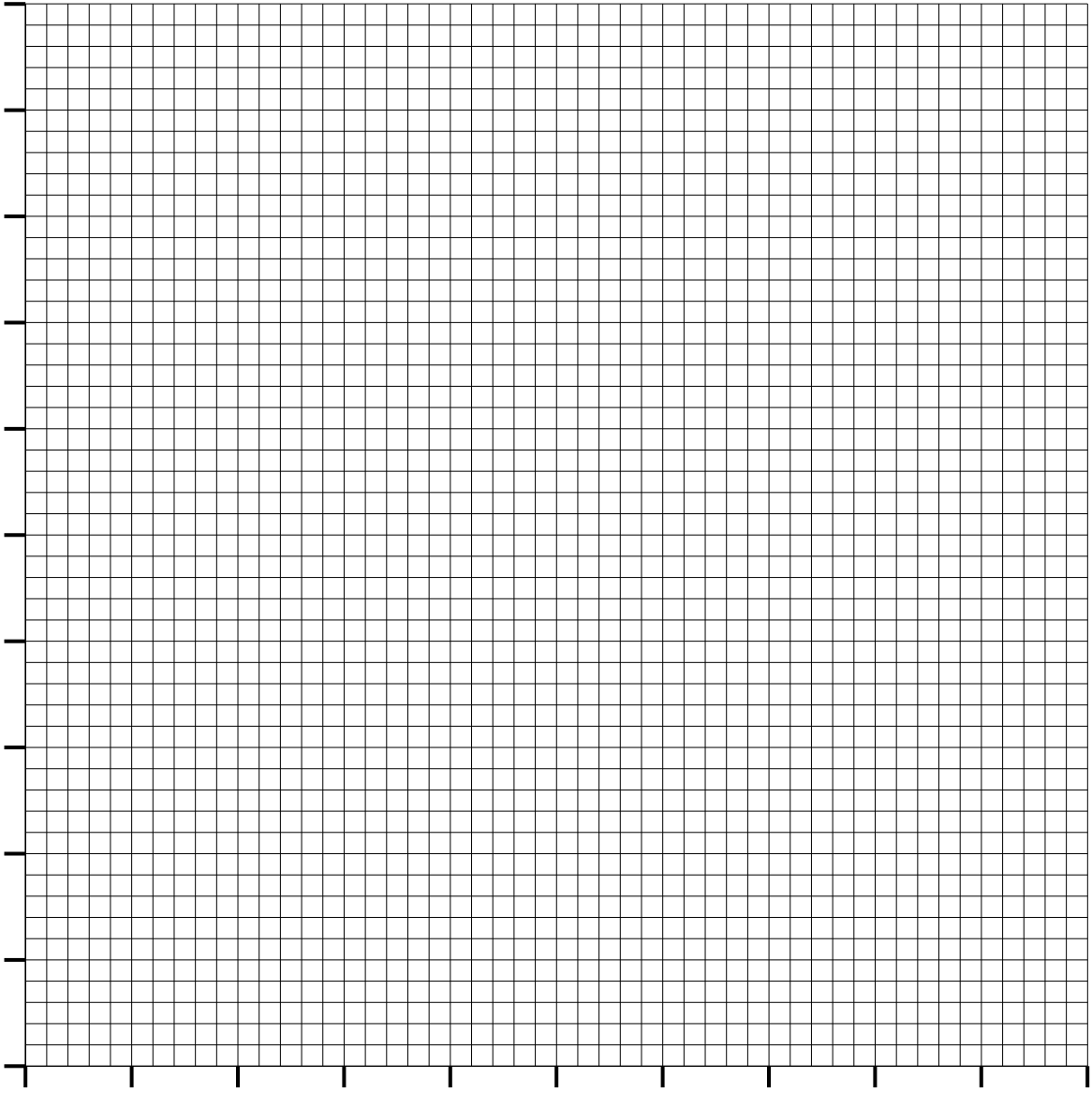(**i**) $Q_2 - Q_1$

(**ii**) $Q_3 - Q_2$

Is the distribution negatively skewed or symmetrical or positively skewed ?

[**3 marks**]

(**d**)   Suggest how the distribution might be interpreted by a passenger who frequently flies from City *A* to City *B*.

[**1 mark**]

Question 2

A supermarket knows the age of each person who pays for their shopping using a loyalty card. They are interested in the 'age profile' of their customers.

The data reveals that the lowest age was 18 (persons below that age cannot own a loyalty card) and the highest was 98.
The lower quartile was 30 and the upper quartile was 68.
The median age of the shoppers was 44.

(i)   Plot a box and whisker diagram to show this information.

0   10   20   30   40   50   60   70   80   90   100
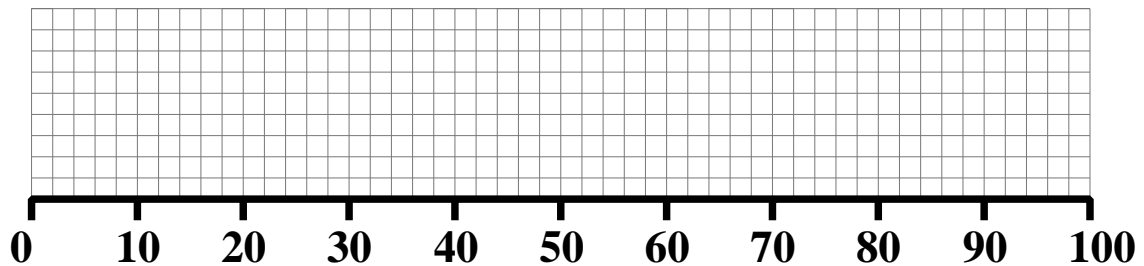
(ii)   The interquartile rage (IQR) = the upper quartile - the lower quartile.

$$IQR = Q_3 - Q_1$$

What is the IQR of the age profile of supermarket customers ?

(iii)   Calculate

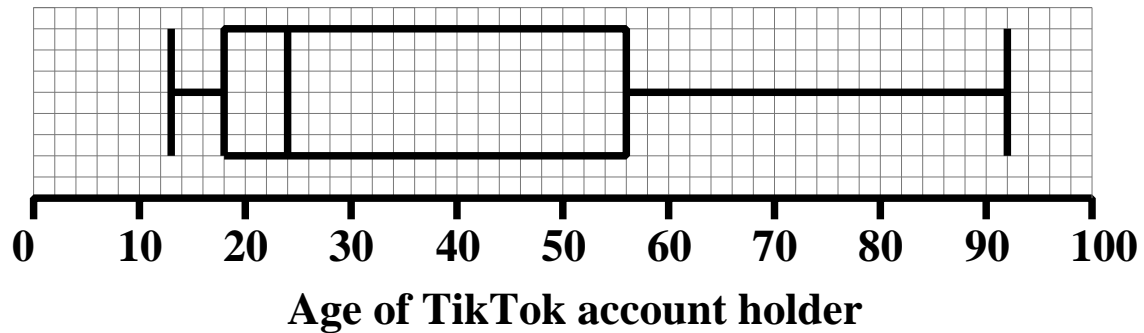   (a) $Q_2 - Q_1$

                        (b) $Q_3 - Q_2$

Is the distribution negatively skewed or symmetrical or positively skewed ?

Question 3

The box plot below shows the age profile of all persons with a TikTok account.

Age of TikTok account holder

State   **(i)**   The lower quartile, Q_1

           (ii)   The upper quartile, Q_3

           (iii)   The interquartile rage, IQR

           (iv)   The median age, Q_2

           (v) $Q_2 - Q_1$

           (vi) $Q_3 - Q_2$

           Is the data is skewed ?

           Justify your answer and give an interpretation of what this reveals in terms of the age profile of TikTok users.

Question 4

S1 examination question, January 2010, Q2 (edited)

The 19 employees of a company take an aptitude test.

The scores out of 40;

07	18	18	24	24	26	28	32	33	33
33	34	35	39	40	40	40	40	40	

Find

(**a**)   the median score

[1 mark]

(**b**)   the interquartile range
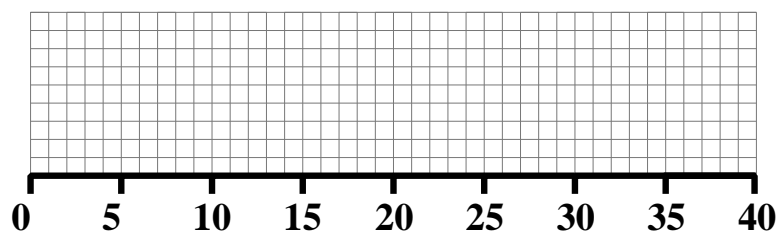
[3 marks]

The company director decides that any employees whose scores are so low
that they are outliers will undergo retraining.
An outlier is an observation whose value is less than the lower quartile
minus 1.0 times the interquartile range.

(**c**)   Explain why there is only one employee who will undergo retraining.

[2 marks]

(**d**)   On the graph paper below draw a box plot to illustrate the employee's scores

[3 marks]

Question 5

In a 100 metre sprint the running time (in seconds) of 23 runners was as follows.

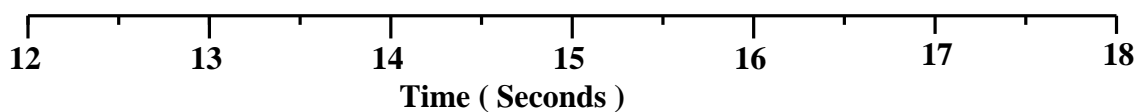|								
15.8	14.2	15.7	15.3	17.0	15.2	12.9	13.8	15.7
16.1	17.6	16.4	12.9	13.4	13.6	14.8	14.6	14.1
16.4	15.7	13.4	14.4	15.2				

(**i**)   Put the data in order.
   Take care to get this correct and try to get a methodical system going.

(**ii**)   Calculate the lower quartile.

(**iii**)   Calculate the upper quartile.

(**iv**)   Calculate the median.

(**v**)   Plot a box and whisker diagram of the data.

Time (Seconds)

(**vi**)   What is the interquartile range for this data ?