1 (a) Describe the main features and uses of a box plot.

(3)

Children from schools *A* and *B* took part in a fun run for charity. The times, to the nearest minute, taken by the children from school *A* are summarised in Figure 1.

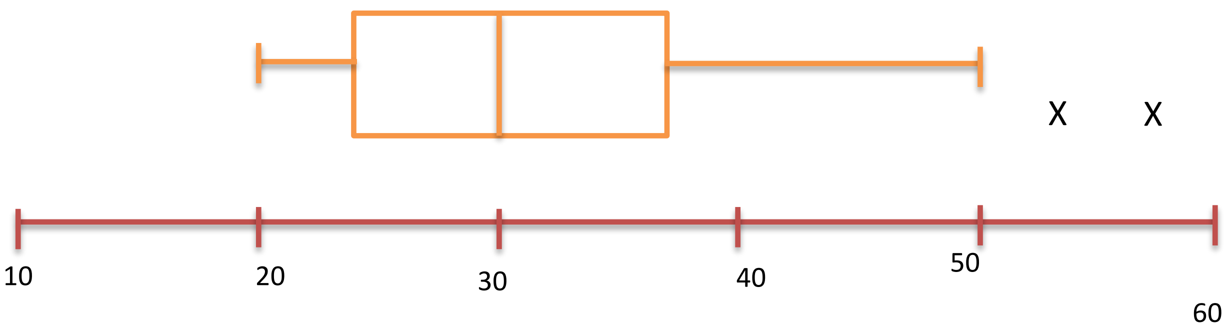


Figure 1

(b) (i) Write down the time by which 75% of the children in school *A* had completed the run.
(ii) State the name given to this value.

(2)

(c) Explain what you understand by the two crosses (x) on Figure 1.

(2)

For school *B* the least time taken by any of the children was 25 minutes and the longest time was 55 minutes. The three quartiles were 30, 37 and 50 respectively.

(d) Draw a box plot to represent the data from school *B*.

(4)

(e) Compare and contrast these two box plots.

(4)

1 a) A box plot shows the upper quartile the median and the lower quartile. The line through the middle indicates the range of the results and the crosses show data points which lie outside the given range and are considered to be outliers. The shape shows the skewness of the data.

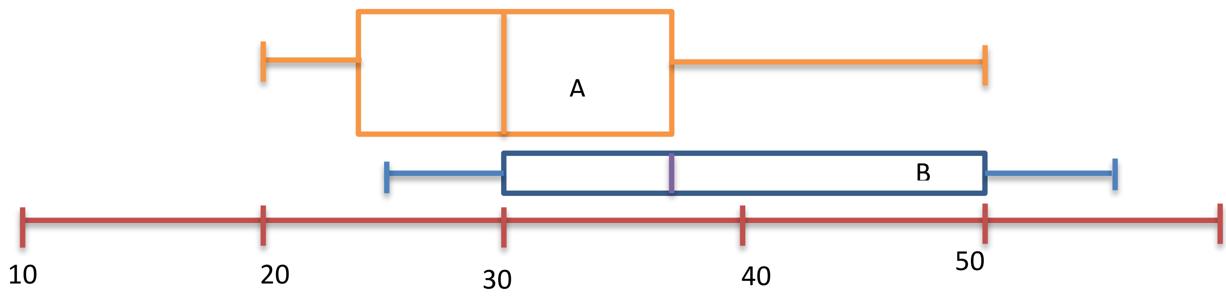
b) 37 mins. This is the third quartile or upper quartile.

c) The crosses are outliers which are considered to be outside a sensible range for the data and need to be considered to see if they should be included in any further calculations. In this case the children probably walked rather than ran.

d) See box plot below.

e) Children from school A generally took less time that team B

There are no outliers in B but there are in A.
Only 50% of the children in school B take less than 37 mins whereas 75% of children from school A take less than 37 mins.
Interquartile range of A is less that the interquartile range of B.



2. Sunita and Shelley talk to one another once a week on the telephone. Over many weeks they recorded, to the nearest minute, the number of minutes spent in conversation on each occasion. The following table summarises their results.

Time (to the nearest minute)	`Number of Conversations	Time Boundaries
5-9	2	4.5-9.5
10-14	9	9.5-14.5
15-19	20	14.5-19.5
20-24	13	19.5-24.5
25-29	8	24.5-29.5
30-34	3	29.5-34.5

Two of the conversations were chosen at random.
(a) Find the probability that both of them were longer than 24.5 minutes.
(2)

The mid-point of each class was represented by *x* and its corresponding frequency by *f*, giving $\Sigma f x$ =1060.
(b) Calculate an estimate of the mean time spent on their conversations.
(2)

During the following 25 weeks they monitored their weekly conversations and found that at the end of the 80 weeks their overall mean length of conversation was 21 minutes.

(c) Find the mean time spent in conversation during these 25 weeks.
(4)

(d) Comment on these two mean values.
(2)

2 a) One longer than 24.5 is that two top values i.e. $\frac{8+3}{2+9+20+13+8+3} = \frac{11}{55} = \frac{1}{5} = 0.2$

Second one is therefore $\frac{10}{54}$

P of both is $\frac{11}{55} \times \frac{10}{54} = \frac{110}{2970} = \frac{1}{27}$

b) Using $\mu = \frac{\sum fx}{n}$

$$\mu = \frac{1060}{55} = 19.3\ (1.d.p)$$

c) Using common sense

$$\frac{55 \times 19.27 + 25\mu_2}{80} = 21$$

$$1059.85 + 25\mu_2 = 1680$$

$$\mu_2 = 24.81\ (2.d.p)$$

d) The mean is much higher in the last 25 weeks making the average mean higher. Maybe they have finished their A-levels!

3. A metallurgist measured the length, *l* mm, of a copper rod at various temperatures, *t* °C, and recorded the following results.

T=x	l	y
20.4	**2461.12**	1.12
27.3	**2461.41**	1.41
32.1	**2461.73**	1.73
39.0	**2461.88**	1.88
42.9	**2462.03**	2.03
49.7	**2462.37**	2.37
58.3	**2462.69**	2.69
67.4	**2463.05**	3.05
$\sum x = 337.1$		$\sum y = 16.28$

The results were then coded such that *x* = *t* and *y* = *l* – 2460.00.
(a) Calculate *Sxy* and *Sxx*.
(You may use Σ*x*^{2}= 15965.01 and Σ*xy* = 757.467)

(5)

(b) Find the equation of the regression line of *y* on *x* in the form *y = a + bx*.

(5)

(c) Estimate the length of the rod at 40 °C.

(3)

(d) Find the equation of the regression line of *l* on *t*.

(2)

(e) Estimate the length of the rod at 90 °C.

(1)

(f) Comment on the reliability of your estimate in part (e).

(2)

3 a) See table for calculations

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$$

$$S_{xy} = 757.467 - \frac{337.1 \times 16.28}{8} = 71.4685$$

And

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{xx} = 15965.01 - \frac{337.1^2}{8} = 1760.45875$$

b) Start by finding b

$$b = \frac{S_{xy}}{S_{xx}}$$

$$b = \frac{71.4685}{1760.45875} = 0.04059652065$$

Find a using

$$a = \bar{y} - b\bar{x}$$

$$a = \frac{\sum y}{n} - b\frac{\sum x}{n} = \frac{16.28}{8} - 0.04059652065\frac{337.1}{8}$$

$$a = 0.324364111$$

Therefore

$$y = 0.324 + 0.0406x \ (3.s.f)$$

c) So x=40 find y from equation

$$y = 0.324 + 0.0406 \times 40 = 1.948$$

Therefore l=y+2460

$$l = 2460 + 1.948 = 2461.95 \ (2.d.p)$$

d) Find L on t, substitute in t for x and l for y

$$l = 0.324 + 0.0406t + 2460$$

$$l = 2460.324 + 0.0406t$$

e) When t=90 just sub in

$$l = 2460.324 + 0.0406 \times 90 = 2463.978$$

f) 90 is considerably outside the dataset and requires considerable extrapolation. Therefore this result in e) is unreliable.

4. The random variable *X* has the discrete uniform distribution

$$P(X = x) = \frac{1}{5}, \quad x = 1, 2, 3, 4, 5$$

(a) Write down the value of E(*X*) and show that Var(*X*) = 2.

(3)

Find

(b) E(3*X* – 2),

(2)

(c) Var(4 – 3*X*).

(2)

4 a) Using $E(X) = \sum xP(X = x)$

$$E(X) = 1 \times \frac{1}{5} + 2 \times \frac{1}{5} + 3 \times \frac{1}{5} + 4 \times \frac{1}{5} + 5 \times \frac{1}{5}$$

$$E(X) = \frac{1}{5} + \frac{2}{5} + \frac{3}{5} + \frac{4}{5} + \frac{5}{5} = \frac{15}{5} = 3$$

$Var(X) = E(X^2) - (E(X))^2$

$$E(X^2) = 1 \times \frac{1}{5} + 4 \times \frac{1}{5} + 9 \times \frac{1}{5} + 16 \times \frac{1}{5} + 25 \times \frac{1}{5}$$

$$E(X^2) = \frac{1}{5} + \frac{4}{5} + \frac{9}{5} + \frac{16}{5} + \frac{25}{5} = 11$$

$Var(X) = E(X^2) - (E(X))^2$ $\quad Var(X) = 11 - 3^2 = 2$

b) Using $E(aX + b) = aE(X) + b$ $\quad E(3X - 2) = 3E(X) - 2 = 9 - 2 = 7$

c) Using $Var(aX + b) = a^2Var(X)$ $\quad Var(4 - 3X) = 9Var(X) = 18$

5. From experience a high-jumper knows that he can clear a height of at least 1.78 m once in 5 attempts. He also knows that he can clear a height of at least 1.65 m on 7 out of 10 attempts.

Assuming that the heights the high-jumper can reach follow a Normal distribution,
(a) draw a sketch to illustrate the above information,

(3)

(b) find, to 3 decimal places, the mean and the standard deviation of the heights the high-jumper can reach,

(6)

(c) calculate the probability that he can jump at least 1.74 m.

(3)

5 a) Turn the question into equations $\quad P(X > 1.78) = \frac{1}{5} = 0.2 \text{ or } P(X < 1.78) = 0.8$

Sketch with 1.78 on right hand side and 1.65 on left. $\quad P(X > 1.65) = \frac{7}{10} = 0.7 \text{ or } P(X < 1.65) = 0.3$

b)

$$P\left(z < \frac{1.78 - \mu}{\sigma}\right) = 0.8$$

$$1.78 - \mu = 0.8416\sigma \quad (1)$$

$$P\left(z < \frac{1.65 - \mu}{\sigma}\right) = 0.3$$

$$1.65 - \mu = -0.5244\sigma \quad (2)$$

Subtract (1) divide by (2) $\quad 0.5244(1.78 - \mu) = -0.8416(1.65 - \mu)$

And

$$\mu(-0.5244 - 0.8416) = -0.8416 \times 1.65 - 1.78 \times 0.5244$$
$$1.366\mu = 2.322072$$

$$\mu = 1.69990 = 1.7\ (2.s.f)$$

Sub back into (1)

$$1.78 - 1.7 = 0.8416\sigma \qquad \frac{0.08}{0.8416} = \sigma \quad \sigma = 0.09506$$

c)

$$P(X > 1.74) = 1 - P(X < 1.74)$$

$$P(X > 1.74) = 1 - P\left(z < \frac{1.74 - 1.7}{0.9506}\right)$$
$$= 1 - P(z < 0.421) = 1 - 0.6628$$
$$= 0.3372$$

6. A group of 100 people produced the following information relating to three attributes.

The attributes were wearing glasses, being left handed and having dark hair.
Glasses were worn by 36 people, 28 were left handed and 36 had dark hair. There were 17 who wore glasses and were left handed, 19 who wore glasses and had dark hair and 15 who were left handed and had dark hair. Only 10 people wore glasses, were left handed and had dark hair.

(a) Represent these data on a Venn diagram.

(6)

A person was selected at random from this group.

Find the probability that this person
(b) wore glasses but was not left handed and did not have dark hair,

(1)

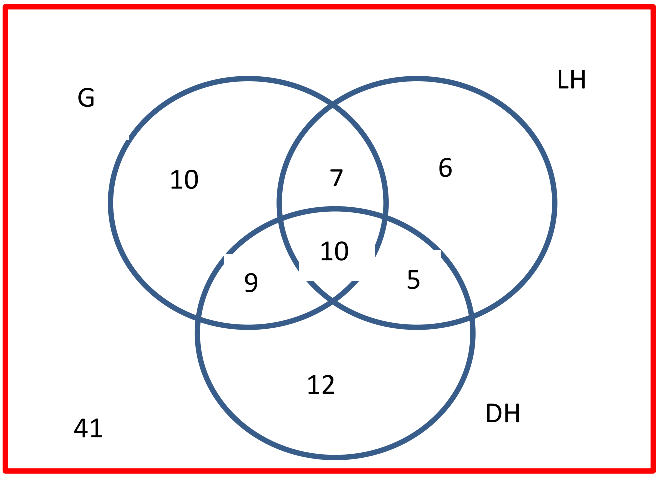
(c) did not wear glasses, was not left handed and did not have dark hair,

(1)

(d) had only two of the attributes,

(2)

(e) wore glasses given that they were left handed and had dark hair.



(3)

6 a) Start with the overlap numbers and work backwards. i.e start with 10 in the middle, carry 19-10,17-10,15-10 and then find the numbers left for each circle. Finally work out how many people have all three attributes and find how many do not.

b)
$$P(G,\overline{LH},\overline{DH}) = \frac{10}{100} = 0.1$$

c)
$$P(\bar{G},\overline{LH},\overline{DH}) = \frac{41}{100} = 0.41$$

d)
$$P(two\ attributes) = \frac{9+7+5}{100} = 0.21$$

e) P (LH and DH)=15
P(G and LH and DH)=10

$$P(G|LH\ AND\ DH) = \frac{10}{15} = \frac{2}{3}$$