5. Tomas is studying the relationship between temperature and hours of sunshine in *Seapron*. He records the midday temperature, t °C, and the hours of sunshine, s hours, for a random sample of 9 days in October. He calculated the following statistics

$$\sum s = 15 \qquad \sum s^2 = 44.22 \qquad \sum t = 127 \qquad S_{tt} = 10.89$$

(a) Calculate S_{ss}

(2)

Tomas calculated the product moment correlation coefficient between s and t to be 0.832 correct to 3 decimal places.

(b) State, giving a reason, whether or not this correlation coefficient supports the use of a linear regression model to describe the relationship between midday temperature and hours of sunshine.

(1)

(c) State, giving a reason, why the hours of sunshine would be the explanatory variable in a linear regression model between midday temperature and hours of sunshine.

(1)

(d) Find S_{st}

(3)

(e) Calculate a suitable linear regression equation to model the relationship between midday temperature and hours of sunshine.

(4)

(f) Calculate the standard deviation of s

(1)

Tomas uses this model to estimate the midday temperature in *Seapron* for a day in October with 5 hours of sunshine.

(g) State the value of Tomas' estimate.

(1)

Given that the values of s are all within 2 standard deviations of the mean,

(h) comment, giving your reason, on the reliability of this estimate.

(2)

2.

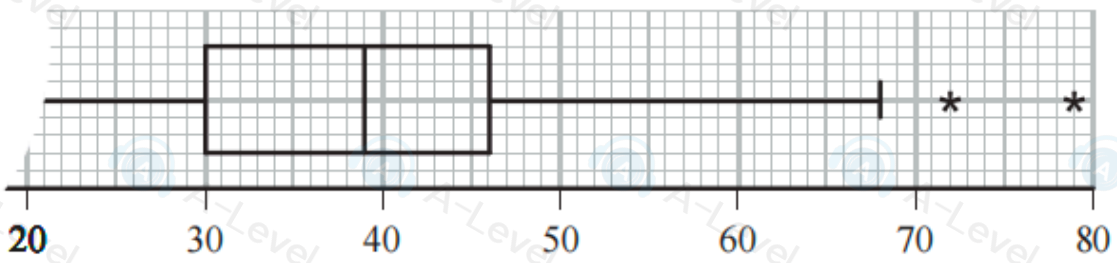


Figure 1

Figure 1 shows part of a box and whisker plot for the marks in an examination with a large number of candidates. Part of the lower whisker has been torn off.

(a) Given that 75% of the candidates passed the examination, state the lowest mark for the award of a pass.

(1)

(b) Given that the top 25% of the candidates achieved a merit grade, state the lowest mark for the award of a merit grade.

(1)

An outlier is defined as any value greater than c or any value less than d where

$$c = Q_3 + 1.5(Q_3 - Q_1)$$
$$d = Q_1 - 1.5(Q_3 - Q_1)$$

(c) Find the value of c and the value of d.

(2)

(d) Write down the 3 highest marks scored in the examination.

(2)

The 3 lowest marks in the examination were 5, 10 and 15

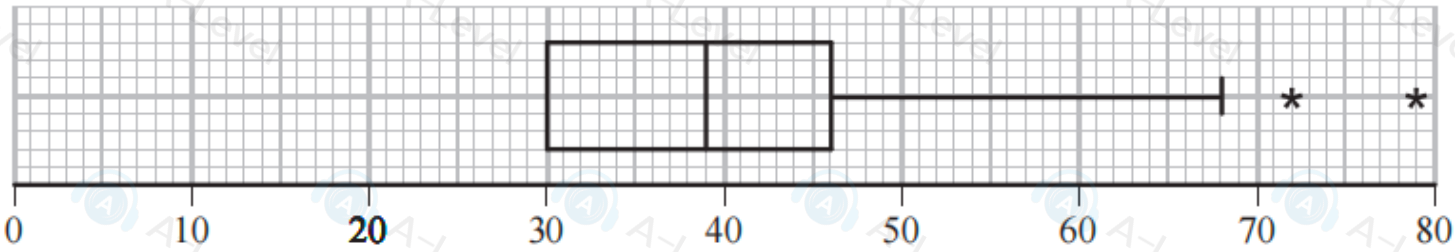
(e) On the diagram on page 7, complete the box and whisker plot.

(3)

Three candidates are selected at random from those who took this examination.

(f) Find the probability that all 3 of these candidates passed the examination but only 2 achieved a merit grade.

(3)



Turn over for a spare diagram if you need to redraw your plot.

1. A random sample of 10 cars of different makes and sizes is taken and the published miles per gallon, p, and the actual miles per gallon, m, are recorded. The data are coded using variables $x = \frac{p}{10}$ and $y = m - 25$

The results for the coded data are summarised below.

x	6.89	3.67	5.92	5.04	4.87	3.92	4.71	5.14	3.65	5.23
y	30	3	22	15	13	8	15	13.5	3	19

(You may use $\sum y^2 = 2628.25$ $\sum xy = 768.58$ $S_{xx} = 9.25924$ $S_{xy} = 74.664$)

(a) Show that $S_{yy} = 626.025$

(2)

(b) Find the product moment correlation coefficient between x and y.

(2)

(c) Give a reason to support fitting a regression model of the form $y = a + bx$ to these data.

(1)

(d) Find the equation of the regression line of y on x, giving your answer in the form $y = a + bx$.
Give the value of a and the value of b to 3 significant figures.

(3)

A car's published miles per gallon is 44

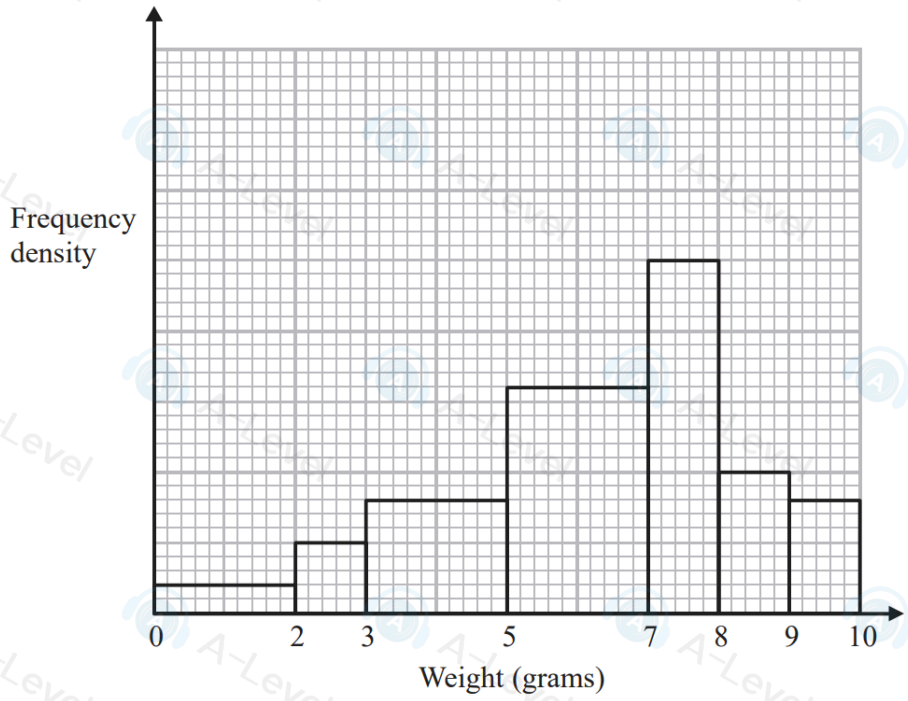
(e) Estimate the actual miles per gallon for this particular car.

(3)

(f) Comment on the reliability of your estimate in part (e). Give a reason for your answer.

(2)

1. Ralph records the weights, in grams, of 100 tomatoes. This information is displayed in the histogram below.



Given that 5 of the tomatoes have a weight between 2 and 3 grams,

(a) find the number of tomatoes with a weight between 0 and 2 grams.

(2)

One of the tomatoes is selected at random.

(b) Find the probability that it weighs more than 3 grams.

(2)

(c) Estimate the proportion of the tomatoes with a weight greater than 6.25 grams.

(2)

(d) Using your answer to part (c), explain whether or not the median is greater than 6.25 grams.

(1)

Given that the mean weight of these tomatoes is 6.25 grams and using your answer to part (d),

(e) describe the skewness of the distribution of the weights of these tomatoes. Give a reason for your answer.

(1)

Two of these 100 tomatoes are selected at random.

(f) Estimate the probability that both tomatoes weigh within 0.75 grams of the mean.

(4)

3. The table shows the price of a bottle of milk, m pence, and the price of a loaf of bread, b pence, for 8 different years.

m	29	29	35	39	41	43	44	46
b	75	83	91	121	120	126	119	126

(You may use $S_{bb} = 3083.875$ and $S_{mm} = 305.5$)

(a) Find the exact value of $\sum bm$

(1)

(b) Find S_{bm}

(3)

(c) Calculate the product moment correlation coefficient between b and m

(2)

(d) Interpret the value of the correlation coefficient.

(1)

A ninth year is added to the data set. In this year the price of the bottle of milk is 46 pence and the price of a loaf of bread is 175 pence.

(e) Without further calculation, state whether the value of the product moment correlation coefficient will increase, decrease or stay the same when all nine years are used. Give a reason for your answer.

(2)

2. A biologist records the length, y cm, and the weight, w kg, of 50 rabbits.
The following summary statistics are calculated from these data.

$$\sum y = 2015 \quad \sum y^2 = 81938.5 \quad \sum w = 125 \quad S_{ww} = 72.25 \quad S_{yw} = 219.55$$

(a) (i) Show that $S_{yy} = 734$

(ii) Calculate the product moment correlation coefficient for these data.
Give your answer to 3 decimal places.

(3)

(b) Interpret your value of the product moment correlation coefficient.

(1)

The biologist believes that a linear regression model may be appropriate to describe these data.

(c) State, with a reason, whether or not your value of the product moment correlation coefficient is consistent with the biologist's belief.

(1)

(d) Find the equation of the regression line of w on y, giving your answer in the form $w = a + by$

(4)

Jeff has a pet rabbit of length 45 cm.

(e) Use your regression equation to estimate the weight of Jeff's rabbit.

(2)

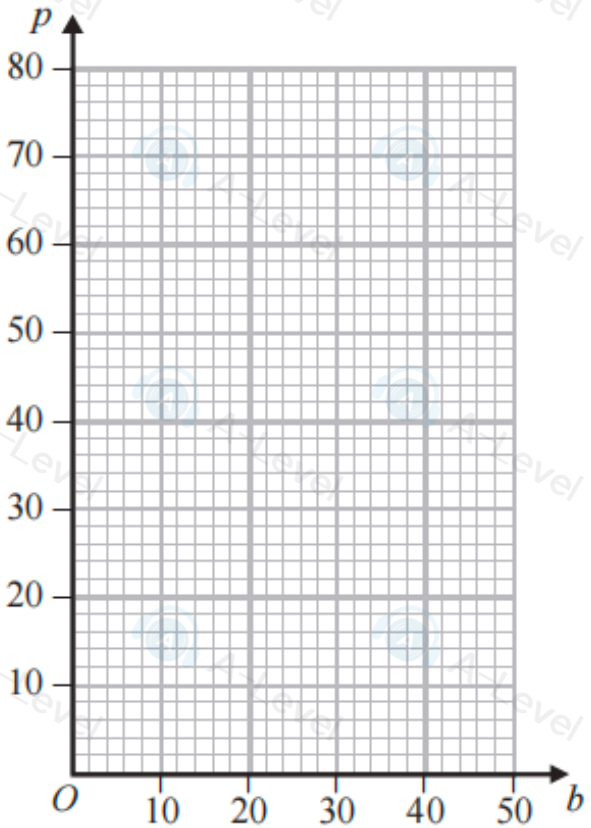
7. A doctor is investigating the correlation between blood protein, p, and body mass index, b.

He takes a random sample of 8 patients and the data are shown in the table below.

Patient	A	B	C	D	E	F	G	H
b	32	36	40	44	42	21	27	37
p	18	21	31	39	21	12	19	70

(a) Draw a scatter diagram of these data on the axes provided.

(2)



The doctor decides to leave out patient H from his calculations.

(b) Give a reason for the doctor's decision.

(1)

For the 7 patients A, B, C, D, E, F and G,

$$S_{bp} = 369, \quad S_{pp} = 490 \quad \text{and} \quad S_{bb} = 423\frac{5}{7}$$

(c) Find the product moment correlation coefficient, r, for these 7 patients.

(2)

(d) Without any further calculations, state how r would differ from your answer in part (c) if it was calculated for all 8 patients.

(1)