

3. Jim records the length, l mm, of 81 salmon. The data are coded using $x = l - 600$ and the following summary statistics are obtained.

$$n = 81 \quad \sum x = 3711 \quad \sum x^2 = 475\,181$$

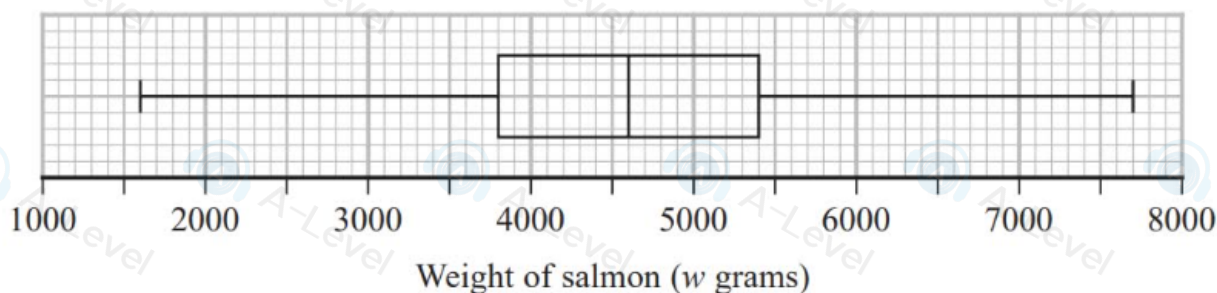
- (a) Find the mean length of these salmon.

(3)

- (b) Find the variance of the lengths of these salmon.

(2)

The weight, w grams, of each of the 81 salmon is recorded to the nearest gram. The recorded results for the 81 salmon are summarised in the box plot below.



- (c) Find the maximum number of salmon that have weights in the interval

$$4600 < w \leq 7700$$

(1)

Raj says that the box plot is incorrect as Jim has not included outliers.

For these data an outlier is defined as a value that is more than

$$1.5 \times \text{IQR} \text{ above the upper quartile or } 1.5 \times \text{IQR} \text{ below the lower quartile}$$

- (d) Show that there are no outliers.

(3)

6. In a factory, a machine is used to fill bags of rice. The weights of bags of rice are modelled using a normal distribution with mean 250g.

Only 1% of the bags of rice weigh more than 256g.

- (a) Write down the percentage of bags of rice with weights between 244g and 256g.

(1)

- (b) Find the standard deviation of the weights of the bags of rice.

(3)

An inspection consists of selecting a bag of rice at random and checking if its weight is within 4g of the mean. If the weight is more than 4g away from the mean, then a second bag of rice is selected at random and checked. If the weight of each of the 2 bags of rice is more than 4g away from the mean, then the machine is shut down.

- (c) Find the probability that the machine is shut down following an inspection.

(4)

8. A manager records the number of hours of overtime claimed by 40 staff in a month.

The histogram in Figure 1 represents the results.

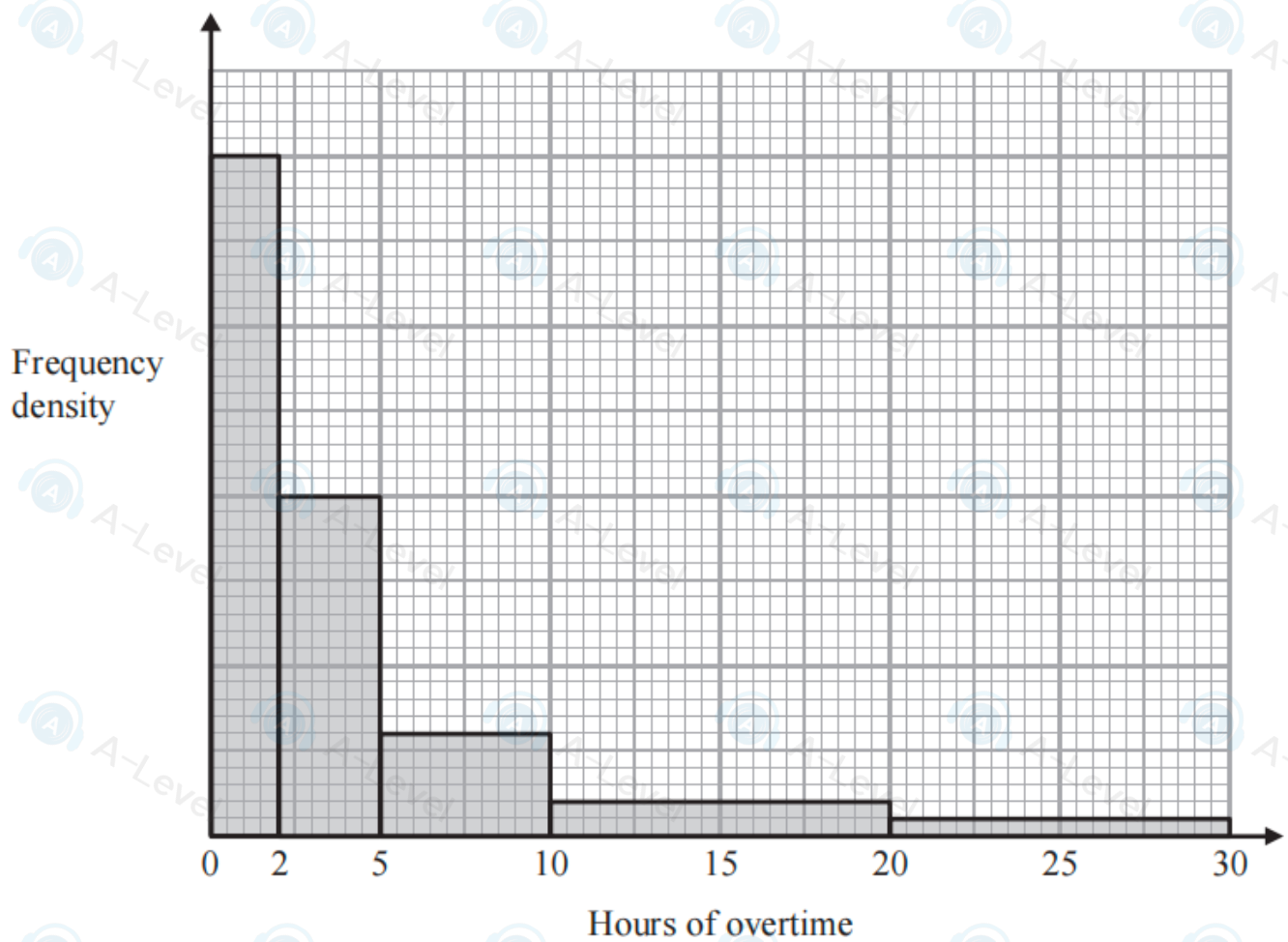


Figure 1

- (a) Calculate the number of staff who have claimed less than 10 hours of overtime in the month. (4)
- (b) Estimate the median number of hours of overtime claimed by these 40 staff in the month. (2)
- (c) Estimate the mean number of hours of overtime claimed by these 40 staff in the month. (2)

The manager wants to compare these data with overtime data he collected earlier to find out if the overtime claimed by staff has decreased.

- (d) State, giving a reason, whether the manager should use the median or the mean to compare the overtime claimed by staff. (2)

1. Jeremiah is investigating the relationship between the annual heating bill, h dollars, and the total floor area, f square metres, of buildings.

A random sample of 8 buildings is taken and the data for each building are coded using

$$x = \frac{f - 3500}{80} \text{ and } y = \frac{h - 4500}{80}$$

The results for the coded data are summarised below

$$\sum x = 5 \quad \sum y = 0 \quad \sum xy = 1818 \quad S_{xx} = 1754$$

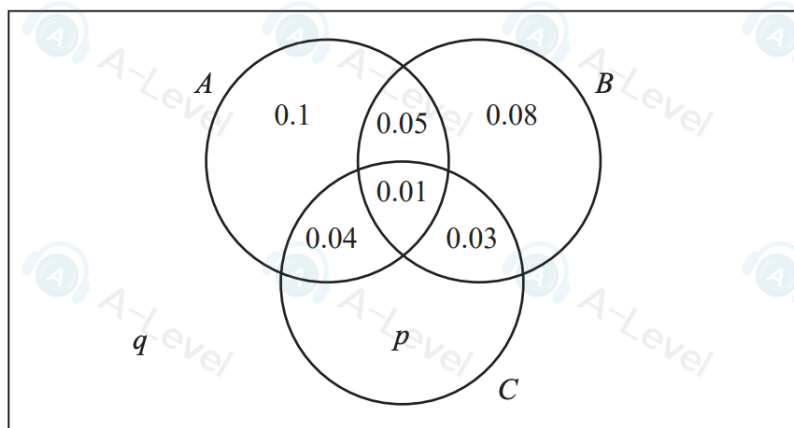
- (a) Calculate S_{xy} (1)
- (b) Find the equation of the regression line of h on f in the form $h = a + bf$ (6)
- (c) Give an interpretation of (i) the value of b in your regression line, (2)
(ii) the value of a in your regression line. (2)
- (d) Estimate the annual heating bill for a building with a total floor area of 4600 square metres. (2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

4. Pieces of wood cladding are produced by a timber merchant. There are three types of fault, A , B and C , that can appear in each piece of wood cladding.

The Venn diagram shows the probabilities of a piece of wood cladding having the various types of fault.



A piece of wood cladding is chosen at random.

- (a) Find the probability that the piece of wood cladding has more than one type of fault. (1)

Fault types A and C occur independently.

- (b) Find the probability that the piece of wood cladding has no faults. (4)

Given that the piece of wood cladding has fault A ,

- (c) find the probability that it also has fault B but not fault C . (2)

Two pieces of the wood cladding are selected at random.

- (d) Find the probability that both have exactly 2 types of fault. (3)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO

8. An orchard produces apples.

The weights, A grams, of its apples are normally distributed with mean μ grams and standard deviation σ grams.

It is known that

$$P(A < 162) = 0.1 \text{ and } P(162 < A < 175) = 0.7508$$

- (a) Calculate the value of μ and the value of σ (5)

A second orchard also produces apples.

The weights, B grams, of its apples have distribution $B \sim N(215, 10^2)$

An outlier is a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$

An apple is selected at random from this second orchard.

Using $Q_3 = 221.74$ grams,

- (b) find the probability that this apple is an outlier. (5)

Leave blank

5. The random variable X has a normal distribution with mean 10 and standard deviation 6

- (a) Find $P(X < 7)$ (3)

- (b) Find the value of k such that

$$P(10 - k < X < 10 + k) = 0.60 \quad (3)$$

A single observation x , of X , is to be taken.

A rectangle is drawn on a centimetre grid with vertices having coordinates $(0, 0)$, $(x, 0)$, $(x, x - 3)$ and $(0, x - 3)$

- (c) Find the probability that the area of this rectangle is more than 40 cm^2 (8)

1. The discrete random variable X has probability function $p(x)$ and cumulative distribution function $F(x)$ given in the table below.

x	1	2	3	4	5
$p(x)$	0.10	a	0.28	c	0.24
$F(x)$	0.10	0.26	b	0.76	d

- (a) Write down the value of d (1)

- (b) Find the values of a , b and c (3)

- (c) Write down the value of $P(X > 4)$ (1)

Two independent observations, X_1 and X_2 , are taken from the distribution of X .

- (d) Find the probability that X_1 and X_2 are both odd. (2)

Given that X_1 and X_2 are both odd,

- (e) find the probability that the sum of X_1 and X_2 is 6
Give your answer to 3 significant figures. (3)

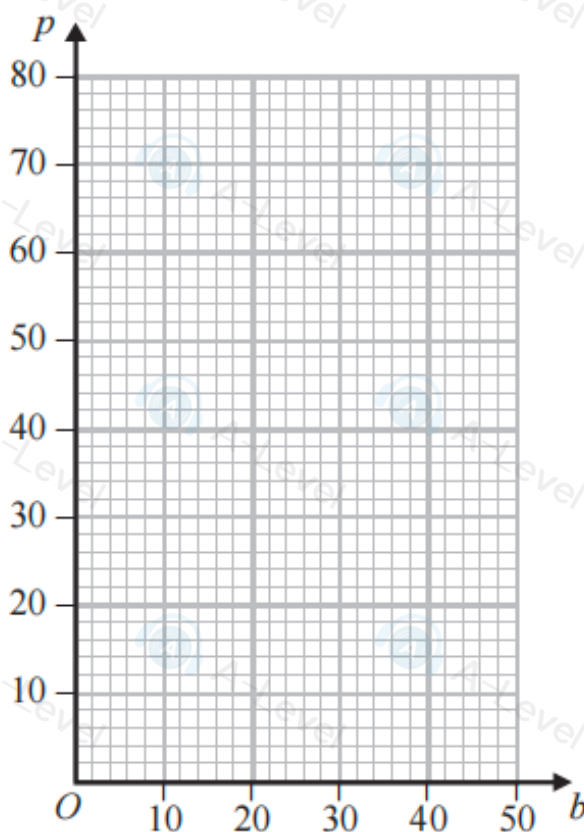
7. A doctor is investigating the correlation between blood protein, p , and body mass index, b .

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)