

A LEVEL SECTION TESTS

unit S1

mathematical models in statistics

S1/1

1. Give a brief definition of each of the following statistical terms:
 - (a) a population,
 - (b) a statistical model,
 - (c) a refinement of a model,
 - (d) sample space.
2.
 - (a) Distinguish between **quantitative** and **qualitative** data, giving an example of each.
 - (b) Distinguish between **discrete** and **continuous** data, giving an example of each.
3. The times taken by a large group of people to run a marathon were recorded. A list was drawn up showing each runner's time to the nearest hour, the nearest second and the nearest tenth of a second.
 - (a) State, for each of these degrees of accuracy, whether it would be more appropriate to model the rounded data as continuous or discrete.
 - (b) State an advantage of modelling data which is actually discrete as continuous.
4. Some numerical data is to be collected, organised, displayed and analysed.

Answer questions (a) to (g) in **each** of the two cases :

 - (i) the data consists of the total scores obtained when two biased dice are thrown together,
 - (ii) the data consists of the heights of the residents of a medium-sized town.
 - (a) Suggest a suitable method for collecting the raw data.
 - (b) Describe a method of organising the data and a way of presenting the organised data in an easy-to-understand numerical format.
 - (c) State whether the data should be treated as discrete or continuous, and suggest two appropriate methods of illustrating the data using a diagram.
 - (d) Name the measures of central tendency and of spread that can most easily and usefully be obtained from the data.
 - (e) Suggest a suitable statistical model for the distribution of the data.
 - (f) Describe the type of predictions that can be made using the chosen model.
 - (g) Briefly explain how the model could be tested to decide whether it is appropriate.

A LEVEL SECTION TESTS

unit S1

representation and summary of data

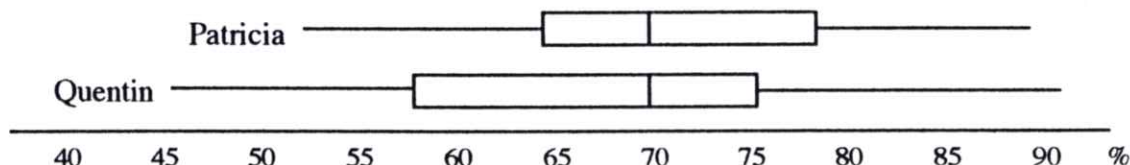
S1/2a

Give all answers correct to 3 significant figures where necessary.

1. The following data are the times, to the nearest minute, taken by 60 people to solve a puzzle.

11	46	33	59	29	12	43	23	15	37	42	51
69	24	40	55	38	21	16	43	52	27	32	10
20	49	32	45	57	36	43	29	21	40	12	50
37	28	26	47	52	41	34	50	19	36	44	13
23	30	35	21	16	53	22	14	31	51	42	39

- Construct a stem and leaf diagram to represent this data, using the groupings 10 - 14, 15 - 19, etc and stating the frequency for each group.
 - State (i) the modal class, (ii) the modal times.
 - Estimate the median and quartiles of the data, making your method clear.
 - On graph paper, showing your scale, draw a box and whisker plot of the same data.
 - Identify a probable outlier among the data.
2. The box plots below represent the marks obtained by two students, Patricia and Quentin, in a series of 50 tests taken throughout a year.



- Briefly compare these two students' test performances using the information in the diagram.
 - Sketch**, on a single diagram, cumulative frequency graphs for the two students' marks.
3. A tour operator's booking form asks for the age, in completed years, of each person who makes a booking. The information collected from 800 people in one year was summarised in a grouped frequency table and a histogram was drawn to represent the distribution. The modal class was 30 - 34, with frequency 160 people. The highest frequency in a class was 200 people, for the 50 - 69 age group. The rectangles representing these two classes were respectively 6.4 cm high and 8 cm wide. Find the total area of the histogram, in cm^2 .
4. The following data shows the heights, in centimetres, of 90 plants.

Height, x cm	0 -	20 -	30 -	35 -	40 -	50 -	60 - 90
Number of plants	8	16	14	18	11	14	9

- Represent this data by (i) a histogram, (ii) a cumulative frequency graph.
- Use your graphs to estimate (i) the mode of the heights, (ii) the median height, (iii) the number of plants between 25 and 42 cm in height.

A LEVEL SECTION TESTS

unit S1

representation and summary of data

S1/2b

Give all answers correct to 3 significant figures where necessary.

1. (a) Find the mean and the standard deviation of the numbers
 $-7, -5, 6, 7, 9, 11, 14$.
 (b) Using your answers to part (a), and showing your method clearly, deduce the mean and the standard deviation of (i) $0, 2, 13, 14, 16, 18, 21$, (ii) $-35, -25, 30, 35, 45, 55, 70$.

2. (a) Write down the median, M , of the following values of a variable X :
 $210, 230, 240, 270, 290, 300, 310, 310$.
 (b) Given that $y = x - M$, find the mean and standard deviation of the eight values of Y .
 (c) Deduce the mean, μ , and the standard deviation, σ , of the eight given values of X .
 (d) Find the value of $\sum (x - \mu)^2$.

3. The variable X takes the values of the class mid-points of the following grouped continuous data

Class	5 -	10 -	15 -	20 -	25 -	30 -	35 -	40 -	45 - 50
Frequency	5	14	16	24	35	5	3	2	6

 (a) Write down the upper class boundaries of the first eight classes
 (b) Use the coding $y = \frac{x - 27.5}{5}$ to calculate estimates of the mean and standard deviation.
 (c) Explain why your answers are only estimates.
 (d) State, giving a reason, whether the data would be more effectively summarised by the mean and standard deviation or by the median and quartiles.

4. The following table gives the weights, to the nearest kg, of the loads carried by ninety vans.

Weight (kg)	0 - 19	20 - 39	40 - 59	60 - 79	80 - 99	100 - 119	120 - 139
No. of loads	3	8	14	16	20	20	9

- (a) Use interpolation to calculate estimated values of (i) the median, (ii) the quartiles, (iii) the interquartile range, (iv) the seventeenth percentile.
- (b) Given that the smallest load weighed 8 kg and the range of weights was 120 kg, draw an accurate box and whisker plot of the data. Describe the skewness of the distribution.
- (c) If outliers are defined to be outside the range $Q_1 - 1.5(Q_3 - Q_1)$ to $Q_3 + 1.5(Q_3 - Q_1)$, determine whether there are any outliers among the given data.

For another set of loads, whose weights ranged from 11 to 125 kg, the median was 79 kg, the lower quartile was 51 kg and the interquartile range was 49 kg.

- (d) Draw a further box plot for these data on the same diagram. Briefly compare the two sets of data using your plots.

A LEVEL SECTION TESTS

unit S1

probability

S1/3

Give all answers correct to 3 significant figures where necessary.

- E, F and G are three events such that $P(E) = \frac{1}{3}$, $P(F) = \frac{1}{6}$, $P(E \cup G) = \frac{2}{3}$, $P(F \cup G) = \frac{5}{8}$.
Given that F and G are mutually exclusive,
(a) find (i) $P(G)$, (ii) $P(E \cap G)$.
(b) Determine, with explanation, whether E and G are independent events.
(c) Find (i) $P(G|E)$, (ii) $P(G|E^c)$.
- A computer generates two numbers at random. The first number can be 1, 2 or 3 and the second number can be 0, 2 or 3. In each case the outcomes are equally likely.
Using a sample space, or otherwise, find the probability that
(a) the sum of the two numbers is even, (b) the product of the two numbers is less than 4.
- A depleted pack of cards contains 8 black cards and 10 red cards. 2 of the black cards and 3 of the red cards are aces. George picks one card at random. Find the probability that it is
(a) an ace, (b) not an ace, given that it is red, (c) red, given that it is not an ace.
- In a large Sixth Form, 40% of the students do Mathematics, 30% do Physics and 24% do both Mathematics and Physics.
(a) Illustrate this information using a Venn diagram.
(b) Hence or otherwise find the probability that a student, chosen at random,
(i) does Physics but not Mathematics, (ii) does neither Mathematics nor Physics,
(iii) does not do Physics, given that he or she does not do Mathematics.
- The employees of a large company are awarded an annual bonus, which they can take in cash, vouchers or shares. The probabilities that a randomly chosen employee will take the bonus in cash and in vouchers are 0.3 and 0.45 respectively.
(a) Find the probability that a randomly chosen employee will take the bonus in shares.
(b) If two employees are chosen at random, find the probability that they take their bonuses in different forms.
(c) If three employees are chosen at random, find the probability that exactly two take cash.
(d) If four employees are chosen at random, find the probability that they all take vouchers.
30% of the bonuses are over £100, and of these, 70% are taken in shares.
(e) Find the probability that an employee who takes shares gets a bonus of over £100.
- A cubical die is biased so that the scores 1, 3 and 5 are all equally likely, the scores 2, 4 and 6 are all equally likely, and an odd number is twice as likely to be scored as an even number. If the die is thrown 12 times, find the probability that
(a) all 12 scores are odd, (b) the scores are alternately odd and even.

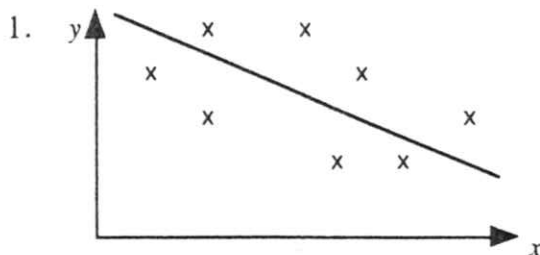
A LEVEL SECTION TESTS

unit S1

correlation and regression

S1/4

Give all answers correct to 3 significant figures where necessary.



The diagram shows a scatter graph on which values of a sample of bivariate data (x, y) have been plotted, together with the regression line of y on x .

- (a) Describe the type of correlation shown by the line.
- (b) Would you use this line to predict values of x for given values of y or to predict values of y for given values of x ?

(c) Comment on the accuracy you might expect if the line were used to make such predictions.

2. The following data was collected for a sample of six cars, showing their engine size x litres and average fuel consumption y litres per 100 km.

Car	A	B	C	D	E	F
x	0.9	1.2	1.4	1.7	2.3	2.5
y	5.2	6.3	5.9	6.4	7.3	8.9

- (a) Calculate the equations of the regression line of (i) y on x , (ii) x on y , expressing both answers in the form $y = ax + b$.

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

- (c) Use the equations of the regression lines to estimate the values of

- (i) y when $x = 2$, (ii) x when $y = 9.6$.

State, with a reason, how accurate you would expect these estimates to be.

- (d) Comment on the use of the lines to find values of x as y gets very small.

- (e) Write down a pair of values (x, y) which satisfy the equations of both regression lines and state their relevance to the data in the table.

3. For a pair of related variables p and q , the regression line of q on p has equation $q = 2.5p + 15$. The equation of the regression line of p on q can be written in the form $q = mp - 37$. The mean value of p is 13. Find (a) the mean value of q , (b) the value of m , (c) the product moment correlation coefficient between p and q , commenting on its significance. (d) Given that $x = 2p + 3$ and $y = 3q - 2$, state the product moment correlation coefficient between x and y .

4. Two variables x and y are such that, for a sample of six pairs of values,

$$\sum x = 93.3, \quad \sum y = 90.6, \quad \sum x^2 = 2066.6, \quad \sum y^2 = 5252.5.$$

The regression line of y on x has gradient 2.4. Find

- (a) $\sum xy$, (b) the equation of the regression line of x on y ,
- (c) the product moment correlation coefficient between y and x .

A LEVEL SECTION TESTS

unit S1

discrete random variables

S1/5

Give all answers correct to 3 significant figures where necessary.

1. The discrete random variable X has the following probability distribution :

x	0	1	2	3	4
$P(X = x)$	0.2	0.3	0.25		0.1

- (a) Find (i) $P(X = 3)$, (ii) $P(0 < X \leq 2)$. (b) Find the mean and the variance of X .
 (c) Write down a table to represent the cumulative distribution function $F(x)$.

2. X and Y are two independent random variables whose probability distributions are as follows:

x	1	2	3
$P(X = x)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$

y	1	2	3	4
$P(Y = y)$	$\frac{1}{2}$	p	p	$\frac{1}{4}$

- (a) Find (i) $E(X)$, (ii) $E(X^2)$, (iii) $\text{Var}(X)$. (b) Show that $p = \frac{1}{8}$.

The random variable Z is defined by $Z = X + Y$.

- (c) Tabulate the probability distribution of Z .

3. The random variable X has mean μ and standard deviation σ .

Express each of the following in terms of μ and σ as simply as possible:

- (a) $E(4X)$, (b) $\text{Var}(4X)$, (c) $E(2X - 3)$, (d) $\text{Var}(2X - 3)$.

4. The random variable X has the discrete uniform distribution on the set $\{1, 2, 3, \dots, n\}$.

- (a) If $n = 15$, find the mean and the variance of X .
 (b) If $n = 20$, find $P(3 < X \leq 11)$.
 (c) If the variance of X is 10, find (i) n , (ii) $E(X)$, (iii) $\text{Var}(3X + 2)$.

5. A spinner in the form of a regular hexagon is weighted so that the probability of each number being scored is proportional to that number.

The random variable X is the score, which can take the values 1, 2, 3, 4, 5 and 6.

- (a) If $P(X = x) = kx$, show that $k = \frac{1}{21}$.

- (b) Calculate the mean and the variance of X .

6. The random variable X takes the values, in pence, of the coins in a cash till. The proportions of the various coins are:

x	1	2	5	10	20	50	100	200
$P(X = x)$	0.1	0.1	q	q	r	0.35	r	0.05

Given that $E(X) = 39.2$, find (a) the values of q and r , (b) the standard deviation of X .

A LEVEL SECTION TESTS

unit S1

the normal distribution

S1/6

Give all answers correct to 3 significant figures where necessary.

- The random variable Z is normally distributed with mean 0 and variance 1.
 - Sketch the probability function of Z , indicating the mean and the standard deviation on the z -axis.
 - Find (i) $P(Z \leq 0.5)$, (ii) $P(Z > 1.5)$, (iii) $P(-1 < Z < 1)$,
(iv) $E(2Z + 3)$, (v) $\text{Var}(1 - 3Z)$.
- The random variable X is normally distributed with mean 3 and variance 4.
 - Sketch the probability function of X , indicating the mean and the standard deviation on the x -axis.
 - Express X in terms of Z , where Z is the random variable defined in Question 1.
 - Find (i) $P(X > 2)$, (ii) $P(X < 0)$, (iii) $P(1 < X < 5)$,
(iv) $E(2X + 3)$, (v) the standard deviation of $(1 - 3X)$.
- Given that $X \sim N(4.5, 2.5^2)$, find
 - $P(X < 4.2)$, (b) $P(X > 5)$, (c) $P(|X - 4| < 1.5)$.
- The times taken by a group of people to run a marathon are modelled as a continuous random variable having the normal distribution with mean 5 hours and standard deviation 1.5 hours.
 - Use this model to calculate
 - the probability that a runner chosen at random took between 4 and 7 hours,
 - the range, symmetrical about the mean, within which 85% of the runners' times lie.
 - Comment on the suitability of the normal distribution as a model in this case.
- The random variable X is distributed $N(\mu, \sigma^2)$.
Given that $P(X < 2) = 0.125$ and $P(X > 3) = 0.271$, find the values of μ and σ .
- The random variable X is normally distributed with variance 900. The probability that X is greater than 325 is 0.791, to 3 significant figures.
 - Calculate the mean of X .
 - In forty independent observations of X , how many would you expect to be less than 325?