Review exercise 1

E **1** A researcher is hired by a cleaning company to survey the opinions of employees on a proposed pension scheme. The company employs 55 managers and 495 cleaners.

a Explain what is meant by a census and give one disadvantage of using it in this context. **(2)**

To collect data the researcher decides to give a questionnaire to the first 50 cleaners to leave at the end of the day.

b State the sampling method used by the researcher. **(1)**

c Give 2 reasons why this method is likely to produce biased results. **(2)**

d Explain briefly how the researcher could select a sample of 50 employees using:

i a systematic sample

ii a stratified sample. **(2)**

← Sections 1.1, 1.2, 1.3

P **2** Data on the daily maximum relative humidity in Camborne during 1987 is gathered from the large data set. The daily maximum relative humidity, h, on the first five days in May is given below:

Day	1st	2nd	3rd	4th	5th
h	100	91	77	83	86

a Explain what is meant by opportunity sampling and describe one limitation of the method in this context. **(2)**

b Calculate an estimate for the mean daily maximum relative humidity for these five days. **(1)**

Joanna concludes that it is not likely to be misty in Camborne in May.

c State, with reasons, whether your answer in part **b** supports Joanna's conclusion. **(2)**

← Sections 1.3, 1.5, 2.1

E **3** Summarised below are the distances, to the nearest mile, travelled to work by a random sample of 120 commuters.

Distance, x miles	Number of commuters
$0 < x \leqslant 10$	10
$10 < x \leqslant 20$	19
$20 < x \leqslant 30$	43
$30 < x \leqslant 40$	25
$40 < x \leqslant 50$	8
$50 < x \leqslant 60$	6
$60 < x \leqslant 70$	5
$70 < x \leqslant 80$	3
$80 < x \leqslant 90$	1

a For this distribution, use linear interpolation to estimate its median. **(3)**

The midpoint of each class was represented by x and its corresponding frequency f giving

$\Sigma fx = 3610$ and $\Sigma fx^2 = 141\,600$

b Estimate the mean and standard deviation of this distribution. **(2)**

← Sections 2.1, 2.4

E/P **4** From the large data set, the daily total sunshine, s, in Leeming during June 1987 is recorded.

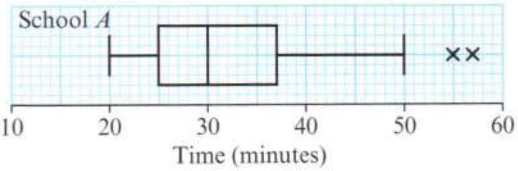
The data is coded using $x = 10s + 1$ and the following summary statistics are obtained.

$n = 30 \qquad \Sigma x = 947 \qquad S_{xx} = 33\,065.37$

Find the mean and standard deviation of the daily total sunshine. **(4)**

← Section 2.5

(E) **5** 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 below.



a 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)**

b Explain what you understand by the two crosses (×) 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.

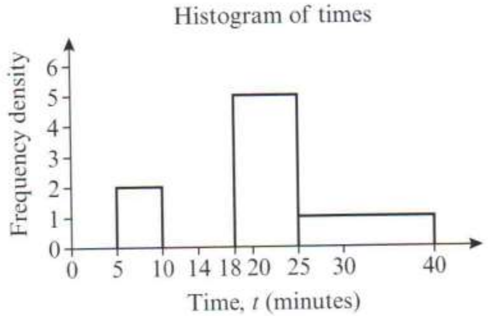
c On graph paper, draw a box plot to represent the data from school B. **(3)**

d Compare and contrast these two box plots. **(2)**

←Sections 3.1, 3.2, 3.5

(E/P) **6** The histogram below shows the times taken, t, in minutes, by a group of people to swim 500 m.

Histogram of times



a Copy and complete the frequency table and histogram for t. **(4)**

t	f
$5 \leqslant t < 10$	10
$10 \leqslant t < 14$	16
$14 \leqslant t < 18$	24
$18 \leqslant t < 25$	
$25 \leqslant t < 40$	

b Find the probability that a person chosen at random took longer than 20 minutes to swim 500 m. **(3)**

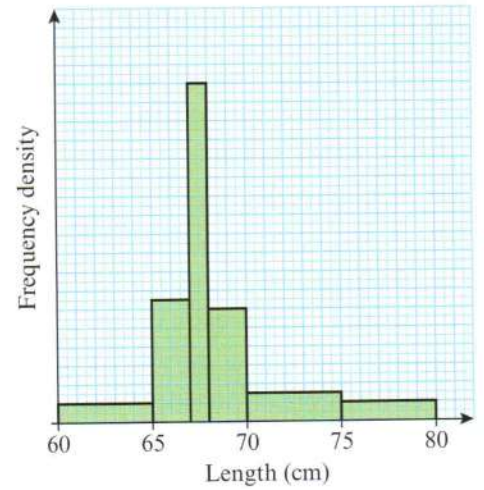
c Find an estimate of the mean time taken. **(2)**

d Find an estimate for the standard deviation of t. **(3)**

e Find an estimate for the median of t. **(2)**

←Sections 2.1, 2.4, 3.4

(E/P) **7** An ornithologist is collecting data on the lengths, in cm, of snowy owls. She displays the information in a histogram as shown below.



Given that there are 26 owls in the 65 to 67 cm class, estimate the probability that an owl, chosen at random is between 63 and 73 cm long. **(4)**

← Section 3.4

(E/P) **8** The daily maximum temperature is recorded in a UK city during May 2015.

t	f
$10 \leqslant t < 13$	1
$13 \leqslant t < 16$	7
$16 \leqslant t < 19$	13
$19 \leqslant t < 22$	10

a Use linear interpolation to find an estimate for the 20% to 80% interpercentile range. **(3)**

b Draw a cumulative frequency diagram to display this data. **(3)**

c Use your diagram to estimate the 20% to 80% interpercentile range and compare your answer to part **a**. Which estimate is likely to be more accurate? **(3)**

d Estimate the number of days in May 2015 where the daily maximum temperature in this city is greater than 15 °C. **(2)**

← Sections 2.3, 3.3

E **9** A manufacturer stores drums of chemicals. During storage, evaporation take place. A random sample of 10 drums was taken and the time in storage, x weeks, and the evaporation loss, y ml, are shown in the table below.

x	3	5	6	8	10	12	13	15	16	18
y	36	50	53	61	69	79	82	90	88	96

a On graph paper, draw a scatter diagram to represent these data. **(3)**

b Give a reason to support fitting a regression model of the form $y = a + bx$ to these data. **(1)**

The equation of the regression line of y on x is $y = 29.02 + 3.9x$.

c Give an interpretation of the value of the gradient in the equation of the regression line. **(1)**

The manufacturer uses this model to predict the amount of evaporation that would take place after 19 weeks and after 35 weeks.

d Comment, with a reason, on the reliability of each of these predictions. **(2)**

← Sections 4.2

P **10** The table shows average monthly temperature, t (°C) and the number of ice creams, c, in 100s, a riverside snack barge sells each month.

t	7	8	10	45	14	17	20	21	15	13	9	5
c	4	7	13	27	30	35	42	41	36	24	9	3

The following statistics were calculated for the data on temperature: mean = 15.3, standard deviation = 10.2 (both correct to 3 s.f.)

An outlier is an observation which lies ±2 standard deviations from the mean.

a Show that $t = 45$ is an outlier. **(1)**

b Give a reason whether or not this outlier should be omitted from the data. **(1)**

This value is omitted from the data, and the equation of the regression line of c on t for the remaining data is calculated as $c = 2.81t - 13.3$.

c Give an interpretation of the value of 2.81 in this regression equation. **(1)**

d State, with a reason, why using the regression line to estimate the number of ice creams sold when the average monthly temperature is 2 °C would not be appropriate. **(2)**

← Sections 3.1, 4.2

E **11** A bag contains nine blue balls and three red balls. A ball is selected at random from the bag and its colour is recorded. The ball is not replaced. A second ball is selected at random and its colour is recorded.

a Draw a tree diagram to represent the information **(3)**

Find the probability that:

b the second ball selected is red **(1)**

c the balls are different colours. **(3)**

← Section 5.4

E/P **12** For events A and B, P(A but not B) = 0.32, P(B but not A) = 0.11 and P(A or B) = 0.65.

a Draw a Venn diagram to illustrate the complete sample space for the events A and B. **(3)**

b Write down the value of P(A) and the value of P(B). **(2)**

c Determine whether or not A and B are independent. **(2)**

← Sections 5.2, 5.3

E **13** A company assembles drills using components from two sources. Goodbuy supplies 85% of the components and

Amart supplies the rest. It is known that 3% of the components supplied by Goodbuy are faulty and 6% of those supplied by Amart are faulty.

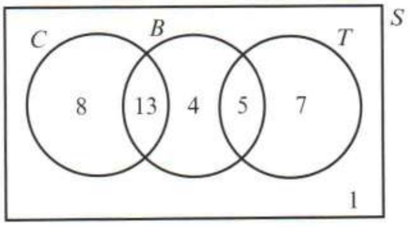
a Represent this information on a tree diagram. **(2)**

An assembled drill is selected at random.

b Find the probability that it is not faulty. **(3)**

← Section 5.4

E/P **14** The Venn diagram shows the number of children who like comics (C), books (B) or television (T).



a Which two hobbies are mutually exclusive? **(1)**

b Determine whether the events 'likes comics' and 'likes books' are independent. **(3)**

← Sections 5.2, 5.3

E/P **15** A fair coin is tossed 4 times.

Find the probability that:

a an equal number of heads and tails occur **(3)**

b all the outcomes are the same **(2)**

c the first tail occurs on the third throw. **(2)**

← Section 6.1

E/P **16** The probability of a bolt being faulty is 0.3. Find the probability that in a random sample of 20 bolts there are:

a exactly 2 faulty bolts **(2)**

b more that 3 faulty bolts. **(2)**

These bolts are sold in bags of 20. John buys 10 bags.

c Find the probability that exactly 6 of these bags contain more than 3 faulty bolts. **(3)**

← Sections 6.2, 6.3

E **17** The random variable X has probability function

$$P(X = x) = \frac{(2x - 1)}{36} \quad x = 1, 2, 3, 4, 5, 6.$$

a Construct a table giving the probability distribution of X. **(3)**

Find

b $P(2 < X \leqslant 5)$ **(2)**

← Section 6.1

E/P **18** A fair five-sided spinner has sectors numbered from 1 to 5. The spinner is spun and the number showing, X, is recorded.

a State the distribution of X. **(1)**

The spinner is now spun until it lands on an odd number, or until it has been spun four times. The random variable Y is defined as the number of spins in this experiment.

b Write down, in table form, the probability distribution of Y. **(3)**

c Find $P(Y > 2)$. **(2)**

← Section 6.1

E **19** The random variable $X \sim B(15, 0.32)$. Find:

a $P(X = 7)$ **b** $P(X \leqslant 4)$

c $P(X < 8)$ **d** $P(X \geqslant 6)$ **(4)**

← Section 6.3

E/P **20** A single observation is taken from a test statistic $X \sim B(40, p)$. Given that H_0: $p = 0.3$ and H_1: $p \neq 0.3$,

a Find the critical region for the test using a 2.5% significance level. (The probability in each tail should be as close as possible to 1.25 %) **(4)**

b State the probability of incorrectly rejecting the null hypothesis using this test. **(1)**

← Section 7.2

E/P **21** A drugs company claims that 75% of patients suffering from depression recover when treated with a new drug.

A random sample of 10 patients with depression is taken from a doctor's records.

a Write down a suitable distribution to model the number of patients in this sample who recover when treated with the new drug. **(1)**

Given that the claim is correct,

b find the probability that the treatment will be successful for exactly 6 patients. **(2)**

The doctor believes that the claim is incorrect and the percentage who will recover is lower. From her records she took a random sample of 20 patients who had been treated with the new drug. She found that 13 had recovered.

c Stating your hypotheses clearly, test, at the 5% level of significance, the doctor's belief. **(7)**

d From a sample of size 20, find the greatest number of patients who need to recover for the test in part **c** to be significant at the 1% level. **(3)**

← Sections 6.3, 7.1, 7.2, 7.3

E/P **22** Dhriti grows tomatoes. Over a period of time, she has found that there is a probability 0.3 of a ripe tomato having a diameter greater than 4 cm. Dhriti wants to test whether a new fertiliser increases the size of her tomatoes. She takes a sample of 40 ripe tomatoes that have been treated with the new fertiliser.

a Write down suitable hypotheses for her test. **(1)**

b Using a 5% significance level, find the critical region for her test. **(4)**

c Write down the actual significance level of the test. **(1)**

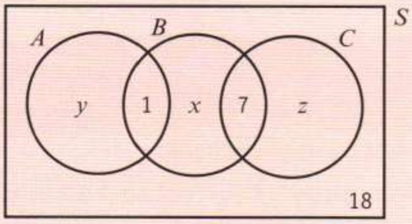
Dhriti finds that 18 out of the 40 tomatoes have a diameter greater than 4 cm.

d Comment on Dhriti's observed value in light of your answer to part **b**. **(2)**

← Sections 7.2, 7.3

Challenge

1 The Venn diagram shows the number of sports club members liking three different sports.



Given that there are 50 members in total, $P(C) = 3P(A)$ and $P(\text{not } B) = 0.76$, find the values of x, y and z. ← Section 5.2

2 A test statistic has binomial distribution $B(30, p)$. Given that $H_0: p = 0.65$, $H_1: p < 0.65$,

a find the critical region for the test statistic such that the probability is as close as possible to 10%.

William takes two observations of the test statistic and finds that they both fall inside the critical region. William decides to reject H_0.

b Find the probability that William has incorrectly rejected H_0. ← Sections 7.2, 7.3