

Populations and Samples

1 (2 marks)

Population: all 1200 students in the school

B1

Sample: the 60 students chosen from the register

B1

Answer: Population: the 1200 students; sample: the 60 students asked

2 marks

2 (2 marks)

Advantage: every member is included, so there is no sampling error and the result is exact for the population

B1

Disadvantage: it is slow and expensive (and impossible if testing destroys the item)

B1

Answer: Advantage: accurate and unbiased; disadvantage: time consuming and costly

2 marks

3 (4 marks)

(a) It describes the whole population, so it is a parameter

B1

(b) It is calculated from a sample, so it is a statistic

B1

(c) It describes the whole population, so it is a parameter

B1

(d) It is calculated from a sample, so it is a statistic

B1

Answer: (a) parameter, (b) statistic, (c) parameter, (d) statistic

4 marks

4 (2 marks)

(a) Census: the group is small so everyone can be asked

B1

(b) Sample: testing destroys the batteries and a census would take too long

B1

Answer: (a) census, (b) sample

2 marks

5 (2 marks)

$$\text{Proportion in sample} = \frac{15}{60} = \frac{1}{4}$$

M1

$$\text{Estimate} = \frac{1}{4} \times 1500 = 375$$

A1

Answer: 375

2 marks

6 (2 marks)

A sampling frame is a list of all the members (sampling units) of the population from which the sample is chosen

B1

For example, the library's list of registered members

B1

Answer: A list of all members of the population, e.g. the membership register

2 marks

7 (3 marks)

(a) Testing a bulb until it fails destroys it

B1

A census would leave no bulbs to sell (and take far too long)

B1

Answer: A census would destroy the whole production

2 marks

(b) A single light bulb

B1

Answer: One light bulb

1 mark

8 (4 marks)

(a) Sampling frame

B1

Answer: Sampling frame

1 mark

(b) It excludes residents under 18 or people who have not registered

B1

Answer: Excludes under 18s / unregistered residents

1 mark

(c) $\bar{x}$ is calculated from a sample, so it is a statistic; μ describes the population, so it is a parameter

B1

Answer: $\bar{x}$ is the statistic; μ is the parameter

1 mark

(d) Different residents are selected, so the sample mean varies from sample to sample

E1

Answer: Sampling variability

1 mark

9 (2 marks)

$$\text{Sum} = 96.5, \text{ so mean} = \frac{96.5}{8}$$

M1

= 12.1 g (3 s.f.); it estimates the mean yolk mass of all the eggs from the farm

A1

Answer: 12.1 g, an estimate of the population mean

2 marks

10 (3 marks)

(a) $2400 \times 12 = 28\,800$ minutes

M1

$$28\,800 \div 60 = 480 \text{ hours}$$

A1

Answer: 480 hours

2 marks

(b) $150 \times 12 \div 60 = 30$ hours

B1

Answer: 30 hours

1 mark

11 (2 marks)

(a) It is calculated from a sample, so it is a statistic

B1

Answer: A statistic

1 mark

(b) Sample means vary by chance, so a mean of 996 g could arise even if the population mean is 1000 g

E1

Answer: Sampling variability

1 mark

12 (3 marks)

$$\frac{8}{50} = \frac{80}{N}$$

M1

$$N = \frac{80 \times 50}{8} = 500$$

A1

Assumption: the tagged deer have mixed randomly with the rest and the population is unchanged

B1

Answer: 500

3 marks

13 (3 marks)

$$\mu = \frac{4 + 7 + 9 + 12}{4} = 8$$

B1

$$\sigma^2 = \frac{16 + 49 + 81 + 144}{4} - 8^2$$

M1

$$= 72.5 - 64 = 8.5$$

A1

Answer: $\mu = 8$, $\sigma^2 = 8.5$

3 marks

14 (7 marks)

(a) $\mu = \frac{2 + 3 + 5 + 8 + 12}{5}$

M1

$$= 6$$

A1

Answer: 6

2 marks

(b) There are 10 possible samples (5 choose 2 = 10)

M1

(2, 3): 2.5, (2, 5): 3.5, (2, 8): 5, (2, 12): 7, (3, 5): 4, (3, 8): 5.5, (3, 12): 7.5, (5, 8): 6.5, (5, 12): 8.5, (8, 12): 10

A1

Answer: 10 samples; means 2.5, 3.5, 5, 7, 4, 5.5, 7.5, 6.5, 8.5, 10

2 marks

(c) Means from 5 to 7 inclusive: 5, 5.5, 6.5, 7, so 4 of 10 samples

M1

$$\text{Probability} = \frac{4}{10} = \frac{2}{5}$$

A1

Answer: $\frac{2}{5}$

2 marks

(d) Sample means vary widely (2.5 to 10) and only some are close to μ , so one sample may not be reliable

E1

Answer: Sample means vary from sample to sample

1 mark

15 (5 marks)

(a) $n = 10: 60 - 41 = 19$

B1

$$n = 100: 52.1 - 49.6 = 2.5$$

B1

Answer: 19 and 2.5

2 marks

(b) Samples of size 100

E1

Their means vary far less, so each is likely to be closer to the population mean

Answer: n = 100, smaller variation

E1

2 marks

(c) $\frac{51.2 + 50.4 + 52.1 + 49.6 + 50.9}{5} = 50.84$, so about 50.8

B1

Answer: 50.8

1 mark

16 (5 marks)

(a) 9 equally likely samples (1,1), (1,2), ..., (3,3)

M1

Sample means and frequencies out of 9 (mean: number of samples): 1: 1, 1.5: 2, 2: 3, 2.5: 2, 3: 1

M1

$P(\bar{X} = 1) = \frac{1}{9}$, $P(1.5) = \frac{2}{9}$, $P(2) = \frac{3}{9}$, $P(2.5) = \frac{2}{9}$, $P(3) = \frac{1}{9}$

A1

Answer: $\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9}$

3 marks

(b) Population mean = $\frac{1 + 2 + 3}{3} = 2$

B1

$E(\bar{X}) = 1(\frac{1}{9}) + 1.5(\frac{2}{9}) + 2(\frac{3}{9}) + 2.5(\frac{2}{9}) + 3(\frac{1}{9}) = \frac{18}{9} = 2$

A1

Answer: Both equal 2

2 marks

17 (6 marks)

(a) There are $C(5,2) = 10$ equally likely samples

M1

$M = k$ occurs in $(k - 1)$ samples, so $P(M = k) = \frac{k - 1}{10}$

M1

$P(M = 2, 3, 4, 5) = \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}$

A1

Answer: $\frac{1}{10}, \frac{1}{5}, \frac{3}{10}, \frac{2}{5}$

3 marks

(b) $E(M) = 2(\frac{1}{10}) + 3(\frac{2}{10}) + 4(\frac{3}{10}) + 5(\frac{4}{10}) = \frac{40}{10} = 4$

M1 A1

The population mean is 3, so M overestimates it on average (it is biased)

E1

Answer: $E(M) = 4$; biased upwards

3 marks

18 (4 marks)

(a) $\frac{5130}{40} = £128.25$

B1

Answer: £128.25

1 mark

(b) 600×128.25

M1

$= £76950$

A1

Answer: £76950

2 marks

(c) A different sample would give a different mean, and a few very large invoices could distort the sample

E1

Answer: Sampling variability

1 mark

19 (4 marks)

(a) As the sample size increases the range decreases, so sample means vary less

E1

Answer: Larger sample, smaller variation

1 mark

(b) $250 + 4n \leq 500$

M1

$n \leq 62.5$, so the largest sample is 62

A1

A sample of 62 is the most reliable the budget allows since larger samples vary less

E1

Answer: 62

3 marks

20 (4 marks)

(a) $0.05 \times 2400 = 120$

B1

Answer: 120

1 mark

(b) $\frac{120}{6} = 20$

B1

Answer: 20

1 mark

(c) The year groups are not equal in size

E1

So larger years are under-represented and smaller years over-represented; the sample should be proportional

E1

Answer: Not proportional to year size**2 marks****21 (6 marks)****(a)** (i) Statistic: it uses only the sample values**B1**(ii) Not a statistic: it contains the unknown μ **B1**

(iii) Statistic: it is found from the sample alone

B1(iv) Not a statistic: it contains the unknown σ **B1****Answer: (i) yes, (ii) no, (iii) yes, (iv) no****4 marks****(b)** A different sample gives a different value of the sample mean, so it varies from sample to sample**E1** μ is a fixed constant describing the whole population**E1****Answer: The sample mean varies with the sample; μ is fixed****2 marks****22 (7 marks)****(a)** $24\,000 \times 3.20 = \text{£}76\,800$ **B1****Answer: £76 800****1 mark****(b)** One third of 76 800 = 25 600**M1** $1600 + 3.2n = 25\,600$ **M1**

$$n = \frac{24\,000}{3.2} = 7500$$

A1**Answer: 7500****3 marks****(c)** Proportion = $\frac{4125}{7500} = 0.55$ **M1** $0.55 \times 24\,000 = 13\,200$ **A1****Answer: 13 200****2 marks****(d)** A different sample of the same size would give a different proportion (sampling variability)**E1****Answer: Sampling variability****1 mark****23 (6 marks)****(a)** Total samples = $C(8,3) = 56$ **B1**Samples containing the person = $C(7,2) = 21$ **M1**

$$\text{Probability} = \frac{21}{56} = \frac{3}{8}$$

A1**Answer: Shown****3 marks****(b)** At least 2 girls: 2 girls and 1 boy, or 3 girls**M1** $C(5,2) \times C(3,1) + C(5,3) = 30 + 10 = 40$ **M1**

$$\text{Probability} = \frac{40}{56} = \frac{5}{7}$$

A1**Answer: $\frac{5}{7}$** **3 marks****24 (8 marks)****(a)** Each person is equally likely to be in the sample, so the probability is $\frac{4}{10}$ **M1**

$$= \frac{2}{5}$$

A1**Answer: $\frac{2}{5}$** **2 marks****(b)** $P(\text{Sam is not chosen on one selection}) = \frac{9}{10}$ **B1** $P(\text{Sam is never chosen}) = \left(\frac{9}{10}\right)^4 = 0.6561$ **M1** $P(\text{at least once}) = 1 - 0.6561 = 0.3439$ **A1****Answer: 0.3439****3 marks**

$$\text{(c)} \frac{10 \times 9 \times 8 \times 7}{10^4}$$

M1

$$= \frac{5040}{10\,000} = 0.504$$

A1**Answer: 0.504****2 marks**

(d) With replacement, nearly half of the samples of four would contain a repeated person, so the sample would give less information about the population	E1
Answer: Repeats waste sample information	1 mark
25 (5 marks)	
(a) $52 - 47 = 5$	B1
Answer: 5	1 mark
(b) All polls have equal size so the mean is used	M1
$\frac{48 + 52 + 51 + 47 + 50}{5} = 49.6\%$	A1
Answer: 49.6%	2 marks
(c) The other polls give figures below 50%, so a single poll can be misleading	E1
Poll results vary from sample to sample, so the combined result of about 49.6% suggests no clear majority	E1
Answer: Not justified	2 marks
26 (5 marks)	
(a) Samples: $C(6,3) = 20$	B1
Sum ≥ 12 needed. Triples: (1,5,6), (2,4,6), (2,5,6), (3,4,5), (3,4,6), (3,5,6), (4,5,6)	M1
Probability = $\frac{7}{20} = \frac{7}{20}$	A1
Answer: $\frac{7}{20}$	3 marks
(b) Each value appears in $C(5,2) = 10$ of the 20 samples, so the total of all sample sums is $10 \times 21 = 210$	M1
Mean of sample means = $\frac{210}{20 \times 3} = 3.5 =$ population mean $\frac{21}{6}$	A1
Answer: Shown	2 marks
27 (9 marks)	
(a) Proportion tagged in the sample = $\frac{18}{120} = 0.15$	M1
Assume this equals the proportion tagged in the lake: $\frac{150}{N} = 0.15$	M1
$N = 1000$	A1
Answer: 1000	3 marks
(b) The number of fish in the lake is constant (no births, deaths or migration)	B1
Tagged fish mix evenly with the others, and tags are not lost	B1
Answer: Constant population; tagged fish well mixed	2 marks
(c) More tagged fish are caught than expected, so the proportion tagged is too large	E1
$N = \frac{150}{\text{proportion}}$ is therefore too small: the lake is underestimated	E1
Answer: Underestimate	2 marks
(d) Tagged in total: $18 + 5 = 23$ out of 150 fish, so $\frac{23}{150} = \frac{150}{N}$	M1
$N = \frac{150 \times 150}{23} = 978.26\dots$, so about 978	A1
Answer: 978	2 marks
28 (8 marks)	
(a) Not correct	E1
How the sample is chosen matters more than its size	E1
A badly chosen 10% sample (for example only people at a golf club) is biased, whereas a random 5% sample is not	E1
Answer: Not correct: selection method matters	3 marks
(b) Not correct	E1
Bias is a feature of the method of selection, not of one result	E1
Random samples vary, so a sample mean differs from μ by chance even with an unbiased method	E1
Answer: Not correct: this is sampling variability	3 marks

(c) Not correct

E1

Different members are chosen, so the sample means will usually differ

E1

Answer: Not correct

2 marks

29 (10 marks)

(a) $C(6,3) = 20$

B1

The median is the middle value; $M = k$ needs $(k - 1)$ smaller and $(6 - k)$ larger values:

M1

$(k-1)(6-k)$ samples

$k = 2: 1 \times 4 = 4, k = 3: 2 \times 3 = 6, k = 4: 3 \times 2 = 6, k = 5: 4 \times 1 = 4$

M1

$P(M = 2, 3, 4, 5) = \frac{1}{5}, \frac{3}{10}, \frac{3}{10}, \frac{1}{5}$

A1

Answer: $\frac{1}{5}, \frac{3}{10}, \frac{3}{10}, \frac{1}{5}$

4 marks

(b) $E(M) = 3.5$ by symmetry

B1

$E(M^2) = 4(\frac{1}{5}) + 9(\frac{3}{10}) + 16(\frac{3}{10}) + 25(\frac{1}{5}) = 13.3$

M1

$\text{Var}(M) = 13.3 - 3.5^2 = 1.05$

A1

Answer: 1.05

3 marks

(c) Both have expectation 3.5 (unbiased)

B1

$\text{Var}(\bar{X}) = 0.583333$ is smaller than $\text{Var}(M) = 1.05$

M1

So $\bar{X}$ varies less from sample to sample and is the better estimator

E1

Answer: The sample mean

3 marks

30 (7 marks)

(a) Mean = $\frac{14 + 52 + 33 + 71 + 40}{5} = 42$

M1

$2 \times 42 - 1 = 83$

A1

Answer: 83

2 marks

(b) $m = 71, n = 5$

B1

$71 + \frac{71}{5} - 1 = 84.2$

A1

Answer: 84.2

2 marks

(c) The largest number observed can never exceed N , so using it alone underestimates N

E1

Anna's method can give a value below an observed number (for example 1, 2, 3, 4, 50 gives $2 \times$

E1

$12 - 1 = 23 < 50$), which is impossible

Ben's estimate is always above the largest observed number, so it is preferred

E1

Answer: Ben's method

3 marks

Sampling Methods

1 (2 marks)

Number the members 1 to 80 and generate random numbers between 1 and 80 (ignoring repeats)

B1

Select the 5 members whose numbers appear

B1

Answer: Number the members; choose 5 different random numbers

2 marks

2 (3 marks)

(a) $\frac{600}{40} = 15$

B1

Answer: 15

1 mark

(b) $7 + 15 = 22$, then 37

B1

Answer: 22 and 37

1 mark

(c) $7 + 39 \times 15 = 592$

B1

Answer: 592

1 mark

3 (3 marks)

- (a) Systematic sampling B1
 (b) Simple random sampling B1
 (c) Quota sampling B1

Answer: (a) systematic (b) simple random (c) quota 3 marks

4 (2 marks)

Girls: $\frac{500}{800} \times 80 = 50$ M1

Boys: $\frac{300}{800} \times 80 = 30$ A1

Answer: 50 girls and 30 boys 2 marks

5 (2 marks)

The population is divided into groups (strata), and a random sample is taken from each group in proportion to its size B1

It is better when the groups are very different in size or character, so that every group is sure to be represented B1

Answer: Random sample from each stratum in proportion to its size; better when the strata differ 2 marks

6 (2 marks)

Advantage: quick and cheap, and no sampling frame is needed B1

Disadvantage: not random, so interviewers' choices can introduce bias and sampling error cannot be estimated B1

Answer: Quick and cheap; not random so may be biased 2 marks

7 (3 marks)

Sampling fraction = $\frac{120}{1200} = \frac{1}{10}$ M1

Years 7, 8, 9: 27, 25, 24 students A1

Years 10, 11: 23, 21 students (total 120) A1

Answer: 27, 25, 24, 23, 21 3 marks

8 (4 marks)

(a) Quota sampling B1

Answer: Quota sampling 1 mark

(b) Quick and cheap, and no sampling frame is needed B1

Answer: Quick, cheap, no sampling frame 1 mark

(c) Not random, so the interviewer's choices can introduce bias B1

Sampling error cannot be estimated, and non-response is not recorded B1

Answer: Not random; interviewer bias 2 marks

9 (3 marks)

Reject 83 as it is bigger than 60 B1

Reject the second 12 as it is a repeat B1

Members chosen: 47, 12, 05, 59 B1

Answer: 47, 12, 05, 59 3 marks

10 (2 marks)

(a) Opportunity (convenience) sampling B1

Answer: Opportunity sampling 1 mark

(b) People at the library are likely to study more than average, so the sample is unrepresentative E1

Answer: Not representative 1 mark

11 (4 marks)

(a) Year 12: $\frac{180}{300} \times 30 = 18$

M1

Year 13: $\frac{120}{300} \times 30 = 12$

A1

Answer: 18 from Year 12 and 12 from Year 13

2 marks

(b) Number the Year 13 students 1 to 120 from a list of the year group

B1

Use a random number generator (or table) to choose 12 different numbers and take the students with those numbers

B1

Answer: Number them 1 to 120 and choose 12 at random

2 marks

12 (3 marks)

Men: $0.48 \times 100 = 48$

B1

Under 30: $0.25 \times 100 = 25$

B1

e.g. employment status / ethnicity / social class

B1

Answer: 48 and 25; another factor such as occupation

3 marks

13 (3 marks)

Self-selected sampling

B1

Only people who listen to the station, and who choose to text, take part

E1

These are unlikely to be typical of all the residents

E1

Answer: Self-selected; unrepresentative

3 marks

14 (6 marks)

(a) Systematic sampling (interval 20, random start)

B1

A list exists and this is quick without needing 120 random numbers

B1

(b) Stratified sampling

B1

Each group is sampled in proportion to its size

B1

(c) Quota sampling

B1

No sampling frame is needed and the quotas can be filled quickly

B1

Answer: (a) systematic; (b) stratified; (c) quota

6 marks

15 (6 marks)

(a) Sampling fraction = $\frac{55}{440} = \frac{1}{8}$

M1

Admin 5, Sales 15

A1

Production 25, Research 10

A1

Answer: 5, 15, 25, 10

3 marks

(b) List the 80 employees, number them and pick 10 using random numbers

B1

Answer: Simple random sample of 10 from 80

1 mark

(c) Every department is guaranteed to be represented

E1

Each department is represented in proportion to its size, so the sample is more representative

E1

Answer: Guaranteed representation in proportion

2 marks

16 (5 marks)

(a) Probability = $\frac{C(380,10)}{C(400,10)}$

M1 M1

= 0.595

A1

Answer: 0.595

3 marks

(b) Stratified sampling takes a fixed number from each group in proportion to its size

E1

So the sixth form is always represented, with $\frac{20}{400} \times$ sample size students

E1

Answer: Guarantees each group is represented

2 marks

17 (2 marks)

Every sampled day would be the same day of the week

E1

Takings follow a weekly pattern (for example busier at weekends), so the sample is not representative

E1

Answer: Always the same weekday

2 marks

18 (5 marks)

(a) Sampling fraction = $\frac{40}{1000}$

M1

18, 14, 8

A1

Answer: 18, 14, 8

3 marks

(b) $\frac{25}{1000} \times 450 = 11.25, \times 350 = 8.75, \times 200 = 5$

M1

Rounding gives 11, 9, 5 which totals 25

A1

The exact values are not whole numbers, so they must be rounded and the proportions are only approximate

E1

Answer: 11, 9, 5; rounding needed

3 marks

19 (7 marks)

(a) Under 35: $\frac{5400}{10\,000} \times 200 = 108$

M1

35 to 59: $\frac{3600}{10\,000} \times 200 = 72$

A1

60 or over: $\frac{1000}{10\,000} \times 200 = 20$

A1

Answer: 108, 72 and 20

3 marks

(b) In a stratified sample the individuals are chosen at random from a sampling frame; in a quota sample the interviewer chooses whoever is available

B1

A quota sample needs no sampling frame and can be done quickly, so it is cheaper, but it is not random

B1

Answer: Random with a frame (stratified) against interviewer's choice, no frame (quota)

2 marks

(c) Within each quota the people are not chosen at random

B1

Interviewers may approach people who look friendly or who are in the same place at the same time, so some people are less likely to be chosen

B1

Answer: Selection within each quota is not random

2 marks

20 (4 marks)

Make a list (sampling frame) of employees and number them 001 to 340

B1

Read three-digit random numbers from the table in order

M1

Ignore numbers outside 001 to 340 and ignore repeats

M1

Continue until 12 different numbers are obtained and select those employees

A1

Answer: Number 001 to 340, use random 3-digit numbers, reject out of range and repeats, until 12 4 marks**21 (8 marks)**

(a) Interval = $\frac{1000}{25} = 40$

B1

Choose a random starting number from 1 to 40 and then every 40th customer after it

B1

Answer: Interval 40, random start from 1 to 40

2 marks

(b) $P(\text{both}) = \frac{25}{1000} \times \frac{24}{999}$

M1

$= \frac{1}{1665}$

A1

Answer: $\frac{1}{1665}$

2 marks

(c) 0, since selected customers are 40 apart

B1

Systematic sampling does not give every possible sample an equal chance, unlike simple random sampling

E1

Answer: 0

2 marks

(d) Every 8th product is also a multiple of 4

E1

So the inspections either all find a fault or all miss them, depending on the start, which is a periodic pattern matching the interval

E1

Answer: Interval matches a pattern in the process

2 marks

22 (8 marks)

(a) $\frac{40}{90} = \frac{2000}{N}$

M1

$N = 4500$

A1

Answer: 4500

2 marks

(b) A: $\frac{18}{90} \times 4500 = 900$

A1

C: $\frac{32}{90} \times 4500 = 1600$

A1

Answer: 900 and 1600

2 marks

(c) Weight each stratum mean by its population

M1

$$\frac{900 \times 24.5 + 2000 \times 31.2 + 1600 \times 28.6}{4500}$$

M1

$= 28.9$ (3 s.f.)

A1

Answer: £28.9

3 marks

(d) The sample sizes are proportional to the strata sizes, so each sampled person has equal weight

E1

Answer: Sample is proportional to stratum size

1 mark

23 (7 marks)

(a) Total staff = 400, sampling fraction $\frac{50}{400} = \frac{1}{8}$

B1

Sales: $\frac{100}{8} = 12.5$, Admin: $\frac{150}{8} = 18.75$, Tech: $\frac{150}{8} = 18.75$

M1

Rounding sensibly to total 50: Sales 12, Admin 19, Tech 19

A1

Answer: 12 (or 13), 19, 19

3 marks

(b) Women: $\frac{30}{150} \times 19 = 3.8 \approx 4$

M1

Men: $19 - 4 = 15$

A1

Answer: 4 women, 15 men

2 marks

(c) Numbers of people must be whole

E1

Select at random, e.g. toss a coin

E1

Answer: Whole numbers; choose randomly

2 marks

24 (6 marks)

(a) Interval = 4, four possible starts, each equally likely

M1

Each member is in the sample for exactly one start, so probability $\frac{1}{4}$

A1

Answer: Shown

2 marks

(b) Selected members are 4 apart, so consecutive members cannot both be chosen

E1

Answer: Members are 4 apart

1 mark

(c) Samples containing both = $C(18,3) = 816$, total = $C(20,5) = 15\,504$

M1

Probability = $\frac{816}{15504} = \frac{1}{19}$

A1

A simple random sample allows adjacent members together; systematic sampling does not

E1

Answer: $\frac{1}{19}$

3 marks

25 (5 marks)

(a) $\frac{900}{10} = 90$

B1

Answer: 90

1 mark

(b) In a simple random sample every possible sample of 90 is equally likely

E1

Here only 10 possible samples (one per start) can occur, so most groups of 90 members cannot be chosen

E1

Answer: Only 10 possible samples

2 marks

(c) Satisfactory when the list order is unrelated to the variable, e.g. alphabetical order and opinions

B1

Not satisfactory if the list has a pattern matching the interval, e.g. couples listed together with the interval 2

B1

Answer: See solution

2 marks

26 (6 marks)

(a) $P(\text{school chosen}) = \frac{3}{12}$ M1

$P = \frac{3}{12} \times \frac{20}{400} = \frac{1}{80}$ A1

Answer: $\frac{1}{80}$ 2 marks

(b) $\frac{3}{12} \times \frac{20}{100} = \frac{1}{20}$ B1

Answer: $\frac{1}{20}$ 1 mark

(c) Students at school B are 4 times as likely to be chosen as those at school A E1

So not every student has the same chance E1

Choose the number of students from each school in proportion to its size (stratified) E1

Answer: Unequal chances; use proportional numbers 3 marks

27 (9 marks)

(a) Number of samples = $C(10, 5) = 252$ M1

Favourable = $C(6, 3) \times C(4, 2) = 20 \times 6 = 120$ M1

Probability = $\frac{120}{252} = \frac{10}{21}$ A1

Answer: $\frac{10}{21}$ 3 marks

(b) 1 B1

Strata sizes are in the ratio 6 : 4, so the sample must contain 3 women and 2 men E1

Answer: 1 2 marks

(c) Simple random: $\frac{5}{10} = \frac{1}{2}$ B1

Stratified, a woman: $\frac{3}{6} = \frac{1}{2}$ B1

Stratified, a man: $\frac{2}{4} = \frac{1}{2}$ B1

Answer: $\frac{1}{2}$ in both cases 3 marks

(d) Although individuals have equal chances, the stratified sample is certain to have the correct E1

mix, whereas a simple random sample has this only $\frac{10}{21}$ of the time

Answer: Guarantees correct composition 1 mark

28 (9 marks)

(a) $\frac{150}{3000} = \frac{1}{20}$ M1

60, 45, 30, 15 A1

Answer: 60, 45, 30, 15 2 marks

(b) Interval = $\frac{1200}{60} = 20$ B1

Pick a random start between 1 and 20, then every 20th name B1

Answer: Interval 20, random start 2 marks

(c) People leaving a library are more likely to study a lot (not representative) E1

Faculties are not represented in proportion and the choice is not random E1

Answer: Biased towards keen students; not random 2 marks

(d) Weights $\frac{1200}{3000}, \frac{900}{3000}, \frac{600}{3000}, \frac{300}{3000} = 0.4, 0.3, 0.2, 0.1$ M1

$0.4 \times 14.2 + 0.3 \times 11.5 + 0.2 \times 9.8 + 0.1 \times 16.4 = 12.73$ hours A1

Answer: 12.73 hours 2 marks

(e) It gives the small Law faculty the same weight as the large Science faculty E1

Answer: Ignores the different faculty sizes 1 mark

29 (8 marks)

(a) Sampling fraction = $\frac{60}{2000} = 0.03$

M1

$0.03 \times 1200 = 36$, $0.03 \times 600 = 18$, $0.03 \times 200 = 6$

A1

Answer: Verified

2 marks

(b) Overall sample mean = $\frac{36 \times 52 + 18 \times 64 + 6 \times 80}{60}$

M1

= $\frac{1872 + 1152 + 480}{60}$

M1

= 58.40

A1

Answer: 58.40

3 marks

(c) It gives each stratum equal weight although the strata differ greatly in size

E1

$\frac{52 + 64 + 80}{3} = 65.33$

M1

Difference = $65.33 - 58.40 = 6.93$

A1

Answer: Overstates by 6.93

3 marks

30 (9 marks)

(a) Samples containing a given unit: C(59,9); total C(60,10)

M1

$\frac{C(59,9)}{C(60,10)} = \frac{10}{60}$

A1

Samples containing two given units: C(58,8), and $\frac{C(58,8)}{C(60,10)} = \frac{10 \times 9}{60 \times 59}$

M1

= $\frac{3}{118}$

A1

Answer: $\frac{3}{118}$

4 marks

(b) Units 1 and 7 differ by 6, so they are both chosen exactly when the start is 1: probability $\frac{1}{6}$

M1 A1

Units 1 and 2 cannot both be selected: probability 0

B1

Answer: $\frac{1}{6}$ and 0

3 marks

(c) In a simple random sample the chance two given units are chosen is the same for every pair

E1

In a systematic sample it depends on the gap between the units (0 or $\frac{1}{6}$ here), so the sample is not equivalent to a simple random sample

E1

Answer: Pair probabilities differ

2 marks

Sampling Bias**1 (2 marks)**

(a) A sample that does not fairly represent the population, so it tends to give results that are systematically wrong in one direction

E1

Answer: A sample that systematically misrepresents the population

1 mark

(b) Early arrivals may be more likely to buy snacks or be regular customers, and they are not randomly chosen

E1

Answer: Not random; early arrivals differ from other customers

1 mark

2 (3 marks)

(a) Self-selection: only people with strong views phone in

B1

(b) The sample is made up of people who already exercise (wrong place and time)

B1

(c) The question is leading

B1

Answer: (a) self-selection; (b) unrepresentative location; (c) leading question

3 marks

3 (4 marks)

- (a) It is a leading question: it suggests that the service was excellent **B1**
 Neutral version: "How would you rate our service?" with a range of options **B1**
 (b) It asks two questions at once, so a single answer cannot be interpreted **B1**
 Better: ask "Do you eat healthily?" and "Do you exercise regularly?" separately **B1**
Answer: (a) leading question; (b) two questions in one **4 marks**

4 (1 mark)

- Its customers are not typical of the whole population, as they are more likely to eat fast food **E1**
Answer: Customers are unrepresentative **1 mark**

5 (2 marks)

- The question is leading: it suggests the answer the interviewer expects **E1**
 Neutral version: "Do you think the new bypass is good value for money?" with options such as yes, no, don't know **B1**
Answer: Leading question; rewrite neutrally **2 marks**

6 (2 marks)

- $\frac{50}{200} \times 100 = 25\%$ **B1**
 People who reply may have stronger views than those who do not **E1**
Answer: 25% **2 marks**

7 (4 marks)

- (a) It is a leading question that pushes the respondent towards answering yes **E1**
Answer: Leading question **1 mark**
 (b) For example: "Should council tax be reduced, kept the same or increased?" **B1**
Answer: A neutral question with balanced options **1 mark**
 (c) $\frac{80}{500} = 16\%$ **B1**
Answer: 16% **1 mark**
 (d) Those who reply may have stronger opinions than those who do not, so the replies may not represent all residents **E1**
Answer: Non-response bias **1 mark**

8 (4 marks)

- (a) Not everyone is listed (mobile-only households and those who opt out are missing) **B1**
 Those listed tend to be older, so the frame under-represents young people who spend more time online **B1**
Answer: Frame excludes many people, especially younger ones **2 marks**
 (b) Only people who are online and visit that website can respond (frame bias) **B1**
 Online shoppers are less likely to be town-centre shoppers; respondents choose themselves (self-selection) **B1**
Answer: Excludes non-internet users; self-selected and unrepresentative **2 marks**

9 (2 marks)

- Households without a landline or with unlisted numbers are excluded **E1**
 Those without landlines (for example younger people) probably shop online differently **E1**
Answer: Excludes people with no landline / unlisted numbers **2 marks**

10 (2 marks)

- At 10 am many commuters will already have travelled; those present are likely to have different working hours **E1**
 Only people who use the station are asked, so drivers are excluded **E1**
Answer: Time of day; only rail users **2 marks**

11 (3 marks)

(a) $\frac{180}{240} = 75\%$

B1

Answer: 75%

1 mark

(b) Self-selected sample: only those who visit and choose to vote are included

E1

They are likely to hold stronger views than the population as a whole

E1

Answer: Self-selection bias

2 marks

12 (2 marks)

The options overlap: £5, £10 and £15 each belong to two groups

E1

Group D has no upper limit, so the detail of high spenders is lost (and no option for people who bring lunch)

E1

Answer: Overlapping options

2 marks

13 (2 marks)

It is a systematic error (bias) in the measuring instrument

E1

True mean estimate = $54.3 - 0.5 = 53.8$ kg

B1

Answer: 53.8 kg

2 marks

14 (6 marks)

(a) Response rate = $\frac{240}{1500} = 16\%$

B1

In favour = $\frac{168}{240} = 70\%$

B1

Answer: 16% and 70%

2 marks

(b) The 84% who did not reply may hold different views from those who did (non-response bias)

E1

Answer: Non-response bias

1 mark

(c) Lowest: none of the 1260 non-respondents in favour: $\frac{168}{1500} = 11.2\%$

M1 A1

Highest: all non-respondents in favour: $\frac{168 + 1260}{1500} = 95.2\%$

A1

Answer: 11.2% and 95.2%

3 marks

15 (6 marks)

(a) $\frac{18 + 28}{100} = 0.46$

B1

Answer: 0.46

1 mark

(b) Girls are 60% of the school but only 30% of the sample

E1

Answer: Girls are under-represented

1 mark

(c) Girls: $\frac{18}{30} = 0.6$ in favour, boys: $\frac{28}{70} = 0.4$ in favour

M1

Weights: girls $\frac{540}{900} = 0.6$, boys 0.4

M1

$0.6 \times 0.6 + 0.4 \times 0.4 = 0.52$

A1

Answer: 0.52

3 marks

(d) The raw sample suggests less than half are in favour, but the corrected estimate shows a small majority

E1

Answer: Conclusion reverses

1 mark

16 (5 marks)

(a) $\frac{120}{340} = 0.353$

B1

Answer: 0.353

1 mark

(b) Lowest: none of the 60 non-respondents would say yes: $\frac{120}{400} = 0.3$

M1 A1

Highest: all 60 would say yes: $\frac{180}{400} = 0.45$

A1

Answer: 0.3 to 0.45

3 marks

(c) Non-response means the true proportion could lie between 30% and 45%, so the estimate is uncertain

E1

Answer: Unreliable**1 mark****17 (5 marks)**

(a) $0.6 \times 70 + 0.4 \times 30 = 42 + 12 = 54$

M1

54%

A1**Answer: 54%****2 marks**

(b) Weight men and women equally: $0.5 \times 60 + 0.5 \times 40$

M1 $= 50\%$ **A1****Answer: 50%****2 marks****(c)** Women are over-represented and women are more likely to support the proposal**E1****Answer: Women over-represented****1 mark****18 (6 marks)**

(a) 30, 35, 35

B1**Answer: 30, 35, 35****1 mark****(b)** 55+ over-represented (58 against 35); 18 to 34 under-represented (12 against 30)**B1**

Possibly the survey was by post/phone during the day, which younger people do not answer

E1**Answer: 55+ over; 18-34 under****2 marks**

(c) Sample: $0.25 \times 12 + 0.5 \times 30 + 0.7 \times 58 = 3 + 15 + 40.6 = 58.6\%$

M1 A1

Weighted: $0.30 \times 25 + 0.35 \times 50 + 0.35 \times 70 = 7.5 + 17.5 + 24.5 = 49.5\%$

A1**Answer: 58.6% and 49.5%****3 marks****19 (5 marks)**

(a) Sum = 506.6 so mean = 50.66

M1 A1

Bias = $50.66 - 50.0 = 0.66$ g

A1**Answer: 0.66 g too high****3 marks**

(b) $72.15 - 0.66 = 71.49$ g

B1

More readings reduce random variation but the systematic error is the same in every reading

E1**Answer: 71.49 g****2 marks****20 (4 marks)****(a)** C**B1****(b)** D**B1****(c)** A**B1****(d)** B**B1****Answer: (a) C (b) D (c) A (d) B****4 marks****21 (8 marks)**

(a) $\frac{10 + 20 + 90}{3}$

M1 $= 40$ **A1****Answer: 40****2 marks**

(b) $P(X = 10) = \frac{10}{120} = \frac{1}{12}$, $P(X = 20) = \frac{20}{120} = \frac{1}{6}$

M1

$P(X = 90) = \frac{90}{120} = \frac{3}{4}$

A1

$E(X) = 10 \times \frac{1}{12} + 20 \times \frac{1}{6} + 90 \times \frac{3}{4}$

M1 $= 71.7$ (3 s.f.)**A1****Answer: 71.7****4 marks****(c)** Large classes contain more students, so a randomly chosen student is much more likely to come from a large class**E1****Answer: Larger classes are more likely to be selected****1 mark****(d)** Choose classes (not students) at random, so each class has the same chance of selection**E1****Answer: Sample the classes, not the students****1 mark**

22 (7 marks)**(a)** Non-respondents: $400 - 260 = 140$

M1

$$\frac{260 \times 34.50 + 140 \times 21.30}{400}$$

M1

 $= £29.88$

A1

Answer: £29.88

3 marks

(b) $34.50 - 29.88 = 4.62$

M1

The replies overestimate the mean by £4.62

A1

Answer: Overestimate by £4.62

2 marks

(c) Only 40 of the 140 non-respondents were contacted, and they may be unrepresentative of all of them

E1

The 40 customers are a sample, so the true mean for non-respondents differs by chance

E1

Answer: Based on a small sample of non-respondents

2 marks

23 (6 marks)**(a)** Plan 1: $\frac{31}{62} = 0.5$

B1

Plan 2: $\frac{328}{400} = 0.82$

B1

Answer: 0.5 and 0.82

2 marks

(b) Plan 1

B1

It is a random sample, so every register member has an equal chance of selection

E1

Plan 2 is self-selected, so people with strong views (cyclists) are over-represented even though the sample is larger

E1

Answer: Plan 1

3 marks

(c) A response rate of 62% means non-respondents may differ from respondents

E1

Answer: Non-response

1 mark

24 (6 marks)**(a)** $\frac{360}{600} = 0.6$

B1

 $\frac{10}{40} = 0.25$

B1

Answer: 0.6 and 0.25

2 marks

(b) Respondents: 360

B1

Non-respondents: $0.25 \times 400 = 100$

M1

Total = 460, so proportion 0.46

A1

Answer: 460 households (46%)

3 marks

(c) It ignored the 400 non-respondents, who are less supportive; non-response bias

E1

Answer: Non-response bias

1 mark

25 (7 marks)**(a)** Post: $\frac{240}{0.30} = 800$

M1

Hand: $\frac{240}{0.75} = 320$

A1

Answer: 800 by post, 320 by hand

2 marks

(b) Post: $800 \times 0.80 = £640$

B1

Hand: $320 \times 3 = £960$

B1

Answer: £640 and £960

2 marks

(c) Higher response rate (75%) means less non-response bias

B1

It costs £320 more, and the extra accuracy may not matter if the charity only needs a rough idea

B1

Answer: Less non-response bias; £320 extra

2 marks

(d) The people who do not reply may differ from those who do, however many are sent

E1

Answer: Non-respondents may differ from respondents

1 mark

26 (5 marks)

(a) $0.8 \times 70 + 0.2 \times 40$ M1

$= 56 + 8 = 64\%$ A1

Answer: 64% 2 marks

(b) Telephone survey only reaches landlines: 70% B1

Bias = $70 - 64 = 6$ percentage points B1

Answer: 70%; bias 6 percentage points 2 marks

(c) Include mobile numbers (or a mixed method such as post or door to door) B1

Answer: Sample mobile numbers too 1 mark**27 (10 marks)**

(a) Students may lie if asked directly, because of embarrassment or fear of punishment E1

Since nobody knows whether a "yes" came from the coin, answers are protected and students are more likely to be honest E1

Answer: Protects privacy so answers are more honest 2 marks

(b) $P(\text{yes}) = P(\text{heads}) \times 1 + P(\text{tails}) \times p$ M1

$= \frac{1}{2} + \frac{1}{2}p$ A1

Answer: Shown 2 marks

(c) $\frac{130}{200} = 0.65 = \frac{1}{2} + \frac{1}{2}p$ M1

$p = 0.3$ A1

Answer: 0.3 2 marks

(d) $P(\text{yes}) = 0.6 + 0.4p = 0.65$ M1

$p = \frac{1}{8}$ A1

Answer: $\frac{1}{8}$ 2 marks

(e) The observed proportion is divided by 0.4 (rather than 0.5) to obtain p, so any sampling error in it is magnified more E1

Fewer students (40% instead of 50%) give information about p E1

Answer: Fewer students answer truthfully, so more variability in the estimate of p 2 marks**28 (10 marks)**

(a) Location: people at a gym are far more likely to exercise E1

Time and day: only those attending at one time are included E1

The sample (40) is small and not random: people are chosen by convenience E1

She generalises to the whole town although the sample comes from a different population (gym users) E1

Answer: Convenience sample of gym users; small; not random; wrong population 4 marks

(b) $\frac{400}{60\,000} = \frac{1}{150}$ M1

140, 160, 100 A1

Answer: 140, 160, 100 2 marks

(c) Use the electoral register or another list of the group's adults as a sampling frame B1

Select the required number using random numbers (simple random sampling) B1

Answer: Random selection from a list of the group 2 marks

(d) Total = $56 + 48 + 22 = 126$ out of 400 M1

$\frac{126}{400} = 31.5\%$ A1

Answer: 31.5% 2 marks**29 (9 marks)**

(a) Mean = 2 B1

Variance = $\frac{(1-2)^2 + 0 + (3-2)^2}{3} = \frac{2}{3}$ M1

Answer: Shown 2 marks

(b) Samples (1, 1), (2, 2), (3, 3) have variance 0 B1

Samples (1, 2), (2, 1), (2, 3), (3, 2): mean 1.5 or 2.5, deviations ± 0.5 , variance $\frac{0.25 + 0.25}{2} = 0.25$ **M1**

Samples (1, 3) and (3, 1): mean 2, variance $\frac{1+1}{2} = 1$ **A1**

Answer: 0 (three times), 0.25 (four times), 1 (twice) **3 marks**

(c) Mean = $\frac{0 \times 3 + 0.25 \times 4 + 1 \times 2}{9} = \frac{3}{9} = \frac{1}{3}$ **M1 A1**

This is only half of the population variance $\frac{2}{3}$, so on average the sample variance is too small **E1**

Answer: $\frac{1}{3}$; too small on average **3 marks**

(d) Multiplying by 2 gives a mean of $\frac{2}{3}$, which equals the population variance **B1**

Answer: Mean becomes $\frac{2}{3}$ **1 mark**

30 (9 marks)

(a) Weight by respondents: $14 \times 20 + 30 \times 50 + 56 \times 30 = 280 + 1500 + 1680 = 3460$ **M1 A1**

$\frac{3460}{100} = 34.6\%$ **A1**

Answer: 34.6% **3 marks**

(b) $0.40 \times 14 + 0.40 \times 30 + 0.20 \times 56$ **M1**

$= 5.6 + 12 + 11.2 = 28.8\%$ **A1**

Answer: 28.8% **2 marks**

(c) $34.6 - 28.8 = 5.8$ percentage points **M1**

Older respondents are over-represented, and they use the app more **E1**

Answer: 5.8 percentage points too high **2 marks**

(d) Stratify by age group **E1**

Take 40, 40 and 20 respondents from the three groups, in proportion to population size **E1**

Answer: Stratified sample 40:40:20 **2 marks**

Measures of Location and Spread

1 (3 marks)

Mean = $\frac{63}{9} = 7$ **B1**

Ordered: 3, 4, 5, 6, 8, 8, 8, 9, 12, so median = 8 **B1**

Mode = 8 **B1**

Answer: Mean 7, median 8, mode 8 **3 marks**

2 (3 marks)

Range = $35 - 12 = 23$ **B1**

$Q_1 = \frac{15 + 17}{2} = 16$ and $Q_3 = \frac{24 + 29}{2} = 26.5$ **B1**

IQR = $26.5 - 16 = 10.5$ **B1**

Answer: Range 23, IQR 10.5 **3 marks**

3 (3 marks)

15 values, so the median is the 8th value. Counting along the rows (3, then 5, then 6 values), the 8th value is 29 **B1**

Mode: 35 occurs three times **B1**

Range = $41 - 12 = 29$ **B1**

Answer: median 29, mode 35, range 29 **3 marks**

4 (3 marks)

Mode = 5 (highest frequency, 11)

B1

Range = $7 - 3 = 4$

B1

Median is the $\frac{15^{\text{th}} + 16^{\text{th}}}{2}$ value, both in size 5 (cumulative 2, 9, 20), so 5

B1

Answer: mode 5, range 4, median 5

3 marks

5 (3 marks)

Mean = 12

B1

Variance = $\frac{1580}{10} - 12^2 = 158 - 144 = 14$

M1

Standard deviation = $\sqrt{14} = 3.74$

A1

Answer: $\sigma^2 = 14$, $\sigma = 3.74$

3 marks

6 (2 marks) $\Sigma f x = 0 + 9 + 24 + 24 + 8 = 65$, $n = 35$

M1

Mean = $\frac{65}{35} = 1.86$

A1

Answer: 1.86

2 marks

7 (4 marks) $\Sigma f x = 0 + 9 + 24 + 24 + 20 + 10 = 87$

M1

Mean = $\frac{87}{40} = 2.175$

A1

Cumulative frequencies 4, 13, 25: the 20th and 21st values are both 2, so median = 2

B1

Mode = 2 (highest frequency 12)

B1

Answer: Mean 2.175, median 2, mode 2

4 marks

8 (5 marks)(a) Midpoints 5, 15, 25, 35, 50: $\Sigma f x = 30 + 210 + 550 + 385 + 350 = 1525$

M1

Mean = $\frac{1525}{60} = 25.4$ min

A1

Answer: 25.4 minutes

2 marks

(b) Cumulative frequencies 6, 20, 42: the 30th value lies in 20-30

B1

Answer: $20 \leq t < 30$

1 mark

(c) $\Sigma f x^2 = 150 + 3150 + 13\,750 + 13\,475 + 17\,500 = 48\,025$

M1

 $\sigma = \sqrt{\frac{48\,025}{60} - (25.42)^2} = 12.4$ min

A1

Answer: 12.4 minutes

2 marks

9 (4 marks) $n = 11$, median is the 6th value = 31

B1

Lower half: 23, 25, 26, 28, 29 so $Q_1 = 26$

M1

Upper half: 33, 34, 36, 40, 45 so $Q_3 = 36$

A1

IQR = $36 - 26 = 10$

B1

Answer: median 31, Q1 26, Q3 36, IQR 10

4 marks

10 (4 marks)Mean = $\frac{4 + 7 + 9 + 12 + 13 + 17}{6} = 10.3333$

B1

Deviations: -6.33333 , -3.33333 , -1.33333 , 1.66667 , 2.66667 , 6.66667

M1

 $\Sigma(x - \bar{x})^2 = 107.333$

A1

Variance = $107.333 \div 6 = 17.89$, so $\sigma = 4.23$

A1

Answer: $\sigma^2 = 17.89$, $\sigma = 4.23$

4 marks

11 (3 marks)

Midpoints 5, 15, 25, 40

M1

$$\Sigma f x = 30 + 210 + 625 + 600 = 1465$$

M1

Estimated mean = $\frac{1465}{60} = 24.417$; modal class $20 \leq t < 30$ (highest frequency, but classes differ in width)

A1

Answer: 24.417 s; $20 \leq t < 30$

3 marks

12 (3 marks)

(a) Mean = $\frac{330}{7} = 47.1$

B1

Median = 35

B1

Answer: mean 47.1, median 35

2 marks

(b) The outlier 120 pulls the mean up to a value that is not typical; the median is unaffected

E1

Answer: Median is not affected by the outlier

1 mark

13 (5 marks)(a) Heights are rounded to the nearest cm, so 4 to 6 means $3.5 \leq h < 6.5$

B1

The class 10 means $9.5 \leq h < 10.5$

B1

Answer: 3.5 and 6.5; 9.5 and 10.5

2 marks

(b) Midpoints from the boundaries: 5, 8, 10, 12.5, 17.5

B1

$$\Sigma f x = 40 + 120 + 120 + 250 + 87.5 = 617.5$$

M1

$$\text{Mean} \approx \frac{617.5}{60} = 10.2917 \text{ cm}$$

A1

Answer: 10.2917 cm

3 marks

14 (6 marks)

(a) Mean = $\frac{480}{12} = 40$

B1

$$\text{Variance} = \frac{19\,800}{12} - 40^2 = 1650 - 1600 = 50$$

M1

$$\sigma = \sqrt{50} = 7.07$$

A1

Answer: Mean 40, $\sigma = 7.07$

3 marks

(b) $\Sigma x = 541$, $\Sigma x^2 = 19\,800 + 61^2 = 23\,521$

M1

$$\text{Mean} = \frac{541}{13} = 41.62$$

A1

$$\sigma = \sqrt{\frac{23\,521}{13} - (41.62)^2} = 8.80$$

A1

Answer: Mean 41.6, $\sigma = 8.80$

3 marks

15 (6 marks)

$$\Sigma x = 20 \times 62 + 30 \times 55 = 2890, \text{ so mean} = \frac{2890}{50} = 57.8$$

M1 A1

$$\text{Class A: } \Sigma x^2 = 20(8^2 + 62^2) = 78\,160$$

M1

$$\text{Class B: } \Sigma x^2 = 30(10^2 + 55^2) = 93\,750$$

A1

$$\text{Variance} = \frac{78\,160 + 93\,750}{50} - 57.8^2 = 97.36$$

M1

$$\sigma = 9.87$$

A1

Answer: Mean 57.8, $\sigma = 9.87$

6 marks

16 (5 marks)

(a) Midpoints 2.5, 7.5, 15, 25, 35

M1

$$\Sigma f x = 680, \text{ mean} = 13.6 \text{ km}$$

A1

Answer: 13.6 km

2 marks

(b) $\Sigma f x^2 = 12675$

M1

$$\text{Variance} = \frac{12675}{50} - 13.6^2 = 68.54$$

M1

$$\sigma = 8.28 \text{ km}$$

A1

Answer: 8.28 km

3 marks

17 (5 marks)

Cumulative frequencies 10, 32, 60, 76, 80

B1

Median is the 40th value, in $40 \leq a < 50$: $40 + \frac{40 - 32}{28} \times 10 = 42.8571$

M1 A1

 Q_1 : 20th value in $30 \leq a < 40$: $30 + \frac{20 - 10}{22} \times 10 = 34.5455$; Q_3 : 60th value = 50

M1

IQR = $50 - 34.5455 = 15.455$ years

A1

Answer: median 42.857, IQR 15.455

5 marks

18 (7 marks)(a) $\Sigma x = 163.1$ and $\Sigma x^2 = 2461.01$

B1

Mean = $\frac{163.1}{11} = 14.8273$

M1

 $\sigma = \sqrt{\frac{2461.01}{11} - 14.8273^2} = 1.97$

A1

Answer: mean 14.8273, $\sigma = 1.97$

3 marks

(b) The 6th of the 11 ordered values is 14.1

B1

Answer: 14.1

1 mark

(c) $3(14.8273 - 14.1) = 2.1818$

M1

Coefficient = $\frac{2.1818}{1.97} = 1.108$

A1

The coefficient is positive, so the data are positively skewed

E1

Answer: 1.108, positive skew

3 marks

19 (6 marks)(a) Group A: median = $\frac{36 + 38}{2} = 37$

B1

Group A: $Q_1 = \frac{28 + 31}{2} = 29.5$, $Q_3 = \frac{44 + 46}{2} = 45$, IQR = 15.5

B1

Group B: median = $\frac{45 + 47}{2} = 46$

B1

Group B: $Q_1 = \frac{38 + 42}{2} = 40$, $Q_3 = \frac{51 + 53}{2} = 52$, IQR = 12

B1

Answer: A: 37 and 15.5; B: 46 and 12

4 marks

(b) Group B has the higher median (46 against 37), so its marks are higher on average

E1

Group B has the smaller IQR (12 against 15.5), so its marks are more consistent

E1

Answer: B higher and less spread

2 marks

20 (5 marks) $a + 12 + 15 + b + 3 = 40$, so $a + b = 10$

M1 A1

 $\Sigma f x = 12 + 30 + 3b + 12 = 54 + 3b$

M1

 $54 + 3b = 1.95 \times 40 = 78$

M1

 $b = 8$ and $a = 2$

A1

Answer: a = 2, b = 8

5 marks

21 (7 marks)

(a) Cumulative frequencies: 8, 28, 54, 72, 80

 Q_1 : 20th value in 20-25: $20 + \frac{20 - 8}{20} \times 5 = 23$

M1 A1

Median: 40th value in 25-30: $25 + \frac{40 - 28}{26} \times 5 = 27.3077$

M1 A1

 Q_3 : 60th value in 30-40: $30 + \frac{60 - 54}{18} \times 10 = 33.3333$

M1 A1

Answer: $Q_1 = 23$, median ≈ 27.3 , $Q_3 \approx 33.3$

6 marks

(b) $Q_3 - \text{median} = 6.026$ is greater than $\text{median} - Q_1 = 4.308$, so positive skew

E1

Answer: Positive skew

1 mark

22 (7 marks)

(a) $\Sigma x = 4 + 6 + a + b + 10 = 5 \times 7 = 35$

M1

$a + b = 35 - 20 = 15$

A1

Answer: Shown

2 marks

(b) $\frac{\Sigma x^2}{5} - 7^2 = 4$, so $\Sigma x^2 = 265$

M1

$16 + 36 + 100 + a^2 + b^2 = 265$, so $a^2 + b^2 = 113$

A1

Answer: Shown

2 marks

(c) $(a + b)^2 = 225$, so $2ab = 225 - 113 = 112$ and $ab = 56$

M1

a and b are roots of $t^2 - 15t + 56 = 0$, so $(t - 7)(t - 8) = 0$

M1

$a = 7$, $b = 8$

A1

Answer: a = 7, b = 8

3 marks

23 (9 marks)

(a) $\Sigma x = 694$, mean = 57.8333

M1 A1

$\Sigma x^2 = 41328$

M1

$\sigma = 9.965$

A1

Answer: mean 57.8333, σ 9.965

4 marks

(b) Limits: $57.8333 \pm 2 \times 9.965 = (37.9029, 77.7638)$

M1

Only 89 lies outside, so 89 is an outlier

A1

Answer: 89 only

2 marks

(c) New mean = $\frac{694 - 89}{11} = 55$

M1 A1

The mean falls from 57.833 to 55: the outlier distorted the mean

E1

Answer: 55

3 marks

24 (7 marks)

(a) $2k + (k + 4) + (3k - 1) + (k + 1) + 2 = 7k + 6$

M1

$7k + 6 = 34$, so $k = 4$

A1

Answer: k = 4

2 marks

(b) Frequencies 8, 8, 11, 5, 2; $\Sigma f x = 87$

M1

Mean = $\frac{87}{34} = 2.5588$

A1

$\Sigma f x^2 = 269$

M1

$\sigma = \sqrt{\frac{269}{34} - 2.5588^2} = 1.17$

A1

Answer: mean 2.5588, σ 1.17

4 marks

(c) Cumulative 8, 16, 27, ...; the 17th and 18th values are both 3

B1

Answer: 3

1 mark

25 (6 marks)

(a) A: mean = $\frac{2016}{40} = 50.4$, variance = $\frac{101\ 650}{40} - 50.4^2 = 1.09$, $\sigma = 1.04$

M1 A1

B: mean = 50.1, variance = $\frac{100\ 631}{40} - 50.1^2 = 5.765$, $\sigma = 2.40$

M1 A1

Answer: A: 50.4, 1.04; B: 50.1, 2.40

4 marks

(b) B has the mean closer to 50 g (50.1 against 50.4)

E1

A is more consistent because its standard deviation is smaller (1.04 against 2.40)

E1

Answer: B nearer target, A more consistent

2 marks

26 (7 marks)

(a) Total of all marks = $30 \times 58 = 1740$

M1

Total for the boys = $18 \times 61 = 1098$, so total for the girls = 642

M1

Mean = $\frac{642}{12} = 53.5$

A1

Answer: 53.5

3 marks

(b) Σx^2 for the class = $30(12^2 + 58^2) = 105\ 240$

M1

$$\Sigma x^2 \text{ for the boys} = 18(10^2 + 61^2) = 68\,778$$

M1

$$\text{Girls: } \Sigma x^2 = 36\,462, \text{ so variance} = \frac{36\,462}{12} - 53.5^2 = 176.25$$

M1

$$\sigma = 13.2759$$

A1

Answer: $\sigma = 13.2759$

4 marks

27 (9 marks)

$$(a) \Sigma x = 25 \times 48 = 1200$$

B1

$$\Sigma x^2/25 = 6^2 + 48^2 = 2340$$

M1

$$\Sigma x^2 = 58\,500$$

A1

Answer: Shown

3 marks

$$(b) \text{Corrected } \Sigma x = 1200 - 70 + 40 = 1170$$

M1

$$\text{Mean} = \frac{1170}{25} = 46.8$$

A1

Answer: 46.8

2 marks

$$(c) \text{Corrected } \Sigma x^2 = 58\,500 - 70^2 + 40^2 = 55\,200$$

M1

$$\text{Variance} = \frac{55\,200}{25} - 46.8^2 = 2208 - 2190.24 = 17.76$$

M1

$$\sigma = 4.21$$

A1

Answer: 4.21

3 marks

(d) The incorrect value 70 was 22 above the old mean, but the correct value is much closer to the centre, so the data are less spread out

E1

Answer: The extreme value has been replaced by a central value

1 mark

28 (9 marks)

$$(a) \Sigma(x - \bar{x})^2 = \Sigma x^2 - 2\bar{x}\Sigma x + \Sigma \bar{x}^2$$

M1

$$\Sigma x = n\bar{x} \text{ and } \Sigma \bar{x}^2 = n\bar{x}^2$$

A1

$$= \Sigma x^2 - 2n\bar{x}^2 + n\bar{x}^2 = \Sigma x^2 - n\bar{x}^2$$

A1

Answer: Shown

3 marks

$$(b) \text{Write } x - a = (x - \bar{x}) + (\bar{x} - a) \text{ and square}$$

M1

$$\Sigma(x - a)^2 = \Sigma(x - \bar{x})^2 + 2(\bar{x} - a)\Sigma(x - \bar{x}) + n(\bar{x} - a)^2$$

A1

$$\Sigma(x - \bar{x}) = 0, \text{ giving the result}$$

A1

Answer: Shown

3 marks

$$(c) \text{The last term is never negative and is zero only when } a = \bar{x}, \text{ so the sum is least when } a = \bar{x}$$

E1

Answer: $a = \bar{x}$

1 mark

$$(d) \bar{x} = 6 \text{ and } \Sigma(x - 6)^2 = 16 + 4 + 0 + 4 + 16 = 40$$

M1

$$\Sigma(x - 4)^2 = 40 + 5 \times (6 - 4)^2 = 60$$

A1

Answer: 60

2 marks

29 (9 marks)

$$(a) \text{The lower half is the first 7 values: 23, 28, 31, 34, 36, 39, 40}$$

M1

$$Q_1 = \text{the 4th value} = 34$$

A1

Answer: 34

2 marks

$$(b) \text{The median is the 8th value, which is } 4a$$

M1

$$43 = 4a \text{ so } a = 3$$

A1

Answer: $a = 3$

2 marks

$$(c) Q_3 = Q_1 + \text{IQR} = 34 + 19 = 53$$

M1

The upper half is the 9th to 15th values; Q_3 is the 12th value, which is 5b

M1

$$\text{So } b = 3$$

A1

Answer: $b = 3$

3 marks

$$(d) \text{Upper fence} = 53 + 1.5 \times 19 = 81.5 \text{ and lower fence} = 34 - 28.5 = 5.5$$

M1

All the marks lie between 23 and 66, so there are no outliers

A1

Answer: Shown

2 marks

30 (10 marks)

$$(a) \text{ Mean} = \frac{n+1}{2} \quad \text{B1}$$

$$\text{Mean of squares} = \frac{(n+1)(2n+1)}{6} \quad \text{M1}$$

$$\text{Variance} = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \quad \text{M1}$$

$$= \frac{(n+1)[2(2n+1) - 3(n+1)]}{12} \quad \text{M1}$$

$$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12} \quad \text{A1}$$

Answer: Shown 5 marks

$$(b) \text{ Variance} = \frac{399}{12} = 33.25 \quad \text{M1}$$

$$\sigma = 5.77 \quad \text{A1}$$

Answer: 5.77 2 marks

(c) These are 5 times the values 1, ..., 20 M1

$$\text{Mean} = 5 \times 10.5 = 52.5 \quad \text{B1}$$

$$\sigma = 5 \times 5.766 = 28.8 \quad \text{A1}$$

Answer: mean 52.5, σ 28.8 3 marks

Histograms and Cumulative Frequency

1 (3 marks)

$$\text{Frequency density} = \frac{\text{frequency}}{\text{class width}} \quad \text{M1}$$

$$140-150: \frac{12}{10} = 1.2, \text{ and } 150-160: \frac{30}{10} = 3 \quad \text{A1}$$

$$160-180: \frac{36}{20} = 1.8, \text{ and } 180-200: \frac{10}{20} = 0.5 \quad \text{A1}$$

Answer: 1.2, 3, 1.8, 0.5 3 marks

2 (3 marks)

$$(a) 5, 5 + 12 = 17, 17 + 20 = 37, 37 + 9 = 46 \quad \text{M1}$$

$$\text{and } 46 + 4 = 50 \quad \text{A1}$$

Answer: 5, 17, 37, 46, 50 2 marks

(b) Plot at the upper class boundary: (20, 5) B1

Answer: (20, 5) 1 mark

3 (3 marks)

Times are rounded to the nearest minute, so the class 1 to 3 starts at 0.5, and the boundaries are 0.5, 3.5, 6.5, 10.5 B2

The last class ends at 20.5, so the widths are 3, 3, 4 and 10 B1

Answer: 0.5 to 3.5, 3.5 to 6.5, 6.5 to 10.5, 10.5 to 20.5; widths 3, 3, 4, 10 3 marks

4 (2 marks)

$$\text{Frequency} = 4 \times 2.5 = 10 \quad \text{B1}$$

$$\text{Height} = \frac{35}{10} = 3.5 \quad \text{B1}$$

Answer: 10 and 3.5 2 marks

5 (2 marks)

$$\text{Area } 2 \times 6 = 12 \text{ cm}^2 \text{ represents } 30, \text{ so } 1 \text{ cm}^2 \text{ represents } 2.5 \quad \text{M1}$$

$$\text{Second area} = 3.5 \times 4 = 14 \text{ cm}^2, \text{ so frequency} = 14 \times 2.5 = 35 \quad \text{A1}$$

Answer: 35 2 marks

6 (2 marks)

Median is the 30th value: read across from 30 on the vertical axis, about 24.545 g

B1

At 35 g the graph is at about 49

B1

Answer: about 24.545 g; about 49 eggs

2 marks

7 (4 marks)

(a) Frequency = frequency density $\times$ class width

M1

$$2.4 \times 10 = 24$$

A1

Answer: 24

2 marks

(b) Frequencies: $1.6 \times 10 = 16$, $3.2 \times 5 = 16$, $2.4 \times 10 = 24$, $0.8 \times 15 = 12$

M1

Total = 68

A1

Answer: 68

2 marks

8 (5 marks)

Frequency densities: $\frac{8}{5} = 1.6$, $\frac{20}{5} = 4$, $\frac{30}{10} = 3$

M1

$$\frac{24}{20} = 1.2 \text{ and } \frac{6}{20} = 0.3$$

A1

Bars drawn with no gaps and correct widths

B1

Bar heights correct

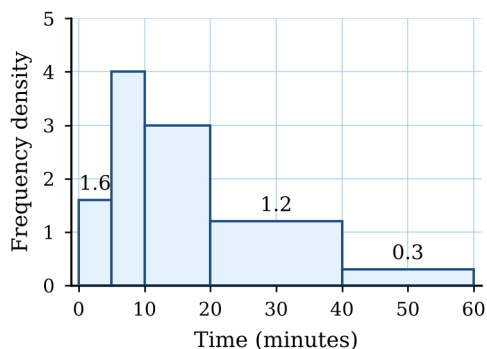
B1

Both axes labelled, with frequency density on the vertical axis

B1

Answer: See histogram

5 marks

**9 (4 marks)**

(a) Frequency = fd $\times$ width

M1

$$1.2 \times 10 = 12, 3 \times 10 = 30$$

A1

$$1.5 \times 20 = 30, 0.4 \times 20 = 8, \text{ total } 80$$

A1

Answer: 12, 30, 30, 8

3 marks

(b) The tallest bar is 10 to 20 minutes

B1

Answer: 10 to 20 minutes

1 mark

10 (4 marks)

$$\text{Frequency density} = \frac{\text{frequency}}{\text{width}}$$

M1

$$\frac{15}{5} = 3, \frac{28}{10} = 2.8$$

A1

$$4.2 \times 5 = 21$$

A1

$$0.6 \times 20 = 12$$

A1

Answer: 3, 21, 2.8, 12

4 marks

11 (3 marks)

$$24 \text{ to } 30: (30 - 24) \times 2.2 = 13.2$$

M1

$$30 \text{ to } 40: (40 - 30) \times 1.4 = 14$$

M1

Total = 27.2, so about 27

A1

Answer: about 27

3 marks

12 (3 marks)**(a)** 50th value: about 49.524**B1****Answer: about 49.5****1 mark****(b)** At score 70, cumulative frequency $\approx 72 + \frac{70 - 60}{20} \times 23 = 83.5$ **M1**

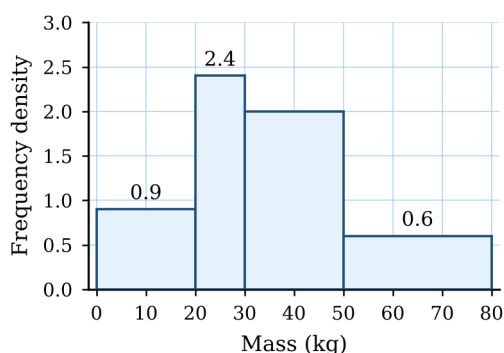
100 – 83.5 = 16.5, about 16 or 17

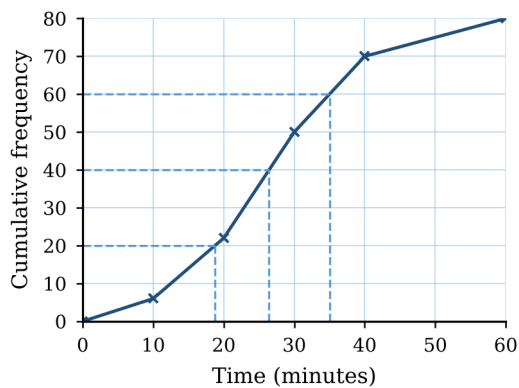
A1**Answer: about 16 or 17****2 marks****13 (4 marks)**Frequencies: $2 \times 5 = 10$, $3.6 \times 5 = 18$, $1.8 \times 10 = 18$, $0.6 \times 10 = 6$ **M1 A1** $\Sigma f x = 10(2.5) + 18(7.5) + 18(15) + 6(25) = 580$ **M1**Mean $\approx \frac{580}{52} = 11.154$ min**A1****Answer: 11.154 min****4 marks****14 (6 marks)****(a)** Known frequencies: $0.9 \times 20 = 18$, $2.0 \times 20 = 40$, $0.6 \times 30 = 18$, total 76**M1**

100 – 76 = 24

A1**Answer: 24****2 marks****(b)** Frequency density = $\frac{24}{10} = 2.4$ **B1****Answer: 2.4****1 mark****(c)** 40 to 50: $10 \times 2.0 = 20$ **M1**50 to 65: $15 \times 0.6 = 9$ **M1**

Total = 29

A1**Answer: 29****3 marks****15 (6 marks)****(a)** 40th value lies between (20, 22) and (30, 50)**M1** $20 + \frac{40 - 22}{28} \times 10 = 26.4$ minutes**A1****Answer: 26.4 minutes****2 marks****(b)** Q_1 : 20th value between (10, 6) and (20, 22): $10 + \frac{14}{16} \times 10 = 18.75$ **M1 A1** Q_3 : 60th value between (30, 50) and (40, 70): $30 + \frac{10}{20} \times 10 = 35$, so IQR = 16.25 minutes**A1****Answer: 16.25 minutes****3 marks****(c)** At 45: $70 + \frac{5}{20} \times 10 = 72.5$, so $80 - 72.5 = 7.5$, about 8 runners**B1****Answer: About 8****1 mark**

**16 (9 marks)**

(a) Boundaries 0.5, 3.5, 7.5, 9.5, 14.5, 23.5, so the widths are 3, 4, 2, 5, 9

B1 M1

Frequency densities $\frac{15}{3} = 5$, $\frac{28}{4} = 7$, $\frac{24}{2} = 12$, $\frac{20}{5} = 4$, $\frac{18}{9} = 2$

A1

Answer: 5, 7, 12, 4, 2

3 marks

(b) 5.5 to 7.5: $2 \times 7 = 14$

M1

7.5 to 9.5: 24, and 9.5 to 11.5: $2 \times 4 = 8$

M1

Total = 46

A1

Answer: 46

3 marks

(c) Cumulative frequencies 15, 43, 67, so the median is the 52.5th value in 7.5 to 9.5

M1

$$7.5 + \frac{52.5 - 43}{24} \times 2$$

M1

= 8.292 hours

A1

Answer: 8.292 hours

3 marks

17 (6 marks)

(a) Median: 60th value, between (40, 58) and (50, 90): $40 + \frac{2}{32} \times 10 = 40.625$

M1 A1

Q_1 : 30th value: $30 + \frac{30 - 26}{32} \times 10 = 31.25$; Q_3 : 90th value = 50

M1

IQR = $50 - 31.25 = 18.75$

A1

Answer: median 40.625, IQR 18.75

4 marks

(b) 90% of 120 = 108th value, between (50, 90) and (60, 110)

M1

$$50 + \frac{108 - 90}{20} \times 10 = 59$$

A1

Answer: 59

2 marks

18 (5 marks)

(a) Frequencies $2.5 \times 4 = 10$, $4 \times 4 = 16$, $3 \times 4 = 12$, $1 \times 8 = 8$

M1

Total = 46

A1

Answer: Shown

2 marks

(b) Median is the 23rd value; cumulative 10, 26

M1

$$\text{Median in } 4 \leq x < 8: 4 + \frac{23 - 10}{16} \times 4$$

M1

= 7.25

A1

Answer: 7.25

3 marks

19 (5 marks)

(a) Frequencies: 5, 13, 28, 24, 10

M1 A1

Frequency densities: 0.5, 1.3, 2.8, 1.2, $\frac{10}{30} = 0.333$

A1

Answer: f: 5, 13, 28, 24, 10

3 marks

(b) 25 to 30: $5 \times 2.8 = 14$

M1

30 to 40: $10 \times 1.2 = 12$, total 26

A1

Answer: 26

2 marks

20 (5 marks)**(a)** Frequencies: 10k, 30k, 40k, 20k

M1 A1

Total 100k = 120, so $k = 1.2$

A1

Answer: $k = 1.2$

3 marks

(b) $10k + 30k + 10 \times 2k = 60k$

M1

 $= 72$

A1

Answer: 72

2 marks

21 (9 marks)**(a)** Midpoints 15, 22.5, 27.5, 32.5, 37.5, 50: $\Sigma fx = 3732.5$

M1

Mean = $\frac{3732.5}{120} = 31.10417$ cm

A1

 $\Sigma fx^2 = 123306$, so variance = $\frac{123306}{120} - 31.10417^2 = 60.0829$

M1

 $\sigma = 7.7513$ cm

A1

Answer: mean 31.1042 cm, $\sigma = 7.751$ cm

4 marks

(b) Limits 23.3529 and 38.8555

B1

Number below 38.8555 ≈ 106.4219 and below 23.3529 ≈ 16.0586

M1

Number within one $\sigma \approx 90.363$, which is 75.303%

A1

Answer: about 75.303%

3 marks

(c) The data have more values near the mean (75.303% against 68%)

B1

The median 30.6944 is close to the mean 31.1042, so the data are roughly symmetrical, but the distribution is more peaked than a Normal distribution, so the model is only rough

E1

Answer: More concentrated than Normal; only a rough model

2 marks

22 (8 marks)**(a)** 4 units of t are represented by 2 cm, so 1 unit is 0.5 cm

M1

Width = $12 \times 0.5 = 6$ cm

A1

Answer: 6 cm

2 marks

(b) The area $2 \times 4.5 = 9$ cm² represents 18, so 1 cm² represents 2 observations

M1

Area = $\frac{30}{2} = 15$ cm², height = $\frac{15}{6} = 2.5$ cm

A1

Answer: 2.5 cm

2 marks

(c) Width = $24 \times 0.5 = 12$ cm

B1

Area = $\frac{18}{2} = 9$ cm², height = $\frac{9}{12} = 0.75$ cm

B1

Answer: 12 cm by 0.75 cm

2 marks

(d) Frequency = $100 - 18 - 30 - 18 = 34$, area = 17 cm², width = 10 cm

M1

Height = $\frac{17}{10} = 1.7$ cm

A1

Answer: 10 cm by 1.7 cm

2 marks

23 (9 marks)**(a)** $fd \times$ width: 0.5×10 , 2.4×5 , 3.2×5 , 1.3×10 , 0.35×20

M1

5, 12, 16, 13, 7

A1

Answer: 5, 12, 16, 13, 7

2 marks

(b) Midpoints 5, 12.5, 17.5, 25, 40: $\Sigma f x = 1060$

M1

Mean = 20

A1

 $\Sigma f x^2 = 26225$, variance = $494.8113 - 20^2 = 94.8113$

M1

 $\sigma = 9.74$

A1

Answer: mean 20, σ 9.74

4 marks

(c) Cumulative 5, 17, 33: median is the 26.5th value in $15 \leq x < 20$: $15 + \frac{26.5 - 17}{16} \times 5 =$

M1 A1

17.9688

Mean 20 > median 17.969: positive skew

E1

Answer: median 17.969; positive skew

3 marks

24 (9 marks)**(a)** Boundaries 0.5, 2.5, 4.5, 5.5, 8.5, 14.5 give midpoints 1.5, 3.5, 5, 7, 11.5**B1**

$$\Sigma fx = 504, \text{ so mean} = \frac{504}{82} = 6.1463$$

M1 A1

$$\sigma = \sqrt{\frac{3981}{82} - 6.1463^2} = 3.282$$

A1**Answer: mean 6.1463, $\sigma = 3.282$** **4 marks****(b)** Cumulative frequencies 10, 26, 46: the 41st value is in 4.5 to 5.5**M1**

$$4.5 + \frac{41 - 26}{20} = 5.25 \text{ minutes}$$

A1**Answer: 5.25 minutes****2 marks****(c)** Class 1 to 2 has width 2 and base 2.5 cm, so 1.25 cm represents 1 minute; the class 9 to 14 has width 6, so base = 7.5 cm**M1 A1**

$$\text{Area } 10 \text{ cm}^2 \text{ represents } 10, \text{ so } 18 \text{ is represented by } 18 \text{ cm}^2 \text{ and height} = \frac{18}{7.5} = 2.4 \text{ cm}$$

A1**Answer: base 7.5 cm, height 2.4 cm****3 marks****25 (7 marks)****(a)** 14, 60, 112, 160, 190, 200**B1****Answer: 14, 60, 112, 160, 190, 200****1 mark****(b)** 35% fail: $0.35 \times 200 = 70$ th value**M1**

$$\text{In } 40 \leq x < 50: 40 + \frac{70 - 60}{52} \times 10$$

M1

$$= 41.9231$$

A1**Answer: 41.9231****3 marks****(c)** 180th value**M1**

$$\text{In } 60 \leq x < 80: 60 + \frac{180 - 160}{30} \times 20$$

M1

$$= 73.3333$$

A1**Answer: 73.3333****3 marks****26 (6 marks)**

$$\text{Total} = 10 + x + 24 + 16 = 50 + x, \text{ so median is the } (25 + \frac{x}{2})\text{th value}$$

M1It lies in $20 \leq m < 30$ provided the cumulative frequency to 20 is $10 + x$ **B1**

$$\text{Interpolating: } 20 + \frac{(25 + \frac{x}{2}) - (10 + x)}{24} \times 10 = 20 + \frac{15 - \frac{x}{2}}{24} \times 10$$

M1 A1

$$\text{Set equal to } 23.75: \frac{15 - \frac{x}{2}}{24} \times 10 = 3.75$$

M1

$$15 - \frac{x}{2} = 9, \text{ so } x = 12$$

A1**Answer: $x = 12$** **6 marks****27 (10 marks)****(a)** Frequencies: 20, 10h, 30, 16, total $66 + 10h$ **M1**

$$\Sigma fx = 5 \times 20 + 15 \times 10h + 25 \times 30 + 40 \times 16 = 1490 + 150h$$

M1

$$\frac{1490 + 150h}{66 + 10h} = 20$$

M1

$$1490 + 150h = 1320 + 200h, \text{ so } h = 3.4$$

A1**Answer: $h = 3.4$** **4 marks****(b)** Total = 100; cumulative frequencies 20, 54, 84, 100; the 50th value is in 10-20**M1**

$$10 + \frac{50 - 20}{34} \times 10 = 18.8$$

A1**Answer: 18.8****2 marks****(c)** 15 to 20: $5 \times 3.4 = 17$, 20 to 30: 30, 30 to 35: $5 \times 0.8 = 4$ **M1 M1**

$$\text{Probability} = \frac{51}{100} = 0.51$$

A1**Answer: 0.51****3 marks**

(d) The tail towards large x (positive skew) pulls the mean above the median

Answer: Positive skew pulls the mean up

E1

1 mark

28 (11 marks)

(a) A: 100th value in 50-60: $50 + \frac{100 - 84}{56} \times 10 = 52.9$

M1 A1

B: 80th value in 40-50: $40 + \frac{80 - 56}{32} \times 10 = 47.5$

A1

Answer: A: 52.9, B: 47.5

3 marks

(b) A: Q_1 (50th) = $40 + \frac{10}{44} \times 10 = 42.2727$; Q_3 (150th) = $60 + \frac{10}{48} \times 20 = 64.1667$

M1

IQR (A) = 21.9

A1

B: Q_1 (40th) = $20 + \frac{24}{40} \times 20 = 32$; Q_3 (120th) = $60 + \frac{8}{36} \times 20 = 64.4444$, IQR (B) = 32.4

A1

Answer: A: 21.9, B: 32.4

3 marks

(c) A: at 70, cumulative = $140 + \frac{10}{20} \times 48 = 164$, so $\frac{36}{200} = 18\%$

M1 A1

B: at 70, cumulative = $112 + \frac{10}{20} \times 36 = 130$, so $\frac{30}{160} = 18.75\%$

A1

Answer: A: 18%, B: 18.75%

3 marks

(d) School A has the higher median (52.9 against 47.5), so its typical score is higher

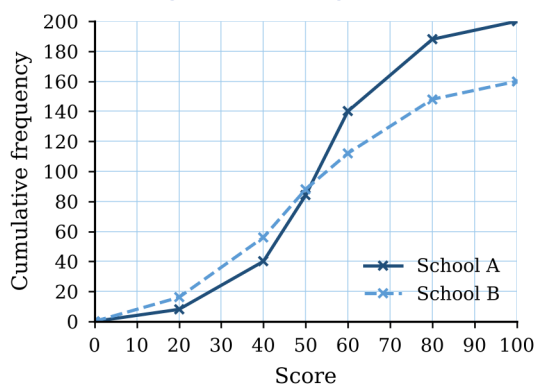
E1

School A has the smaller IQR (21.9 against 32.4), so its scores are more consistent, although B has a slightly larger percentage above 70

E1

Answer: A higher on average and more consistent

2 marks



29 (9 marks)

(a) $1.5 \times 10 = 15$, $2.5 \times 10 = 25$, $0.5 \times 20 = 10$

M1 A1

Answer: 15, 25, 10

2 marks

(b) Waited more than 10 min: $25 + 10 = 35$

B1

More than 25 min: 25 to 40 gives $15 \times 0.5 = 7.5$

M1 A1

Probability = $\frac{7.5}{35} = \frac{3}{14}$

A1

Answer: $\frac{3}{14}$

4 marks

(c) $P(\text{first}) = \frac{15}{50}$

B1

$P(\text{second}) = \frac{14}{49}$

M1

Probability = $\frac{15}{50} \times \frac{14}{49} = \frac{3}{35}$

A1

Answer: $\frac{3}{35}$

3 marks

30 (10 marks)

(a) Midpoints 5, 15, 25, 40 M1

$$\Sigma f x = 40 + 270 + 525 + 520 = 1355 \quad \text{M1}$$

$$\text{Mean} \approx \frac{1355}{60} = 22.5833 \quad \text{A1}$$

Answer: 22.5833 3 marks

(b) Smallest: use lower class boundaries: $\frac{0 \times 8 + 10 \times 18 + 20 \times 21 + 30 \times 13}{60} = \frac{990}{60} = 16.5$ M1 A1

Largest: upper boundaries: $\frac{1720}{60} = 28.6667$ A1

Answer: 16.5 to 28.6667 3 marks

(c) True mean = $\frac{1300}{60} = 21.6667$ B1

Error = $22.5833 - 21.6667 = 0.917$ years (an overestimate) A1

Answer: True mean 21.6667; estimate too high by 0.917 2 marks

(d) Every age in a class lies between its boundaries, so the true total lies between the two sums E1

Midpoints assume that values are spread evenly so that errors cancel E1

Answer: Boundaries give extremes 2 marks

Box Plots and Outliers

1 (3 marks)

Median = 26 minutes B1

IQR = $35 - 20 = 15$ minutes B1

Range = $50 - 12 = 38$ minutes B1

Answer: 26, 15, 38 3 marks

2 (3 marks)

Minimum 4 and maximum 30 B1

Median = $\frac{12 + 14}{2} = 13$ B1

$Q_1 = \frac{7 + 9}{2} = 8$ and $Q_3 = \frac{18 + 21}{2} = 19.5$ B1

Answer: 4, 8, 13, 19.5, 30 3 marks

3 (3 marks)

(a) 8 kg is the median, so half of the 80 parcels: 40 B1

Answer: 40 1 mark

(b) The quartiles are 5 and 14, so half of the parcels lie between them: 40 B1

Answer: 40 1 mark

(c) 14 kg is the upper quartile, so 25% B1

Answer: 25% 1 mark

4 (3 marks)

IQR = $32 - 20 = 12$, $1.5 \times \text{IQR} = 18$ M1

Lower limit = $20 - 18 = 2$ A1

Upper limit = $32 + 18 = 50$ A1

Answer: 2 and 50 3 marks

5 (3 marks)

Median is the 5th value = 15 B1

Lower half 6, 9, 11, 14: $Q_1 = 10$ B1

Upper half 18, 21, 22, 30: $Q_3 = 21.5$ B1

Answer: median 15, Q1 10, Q3 21.5 3 marks

6 (3 marks)

IQR = 12, so limits are $30 - 18 = 12$ and $42 + 18 = 60$

M1 A1

Outliers: 8, 11 (below 12) and 70 and 62 (above 60), so all four

A1

Answer: all four

3 marks

7 (4 marks)

(a) IQR = $46 - 34 = 12$, so $1.5 \times \text{IQR} = 18$

B1

Fences: $34 - 18 = 16$ and $46 + 18 = 64$

M1

$12 < 16$ and $71 > 64$ so 12 and 71 are outliers

A1

Answer: 12 and 71

3 marks

(b) The whiskers end at the most extreme non-outliers: 20 and 58

B1

Answer: 20 and 58

1 mark

8 (4 marks)

(a) Class A: $61 - 35 = 26$

B1

Class B: $56 - 41 = 15$

B1

Answer: A: 26, B: 15

2 marks

(b) Class B has the slightly higher median (50 against 48), so its typical mark is a little higher

E1

Class B has the smaller IQR (and range), so its marks are more consistent

E1

Answer: B higher median and less spread

2 marks

9 (3 marks)

Axis with a suitable scale

B1

Box from 20 to 34 with a line at the median 26

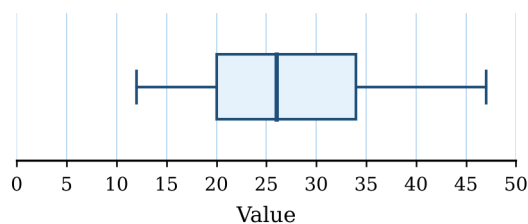
B1

Whiskers to 12 and 47

B1

Answer: See diagram

3 marks

**10 (7 marks)**

(a) Median is the 5th value = 22

B1

Lower half 14, 17, 18, 21: $Q_1 = \frac{17 + 18}{2} = 17.5$

B1

Upper half 25, 26, 29, 58: $Q_3 = \frac{26 + 29}{2} = 27.5$

B1

Answer: 22, 17.5, 27.5

3 marks

(b) IQR = $27.5 - 17.5 = 10$, so the upper fence is $27.5 + 15 = 42.5$

M1

$58 > 42.5$ so it is an outlier

A1

Answer: Shown

2 marks

(c) Box from 17.5 to 27.5 with the median at 22, whiskers to 14 and 29

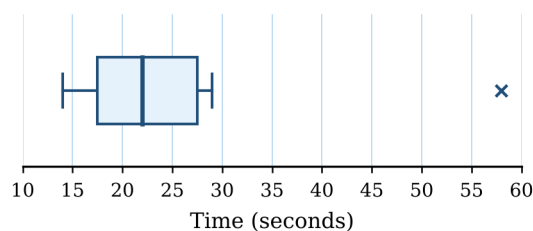
B1

Cross at 58

B1

Answer: See diagram

2 marks



11 (3 marks)

$$Q_2 - Q_1 = 4 \text{ and } Q_3 - Q_2 = 13$$

M1 A1

$$Q_3 - Q_2 > Q_2 - Q_1, \text{ so positive skew}$$

E1

Answer: Positive skew

3 marks

12 (3 marks)

$$\text{Limits: } 64 \pm 2 \times 6 = 52 \text{ and } 76$$

M1 A1

$$49 < 52 \text{ so } 49 \text{ is an outlier; } 75 < 76 \text{ so } 75 \text{ is not}$$

A1

Answer: 49 yes, 75 no

3 marks

13 (5 marks)

$$\text{Median} = \frac{11 + 12}{2} = 11.5$$

B1

$$Q_1 = \frac{6 + 8}{2} = 7, Q_3 = \frac{15 + 17}{2} = 16$$

M1 A1

$$\text{Upper limit} = 16 + 1.5 \times 9 = 29.5; \text{ lower limit} = -6.5$$

M1

Only 45 is an outlier

A1

Answer: 45 is the only outlier

5 marks

14 (6 marks)

$$\text{(a) IQR} = 55 - 40 = 15, \text{ so } 1.5 \times \text{IQR} = 22.5$$

B1

$$\text{Fences: } 40 - 22.5 = 17.5 \text{ and } 55 + 22.5 = 77.5$$

M1

$$84 > 77.5 \text{ so it is an outlier, but } 18 > 17.5 \text{ so it is not}$$

A1

Answer: Shown

3 marks

(b) Box from 40 to 55 with the median at 47

B1

Whiskers to 18 and 66

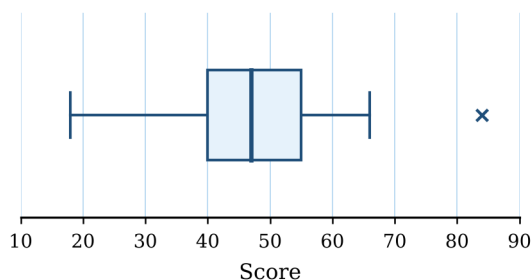
B1

Cross at 84

B1

Answer: See diagram

3 marks

**15 (6 marks)**

$$\text{(a) } 34 - 22 = 12$$

B1

Answer: 12

1 mark

$$\text{(b) Upper fence} = 34 + 1.5 \times 12 = 52$$

M1

$$61 > 52$$

A1

Answer: Shown

2 marks

(c) The whisker ends at the largest value that is not an outlier, 49

B1

So no data value lies between 49 and 52, and every value above 52 is an outlier

E1

Answer: No values between 49 and 52

2 marks

(d) Lower fence = $22 - 18 = 4$, and the smallest value 8 is above it

B1

Answer: Shown

1 mark

16 (8 marks)**(a)** Stems 1 to 8, with the leaves attached to the correct stems

B1

Leaves in increasing order along each row

B1

A key, for example 2 | 6 means 26

B1

Answer: See diagram

3 marks

(b) Median is the 8th value = 26

B1

$$\text{Lower half } 12, 15, 17, 19, 21, 22, 24: Q_1 = 19$$

B1

$$\text{Upper half } 28, 30, 33, 38, 47, 55, 81: Q_3 = 38$$

B1

Answer: 26, 19, 38**3 marks****(c)** $IQR = 38 - 19 = 19$, upper fence = $38 + 28.5 = 66.5$ **M1**81 > 66.5 so 81 is an outlier, and the lower fence $19 - 28.5 < 0$ gives no low outliers**A1****Answer: 81 only****2 marks**

1	2 5 7 9
2	1 2 4 6 8
3	0 3 8
4	7
5	5
6	
7	
8	1

Key: 2 | 6 means 26

17 (6 marks)**(a)** 14 values: $\frac{22 + 23}{2} = 22.5$ **B1**

13 values: the 7th value = 22

B1**Answer: 22.5 and 22****2 marks****(b)** Sum of 14 = 388, mean = 27.7143**M1 A1**

Sum of 13 = 290, mean = 22.3077

A1

The mean changes by 5.407 but the median by only 0.5, so the median is more resistant to outliers

E1**Answer: 27.7143 to 22.3077****4 marks****18 (7 marks)****(a)** Median is the 5th value = 19**B1**

$$Q_1 = \frac{14 + 15}{2} = 14.5, Q_3 = \frac{22 + 23}{2} = 22.5$$

M1 A1

$$IQR = 22.5 - 14.5 = 8$$

A1**Answer: Q1 = 14.5, Q3 = 22.5, IQR = 8****4 marks****(b)** Upper limit = $22.5 + 1.5 \times 8 = 34.5$ **M1 A1**x is an outlier when it exceeds the upper limit, so $x > 34.5$ **A1****Answer: $x > 34.5$** **3 marks****19 (6 marks)****(a)** IQRs: Group 1 = 24, Group 2 = 12, Group 3 = 24**M1**

Groups 1 and 3 tie with 24

A1**Answer: Groups 1 and 3, 24****2 marks****(b)** $Q_2 - Q_1 = 6$, $Q_3 - Q_2 = 18$ **M1**

Positive skew

A1**Answer: Positive skew****2 marks****(c)** Upper limit = $48 + 1.5 \times 12 = 66$ **M1**

85 > 66

A1**Answer: Shown****2 marks****20 (4 marks)****(a)** A quarter of 120 = 30**B1****Answer: 30****1 mark****(b)** Half of 120 = 60**B1****Answer: 60****1 mark****(c)** Between Q_1 and the median is one quarter of the weeks, not half**E1**

Correct: 30 weeks, which is a quarter

B1**Answer: A quarter, 30 weeks****2 marks**

21 (9 marks)

(a) Median = $\frac{41 + 43}{2} = 42$

B1

$Q_1 = \frac{36 + 38}{2} = 37$

B1

$Q_3 = \frac{47 + 49}{2} = 48$

B1

Answer: 37, 42, 48

3 marks

(b) A: IQR = 11, upper fence = $48 + 16.5 = 64.5 < 78$, so 78 is an outlier

M1

B: IQR = 20, fences 5 and 85, and all values lie between them

A1

Answer: Shown

2 marks

(c) Group A: box 37 to 48, median 42, whiskers 31 and 52, cross at 78

B1

Group B: box 35 to 55, median 45, whiskers 22 and 64

B1

Answer: See diagram

2 marks

(d) Group B is slower on average (median 45 against 42) and more spread out (IQR 20 against 11)

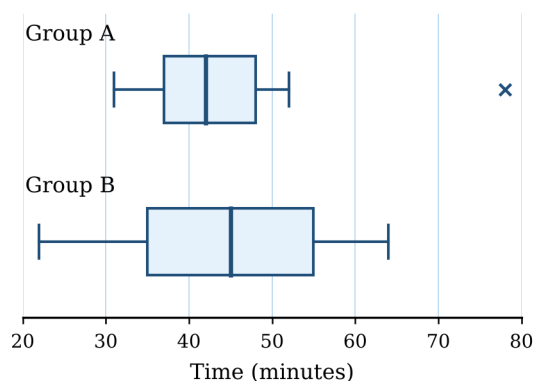
E1

The outlier 78 inflates the mean of A, so the median is the better measure

E1

Answer: Median, because of the outlier

2 marks

**22 (7 marks)**

(a) $Q_3 + 1.5(Q_3 - 34) = 64$

M1

$2.5Q_3 - 51 = 64$

M1

$Q_3 = 46$

A1

Answer: 46

3 marks

(b) Lower fence = $34 - 1.5 \times 12 = 16$

M1

$15 < 16$ so it is an outlier

A1

Answer: Yes

2 marks

(c) Median = $1.8 \times 40 + 32 = 104^\circ\text{F}$

B1

IQR = $1.8 \times 12 = 21.6^\circ\text{F}$ (the constant 32 does not affect spread)

B1

Answer: 104°F and 21.6°F

2 marks

23 (8 marks)

(a) Mean = $\frac{757}{25} = 30.28$

M1 A1

$\sigma = \sqrt{\frac{27631}{25} - 30.28^2} = 13.72$

A1

Answer: mean 30.28, σ 13.72

3 marks

(b) $95 - 30.28 = 64.72 > 2 \times 13.72 = 27.449$

B1

Answer: Shown

1 mark

(c) $\Sigma x = 662$, $\Sigma x^2 = 18606$, $n = 24$

B1

Mean = 27.5833

A1

$\sigma = 3.80$

A1

Answer: mean 27.5833, σ 3.80

3 marks

(d) The outlier increases the mean slightly but increases the standard deviation by a factor of about 3.62

E1

Answer: Inflates the SD more

1 mark

24 (7 marks)

Cumulative frequencies 10, 34, 64, 84, 96, 100

B1

 Q_1 : 25th value in $10 \leq t < 20$: $10 + \frac{15}{24} \times 10 = 16.25$

M1 A1

 Q_3 : 75th value in $30 \leq t < 40$: $30 + \frac{11}{20} \times 10 = 35.5$

A1

IQR = 19.25; upper limit = $35.5 + 1.5 \times 19.25 = 64.375$

M1 A1

Values above 64.375 are outliers, so the whisker ends at 64.375 or below (the largest value not above this limit)

E1

Answer: Q1 16.25, Q3 35.5; whisker at most 64.375

7 marks

25 (6 marks)(a) IQR = 18, upper limit = $46 + 27 = 73$

M1 A1

 $62 < 73$ so not an outlier, 90 and $96 > 73$ are outliers

A1

Answer: Verified

3 marks

(b) Median rainfall was higher in 2021 (36 against 30)

E1

IQR larger in 2021 (18 against 16) so more variable

E1

2021 had more extreme months: two outliers against one

E1

Answer: Higher, more variable, more extremes

3 marks

26 (9 marks)

(a) Median is the 12th value = 42

B1

 Q_1 = the 6th value = 31 and Q_3 = the 18th value = 50

B1

IQR = $50 - 31 = 19$

B1

Answer: median 42, IQR 19

3 marks

(b) $Q_1 - \text{IQR} = 31 - 19 = 12$

M1

Only the score 8 is below 12 (the next lowest score is 18)

A1

Answer: Shown

2 marks

(c) Lower fence = $31 - 1.5 \times 19 = 2.5$

M1

 $8 > 2.5$ so it is not an outlier

A1

Answer: No

2 marks

(d) $Q_3 - \text{median} = 8$ and $\text{median} - Q_1 = 11$

M1

The lower part is more spread out, so the data are negatively skewed

E1

Answer: Negative skew

2 marks

27 (9 marks)(a) $Q_1 = -0.6745$ and $Q_3 = 0.6745$, so IQR = 1.349

B1

Answer: 1.349

1 mark

(b) Upper fence = $0.6745 + 1.5 \times 1.349$

M1

= 2.698, and by symmetry the lower fence is -2.698

A1

Answer: ± 2.698

2 marks

(c) $P(|Z| > 2.698) = 2 \times P(Z > 2.698)$

M1

= 0.00698 (from calculator)

A1

Answer: 0.00698

2 marks

(d) Fences: $50 \pm 2.698 \times 8$

M1

28.416 and 71.584

A1

Answer: 28.416 and 71.584

2 marks

(e) Expected number = 500×0.00698

M1

= 3.488, so about 3 or 4 outliers

A1

Answer: About 3 or 4

2 marks

28 (10 marks)

$$(a) \text{ Median} = \frac{33 + 34}{2} = 33.5 \quad \text{B1}$$

$$Q_1 = \frac{28 + 29}{2} = 28.5 \quad \text{B1}$$

$$Q_3 = \frac{39 + 41}{2} = 40 \quad \text{B1}$$

Answer: 28.5, 33.5, 40 3 marks

(b) IQR = 11.5, so $1.5 \times \text{IQR} = 17.25$ and the fences are 11.25 and 57.25 M1

Only 96 lies outside, since $21 > 11.25$ and $47 < 57.25$ A1

Answer: Shown 2 marks

(c) (i) $\Sigma x = 732$ and $\Sigma x^2 = 31418$, so mean = 36.6 and $\sigma = 15.2099$ M1 A1

(ii) $\Sigma x = 636$ and $\Sigma x^2 = 22202$, so mean = 33.4737 and $\sigma = 6.931$ A1

Answer: (i) 36.6, 15.21; (ii) 33.474, 6.931 3 marks

(d) The outlier raises the mean by 3.126 and multiplies the standard deviation by about 2.19 E1

Median and IQR are barely affected, so they are the better measures to report for typical hours E1

Answer: Median and IQR 2 marks

29 (9 marks)

$$(a) \text{ IQR} = 6k - 2k = 4k \quad \text{B1}$$

$$\text{Upper limit} = 6k + 1.5 \times 4k = 12k \quad \text{M1}$$

$$\text{Lower limit} = 2k - 6k = -4k \quad \text{A1}$$

Answer: -4k 3 marks

$$(b) 50 > 12k \text{ so } k < \frac{50}{12} = 4.17 \quad \text{M1 A1}$$

$$42 \leq 12k \text{ so } k \geq 3.5 \quad \text{M1}$$

$$3.5 \leq k < \frac{25}{6} \quad \text{A1}$$

Answer: $3.5 \leq k < 25/6$ 4 marks

$$(c) k = 4 \quad \text{B1}$$

$$Q_1 = 8 \text{ and } Q_3 = 24 \quad \text{B1}$$

Answer: $k = 4$, $Q_1 = 8$, $Q_3 = 24$ 2 marks

30 (11 marks)

$$(a) \text{ Median} = 27, Q_1 = 24, Q_3 = 31 \quad \text{B1}$$

$$\text{IQR} = 7, \text{ upper limit} = 31 + 1.5 \times 7 = 41.5 \quad \text{M1 A1}$$

$$70 \text{ and } 95 \text{ both exceed } 41.5 \quad \text{A1}$$

Answer: 70 and 95 4 marks

$$(b) \Sigma x = 507, \text{ mean} = 33.8 \quad \text{M1 A1}$$

$$\Sigma x^2 = 23095, \sigma = 19.9305 \quad \text{M1 A1}$$

$$\text{Upper limit} = 33.8 + 2 \times 19.9305 = 73.6611, \text{ so only } 95 \text{ is an outlier (70 is not)} \quad \text{A1}$$

Answer: 95 only 5 marks

(c) The outliers inflate the mean and standard deviation, so 70 is hidden (masking) E1

Quartiles are not affected by extreme values, so Rule A is more reliable E1

Answer: Rule A 2 marks

Coding of Data**1 (3 marks)**

$$(a) 3, 7, 11, 2, 7 \quad \text{B1}$$

Answer: 3, 7, 11, 2, 7 1 mark

$$(b) \frac{3 + 7 + 11 + 2 + 7}{5} = 6 \quad \text{B1}$$

Answer: 6 1 mark

$$(c) x = y + 20, \text{ so mean of } x = 26 \quad \text{B1}$$

Answer: 26 1 mark

2 (3 marks)

$$x = 10y + 50$$

B1

$$\text{Mean of } x = 10 \times 2.4 + 50 = 74$$

B1

$$\text{Standard deviation of } x = 10 \times 1.3 = 13 \text{ (adding 50 does not change the spread)}$$

B1

Answer: Mean 74, sd 13

3 marks

3 (3 marks)

$$\text{Standard deviation of } x = \sqrt{16} = 4$$

B1

$$\text{Standard deviation of } y = 3 \times 4 = 12, \text{ since adding 5 has no effect on spread}$$

M1

$$\text{Variance of } z = \frac{16}{4^2} = 1$$

A1

Answer: 12 and 1

3 marks

4 (2 marks)

$$\text{Mean of } y = 2 \times 6 + 5 = 17$$

B1

$$\text{Standard deviation of } y = 2 \times 1.5 = 3$$

B1

Answer: 17 and 3

2 marks

5 (2 marks)

$$\text{Divide by 100: mean} = 1.65 \text{ m}$$

B1

$$\text{Standard deviation} = 0.08 \text{ m}$$

B1

Answer: 1.65 m and 0.08 m

2 marks

6 (2 marks)

$$\text{Mean of } y = \frac{48}{12} = 4$$

M1

$$\text{Mean of } x = 5 \times 4 = 20$$

A1

Answer: 20

2 marks

7 (4 marks)

$$\text{Mean of } (x - 100) = \frac{-35}{20} = -1.75, \text{ so mean of } x = 98.25$$

M1 A1

$$\text{Variance} = \frac{500}{20} - (-1.75)^2 = 25 - 3.0625 = 21.9375$$

M1

$$\text{Standard deviation} = 4.68$$

A1

Answer: Mean 98.25, sd 4.68

4 marks

8 (4 marks)

$$\text{(a) Mean} = 1.8 \times 17.5 + 32 = 63.5 \text{ } ^\circ\text{F}$$

B1

$$\text{Standard deviation} = 1.8 \times 3.2 = 5.76 \text{ } ^\circ\text{F}$$

B1

Answer: 63.5 °F and 5.76 °F

2 marks

$$\text{(b) } C = \frac{F - 32}{1.8}, \text{ so mean} = \frac{59 - 32}{1.8} = 15 \text{ } ^\circ\text{C}$$

B1

$$\text{Standard deviation} = \frac{9}{1.8} = 5 \text{ } ^\circ\text{C}$$

B1

Answer: 15 °C and 5 °C

2 marks

9 (4 marks)

$$\text{Mean of } y = \frac{53 - a}{4} = 2.5$$

M1

$$53 - a = 10, \text{ so } a = 43$$

A1

$$\text{Standard deviation of } x = 4 \times 1.8 = 7.2$$

B1

$$\text{Variance of } x = 7.2^2 = 51.84$$

B1

Answer: a = 43, σ = 7.2, variance 51.84

4 marks

10 (5 marks)

$$\Sigma f y = -3, \Sigma f y^2 = 51, n = 35$$

M1

$$\text{Mean of } y = -0.0857$$

A1

$$\text{Variance of } y = 1.4571 - (-0.0857)^2 = 1.4498, \text{SD of } y = 1.2041$$

M1 A1

$$\text{Mean of } x = 150 + 5(-0.0857) = 149.5714 \text{ mm; SD of } x = 5 \times 1.2041 = 6.02 \text{ mm}$$

A1

Answer: mean 149.5714 mm, SD 6.02 mm

5 marks

11 (3 marks)

$$\text{New mean} = 1.2 \times 25 = 30$$

B1

$$\text{New SD} = 1.2 \times 4 = 4.8$$

M1

$$\text{Variance} = 4.8^2 = 23.04$$

A1

Answer: mean £30, variance 23.04

3 marks

12 (3 marks)

$$z = \frac{57 - 45}{6} = 2$$

M1 A1

$$-1.5 = \frac{x - 45}{6} \text{ gives } x = 45 - 9 = 36$$

A1

Answer: z = 2; x = 36

3 marks

13 (3 marks)

$$\text{Median of } y = 3 \times 14 - 2 = 40$$

M1 A1

$$\text{IQR of } y = 3 \times 6 = 18 \text{ (the constant has no effect on spread)}$$

A1

Answer: median 40, IQR 18

3 marks

14 (6 marks)

$$\text{(a) } y = -2, -1, 0, 1, 2: \Sigma f y = -8 - 11 + 0 + 12 + 8 = 1$$

B1

$$\Sigma f y^2 = 16 + 11 + 0 + 12 + 16 = 55$$

B1

$$\text{Answer: } \Sigma f y = 1, \Sigma f y^2 = 55$$

2 marks

$$\text{(b) Mean of } y = \frac{1}{50} = 0.02$$

M1

$$x = 10y + 30, \text{ so mean of } x = 30.2 \text{ g}$$

A1

Answer: 30.2 g

2 marks

$$\text{(c) Variance of } y = \frac{55}{50} - 0.02^2 = 1.0996$$

M1

$$\text{sd of } x = 10 \times \sqrt{1.0996} = 10.5 \text{ g}$$

A1

Answer: 10.5 g

2 marks

15 (6 marks)

$$\text{Mean of } y = \frac{250}{25} = 10$$

B1

$$4\bar{x} - 30 = 10, \text{ so } \bar{x} = 10 \text{ seconds}$$

M1 A1

$$\text{Variance of } y = \frac{3400}{25} - 10^2 = 36, \text{ so sd of } y = 6$$

M1

$$y = 4x - 30 \text{ so sd of } y = 4 \times \text{sd of } x$$

M1

$$\text{sd of } x = \frac{6}{4} = 1.5 \text{ seconds}$$

A1

Answer: Mean 10 s, sd 1.5 s

6 marks

16 (7 marks)

$$\text{(a) Mean of } (x - 30) = 2.5, \text{ so mean of } x = 32.5$$

M1 A1

$$\text{Variance} = \frac{140}{10} - 2.5^2 = 7.75$$

M1 A1

Answer: mean 32.5, variance 7.75

4 marks

$$\text{(b) } \Sigma x = \Sigma(x - 30) + 10 \times 30 = 25 + 300 = 325$$

B1

$$\Sigma(x - 30)^2 = \Sigma x^2 - 60 \Sigma x + 900n$$

M1

$$\Sigma x^2 = 140 + 60 \times 325 - 9000 = 10640$$

A1

Answer: $\Sigma x = 325$, $\Sigma x^2 = 10640$

3 marks

17 (6 marks)

$$(a) \text{ Mean} = \frac{5}{9}(68.9 - 32) = \frac{5}{9} \times 36.9$$

M1

$$= 20.5 \text{ }^{\circ}\text{C}$$

A1

$$\text{SD} = \frac{5}{9} \times 7.2$$

M1

$$= 4.00 \text{ }^{\circ}\text{C}$$

A1

Answer: 20.5 °C and 4.00 °C

4 marks

$$(b) \frac{5}{9}(83.3 - 32) = 28.5 \text{ }^{\circ}\text{C}$$

M1

$$\frac{83.3 - 68.9}{7.2} = 2 \text{ standard deviations}$$

A1

Answer: 28.5 °C; 2 SD

2 marks

18 (5 marks)

$$(a) \Sigma y = 410 - 20 \times 20 = 10$$

B1

$$\Sigma y^2 = \Sigma x^2 - 40 \Sigma x + 400 \times 20$$

M1

$$= 8800 - 16\,400 + 8000 = 400$$

A1

Answer: $\Sigma y = 10$, $\Sigma y^2 = 400$

3 marks

$$(b) \text{ Variance of } y = \frac{400}{20} - \left(\frac{10}{20}\right)^2 = 19.75$$

M1

Adding a constant does not change the variance: variance of $x = 19.75$

A1

Answer: 19.75 for both

2 marks

19 (5 marks)

$$\text{Mean of } y = \frac{-16}{40} = -0.4$$

B1

$$\text{Variance of } y = \frac{210}{40} - (-0.4)^2 = 5.09$$

M1 A1

$$\text{Mean of } x = 60 + 5(-0.4) = 58$$

M1

$$\text{SD of } x = 5\sqrt{5.09} = 11.3$$

A1

Answer: mean 58, SD 11.3

5 marks

20 (8 marks)

$$(a) \text{ Coded midpoints are } -2, -1, 0, 1, 2: \Sigma fy = -18 - 31 + 0 + 38 + 40 = 29$$

M1

$$\Sigma fy^2 = 36 + 31 + 0 + 38 + 80 = 185$$

A1

$$\text{Mean of } y = \frac{29}{150} = 0.1933 \text{ and } \sigma_y = \sqrt{\frac{185}{150} - (0.1933)^2} = 1.0936$$

A1

Answer: mean 0.1933, $\sigma = 1.0936$

3 marks

$$(b) x = 1.05 + \frac{y}{50}, \text{ so mean of } x = 1.05 + \frac{0.1933}{50} = 1.05387$$

M1

$$\sigma_x = \frac{1.0936}{50} = 0.022$$

A1

Answer: mean 1.0539 m, $\sigma = 0.022$ m

2 marks

$$(c) \text{ Median is the 75th value, in } 1.04 \text{ to } 1.06: 1.04 + \frac{75 - 40}{52} \times 0.02 = 1.05346$$

M1 A1

$$\text{Coefficient} = \frac{3(1.05387 - 1.05346)}{0.022} = 0.06, \text{ which is almost zero, so the distribution is nearly}$$

A1

symmetrical

Answer: median 1.0535 m; almost symmetrical

3 marks

21 (7 marks)

$$(a) \text{ sd of } y = a \times \text{sd of } x: 10 = 4a$$

M1

$$a = 2.5$$

A1

$$100 = 2.5 \times 12 + b$$

M1

$$b = 70$$

A1

Answer: $a = 2.5$, $b = 70$

4 marks

$$(b) y = 2.5 \times 18 + 70 = 115$$

B1

Answer: 115**1 mark**(c) $130 = 2.5x + 70$, so $x = 24$, and $\frac{24 - 12}{4} = 3$ **M1** $\frac{130 - 100}{10} = 3$ also**A1****Answer: $x = 24$, both 3 standard deviations****2 marks****22 (7 marks)**(a) $4.52 - 0.3 = 4.22$ kg**B1****Answer: 4.22 kg****1 mark**

(b) 0.40 kg

B1

Subtracting the same constant from every value does not change the spread

E1**Answer: 0.40 kg****2 marks**(c) Mean = $4.22 \times 2.2 = 9.284$ lb**B1**Standard deviation = $0.40 \times 2.2 = 0.88$ lb**B1****Answer: 9.284 lb and 0.88 lb****2 marks**(d) Correct median = $4.45 - 0.3 = 4.15$ kg**M1** $4.15 \times 2.2 = 9.13$ lb**A1****Answer: 9.13 lb****2 marks****23 (7 marks)**(a) Group 1: $50 + \frac{40}{20} = 52$ **B1**Group 2: $50 - \frac{30}{30} = 49$ **B1****Answer: 52 and 49****2 marks**(b) $\Sigma(x - 50) = 40 - 30 = 10$, $\Sigma(x - 50)^2 = 1300$, $n = 50$ **M1 A1**Mean of $x = 50 + \frac{10}{50} = 50.2$ **A1**Variance = $\frac{1300}{50} - 0.2^2 = 25.96$ **M1** $\sigma = 5.10$ **A1****Answer: mean 50.2, σ 5.10****5 marks****24 (5 marks)**(a) Mean of $y = 10 - 3 \times 4 = -2$ **M1 A1**Variance of $y = (-3)^2 \times 4 = 36$ **A1****Answer: mean -2, variance 36****3 marks**(b) Standard deviation is a square root of a variance, so it cannot be negative; it scales by $|-3| = 3$ **E1**

Correct value: 6

B1**Answer: 6****2 marks****25 (7 marks)**(a) Mean of $y = 2 \times 8 - 7 = 9$ **M1 A1**Fifth value: $45 - 44 = 1$ **M1** $x = \frac{y+7}{2}$: 6, 8, 10, 12, 4**A1****Answer: $y = 1$; $x = 6, 8, 10, 12, 4$** **4 marks**(b) $\Sigma(x - 8)^2 = 4 + 0 + 4 + 16 + 16 = 40$ **M1 A1** $\sigma = \sqrt{40/5} = 2.83$ **A1****Answer: 2.83****3 marks****26 (8 marks)**(a) $\Sigma x = 6n$ **B1**Variance = $\frac{\Sigma x^2}{n} - 6^2 = 4$ **M1** $\Sigma x^2 = 40n$ **A1****Answer: $\Sigma x = 6n$, $\Sigma x^2 = 40n$** **3 marks**

$$\begin{aligned} \text{(b)} \quad \Sigma(x - 2)^2 &= \Sigma x^2 - 4\Sigma x + 4n \\ &= 40n - 24n + 4n = 20n \end{aligned}$$

Answer: Shown

M1
A1
2 marks

$$\begin{aligned} \text{(c)} \quad \Sigma(x - k)^2 &= n[4 + (6 - k)^2] \\ 4 + (6 - k)^2 &= 20 \text{ so } (6 - k)^2 = 16 \\ k &= 2 \text{ or } k = 10 \end{aligned}$$

Answer: k = 2 or k = 10

M1
M1
A1
3 marks

27 (9 marks)

$$\text{(a)} \quad \Sigma y = \Sigma(ax + b) = a\Sigma x + nb$$

$$\text{Dividing by } n: \bar{y} = a\bar{x} + b$$

Answer: Shown

M1
A1
2 marks

$$\text{(b)} \quad y - \bar{y} = ax + b - a\bar{x} - b = a(x - \bar{x})$$

$$\begin{aligned} \text{Variance of } y &= \frac{\Sigma[a(x - \bar{x})]^2}{n} = a^2 \times \frac{\Sigma(x - \bar{x})^2}{n} \\ &= a^2 \sigma^2 \end{aligned}$$

Answer: Shown

M1
M1
A1
3 marks

$$\text{(c)} \quad \text{Mean} = \frac{1260}{40} = 31.5$$

$$\text{Variance} = \frac{41\,130}{40} - 31.5^2 = 1028.25 - 992.25 = 36, \text{ so sd} = 6$$

Answer: Mean 31.5, sd 6

B1
M1
2 marks

$$\text{(d)} \quad \text{sd: } a \times 6 = 1, \text{ so } a = \frac{1}{6}$$

$$\text{Mean: } \frac{1}{6} \times 31.5 + b = 0, \text{ so } b = -5.25$$

Answer: a = $\frac{1}{6}$, b = -5.25

M1
A1
2 marks

28 (10 marks)

$$\text{(a)} \quad \text{Mean of } (x - c) \text{ is } \frac{60}{n}$$

$$\text{Variance of } x = \text{variance of } (x - c) = \frac{375}{n} - \left(\frac{60}{n}\right)^2$$

$$\text{This equals } 3^2 = 9$$

$$\text{Multiplying by } n^2: 375n - 3600 = 9n^2$$

$$\text{Rearranging and dividing by 3: } 3n^2 - 125n + 1200 = 0$$

Answer: Shown

M1
M1
A1
M1
A1
5 marks

$$\text{(b)} \quad (3n - 80)(n - 15) = 0$$

$$n = 15 \text{ or } n = \frac{80}{3}$$

n must be a whole number, so n = 15

Answer: n = 15

M1
A1
3 marks

$$\text{(c)} \quad \text{Mean of } (x - c) = \frac{60}{15} = 4$$

$$24 - c = 4, \text{ so } c = 20$$

Answer: c = 20

M1
A1
2 marks

29 (10 marks)

$$\text{(a)} \quad z = \frac{1}{\sigma}x - \frac{\mu}{\sigma}, \text{ so mean of } z = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0$$

$$\text{Variance of } z = \left(\frac{1}{\sigma}\right)^2 \sigma^2$$

= 1, so the standard deviation is 1

Answer: Shown

M1 A1
M1
A1
4 marks

$$\text{(b)} \quad \text{Maths: } z = \frac{74 - 62}{12} = 1$$

M1 A1

$$\text{Physics: } z = \frac{63 - 54}{6} = 1.5$$

A1

She did relatively better in Physics, being 1.5 standard deviations above the mean

E1

Answer: Physics (z = 1.5 against 1)

4 marks

(c) $62 + 1.5 \times 12$

M1

$= 80$

A1

Answer: 80

2 marks

30 (10 marks)

(a) SD: $12 = 8a$ so $a = 1.5$

M1 A1

Mean: $60 = 1.5 \times 48 + b$

M1

$b = -12$

A1

Answer: a = 1.5, b = -12

4 marks

(b) $m = 1.5m - 12$ gives $0.5m = 12$, so $m = 24$

M1 A1

Students scoring above 24 gain marks and those below lose marks

E1

Answer: 24

3 marks

(c) $1.5m - 12 \leq 100$

M1

$m \leq 74.6667$, so the highest original mark is 74

A1

Higher original marks would exceed 100, so the adjustment cannot be applied without capping, which would change the mean and SD

E1

Answer: 74

3 marks

Large Data Sets

1 (3 marks)

(a) Continuous quantitative

B1

(b) Discrete quantitative

B1

(c) Qualitative (categorical)

B1

Answer: (a) continuous; (b) discrete; (c) qualitative

3 marks

2 (3 marks)

(a) The two "tr" entries are not exact values, only known to lie between 0 and 0.05 mm

E1

Answer: "tr" is not an exact value

1 mark

(b) Known values: $0 + 2.4 + 0.2 + 11.6 + 0 + 5.0 = 19.2$ mm, so the least total is just over 19.2 mm

B1

Each "tr" adds less than 0.05, so the total is less than $19.2 + 0.1 = 19.3$ mm

B1

Answer: Between 19.2 mm and 19.3 mm

2 marks

3 (3 marks)

(a) $\frac{6}{8} = \frac{3}{4}$

B1

Answer: $\frac{3}{4}$

1 mark

(b) The value is impossible, so it is an error in the data

E1

Check against the source or remove it and treat the value as missing

B1

Answer: Error: check or remove

2 marks

4 (2 marks)

Wednesday

B1

The maximum (9.8) is below the minimum (15.3), which cannot happen

E1

Answer: Wednesday: max below min

2 marks

5 (4 marks)

$$\Sigma x = 42.0, \text{ so mean} = \frac{42}{8} = 5.25 \text{ hours}$$

M1 A1

$$\text{Ordered: } 0, 0.5, 2.3, 3.4, 6.2, 8.1, 9.9, 11.6, \text{ so median} = \frac{3.4 + 6.2}{2} = 4.8 \text{ hours}$$

B1

4 of the 8 days had at least 6 hours, which is 50%

B1

Answer: mean 5.25 hours, median 4.8 hours, 50%

4 marks

6 (2 marks)

Wet days (≥ 1 mm): 3.2, 7.5, 1.0, 12.6, 1.8, 4.4, 2.2, 5.6, which is 8 days

M1

$$\frac{8}{20} \times 100 = 40\%$$

A1

Answer: 40%

2 marks

7 (4 marks)

(a) $\Sigma x = 247.5$

M1

$$\text{Mean} = \frac{247.5}{12} = 20.625 \approx 20.6 \text{ }^{\circ}\text{C}$$

A1

Answer: 20.6 }^{\circ}\text{C}

2 marks

(b) Ordered, the 6th and 7th values are 20.1 and 20.7: median = 20.4 }^{\circ}\text{C}

B1

Answer: 20.4 }^{\circ}\text{C}

1 mark

(c) The days are consecutive in one month, so they are not a random or representative sample of the whole six months (summer is warmer than May or October)

E1

Answer: Not representative of the six months

1 mark

8 (4 marks)

(a) Leave them out, and divide by the number of valid readings (not 7)

B1

Answer: Exclude them

1 mark

(b) 61.2 }^{\circ}\text{C is impossible for a daily mean temperature in the UK}

B1

Remove it (or check the source for a typing error such as 16.2 or 6.12)

B1

Answer: 61.2: remove or correct

2 marks

(c) $\frac{14.3 + 15.1 + 13.9 + 16.0}{4} = 14.825 \text{ }^{\circ}\text{C}$

B1

Answer: 14.825 }^{\circ}\text{C}

1 mark

9 (3 marks)

(a) Sum = 169.8

M1

$$\text{Mean} = 16.98 \text{ }^{\circ}\text{C}$$

A1

Answer: 16.98 }^{\circ}\text{C}

2 marks

(b) The sample mean differs by 0.18 because samples vary; it is close to the population mean

E1

Answer: Sampling variability

1 mark

10 (4 marks)

Reject 42 (over 31) and the second 17 (repeat)

B1

Days chosen: 17, 8, 31 and 29

B1 B1

The first four days are not random and may all share one weather pattern, so are unrepresentative

E1

Answer: 17, 8, 31, 29

4 marks

11 (3 marks)

The inland station is warmer on average (19.4 against 17.8)

E1

Its temperatures vary more (SD 4.6 against 3.1)

E1

The sea moderates temperatures at the coast, giving lower and steadier maxima

E1

Answer: Inland warmer, more variable

3 marks

12 (4 marks)

Mode = 8 (14 days)

B1

Cumulative frequencies 1, 1, 3, 6, 8, 11, 17, 26 so the 20th and 21st values are both 7

M1

Median = 7

A1

$$\text{Mean} = \frac{247}{40} = 6.175$$

B1

Answer: mode 8, median 7, mean 6.175

4 marks

13 (4 marks)

$$\text{Each day is } \frac{360}{36} = 10^\circ$$

M1

Angles: 60° , 150° , 100° , 50°

A1

Days with at least 20°C : $10 + 5 = 15$

M1

$$\frac{15}{36} \times 100 = 41.667\%$$

A1

Answer: 60° , 150° , 100° , 50° ; 41.667%

4 marks

14 (6 marks)

$$\text{(a) Days are 3, 10, 17, ..., } 3 + 7k \leq 184, \text{ so } k \leq \frac{181}{7} = 25.86$$

M1

 $k = 0, 1, \dots, 25$, giving 26 days

A1

Answer: 26

2 marks

$$\text{(b) Last day} = 3 + 7 \times 25 = 178$$

M1

Days up to 30 September are numbered 1 to 153, so day 178 is 25 October

A1

Answer: 25 October

2 marks

(c) Every selected day is the same day of the week, and the sample comes from only one season and one year

E1

It comes from one weather station only, which may not represent the whole region

E1

Answer: One station, one year, fixed weekday

2 marks

15 (6 marks)

$$\text{(a) A: } \frac{21.0 - 15.3}{3.1} = 1.839$$

M1 A1

$$\text{B: } \frac{18.5 - 13.8}{2.4} = 1.958$$

A1

Answer: A: 1.839, B: 1.958

3 marks

(b) B's reading (it is further above its mean in standard deviations)

B1

Answer: Station B's reading

1 mark

(c) The means only describe typical values, not every day

E1

The spreads are large compared with the 1.5°C difference in means, so the daily values overlap (for example B's 18.5°C is above A's mean)

E1

Answer: Overlapping distributions

2 marks

16 (5 marks)

(a) Strong positive linear correlation

B1

More sunshine is associated with higher maximum temperatures

E1

Answer: Strong positive

2 marks

(b) There may be a recording error, or a cold wind/early frost kept temperatures low despite the sunshine

E1

Check the source data before deciding whether to include it

B1

Answer: Check source; possible cold wind or error

2 marks

(c) Both may be linked to a third factor such as season or air mass, so correlation does not imply causation

E1

Answer: Correlation is not causation

1 mark

17 (6 marks)

(a) Mean = $20 + \frac{62}{31} = 22$

M1 A1

Variance = $\frac{412}{31} - 2^2 = 9.2903$

M1

$\sigma = 3.05$ °C

A1

Answer: mean 22 °C, σ 3.05 °C

4 marks

(b) September was cooler on average by 3.80 °C

E1

September's temperatures were less spread out (2.1 against 3.05)

E1

Answer: Cooler and less variable

2 marks

18 (5 marks)

(a) Rain days: $38 + 6 + 14 = 58$, so $\frac{58}{184} = 0.315$

B1

Answer: 0.315

1 mark

(b) Westerly days: $38 + 46 = 84$

M1

$\frac{38}{84} = 0.452$

A1

Answer: 0.452

2 marks

(c) Independent if $P(\text{rain} | \text{westerly}) = P(\text{rain})$

M1

$0.452 \neq 0.315$, so not independent: rain is more likely on westerly days

A1

Answer: Not independent

2 marks

19 (5 marks)

(a) IQR = 12

B1

Upper limit = $21 + 1.5 \times 12 = 39$

M1

$48.5 > 39$ and $52.0 > 39$

A1

Answer: Shown

3 marks

(b) Exceptionally clear air after rain or a change of air mass

E1

They are genuine readings so should not be removed without evidence of error

E1

Answer: Genuine; keep

2 marks

20 (4 marks)

(a) $16.2 - 15.5 = 0.7$

M1

$\frac{0.7}{15.5} \times 100 = 4.516\%$

A1

Answer: 4.516%

2 marks

(b) One year at one station is too little data to show a trend

E1

A year-to-year variation of this size can occur by chance (sampling variability)

E1

Answer: One year, one station

2 marks

21 (8 marks)

(a) Median = $\frac{0.2 + 0.4}{2} = 0.3$

B1

Q_1 : lower half has nine zeros and 0.2, so $Q_1 = 0$

B1

$Q_3 = \frac{3.4 + 5.6}{2} = 4.5$

B1

Answer: $Q_1 = 0$, median 0.3, $Q_3 = 4.5$

3 marks

(b) IQR = 4.5, upper fence = $4.5 + 6.75 = 11.25$ (the lower fence is negative)

M1

Outliers are 14.6 and 21.0

A1

Answer: 14.6 and 21.0

2 marks

(c) At least a quarter of the days had no measurable rain

B1

Answer: At least 25% of days were dry

1 mark

(d) The lower quartile coincides with the minimum so the lower whisker has no length, and the data are heavily skewed

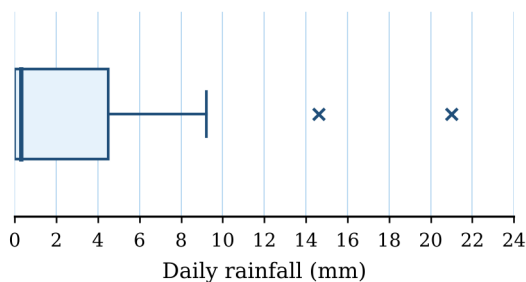
E1

Give the proportion of dry days, $\frac{9}{20} = 45\%$

E1

Answer: Report the proportion of dry days

2 marks

**22 (8 marks)**

(a) $\frac{41 + 33}{184} = \frac{74}{184} = 0.402$

B1**Answer: 0.402****1 mark**

(b) $P(\text{rain} | \text{cool}) = \frac{33}{64}$

M1

$= 0.516$

A1**Answer: 0.516****2 marks**

(c) If independent, $P(\text{rain} | \text{cool})$ would equal $P(\text{rain}) = 0.402$

M1

$0.516 \neq 0.402$, so they are not independent: cool days are more likely to be wet

A1**Answer: Not independent****2 marks**

(d) $\frac{74}{184} \times \frac{73}{183}$

M1

$= 0.16$

A1**Answer: 0.16****2 marks**

(e) The data only cover May to October

E1**Answer: Only six months of the year****1 mark****23 (6 marks)**

(a) Sampling fraction $= \frac{36}{184}$

B1

May: $\frac{36}{184} \times 31 = 6.065$; June: $\frac{36}{184} \times 30 = 5.87$

M1

Both round to 6, and $6 \times 6 = 36$

A1**Answer: Shown****3 marks**

(b) Number the May days 1 to 31

B1

Use random numbers to select 6 different days

B1**Answer: Random selection from 31 numbered days****2 marks**

(c) $\frac{184}{36} = 5.1...$ is not an integer

E1**Answer: 184/36 is not an integer****1 mark****24 (6 marks)**

(a) $28 + 71 + 56 + 13 = 168$

B1**Answer: 168****1 mark**

(b) Midpoints 5, 15, 30, 50

M1

$\Sigma f x = 140 + 1065 + 1680 + 650 = 3535$

M1

Mean $\approx \frac{3535}{168} = 21.042$ km

A1**Answer: 21.042 km****3 marks**

(c) Days may be missing for a reason connected with visibility (for example equipment failure in fog)

E1

Exclude them from the calculation and state that this was done, rather than counting them as zero

E1**Answer: Exclude and state****2 marks**

25 (7 marks)

(a) May: $\frac{62.4}{31} = 2.013$ mm; $\frac{62.4}{9} = 6.933$ mm per wet day M1 A1

October: $\frac{118.9}{31} = 3.835$ mm; $\frac{118.9}{14} = 8.493$ mm per wet day M1 A1

Answer: May 2.01, 6.93; Oct 3.84, 8.49 4 marks

(b) October had more wet days (14 against 9), more rain per wet day and more total rainfall E1

Only one year and one station are used E1

Rainfall varies from year to year, so this may be due to chance E1

Answer: Supported by this data only 3 marks

26 (7 marks)

(a) Mean = $\frac{1012}{92} = 11$ B1

Variance = $\frac{13100}{92} - 11^2 = 21.3913$ M1

$\sigma = 4.63$ knots A1

Answer: mean 11, σ 4.63 3 marks

(b) Mean = $1.15 \times 11 = 12.65$ mph M1

$\sigma = 1.15 \times 4.63 = 5.32$ mph A1

Answer: mean 12.65 mph, σ 5.32 mph 2 marks

(c) $\frac{24 - 11}{4.63}$ M1

= 2.811 standard deviations A1

Answer: 2.811 2 marks

27 (10 marks)

(a) Midpoints 2, 6, 10, 16: $\Sigma fx = 20 + 192 + 300 + 320 = 832$ M1

Mean = $\frac{832}{92} = 9.0435$ knots A1

$\Sigma fx^2 = 40 + 1152 + 3000 + 5120 = 9312$ M1

$\sigma = \sqrt{\frac{9312}{92} - 9.0435^2} = 4.4083$ knots A1

Answer: Mean 9.043 knots, sd 4.408 knots 4 marks

(b) Mean = $1.15 \times 9.0435 = 10.4$ mph B1

sd = $1.15 \times 4.4083 = 5.0695$ mph B1

Answer: 10.4 mph and 5.07 mph 2 marks

(c) Σx for B = $92 \times 7.5 = 690$, so mean = $\frac{832 + 690}{184} = 8.2717$ M1 A1

Σx^2 for B = $92(3.6^2 + 7.5^2) = 6367.32$ M1

$\sigma = \sqrt{\frac{9312 + 6367.32}{184} - 8.2717^2} = 4.0978$ knots A1

Answer: Mean 8.272 knots, sd 4.098 knots 4 marks

28 (9 marks)

(a) Camborne: $\frac{78}{184} = 42.391\%$ B1

Leuchars: $\frac{61}{184} = 33.152\%$ B1

Answer: 42.391% and 33.152% 2 marks

(b) Camborne has the greater median (3.0 mm against 2.1 mm), so wet days there are typically wetter E1

Camborne's IQR (5.6 mm) is larger than Leuchars's (3.6 mm) and it has more and larger outliers, so it is more variable E1

Answer: Camborne wetter and more variable 2 marks

(c) Upper fence = $4.4 + 1.5 \times 3.6 = 9.8$ M1

$12.4 > 9.8$ so it is an outlier, and the whisker ends at $9.5 < 9.8$ A1

Answer: Shown 2 marks

(d) The outliers (up to 31.4 mm) would raise the mean a lot, whereas the median is not affected

E1

Answer: The mean is distorted by outliers

1 mark

(e) The data only cover May to October

E1

Only one year (and one station for each place) is used, which may not be typical

E1

Answer: Only six months and one year

2 marks

29 (12 marks)

(a) Rain today: $45 + 32 = 77$, so $\frac{77}{183} = 0.421$

M1 A1

Rain yesterday: $45 + 33 = 78$, so $\frac{45}{78} = 0.577$

M1 A1

Answer: 0.421 and 0.577

4 marks

(b) $\frac{32}{105} = 0.305$

M1 A1

These three probabilities differ, so successive days are not independent: rain tends to persist

E1

Answer: 0.305

3 marks

(c) Independent model: $0.42^7 = 0.002$

M1 A1

Dependent model: $0.42 \times 0.577^6 = 0.015$

M1 A1

The dependent model gives a probability about 6.72 times greater, so long wet spells are more likely than independence suggests

E1

Answer: 0.002 and 0.015

5 marks

30 (9 marks)

(a) IQR = 6.5, upper limit = $18.5 + 9.75 = 28.25$

M1 A1

$24 < 28.25$ so it is not an outlier; 29.5 and 31.0 > 28.25 are outliers

A1

Answer: Verified

3 marks

(b) Median: C is 2.5 °C higher (17.5 against 15.0)

E1

IQR: C = 5.0, L = 6.5, so L is more variable

B1 E1

Answer: C warmer, L more variable

3 marks

(c) "Always" is wrong: the boxes overlap (Q_3 of L is 18.5 and Q_1 of C is 15), so many days at L are warmer than days at C

E1

Outliers at L (29.5, 31.0) exceed the maximum at C's whisker

E1

Use paired same-day differences or a test over several years

E1

Answer: Not justified

3 marks