



Cambridge O Level

CANDIDATE
NAME

CENTRE
NUMBER

--	--	--	--	--

CANDIDATE
NUMBER

--	--	--	--

STATISTICS

4040/12

Paper 1

October/November 2021

2 hours 15 minutes

You must answer on the question paper.

You will need: Calculator
Pair of compasses
Protractor

INSTRUCTIONS

- Answer **all** questions.
- Use a black or dark blue pen. You may use an HB pencil for any diagrams or graphs.
- Write your name, centre number and candidate number in the boxes at the top of the page.
- Write your answer to each question in the space provided.
- Do **not** use an erasable pen or correction fluid.
- Do **not** write on any bar codes.
- You should use a calculator where appropriate.
- You must show all necessary working clearly.

INFORMATION

- The total mark for this paper is 100.
- The number of marks for each question or part question is shown in brackets [].

This document has **16** pages.



1 Seven terms which are used in statistical analysis are

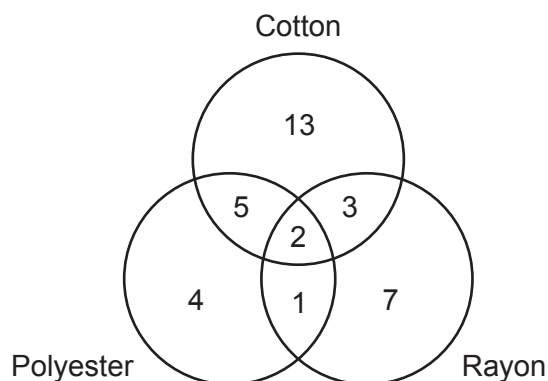
classification,
interpolation,
transformation,
population,
correlation,
expectation,
and representation.

In each of the following statements one of these terms has been omitted.
Complete each statement by inserting the appropriate term.

- (a) When a census is conducted, information is collected on every item in the [1]
- (b) A scatter diagram will display the amount of between the variables in a bivariate distribution. [1]
- (c) To calculate the median from a grouped frequency distribution table, the method of linear can be used. [1]
- (d) A bar chart is preferred to a histogram for the pictorial of discrete quantitative data. [1]
- (e) To compare a pupil's performance in two examinations, relative to all the pupils, her scores can be converted by linear to a given mean and standard deviation. [1]

3

- 2 A manufacturer designs shirts made from various fabrics. The diagram shows the number of designs made using one or more of the fabrics cotton, polyester and rayon.



- (a) Interpret the value 2 in the diagram.

.....
 [1]

- (b) Find the number of designs made using

- (i) rayon,

..... [1]

- (ii) cotton or polyester or both,

..... [1]

- (iii) exactly two of these fabrics.

..... [1]

- (c) How many more designs are made using cotton than are made using polyester?

..... [1]

- (d) Of the designs made using polyester, find the percentage that also use rayon.

..... [1]

- 3 A perfume company is creating a new product with the intention that 20% of its volume will be aromatic oils. It is important that this is achieved in production, and that any variation from it is as small as possible, to ensure the consistency of the product.

Five 100 millilitre (ml) bottles are tested, and the exact amounts of aromatic oils they contain, in millilitres, are as follows.

20.1 19.8 19.7 21.3 20.8

The company's chemist rules that, for the volume of aromatic oils in a test sample of five such bottles:

- the mean should not be larger than 20.5 ml, and not smaller than 19.5 ml
- the standard deviation should not be larger than 0.62 ml.

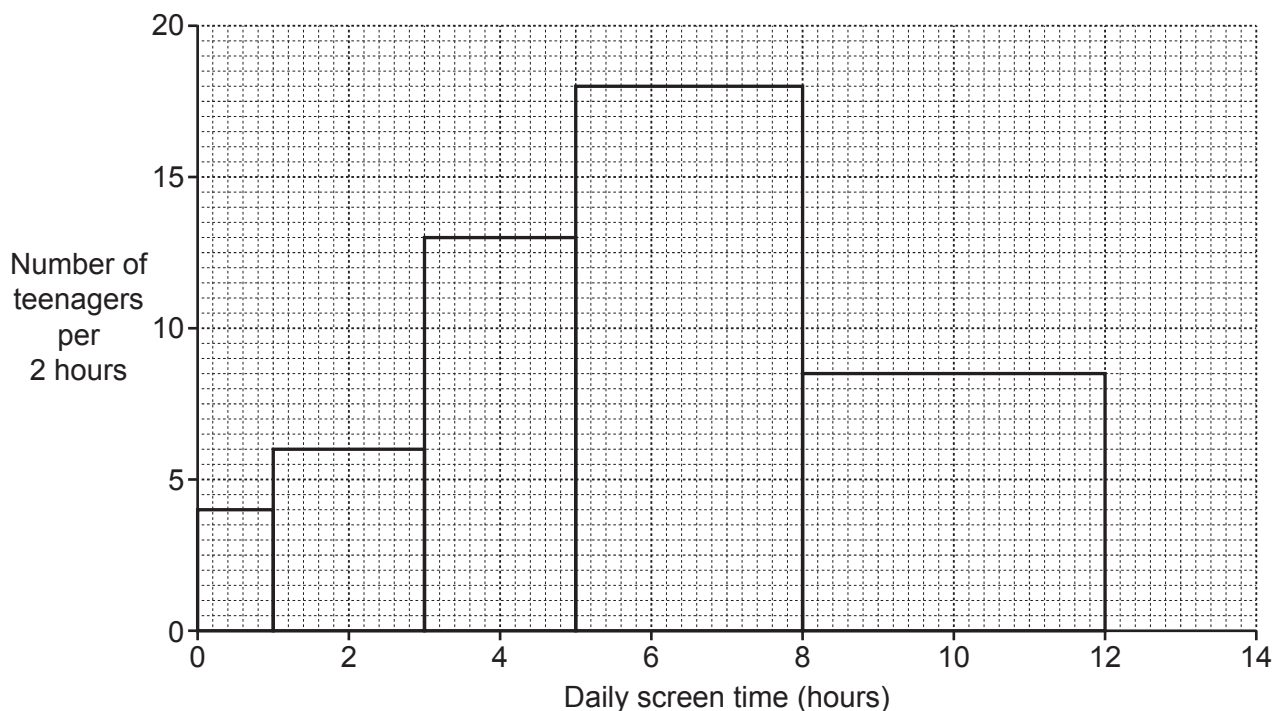
- (a) Does this sample satisfy the chemist's ruling?

..... [5]

- (b) Explain whether or not the chemist should also rule that the standard deviation of such a test sample should not be **smaller** than a certain amount.

.....
 [1]

- 4 Leila is a sociologist. She surveys the amount of time teenagers spend each day looking at screens (television, computers, smartphones etc.). The following histogram summarises her results for the recorded 'daily screen time' of a sample of teenagers.



- (a) Use the histogram to estimate the number of teenagers whose daily screen time is

- (i) up to 3 hours,

..... [2]

- (ii) from 6 hours up to 12 hours.

..... [3]

The teenagers themselves had recorded their own estimates of daily screen time for answers to Leila's survey. She suspects that they will have underestimated the values by one hour.

- (b) Estimate the number of teenagers whose daily screen time is from 3 hours up to 6 hours, if Leila's suspicion is correct.

..... [2]

- 5 The following table shows the number of university applications made by 39 final-year pupils at a particular school.

<i>Number of university applications, x</i>	0	1	2	3	4	5 or more
<i>Number of pupils, f</i>	7	5	6	8	4	9

For this distribution,

- (a) name a measure of dispersion which **cannot** be found exactly,

..... [1]

- (b) name a measure of dispersion which **can** be found exactly,

..... [1]

- (c) name, and find the value of, a measure of central tendency which **can** be found exactly.

Name

Value

[3]

Three pupils were absent when the data were collected. All had made no university applications. This extra data is now included.

- (d) For the 42 pupils, find the value of the measure of central tendency you have named in part (c).

..... [2]

- (e) For the 42 pupils, there is a measure of central tendency which can now be found that could **not** be found for the original 39 pupils.

Name, and state the value of, this measure of central tendency.

Name

Value

[1]

- 6** In 3×3 basketball there are three players on each of two teams.
The manager of Stats Chancers basketball team has a squad of 12 players. She categorises them into three types according to their most outstanding characteristic: there are 4 strong players, 5 creative players, and 3 fast players.

She selects three players at random from the squad for a match.

Find the probability of selecting

- (a)** three creative players,

..... [2]

- (b)** three different types of player.

..... [3]

Later, for a practice match, she selects three players at random from the squad to form a red team and three players at random from the remaining members of the squad to form a blue team.

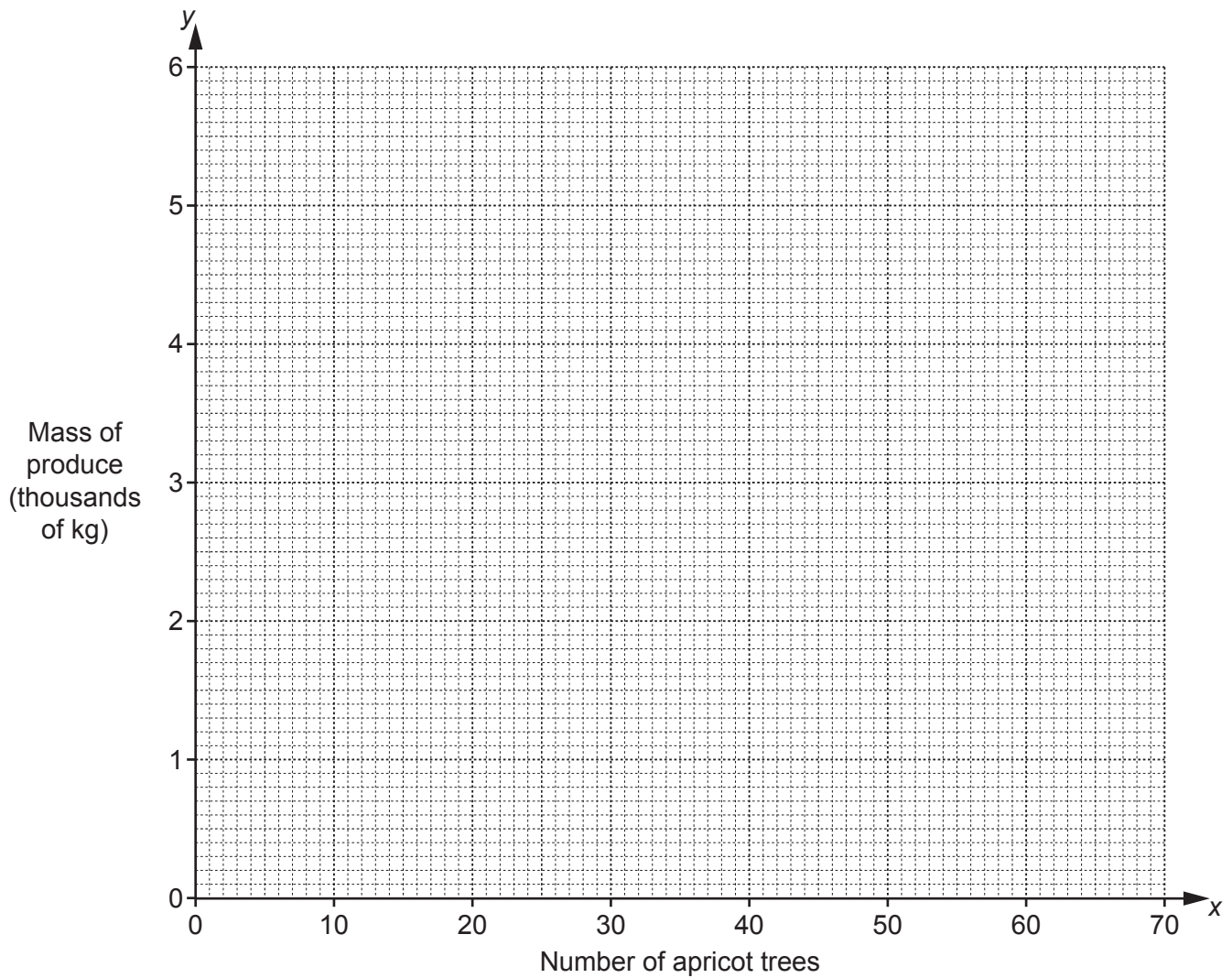
- (c)** Find the probability that the red team contains exactly one strong player and the blue team contains no strong players.

..... [3]

- 7 An agricultural institute conducts research on the produce in one growing season from apricot orchards in a particular region of a country. The following data are collected on eight orchards.

Orchard	A	B	C	D	E	F	G	H
Number of apricot trees, x	30	60	50	35	20	70	60	40
Mass of produce, y (thousands of kg)	3.5	1.5	4.5	3.0	2.0	6.0	5.5	1.0

- (a) Plot these data on the grid below.



[2]

The data have an overall mean of (45.625, 3.375), and an upper semi-average of (60, 4.375).

- (b) Find the lower semi-average.

..... [2]

- (c) Without plotting the averages, and without drawing the line, find the equation of the line of best fit in the form $y = mx + c$.

..... [3]

Two of the orchards, B and H, contain only young trees, and the remaining orchards contain only mature trees.

- (d) Identify a difference between young and mature trees that is shown by your answer to part (a).

.....
 [1]

- (e) Ignoring the orchards B and H draw, by eye, on the grid opposite, a line of best fit through the remaining six points. [1]

- (f) Use the line you have drawn in part (e) to estimate the mass of apricots produced in one growing season from an orchard of 40 mature trees.

..... kg [2]

Fresh apricots are often dried before selling. In the drying process about 80% of the mass is lost.

- (g) If an orchard plans to produce 1000 kg of dried apricots in one growing season, estimate the number of mature trees it should have.

..... [3]

8 In this question all injury rates are expressed as injuries per thousand workers.

Workplace injury rates often depend on job tenure (the length of time employees have worked in their current jobs). The table below gives information on the workplace injury rates in a city during 2019 for workers in three areas of employment, together with the standard population of workers in all areas of employment in the city.

<i>Job tenure group (years)</i>	<i>Job tenure group injury rate</i>			<i>Standard population of workers (%)</i>
	Manufacturing	Construction	Services	
under 1	50.3	88.0	55.9	10
1–under 2	41.7	42.5	38.0	15
2–under 5	34.6	35.7	27.7	50
5 or more	29.3	32.4	20.1	25

- (a) Explain how the information shows that, for all these areas of employment, the correlation between job tenure and injury rate was negative.

.....
 [1]

- (b) Without calculation, explain how the information shows that the standardised injury rate for Construction was the highest of the standardised injury rates for the three areas of employment.

.....
 [1]

- (c) (i) Calculate the standardised injury rate for Manufacturing and the standardised injury rate for Services.

Manufacturing

Services

[5]

- (ii) State, with a reason, which of these two areas of employment seems to have provided the safer working environment.

..... [1]

The table below gives information on the number of workers in the city in Manufacturing and Construction, in 2019.

<i>Job tenure group (years)</i>	<i>Number of workers in job tenure group</i>	
	Manufacturing	Construction
under 1	800	700
1–under 2	1100	900
2–under 5	4300	2600
5 or more	2500	1200

- (d) Find which of these two areas of employment suffered the greater number of injuries, and by how many.

..... [5]

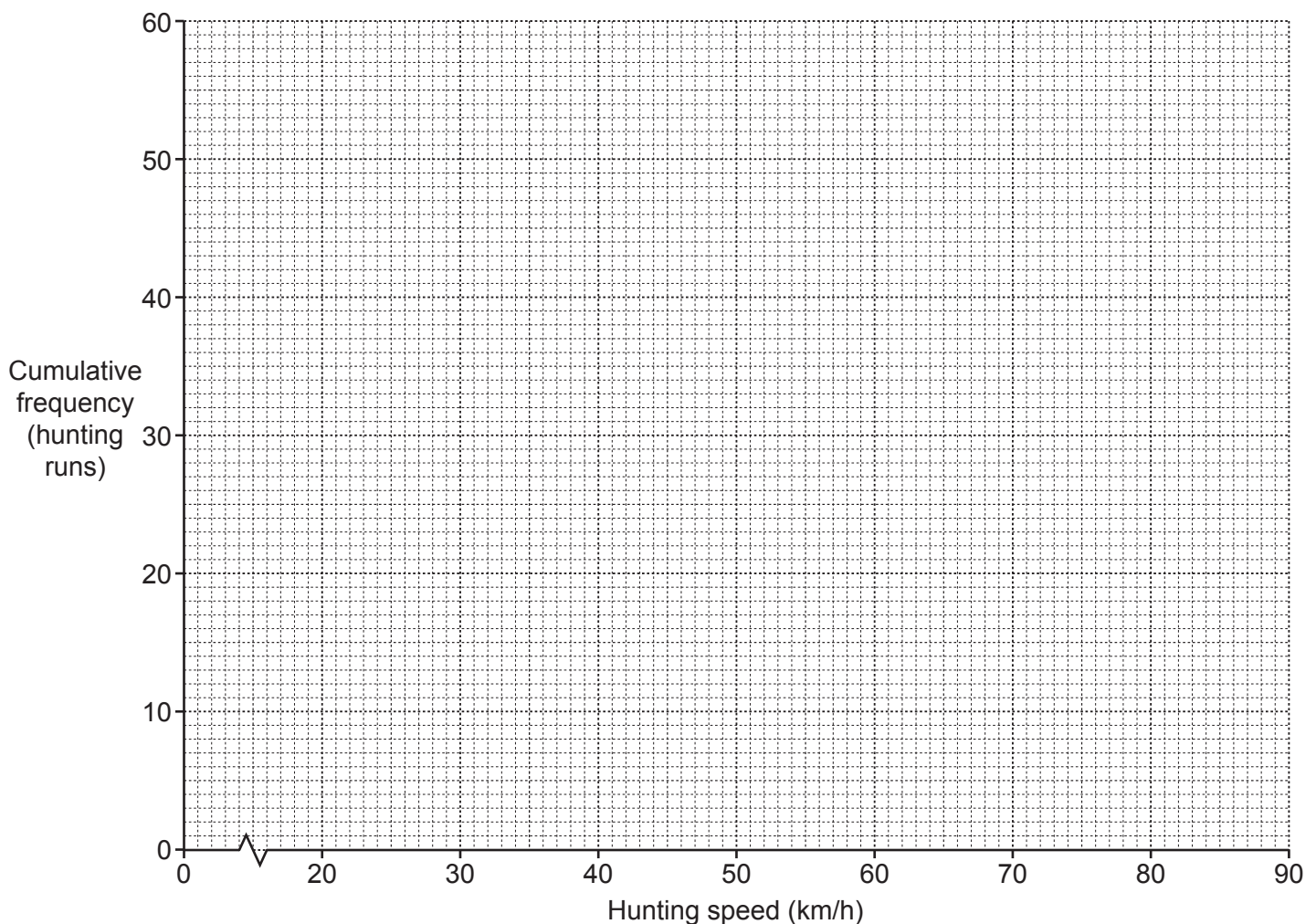
- (e) Calculate the crude injury rate for Construction.

..... [2]

- 9 The cheetah is the world's fastest land animal. To measure its speed when hunting, a wild cheetah, given the name Wilfred, was fitted with an electronic tracking collar. The following table summarises his hunting speeds during 56 hunting runs.

<i>Hunting speed (km/h)</i>	<i>Number of hunting runs</i>	<i>Cumulative frequency</i>
20–under 30	2	
30–under 40	8	
40–under 50	20	
50–under 60	15	
60–under 70	7	
70–under 80	3	
80–under 90	1	

- (a) Complete the cumulative frequency column in the table. [1]
- (b) Plot the cumulative frequencies on the grid below, joining the points with a smooth curve.



[3]

(c) Use the graph to estimate, for Wilfred's hunting speeds,

(i) the median,

..... [1]

(ii) the upper quartile,

..... [2]

(iii) the value of p if the p th percentile is 44 km/h.

..... [2]

A cheetah's hunting speed depends on the speed which the intended prey can maintain when trying to escape. A particular type of gazelle, which is common prey for a cheetah, can maintain a speed of 50 km/h.

(d) Use the graph to estimate the lower quartile of Wilfred's hunting speed for speeds of 50 km/h or more.

..... [2]

A cheetah can maintain a high speed only over a short distance, so the intended prey often escapes. Of Wilfred's 56 hunting runs, 17 ended in success. The remainder ended in failure.

(e) For Wilfred's next three hunting runs, estimate the probability that he will have more successes than failures.

..... [4]

- 10** Two tourist attractions in a city are the Aquarium and the Museum. A researcher from the tourist office surveyed samples of visitors to these attractions and recorded visitors' opinions of them. Visitors were asked to choose between 'Good', 'Fair' and 'Poor'. The results are shown below.

<i>Opinion</i>	<i>Number of Visitors</i>	
	Aquarium	Museum
Good	69	85
Fair	35	71
Poor	16	24
TOTAL	120	180

The opinions are to be represented in comparative pie charts drawn to scale, one for the Aquarium, and one for the Museum. The chart for the Aquarium has already been drawn for you.

- (a)** Draw the chart for the Museum.

Opinions of visitors to the Aquarium



Opinions of visitors to the Museum

- (b) Write down three observations that can be made by comparing the two pie charts that could not easily be seen in the original data.

1

2

3

[3]

The researcher recorded also the ages of the visitors surveyed. The table below summarises the ages of visitors to each attraction whose opinion had been recorded as 'Good'.

Age (years)	Number of Visitors	
	Aquarium	Museum
0–under 20	13	9
20–under 40	26	23
40–under 60	21	34
60–under 80	9	19

- (c) Estimate, to 3 significant figures, the mean age of these visitors to each attraction. Write down a conclusion which may be made from your answers.

Aquarium

Museum

Conclusion

.....

[5]

[Question 10 continues on the next page]

The survey was conducted by interviewing every tenth visitor as they left each attraction over a fixed period of time.

(d) State the name of this type of sampling.

..... [1]

In conducting a survey, open or closed questions may be used.

(e) State which of these types of question were used when recording the opinions of visitors. Explain your answer.

.....
..... [1]

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced online in the Cambridge Assessment International Education Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download at www.cambridgeinternational.org after the live examination series.

Cambridge Assessment International Education is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate (UCLES), which itself is a department of the University of Cambridge.