

Please check the examination details below before entering your candidate information

Candidate surname		Other names	
Centre Number	Candidate Number		
<input type="text"/>	<input type="text"/>		

Pearson Edexcel Level 3 GCE

Wednesday 21 June 2023

Afternoon (Time: 2 hours)	Paper reference	9ST0/03
---------------------------	-----------------	----------------

Statistics

Advanced

PAPER 3: Statistics in Practice

You must have: Statistical formulae and tables booklet Calculator	Total Marks <input style="width: 50px; height: 30px;" type="text"/>
--------------------------------------------------------------------------------	------------------------------------------------------------------------

Candidates may use any calculator allowed by Pearson regulations.
Calculators must not have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided
– *there may be more space than you need.*
- You should show sufficient working to make your methods clear.
Answers without working may not gain full credit.
- Unless otherwise stated, inexact answers should be given to three significant figures.
- Unless otherwise stated, statistical tests should be carried out at the 5% significance level.

Information

- A booklet 'Statistical formulae and tables' is provided.
- There are 6 questions in this question paper. The total mark for this paper is 80.
- The marks for **each** question are shown in brackets
– *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.

Turn over ►

R72911A

©2023 Pearson Education Ltd.
N:1/1/1/1/

Answer ALL questions. Write your answers in the spaces provided.

- 1 In a video game you can mine blocks of coal. The yield of a block, when mined, is determined randomly.

Mining a block of coal yields different numbers of pieces of coal, which follow the distribution shown in **Figure 1**.

The probabilities, p , of yielding 2, 3 or 4 pieces are equal.

Pieces of coal	1	2	3	4
Probability	0.4	p	p	p

[Source: <https://minecraft.fandom.com/wiki/Fortune>]

Figure 1

- (a) Find the value of p (1)
- (b) Show the expected yield from mining one block of coal is 2.2 pieces of coal. (1)
- (c) Find the variance of the yield from mining one block of coal. (1)

Question 1 continued

- (d) Find the probability that the yield of a block is **one** piece of coal, given that the yield was an odd number of pieces of coal.

(2)

- (e) State the central limit theorem.

(3)

- (f) Estimate the probability that, when 60 blocks of coal are mined, the average yield per block is more than 2.25

(2)

(Total for Question 1 is 10 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

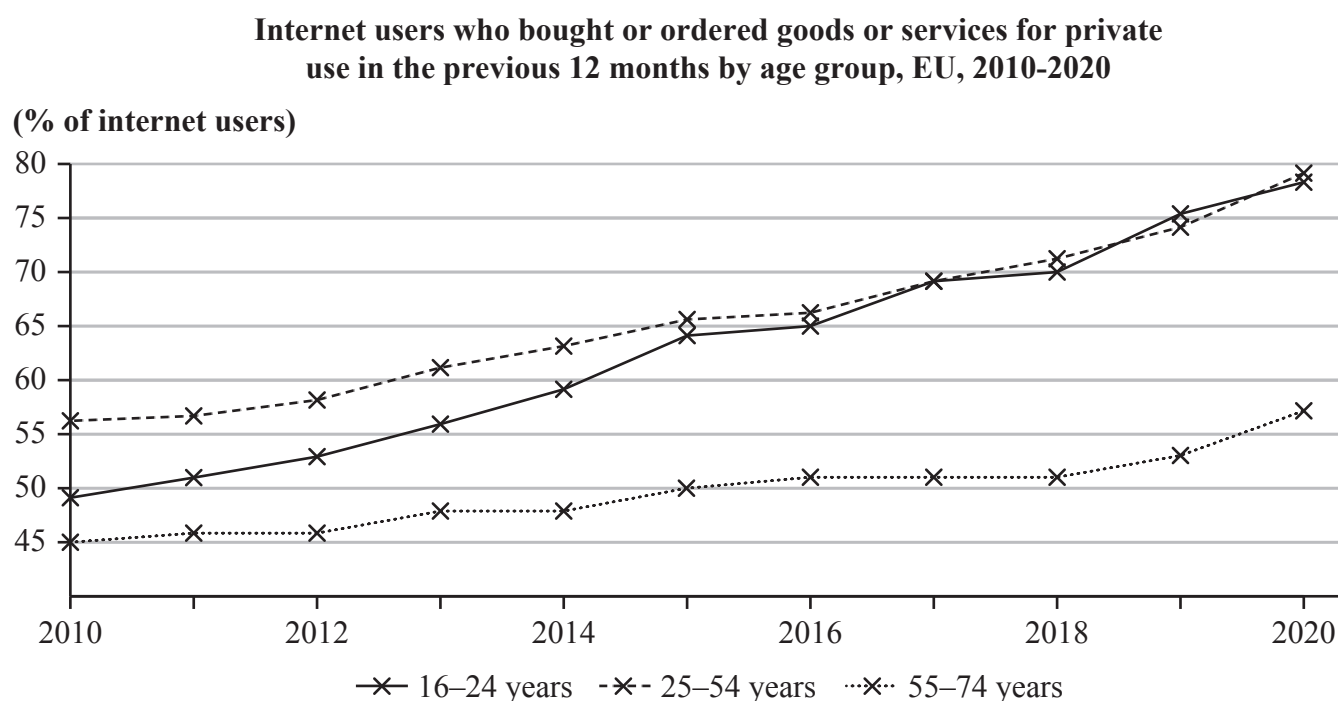
DO NOT WRITE IN THIS AREA

BLANK PAGE



- 2 Brett runs a business based in the European Union (EU). He is planning on launching a website to sell products online and is carrying out research to see if this is a feasible idea.

He finds data on the online shopping habits of different age groups, shown in Figure 2.



[Source: Eurostat]

Figure 2

Brett states that there has been an increase of more than 25 per cent in 55-74 year olds who have made an online purchase in the previous 12 months between 2010 and 2020.

His working is shown below.

$$\text{Difference: } 57 - 45 = 12$$

$$\frac{12}{45} = 0.267 \text{ (3 s.f.)}$$
$$= 26.7\%$$

Brett's colleague, Elias, states that the increase is only 12 per cent.

- (a) Briefly discuss Brett's and Elias's statements.

(2)

Question 2 continued

Figure 3 shows the percentage of people within European countries that used the internet during 2020, and the percentage of people that used the internet to make an online purchase during 2020.

Percentage of internet use and online purchases, 2020
(% of individuals aged 16 to 74)

	Proportion of individuals who:	
	Used within the last 12 months	Purchased online within the last 12 months
EU	89	65
Belgium	92	73
Bulgaria	74	31
Czechia	89	72
Denmark	99	89
Germany	95	83
Estonia	90	68
Ireland	92	74
Greece	79	46
Spain	93	63
France	91	70
Croatia	80	55
Italy	81	44
Cyprus	91	47
Latvia	90	56
Lithuania	84	54
Luxembourg	99	79
Hungary	86	60
Malta	87	87
Netherlands	95	66
Austria	89	66
Poland	85	61
Portugal	79	61
Romania	85	45
Slovenia	88	63
Slovakia	91	62
Finland	97	76
Sweden	97	84
Iceland	99	83
Norway	98	85
Switzerland	97	80
Montenegro	79	23
North Macedonia	84	34
Serbia	79	38
Turkey	78	33
Bosnia and Herzegovina	74	28
Kosovo	97	46

Figure 3

Question 2 continued

- (b) Considering the information in **Figure 3** only, suggest a country to Brett that would be useful to advertise in.

Explain your answer.

(2)

- (c) Explain why Brett should **not** make his decision on where to advertise **only** using the information in **Figure 3**.

You should make **three** separate comments.

(3)

Question 2 continued

Also available is data on employment status, which is included with some additional data in **Figure 4**.

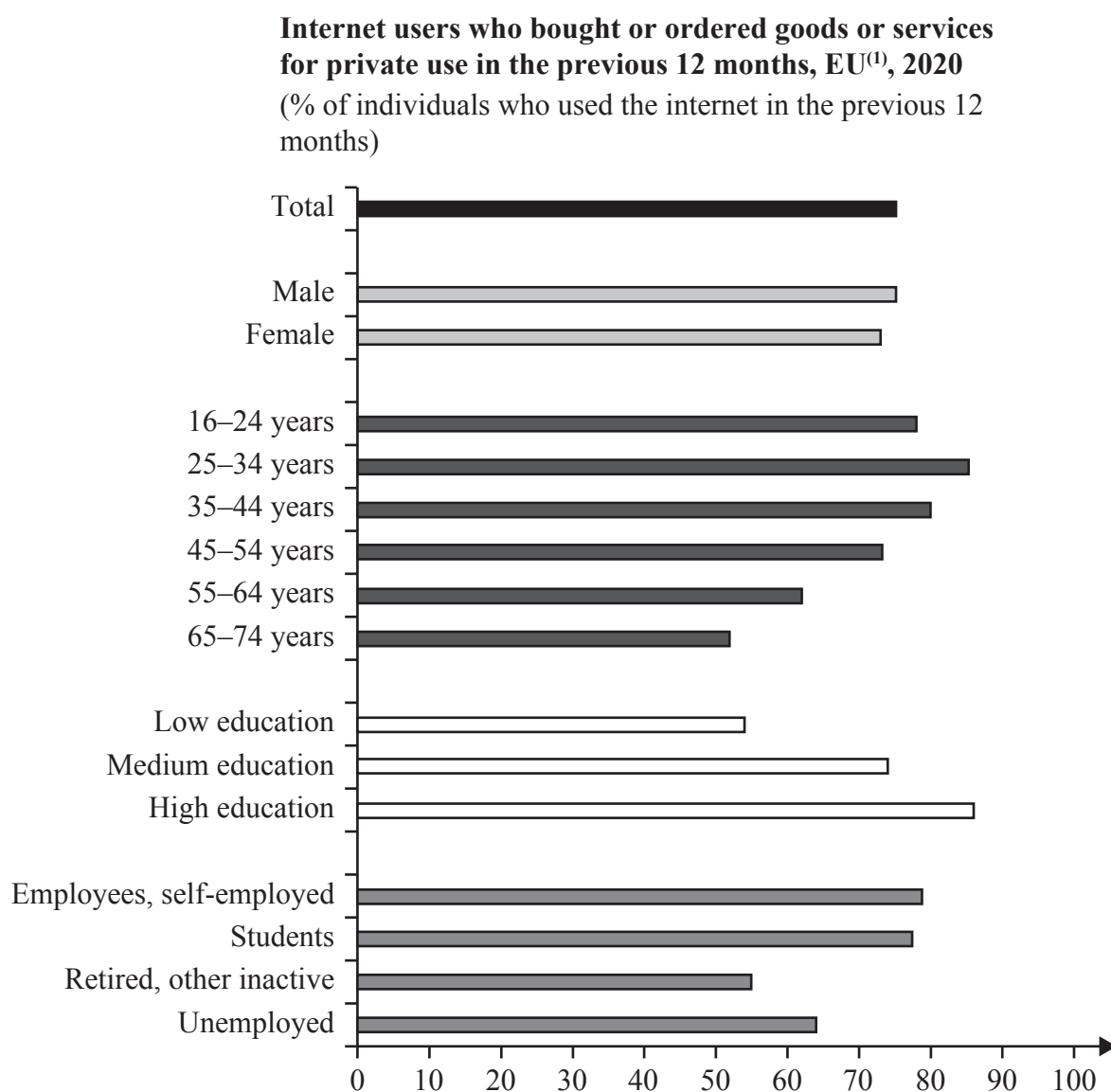


Figure 4

Question 2 continued

(d) Make **three** comments on the data in **Figure 4**.

(3)

(Total for Question 2 is 10 marks)

3 Upulani works for a lawnmower manufacturing company.

The company is considering several new batteries to use in a cordless lawnmower.

Upulani measures the lifespan of a random sample of batteries made from four different materials at three different operating temperatures (15°C, 40°C and 65°C).

A longer battery lifespan is more desirable.

She records the lifespan of each battery in hours. Her results are shown in **Figure 5**.

		Temperature			Total
		15°C	40°C	65°C	
Material	A	10.5	6.9	2.2	19.6
	B	10.0	6.6	2.1	18.7
	C	13.7	9.0	2.9	25.6
	D	12.8	8.5	2.8	24.1
Total		47	31	10	88

Figure 5

The data produced the following summary statistic

$$\sum_i \sum_j x_{ij}^2 = 831.7$$

Upulani claims that there is a difference between average battery lifespan for the four different materials.

(a) Name the technique Upulani should use to carry out this test.

(2)

(b) State **two** necessary assumptions to carry out such a test.

(2)

Question 3 continued

- (c) Carry out a hypothesis test to investigate Upulani's claim.

You may find it helpful to use the table provided.

(11)

Source	Sums of squares	Degrees of freedom	Mean square	<i>F</i> ratio
Rows				
Columns				
Error				
Total				

Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 3 continued

(d) Following your analysis in (c), make a recommendation to the lawnmower company. (1)

(e) Complete a further analysis to determine whether the blocking factor was effective. (4)

(Total for Question 3 is 20 marks)

- 4 Anna is a statistician. She is employed to advise companies on suitable statistical models to use in a variety of scenarios.

Recently, she has been asked to model three scenarios.

- (a) For each of the following scenarios, explain which model you would expect Anna to use, and justify your answers.

(i) The number of days in January on which it snows,

(2)

(ii) The number of patients attending A&E between 9pm and 10pm on a Monday evening,

(2)

(iii) The amount of time between successive visits to a company's website.

(3)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

- (b) For **one** of the scenarios in (a), give a reason, in context, why the distribution you suggested may **not** be appropriate to model the situation.

Clearly state which model you are referring to.

(1)

(Total for Question 4 is 8 marks)

- 5 Anja is investigating ‘Speedcubing’, a competition in which participants attempt to solve puzzle cubes in the fastest time possible.

One category in Speedcubing is ‘Individual time’. Individual time considers the time taken to solve a single scramble of a cube puzzle.

Another category, ‘Average time’, considers the average time taken when solving 5 different scrambles of the same cube puzzle.

Anja finds the results from one Speedcubing competition.

For a random sample of participants in this competition, Anja calculates the least squares regression line between Individual time, x seconds, and Average time, y seconds in the form $y = a + bx$

- (a) Explain why Anja should expect her value of b to be positive.

(1)

Anja’s regression line has equation $y = -0.374 + 1.564x$

[Source: <https://www.worldcubeassociation.org/>]

- (b) Estimate the Average time for a competitor with an Individual time of 9 seconds.

(2)

- (c) Explain why there is no sensible interpretation of the value of -0.374 in Anja’s model.

(1)



Question 5 continued

- (d) Interpret the value of 1.564 in context.

(1)

Anja finds a database of results of all Speedcubing competitions.

Each competitor has a unique ID, so that if there are two competitors with the same name they can identify the times.

Records containing the competitors and their unique IDs are listed in one table, and a competitor's best times in the Individual time category are listed in another table, also with their unique ID.

- (e) Explain how Anja could obtain a list of best times, from her database, with each time shown beside the competitor's name instead of their unique ID.

(3)

(Total for Question 5 is 8 marks)

- 6 In field hockey, some fouls result in the awarding of a penalty corner. A player then has the opportunity to score a goal from the penalty corner.

A study of 84 games in the 1998 Field Hockey World Cup found that

- the mean number of penalty corners awarded per game was 2.976
- the standard deviation of penalty corners awarded per game was 0.98
- 114 penalty corners resulted in a goal, and 136 did not.

[Source: https://www.researchgate.net/figure/Data-of-successfulpenalty-corners_tbl2_233698670]

Douglas plays for a field hockey team.

Sam, a sports data analyst, decides to analyse the average number of penalty corners awarded per game for Douglas's team.

Sam studies data from a random sample of 15 games played by Douglas's team over the last three years.

Sam found that

- the mean number of penalty corners awarded per game was 3.25
- $s = 1.2$

- (a) Making any necessary assumptions, calculate a 95% confidence interval for the mean number of penalty corners awarded per game for Douglas's team.

(3)

- (b) State the necessary distributional assumption required for the confidence interval found in (a) to be valid.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 6 continued

- (c) By making reference to the confidence interval found in (a), comment on the claim that, over the last three years, the number of penalty corners awarded to Douglas's team per game, on average, differs to that awarded to teams at the 1998 Field Hockey World Cup.

(2)

Douglas believes that a **lower** proportion of penalty corners that his team is awarded result in goals compared to that for teams playing in the 1998 Field Hockey World Cup.

Last year, his team was awarded 35 penalty corners resulting in 7 goals.

Douglas decides to test the following hypotheses:

$$H_0: \pi = 0.456$$

$$H_0: \pi < 0.456$$

- (d) Show how Douglas obtained the value of 0.456 for his hypothesis test.

(1)

Question 6 continued

- (e) Explain why the critical region for Douglas's test is $X \leq 10$, where X represents the number of goals scored from 35 penalty corners.

(2)

- (f) Complete Douglas's hypothesis test.

(2)

- (g) Give one advantage of using a critical region instead of a p -value when carrying out hypothesis tests.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued

(h) Describe, in context, the meaning of a Type I error for Douglas's test.

(2)

(i) Write down the exact probability that Douglas makes a Type I error when carrying out his test.

(1)

(j) If the probability that, for Douglas's team, a penalty corner results in a goal is actually 0.15, find the power of his test.

(3)

Question 6 continued

- (k) State **two** necessary assumptions that you made when selecting the distribution to use in (e).

You should explain whether you believe **each** to be valid or not.

(4)

During the 1998 Field Hockey World Cup, specialist penalty corner players were brought on to substitute for other players when a penalty corner was awarded.

In 1999, a new rule was introduced that specialist penalty corner players could not be brought on, unless a player was injured.

- (l) Explain how this information may affect the validity of the conclusion you made in (f).

(2)

(Total Marks for Question 6 is 24 marks)

(TOTAL IS 80 MARKS)





DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

