

Please check the examination details below before entering your candidate information

Candidate surname

Other names

**Pearson Edexcel
Level 3 GCE**

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

Wednesday 21 October 2020

Afternoon (Time: 1 hour 30 minutes)

Paper Reference **8ST0/01**

Statistics

**Advanced Subsidiary
Paper 1**

You must have:

Statistical formulae and tables booklet
Calculator

Total Marks

--

**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 7 questions in this question paper. The total mark for this paper is 60.
- 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 ►

P62296A

©2020 Pearson Education Ltd.

1/1/1/1/



Pearson

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

- 1 Nadiya and John are lecturers who teach a popular university course. They decide to compare the test marks obtained by their students.

Nadiya presents her students' results using a box plot whereas John presents his results using a cumulative frequency diagram.

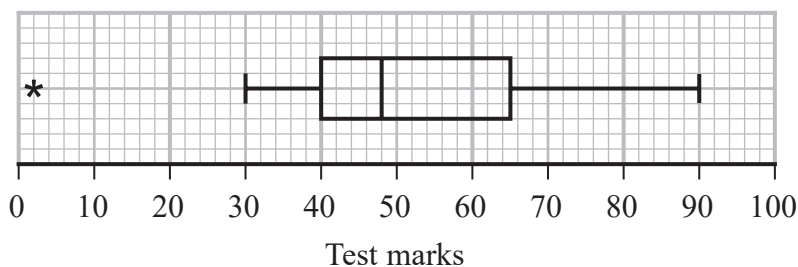


Figure 1

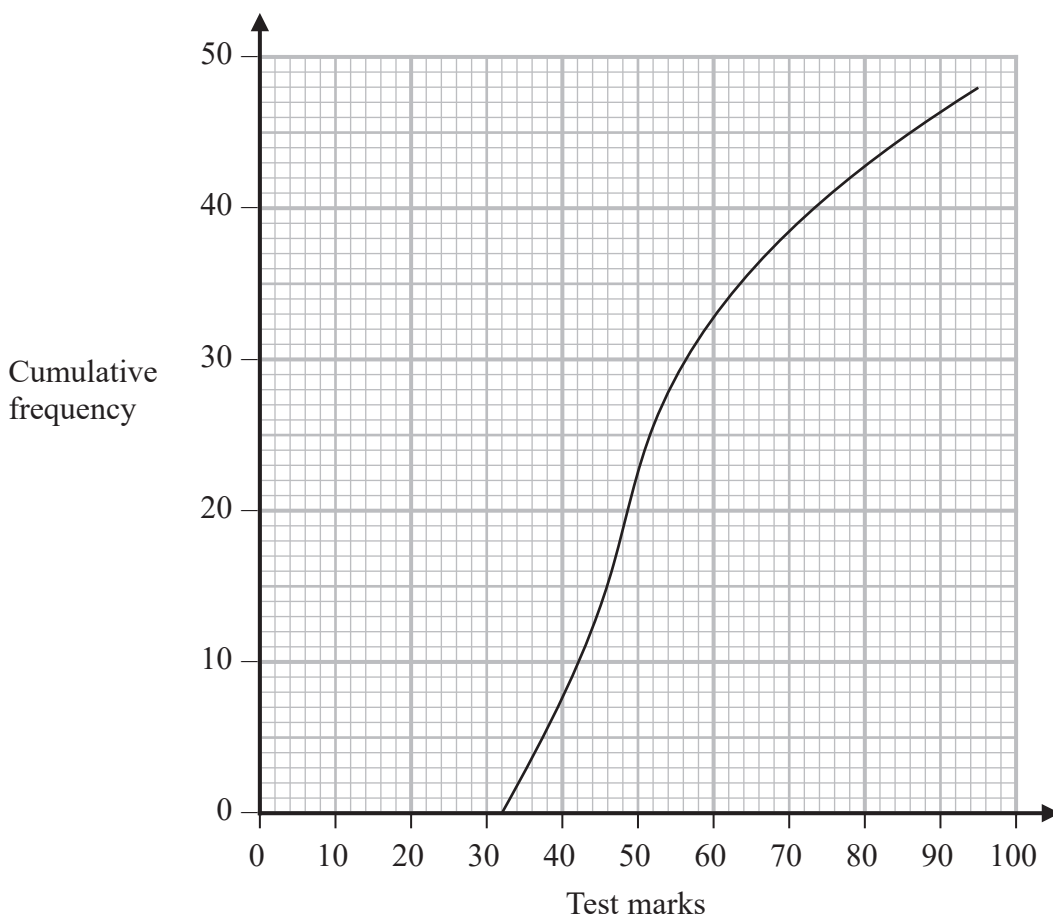


Figure 2

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 1 continued

- (a) Using information from **Figure 1** and **Figure 2**, compare the results for the students taught by Nadiya and the students taught by John.

You should make **five** different comments.

(5)

Handwriting practice area consisting of horizontal lines.



Question 1 continued

Nadiya is an experienced teacher who has been lecturing for many years. She used her students’ results from the past 10 years to construct her box plot.

John is a newly qualified teacher who is in his first year of teaching.

(b) Explain why this information makes it difficult to identify who, of Nadiya and John, is the better teacher.

You should make **two** separate points.

(2)

(Total for Question 1 is 7 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

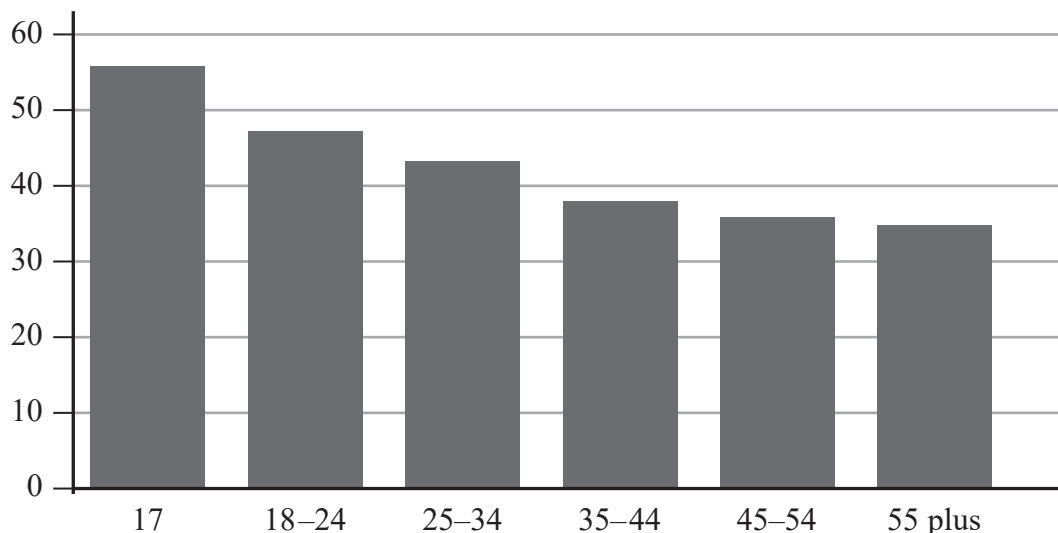
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

2 An excerpt from an article on driving tests is shown in **Figure 3**.

Driving test passes by age-group

Percentage passing test in Great Britain, 2017–18



Source: Driving and Vehicle Standards Agency



Two-thirds of people sitting their test are 25 and under, and in terms of pass rates, they are the most successful.

The performance of other age groups seems to suggest the test becomes more difficult to pass the older you get.

(Excerpt from: <https://www.bbc.co.uk/news/uk-46374981>)

Figure 3

(a) Give **two** reasons why the statement ‘the test becomes more difficult to pass the older you get’ made in the article may be **incorrect**.

(2)



P 6 2 2 9 6 A 0 5 2 4

Question 2 continued

Data is also available regarding the attempt on which a person passed their driving test. The probability that a person passed their driving test on their n th attempt is given in **Figure 4**.

Attempt on which person passed (n)	Probability
1	0.45
2	0.3
3	0.1
4+	0.15

Figure 4

- (b) Philippa states that 45% of people in the UK have sat exactly **one** driving test. Explain why the probabilities in **Figure 4** do **not** support Philippa's statement.

(1)

- (c) Explain why the expected number of attempts to pass a driving test is **greater than** 1.95

(2)

(Total for Question 2 is 5 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

3 10 fair coins are flipped, and the number of heads obtained is counted.

The following events are defined:

Event A: there are exactly 2 heads obtained in the 10 coin flips

Event B: there is exactly 1 head obtained in the **first** 5 coin flips and exactly 1 head obtained in the **second** 5 coin flips

Selasi claims that the probability of Event A is **equal to** the probability of Event B

Tim claims that the probability of Event A is **bigger than** the probability of Event B

Alexandra claims that the probability of Event A is **smaller than** the probability of Event B

(a) **Without performing any calculations**, explain whose claim is correct.

(2)



Question 3 continued

Alexandra then flips a fair coin 5 times and obtains a head on every flip.

Based on this outcome, Alexandra makes a new claim. She claims that it is more likely that she will obtain a tail rather than a head on her sixth flip on the coin.

(b) Comment, with an explanation, on Alexandra’s new claim.

(2)

(Total for Question 3 is 4 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

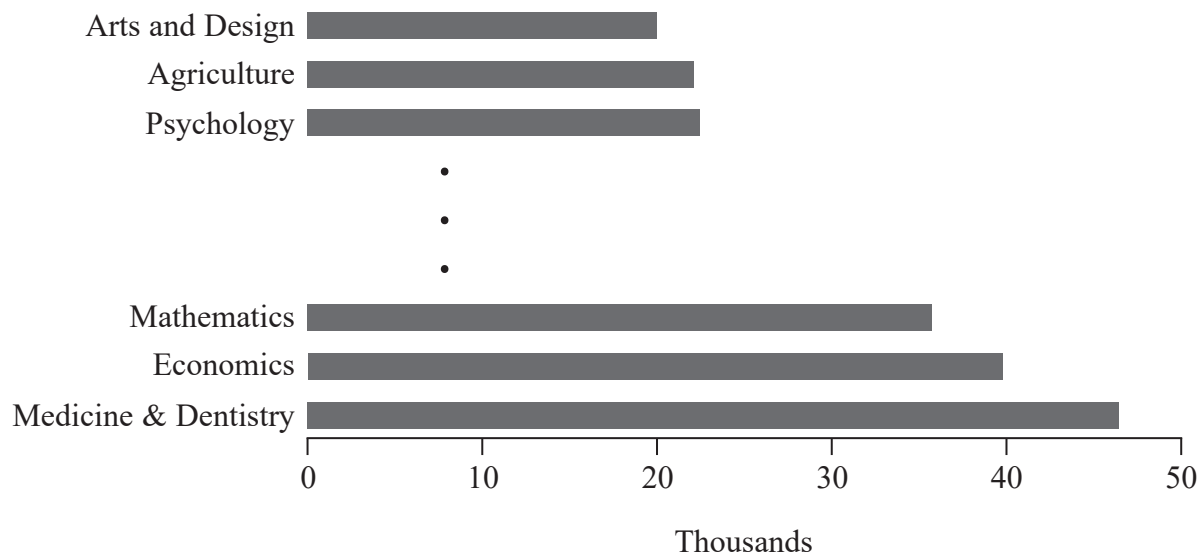
DO NOT WRITE IN THIS AREA

- 4 Bruno read an article about the different annual earnings for graduates in various degree subjects five years after graduation. He decides to investigate whether there is a difference in median annual earnings of graduates in 'Humanities or Arts' subjects and those in 'Science' subjects.

An excerpt from the article is shown in **Figure 5**.

Which graduates earn the most?

Median annual earnings (£) by subject, five years after graduating



(Source: Institute for Fiscal Studies BBC)

Figure 5

Bruno carries out his own investigation, and makes a random selection of current degree subjects and classifies each subject as either 'Humanities or Arts' or 'Science'. He then ranks the subjects, with rank 9 indicating the highest annual earnings, and rank 1 the lowest.

Figure 6 shows Bruno's rankings.

	Degree subject	
	Humanities or Arts	Science
Rank of Annual Earnings of Graduate	2, 5, 6, 9	1, 3, 4, 7, 8

Figure 6



P 6 2 2 9 6 A 0 9 2 4

Question 4 continued

(a) Explain why Bruno would **not** be correct to use a Wilcoxon signed-rank test to analyse his data in **Figure 6**.

(1)

(b) Suggest an alternative test that Bruno could use to analyse his data in **Figure 6**.

(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 4 continued

Bruno carries out a suitable test on his data and concludes that there is no difference in median annual earnings of graduates in ‘Humanities or Arts’ and those in ‘Science’ subjects.

Bruno’s friend Colin carries out a similar investigation using the same test as Bruno but he includes many more subjects. Colin concludes that there is a difference in median annual earnings of graduates in ‘Humanities or Arts’ and those in ‘Science’ subjects.

(c) State which of Bruno or Colin’s conclusions is likely to be more reliable.

Give a reason for your answer.

(1)

(d) Why did Bruno and Colin **not** need any information about the distribution of the data contained in the article in order to carry out their analyses?

(1)

(Total for Question 4 is 4 marks)



P 6 2 2 9 6 A 0 1 1 2 4

- 5 Roberta wanted to investigate the nationalities of players signed by football clubs in England and Wales.

These clubs are in one of three leagues, 'Premier League', 'EFL Championship' or 'EFL League One'.

Roberta obtained data for all the London-based clubs in these three leagues.

She formed a sample by counting the number of players of each nationality signed by the London-based clubs in one season.

- (a) State, in context, one possible advantage and one possible disadvantage of Roberta's sampling technique.

(2)

The results of Roberta's investigation are shown in **Figure 7**.

		Nationality of player signed	
		British	Non-British
League	Premier League	6	20
	EFL Championship	9	8
	EFL League One	17	6

Figure 7

Roberta claims that there is an association between the league in which a club plays and the nationality of the player signed.

- (b) Treating the data in **Figure 7** as a random sample, carry out a hypothesis test to investigate Roberta's claim.

(10)



Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Lined writing area for the response to Question 5.



P 6 2 2 9 6 A 0 1 3 2 4

Question 5 continued

- (c) Interpret your conclusion in part (b) by identifying the **greatest** source of association between the league in which a club plays and the nationality of the player signed.

(1)

Alan carried out the same analysis several years earlier. He believes that Roberta's claim is not valid because he found no significant association between the league in which a club plays and the nationality of the player signed.

- (d) Explain why Alan's belief may be wrong.

(1)

- (e) With reference to her sample, suggest **two** ways in which Roberta could improve her investigation.

(2)

(Total for Question 5 is 16 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

6 A scientist is investigating various species of gecko. The lengths of a random sample of 50 adult geckos of **Species A** are recorded, and shown in **Figure 8**.

Length (cm)	Frequency
$x < 10$	0
$10 \leq x < 11$	10
$11 \leq x < 12$	16
$12 \leq x < 14$	8
$14 \leq x < 16$	12
$16 \leq x < 20$	4
$x \geq 20$	0

Figure 8

(a) Explain why a normal distribution is **not** appropriate for modelling the lengths of these geckos.

(1)



Question 6 continued

Rita remembers reading an article where it claimed that geckos of **Species B** had lengths that are normally distributed with mean 11 cm and standard deviation 6 cm.

(b) Comment on the validity of the claim in this article, as remembered by Rita.

(2)

The lengths of geckos of **Species C** have a mean of 14.2 cm and a standard deviation of 2.1 cm. It is known that these lengths can be modelled by a normal distribution.

(c) For geckos of **Species C**, find the probability that:

(i) a randomly selected gecko is longer than 16.0 cm,

(1)



7 Derek wants to buy a new smartphone and is investigating which model of smartphone to buy. He observes that models of smartphone with a greater battery energy charge seem to be heavier. Derek decides to investigate his observation.

Derek makes a list of 110 models of smartphone and assigns the number 1, 2, ..., 110 to each model, in turn, working through the list.

He randomly selects a number between 1 and 10.

He selects the model that has this number as its identifier in the list and every 10th subsequent model in the list. These models form his sample.

(a) Name the type of sampling Derek has used, and explain why it does not give a random sample.

(2)

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 7 continued

(b) For Derek's sample, find the probability that:

(i) it contains the model of smartphone with identifier 17

(1)

(ii) it contains **both** the models of smartphone whose identifiers are 17 and 27

(1)

(iii) it contains the model of smartphone with identifier 18, given that it contains the model of smartphone with identifier 17

(1)



Question 7 continued

Derek obtains data about each of the models in his sample from a website and enters his sample data into a spreadsheet, as shown in **Figure 9**.

	A	B	
1	Battery energy charge (milliamp hours)	Weight (g)	
2	2300	143	
3	2300	147	
4	2200	131	
5	2200	140	
6	2400	147	
7	3080	145	
8	2500	165	
9	3300	178	
10	2730	137	
11	3000	152	
12	4000	178	

(Data Source: Productchart.co.uk)

Figure 9

- (c) Explain how Derek could use his spreadsheet to place the models of smartphone in order of weight, from lightest to heaviest.

(2)



Question 7 continued

Later, Derek realises that he had made an error and some models of smartphone appeared twice in his original list of 110

(e) How would this information affect the conclusion you made in part (d)?

(1)

Derek's friend, Amina, carried out her own investigation. The result of her test suggested that there is significant evidence that models of smartphone with larger screens weigh more than those with smaller screens.

Amina claims that the weight of a model of smartphone depends on the size of its screen.

(f) Comment on Amina's claim.

(1)

(Total for Question 7 is 17 marks)

TOTAL FOR PAPER IS 60 MARKS



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

Every effort has been made to contact copyright holders to obtain their permission for the use of copyright material. Pearson Education Ltd. will, if notified, be happy to rectify any errors or omissions and include any such rectifications in future editions.

