

Please check the examination details below before entering your candidate information

Candidate surname

Other names

Centre Number

Candidate Number

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

--	--	--	--	--

--	--	--	--

Tuesday 18 June 2019

Morning (Time: 1 hour 30 minutes)

Paper Reference **1ST0/2H**

Statistics

Paper 2
Higher Tier

You must have:

Ruler graduated in centimetres and millimetres, protractor, pair of compasses, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided
– *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.



Information

- 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.

Turn over ►

P60683A

©2019 Pearson Education Ltd.

6/6/1/1/




Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Rates of change (e.g. Crude birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

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

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 At a university, 70% of students are undergraduates and 30% of students are postgraduates.

Amy and Robert want to do a survey.

Amy decides to use simple random sampling to collect a sample of 100 students.

She uses the university database as a sample frame and she numbers each student on the database.

She then generates exactly 100 random numbers and uses these random numbers to select her sample.

(a) Give **two** reasons why Amy's method may **not** produce a sample of 100 students.

.....

.....

.....

.....

(2)

Robert decides to use quota sampling to collect a sample of 100 students.

He plans to stand outside the main building until he has interviewed 70 undergraduates and 30 postgraduates.

(b) Give **two** advantages of using quota sampling.

.....

.....

.....

.....

(2)

(c) Explain why this quota sample is **not** a random sample.

.....

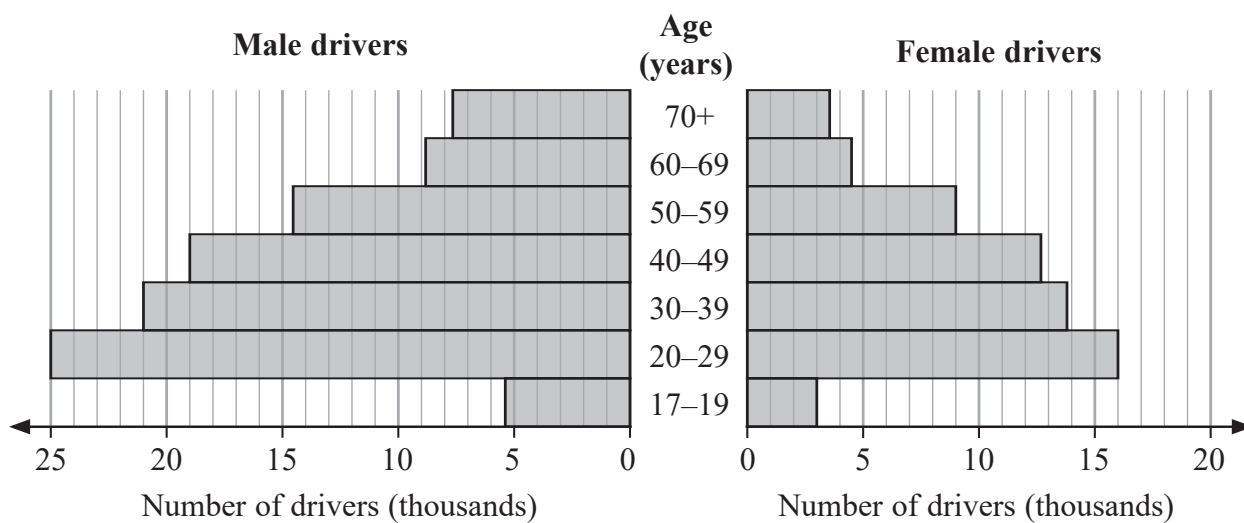
.....

(1)

(Total for Question 1 is 5 marks)



2 The population pyramid shows information about the numbers (in thousands) of drivers of each gender who made car insurance claims in the UK in 2015



(Source: Department for Transport)

(a) How many female drivers aged 50–59 in the UK in 2015 made car insurance claims?

..... thousand
(1)

The population pyramid shows that the age group which has the fewest number of drivers who made car insurance claims is the 17–19 age group.

(b) Suggest a reason why this should be so.

.....
.....
(1)

In 2014, the number of male drivers aged 20–49 in the UK who made car insurance claims was 66 700

(c) Compare the number of male drivers aged 20–49 in the UK who made car insurance claims in 2014 with the number of male drivers aged 20–49 in the UK who made car insurance claims in 2015
You must show your working.

.....
.....
(3)



The SafeDrive insurance company charges young male drivers more for car insurance than it charges all other drivers.

- (d) Explain **two** features of the population pyramid which SafeDrive might use as its justification for doing this.

.....

.....

.....

.....

(2)

Jeremy says,

“The population pyramid shows that the total number of male drivers in the UK in 2015 is greater than the total number of female drivers in the UK in 2015”

- (e) Explain whether or not Jeremy’s conclusion is appropriate.

.....

.....

(1)

Vicki says,

“In the UK in 2019, there will be more male drivers who make car insurance claims than female drivers who make car insurance claims”

- (f) Explain whether or not the information in the population pyramid can be used to support Vicki’s statement.

.....

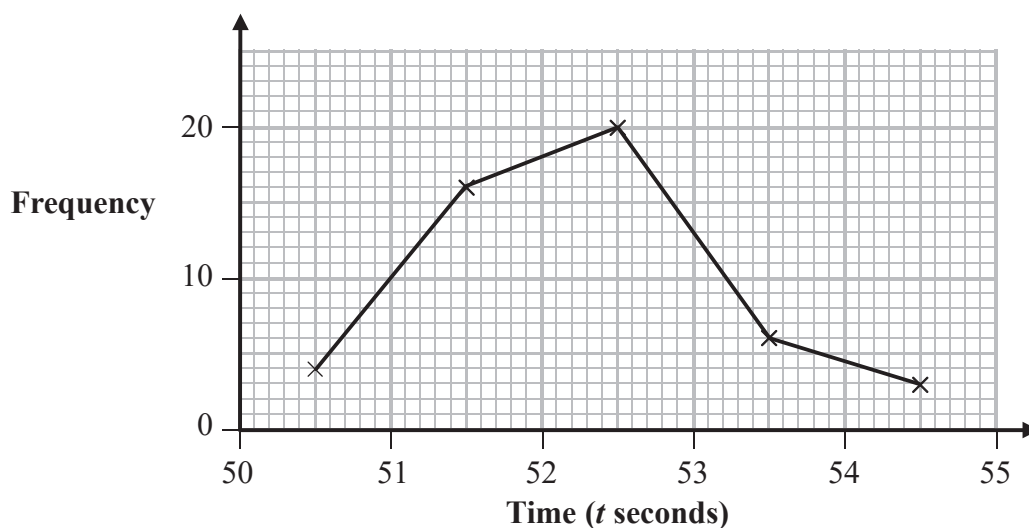
.....

(1)

(Total for Question 2 is 9 marks)



3 The frequency polygon shows information about the times taken by 49 athletes to run 400 metres at the 2017 World Championships.



(Source: www.iaaf.org)

(a) Use the information in the frequency polygon to complete the table by writing in the missing times.

Time (t seconds)	Frequency
$\dots < t \leq \dots$	6

(1)

None of the athletes ran 400 metres in exactly 53 seconds.

(b) Find the number of athletes who ran 400 metres in less than 53 seconds.

.....

(2)

(c) Calculate an estimate for the mean time of the 49 athletes.

..... seconds

(3)

(Total for Question 3 is 6 marks)



4 Randall used the internet to find the number of points scored by each of the winning teams in the last 30 years of the American Super Bowl.

He drew this stem and leaf diagram for his data.

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

(Source: nfl.com)

(a) Write down two improvements Randall should make to the stem and leaf diagram.

1.....

.....

2.....

.....

(2)

(b) Assess the reliability of using the internet as the source of Randall’s data.

.....

.....

(1)

Randall believes that the number of points scored by the winning team is increasing over time.

(c) Comment on whether or not Randall has represented his data in an appropriate diagram in order to support his belief.

.....

.....

.....

(1)

(Total for Question 4 is 4 marks)



- 5 The figures below show the amount, in £ millions, of UK aid given to Ethiopia for each of the years 2009 to 2015

219 263 344 265 329 321 338

The table gives a summary of the amount, in £ millions, of UK aid given to Pakistan for the years 2009 to 2015

Mean	Standard deviation	Greatest amount
243.7	79.5	374

(Source: *dfid.gov.uk*)

Compare the amount of UK aid given to Ethiopia and the amount of UK aid given to Pakistan for the years 2009 to 2015

You may use $219^2 + 263^2 + 344^2 + 265^2 + 329^2 + 321^2 + 338^2 = 631217$

(Total for Question 5 is 5 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

6 Using July 2015 as the base month, the table shows some of the index numbers for the consumer price index (CPI) from July 2015 to July 2017

	July 2015	Jan 2016	July 2016	Jan 2017	July 2017
Consumer price index (CPI)	100	99.5	100.6	101.4	103.2

(Source: *ons.gov.uk*)

(a) Give an interpretation of the index number 103.2 in the table.

.....

.....

.....

.....

(2)

The CPI is a weighted index. It measures the changes in the prices of consumer goods and of services only using the weightings as shown in the table below.

	weighting
goods	525
services	475

Using July 2015 as the base month, the index number for goods in July 2017 was 100.7

(b) Calculate the index number for services in July 2017
Give your answer to the nearest whole number.

.....

(2)

One of the items included in the goods index is food.

(c) Assess whether or not it is appropriate to use 100.7 in order to find an estimate of the change in the price of food from July 2015 to July 2017

.....

.....

.....

(2)

(Total for Question 6 is 6 marks)



- 7 Susie is investigating the amount of time students at her school spend watching television. She selects a random sample of 50 students from her school and asks them the following question.

How much time, in hours to the nearest hour, did you spend watching television last Saturday?

2

3

4 to 5

6 to 7

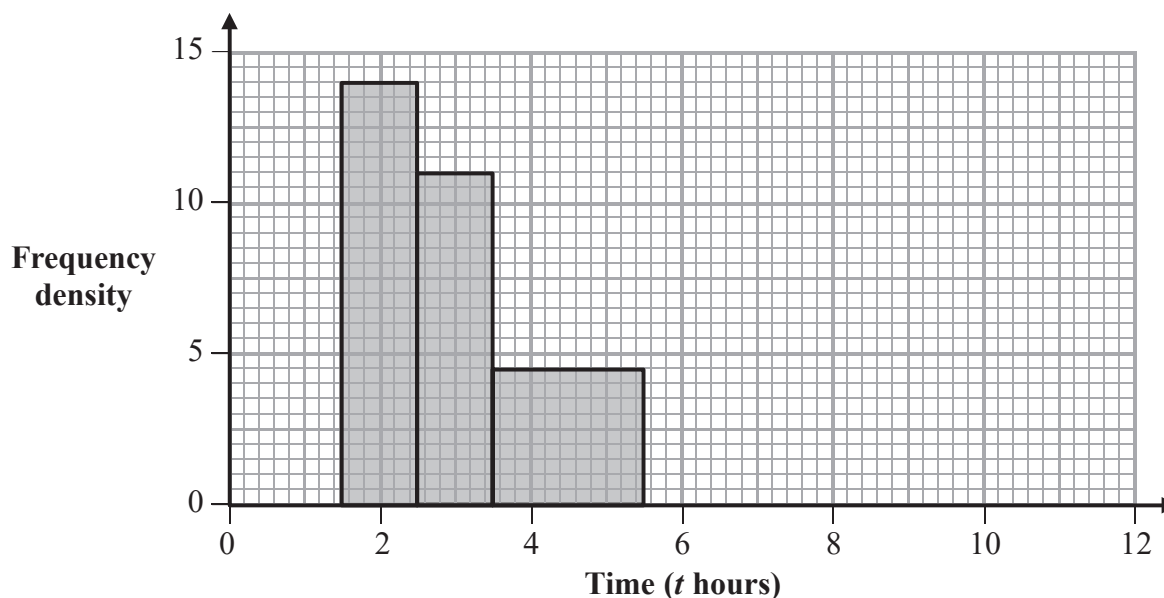
8 to 11

Susie received answers from only 43 of the students.

The table shows a summary of her results.

Time (t hours)	Frequency
2	14
3	11
4 to 5	9
6 to 7	5
8 to 11	4

The incomplete histogram shows information about some of Susie's results.



- (a) Give a reason to support using a histogram to represent the amounts of time students at her school spend watching television.

(1)



(b) Complete the histogram.

(3)

A student is selected at random from those students in the sample who watched less than 5.5 hours of television last Saturday.

(c) Estimate the probability that this student watched less than 2 hours of television last Saturday.

.....
(2)

Using class midpoints, Susie calculated the mean time spent watching television to be 4 hours. She concludes that the mean time spent watching television each day is 4 hours for **all** students.

(d) Assess how Susie's question affects the accuracy of the mean time.

.....
(2)

(e) Assess how Susie's method for collecting her data affects the validity of her conclusion.

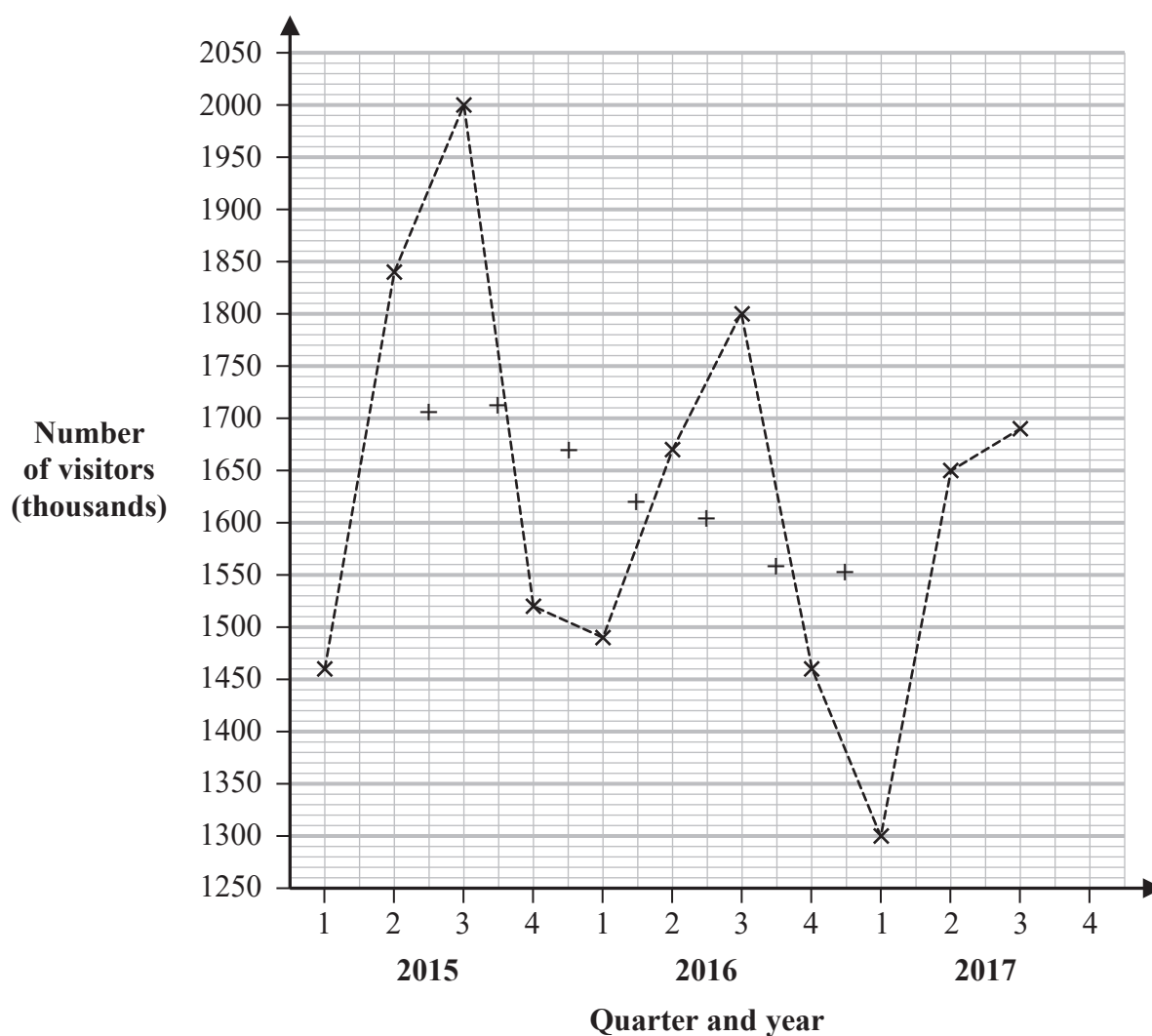
.....
(3)

(Total for Question 7 is 11 marks)



- 8 The table and time series graph give information about the number, in thousands, of visitors to the British Museum each quarter from Quarter 1, 2015 to Quarter 3, 2017

Year	Quarter			
	1	2	3	4
2015	1460	1840	2000	1520
2016	1490	1670	1800	1460
2017	1300	1650	1690	



(Source: Department for Digital, Culture, Media and Sport)

Seven of the eight 4-point moving averages for the data are also plotted on the grid.

- (a) Explain why it is appropriate to calculate 4-point moving averages for this set of data.

(1)



The eighth 4-point moving average for the data is 1525

(b) Plot this moving average on the grid. (1)

(c) Draw a trend line for the time series graph. (1)

The average seasonal effect for Quarter 4 is found to be -140 thousand.

(d) (i) Give an interpretation of this value. (1)

(ii) Calculate an estimate of the number of visitors to the British Museum in Quarter 4, 2017

..... thousand (2)

(e) Write down two assumptions that have to be made so that your estimate in part (d)(ii) is valid.

1

2

(2)

(Total for Question 8 is 8 marks)



9 To apply for a job at a bank, an applicant must first take a numeracy test.

The table gives information about the results of the test taken by all applicants.

Mean	Standard deviation
55	8

To get an interview, an applicant must score a standardised score of at least 1.5 for the test.

Mithra scored 68 in the test.

(a) Determine whether or not Mithra gets an interview.

(3)

Alexi's standardised score for the test is -1.25

Fiona's standardised score for the test is -1

(b) Compare Alexi's performance in the test with Fiona's performance in the test.
Give a reason for your answer.

(2)

(Total for Question 9 is 5 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

10 Noah is investigating the reaction times of some athletes.

He collects information from 50 athletes and records the data in a spreadsheet. He sorts the data from the least reaction time to the greatest reaction time.

The first few rows and last few rows of his spreadsheet are shown below.

Reaction time, s seconds
0.3
0.4
0.7
0.8
~~~~~
~~~~~
3.0
3.1
3.6
5.3
5.7

Here are some of the summary statistics for all the data.

$$\sum s = 88 \quad \sum s^2 = 210 \quad n = 50 \quad \text{standard deviation} = 1.05 \quad \text{range} = 5.4$$

Noah decides to clean the data before proceeding with his investigation. He models the reaction times using a normal distribution.

Using the summary statistics, describe how Noah should clean the data and justify why he needs to clean the data.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Total for Question 10 is 6 marks)



11 Gisele collected data about the age and the salary of each employee at a small company. She used statistical software to draw a scatter diagram for her data.

(a) Give one advantage of using statistical software when representing data.

(1)

Gisele calculated correlation coefficients for her data. She obtained the following results.

Spearman's rank correlation coefficient	0.95
Pearson's product moment correlation coefficient	0.77

(b) (i) Describe and interpret the type of correlation represented by 0.95 in the table.

(2)

(ii) Which of the two correlation coefficients in the table represents the stronger correlation? You must give a reason for your answer.

(1)



Figure 1 and Figure 2 show two possible scatter diagrams for Gisele's data.

Figure 1

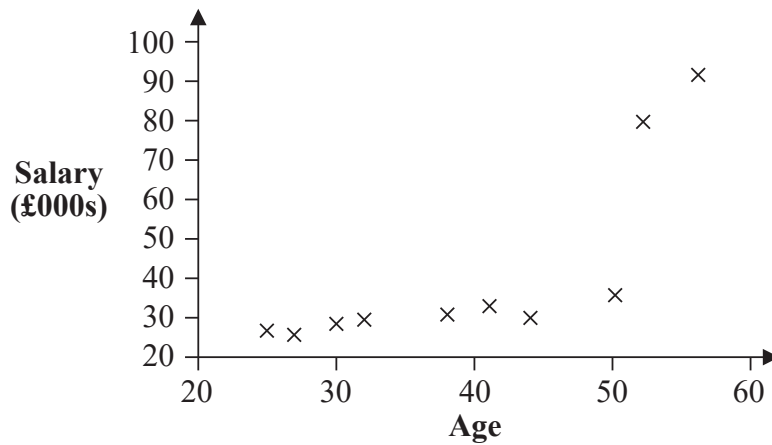
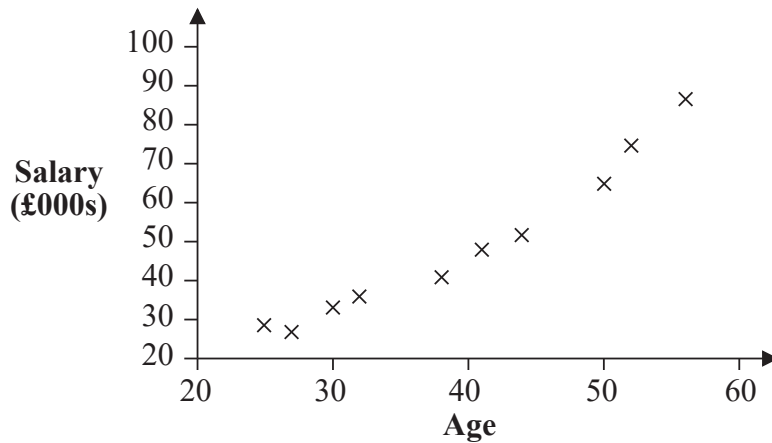


Figure 2



- (c) Which one of these two diagrams most likely represents the data?
You must give a reason for your answer.

(2)

Brett wants to use a Pearson's product moment correlation coefficient (PMCC) to compare the salaries of male employees with the salaries of female employees.

- (d) Explain whether or not it is appropriate to use the PMCC to make this comparison.

(1)

(Total for Question 11 is 7 marks)



12 Jasper has 3 coins.

In an experiment, Jasper flips each of the 3 coins and records the total number of heads that he gets.

Jasper believes that each coin is biased so that the number of heads he gets can be modelled by the binomial distribution, $B(3, 0.4)$.

(a) Show that $P(0 \text{ heads}) = 0.216$

(1)

(b) Work out the probability that the outcome of the experiment is exactly 1 head.

.....
(2)



Jasper carries out the experiment 100 times.

The table shows information about the number of heads he gets for each experiment.

Number of heads	Frequency
0	24
1	41
2	30
3	5

(c) Determine whether or not the model $B(3, 0.4)$ is suitable for Jasper's experiment.

(5)

(Total for Question 12 is 8 marks)

TOTAL FOR PAPER IS 80 MARKS



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

