

Write your name here

Surname

Other names

Centre Number

Candidate Number

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

Statistics

Paper 1

Higher Tier

Sample assessment material for first teaching September 2017

Paper Reference

Time: 1 hour 30 minutes

1ST0/1H

You must have:

Ruler graduated in centimetres and millimetres, protractor, 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 ►

S54191A

©2018 Pearson Education Ltd.

1/1/1/1



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. 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

Answer ALL the questions.

Write your answers in the spaces provided.

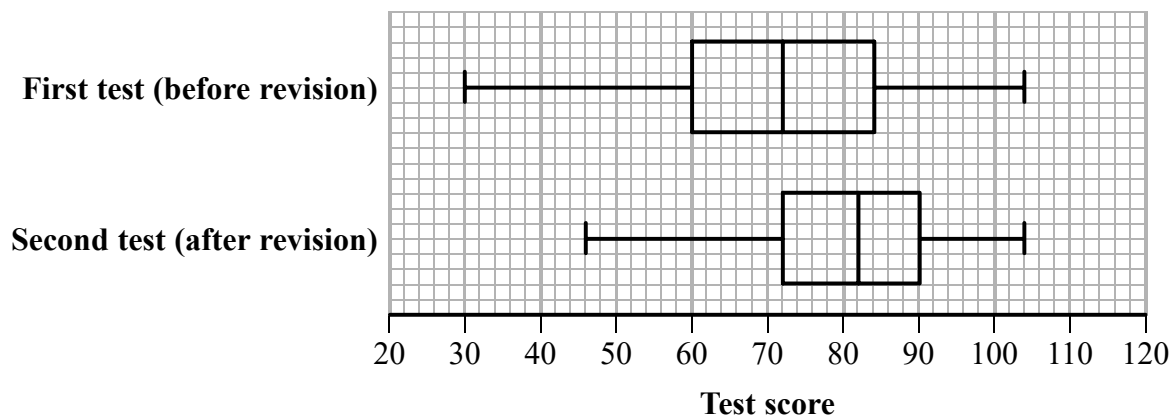
You must write down all the stages in your working.

- 1** A teacher gave her students two mathematics tests on the same topics.

The first test was given before revision of the topics and the second test was given after revision of the topics.

The two tests had the same maximum mark and the same difficulty.

The box plots give information about the students' scores in these tests.



Describe what effect the revision had on the test scores of these students.

You must explain how you reach your conclusions.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Total for Question 1 is 4 marks)

- 2 Kerry is investigating whether there is a difference in the lengths of the text messages sent by boys and sent by girls at her school.

She writes the following hypothesis for the investigation.

“The length of text messages sent by girls is greater than the length of text messages sent by boys”.

Kerry decides to use a census of the 800 students in her school.

She is going to ask each student to record the number of characters in their last text message.

Kerry then collects this information from each student through an online database.

Some of the database is shown below.

	Gender	Length of text message
1	male	73
2	F	68
3	girl	thirty five
4	boy	114,
5	boy	85
6	girl	
7	M	56
8	boy	48
9	girl	5
10	G	75
11	B	41
12	girl	28

- (a) Give **two** reasons why Kerry must clean the data before processing it.

Reason 1:

.....

Reason 2:

.....

(2)

(b) Discuss how Kerry's data collection plan could affect the reliability of her conclusions.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(2)

(Total for Question 2 is 4 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

3 The table shows information about houses for sale in Oxford.

Number of bedrooms	1	2	3	4	5 or more	Total
Number of houses for sale	140	300	420	240	100	1200

(Source: adapted from *rightmove.co.uk*)

An estate agent says the mode of the number of bedrooms for these houses is 3

(a) Explain how she knows this.

(1)

The estate agent wants to investigate the prices of these houses.

She takes a stratified sample of 60 houses according to the number of bedrooms.

(b) Work out the number of houses in her sample for each number of bedrooms.

(3)

Number of bedrooms	1	2	3	4	5 or more
Number of houses in the sample					

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(c) Describe how to select the 60 houses in the sample.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 3 is 7 marks)

4 Morgan is investigating the 180 Year 11 students in his school.

He collected information from the 30 students in his class.

Part of the spreadsheet he used to collect the information is shown below.

Student number	Name	Left-handed (1 = yes, 0 = no)	Height (cm)
1	Jason	0	169
2	Rami	1	165
29	Youen	0	164
30	Elena	0	162
Total		4	5031

Morgan uses these results to find estimates for all Year 11 students.

(a) Find his estimate for the number of left-handed students in Year 11

.....
(2)

(b) Explain how Morgan can use the information in the spreadsheet to estimate the mean height of all the students in Year 11

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

(Total for Question 4 is 3 marks)

- 5 The table shows information about the retail price index (RPI) and the price of a second-class stamp (in pence) in the United Kingdom for January 1996, January 2006 and January 2016

	Jan 1996	Jan 2006	Jan 2016
Retail price index (RPI)	100	129	172
Price of second-class stamp (pence)	20	23	54

(Sources: *ons.gov.uk* and *royalmail.com*)

Describe how the increase in the price of a second-class stamp compares with the RPI over the ten years to January 2006 **and** over the twenty years to January 2016

(Total for Question 5 is 5 marks)

6 Rahul, Lisa and Paul are investigating how much the workers in a company earn.

They have been told that in a week the workers earn £260 or £370 or £510

Last week

20% of the workers earned £260

35% of the workers earned £370

45% of the workers earned £510

Rahul, Lisa and Paul want to work out the average earnings for these workers last week.

Rahul thinks that they should find the mean of £260, £370 and £510

Lisa thinks that they should find the median of £260, £370 and £510

Paul thinks that they should find the weighted mean of the earnings.

(a) Which one of these three averages should they use?

Give a reason for your answer.

.....
.....
.....
.....

(2)

Rahul works out that the mean of the earnings is £380

Lisa finds that the median of the earnings is £370

(b) Work out the weighted mean of the earnings for Paul.

£.....

(2)

(Total for Question 6 is 4 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

7 A fitness company is planning to open a new gym in a town.

The company wants to collect information about the health of people in the town.

The company plans to interview people in the town centre and ask them questions from a questionnaire.

Here are two of the questions from the questionnaire.

A How old are you?

B What is your weight?

Discuss whether using these questions in an interview is an appropriate way to collect this information.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(Total for Question 7 is 3 marks)

- 8 The table gives information about the monthly average price per litre, in pence, of diesel over a period of five months.

The table also gives some of the chain base index numbers, correct to one decimal place, for this information.

	May	June	July	August	September
Monthly average price (p)	109.1	111.8	112.7	111.2	113.1
Chain base index number		102.5	100.8		

(Source: *theaa.com*)

- (a) Calculate the chain base index numbers for August and September and write them in the table.
Give each value correct to one decimal place.

(2)

- (b) (i) Calculate the geometric mean of the four chain base index numbers.
You must show your working.
Give your answer correct to one decimal place.

(2)

- (ii) Interpret your answer.

(2)

(Total for Question 8 is 6 marks)

9 Tania wants to estimate the number of snails in a pond.
She takes a sample of 10 snails from the pond.
She marks each snail with a waterproof dye and then puts the snail back in the pond.

Two weeks later, Tania takes another sample of 10 snails from the pond.
She finds that only one of the snails is marked with the dye.

Tania says,

“I estimate there are 100 snails in the pond.”

How reliable is Tania’s estimate?

Give reasons for your answer.

You are not required to check Tania’s calculation.

.....

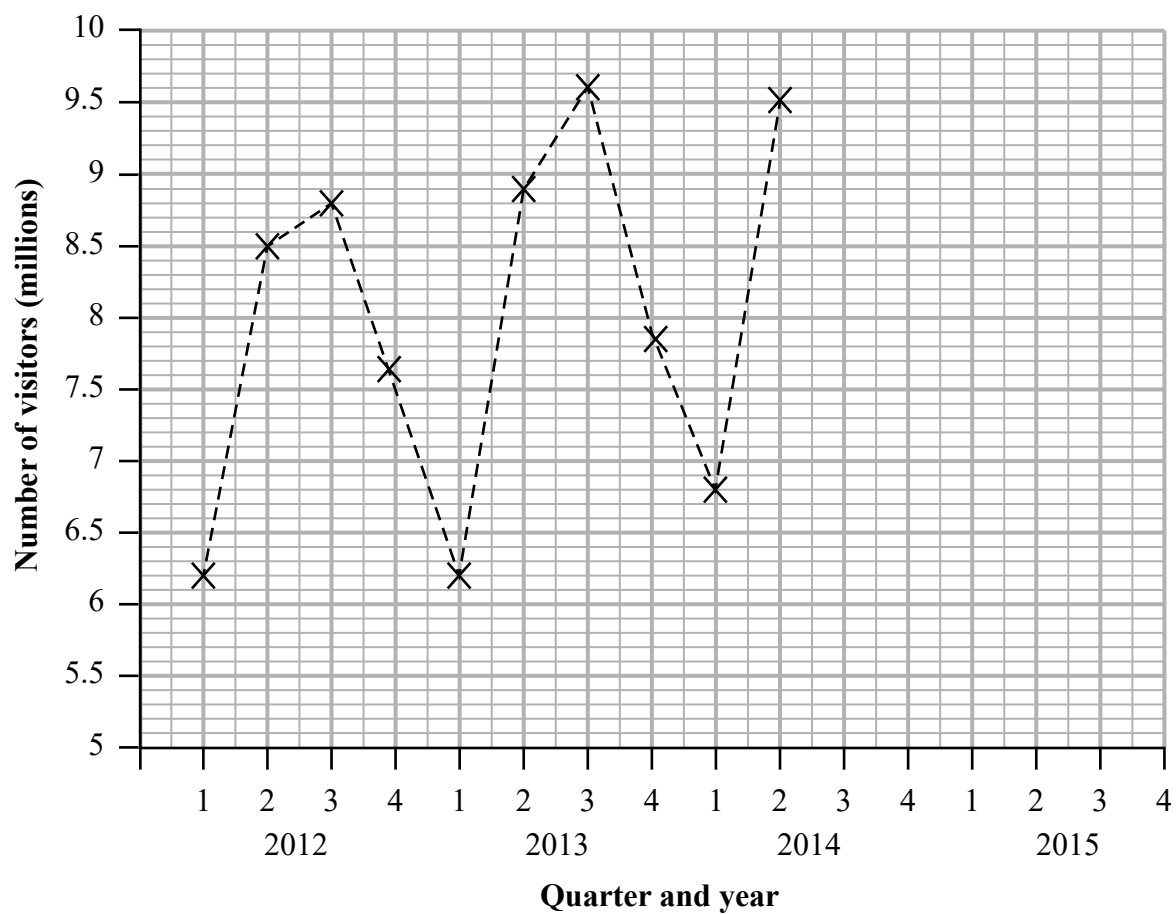
.....

.....

.....

(Total for Question 9 is 2 marks)

- 10 The time series graph shows some information about the numbers of visitors to the UK from 2012 to 2014



(Source: *visitbritain.org*)

- (a) Identify **and** interpret in context **one** example of the seasonality shown by the time series graph.

(2)

Tony calculates the 4-point moving averages for the information shown in the time series graph.

Here are his results.

7.77 7.79 7.90 8.11 8.17 8.29 8.45

(b) Plot these moving averages on the time series graph and hence draw a trend line for the number of visitors to the UK from 2012 to 2014

(3)

(c) Describe the trend.

.....
(1)

Tony says,

“Using the time series graph I estimate that there were approximately 7.35 million visitors to the UK in Quarter 1 of 2015”.

(d) Use the average seasonal effect for Quarter 1 to show that Tony’s estimate is reasonable.

(3)

Tony calculated 4-point moving averages for the information shown in the time series graph.

(e) Explain why this is appropriate.

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

(Total for Question 10 is 10 marks)

- 11 In a television talent contest, 9 acts were given the following ranks by the judges and by a public telephone vote.

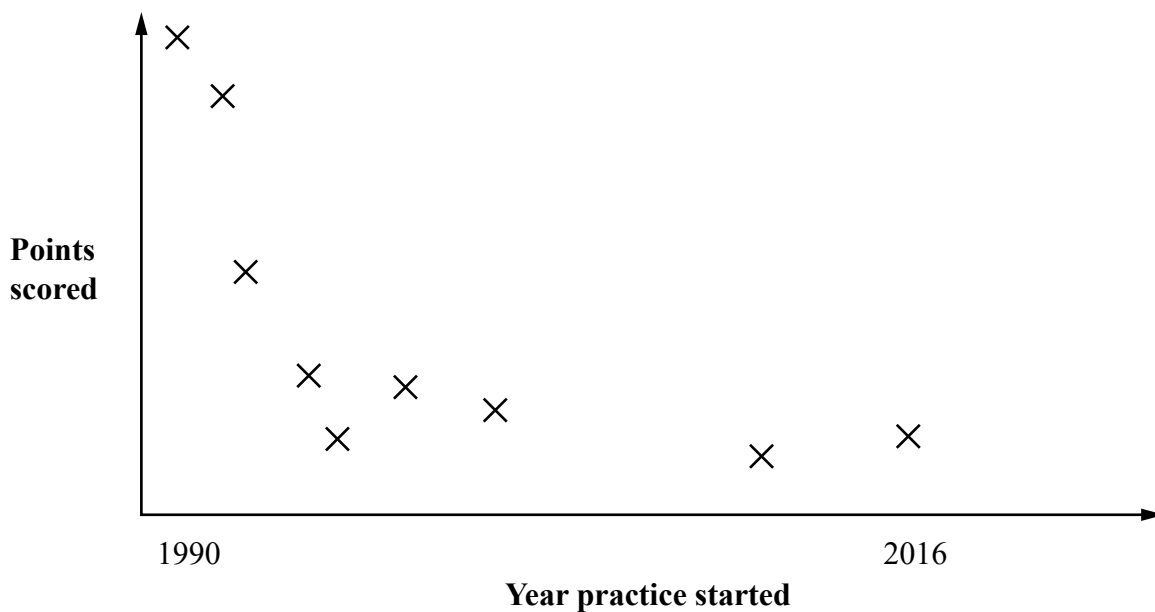
Act	Judges' rank	Public vote rank		
A	9	9		
B	3	2		
C	1	3		
D	4	5		
E	5	8		
F	8	6		
G	6	4		
H	2	1		
I	7	7		

- (a) Use calculations to determine how much agreement there is between the judges and the public.

(5)

Gurdeep was investigating the relationship between the number of points scored and the year practice started for each of 9 acts.

This is the scatter diagram he obtained using statistical software.



The statistical software also calculated two correlation coefficients.

Spearman's rank correlation coefficient

Pearson's product moment correlation coefficient

- (b) (i) Circle **one** value in each row below to show the most likely **pair** of correlation coefficients for this data.

Spearman's rank

correlation coefficient: -0.9 -0.7 0 0.7 0.9

Pearson's product moment

correlation coefficient: -0.9 -0.7 0 0.7 0.9

(2)

- (ii) Explain your choice of answers in part (i).

.....

.....

.....

(1)

(Total for Question 11 is 8 marks)

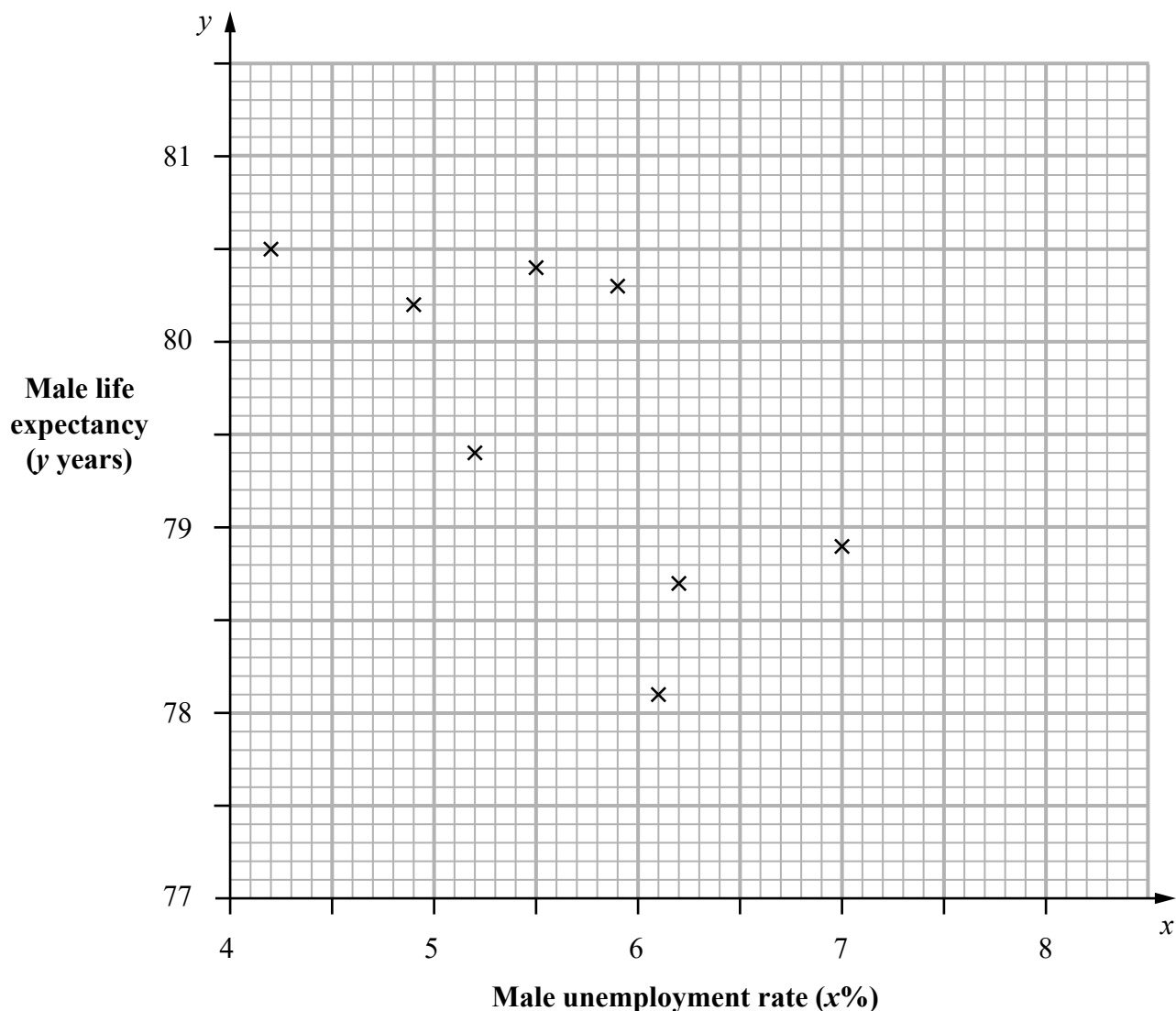
12 Rajesh is investigating to see if there is an association between the male unemployment rate and male life expectancy in the regions of England for 2014

His hypothesis is

“Where the male unemployment rate is high, the male life expectancy is low.”

Rajesh finds the male unemployment rate ($x\%$) and the male life expectancy (y years) for each of eight of the nine regions of England for 2014

This information is shown on the scatter diagram below.



(Source: *ons.gov.uk*)

(a) Explain, giving a statistical reason, whether or not this scatter diagram supports Rajesh’s hypothesis.

.....

.....

.....

(2)

The equation of the regression line for the data in the scatter diagram is $y = 83.5 - 0.7x$

(b) Draw this line on the scatter diagram.

(2)

The male unemployment rates for these eight regions have a mean of 5.6%

(c) Use this information to find the mean of the male life expectancies for these regions.

..... years

(1)

(d) Interpret the value of the gradient of this regression line.

.....
.....
.....
.....

(2)

Rajesh now finds that the male unemployment rate for the missing ninth region is 8.0%

(e) Give **two** reasons why Rajesh should not use the regression line to predict the male life expectancy in this region.

.....
.....
.....
.....
.....

(2)

(Total for Question 12 is 9 marks)

13 A film company employs Gary to investigate the film-watching habits of people living in the UK.

Gary is going to use a questionnaire.

Here is Question 1 on Gary's questionnaire.

Question 1
Spin a fair coin.
If you get **Heads**, tick box A.
If you get **Tails**, answer this question.

Have you downloaded a film illegally during the last month?

If **yes**, tick box A. If **no**, tick box B.

A B

The method used to decide whether or not to answer a question by spinning a coin is called the random response technique.

(a) Explain why this method is used.

.....

.....

.....

.....

(1)

Gary sends the questionnaire to a sample of people living in a town.

He uses a telephone directory as the sample frame.

For Question 1

743 people ticked box A

679 people ticked box B

- (b) Calculate an estimate of the proportion of the people in the sample who had downloaded a film illegally during the last month.

.....
(3)

Gary is going to write a report on the outcome of Question 1.

He is going to use the answer to part (b) as an estimate of the proportion of all the people living in the UK who had downloaded a film illegally during the last month.

- (c) Is it appropriate for Gary to use the answer to part (b)?
Give **two** reasons for your answer.

.....
(3)

(Total for Question 13 is 7 marks)

14 The table gives information about blood donations in the UK.

It shows the blood groups O, A, B and AB and the number of donations for each blood group expressed as a percentage of the total number of all donations.

Blood group	O	A	B	AB
Percentage of all donations	48%	38%	10%	4%

(Source: *blood.co.uk*)

6 people attend a clinic on Monday to donate blood.

(a) (i) Name the probability distribution that can be used to model the number of people from these 6 people who will have blood group AB.

.....

(ii) Write down one condition needed so that this distribution is a suitable model.

.....

.....

(2)

(b) Work out the probability that exactly one of these 6 people will have blood group AB. Give your answer correct to 3 decimal places.

.....

(3)

On Tuesday n people attend the clinic to donate blood.

The probability that at least one of these n people will have blood group AB is greater than 0.5

- (c) What can you conclude, if anything, about the value of n ?
You must show your working.

(3)

(Total for Question 14 is 8 marks)

TOTAL FOR PAPER IS 80 MARKS