

**ADVANCED SUBSIDIARY GCE
MATHEMATICS**

Probability & Statistics 1

4732

QUESTION PAPER

Candidates answer on the printed answer book.

OCR supplied materials:

- Printed answer book 4732
- List of Formulae (MF1)

Other materials required:

- Scientific or graphical calculator

**Thursday 26 May 2011
Morning**

Duration: 1 hour 30 minutes

MODIFIED LANGUAGE

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the printed answer book and the question paper.

- The question paper will be found in the centre of the printed answer book.
- Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Give non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question or is clearly appropriate.

INFORMATION FOR CANDIDATES

This information is the same on the printed answer book and the question paper.

- The number of marks is given in brackets [] at the end of each question or part question on the question paper.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- The printed answer book consists of **12** pages. The question paper consists of **8** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this question paper for marking; it should be retained in the centre or destroyed.

- 1 Five salesmen from a firm were selected at random for a survey. For each salesman, the annual income, x thousand pounds, and the distance driven last year, y thousand miles, were recorded.

The results were summarised as follows.

$$n = 5 \quad \Sigma x = 251 \quad \Sigma x^2 = 14\,323 \quad \Sigma y = 65 \quad \Sigma y^2 = 855 \quad \Sigma xy = 3247$$

- (i) (a) Show that the product moment correlation coefficient, r , between x and y is -0.122 , correct to 3 significant figures. [3]
- (b) State what this value of r shows about the relationship between annual income and distance driven last year for these five salesmen. [1]
- (c) It was decided to recalculate r with the distances measured in kilometres instead of miles. State what effect, if any, this would have on the value of r . [1]
- (ii) Another salesman from the firm is selected at random. His annual income is known to be £52 000, but the distance that he drove last year is unknown. In order to estimate this distance, a regression line based on the above data is used.

Comment on the reliability of such an estimate. [2]

- 2 The orders in which 4 contestants, P , Q , R and S , were placed in two competitions are shown in the table.

Position	1st	2nd	3rd	4th
Competition 1	Q	R	S	P
Competition 2	Q	P	R	S

Calculate Spearman's rank correlation coefficient between these two orders. [5]

- 3 (i) A random variable, X , has the distribution $B(12, 0.85)$. Find
- (a) $P(X > 10)$, [2]
- (b) $P(X = 10)$, [2]
- (c) $\text{Var}(X)$. [2]
- (ii) A random variable, Y , has the distribution $B(2, \frac{1}{4})$. Two independent values of Y are found. Find the probability that the sum of these two values is 1. [4]

- 4 The table shows information about the time, t minutes correct to the nearest minute, taken by 50 people to complete a race.

Time (minutes)	$t \leq 27$	$28 \leq t \leq 30$	$31 \leq t \leq 35$	$36 \leq t \leq 45$	$46 \leq t \leq 60$	$t \geq 61$
Number of people	0	4	28	14	4	0

- (i) In a histogram illustrating the data, the height of the block for the $31 \leq t \leq 35$ class is 5.6 cm. Find the height of the block for the $28 \leq t \leq 30$ class. (There is no need to draw the histogram.) [3]

- (ii) The data in the table are used to estimate the median time.

State, with a reason, whether the estimated median time is more than 33 minutes, less than 33 minutes or equal to 33 minutes. [3]

- (iii) Calculate estimates of the mean and standard deviation of the data. [6]

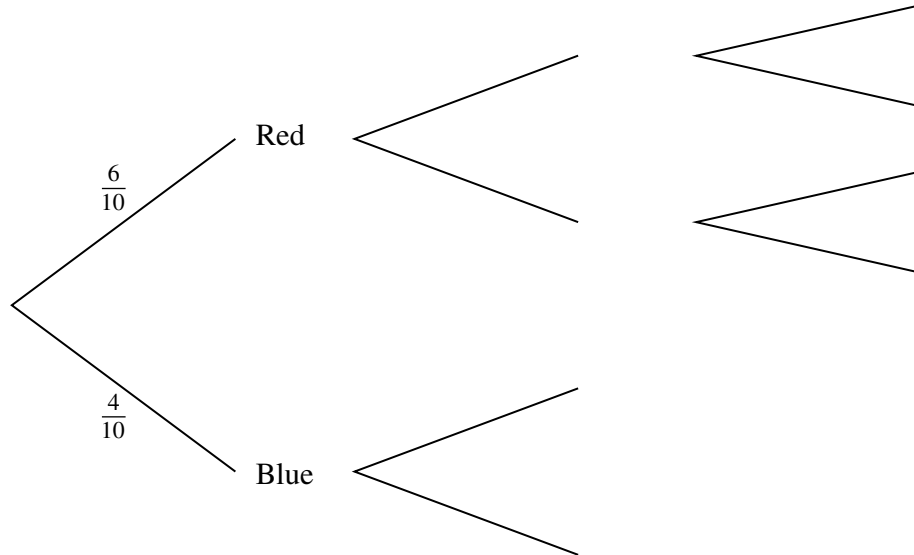
- (iv) It was found that the winner's time had been incorrectly recorded and that it was actually less than 27 minutes 30 seconds.

State whether each of the following will increase, decrease or remain the same:

- (a) the mean, [1]
 (b) the standard deviation, [1]
 (c) the median, [1]
 (d) the interquartile range. [1]

- 5 A bag contains 4 blue discs and 6 red discs. Chloe takes a disc from the bag. If this disc is red, she takes 2 more discs. If it is blue, she takes 1 more disc. Each disc is taken at random and no discs are replaced.

(i) Complete the probability tree diagram in your Answer Book, showing all the probabilities. [2]



The total number of blue discs that Chloe takes is denoted by X .

(ii) Show that $P(X = 1) = \frac{3}{5}$. [2]

The complete probability distribution of X is given below.

x	0	1	2
$P(X = x)$	$\frac{1}{6}$	$\frac{3}{5}$	$\frac{7}{30}$

(iii) Calculate $E(X)$ and $\text{Var}(X)$. [5]

- 6 A group of 7 students sit in random order on a bench.

(i) (a) Find the number of orders in which they can sit. [1]

(b) The 7 students include Tom and Jerry.

Find the probability that Tom and Jerry sit next to each other. [3]

(ii) The students consist of 3 girls and 4 boys.

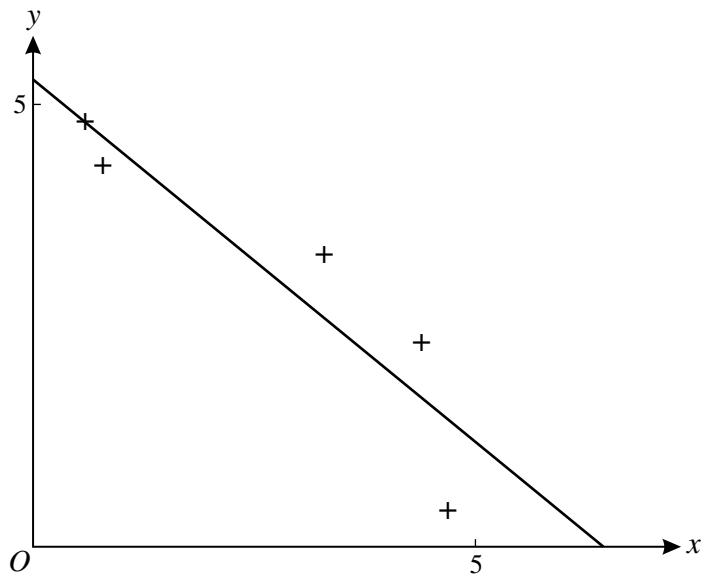
Find the probability that

(a) no two boys sit next to each other, [2]

(b) all three girls sit next to each other. [3]

7 The diagram shows the results of an experiment involving some bivariate data.

The least squares regression line of y on x for these results is also shown.



(i) Given that the least squares regression line of y on x is used for an estimation, state which of x or y is treated as the independent variable. [1]

(ii) Use the diagram to explain what is meant by 'least squares'. [2]

(iii) State, with a reason, the value of Spearman's rank correlation coefficient for these data. [2]

(iv) What can be said about the value of the product moment correlation coefficient for these data? [1]

8 Ann, Bill, Chris and Dipak play a game with a fair six-sided dice. Starting with Ann they take turns, in alphabetical order, to throw the dice. This process is repeated as many times as necessary until a player throws a 6. When this happens, the game stops and this player is the winner.

Find the probability that

(i) Chris wins on his first throw, [1]

(ii) Dipak wins on his second throw, [3]

(iii) Ann gets a third throw, [2]

(iv) Bill throws the dice exactly three times. [4]

BLANK PAGE

BLANK PAGE

**Copyright Information**

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

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