OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2641

Probability & Statistics 1

Thursday **5 JUNE 2003** Morning 1 hour 20 minutes

Additional materials:
Answer booklet
Graph paper
List of Formulae (MF8)

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

- Write your Name, Centre Number and Candidate Number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- 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.
- You are permitted to use a graphic calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is 60.
- Questions carrying smaller numbers of marks are printed earlier in the paper, and questions carrying larger numbers of marks later in the paper.
- **You are reminded of the need for clear presentation in your answers.**

This question paper consists of 4 printed pages.

Registered Charity Number: 1066969

[Turn over

1 Seven A-level projects were marked by a teacher and a moderator. The marks are given in the table below.

Project	A	B	C	D	E	F	G
Teacher's mark	21	19	16	5	11	17	15
Moderator's mark	19	20	11	7	14	18	16

(i) Calculate Spearman's rank correlation coefficient for the data. [4]

(ii) What can be said about the level of agreement between the teacher and the moderator in their assessments of the projects? [1]

2 The random variable X has a Geo(0.4) distribution.

(i) Show that $\mathrm{P}(X = 4) = 0.0864$. [2]

(ii) Calculate the probability $\mathrm{P}(4 \leqslant X < 9)$. [3]

(iii) State the value of $\mathrm{E}(X)$. [1]

3 A standard pack of playing cards consists of 52 distinct cards. Five different cards are selected at random. The order in which the cards are selected does not matter.

(i) Find the number of different possible selections of 5 cards. [1]

(ii) There are 4 suits (clubs, diamonds, hearts and spades) and each suit consists of 13 cards. How many of the selections in part **(i)** consist of 3 spades and 2 clubs? [2]

(iii) How many of the selections in part **(i)** contain exactly 3 spades? [2]

(iv) Calculate the probability that 5 cards selected at random will consist of 3 spades and 2 clubs. [2]

4 For a sample of divorced women, the table below gives the age, a years, at which each woman was divorced.

a	Frequency	Frequency density
$16 \leqslant a < 20$	8	2
$20 \leqslant a < 30$	30	
$30 \leqslant a < 50$	40	
$50 \leqslant a < 70$	16	
$70 \leqslant a < 90$	6	

(i) It is intended to draw a histogram to represent the data. The frequency density of the first class is 2, as shown in the table. Find the frequency density for each of the remaining classes, and draw a histogram on graph paper to represent the data. You should use scales of 1 cm to 5 years on the age axis and 1 cm to 0.2 units on the frequency density axis. [5]

(ii) Calculate estimates of the mean and standard deviation of the ages at which women in the sample were divorced. [5]

5 A jar contains 3 red discs and 2 white discs. A disc is taken at random from the jar and its colour is noted. The disc is **not** replaced. This process is repeated until a white disc is taken. Let D be the number of discs taken, up to and including the white disc.

(i) Show that $\mathrm{P}(D = 2) = \frac{3}{10}$. [2]

(ii) Copy and complete the probability distribution table which is given in a partially completed form below. [3]

d	1	2	3	4
$\mathrm{P}(D = d)$	$\frac{2}{5}$	$\frac{3}{10}$		

(iii) Use the table found in part **(ii)** to calculate $\mathrm{E}(D)$ and $\mathrm{Var}(D)$. [5]

6 Every day I try to 'log on' to the internet. Over 100 days I found that I was successful at the first attempt on 88 days. I will try to log on to the internet each day for the next seven days. Let S be the number of days, out of the seven, on which I am successful at the first attempt.

(i) Suggest a model for the distribution of S, giving the values of any parameters. [2]

(ii) State two assumptions, **in context**, which are required to make this a good model. [2]

(iii) Calculate $\mathrm{P}(S = 4)$. [3]

(iv) Calculate $\mathrm{P}(S > E(S))$. [3]

[Question 7 is printed overleaf.]

7 A researcher took a random sample of 10 GCSE students who had just received their results. She recorded for each student the mean number of hours per week, x, that they spent watching television during their revision period and the number of GCSE 'points', y, that they obtained. The results are given in the table below.

x	15.5	16.9	18.5	19.2	20.3	22.0	25.0	27.5	30.0	30.9
y	49	46	43	40	39	46	25	25	20	18

[$n = 10$, $\Sigma x = 225.8$, $\Sigma y = 351$, $\Sigma x^2 = 5368.90$, $\Sigma y^2 = 13\,577$, $\Sigma xy = 7372.8$.]

(i) On graph paper draw a scatter diagram which illustrates the data. [3]

(ii) Calculate the product moment correlation coefficient for the data. [2]

(iii) Describe the relationship between the mean number of hours of television watched during the students' revision periods and the number of GCSE points that they achieve. State the evidence for your answer. [2]

(iv) Calculate the equation of an appropriate regression line to predict a student's GCSE points score from the mean number of hours of television which the student watched per week during their revision period. [3]

(v) Use the equation of the line found in part **(iv)** to predict the number of GCSE points which a student would achieve if the mean number of hours of television watched per week was 28.1 hours. [1]

(vi) A second researcher also used the set of data in the table but decided to transform the data before doing any calculations. She used $u = \frac{x - 15}{2}$ and $v = y - 50$. State the product moment correlation coefficient for the transformed data used by the second researcher. [1]

1

Project	A	B	C	D	E	F	G
Teacher's Rank	1	2	4	7	6	3	5
Moderator's Rank	2	1	6	7	5	3	4
d_i	-1	1	-2	0	1	0	1

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} = 1 - \frac{6\times 8}{7\times 48} = \frac{6}{7} = \mathbf{0.857} \qquad \text{(3 s.f.)}$$ [4]

There is a **reasonably high level of agreement in their rankings** (maybe not so good agreement in their actual mark allocation) [1]

2 $X \sim \text{Geo}(0.4)$

$$\text{p}(X=4) = (0.6)^3(0.4) = \mathbf{0.0864}$$ [2]

$$\text{p}(4 \le X < 9) = \text{p}(X>3) - \text{p}(X>8) = (0.6)^3 - (0.6)^8 = 0.199203... = \mathbf{0.199} \qquad \text{(3 s.f.)}$$ [3]

$$\text{E}[X] = \tfrac{1}{p} = \mathbf{2.5}$$ [1]

3 number of different selections $= {}^{52}C_5 = \mathbf{2\,598\,960}$ [1]

no. of selections with 3 ♠ and 2 ♣ $= {}^{13}C_3 . {}^{13}C_2 = 286 \times 78 = \mathbf{22\,308}$ [2]

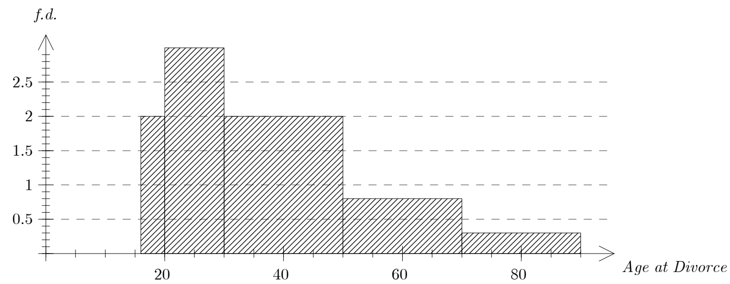
no. of selections with 3 ♠ $= {}^{13}C_3 . {}^{39}C_2 = \mathbf{211\,926}$ [2]

p(3 ♠ and 2 ♣) $= \dfrac{(ii)}{(i)} = \dfrac{22308}{2\,598\,960} = \mathbf{0.00858}$ (3 s.f.) [2]

4

remaining f.d.'s

3, 2, 0.8, 0.3



[5]

$$\text{estimate of mean} = \frac{8\times 18 + 30\times 25 + ... + 6\times 80}{100} = \frac{3934}{100} = \mathbf{39.3} \text{ (3 s.f.)}$$

$$\text{s.d.} = \sqrt{\frac{18^2\times 8 + ... + 80^2 \times 6}{100} - \left(\frac{3934}{100}\right)^2} = \sqrt{\frac{181\,342}{100} - \left(\frac{3934}{100}\right)^2} = \sqrt{265.7844} = \mathbf{16.3}$$ [5]

5 $p(D=2) = p(\text{red } \textit{then} \text{ white}) = \frac{3}{5}\cdot\frac{2}{4} = \mathbf{\frac{3}{10}}$ (show) [2]

d	1	2	3	4
$p(D=d)$	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{3}{5}\cdot\frac{1}{2}\cdot\frac{2}{3}=\mathbf{\frac{1}{5}}$	$\frac{3}{5}\cdot\frac{1}{2}\cdot\frac{1}{3}\cdot 1=\mathbf{\frac{1}{10}}$

[3]

$\mathrm{E}[D] = \sum d_i p_i = \frac{2}{5}+\frac{3}{5}+\frac{3}{5}+\frac{2}{5} = \mathbf{2}$

$\mathrm{Var}[D] = \sum d_i^2 p_i - \mathrm{E}[D]^2 = \left(\frac{2}{5}+\frac{6}{5}+\frac{9}{5}+\frac{8}{5}\right) - 4 = \mathbf{1}$

[5]

6 $S \sim \mathrm{B}(7,\ 0.88)$ [2]

Assumptions

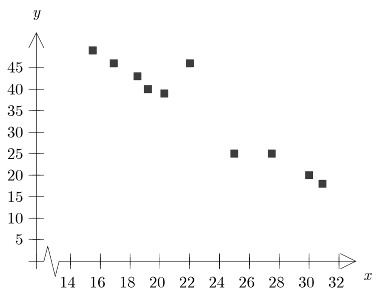
1. constant prob. of success (i.e. there aren't 'good' days & 'bad' days)
2. success on one day independent of success on other days

[2]

$p(S=4) = {}^7C_4\,(0.88)^4\,(0.12)^3 = 0.036269... = \mathbf{0.0363}$ (3 s.f.) [3]

$p(S > \mathrm{E}[S]) = p(S > 6.16) = p(S=7) = 0.88^7 = \mathbf{0.409}$ (3 s.f.) [3]

7



scatter[3]

$$r = \frac{\frac{1}{n}\sum xy - \bar{x}\,\bar{y}}{\sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}}$$

$$= \frac{-55.278}{\sqrt{27.0336}\sqrt{125.69}} = -0.948309... = \mathbf{-0.948}$$

[2]

strong, negative, linear correlation [2]

regression line of y on x

$$y - \bar{y} = \frac{\frac{1}{n}\sum xy - \bar{x}\,\bar{y}}{\frac{1}{n}\sum x^2 - \bar{x}^2}(x - \bar{x}) \qquad y - 35.1 = \frac{-55.278}{27.0336}(x - 22.58)$$

$\boldsymbol{y = 81.27 - 2.045x}$ [3]

$x = 28.1$ $\quad \boldsymbol{\hat{y} = 23.8}$ [1]

pmcc unchanged by linear transformations of x and y, so $\boldsymbol{r = -0.948}$ [1]

[Total 60]