OCR

RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS **2643**

Probability & Statistics 3

Wednesday **23 JANUARY 2002** Afternoon 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 3 printed pages and 1 blank page.

[Turn over

1 In a survey of a random sample of 850 UK teenagers, 325 said that they expected eventually to have a higher standard of living than their parents now have. Calculate an approximate 90% confidence interval for the population proportion of UK teenagers who expect eventually to have a higher standard of living than their parents now have. [5]

2 Sweets of a certain make have wrappers which are coloured either red, blue, green, silver or gold. In a box of 100 of these sweets the numbers of wrappers of each colour are as shown below.

Colour	Red	Blue	Green	Silver	Gold
Frequency	18	13	25	30	14

Assuming that the sweets in the box are an appropriate random sample, use a χ^2 test to test, at the 5% significance level, whether wrappers of the five different colours are equally likely. [5]

3 The eggs laid by hens at a farm have masses which are normally distributed. The masses, x grams, of a random sample of 20 eggs from the farm are summarised by

$$\Sigma x = 1366, \qquad \Sigma x^2 = 94\,366.$$

Test, at the 10% significance level, the claim that the mean mass of the farm's eggs is less than 70 grams. [7]

4 In an investigation into a possible association between hair colour and height, a random sample of 200 adult men was taken, and the data shown in the table was obtained.

		Hair colour			
		Dark	Fair	Red	Total
Height	Less than 165 cm	16	7	11	34
	165 cm to 180 cm	46	39	9	94
	More than 180 cm	33	35	4	72
Total		95	81	24	200

It is proposed to carry out a χ^2 test for independence between hair colour and height.

(i) Calculate the expected frequencies. [2]

(ii) Explain briefly why some combining of rows or columns should be carried out. [1]

(iii) Carry out the test, combining suitable rows and using a 5% significance level. [5]

5 A supermarket sells packs of cauliflower and broccoli. Each pack contains one portion of cauliflower and one portion of broccoli. The mass of cauliflower in a pack has a normal distribution with mean 100 g and standard deviation 10 g. The mass of broccoli in a pack has, independently, a normal distribution with mean 120 g and standard deviation 15 g. Find the probability that, in a randomly chosen pack,

(i) the total mass of the cauliflower and the broccoli exceeds 200 g, [5]

(ii) the mass of the cauliflower is less than three-quarters of the mass of the broccoli. [6]

6 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the (cumulative) distribution function of X. [3]

(ii) Hence find the (cumulative) distribution function of Y, where $Y = \frac{1}{X}$, and deduce that the probability density function of Y is given by

$$g(y) = \begin{cases} \dfrac{1}{2y^3} & \frac{1}{2} < y, \\ 0 & \text{otherwise.} \end{cases}$$ [6]

(iii) Find $\mathrm{E}\left(\frac{1}{X}\right)$. [2]

7 A new variety of genetically modified tomato is claimed to have an average shelf-life that is at least 15 days longer than that of an unmodified variety. Data from a random sample of 12 tomatoes of the new variety and a random sample of 14 unmodified tomatoes is summarised by

$$\Sigma x = 386, \quad \Sigma x^2 = 12\,536, \quad n_x = 12,$$
$$\Sigma y = 257, \quad \Sigma y^2 = 4787, \quad n_y = 14,$$

where x and y denote the shelf-lives, in days, of the new variety of tomatoes and the unmodified variety, respectively.

(i) State what needs to be assumed about the distributions of shelf-lives in order to justify the use of a t-test to test the claim about the increase in shelf-life. [2]

(ii) Carry out the test, using a 5% significance level. [7]

(iii) Calculate a 95% confidence interval for the increase in mean shelf-life. [4]

S3 jan 02

1. $X =$ number of positives out of the sample.

$X \sim B(n, p) \approx N(np, npq)$

We use $\frac{X}{n}$ to estimate p.

$\frac{X}{n} \sim N\left(p, \frac{pq}{n}\right)$.

$$P\left(-1.645 < \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} < 1.645\right) = 90\%$$

leading to 90% c.i. of $p_s \pm 1.645\sqrt{\frac{pq}{n}}$

in this case $\frac{325}{850} \pm 1.645\sqrt{\frac{\frac{325}{850} \times \frac{525}{850}}{850}}$

i.e. (0.355, 0.410)

---

2. $H_0: p_i = \frac{1}{5}$ for each $i = 1, ..5$

H_1: they are not equally likely.

each e_i is clearly 20 and

$$X^2 = \frac{2^2}{20} + \frac{7^2}{20} + \frac{5^2}{20} + \frac{10^2}{20} + \frac{6^2}{20} = 10.7$$

$\nu = 4$ and the critical value at the 5% level is 9.488.

There is significant evidence that the different colours are not all equally likely.

---

3. H_0: mean mass of eggs is 70g

H_1: mean mass < 70g

$$s^2 = \frac{20}{19}\left(\frac{94366}{20} - \left(\frac{1366}{20}\right)^2\right) = 56.22..$$

$$t = \frac{68.3 - 70}{\sqrt{\frac{56.22}{20}}} = -1.0139..$$

The critical t- value for $\nu = 19$ is -1.729.

Our statistic is well within the acceptance region, and there is not significant evidence of mean weight under 70g.

---

4. (i)

e_i's are:

16.15	13.77	4.08
44.65	38.07	11.28
34.2	29.16	8.64

(ii) $e_{1,3} < 5$ so rows or columns must be combined, so let's combine rows 1 and 2:

o_i's are

(iii)

62	46	20
33	35	4

while e_i's are:

60.8	51.84	15.36
34.2	29.16	8.64

so $X^2 = 5.79$ and
$\nu = (3-1)(2-1) = 2$
Critical value of χ_2^2 is 5.991 at the 5% level, and this is just within acceptance region, so there is not quite enough evidence to support a link between hair colour and height.

---

5. (i) With usual definitions, $B + C \sim N(120 + 100, 15^2 + 10^2)$
i.e. $N(220, 325)$.

$$P(B + C > 200) = 1 - \Phi\left(\frac{200 - 220}{\sqrt{325}}\right)$$
$= 1 - \Phi(-1.109..)$
$= 0.866$

(ii) $P\left(C < \frac{3}{4}B\right) = P(4C - 3B < 0)$
$4C - 3B \sim N(40, 4^2 \times 10^2 + (-3)^2 \times 15^2)$
i.e. $N(40, 3625)$
so $P\left(C < \frac{3}{4}B\right) = \Phi\left(\frac{0-40}{\sqrt{3625}}\right)$
$= \Phi(-0.6643..)$
$= 0.253$

---

6. (i)

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x^2 & 0 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

(ii) $F(y) = P(Y < y)$
$= P\left(\frac{1}{X} < y\right)$
clearly if $y < 0$ this probability is $= 0$.
For $y > 0$, the probability becomes
$= P(X > \frac{1}{y})$.
$F(y)$ will be a tri-partite function with different forms in domains given by

$\frac{1}{y} < 0$

$0 \leq \frac{1}{y} \leq 2$

$\frac{1}{y} > 2$

which becomes

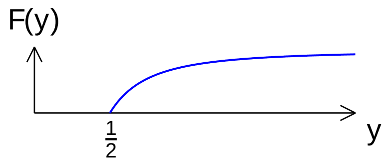
$y < 0$

$y > \frac{1}{2}$

$0 \leq y \leq \frac{1}{2}$

and so

$$F(y) = \begin{cases} 0 & y < 0 \\ 0 & 0 \leq y \leq \frac{1}{2} \\ 1 - \frac{1}{4y^2} & y > \frac{1}{2} \end{cases}$$



Differentiating in each part of the domain gives the p.d.f. which is

$$f(y) = \begin{cases} 0 & y \leq \frac{1}{2} \\ \frac{1}{2y^3} & y > \frac{1}{2} \end{cases}$$

(iii) $E\left(\frac{1}{X}\right)$ can be done by noting it is $E(Y)$ which is

$$\int_{\frac{1}{2}}^{\infty} y \times \frac{1}{2y^3}\,dy$$

but more simply by $E\left(\frac{1}{X}\right) = \int_0^2 \frac{1}{x} \times \frac{1}{2} x\,dx = 1$

---

7. (i) The shelf lives need to be distributed Normally; we also need to assume common variance for this test.

(ii) $H_0: \mu_B - \mu_A = 15$

$H_1: \mu_B - \mu_A < 15$

where μ_A and μ_B are the expected shelf-lives of unmodified and modified tomatoes respectively.

$$est(\sigma^2) = \frac{12536 - \frac{1}{12} \times 386^2 + 4787 - \frac{1}{14} \times 257^2}{12 + 14 - 2}$$

$= 7.87..$

$$t_{24} = \frac{\bar{x} - \bar{y} - 15}{\hat{\sigma}\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$$

$$= \frac{(32.16 - 18.36) - 15}{\sqrt{7.87\left(\frac{1}{12} + \frac{1}{14}\right)}}$$

$= -1.0787..$

$t_{crit} = -1.711$

so there is not significant evidence to reject the claim of 15 days longer shelf life.

(iii)

$$P\left(-2.064 < \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\hat{\sigma}\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} < 2.064\right) = 95\%$$

leading to 95% confidence interval for increase in mean shelf-lifeof

$$13.80952.. \pm 2.064 \times \sqrt{7.87\left(\frac{1}{12} + \frac{1}{14}\right)}$$

i.e. $(11.5, 16.1)$ days