Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Thursday 15 May 2008 – Morning

Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Green)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
There are 7 questions in this question paper. The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner.
Answers without working may gain no credit.

1. A disease is known to be present in 2% of a population. A test is developed to help determine whether or not someone has the disease.

Given that a person has the disease, the test is positive with probability 0.95.

Given that a person does not have the disease, the test is positive with probability 0.03.

(*a*) Draw a tree diagram to represent this information.

(3)

A person is selected at random from the population and tested for this disease.

(*b*) Find the probability that the test is positive.

(3)

A doctor randomly selects a person from the population and tests him for the disease. Given that the test is positive,

(*c*) find the probability that he does not have the disease.

(2)

(*d*) Comment on the usefulness of this test.

(1)

---

2. The age in years of the residents of two hotels are shown in the back to back stem and leaf diagram below.

Abbey Hotel 8 | 5 | 0 means 58 years in Abbey Hotel and 50 years in Balmoral Hotel Balmoral Hotel

(1)	2	0		
(4)	9 7 5 1	1		
(4)	9 8 3 1	2	6	(1)
(11)	9 9 9 9 7 6 6 5 3 3 2	3	4 4 7	(3)
(6)	9 8 7 7 5 0	4	0 0 5 5 6 9	(6)
(1)	8	5	0 0 0 0 1 3 6 6 7	(9)
		6	2 3 3 4 5 7	(6)
		7	0 1 5	(3)

For the Balmoral Hotel,

(*a*) write down the mode of the age of the residents,

(1)

(*b*) find the values of the lower quartile, the median and the upper quartile.

(3)

(*c*) (i) Find the mean, $\bar{x}$, of the age of the residents.

(ii) Given that $\sum x^2 = 81\,213$, find the standard deviation of the age of the residents.

(4)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

(*d*) Evaluate this measure for the Balmoral Hotel.

(2)

For the Abbey Hotel, the mode is 39, the mean is 33.2, the standard deviation is 12.7 and the measure of skewness is –0.454.

(*e*) Compare the two age distributions of the residents of each hotel.

(3)

3. The random variable X has probability distribution given in the table below.

x	–1	0	1	2	3
$P(X = x)$	p	q	0.2	0.15	0.15

Given that $E(X) = 0.55$, find

(a) the value of p and the value of q,

(5)

(b) Var (X),

(4)

(c) $E(2X - 4)$.

(2)

---

4. Crickets make a noise. The pitch, v kHz, of the noise made by a cricket was recorded at 15 different temperatures, t °C. These data are summarised below.

$$\sum t^2 = 10\,922.81, \quad \sum v^2 = 42.3356, \quad \sum tv = 677.971, \quad \sum t = 401.3, \quad \sum v = 25.08$$

(a) Find S_{tt}, S_{vv} and S_{tv} for these data.

(4)

(b) Find the product moment correlation coefficient between t and v.

(3)

(c) State, with a reason, which variable is the explanatory variable.

(2)

(d) Give a reason to support fitting a regression model of the form $v = a + bt$ to these data.

(1)

(e) Find the value of a and the value of b. Give your answers to 3 significant figures.

(4)

(f) Using this model, predict the pitch of the noise at 19 °C.

(1)

---

5. A person's blood group is determined by whether or not it contains any of 3 substances *A*, *B* and *C*.

A doctor surveyed 300 patients' blood and produced the table below.

Blood contains	**No. of Patients**
only *C*	100
A and *C* but not *B*	100
only *A*	30
B and *C* but not *A*	25
only *B*	12
A, *B* and *C*	10
A and *B* but not *C*	3

(*a*) Draw a Venn diagram to represent this information.

(4)

(*b*) Find the probability that a randomly chosen patient's blood contains substance *C*.

(2)

Harry is one of the patients. Given that his blood contains substance *A*,

(*c*) find the probability that his blood contains all 3 substances.

(2)

Patients whose blood contains none of these substances are called universal blood donors.

(*d*) Find the probability that a randomly chosen patient is a universal blood donor.

(2)

6. The discrete random variable X can take only the values 2, 3 or 4. For these values the cumulative distribution function is defined by

$$F(x) = \frac{(x+k)^2}{25} \text{ for } x = 2, 3, 4,$$

where k is a positive integer.

(*a*) Find k.

(2)

(*b*) Find the probability distribution of X.

(3)

---

7. A packing plant fills bags with cement. The weight X kg of a bag of cement can be modelled by a normal distribution with mean 50 kg and standard deviation 2 kg.

(*a*) Find $P(X > 53)$.

(3)

(*b*) Find the weight that is exceeded by 99% of the bags.

(5)

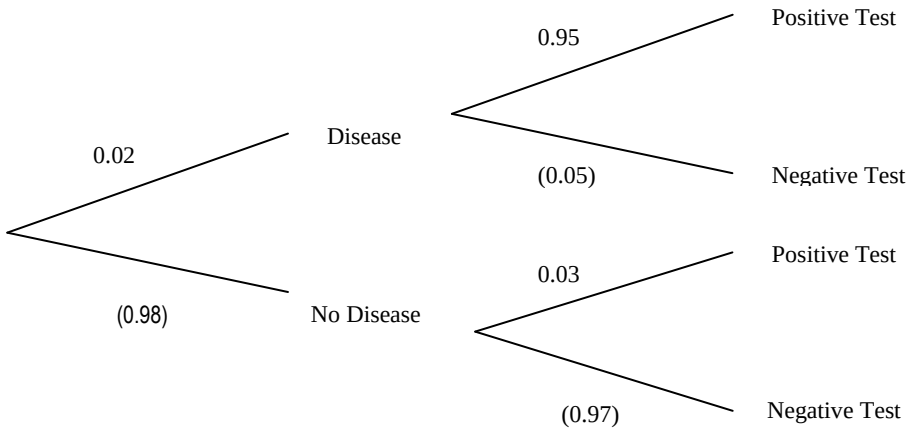
Three bags are selected at random.

(*c*) Find the probability that two weigh more than 53 kg and one weighs less than 53 kg.

(4)

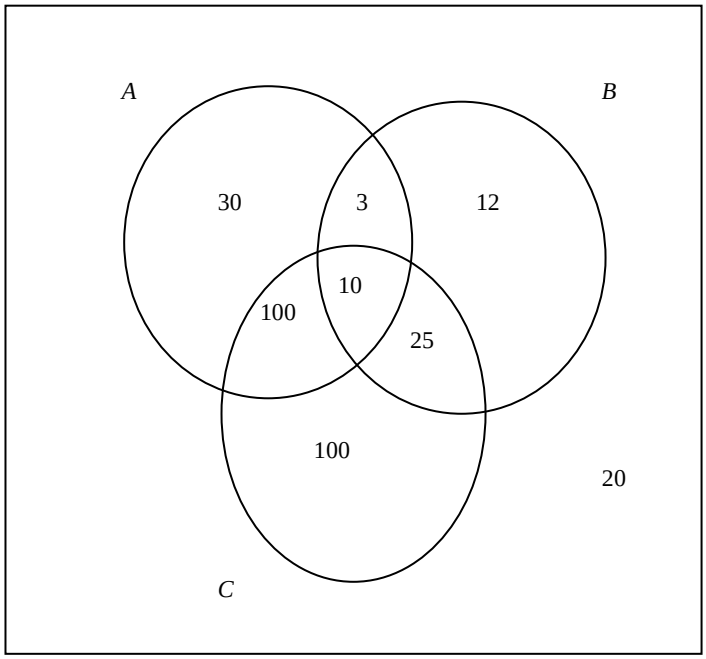
TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)		
	Tree without probabilities or labels	M1
	0.02(Disease), 0.95(Positive) on correct branches	A1
	0.03(Positive) on correct branch.	A1 (3)
(b)	P(Positive Test) $=0.02\times0.95+0.98\times0.03$	M1A1ft
	$=0.0484$	A1 (3)
(c)	P(Do not have disease | Postive test) $=\dfrac{0.98\times0.03}{0.0484}$	M1
	$=0.607438..$ awrt 0.607	A1 (2)
(d)	Test not very useful *or* High probability of not having the disease for a person with a positive test	B1 (1)
		(9 marks)

Question Number		Scheme	Marks
2.	(a)	50	B1 (1)
	(b)	$Q_1 = 45$	B1
		$Q_2 = 50.5$ ONLY	B1
		$Q_3 = 63$	B1 (3)
	(c)	Mean $= \frac{1469}{28} = 52.464286..$ awrt 52.5	M1 A1
		Sd $= \sqrt{\frac{81213}{28} - \left(\frac{1469}{28}\right)^2}$	M1
		$=12.164\ldots$ or $12.387216\ldots$ for divisor n-1 awrt 12.2 or 12.4	A1 (4)
	(d)	$\frac{52.46.. - 50}{sd} =$ awrt 0.20 or 0.21	M1 A1 (2)
	(e)	1. mode/median/mean Balmoral>mode/median/mean Abbey 2. Balmoral sd < Abbey sd or similar sd or correct comment from their values, Balmoral range<Abbey range, Balmoral IQR>Abbey IQR or similar IQR 3. Balmoral positive skew or almost symmetrical AND Abbey negative skew, Balmoral is less skew than Abbey or correct comment from their value in (d) 4. Balmoral residents generally older than Abbey residents or equivalent. Only one comment of each type max 3 marks	B1 B1 B1 (3)
			(13 marks)
3.	(a)	$-1\times p + 1\times 0.2 + 2\times 0.15 + 3\times 0.15 = 0.55$	M1 M1
		$p = 0.4$	A1
		$p + q + 0.2 + 0.15 + 0.15 = 1$	M1
		$q = 0.1$	A1 (5)
	(b)	$\text{Var}(X) = (-1)^2 \times p + 1^2 \times 0.2 + 2^2 \times 0.15 + 3^2 \times 0.15, -0.55^2$	M1A1,M1
		$= 2.55 - 0.3025 = 2.2475$ awrt 2.25	A1 (4)
	(c)	E(2X-4) =2E(X)-4	M1
		= –2.9	A1 (2)
			(11 marks)

Question Number		Scheme		Marks
4.	(a)	$S_{tt} = 10922.81 - \frac{401.3^2}{15} = 186.6973$	awrt 187	M1A1
		$S_{vv} = 42.3356 - \frac{25.08^2}{15} = 0.40184$	awrt 0.402	A1
		$S_{tv} = 677.971 - \frac{401.3 \times 25.08}{15} = 6.9974$	awrt 7.00	A1 (4)
	(b)	$r = \frac{6.9974}{\sqrt{186.6973 \times 0.40184}}$		M1A1ft
		$= 0.807869$	awrt 0.808	A1 (3)
	(c)	*t* is the explanatory variable as we can control temperature but not frequency of noise or equivalent comment		B1 B1 (2)
	(d)	High value of *r* or *r* close to 1 or strong correlation		B1 (1)
	(e)	$b = \frac{6.9974}{186.6973} = 0.03748$	awrt 0.0375	M1 A1
		$a = \frac{25.08}{15} - b \times \frac{401.3}{15} = 0.6692874$	awrt 0.669	M1 A1 (4)
	(f)	$t = 19$, $v = 0.6692874 + 0.03748 \times 19 = 1.381406$	awrt 1.4	B1 (1)
				(15 marks)

Question Number	Scheme	Marks
5. (a)	 3 closed intersecting curves with labels 100 100,30 12,10,3,25 Box	M1 A1 A1 B1 (4)
(b)	P(Substance C) $= \frac{100+100+10+25}{300} = \frac{235}{300} = \frac{47}{60}$ or exact equivalent	M1 A1ft (2)
(c)	P(All 3 $\mid A$) $= \frac{10}{30+3+10+100} = \frac{10}{143}$ or exact equivalent	M1 A1ft (2)
(d)	P(Universal donor) $= \frac{20}{300} = \frac{1}{15}$ or exact equivalent	M1 A1 cao (2) **(10 marks)**
6. (a)	F(4) = 1, $(4+k)^2 = 25$ $k = 1$ as $k > 0$	M1 A1 (2)
(b)	x: 2, 3, 4 P(X=x): $\frac{9}{25}$, $\frac{7}{25}$, $\frac{9}{25}$	B1ft B1 B1 (2) **(5 marks)**

Question Number		Scheme		Marks
7.	(a)	$z = \frac{53-50}{2}$	Attempt to standardise	M1
		$P(X>53) = 1 - P(Z < 1.5)$	1-probability required can be implied	B1
		$=1-0.9332$		A1 (3)
		$=0.0668$		
	(b)	$P(X \le x_0) = 0.01$		M1
		$\frac{x_0-50}{2} = -2.3263$		M1 B1
		$x_0 = 45.3474$	awrt 45.3 or 45.4	M1 A1 (5)
	(c)	$P(\text{2 weigh more than 53kg and 1 less}) = 3\times 0.0668^2(1-0.0668)$		B1 M1 A1ft
		$= 0.012492487..$	awrt 0.012	A1 (4)
				(12 marks)