Edexcel GCE

Statistics S1

Advanced Subsidiary

Tuesday 15 January 2008 – Morning

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Green)	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

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name 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.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions on this 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 personnel manager wants to find out if a test carried out during an employee's interview and a skills assessment at the end of basic training is a guide to performance after working for the company for one year.

The table below shows the results of the interview test of 10 employees and their performance after one year.

Employee	A	B	C	D	E	F	G	H	I	J
Interview test, x %	65	71	79	77	85	78	85	90	81	62
Performance after one year, y %	65	74	82	64	87	78	61	65	79	69

[You may use $\sum x^2 = 60\,475$, $\sum y^2 = 53\,122$, $\sum xy = 56\,076$]

(a) Showing your working clearly, calculate the product moment correlation coefficient between the interview test and the performance after one year.

(5)

The product moment correlation coefficient between the skills assessment and the performance after one year is –0.156 to 3 significant figures.

(b) Use your answer to part (a) to comment on whether or not the interview test and skills assessment are a guide to the performance after one year. Give clear reasons for your answers.

(2)

2. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	A	B	C	D	E	F	G	H	I	J	K	L
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724\,961$]

(a) Find the mean and standard deviation of the level of cotinine in a patient's blood.

(4)

(b) Find the median, upper and lower quartiles of these data.

(3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3 + 1.5(Q_3 - Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

(c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.

(4)

Research suggests that cotinine levels in the blood form a skewed distribution.

One measure of skewness is found using $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$.

(d) Evaluate this measure and describe the skewness of these data.

(3)

3

3. The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

Frequency density

10

9

8

7

6

5

4

3

2

1

60.5 62.5 64.5 66.5 68.5 70.5 72.5 74.5 76.5 78.5 80.5 82.5 84.5 86.5 88.5 90.5   time
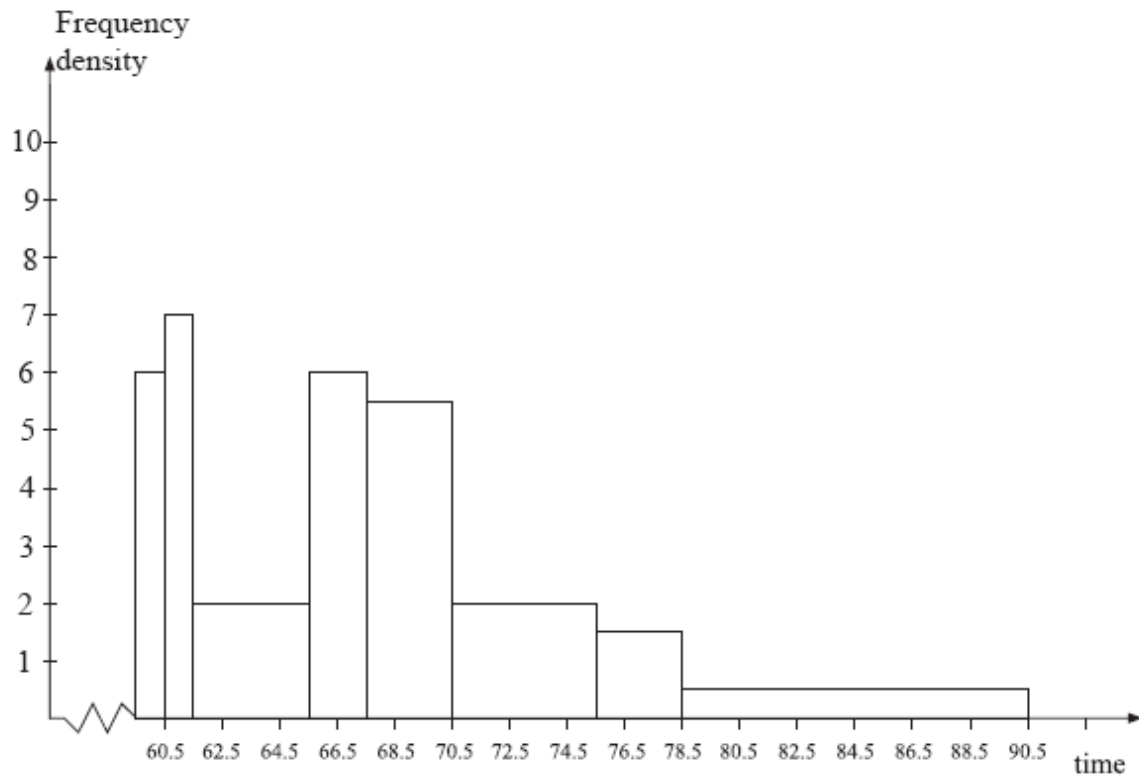
Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(5)

4. A second hand car dealer has 10 cars for sale. She decides to investigate the link between the age of the cars, x years, and the mileage, y thousand miles. The data collected from the cars are shown in the table below.

Age, x (years)	2	2.5	3	4	4.5	4.5	5	3	6	6.5
Mileage, y (thousands)	22	34	33	37	40	45	49	30	58	58

[You may assume that $\sum x = 41,\ \sum y = 406,\ \sum x^2 = 188,\ \sum xy = 1818.5$]

(a) Find S_{xx} and S_{xy}.

(3)

(b) Find the equation of the least squares regression line in the form $y = a + bx$. Give the values of a and b to 2 decimal places.

(4)

(c) Give a practical interpretation of the slope b.

(1)

(d) Using your answer to part (b), find the mileage predicted by the regression line for a 5 year old car.

(2)

5.   The following shows the results of a wine tasting survey of 100 people.

96 like wine *A*,
93 like wine *B*,
96 like wine *C*,
92 like *A* and *B*,
91 like *B* and *C*,
93 like *A* and *C*,
90 like all three wines.

(*a*) Draw a Venn Diagram to represent these data.

(6)

Find the probability that a randomly selected person from the survey likes

(*b*) none of the three wines,

(1)

(*c*) wine *A* but not wine *B*,

(2)

(*d*) any wine in the survey except wine *C*,

(2)

(*e*) exactly two of the three kinds of wine.

(2)

Given that a person from the survey likes wine *A*,

(*f*) find the probability that the person likes wine *C*.

(3)

---

6.   The weights of bags of popcorn are normally distributed with mean of 200 g and 60% of all bags weighing between 190 g and 210 g.

(*a*) Write down the median weight of the bags of popcorn.

(1)

(*b*) Find the standard deviation of the weights of the bags of popcorn.

(5)

A shopkeeper finds that customers will complain if their bag of popcorn weighs less than 180 g.

(*c*) Find the probability that a customer will complain.

(3)

---

7. Tetrahedral dice have four faces. Two fair tetrahedral dice, one red and one blue, have faces numbered 0, 1, 2, and 3 respectively. The dice are rolled and the numbers face down on the two dice are recorded. The random variable R is the score on the red die and the random variable B is the score on the blue die.

(*a*) Find $P(R = 3 \text{ and } B = 0)$.

(2)

The random variable T is R multiplied by B.

(*b*) Complete the diagram below to represent the sample space that shows all the possible values of T.

B				
3				
2		2		
1	0			
0				
R	**0**	**1**	**2**	**3**

Sample space diagram of T

(3)

The table below represents the probability distribution of the random variable T.

t	0	1	2	3	4	6	9
$P(T = t)$	a	b	$\frac{1}{8}$	$\frac{1}{8}$	c	$\frac{1}{8}$	d

(*c*) Find the values of a, b, c and d.

(3)

Find the values of

(*d*) $E(T)$,

(2)

(*e*) $Var(T)$.

(4)

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)	$\sum x = 773, \sum y = 724$ $r = \dfrac{10 \times 56076 - 773 \times 724}{\sqrt{(10 \times 60475 - 773^2)(10 \times 53122 - 724^2)}}$ o.e. $r = 0.155357.....$	**B1, B1** **M1 A1ft** **A1** **(5)**
(b)	Both weak correlation Neither score is a good indication of future performance Interview test is slightly better since correlation is positive	**B1g B1h** **(2)** **Total 7 marks**

NB
 	
(a)
 	
(b)
	$\mathrm{S}_{xx} = 60475 - \dfrac{(773)^2}{10} = 722.1, \ \mathrm{S}_{yy} = 53122 - \dfrac{(724)^2}{10} = 704.4, \ \mathrm{S}_{xy} = 56076 - \dfrac{773 \times 724}{10} = 110.8$
	
	 1st B1 for $\sum x$ and 2nd B1 for $\sum y$, should be seen or implied.
	
	 M1 for at least one correct attempt at one of $\mathrm{S}_{xx}, \ \mathrm{S}_{yy} \ $ or $\ \mathrm{S}_{xy}$ and then using in the correct formula
	
	 1st A1ft for a fully correct expression. (ft their Σx and their Σy) or 3 correct expressions for $\mathrm{S}_{xx}, \mathrm{S}_{xy},$ and S_{yy} but possibly incorrect values for these placed correctly in r.
	 2nd A1 for awrt 0.155
	
	     If $\lvert r \rvert > 0.5$ they can score B1g in (b) for saying that it (skills test) is not a good guide to performance but B0h since a second acceptable comment about both tests is not possible.
	
	     Give B1 for one correct line, B1B1 for any 2.
	 If the only comment is the test(s) <u>are</u> a good guide: scores B0B0
	 If the only comment is the tests are not good: scores B1B0 (second line)
	
	 The third line is for a comment that suggests that the interview test is OK but the skills test is not since one is positive and the other is negative.
	
	 Treat 1st B1 as B1g and 2nd as B1h
	
	     An answer of "no" alone scores B0B0	

Question Number	Scheme	Marks
2.		
(a)	mean is $\dfrac{2757}{12}, = 229.75$ AWRT 230	**M1, A1**
	sd is $\sqrt{\dfrac{724961}{12} - (229.75)^2}, = 87.34045$ AWRT 87.3	**M1, A1**
	[Accept s = AWRT 91.2]	
		(4)
(b)	Ordered list is: 125, 160, 169, 171, 175, 186, 210, 243, 250, 258, 390, 420	
	$Q_2 = \frac{1}{2}(186 + 210) = 198$	**B1**
	$Q_1 = \frac{1}{2}(169 + 171) = 170$	**B1**
	$Q_3 = \frac{1}{2}(250 + 258) = 254$	**B1**
		(3)
(c)	$Q_3 + 1.5(Q_3 - Q_1) = 254 + 1.5(254 - 170), = 380$ Accept AWRT (370-392)	**M1, A1**
	Patients F (420) and B (390) are outliers.	**B1ft B1ft**
		(4)
(d)	$\dfrac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1} = \dfrac{170 - 2 \times 198 + 254}{254 - 170}, = 0.\dot{3}$ AWRT 0.33	**M1, A1**
	Positive skew.	**A1ft**
		(3)
		Total 14 marks

(a)	1st M1 for using $\dfrac{\sum x}{n}$ with a credible numerator and n = 12.	
	2nd M1 for using a correct formula, root required but can ft their mean	
NB	Use of $s = \sqrt{8321.84...} = 91.22...$ is OK for M1A1 here.	
	Answers only from a calculator in (a) can score full marks	
(b)	1st B1 for median= 198 only, 2nd B1 for lower quartile 3rd B1 for upper quartile	
S.C.	If all Q_1 and Q_3 are incorrect but an ordered list (with $\geq$ 6 correctly placed) is seen and used then award B0B1 as a special case for these last two marks.	
(c)	M1 for a clear attempt using their quartiles in given formula,	
	A1 for any value in the range 370 - 392	
	1st B1ft for any one correct decision about B or F - ft their limit in range (258, 420)	
	2nd B1ft for correct decision about both F and B - ft their limit in range (258, 420)	
	If more points are given score B0 here for the second B mark.	
	(Can score M0A0B1B1 here)	
(d)	M1 for an attempt to use their figures in the correct formula – must be seen	
	($\geq$ 2 correct substitutions)	
	1st A1 for AWRT 0.33	
	2nd A1ft for positive skew. Follow through their value/sign of skewness .	
	Ignore any further calculations.	
	"positive correlation" scores A0	

3.	Width	1	1	4	2	3	5	3	12	**M1**
	Freq. Density	6	7	2	6	5.5	2	1.5	0.5	

0.5×12 or 6 **A1**

Total area is $(1\times6)+(1\times7)+(4\times2)+....,= 70$

$$\left(90.5 - 78.5\right)\times\tfrac{1}{2}\times\dfrac{140}{\text{their }70}$$ **M1**

"70 seen anywhere"   **B1**

Number of runners is 12   **A1**

(5)
Total 5 marks

1st M1   for attempt at width of the correct bar (90.5 - 78.5)
             [Maybe on histogram or in table]

1st A1   for 0.5×12 or 6 (may be seen on the histogram. Must be related to the area
             of the bar above 78.5 - 90.5.

2nd M1   for attempting area of correct bar $\times\dfrac{140}{\text{their }70}$

B1   for 70 seen anywhere in their working

2nd A1   for correct answer of 12.

Minimum working required is $2\times0.5\times12$ where the 2 should come from $\dfrac{140}{70}$

Beware 90.5 - 78.5 = 12 (this scores M1A0M0B0A0)

Common answer is $0.5\times12= 6$ (this scores M1A1M0B0A0)

If unsure send to review e.g. $2 \times0.5 \times 12=12$ without 70 being seen

4.				
(a)	$S_{xy} = 1818.5 - \dfrac{41 \times 406}{10}, = 153.9$ (could be seen in (b)) AWRT 154			**M1, A1**
	$S_{xx} = 188 - \dfrac{41^2}{10} = 19.9$ (could be seen in (b))			**A1**
				(3)
(b)	$b = \dfrac{153.9}{19.9}, = 7.733668....$ AWRT 7.73			**M1, A1**
	$a = 40.6 - b \times 4.1(= 8.89796....)$			**M1**
	$y = 8.89 + 7.73x$			**A1**
				(4)
(c)	A typical car will travel 7700 miles every year			**B1ft**
				(1)
(d)	$x = 5, y = 8.89 + 7.73 \times 5(= 47.5 - 47.6)$			**M1**
	So mileage predicted is AWRT 48000			**A1**
				(2)
				Total 10 marks

Accept calculations for S_{xx} **and** S_{xy} **in (a) or (b)**

(a)	M1	for correct attempt or expression for either
	1st A1	for one correct
	2nd A1	for both correct

(b)   **Ignore the epen marks for part (b) they should be awarded as per this scheme**

	1st M1	for $\dfrac{\text{their S}_{xy}}{\text{their S}_{xx}}$
	1st A1	for AWRT 7.73
	2nd M1	for attempt at correct formula for *a* (minus required). Ft their *b*. Quoting a correct formula but making one slip in sub.eg. $\bar{y}$ =406 is OK
	2nd A1	for correct equation with 2dp accuracy. Accept *a* = 8.89, and *b* = 7.73 even if not written as final equation.

Correct answers only (from calc) score 4/4 if correct to 2dp or 3/4 if AWRT 2dp

(c)	B1ft	for their $b \times 1000$ to at least 2 sf. Accept "7.7 thousand" but value is needed
(d)	M1	for substituting $x = 5$ into their final answer to (b).
	A1	for AWRT 48000 (Accept "48 thousands")

5.	**Diagram may be drawn with $B \subset (A \cup C)$ or with the 0 for $B \cap (A \cup C)'$ simply left blank**		
(a)			

Accept decimals or probs. in Venn diagram

3cc   **M1**
90,3,2,1   **A1**
1,(0),2   **M1A1**
1 outside   **A1**
Box   **B1**

(6)

(b)	P(none)=0.01	**B1ft**
		(1)
(c)	P(*A* but not *B*)=0.04	**M1 A1ft**
		(2)
(d)	P(any wine but *C*)=0.03	**M1A1ft**
		(2)
(e)	P(exactly two)=0.06	**M1A1ft**
		(2)
(f)	$P(C\vert A) = \dfrac{P(C \cap A)}{P(A)} =, \dfrac{93}{96}$ or $\dfrac{31}{32}$ or AWRT 0.969	**M1A1ft,A1**
		(3)
		Total 16 marks

(a)	1st M1 for 3 closed, labelled curves that overlap. A1 for the 90, 3, 2 and 1
	2nd M1 for one of 1, 0 or 2 correct <u>or</u> a correct sum of 4 values for *A, B* or *C*
	2nd A1 for all 7 values correct. Accept a blank instead of 0.
	NB final mark is a B1 for the box not an A mark as on EPEN
	In parts (b) to (f) full marks can be scored for correct answers or correct ft

(b)	B1ft	Follow through their '1'from outside divided by 100	
(c)	M1	for correct expression eg $P(A \cup B) - P(B)$ or calculation e.g. 3 + 1 or 4 on top	For M marks in (c) to (e) they must have a fraction
	A1	for a correct probability, follow through with their '3+1' from diagram	
(d)	M1	for correct expression or calculation e.g. 1+2+0 or 99-96 or 3 on top	
	A1	for a correct probability, follow through their '2+1+0' from diagram	
(e)	M1	for a correct expression or calculation e.g. 3+2+1 or 6 on top	
(f)	M1	for a correct expression upto "," and <u>some</u> correct substitution, ft their values. One of these probabilities must be correct or correct ft. If P(*C*) on bottom M0	
	1st A1ft	follow through their $A \cap C$ and their *A* but the ratio must be in (0, 1)	
	2nd A1	for correct answer only. Answer only scores 3/3, but check working P($A \cap C$)/P(C) is M0	

6. (a)	200 or 200g	**B1** **(1)**
(b)	P(190 < X < 210) = 0.6 or P(X < 210) = 0.8 or P(X > 210)= 0.2 or diagram (o.e.) Correct use of 0.8 or 0.2 $Z = (\pm)\dfrac{210 - 200}{\sigma}$ $\dfrac{10}{\sigma} = 0.8416$ 0.8416 $\sigma = 11.882129....$ AWRT 11.9	**M1** **A1** **M1** **B1** **A1** **(5)**
(c)	$P(X < 180) \quad = P\left(Z < \dfrac{180 - 200}{\sigma} \right)$ $= P(Z < -1.6832)$ $= 1 - 0.9535$ $= 0.0465 \quad$ or $\quad$ AWRT 0.046	**M1** **M1** **A1** **(3)** **Total 9 marks**
(a)	"mean = 200g" is B0 but "median = 200" or just "200" alone is B1 **Standardization in (b) and (c).** They must use σ not σ^2 or $\sqrt{\sigma}$.	

(b)	1st M1 for a correct probability statement (as given or eg P(200<X<210)=0.3 o.e.)
	                      or shaded diagram - must have values on z-axis and probability areas shown
	1st A1 for correct use of 0.8 or $p = 0.2$. Need a correct probability statement.
	                      May be implied by a suitable value for z seen (e.g. $z = 0.84$)
	2nd M1 for attempting to standardise. Values for x and μ used in formula.
	                      Don't need $z =$ for this M1 nor a z-value, just mark standardization.
	B1 for $z = 0.8416$ (or better) [$z = 0.84$ usually just loses this mark in (a)]
	2nd A1 for AWRT 11.9	

(c)	1st M1 for attempting to Standardise with 200 and their sd(>0) e.g. $(\pm)\dfrac{180 - 200}{\text{their } \sigma}$
	
	2nd M1 **NB on epen this is an A mark ignore and treat it as 2nd M1**
	                      for 1 – a probability from tables provided compatible with their
	                      probability statement.
	A1 for 0.0465 or AWRT 0.046 (Dependent on both Ms in part (c))	

7.(a)	$P(R = 3 \cap B = 0) = \dfrac{1}{4} \times \dfrac{1}{4}, = \dfrac{1}{16}$		**M1, A1**
			(2)
(b)			

	3	0	3	6	9
	2	0	2	4	6
	1	0	1	2	3
	0	0	0	0	0
B R		0	1	2	3

All 0s   **B1**
All 1,2,3s   **B1**
All 4,6,9s   **B1**

(3)

(c)	$a = \dfrac{7}{16}, b = c = d = \dfrac{1}{16}$	**B1, B1 B1**
		(3)
(d)	$\mathrm{E}(T) = \left(1 \times \dfrac{1}{16}\right) + \left(2 \times \dfrac{1}{8}\right) + \left(3 \times \dfrac{1}{8}\right) + \left(4 \times \dfrac{1}{16}\right) + \dots$	**M1**
	$= 2\dfrac{1}{4}$ or exact equivalent e.g. 2.25, $\dfrac{9}{4}$	**A1**
		(2)
(e)	$\mathrm{Var}(T) = \left(1^2 \times \dfrac{1}{16}\right) + \left(2^2 \times \dfrac{1}{8}\right) + \left(3^2 \times \dfrac{1}{8}\right) + \left(4^2 \times \dfrac{1}{16}\right) + \dots - \left(\dfrac{9}{4}\right)^2$	**M1A1,M1**
	$= \dfrac{49}{4} - \dfrac{81}{16} = 7\dfrac{3}{16}$ or $\dfrac{115}{16}$ (o.e.) AWRT 7.19	**A1** **(4)**
		Total 14 marks

(a)	M1	for $\frac{1}{4} \times \frac{1}{4}$
(c)	1st B1	for $\frac{7}{16}$,
	2nd B1	for only one error in b, c, d ($b = c = d \neq \frac{1}{16}$ or $b = c = \frac{1}{16} \neq d$ etc), 3rd B1 all of b, c, d $= \frac{1}{16}$
(d)	M1	for attempting $\sum t\mathrm{P}(T = t)$, 3 or more terms correct or correct ft. Must Attempt to sum.
		NB calculating $\mathrm{E}(T)$ and then dividing by a number other than 1 scores M0.
(e)	1st M1	for attempt at $\mathrm{E}(T^2)$, 3 or more terms correct or correct ft.
	1st A1	for $\dfrac{49}{4}$ (o.e.) or a fully correct expression (all non-zero terms must be seen)
	2nd M1	for subtracting their $[\mathrm{E}(T)]^2$, Must be some attempt to square $-\frac{9}{4}$ is M0 but $-\frac{9}{16}$ could be M1
	2nd A1	for correct fraction or AWRT 7.19
		Full marks can still be scored in (d) and (e) if a is incorrect