Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced Level

Friday 20 May 2011 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Pink)

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 formulas 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.
This paper has 8 questions.
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 not gain full credit.

1. On a particular day the height above sea level, x metres, and the mid-day temperature, y °C, were recorded in 8 north European towns. These data are summarised below

$$S_{xx} = 3\,535\,237.5 \qquad \sum y = 181 \qquad \sum y^2 = 4305 \qquad S_{xy} = -23\,726.25$$

(*a*) Find S_{yy}.

(2)

(*b*) Calculate, to 3 significant figures, the product moment correlation coefficient for these data.

(2)

(*c*) Give an interpretation of your coefficient.

(1)

A student thought that the calculations would be simpler if the height above sea level, h, was measured in kilometres and used the variable $h = \frac{x}{1000}$ instead of x.

(*d*) Write down the value of S_{hh}.

(1)

(*e*) Write down the value of the correlation coefficient between h and y.

(1)

---

2. The random variable $X \sim N(\mu, 5^2)$ and $P(X < 23) = 0.9192$.

(*a*) Find the value of μ.

(4)

(*b*) Write down the value of $P(\mu < X < 23)$.

(1)

---

3. The discrete random variable Y has the probability distribution

y	1	2	3	4
$P(Y = y)$	a	b	0.3	c

where a, b and c are constants.

The cumulative distribution function $F(y)$ of Y is given in the following table.

y	1	2	3	4
$F(y)$	0.1	0.5	d	1.0

where d is a constant.

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

(5)

(b) Find $P(3Y + 2 \geq 8)$.

(2)

---

4. Past records show that the times, in seconds, taken to run 100 m by children at a school can be modelled by a normal distribution with a mean of 16.12 and a standard deviation of 1.60.

A child from the school is selected at random.

(a) Find the probability that this child runs 100 m in less than 15 s.

(3)

On sports day the school awards certificates to the fastest 30% of the children in the 100 m race.

(b) Estimate, to 2 decimal places, the slowest time taken to run 100 m for which a child will be awarded a certificate.

(4)

---

5. A class of students had a sudoku competition. The time taken for each student to complete the sudoku was recorded to the nearest minute and the results are summarised in the table below.

Time	Mid-point, x	Frequency, f
2 – 8	5	2
9 – 12		7
13 – 15	14	5
16 – 18	17	8
19 – 22	20.5	4
23 – 30	26.5	4

(You may use $\sum fx^2 = 8603.75$)

(*a*) Write down the mid-point for the 9 – 12 interval.

(1)

(*b*) Use linear interpolation to estimate the median time taken by the students.

(2)

(*c*) Estimate the mean and standard deviation of the times taken by the students.

(5)

The teacher suggested that a normal distribution could be used to model the times taken by the students to complete the sudoku.

(*d*) Give a reason to support the use of a normal distribution in this case.

(1)

On another occasion the teacher calculated the quartiles for the times taken by the students to complete a different sudoku and found

$$Q_1 = 8.5 \qquad Q_2 = 13.0 \qquad Q_3 = 21.0$$

(*e*) Describe, giving a reason, the skewness of the times on this occasion.

(2)

---

6. Jake and Kamil are sometimes late for school.
The events J and K are defined as follows

J = the event that Jake is late for school,
K = the event that Kamil is late for school.

$P(J) = 0.25$, $P(J \cap K) = 0.15$ and $P(J' \cap K') = 0.7$.

On a randomly selected day, find the probability that

(*a*) at least one of Jake or Kamil are late for school,

(1)

(*b*) Kamil is late for school.

(2)

Given that Jake is late for school,

(*c*) find the probability that Kamil is late.

(3)

The teacher suspects that Jake being late for school and Kamil being late for school are linked in some way.

(*d*) Determine whether or not J and K are statistically independent.

(2)

(*e*) Comment on the teacher's suspicion in the light of your calculation in part (*d*).

(1)

---

7. A teacher took a random sample of 8 children from a class. For each child the teacher recorded the length of their left foot, f cm, and their height, h cm. The results are given in the table below.

f	23	26	23	22	27	24	20	21
h	135	144	134	136	140	134	130	132

(You may use $\sum f = 186$ $\sum h = 1085$ $S_{ff} = 39.5$ $S_{hh} = 139.875$ $\sum fh = 25\,291$)

(a) Calculate S_{fh}.

(2)

(b) Find the equation of the regression line of h on f in the form $h = a + bf$.
Give the value of a and the value of b correct to 3 significant figures.

(5)

(c) Use your equation to estimate the height of a child with a left foot length of 25 cm.

(2)

(d) Comment on the reliability of your estimate in part (c), giving a reason for your answer.

(2)

The left foot length of the teacher is 25 cm.

(e) Give a reason why the equation in part (b) should not be used to estimate the teacher's height.

(1)

---

8. A spinner is designed so that the score S is given by the following probability distribution.

s	0	1	2	4	5
$P(S = s)$	p	0.25	0.25	0.20	0.20

(a) Find the value of p.

(2)

(b) Find $E(S)$.

(2)

(c) Show that $E(S^2) = 9.45$.

(2)

(d) Find $Var(S)$.

(2)

Tom and Jess play a game with this spinner. The spinner is spun repeatedly and S counters are awarded on the outcome of each spin. If S is even then Tom receives the counters and if S is odd then Jess receives them. The first player to collect 10 or more counters is the winner.

(e) Find the probability that Jess wins after 2 spins.

(2)

(f) Find the probability that Tom wins after exactly 3 spins.

(4)

(g) Find the probability that Jess wins after exactly 3 spins.

(3)

TOTAL FOR PAPER: 75 MARKS

END

edexcel

June 2011
Statistics S1 6683
Mark Scheme

Question Number	Scheme	Marks
1. **(a)**	$S_{yy} = 4305 - \frac{181^2}{8}$ = <u>209.875</u> **(awrt 210)**	M1 A1 (2)
(b)	$r = \frac{(-)23726.25}{\sqrt{3535237.5 \times "209.875"}}$ = <u>−0.87104…</u> **(awrt −0.871)**	M1 A1 (2)
(c)	Higher towns have lower temperature or temp. decreases as height increases	B1 (1)
(d)	$S_{hh} = 3.5352375$ (**awrt 3.54**) (condone 3.53)	B1 (1)
(e)	r = <u>−0.87104…</u> **(awrt −0.871)**	B1ft (1) **(7 marks)**
	Notes	
(a)	M1 for a correct expression. Allow one slip e.g. 4350 for 4305	
(b)	M1 for a correct expression for r, follow through their answer to (a). Condone no "_ " Allow M1 for <u>+</u> 0.87 with no working. (−0.871 is M1A1)	
(c)	B1 Must mention <u>temperature</u> (o.e.) and <u>height</u> (above sea level) and interpret the relationship between them. Must be a correct <u>and sensible</u> comment. e.g. "As temperature increases the height <u>of the sea</u> decreases" is B0. BUT simply stating "As temperature increases the height decreases" is B1 although "As height increases the temperature decreases" would be better. Treat mention of 0.87... as ISW "strong negative correlation between height and temp" is B0 (no interpretation) " as x increases y decreases" is B0 (no mention of height and temperature)	

Question Number	Scheme	Marks
(d)	B1 accept awrt 3.54 and condone 3.53 (i.e truncation)	
(e)	B1ft for awrt -0.871 or ft their final answer to part (b) to the same accuracy (or 3 sf) provided $-1 < r < 1$ Answer to part (e) must be a number "it's the same" is B0	
2. **(a)**	awrt ± 1.40 $\frac{23-\mu}{5} = "1.40"$ (o.e) $\mu = 16$ (or awrt 16.0)	B1 M1A1ft A1 (4)
(b)	0.4192	B1 (1) **5**
	Notes	
(a)	B1 for awrt $\pm$ 1.40 or better seen anywhere. Condone 1.4 instead of 1.40 M1 for attempting to standardise with 23 and 5 and μ, accept $\pm$ e.g. $\frac{23-\mu}{25} = 1.40$ can score B1M0 (since using 25 not 5 for standardising) $\frac{23-\mu}{5} = 0.9192$ can score B0M1 (since have correct standardisation) Can accept equivalent equations e.g. $23-\mu = 5\times "1.40"$ 1st A1ft for standardised expression = to a z value ($\|z\| > 1$). Signs must be compatible. Follow through their z e.g. $\frac{23-\mu}{5}$ = their z where $z > 1$ or $\frac{\mu-23}{5}$ = their z where $z < -1$ 2nd A1 for 16 or awrt 16.0 if they are using a more accurate z **Correct answer only scores 4/4 but if any working is seen apply scheme**	
(b)	B1 for 0.4192 (but accept 3sf accuracy if 0.9192 – 0.5 is seen)	

Question Number	Scheme	Marks
3. **(a)**	$a = \underline{0.1}$ [F(3) = F(2) + P(Y =3) =(0.5 + 0.3)] $d = \underline{0.8}$ b = F(2) - a = 0.5 - 0.1 or $a + b = 0.5$ $b = \underline{0.4}$ c = 1 - F(3) or 1 - ($a + b$ + 0.3) or $a + b + c = 0.7$ c = $\underline{0.2}$	B1 B1 M1 A1 A1 (5)
(b)	P($3Y + 2 \geq 8$) = P($Y \geq 2$) or 1 - P($Y \leq 1$) = b + 0.3 + c or 1 - a = $\underline{0.9}$	M1 A1ft (2) **7**
	Notes	
(a)	**Correct answers with no (or irrelevant) working score full marks** 1st B1 for a = 0.1 2nd B1 for F(3) = 0.8 or d = 0.8 M1 for a method for b or c. E.g. sight of $a + b = 0.5$ or $a + b + c = 0.7$ If their values satisfy one of these equations then score M1 provided their values are genuine probabilities (i.e. $0 < p < 1$) This M1 may be implied by a correct answer for b or c 1st A1 for b or P(2) = 0.4 2nd A1 for c or P(3) = 0.2	
(b)	M1 for rearranging to P($Y \geq 2$) or 1 - P($Y \leq 1$) or selecting cases Y = 2, 3 and 4 A1ft for 0.3 + their b + their c or 1 - their a, provided final answer < 1 and their values are probabilities.	

Question Number	Scheme	Marks
4. **(a)**	$(z = \pm)\dfrac{15-16.12}{1.6}(=-0.70)$	M1
	P(Z < -0.70) = 1 - 0.7580	M1
	= <u>0.2420</u> **(awrt 0.242)**	A1 (3)
(b)	[P($T < t$)=0.30 implies] $z = \dfrac{t-16.12}{1.6} = -0.5244$	M1 A1
	$\dfrac{t-16.12}{1.6} = -0.5244 \Rightarrow t = 16.12 - 1.6 \times "0.5244"$	M1
	t = awrt <u>15.28</u> (allow awrt 15.28/9)	A1 (4) **7**

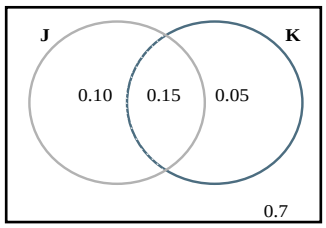
	Notes
	Allow slips e.g. 16.2 for 16.12 for 1st M1 in (a) and (b)
(a)	1st M1 for standardising expression with 15, 16.12 and 1.6 - allow <u>±</u> 2nd M1 for 1 - a probability (> 0.5) from tables or calculator based on their standardised value **Correct answer only scores 3/3**
	In part (b) they can use any letter or symbol instead of *t*
(b)	1st M1 for standardising with t (o.e.), 16.12 and 1.6, allow <u>±</u>, and setting equal to a z value 1st A1 for an equation with z = <u>±</u> 0.5244 or better e.g. $\dfrac{t-16.12}{1.6} = \pm 0.52$ (or 0.525) scores M1 (but A0) 2nd M1 for solving <u>their</u> linear equation as far as $t = a$ <u>±</u> $b \times 1.6$. Not dependent on 1st M1 e.g. solving $\dfrac{t-16.12}{1.6} = 0.3$ to give $t = 16.12 + 1.6 \times 0.3$ scores this M1 Allow $\dfrac{t-16.12}{1.6^2} = 0.3$ to give $t = 16.12 + 1.6^2 \times 0.3$ to score M1 too 2nd A1 dependent on both M marks. Allow awrt 15.28 or awrt 15.29 Condone awrt 15.3 if a correct expression for t = ... is seen. **Answers with no working:** 15.28 is M1A1M1A1, 15.29 is M1A0M1A1, 15.3 is M1A0M1A0

Question Number	Scheme	Marks
5.		
(a)	10.5	B1 (1)
(b)	$(Q_2 =)\ (15.5+)\ \frac{\frac{1}{2}\times 30-14}{8}\times 3$ or $\frac{\frac{1}{2}\times 31-14}{8}\times 3$	M1
	= 15.875 or 16.0625	A1 (2)
(c)	$\bar{x} = \frac{477.5}{30}$ = 15.9 (15.916) [Accept $\frac{191}{12}$ or $15\frac{11}{12}$]	M1, A1
	$\sigma = \sqrt{\frac{8603.75}{30} - \bar{x}^2}$,= 5.78 (accept s = 5.88)	M1A1ft, A1 (5)
(d)	Since mean and median are similar (or equal or very close) a normal distribution may be suitable. [Allow mean or median close to mode/modal class]	B1 (1)
(e)	$Q_3 - Q_2 (=8) > (4.5 =) Q_2 - Q_1$	M1
	Therefore positive skew	A1 (2)
		(11 marks)

	Notes
	In parts (a) to (c) a correct answer with no working scores full marks for that value.
(a)	B1 for 10.5 which may be in the table
(b)	M1 for a correct ratio and times 3, ignore the lower boundary for this mark A1 for awrt 15.9 (if n =30 used) or awrt 16.1 (if n+1 = 31 is used)
(c)	1st M1 for attempt at $\sum fx$ (this may be seen in the table as fx: 10, 73.5, 70, 136, 82, 106 [condone 1 slip] or awrt 500) and use of $\frac{\sum fx}{\sum f}$ or a correct expression for mean. 1st A1 for awrt 15.9 2nd M1 for an attempt at σ or σ^2, can ft their mean, condone mis-labelling $\sigma^2 = \sqrt{...}$ etc Allow use of their $\sum fx^2$ (awrt 9000) 2nd A1ft for a correct expression including square root, ft their mean but not their $\sum fx^2$. No label or correct label is OK but wrong label (e.g. $\sigma^2 = \sqrt{...}$) is A0 3rd A1 for awrt 5.78, allow s = awrt 5.88. **SC** Allow M1A1A0 for awrt 5.79 if $\bar{x}$ correct
(d)	B1 for a reason implying or stating symmetry. "Time is continuous" or "evenly distributed" is B0

Question Number	Scheme	Marks
(e)	M1 for a clear reason or comparison, values not essential but comparison implying they have been found is required. A1 for stating "positive skew". Condone just "positive" but "positive correlation" is A0 **Do not allow arguments based on mean and median since this part relates to a different set of data.**	
6.		
(a)	$P(J \cup K) = 1 - 0.7$ or $0.1 + 0.15 + 0.05 =$ $\underline{0.3}$	B1 (1)
(b)	$P(K) = 0.05 + 0.15$ or "0.3" – 0.25 + 0.15 or "0.3" = 0.25 + $P(K)$ – 0.15 May be seen on Venn diagram $= \underline{0.2}$	M1 A1 (2)
(c)	$\left[P(K \mid J)\right] = \dfrac{P(K \cap J)}{P(J)}$ $= \dfrac{0.15}{0.25}$ $= \underline{\dfrac{3}{5} \text{ or } 0.6}$	M1 A1 A1 (3)
(d)	$P(J) \times P(K) = 0.25 \times 0.2 (= 0.05)$, $P(J \cap K) = 0.15$ $\underline{\text{or}}$ $P(K \mid J) = 0.6$, $P(K) = 0.2$ $\underline{\text{or}}$ may see $P(J\|K) = 0.75$ and $P(J) = 0.25$ not equal therefore not independent	M1 A1ft (2)
(e)	Not independent so confirms the teacher's suspicion $\underline{\text{or}}$ they are linked (This requires a statement about independence in (d) or in (e))	B1ft (1) **(9 marks)**

Question Number	Scheme	Marks
	Notes	
(b)	M1 for a complete method, follow through their 0.3, leading to a linear equation for P(*K*) NB You may see this Venn diagram. A correct diagram (Venn or table) implies M1 in (b) Need not include box or 0.7 **Correct answer only is 2/2**	
(c)	**In parts (c) and (d) they must have defined *A* and *B*** M1 for a correct expression (including ratio) in symbols. 1st A1 for a correct ratio of probabilities (if this is seen the M1 is awarded by implication) Must be in (c). Condone no LHS but wrong LHS (e.g. P(*K*) or P(*J*|*K*)) is M0A0 2nd A1 for correct answer as printed only. **Correct answer only 3/3**	
(d)	**Mark (d) and (e) together** M1 for a correct comparison of known probabilities for an independence test - ft their values. E.g. $P(J)\times P(K)$ with $P(J\cap K)$ <u>or</u> P(*K*|*J*) with P(*K*) [Must have expressions] The values of these probabilities should be given unless they are in the question or stated elsewhere. A1ft for correct calculations and correct comment for their probabilities	
(e)	B1ft ft their conclusion on independence so not independent confirms teacher…independent contradicts teacher. **Methods leading to negative probabilities should score M0**	



Question Number	Scheme	Marks
7.		
(a)	$\left(S_{fh} =\right) 25291 - \frac{186 \times 1085}{8}$	M1
	$= \underline{64.75}$ (accept 64.8)	A1
		(2)
(b)	$b = \frac{"64.75"}{39.5}$, $= \underline{1.6392\ldots}$ (awrt 1.6)	M1, A1
	$a = \frac{1085}{8} - b \times \frac{186}{8}$, $= \underline{97.512\ldots}$ (awrt 97.5)	M1, A1
	$\underline{h = 97.5 + 1.64f}$	A1ft (dep on M1M1)
		(5)
(c)	$h = 97.5 + 1.64 \times 25$, $= \underline{138 \sim 139}$ (final answer in [138, 139])	M1, A1
		(2)
(d)	Should be reliable, since 25 cm(or *f* or footlength) is within the range of the data	B1, B1
		(2)
(e)	Line is for children – a different equation would apply to adults or Children are still growing, height will increase more than foot length	B1
		(1)
		12

Notes

[NB $r = 0.871$ so do not confuse this with question 1]

(a) M1 for attempting a correct expression [allow a copying slip e.g. 25921]

(b) 1st M1 for a correct expression for *b*, ft their part (a) but not $S_{fh} = 25291$

1st A1 for awrt 1.6

2nd M1 for use of $a = \bar{h} - b \times \bar{f}$, ft their value for *b*. Must use $\bar{h}$ and $\bar{f}$ not values from table.

2nd A1 for awrt 97.5 [NB $a = 135 - 1.63 \times 23 = 97.51$ but M0A0 since not using $\bar{h}$ and $\bar{f}$]

3rd A1ft for an equation for *h* and *f* with their coefficients to 3sf. **Dependent on both Ms**

Must be 3sf not awrt. Give this mark if seen in (c). Equation must be in *h* and *f* not *y* and *x*.

(c) M1 for using their equation and $f = 25$ to find *h*

A1 for their final answer in [138, 139]. Can give if they have 137.7... but round to 138

(d) 1st B1 for suggesting it is reliable

2nd B1 for mentioning that 25 cm is within range of data. "interpolation"or"not extrapol'B1

Use of "it" or a comment that height is in range is B0 but apply ISW

(e) B1 for some comment that states a difference between children and teachers(adults)

Must mention teacher/adults and children

e.g. ".teacher is not in same age group as the children", "equation is for children not adults"

"children and adults are different populations"

"teacher will be taller" is B0 since no mention of children.

"equation is only valid for children" is OK since "only" implies not suitable for adults

Or Reference to different growth rates

Question Number	Scheme	Marks
8.		
(a)	$1 = p + (0.25 + 0.25 + 0.2 + 0.2), \Rightarrow p = \underline{\frac{1}{10} \text{ or } 0.1}$	M1, A1 (2)
(b)	$\text{E}(S) = \frac{1}{4} + 2\times\frac{1}{4} + 4\times\frac{1}{5} + 5\times\frac{1}{5}$, (or equiv. in decimals) $= \underline{2.55}$	M1, A1 (2)
(c)	$\text{E}\left(S^2\right) = \frac{1}{4} + \frac{2^2}{4} + \frac{4^2}{5} + \frac{5^2}{5}$ or 0.25 + 1 + 3.2 + 5 $= \underline{9.45}$ (*)	M1, A1cso (2)
(d)	$\text{Var}(S) = 9.45 - (\text{E}(S))^2$, $= \underline{2.9475 \text{ or } \frac{1179}{400}}$ (accept awrt 2.95)	M1, A1 (2)
(e)	$\text{P}(5 \text{ and } 5) = \left(\frac{1}{5}\right)^2$, $= \underline{\frac{1}{25} \text{ or } 0.04}$	M1, A1 (2)
(f)	$\text{P}(4, 4, 2) = \left(\frac{1}{5}\right)^2 \times \frac{1}{4} \times 3$ ($= 0.03$ or $\frac{3}{100}$) $\text{P}(4, 4, 4) = \left(\frac{1}{5}\right)^3$ ($= 0.008$ or $\frac{1}{125}$) P(Tom wins in 3 spins) = $\underline{0.038}$	M1, M1 B1 A1 (4)
(g)	$\text{P}\left(\overline{5} \cap 5 \cap 5\right) + \text{P}(5 \cap \overline{5} \cap 5) = \frac{4}{5} \times \left(\frac{1}{5}\right)^2 \times 2 = \underline{0.064 \text{ or } \frac{8}{125}}$	M1, M1, A1 (3)
		17

	Notes
(a)	M1 for clear attempt to use sum of probabilities = 1 (fractions or decimals) Ans only 2/2
(b)	M1 for at least 2 correct terms ($\neq 0$) of the expression. 2.55 with no working scores M1A1 **Any division by *k* (usually 5) in (b) or (c) or (d) scores M0**
(c)	M1 for at least 3 correct, non-zero terms of the expression seen, allow decimals. A1cso for the full expression (with 9.45) seen. Must be cso but can ignore wrong *p*.
(d)	M1 for a correct expression (9.45 seen), can ft their E(*S*). May see $\sum (x - \text{"2.55"})^2 \times \text{P}(X = x)$ A1 accept awrt 2.95 Answer only can score M1 for correct ft and A1 for awrt 2.95 **Answer only in (e) and (f) is full marks, in (g) is no marks**
(e)	M1 for $\left(\frac{1}{5}\right)^2$ Condone $\text{P}(5) \times \text{P}(5) = 0.25 \times 0.25$. [Beware 0.4 is A0]
(f)	1st M1 for $\left(\frac{1}{5}\right)^2 \times \frac{1}{4}$ or 0.01 seen 2nd M1 for multiplying a $p^2 q$ probability by 3($p, q \in (0,1)$). B1 for $(0.2)^3$ or better seen
(g)	1st M1 for $\frac{4}{5} \times \left(\frac{1}{5}\right)^2$ or all cases considered and correct attempt at probabilities. 2nd M1 for multiplying a $p^2(1-p)$ probability by 2. **Beware $(0.4)^3$ = 0.064 is M0M0A0**