Centre No.						Paper Reference	Surname	Initial(s)
Candidate No.						6 6 8 3 / 0 1	Signature	

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 20 May 2009 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
8	
Total	

Materials required for examination	Items included with question papers
Mathematical Formulae (Orange or 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 above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer for each question in the space following the question.
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 8 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
H34279A
W850/R6683/57570 3/5/4/3

Turn over

edexcel
advancing learning, changing lives

---

1. The volume of a sample of gas is kept constant. The gas is heated and the pressure, p, is measured at 10 different temperatures, t. The results are summarised below.

$\sum p = 445 \quad \sum p^2 = 38\,125 \quad \sum t = 240 \quad \sum t^2 = 27\,520 \quad \sum pt = 26\,830$

(a) Find S_{pp} and S_{pt}. **(3)**

Given that $S_{tt} = 21\,760$,

(b) calculate the product moment correlation coefficient. **(2)**

(c) Give an interpretation of your answer to part (b). **(1)**

a) $Spp = \Sigma p^2 - (\Sigma p)^2 \div n = 18322.5$

$Spt = \Sigma pt - (\Sigma p)(\Sigma t) \div n = 16150$

b) $PMCC, r = \dfrac{Spt}{\sqrt{Spp \times Stt}} = \dfrac{16150}{\sqrt{18322.5 \times 21760}}$

$r = 0.809$

c) Some evidence to suggest positive correlation, so the higher the pressure, the higher the temperature

As pressure increases, temperature increase

---

2. On a randomly chosen day the probability that Bill travels to school by car, by bicycle or on foot is $\frac{1}{2}$, $\frac{1}{6}$ and $\frac{1}{3}$ respectively. The probability of being late when using these methods of travel is $\frac{1}{5}$, $\frac{2}{5}$ and $\frac{1}{10}$ respectively.

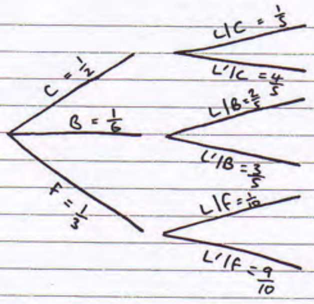
(a) Draw a tree diagram to represent this information. **(3)**

(b) Find the probability that on a randomly chosen day

(i) Bill travels by foot and is late,

(ii) Bill is not late. **(4)**

(c) Given that Bill is late, find the probability that he did not travel on foot. **(4)**



$C \cap L = \frac{1}{10} = \frac{3}{30}$

$C \cap L' = \frac{4}{10} = \frac{12}{30}$

$B \cap L = \frac{2}{30}$

$B \cap L' = \frac{3}{30}$

$F \cap L = \frac{1}{30}$

$F \cap L' = \frac{9}{30}$

b) $P(f \cap L) = \frac{1}{30}$ ii) $P(L') = \frac{12+3+9}{30} = \frac{24}{30} = \frac{4}{5}$

c) $P(F'|L) = \dfrac{P(F' \cap L)}{P(L)} = \dfrac{P(C \cap L) + P(B \cap L)}{1 - \frac{4}{5}}$

$\dfrac{\frac{3}{30} + \frac{2}{30}}{\frac{1}{5}} = \frac{5}{30} \div \frac{1}{5} = \frac{5}{30} \times \frac{5}{1} = \frac{25}{30}$

$= \frac{5}{6}$

---

3. The variable x was measured to the nearest whole number. Forty observations are given in the table below.

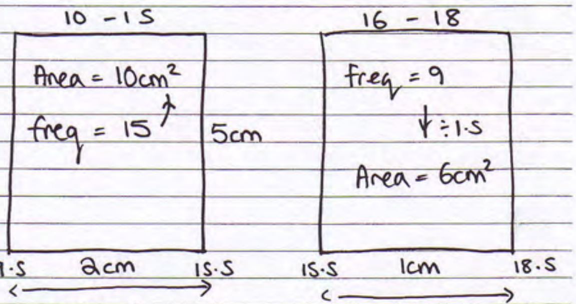
x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 – 15 class has a width of 2 cm and a height of 5 cm. For the 16 – 18 class find

(a) the width, **(1)**

(b) the height **(2)**

of the bar representing this class.



cw ÷ 3 = width $\quad$ 3 ÷ 3 = 1 width = 1cm

freq ÷ 1.5 = Area $\quad$ 9 ÷ 1.5 = 6 Area = 6cm²

$\text{height} = \dfrac{\text{Area}}{\text{width}} = \dfrac{6}{1} = 6\text{cm}$

a) width = 1cm $\quad$ b) height = 6cm.

Leave blank

4. A researcher measured the foot lengths of a random sample of 120 ten-year-old children. The lengths are summarised in the table below.

no gaps →

MP, x	Foot length, l, (cm)	f Number of children
11	$10 \leqslant l < 12$	5
14.5	$12 \leqslant l < 17$	53
18	$17 \leqslant l < 19$	29
20	$19 \leqslant l < 21$	15
22	$21 \leqslant l < 23$	11
24	$23 \leqslant l < 25$	7

CF
0
5
58 ← 60th
87
102
113
120

(a) Use interpolation to estimate the median of this distribution.
(2)

(b) Calculate estimates for the mean and the standard deviation of these data.
(6)

One measure of skewness is given by

$$\text{Coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

(c) Evaluate this coefficient and comment on the skewness of these data.
(3)

Greg suggests that a normal distribution is a suitable model for the foot lengths of ten-year-old children.

(d) Using the value found in part (c), comment on Greg's suggestion, giving a reason for your answer.
(2)

a) $\frac{1}{2}n = 60$ (17-19)

$\xleftrightarrow{2}$
17 Q_2 19
58 60 87
$\xleftrightarrow{2}$
$\xleftrightarrow{29}$

$\frac{Q_2 - 17}{2} = \frac{2}{29}$

$Q_2 = \frac{2}{29} \times 2 + 17$

$Q_2 = \underline{17.14}$

b) $\Sigma fx = 2055.5$
$\Sigma fx^2 = 36500.25$

$\bar{x} = \frac{\Sigma fx}{n} = \underline{17.13}$ $\sigma = \sqrt{\frac{36500.25}{120} - 17.13^2} = \underline{3.28}$

c) $\text{Skew} = \frac{3(17.13 - 17.14)}{3.28} = -0.009$

Symmetrical skew

d) yes normal distribution seems appropri
since symmetrical skew mean ≃ median
95% of data should lie within $\bar{x} \pm 2\sigma$
17.13 ± 6.56 (10.6 - 23.7) which seems t
be the case.

Leave blank

5. The weight, w grams, and the length, l mm, of 10 randomly selected newborn turtles are given in the table below.

l	49.0	52.0	53.0	54.5	54.1	53.4	50.0	51.6	49.5	51.2
w	29	32	34	39	38	35	30	31	29	30

(You may use $S_{ll} = 33.381$ $S_{wl} = 59.99$ $S_{ww} = 120.1$)

(a) Find the equation of the regression line of w on l in the form $w = a + bl$.
(5)

(b) Use your regression line to estimate the weight of a newborn turtle of length 60 mm.
(2)

(c) Comment on the reliability of your estimate giving a reason for your answer.
(2)

$\left\{ \begin{matrix} w = a + bl \\ y = a + bx \end{matrix} \right\}$ $w = y$ $l = x$

$b = \frac{S_{xy}}{S_{xx}} = \frac{S_{lw}}{S_{ll}} = \frac{59.99}{33.381} = 1.797$

$a = \bar{y} - b\bar{x} \Rightarrow a = \bar{w} - b\bar{l}$

$\bar{w} = \frac{\Sigma w}{n} = \frac{518.3}{10} = 51.83$

$\bar{l} = \frac{\Sigma l}{n} = \frac{32.7}{10} = 3.27$

$a = 51.83 - 1.797 \times 3.27 = -60.445$

$w = -60.445 + 1.797l$

b) $l = 60$ $w = 47.375g$

c) Unreliable, no evidence to support this since max length in our data was 54.5cm

---

6. The discrete random variable X has probability function

$$P(X = x) = \begin{cases} a(3 - x) & x = 0,1,2 \\ b & x = 3 \end{cases}$$

(a) Find $P(X = 2)$ and complete the table below.

$\Sigma P = 1 \Rightarrow 6a + b =$
$(\times 3)$

x	0	1	2	3
$P(X = x)$	$3a$	$2a$	a	b

$E(X) = 0 + 2a + 2a + 3b$
(1)

Given that $E(X) = 1.6$ $\Rightarrow 4a + 3b = 1.6$

(b) Find the value of a and the value of b.
(5)

Find

(c) $P(0.5 < X < 3)$,
(2)

(d) $E(3X - 2)$.
(2)

(e) Show that the $\text{Var}(X) = 1.64$
(3)

(f) Calculate $\text{Var}(3X - 2)$.
(2)

x^2	0	1	4	9
P	0.3	0.2	0.1	0.4

$E(x^2) = 0 + 0.2 + 0.4 + 3.6$

$E(x^2) = 4.2$

b) $18a + 3b = 3$ −
$\underline{4a + 3b = 1.6}$

$14a = 1.4 \Rightarrow a = \underline{0.1} \Rightarrow b = \underline{0.4}$

c) $P(0.5 < X < 3) = P(1) + P(2) = 3a = \underline{0.3}$

d) $E(3X - 2) = 3E(x) - 2 = 3 \times 1.6 - 2 = \underline{2.8}$

e) $V(x) = E(x^2) - E(x)^2 = 4.2 - 1.6^2 = \underline{1.64}$

f) $V(3X - 2) = 3^2 \times V(x) = \underline{14.76}$

7. (a) Given that $P(A) = a$ and $P(B) = b$ express $P(A \cup B)$ in terms of a and b when

(i) A and B are mutually exclusive,
(ii) A and B are independent.

(2)

Two events R and Q are such that

$P(R \cap Q') = 0.15$, $P(Q) = 0.35$ and $P(R|Q) = 0.1$

Find the value of

(b) $P(R \cup Q)$,

(1)

(c) $P(R \cap Q)$,

(2)

(d) $P(R)$.

(2)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

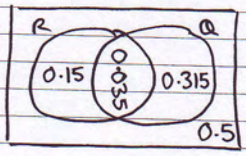
i) Mutually exclusive $\Rightarrow P(A \cap B) = 0$

$P(A \cup B) = a + b$

ii) Independent $\Rightarrow P(A \cap B) = P(A)P(B) = ab$

$P(A \cup B) = a + b - ab.$

ii) $P(R|Q) = \dfrac{P(R \cap Q)}{P(Q)} \Rightarrow 0.1 = \dfrac{P(R \cap Q)}{0.35} \Rightarrow P(R \cap Q) = 0.035$



b) $P(R \cup Q) = 0.15 + 0.035 + 0.315$
$= 0.5$

c) $P(R \cap Q) = 0.035$

d) $P(R) = 0.15 + 0.035$
$\underline{0.185}$

---

8. The lifetimes of bulbs used in a lamp are normally distributed.
A company X sells bulbs with a mean lifetime of 850 hours and a standard deviation of 50 hours.

(a) Find the probability of a bulb, from company X, having a lifetime of less than 830 hours.

(3)

(b) In a box of 500 bulbs, from company X, find the expected number having a lifetime of less than 830 hours.

(2)

A rival company Y sells bulbs with a mean lifetime of 860 hours and 20% of these bulbs have a lifetime of less than 818 hours.

(c) Find the standard deviation of the lifetimes of bulbs from company Y.

(4)

Both companies sell the bulbs for the same price.

(d) State which company you would recommend. Give reasons for your answer.

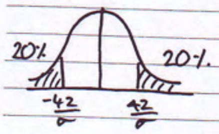
(2)

a) $P(X < 830) \Rightarrow P\left(Z < \dfrac{830-850}{50}\right) = P(Z < -0.4)$

$= 1 - \Phi(0.4) = \underline{0.3446}$

b) $0.3446 \times 500 = 172.3$ $\underline{172 \text{ bulbs}}$

c) $P(X < 818) = 0.2 \Rightarrow P\left(Z < -\dfrac{42}{\sigma}\right) = 0.2$



$P\left(Z > \dfrac{42}{\sigma}\right) = 0.2$ $\underline{\text{\% points}}$

$\dfrac{42}{\sigma} = 0.8416 \Rightarrow \underline{\sigma = 49.9}$

d) Y since it has a higher mean and the standard deviation is almost identical