

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Bronze Level B3

Time: 1 hour 30 minutes

Materials required for examination
papers

Mathematical Formulae (Green)

Items included with question

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

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.

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.

Suggested grade boundaries for this paper:

A*	A	B	C	D	E
72	66	59	52	46	41

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.

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

(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)

May 2009

2. Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$$\Sigma x = 84.6$$

(a) Calculate the mean amount of rainfall during the whole 28 days.

(2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0.

Keith corrects these figures.

(b) State, giving your reason, the effect this will have on the mean.

(2)

January 2011

3. 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

(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)

May 2008

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)

May 2008

5. There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,
70 take developing software,
81 take networking,
35 take developing software and systems support,
28 take networking and developing software,
40 take systems support and networking,
4 take all three extra options.

- (a) Draw a Venn diagram to represent this information. (5)

A student from the course is chosen at random.

Find the probability that the student takes

- (b) none of the three extra options, (1)
- (c) networking only. (1)

Students who want to become technicians take systems support and networking. Given that a randomly chosen student wants to become a technician,

- (d) find the probability that this student takes all three extra options. (2)

January 2010

6. 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)

January 2008

7. A manufacturer carried out a survey of the defects in their soft toys. It is found that the probability of a toy having poor stitching is 0.03 and that a toy with poor stitching has a probability of 0.7 of splitting open. A toy without poor stitching has a probability of 0.02 of splitting open.

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

(3)

(b) Find the probability that a randomly chosen soft toy has exactly one of the two defects, poor stitching or splitting open.

(3)

The manufacturer also finds that soft toys can become faded with probability 0.05 and that this defect is independent of poor stitching or splitting open. A soft toy is chosen at random.

(c) Find the probability that the soft toy has none of these 3 defects.

(2)

(d) Find the probability that the soft toy has exactly one of these 3 defects.

(4)

May 2012

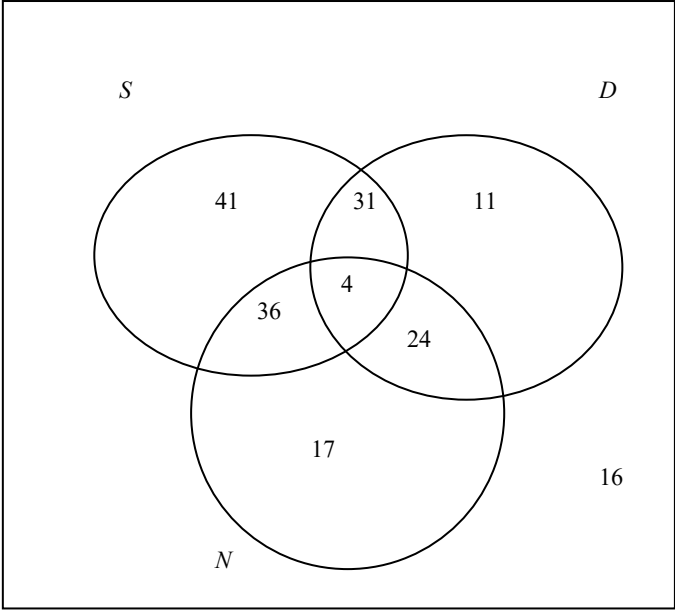
TOTAL FOR PAPER: 75 MARKS

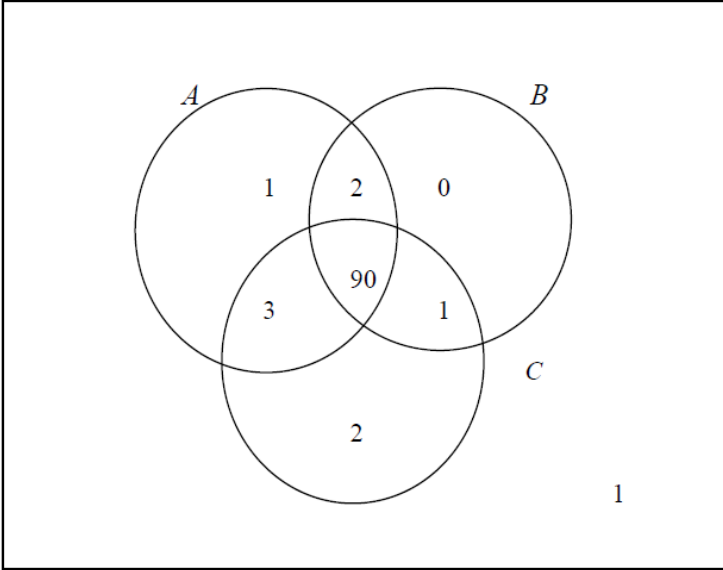
END

Question Number	Scheme	Marks
1. (a)	$(S_{pp}) = 38125 - \frac{445^2}{10}$ $= 18322.5$ awrt 1830 $(S_{pt}) = 26830 - \frac{445 \times 240}{10}$ $= 16150$ awrt 16200	M1 A1 A1 (3)
(b)	$r = \frac{"16150"}{\sqrt{"18322.5" \times 21760}}$ $= 0.8088...$ awrt 0.809	M1 A1 (2)
(c)	As the temperature increases the pressure increases.	B1 (1) [6]
2. (a)	$2.8 + 5.6 + 2.3 + 9.4 + 0.5 + 1.8 + 84.6 = 107$ mean = $107 / 28 (= 3.821...)$ (awrt 3.8)	M1 A1 (2)
(b)	It will have no effect since one is 4.5 under what it should be and the other is 4.5 above what it should be.	B1 dB1 (2) [4]

Question Number	Scheme	Marks
3. (a)	50	B1 (1)
(b)	$Q_1 = 45$ $Q_2 = 50.5$ $Q_3 = 63$	B1 B1 B1 (3)
(c)	Mean = $\frac{1469}{28} = 52.464286..$ awrt 52.5 Sd = $\sqrt{\frac{81213}{28} - \left(\frac{1469}{28}\right)^2}$ = 12.164.... or 12.387216... for divisor $n-1$ awrt 12.2 or 12.4	M1 A1 M1 A1 (4)
(d)	$\frac{52.46.. - 50}{sd} = \text{awrt } 0.20 \text{ or } 0.21$	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]

Question Number	Scheme	Marks
4. (a)	$S_{tt} = 10922.81 - \frac{401.3^2}{15} = 186.6973$ $S_{vv} = 42.3356 - \frac{25.08^2}{15} = 0.40184$ $S_{tv} = 677.971 - \frac{401.3 \times 25.08}{15} = 6.9974$	awrt 187 awrt 0.402 awrt 7.00 M1A1 A1 A1 (4)
(b)	$r = \frac{6.9974}{\sqrt{186.6973 \times 0.40184}}$ $= 0.807869$	M1A1ft 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)	$b = \frac{6.9974}{186.6973} = 0.03748$ $a = \frac{25.08}{15} - b \times \frac{401.3}{15} = 0.6692874$	awrt 0.0375 awrt 0.669 M1 A1 M1 A1 (4)
(f)	$t = 19, v = 0.6692874 + 0.03748 \times 19 = 1.381406$	awrt 1.4 B1 (1)
		[15]

Question Number	Scheme	Marks
5. (a)	 <p>3 closed curves and 4 in centre</p> <p>Evidence of subtraction</p> <p>31,36,24</p> <p>41,17,11</p> <p>Labels on loops, 16 and box</p>	<p>M1</p> <p>M1</p> <p>A1</p> <p>A1</p> <p>B1</p> <p>(5)</p>
(b)	$P(\text{None of the 3 options}) = \frac{16}{180} = \frac{4}{45}$	<p>B1ft</p> <p>(1)</p>
(c)	$P(\text{Networking only}) = \frac{17}{180}$	<p>B1ft</p> <p>(1)</p>
(d)	$P(\text{All 3 options/technician}) = \frac{4}{40} = \frac{1}{10}$	<p>M1 A1</p> <p>(2)</p> <p>[9]</p>

Question Number	Scheme	Marks
6. (a)	<p>Diagram may be drawn with $B \subset (A \cup C)$ or with the 0 for $B \cap (A \cup C)'$ simply left blank</p>  <p>Accept decimals or probs. in Venn diagram</p>	<p>3cc 90,3,2,1 1,(0),2 1 outside Box</p> <p>M1 A1 M1A1 A1 B1</p>
(b)	$P(\text{none})=0.01$	(6) B1ft
(c)	$P(A \text{ but not } B)=0.04$	(1) M1 A1ft
(d)	$P(\text{any wine but } C)=0.03$	(2) M1A1ft
(e)	$P(\text{exactly two})=0.06$	(2) M1A1ft
(f)	$P(C A) = \frac{P(C \cap A)}{P(A)} = \frac{93}{96}$ or $\frac{31}{32}$ or awrt 0.969	(3) M1A1ft, A1 [16]

Question Number	Scheme	Marks
7. (a)		Shape B1 Labels & 0.03 B1 Labels & 0.7,0.02 B1
(b)	$P(\text{Exactly one defect}) = 0.03 \times 0.3 + 0.97 \times 0.02$ <u>or</u> $P(PS \cup Split) - 2P(PS \cap Split)$ $= [0.009 + 0.0194 =]$	(3) M1A1ft A1 cao (3)
(c)	$P(\text{No defects}) = (1 - 0.03) \times (1 - 0.02) \times (1 - 0.05) \quad (\text{or better})$ $= 0.90307$	M1 A1 cao (2)
(d)	$P(\text{Exactly one defect}) = (b) \times (1 - 0.05) + (1 - 0.03) \times (1 - 0.02) \times 0.05$ $= "0.0284" \times 0.95 + 0.97 \times 0.98 \times 0.05$ $= [0.02698 + 0.04753] = 0.07451$	M1 M1 A1ft A1 cao (4) [12]

Examiner reports

Question 1

The vast majority scored full marks in part (a). The most common reason for losing marks for the correlation coefficient was for rounding to less than 3 significant figures without having stated the more accurate answer first. A large proportion of candidates still believe that stating 'it's a high level of correlation' will be enough to gain the mark for interpretation. A fully contextual comment is required here, using the named variables of pressure and temperature and not just the letters p and t .

Question 2

Again the majority of the candidates encountered few problems here and many scored 3 or 4 marks, although a number in part (a) found both means and then simply added them and divided by 2 rather than taking the weighted average. Most chose to find the sum of all 28 records and then simply divided this by 28. A few candidates misinterpreted the 84.6 as the mean for the 21 days despite what should have been an obvious discrepancy with Keith's data.

In part (b) many realised that the changes would have no effect on the mean but sometimes they failed to give any numerical values to support this claim. There were many excellent answers though showing clearly that $9.4 + 0.5 = 4.9 + 5.0$ or some other suitable calculation.

Question 3

Parts (a) to (d) represented a chance for all students who had an average grasp of statistics to score highly. The median in part (b) was incorrectly identified by a significant number of candidates, but the standard deviation was often correct. Part (e) was done surprisingly well with students appearing to have a much greater understanding of what is required for a comparison than in previous years. Often numbers were stated without an actual comparison. Confusion was evident in some responses as skewness was often referred to as correlation. A small minority of candidates had failed to take note of the 'For the Balmoral Hotel' and had done some correct statistics for all 55 students.

Question 4

This was done well by all but the weakest students with most using sufficient accuracy to score highly. Many candidates demonstrated an understanding of the use of the formulae to achieve full marks in part (a) and part (b). By far the main reason for loss of marks was premature approximation. Part (c) and part (d) were done well by good candidates. Only the more able candidates had a correct reason why t was the explanatory variable. Many called v the explanatory variable but gave a correct reason for t . The written parts were not universally done correctly, although the ability of students to deal with this topic has improved considerably in recent examinations. Rounding once again caused issues in part (e), but usually did not have an effect on part (f).

Question 5

There were many good answers to this question. The Venn diagram was often totally correct although a number failed to subtract for the intersections and obtained value of 35, 40 and 28 instead of 31, 36 and 24 for the numbers taking two options. Parts (b) and (c) were answered very well with only a minority of candidates failing to give probabilities. Part (d) proved straightforward for those who knew what was required but some attempted complicated calculations, often involving a product of probabilities, whilst others simply gave their answer as $4/180$.

Question 6

This question was often answered very well. The Venn diagram was usually correct although a few forgot the box and some missed the “1” outside the circles. A small minority of candidates failed to subtract the “90” from the overlaps of each pair and this meant that any attempts to follow through in later parts of the question were hopeless as their probabilities were greater than 1. Part (b) was answered well although some wrote 0.1 instead of 0.01.

Parts (c), (d) and (e) were answered well too but some candidates simply gave integer answers rather than probabilities and a few tried to multiply probabilities together. The conditional probability in part (f) was often identified but some thought that $P(C | A) = \frac{P(C \cap A)}{P(C)}$ and $P(C \cap A) = 0.03$ was another common error.

Question 7

This question was answered well by most candidates. There were a number of accuracy errors with the wrong positioning of the decimal point with 0.03×0.3 giving rise to 0.9, 0.09 or even 0.0009 and 0.047...was sometimes copied from a calculator as 0.47...

Part (a) proved a straightforward start to the question and most candidates completed a correct tree diagram.

Most completed part (b) correctly too but some found $P(\text{a split})$ and others found $P(\text{at least one of the defects})$ but many correct answers were seen here.

In part (c) the majority were able to multiply together the 3 required probabilities but a small minority just gave the answer as 0.97×0.98 .

Numerous candidates found the correct products of three probabilities in part (d) but some just added 0.05 to their answer to part (b). Most candidates though found this question a very rewarding end to the paper.

Statistics for S1 Practice Paper Bronze Level B3

Qu	Max Score	Modal score	Mean %	Mean score for students achieving grade:							
				ALL	A*	A	B	C	D	E	U
1	6		82	4.91		5.46	5.18	4.99	4.80	4.62	3.68
2	4		79	3.17	3.59	3.53	3.17	2.93	2.80	2.65	2.09
3	13		75	9.78		11.69	10.60	9.77	8.91	7.91	6.02
4	15		74	11.14		13.18	11.96	11.25	10.44	9.49	6.87
5	9		74	6.66		7.78	6.83	6.23	5.39	4.78	3.41
6	16		71	11.37		14.39	11.97	9.84	8.39	6.87	4.54
7	12		68	8.11	10.89	10.52	9.14	8.14	7.26	6.50	4.69
	75		74	55.14		66.55	58.85	53.15	47.99	42.82	31.30