Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1
Silver Level S3

Time: 1 hour 30 minutes

Materials required for examination papers	**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 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 8 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
65	58	51	44	36	29

1. A young family were looking for a new 3 bedroom semi-detached house. A local survey recorded the price x, in £1000, and the distance y, in miles, from the station of such houses. The following summary statistics were provided

$$S_{xx} = 113\,573,\ S_{yy} = 8.657,\ S_{xy} = -808.917$$

(a) Use these values to calculate the product moment correlation coefficient.

(2)

(b) Give an interpretation of your answer to part (a).

(1)

Another family asked for the distances to be measured in km rather than miles.

(c) State the value of the product moment correlation coefficient in this case.

(1)

June 2007

---

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

(a) Find the value of μ.

(4)

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

(1)

May 2011

---

3. The discrete random variable X has probability distribution given by

x	-1	0	1	2	3
$P(X = x)$	$\frac{1}{5}$	a	$\frac{1}{10}$	a	$\frac{1}{5}$

where a is a constant.

(*a*) Find the value of a.
(2)

(*b*) Write down $E(X)$.
(1)

(*c*) Find $\mathrm{Var}(X)$.
(3)

The random variable $Y = 6 - 2X$.

(*d*) Find $\mathrm{Var}(Y)$.
(2)

(*e*) Calculate $P(X \geq Y)$.
(3)

May 2010

4. In a company the 200 employees are classified as full-time workers, part-time workers or contractors.

The table below shows the number of employees in each category and whether they walk to work or use some form of transport.

	Walk	Transport
Full-time worker	2	8
Part-time worker	35	75
Contractor	30	50

The events F, H and C are that an employee is a full-time worker, part-time worker or contractor respectively. Let W be the event that an employee walks to work.

An employee is selected at random.

Find

(*a*) $P(H)$
(2)

(*b*) $P([F \cap W]')$
(2)

(*c*) $P(W \mid C)$
(2)

Let B be the event that an employee uses the bus.

Given that 10% of full-time workers use the bus, 30% of part-time workers use the bus and 20% of contractors use the bus,

(*d*) draw a Venn diagram to represent the events F, H, C and B,
(4)

(*e*) find the probability that a randomly selected employee uses the bus to travel to work.
(2)

May 2013

5. The Venn diagram in Figure 1 shows the number of students in a class who read any of 3 popular magazines *A*, *B* and *C*.

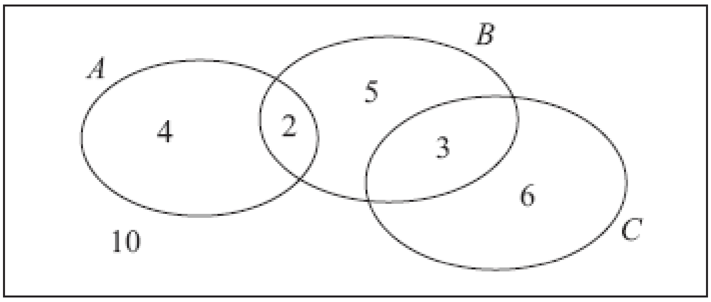


Figure 1

One of these students is selected at random.

(*a*) Show that the probability that the student reads more than one magazine is $\frac{1}{6}$.

(2)

(*b*) Find the probability that the student reads *A* or *B* (or both).

(2)

(*c*) Write down the probability that the student reads both *A* and *C*.

(1)

Given that the student reads at least one of the magazines,

(*d*) find the probability that the student reads *C*.

(2)

(*e*) Determine whether or not reading magazine *B* and reading magazine *C* are statistically independent.

(3)

May 2010

6.

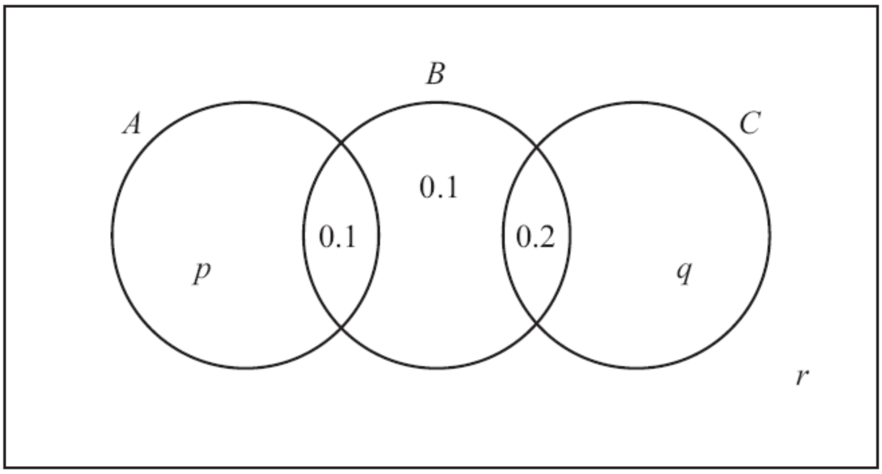


Figure 1

The Venn diagram in Figure 1 shows three events A, B and C and the probabilities associated with each region of B. The constants p, q and r each represent probabilities associated with the three separate regions outside B.

The events A and B are independent.

(a) Find the value of p.

(3)

Given that $\mathrm{P}(B \mid C) = \frac{5}{11}$,

(b) find the value of q and the value of r.

(4)

(c) Find $\mathrm{P}(A \cup C \mid B)$.

(2)

May 2013 (R)

7. The heights of a population of women are normally distributed with mean μ cm and standard deviation σ cm. It is known that 30% of the women are taller than 172 cm and 5% are shorter than 154 cm.

(*a*) Sketch a diagram to show the distribution of heights represented by this information.
(3)

(*b*) Show that $\mu = 154 + 1.6449\sigma$.
(3)

(*c*) Obtain a second equation and hence find the value of μ and the value of σ.
(4)

A woman is chosen at random from the population.

(*d*) Find the probability that she is taller than 160 cm.
(3)

January 2010

---

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.

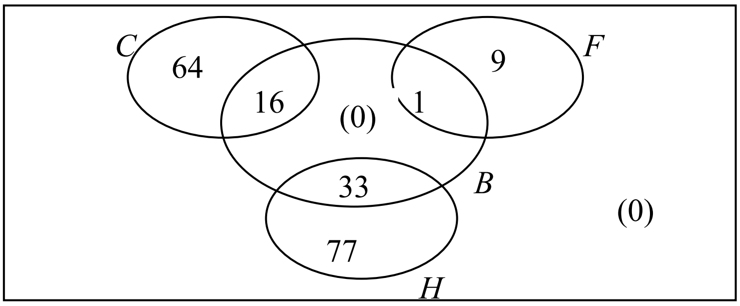
(*d*) State which company you would recommend. Give reasons for your answer.
(2)

May 2009

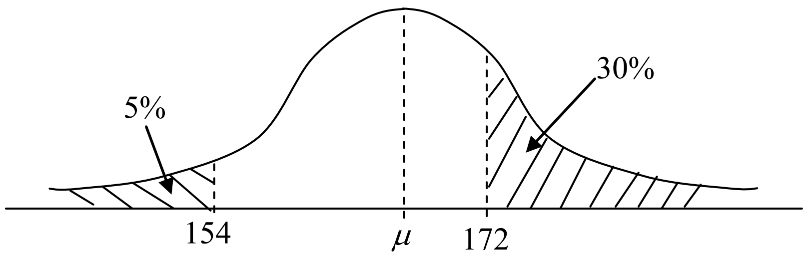
TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)	$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad = \frac{-808.917}{\sqrt{113573 \times 8.657}}$	M1
	$= -0.81579...$	A1
		(2)
(b)	Houses are <u>cheaper</u> further away from the station or equivalent statement	B1
		(1)
(c)	-0.816	B1∫
		(1)
		[4]
2. (a)	awrt $\pm$ 1.40	B1
	$\frac{23-\mu}{5} = \text{"1.40"}$ (o.e)	M1A1ft
	$\underline{\mu = 16}$ (or awrt 16.0)	A1
		(4)
(b)	<u>0.4192</u>	B1
		(1)
		[5]
3. (a)	$2a + \frac{2}{5} + \frac{1}{10} = 1$ (or equivalent)	M1
	$\underline{a = \frac{1}{4}}$ **<u>or 0.25</u>**	A1
		(2)
(b)	E(X) = **<u>1</u>**	B1
		(1)
(c)	$\text{E}(X^2) = 1\times\frac{1}{5} + 1\times\frac{1}{10} + 4\times\frac{1}{4} + 9\times\frac{1}{5}$ (= 3.1)	M1
	Var(X) = $3.1 - 1^2$, **<u>= 2.1 or</u>** $\underline{\frac{21}{10}}$ **<u>oe</u>**	M1 A1
		(3)
(d)	Var(Y) = $(-2)^2$ Var(X) , = **<u>8.4 or</u>** $\underline{\frac{42}{5}}$ **<u>oe</u>**	M1 A1
		(2)
(e)	$X \geq Y$ when X = 3 or 2, so probability = $\text{"}\frac{1}{4}\text{"} + \frac{1}{5}$	M1 A1ft
	$= \underline{\frac{9}{20}}$ **<u>oe</u>**	A1
		(3)
		[11]

Question Number	Scheme		Marks
4. (a)	$\frac{35+75}{200}=0.55$		M1 A1
			(2)
(b)	$\frac{200-2}{200}=0.99$		M1 A1
			(2)
(c)	$\left[\mathrm{P}\left(W \mid C\right)\right]=\frac{\mathrm{P}(W\cap C)}{\mathrm{P}(C)}=\frac{30/200}{80/200}=\frac{30}{80}=0.375$		M1 A1
			(2)
(d)			M1
		9, 1	B1
		77,33	B1
		64,16	B1
	Allow diagrams with intersections between *F, C* and *H* provided these are marked with 0. If their diagram indicates extra empty regions do not treat a blank as 0.		
			(4)
(e)	$\frac{1+16+33}{200}=0.25$		M1 A1
			(2)
			[12]

Question Number	Scheme	Marks
5. (a)	$\frac{2+3}{\text{their total}} = \frac{5}{\text{their total}} = \frac{1}{6}$ **(** given answer**)**	M1 A1cso (2)
(b)	$\frac{4+2+5+3}{\text{total}}, \quad = \frac{14}{30}$ or $\frac{7}{15}$ or $0.4\dot{6}$	M1 A1 (2)
(c)	$\text{P}(A \cap C) = 0$	B1 (1)
(d)	$\text{P(C} \mid \text{reads at least one magazine)} = \frac{6+3}{20} = \frac{9}{20}$	M1 A1 (2)
(e)	$\text{P}(B) = \frac{10}{30} = \frac{1}{3}, \ \text{P}(C) = \frac{9}{30} = \frac{3}{10}, \quad \text{P}(B \cap C) = \frac{3}{30} = \frac{1}{10}$ or $\text{P}(B\|C) = \frac{3}{9}$ $\text{P}(B) \times \text{P}(C) = \frac{1}{3} \times \frac{3}{10} = \frac{1}{10} = \text{P}(B \cap C)$ or $\text{P}(B\|C) = \frac{3}{9} = \frac{1}{3} = \text{P}(B)$ So yes they are statistically independent	 M1 M1 A1cso (3) **[10]**
6. (a)	[P(*B*) = 0.4, P(*A*) = *p* + 0.1 so] $0.4 \times (p + 0.1) = 0.1$ or $0.4 \times \text{P}(A) = 0.1$ $p = \frac{1}{4} - 0.1$ <u>***p* = 0.15**</u>	 M1 M1A1 (3)
(b)	$\frac{5}{11} = \left[\frac{\text{P}(B \cap C)}{\text{P}(C)} = \right] \frac{0.2}{0.2+q}$ or $\frac{5}{11} = \frac{0.2}{\text{P}(C)}$ $11 \times 0.2 = 5 \times (0.2 + q)$ <u>***q* = 0.24**</u> $r = 0.6 - (p + q)$ i.e. <u>***r* = 0.21**</u>	M1 dM1 A1 A1ft (4)
(c)	$\left[\frac{\text{P}((A \cup C) \cap B)}{\text{P}(B)}\right] = \frac{0.3}{0.4}$ = <u>**0.75**</u>	M1 A1 (2) **[9]**

Question Number	Scheme	Marks
7. (a)		
	bell shaped, must have inflexions	B1
	154,172 on axis	B1
	5% and 30%	B1
		(3)
(b)	$\text{P}(X < 154) = 0.05$	M1
	$\frac{154-\mu}{\sigma} = -1.6449$ or $\frac{\mu-154}{\sigma} = 1.6449$	B1
	$\mu = 154 + 1.6449\sigma$ **given**	A1 cso
		(3)
(c)	$172 - \mu = 0.5244\sigma$ or $\frac{172-\mu}{\sigma} = 0.5244$	B1
	(allow z = 0.52 or better here but must be in an equation)	
	Solving gives $\sigma = 8.2976075$ (**awrt 8.30**) and $\mu = 167.64873$ (**awrt 168**)	M1 A1 A1
		(4)
(d)	P(Taller than 160cm) = $\text{P}\left(Z > \frac{160-\mu}{\sigma}\right)$	M1
	$= \text{P}(Z < 0.9217994)$	B1
	$= 0.8212$ **awrt 0.82**	A1
		(3) **[13]**

Question Number	Scheme		Marks
8. (a)	Let the random variable X be the lifetime in hours of bulb		
	P(X < 830) =P(Z< $\frac{\pm(830-850)}{50}$)	Standardising with 850 and 50	M1
	= P(Z <–0.4)		
	= 1 – P(Z < 0.4)	Using 1-(probability>0.5)	M1
	= 1 – 0.6554		
	= 0.3446 or 0.344578 by calculator	awrt 0.345	A1 (3)
(b)	0.3446 × 500	Their (a) x 500	M1
	= 172.3	Accept 172.3 or 172 or 173	A1 (2)
(c)	Standardise with 860 and σ		
	and equate to z value $\frac{\pm(818-860)}{\sigma}$ = z value		M1
	$\frac{818-860}{\sigma}$ = – 0.84(16) or $\frac{860-818}{\sigma}$ = 0.84(16)		
	or $\frac{902-860}{\sigma}$ =0.84(16) or equiv.		A1
		±0.8416(2)	B1
	σ = 49.9	50 or awrt 49.9	A1 (4)
(d)	Company Y as the ***mean*** is greater for Y.	both	B1
	They have (approximately) the same ***standard deviation*** or ***sd***		B1 (2) **[11]**

Examiner reports

Question 1

Most candidates had little trouble with part (a). In part (b) a sizeable minority correctly identified negative correlation, but failed to put it into the context of the question. In part (c) again a sizeable minority did not attempt this part or multiplied their answer for part (a) by some arbitrary factor.

Question 2

Most candidates knew how to standardise and there were very few dividing by 5^2 instead of 5. The usual problem arose with confusion of probabilities and z values and many simply equated their standardised expression to 0.9192. Those who did use 1.4 though invariably solved their equation successfully to reach 16. Part (b) was a 1 mark question but some candidates wrote several lines with various degrees of success. Common errors were to give 0.9192 or 1 – 0.9192.

Question 3

Finding the correct value of a in the first part of the question proved to be relatively straightforward for most candidates. Few errors were seen although some candidates provided very little in the way of working out and did not always make it explicit that they were using the fact that the sum of the probabilities equals one. Similarly, most candidates were able to obtain the correct value of $\text{E}(X)$, though not many deduced this fact by recognising the symmetry of the distribution.

The majority opted to use the formula to calculate $\text{E}(X)$, which resulted in processing errors in some cases. Common errors seen in calculating $\text{Var}(X)$ included forgetting to subtract $[\text{E}(X)]^2$ from $\text{E}(X^2)$ or calculating $\text{E}(X^2) - \text{E}(X)$, although on the whole the correct formula was successfully applied.

Most candidates were able to correctly apply $\text{Var}(aX + b) = a^2\,\text{Var}(X)$ to deduce $\text{Var}(Y) = 4\,\text{Var}(X)$, although $\text{Var}(Y) = 6 - 2\text{Var}(X)$ was a typical error. Quite a number of candidates attempted to calculate $\text{E}(Y^2) - [\text{E}(Y)]^2$ with varying degrees of success. Occasionally, candidates divided their results in part (b), part (c) and part (d) by 5.

The final part of the question proved to be the most challenging of all and was often either completely omitted or poorly attempted with little or no success. Only a minority of candidates knew they would need to equate $6 - 2X$ to X in order to obtain the corresponding values of X and of those who did, only a small number scored full marks, as candidates were generally unable to identify the correct values of X.

Question 4

Part (a) was answered very well but in part (b) a number of candidates failed to spot or consider the complement (giving an answer of 0.01) and others confused the 200 with 100 and gave an answer of 0.98.

The conditional probability in part (c) was answered quite well but a few had $\text{P}(W)$ on their denominator and some assumed independence when calculating the numerator and used $\text{P}(C) \times \text{P}(W) = \text{P}(C \cap W)$.

Part (d) was a little different from the usual Venn diagram and candidates had to consider carefully how to represent the 4 events. Three overlapping circles or 3 separate circles with no indication of set B was quite common and those who did have a correct shape sometimes struggled to place the frequencies or probabilities. Those using frequencies were usually more successful as the probabilities were not always out of 200. A Venn diagram such as this should, of course, always have a box defining the universal set [and ideally a 0 for the region $(F \cup C \cup H \cup B')$] and a few candidates missed this out.

Despite their difficulties with the structure of the diagram for part (d) many candidates were able to interpret the table correctly and score the marks in part (e).

Question 5

Overall this question proved to be quite challenging for candidates and incorrect interpretation of the Venn diagram lost many candidates marks. In spite of this, most candidates had no trouble proving the given probability in part (a).

In part (b), however, quite a number of candidates neglected one of the four components of the numerator, usually the 3, and $\frac{11}{30}$ was consequently an extremely common wrong answer. Other wrong answers included $\frac{9}{30}$, $\frac{13}{30}$ and $\frac{16}{30}$. Some candidates chose to use the addition rule, which was generally written down correctly, although quite often P(A) was given as $\frac{4}{30}$ and P(B) as $\frac{5}{30}$, giving rise to $P(A \cup B) = \frac{7}{30}$.

In contrast, the majority of candidates were able to deduce that $P(A \cap C) = 0$ and quite a few gave explanations as part of their answer, such as ‘there is no overlap’, or ‘no intersection’ and some even discussed the idea of mutual exclusivity. A small proportion of candidates had the right idea but failed to give a probability, giving their answer as ‘nobody’ or in a few cases ‘the empty set’. However, not all of the candidates realised that mutually exclusive events have a probability of 0 of occurring together and some mistakenly thought that $P(A \cap C)$ equalled P(A)P(C) here.

Answers to part (d) were extremely varied. Most candidates did not recognise that a conditional probability was required and consequently did not obtain the correct denominator. Common wrong answers were $\frac{6}{30}$, $\frac{6}{20}$ and $\frac{3}{20}$. A significant number attempted to perform some complex calculations in which they tried unsuccessfully to use the formula for conditional probability. Very few candidates used the Venn diagram to calculate the probability directly.

Testing for independence was generally performed successfully overall, with the majority of candidates carrying out suitable tests. However, some candidates did find this challenging and often the wrong probabilities were compared and some incorrect probabilities were obtained. A number of candidates appeared to be confusing independence with mutual exclusivity. Some candidates merely provided a comment on the perceived nature of independence without performing any calculations at all. Of those candidates who were successful, the most common approach was to test whether $P(B \cap C) = P(B)P(C)$, although there were a few cases where $P(A \cap C)$ was compared with P(A)P(C) by mistake. Rather worryingly, a surprisingly high number of candidates failed to recognise $\frac{3}{30}$ and $\frac{1}{10}$ as equivalent fractions and thus concluded that the events were not independent.

Question 6

In part(a) most used the independence property correctly to show that P(A) = 0.25 but some mistakenly assumed p = P(A). The conditional probability formula was usually used correctly in part (b) to find P(C) and often the value of q as well and many were also able to find the value of r too although occasionally candidates seemed to miss this demand. There were some good responses to part (c) despite the rather unusual nature of the conditional probability. Many candidates were able to write down a correct ratio of probabilities and there weren't too many cases of candidates attempting to evaluate the numerator as 0.69×0.4.

Question 7

Part (a) was usually answered well but the remaining parts of the question proved challenging for many. There was much muddled work in part (b) and although some scored M1B1 for attempts such as $\frac{154-\mu}{\sigma}=1.6449$ very few scored the A1cso for a completely correct derivation without any incorrect statements being seen. Those who fumbled their way to the printed answer in (b) usually came unstuck in part (c). A common error was to write $\frac{172-\mu}{\sigma}=0.5244$ and then replace the 0.5244 with 1-0.5244. Those with a correct pair of equations were usually able to solve them correctly to find σ and μ. Many attempted to standardise in part (d) but even those with correct answers in (c) often failed to score full marks either due to premature rounding or because they thought their final answer was $1 - 0.8212$. Curiously even a correct diagram failed to prevent some of them from making this final error.

Question 8

More able candidates made a good start to this question. Part (a) was well done and part (b) usually gained the method mark. Part (c) proved to be much more of a challenge despite it being very similar to questions set in previous papers on this topic. Many candidates gained 4 marks, and there were a number who could see what was required but could not quite answer fully, submitting solutions which had most of the components, but not in the right sequence. Many adjusted the sign, either losing it during the calculation, or right at the end when -50 did not appear to be correct. Many candidates did not use 0.8416, settling for 0.84. A few used a probability rather than a z value but this was less than in previous years. Some candidates drew diagrams to help their thought processes. In part (d) candidates lost marks as they were not confident in interpreting what their figures meant. Many candidates did not use correct statistical language and thus lost the marks, others commented on the standard deviation for Y being lower than that for X without considering the magnitude of the difference.

Statistics for S1 Practice Paper Silver Level S3

Qu	Max Score	Modal score	Mean %	Mean score for students achieving grade: ALL	A*	A	B	C	D	E	U
1	4		69	2.76		3.40	3.00	2.68	2.47	2.26	1.80
2	5		69	3.43	4.93	4.79	4.28	3.60	2.84	1.99	0.92
3	11		58	6.39	9.72	8.57	7.27	6.43	5.57	4.54	2.61
4	12	12	58	6.99	10.73	10.05	8.08	6.86	5.84	5.03	3.61
5	10		58	5.82	8.81	7.94	6.27	5.38	4.66	4.12	3.38
6	9		80	7.16	8.75	8.40	7.55	6.30	4.88	3.74	1.85
7	13		55	7.14		10.10	7.67	5.47	4.08	2.94	1.43
8	11		53	5.80		8.73	6.95	5.63	4.31	3.01	1.37
	75		**61**	**45.49**		**61.98**	**51.07**	**42.35**	**34.65**	**27.63**	**16.97**