Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Silver Level S4

Time: 1 hour 30 minutes

Materials required for examination papers	**Items included with question**
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 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
62	55	48	41	36	29

1. The histogram in Figure 1 shows the time, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway.
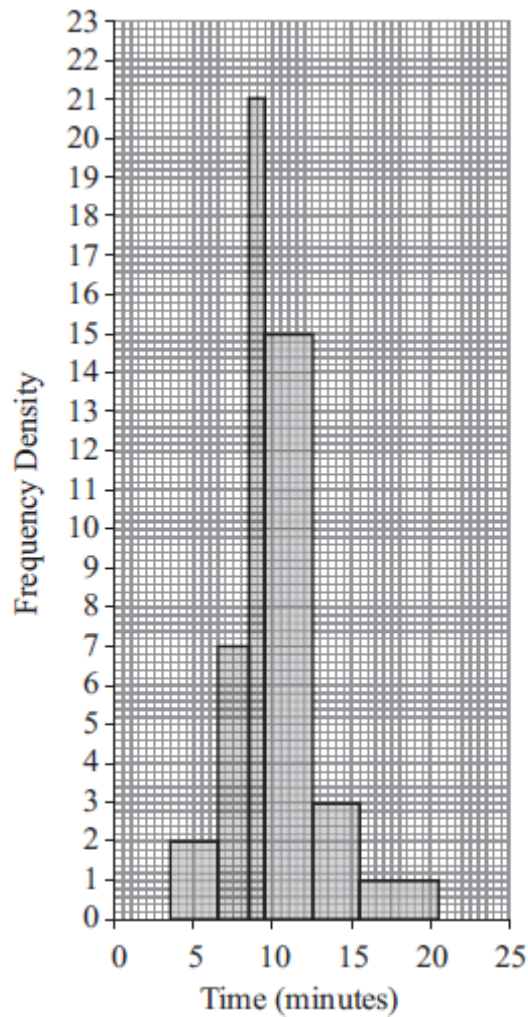

Figure 1

(a) Complete the table.

Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	21
10 – 12	45
13 – 15	9
16 – 20	

(2)

(b) Estimate the number of motorists who were delayed between 8.5 and 13.5 minutes by the roadworks.

(2)

January 2012

2. When Rohit plays a game, the number of points he receives is given by the discrete random variable X with the following probability distribution.

x	0	1	2	3
$P(X = x)$	0.4	0.3	0.2	0.1

(*a*) Find E(X).

(2)

(*b*) Find F(1.5).

(2)

(*c*) Show that Var(X) = 1.

(4)

(*d*) Find Var(5 – 3X).

(2)

Rohit can win a prize if the total number of points he has scored after 5 games is at least 10. After 3 games he has a total of 6 points. You may assume that games are independent.

(*e*) Find the probability that Rohit wins the prize.

(6)

January 2009

---

3. A survey of the reading habits of some students revealed that, on a regular basis, 25% read quality newspapers, 45% read tabloid newspapers and 40% do not read newspapers at all.

(*a*) Find the proportion of students who read both quality and tabloid newspapers.

(3)

(*b*) Draw a Venn diagram to represent this information.

(3)

A student is selected at random. Given that this student reads newspapers on a regular basis,

(*c*) find the probability that this student only reads quality newspapers.

(3)

June 2007

---

4. A survey of 100 households gave the following results for weekly income £y.

Income y (£)	Mid-point	Frequency f
$0 \le y < 200$	100	12
$200 \le y < 240$	220	28
$240 \le y < 320$	280	22
$320 \le y < 400$	360	18
$400 \le y < 600$	500	12
$600 \le y < 800$	700	8

(You may use $\sum fy^2 = 12\,452\,800$)

A histogram was drawn and the class $200 \le y < 240$ was represented by a rectangle of width 2 cm and height 7 cm.

(*a*) Calculate the width and the height of the rectangle representing the class $320 \le y < 400$.

(3)

(*b*) Use linear interpolation to estimate the median weekly income to the nearest pound.

(2)

(*c*) Estimate the mean and the standard deviation of the weekly income for these data.

(4)

One measure of skewness is $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$.

(*d*) Use this measure to calculate the skewness for these data and describe its value.

(2)

Katie suggests using the random variable X which has a normal distribution with mean 320 and standard deviation 150 to model the weekly income for these data.

(*e*) Find P$(240 < X < 400)$.

(2)

(*f*) With reference to your calculations in parts (*d*) and (*e*) and the data in the table, comment on Katie's suggestion.

(2)

January 2013

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

$$J = \text{the event that Jake is late for school,}$$
$$K = \text{the event that Kamil is late for school.}$$

$P(J) = 0.25$, $P(J \cap K) = 0.15$ 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)

May 2011

---

6.   A manufacturer fills jars with coffee. The weight of coffee, W grams, in a jar can be modelled by a normal distribution with mean 232 grams and standard deviation 5 grams.

(*a*) Find $P(W < 224)$.

(3)

(*b*) Find the value of w such that $P(232 < W < w) = 0.20$.

(4)

Two jars of coffee are selected at random.

(*c*) Find the probability that only one of the jars contains between 232 grams and w grams of coffee.

(3)

January 2012

---

7. The weight, X grams, of soup put in a tin by machine A is normally distributed with a mean of 160 g and a standard deviation of 5 g.

A tin is selected at random.

(*a*) Find the probability that this tin contains more than 168 g.

(3)

The weight stated on the tin is w grams.

(*b*) Find w such that $P(X < w) = 0.01$.

(3)

The weight, Y grams, of soup put into a carton by machine B is normally distributed with mean μ grams and standard deviation σ grams.
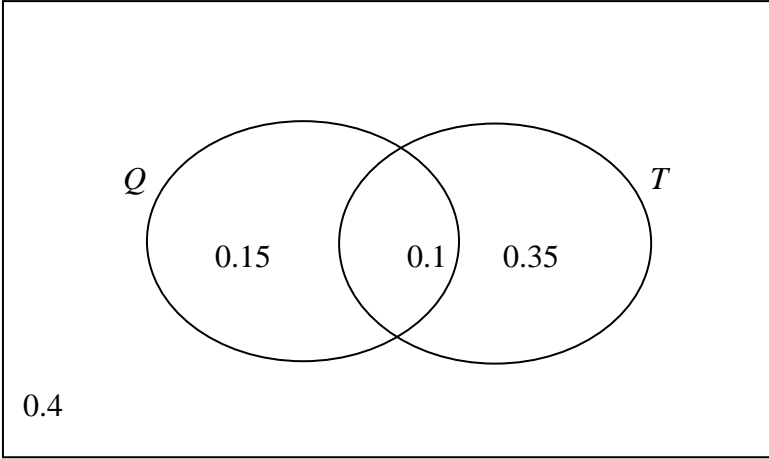
(*c*) Given that $P(Y < 160) = 0.99$ and $P(Y > 152) = 0.90$, find the value of μ and the value of σ.

(6)

January 2011

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)	14, 5	M1 A1 (2)
(b)	$21 + 45 + 3 = 69$	M1 A1 (2) **[4]**
2. (a)	$E(X) = 0 \times 0.4 + 1 \times 0.3 + ... + 3 \times 0.1, \quad = \quad 1$	M1, A1 (2)
(b)	$F(1.5) = \left[P(X \le 1.5) = \right] P(X \le 1), \quad = 0.4 + 0.3 = 0.7$	M1, A1 (2)
(c)	$E(X^2) = 0^2 \times 0.4 + 1^2 \times 0.3 + ... + 3^2 \times 0.1 \quad, = \quad 2$	M1, A1
	$Var(X) = 2 - 1^2 , \quad = 1 \qquad (*)$	M1, A1cso (4)
(d)	$Var(5 - 3X) = \left(-3\right)^2 Var(X), \quad = 9$	M1, A1 (2)
(e)		B1B1B1 M1

Total	Cases	Probability
4	$(X = 3) \cap (X = 1)$	$0.1 \times 0.3 = 0.03$
	$(X = 1) \cap (X = 3)$	$0.3 \times 0.1 = 0.03$
	$(X = 2) \cap (X = 2)$	$0.2 \times 0.2 = 0.04$
5	$(X = 3) \cap (X = 2)$	$0.1 \times 0.2 = 0.02$
	$(X = 2) \cap (X = 3)$	$0.2 \times 0.1 = 0.02$
6	$(X = 3) \cap (X = 3)$	$0.1 \times 0.1 = 0.01$

A1

Total probability $= 0.03 + 0.03 + 0.04 + 0.02 + 0.02 + 0.01 = 0.15$

A1 (6) **[16]**

Question Number	Scheme		Marks
3. (a)	$\mathrm{P}(Q \cup T)$ $\qquad = 0.6$		B1
	$\mathrm{P}(Q) + \mathrm{P}(T) - \mathrm{P}(Q \cap T) \qquad = 0.6$		M1
	$\mathrm{P}(Q \cap T) \qquad = 0.1$		A1
			(3)
(b)		Venn	M1
		0.15, 0.35	A1∫
		0.4 and box	B1
			(3)
(c)	$P(Q \cap T' \vert Q \cup T) = \dfrac{0.15}{0.60} = \dfrac{1}{4}$ or 0.25 or 25%		M1A1∫ A1
			(3)
			[9]

Question Number	Scheme	Marks
4. (a)	Width = 4 (cm)	B1
	Area of 14 cm^2 represents frequency 28 and area of $4h$ represents 18	M1
	Or $\dfrac{4h}{18} = \dfrac{14}{28}$ (o.e.) $\qquad\qquad\qquad\qquad\qquad$ $h = \underline{\mathbf{2.25}}$ (cm)	A1
		(3)
(b)	$m = (240) + \dfrac{10}{22} \times 80$ (o.e.)	M1
	$= 276.36...$ ($\frac{3040}{11}$) $\qquad\qquad$ ((£)$\underline{\mathbf{276}} \le m < $ (£)276.5)	A1
		(2)
(c)	$\sum fy = 31600$ leading to $\overline{y} = 316$	M1A1
	$\sigma_y = \sqrt{\dfrac{12452800}{100} - \left(\overline{y}\right)^2}$ $\;= 157.07...$ (awrt $\underline{\mathbf{157}}$) Allow $s = 157.86...$	M1A1
		(4)
(d)	Skewness $= 0.764...$ $\qquad$ (awrt $\underline{\mathbf{0.76}}$ or $\underline{\mathbf{0.75}}$)	B1
	$\qquad\qquad$ [If $n+1$ used in (b) and $m = £278$ accept awrt 0.73 or 0.72]	
	<u>Positive</u> skew	B1ft
		(2)
(e)	$z = \pm\dfrac{80}{150}$	M1
	P(240< X < 400) = $\underline{\mathbf{0.40 \sim 0.41}}$	A1
		(2)
(f)	(e) suggests a reasonable fit for this range BUT (d) since skew it will not be a good fit overall	B2/1/0
		(2)
		[15]

Question Number	Scheme	Marks
5. (a)	$P(J \cup K) = 1 - 0.7$ or $0.1 + 0.15 + 0.05 = $ <u>0.3</u>	B1 (1)
(b)	$P(K) = 0.05 + 0.15$ or "0.3" $- 0.25 + 0.15$ or "0.3" $= 0.25 + P(K) - 0.15$ $= $ <u>0.2</u>	M1 A1 (2)
(c)	$[P(K\,|\,J)] = \dfrac{P(K \cap J)}{P(J)}$ $= \dfrac{0.15}{0.25}$ $= \dfrac{3}{5}$ or 0.6	M1 A1 A1 (3)
(d)	$P(J) \times P(K) = 0.25 \times 0.2 (= 0.05), \quad P(J \cap K) = 0.15$ <u>or</u> $P(K\,|\,J) = 0.6, \ P(K) = 0.2$ <u>or</u> 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 <u>or</u> they are linked (This requires a statement about independence in (d) or in (e))	B1ft (1) **[9]**
6. (a)	$P(W < 224) \quad = P\left(z < \dfrac{224 - 232}{5} \right)$ $= P(z < -1.6)$ $= 1 - 0.9452$ $= 0.0548$ awrt 0.0548	M1 M1 A1 (3)
(b)	$0.5 - 0.2 = 0.3$ 0.3 or 0.7 seen $\dfrac{w - 232}{5} = 0.5244$ 0.5244 seen; any z $w = 234.622$ awrt 235	M1 B1; M1 A1 (4)
(c)	$0.2 \times (1 - 0.2)$ $2 \times 0.8 \times (1 - 0.8) = 0.32$	M1 M1 A1 (3) **[10]**

Question Number	Scheme	Marks
7. (a)	$P(X > 168) = P\left(Z > \dfrac{168 - 160}{5} \right)$	M1
	$= P(Z > 1.6)$	A1
	$= 0.0548$ **awrt 0.0548**	A1
		(3)
(b)	$P(X < w) = P\left(Z < \dfrac{w - 160}{5} \right)$	
	$\dfrac{w - 160}{5} = -2.3263$	M1 B1
	$w = 148.37$ **awrt 148**	A1
		(3)
(c)	$\dfrac{160 - \mu}{\sigma} = 2.3263$	M1 B1
	$\dfrac{152 - \mu}{\sigma} = -1.2816$	B1
	$160 - \mu = 2.3263\sigma$	
	$152 - \mu = -1.2816\sigma$	
	$8 = 3.6079\sigma$	M1
	$\sigma = 2.21\ldots$ **awrt 2.22**	A1
	$\mu = 154.84\ldots$ **awrt 155**	A1
		(6)
		[12]

Examiner reports

Question 1

Part (a) was generally well answered, with most candidates getting the correct values. Only a few stated 7 and 1 (frequency density) and a few quoted other figures. A simple check that the total number of motorists should add to 100 would have told candidates if their calculations were incorrect. Some could not cope with the different interval widths and/or intervals starting and ending in ".5". This was more evident in part (b) where many struggled. The most common error was simply adding 21 and 45 not realising that 13.5 took them into the next group. A large number of candidates did realise that they needed to add three different frequencies together, although some found it difficult to get the correct fraction of the third class. Some ignored the third class or used its whole frequency.

Question 2

Part (a) was answered well although a small minority of candidates insisted on dividing by n (where n was usually 4). Part (b), on the other hand, caused great confusion. Some interpreted F(1.5) as E(1.5X), others interpolated between P(X = 1) and P(X = 2) and a few thought that F(1.5) was zero since X has a discrete distribution. Although the majority of candidates gained full marks in part (c) the use of notation was often poor. Statements such as Var(X) = 2 = 2-1 = 1 were rife and some wrote Var(X) or $\sum X^2$ when they meant E(X^2). Many candidates can now deal with the algebra of Var(X) but there were the usual errors such as 5Var(X) or 25Var(X) or -3Var(X) and the common -3^2 Var(X) which was condoned if the correct answer followed.

Part (e) was not answered well and some candidates did not attempt it. Those who did appreciate what was required often missed one or more of the possible cases or incorrectly repeated a case such as (2, 2). There were many fully correct responses though often aided by a simple table to identify the 6 cases required.

Question 3

Many candidates were able to determine the correct answer for part (a) but a very common error was to multiply the two probabilities, incorrectly assuming independence. Many candidates used the Venn diagram to attain the correct solution to part (b). The most common errors were to omit a box or add a third circle. In part (c), as is often the case in this type of question, many failed to realise this was a conditional probability

Question 4

In part (a) there is still some uncertainty about the concept of histograms. Most candidates could find the width of the required bar but a number still failed to consider frequency densities to find the height. Some attempts were on the right lines but used the interval [240, 320); others assumed height was linked to frequency alone and 4.5 cm was a common incorrect answer. There has been a clear improvement in the use of interpolation to calculate the median in recent sessions and many had $\frac{10}{22} \times 80$ (with only a small minority using 10.5 instead of 10). The common error was to use 239.5 rather than 240 as the class boundary but there were many fully correct answers seen.

The calculation of the mean in part (c) was usually correct but the standard deviation still causes difficulties for some who failed to use a correct formula. Part (d) was straightforward although some failed to secure the second mark because they did not give a description (e.g. "positive") of their skewness.

In part (e) the insertion of a normal calculation in this context seemed to fox some candidates but many were able to carry out this routine calculation correctly. However the interpretation in part (f) was not answered so well. Some drew the connection between their answer to part (e) and the proportion in the table and deduced that a normal distribution was reasonable (over this range). Slightly more focussed on the fact that in part (d) they had stated that the distribution was skewed and therefore a normal distribution was probably not a suitable model (overall) but few were able to marry these statements up and secure both marks.

Question 5

It was encouraging to see many candidates drawing a correct Venn diagram for the start of this question but the wording "at least one" in part (a) was often misinterpreted as the intersection and some gave the probability of "only K" in part (b). There were many correct answers seen in part (c) although many candidates simply wrote down the correct probability ratio without attempting an expression in symbols. Inevitably some wrote $P(J|K)$ but the identification and calculation of conditional probabilities seems to be improving. In part (d) most candidates opted for the $P(K \cap J) = P(K) \times P(J)$ test whilst others attempted other incorrect comparisons such as $P(K \cap J) = 0$ or even $P(K \cap J) = P(K) + P(J)$. Some merely quoted a standard formula in terms of A and B which, without some identification of what their A and B represented gained no credit. Only the very best realised that they had just calculated $P(K|J)$ and $P(K)$ and since these two were not equal the two events were not independent. Part (e) discriminated very well with good candidates engaging with the context and showing they clearly understood the concept of independence.

Question 6

It was pleasing to see fewer blank pages than in the past although full marks were rarely gained for this question. Part (a) was answered well, if it was incorrect it was usually because candidates standardised with 25 as the standard deviation instead of 5 or did not subtract their probability from 1.

Part (b) was less well done and very few drew a diagram which helped with the areas and probabilities. Those who knew what to do often forgot to use the percentage points table and it was rare to see $z = 0.5244$ used. Some candidates used probabilities instead of z values, and some used the z value of 0.8416, although sign errors were few.

Part (c) was challenging for a large number of candidates and was not attempted if they had struggled with the earlier parts. The value of 0.16 was often seen in the scripts of those who tried it, ignoring the different ways of selection. Some candidates tried using the normal distribution to solve this part, or used values from part (b) such as 0.5793 in their work with lots of elaborate wrong calculations. Those who drew a tree diagram usually scored full marks.

Question 7

There were some good responses to this question and even some of those who had struggled with parts of questions 6 and 7 [*on the January 2011 paper: this question was originally qu.8 on that paper*] were able to pick up a good score here.

Part (a) was answered well and it was good to see diagrams being used to assist the candidates. Part (b) still causes problems for many candidates. They can usually standardise but then far too often equate this expression to 0.01 or 0.99. Those who did use a z value often used the "small" table and found the 2.3263 value but the minus sign was often missed and the final answer was therefore incorrect. Most seemed to try part (c) and the standardising was usually correct and suitable z values were often seen (use of 2.32 or 2.33 and 1.28 were acceptable here) but, when suitable equations were formed, a minus sign was often missing from the second equation. Solving their two linear equations was carried out quite well but full marks were only secured by those who worked carefully and accurately throughout this part.

Statistics for S1 Practice Paper Silver Level S4

Qu	Max Score	Modal score	Mean %	ALL	A*	A	B	C	D	E	U
1	4		68	2.70	3.54	3.34	2.72	2.43	2.03	1.66	1.32
2	16		58	9.24		11.41	9.28	7.74	6.54	5.19	2.53
3	9		58	5.19		7.08	5.45	4.78	4.31	3.87	3.12
4	15	14	62	9.23	12.72	11.86	9.86	8.05	6.51	4.87	2.75
5	9		54	4.87	8.36	7.63	5.88	4.49	3.29	2.29	1.23
6	10		51	5.10	8.65	7.40	5.17	3.78	2.86	2.39	1.28
7	12		55	6.61	10.72	9.38	6.84	5.23	3.90	2.82	1.55
	75		**57**	**42.94**		**58.10**	**45.20**	**36.50**	**29.44**	**23.09**	**13.78**