Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Silver Level S3

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 S2), the paper reference (6684), 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
69	62	53	44	36	26

1. A string AB of length 5 cm is cut, in a random place C, into two pieces. The random variable X is the length of AC.

(*a*) Write down the name of the probability distribution of X and sketch the graph of its probability density function.
(3)

(*b*) Find the values of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.
(3)

(*c*) Find $\mathrm{P}(X > 3)$.
(1)

(*d*) Write down the probability that AC is 3 cm long.
(1)

---

2. The continuous random variable Y has cumulative distribution function

$$\mathrm{F}(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{4}(y^3 - 4y^2 + ky) & 0 \le y \le 2 \\ 1 & y > 2 \end{cases}$$

where k is a constant.

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

(*b*) Find the probability density function of Y, specifying it for all values of y.
(3)

(*c*) Find $\mathrm{P}(Y > 1)$.
(2)

---

3. A robot is programmed to build cars on a production line. The robot breaks down at random at a rate of once every 20 hours.

(*a*) Find the probability that it will work continuously for 5 hours without a breakdown.
(3)

Find the probability that, in an 8 hour period,

(*b*) the robot will break down at least once,
(3)

(*c*) there are exactly 2 breakdowns.
(2)

In a particular 8 hour period, the robot broke down twice.

(*d*) Write down the probability that the robot will break down in the following 8 hour period. Give a reason for your answer.
(2)

---

4. The lifetime, X, in tens of hours, of a battery has a cumulative distribution function F(x) given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 1 \\ \frac{4}{9}(x^2 + 2x - 3) & 1 \le x \le 1.5 \\ 1 & x > 1.5 \end{cases}$$

(*a*) Find the median of X, giving your answer to 3 significant figures.

(3)

(*b*) Find, in full, the probability density function of the random variable X.

(3)

(*c*) Find $\mathrm{P}(X \geq 1.2)$

(2)

A camping lantern runs on 4 batteries, all of which must be working. Four new batteries are put into the lantern.

(*d*) Find the probability that the lantern will still be working after 12 hours.

(2)

---

5. Defects occur at random in planks of wood with a constant rate of 0.5 per 10 cm length. Jim buys a plank of length 100 cm.

(*a*) Find the probability that Jim's plank contains at most 3 defects.

(2)

Shivani buys 6 planks each of length 100 cm.

(*b*) Find the probability that fewer than 2 of Shivani's planks contain at most 3 defects.

(5)

(*c*) Using a suitable approximation, estimate the probability that the total number of defects on Shivani's 6 planks is less than 18.

(6)

---

6. The probability that a sunflower plant grows over 1.5 metres high is 0.25. A random sample of 40 sunflower plants is taken and each sunflower plant is measured and its height recorded.

(*a*) Find the probability that the number of sunflower plants over 1.5 m high is between 8 and 13 (inclusive) using

(i) a Poisson approximation,

(ii) a Normal approximation.

(10)

(*b*) Write down which of the approximations used in part (*a*) is the most accurate estimate of the probability. You must give a reason for your answer.

(2)

---

7. The random variable Y has probability density function f(y) given by

$$f(y) = \begin{cases} ky(a-y) & 0 \le y \le 3 \\ 0 & \text{otherwise} \end{cases}$$

where k and a are positive constants.

(a) (i) Explain why $a \ge 3$.

(ii) Show that $k = \dfrac{2}{9(a-2)}$.

(6)

Given that $E(Y) = 1.75$,

(b) show that $a = 4$ and write down the value of k.

(6)

For these values of a and k,

(c) sketch the probability density function,

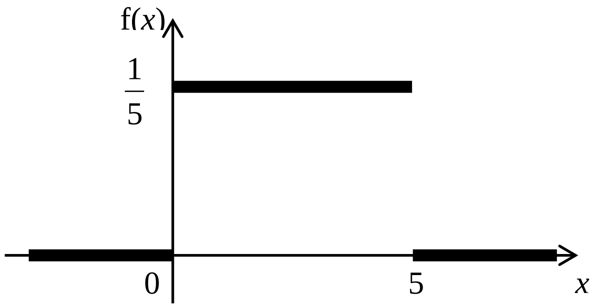
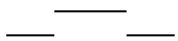
(2)

(d) write down the mode of Y.

(1)

TOTAL FOR PAPER: 75 MARKS

END

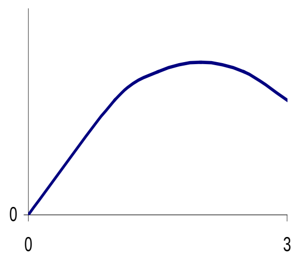
Question Number	Scheme	Marks
1(a)	<u>Continuous</u> <u>uniform</u> distribution *or* <u>rectangular</u> distribution.	B1
	 1/5,(0),5	B1
	0 may be implied by start at y axis 	B1
		(3)
(b)	$E(X) = 2.5$ ft from their a and b, must be a number	B1ft
	$\text{Var}(X) = \frac{1}{12}(5-0)^2$ or attempt to use $\int_0^5 f(x)x^2 dx - \mu^2$ use their f(x)	M1
	$= \frac{25}{12}$ or 2.08 o.e awrt 2.08	A1
		(3)
(c)	$P(X > 3) = \frac{2}{5} = 0.4$ 2 times their 1/5 from diagram	B1ft
		(1)
(d)	$P(X = 3) = 0$	B1
		(1)
		(Total 8)

Question Number	Scheme	Marks
2.		
(a)	F(2) = 1 gives: $\frac{1}{4}\left(2^3 - 4\times 2^2 + 2k\right) = 1$	M1
	$\underline{\boldsymbol{k = 6}}$	A1
		(2)
(b)	f(y) = $\frac{\mathrm{d}}{\mathrm{d}y}\left(\mathrm{F}(y)\right) = \frac{1}{4}\left(3y^2 - 8y + \text{"6"}\right)$	M1A1ft
	$\mathrm{f}(y) = \begin{cases} \frac{1}{4}\left(3y^2 - 8y + 6\right) & 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$	A1
		(3)
(c)	P(Y > 1) = 1 – F(1) = $1 - \frac{1}{4}\left(1^3 - 4\times 1^2 + k\right)$	M1
	$= \frac{1}{4}$ (o.e.)	A1
		(2) **[7]**

Question Number	Scheme	Marks
3. (a)	$Y \sim \mathrm{Po}(0.25)$	B1
	P(Y=0) $= \mathrm{e}^{-0.25}$	M1
	$= 0.7788$	A1
		(3)
(b)	$X \sim \mathrm{Po}(0.4)$	B1
	P(Robot will break down) $= 1 - \mathrm{P}(X = 0)$	
	$= 1 - e^{-0.4}$	M1
	$= 1 - 0.067032$	
	$= 0.3297$	A1
		(3)
(c)	P(X = 2) $= \frac{e^{-0.4}(0.4)^2}{2}$	M1
	$= 0.0536$	A1
		(2)
(d)	0.3297 or answer to part (b)	B1ft
	as Poisson events are <u>independent</u>	B1 dep
		(2)
		Total [10]

Question Number	Scheme	Marks
4. (a)	$\frac{4}{9}\left(m^2+2m-3\right)$ = 0.5	M1
	m2 + 2m – 4.125 = 0	
	m = $\frac{-2\pm\sqrt{4+16.5}}{2}$	M1
	m =1.26, -3.264	
	(median =) 1.26	A1 (3)
(b)	Differentiating $\frac{\mathrm{d}\left(\frac{4}{9}\left(x^2+2x-3\right)\right)}{\mathrm{d}x}=\frac{4}{9}\left(2x+2\right)$	M1 A1
	$\mathrm{f}(x)=\begin{cases}\frac{8}{9}\left(x+1\right) & 1\le x\le 1.5\\ 0 & \text{otherwise}\end{cases}$	B1ft (3)
(c)	P(X > 1.2) = 1 – F(1.2)	M1
	= 1 – 0.3733	
	= $\frac{47}{75}$, 0.6267 awrt 0.627	A1 (2)
(d)	(0.6267)4= 0.154 awrt 0.154 or 0.155	M1 A1 (2)
		[10]

Question Number	Scheme	Marks
5. (a)	X ~Po(5); P($X \le 3$) = 0.2650	M1 A1 (2)
(b)	Let Y = the no.of planks with at most 3 defects, Y ~Binomial Y~B(6, 0.265) P(Y < 2) = P($Y \le 1$) $= \left[0.735^6 + 6 \times 0.265 \times 0.735^5\right]$ = 0.4987…. **awrt 0.499 or 0.498**	M1 A1ft M1 A1 A1 (5)
(c)	Let T = total number of defects on 6 planks, T ~Po(30) so $T \approx S$ ~Normal S~N(30, 30) P(T < 18) = P(S < 17.5) $= P\left(z < \frac{17.5-30}{\sqrt{30}}\right)$ = P(Z < - 2.28…) = 0.01123... **awrt 0.0112 or 0.0113**	M1 A1 M1 M1 A1 A1 (6) **13**
6. (a) (i)	Let X represent the number of sunflower plants more than 1.5m high X~ Po(10) μ=10 P($8 \le X \le 13$) = P($X \le 13$) – P($X \le 7$) = 0.8645 – 0.2202 = 0.6443 awrt 0.644	 **B1** **M1** **A1**
(ii)	X~ N(10,7.5) P($7.5 \le X \le 13.5$) = $\text{P}\left(\frac{17.5-10}{\sqrt{17.5}} \le X \le \frac{13.5-10}{\sqrt{7.5}}\right)$ = P (-0.913≤ $X \le$ 1.278) = 0.8997 – (1 – 0.8186)X = 0.7183 awrt 0.718 or 0.719	**B1** **M1 M1** **A1 A1** **M1** **A1** **(10)**
(b)	Normal approx /not Poisson since (n is large) and p close to half. **or** (np = 10 npq = 7.5) mean ≠ variance **or** np (= 10) and nq (= 30) both >5. or exact binomial = 0.7148	**B1** **B1dep** **(2)**

7. (a)(i)	f(y) $\geq$ 0 or f(3) $\geq$ 0	M1
	$ky(a-y) \geq 0$ or $3k(a-3) \geq 0$ or $(a-y) \geq 0$ or $(a-3) \geq 0$	
	$a \geq 3$	A1 cso
(ii)	$\int_0^3 k(ay-y^2)dy = 1$ — integration	M1
	$\left[k\left(\frac{ay^2}{2}-\frac{y^3}{3}\right)\right]_0^3 = 1$ — answer correct	A1
	$k\left(\frac{9a}{2}-9\right) = 1$ — answer = 1	M1
	$k\left[\frac{9a-18}{2}\right] = 1$	
	$k = \frac{2}{9(a-2)}$ *	A1 cso 6)
(b)	$\int_0^3 k(ay^2-y^3)\text{d}y = 1.75$ — Int $\int xf(x)$	M1
	Correct integration	A1
	$\left[k\left(\frac{ay^3}{3}-\frac{y^4}{4}\right)\right]_0^3 = 1.75$ — $\int xf(x) = 1.75$ and limits 0,3	M1dep
	$k\left(9a-\frac{81}{4}\right) = 1.75$	
	$2\left(9a-\frac{81}{4}\right) = 15.75(a-2)$ — subst k	M1dep
	$2.25a = -31.5 + \frac{81}{2}$	
	a = 4 *	A1cso
	$k = \frac{1}{9}$	B1 (6)
(c)		B1 B1 (2)
(d)	mode = 2	B1 (1) **[15]**

Examiner reports

Question 1

This question was answered very well with many completely accurate solutions. In part (a) some candidates neglected to specify a **continuous** Uniform distribution and in a lot of cases failed to show the graph for values of $x < 0$ and $x > 5$ with thick horizontal lines. In part (b) the mean and the variance were found correctly using the formulae. The integration method was rarely used. Most candidates were able to work out the probability in (c) using geometry but many failed to realise that for a continuous distribution the probability of an integer value was zero.

Question 2

This question was answered well by many candidates. The vast majority used F(2) = 1 and went on to find $k = 6$ and proceeded to a correct version of f(x), maybe with a few algebraic errors. Other errors included integrating between 0 and 2 which resulted in $k = \frac{16}{3}$.

Part (b) was done well by the majority of candidates and those candidates who had lost marks previously were able to differentiate for their value of k.

Some candidates lost the final mark as they failed to specify the probability density function fully or incorrectly; usually saying 0 for $y < 0$ and 1 for $y > 2$.

A few candidates when multiplying out the bracket only multiplied the first term and failed to realise that it is multiplied by all terms in the bracket.

In part (c) most candidates used 1 – F(y) but some integrated f(y) with correct limits. The main errors included candidates using f(x) without integrating or finding the wrong area, for example finding F(1) and not subtracting from 1 or finding F(2).

Question 3

Although there were a minority of candidates who were unable to identify the correct distribution to use the majority of candidates achieved full marks to parts (a) (b) and (c). Part (d) seemed to cause substantial difficulty. In part (a) the majority of candidates identified that a Poisson (rather than the Binomial) distribution was appropriate but some calculated the parameter as 2.5 or 4 rather than 0.4. A few used Po(1) and calculated P(Y=5).

In part (b) and part (c) the most common error was to use Po(2.5). The majority of candidates were able to work out P(X >1) and P(X = 2) using the correct Poisson formula. Many thought that their answer to part (c) was the correct solution while others used or multiplied their answers to both (b) and (c). Whether stating a correct or incorrect solution only a minority used the statistical term "independence" as the reason for their answer.

Question 4

Students seemed to fare better on the continuous distributions with the material spread over two questions rather than all together as previously.

In part (a) a large majority of candidates were able to set F(m) = 0.5 and formulate a correct equation. However then many candidates were unable to manipulate the initial equation into a more suitable form. However candidates who showed their working were able to earn credit for using a correct method on an incorrect equation. Candidates who solved the quadratic on their calculator showed no such method and therefore lost two marks instead of one for the incorrect answer. A variety of methods were then used successfully to solve the equation namely ‘the formula’, ‘completing the square’ and ‘trial and improvement’.

Part (b) was well done with only a few candidates neglecting to put “0 otherwise” in the full definition.

In part(c) the main problem was the result of confusion between discrete and continuous variables. It was not uncommon to see $P(X \geq 2) = 1 - F(1.1)$

Many candidates were able to gain the method mark in part (d). Those who didn’t put $(0.6267)^4$ either repeated their answer to part (c) or multiplied it by 4, seemingly unconcerned by a probability greater than 1.

Question 5

This question was well answered with many candidates gaining full marks. Part (b) required considerable attention to detail, and it is commendable that so many candidates were able to achieve all 5 marks. There were a small number of candidates who earned four of these five marks despite an incorrect answer to part (a). These candidates demonstrated the importance of showing one’s method and including all relevant details. It is often not possible to earn ‘follow through’ marks unless the method and necessary detail is made explicit. The most common errors were to muddle up p and q or forget the 6 in the calculation of P(X = 1).

Candidates generally gave clear, confident and accurate responses to part (c) that demonstrated mastery of both theory and detail. Although there were the usual errors of not using a continuity correction, using the wrong continuity correction or using the wrong mean and variance there were many fully correct solutions.

Question 6

This question was quite well answered with a high proportion of candidates clearly understanding what was required of them. Most candidates used a Poisson distribution with a mean of 10, and then most used tables correctly to get the solution. A small minority used the formula.

(a) (i) The most common error in this part was to interpret:
$(8 \leq X \leq 13)$ as $P(X \leq 13) - P(X \leq 8)$ or even $P(X \leq 13) - (1 - P(X \leq 7))$.

(ii) Many excellent responses were given to finding the Normal approximation showing good understanding of finding the variance, using continuity correction and standardising. Those who lost marks gave the variance as 10, or did not use the continuity correction correctly or did not use it at all. A high proportion of candidates had no problems in finding the correct area.

(b) This part of the question proved to be the most challenging. Conditions for both a Poisson and Normal approximation were quoted freely but in many cases, having stated that the Normal was the most appropriate, they then proceeded to state why the Poisson was not appropriate rather than why the Normal **was** appropriate. Again, a sizeable minority gave the reason for Normal as $np>5$ and $npq > 5$ rather than $np>5$ and $nq>5$.

Question 7

More candidates seemed to score full marks or nearly full marks than usual for this type of question. Some had problems with the concept of proof and some circular arguments were seen in part (b). There were also some problems in manipulating the algebraic fractions

In part (a) many candidates used a proof by contradiction approach rather than starting from $f(y) > 0$. Some wrongly thought that a probability density function cannot be 0 at any point and some thought that it can't be greater than 1. More attempted an explanation in words than a symbolic proof.

Part (b) saw many excellent solutions. There was a lot of detail involved. Yes, the equation to be solved was only linear, but the coefficients were potentially forbidding to those of us who only use a calculator as a last resort. There were many admirable responses, where candidates displayed persistence and excellent command of detail. The quantity of algebraic working seen varied substantially, from a few lines of genuine, succinct and accurate work, to a few pages of laboured inaccurate solutions.

Part (c) was surprisingly badly done. A few candidates were confused by the variable being y rather than the more usual x and so reflected their sketch in $y = x$. The minority who took their time over the sketch got it correct while those who just saw the squared term assumed a parabola intersecting the x-axis at 0 and 3. The mode was usually identified from their sketch although, as ever, there were those who gave the y value rather than the x value.

Statistics for S2 Practice Paper Silver 3

				Mean average scored by candidates achieving grade:							
Qu	**Max Score**	**Modal score**	**Mean %**	**ALL**	**A***	**A**	**B**	**C**	**D**	**E**	**U**
1	8		65.4	5.23		6.40	5.53	4.74	4.00	3.15	1.71
2	7		83.4	5.84	6.83	6.60	5.78	5.06	3.60	2.78	1.72
3	10		66.3	6.63		7.40	5.86	4.40	3.33	1.94	1.31
4	10		72.0	7.20	9.52	8.90	7.50	6.30	4.85	3.34	1.64
5	13		72.5	9.43	12.26	11.64	10.04	8.22	6.06	4.12	1.96
6	12		68.4	8.21		9.53	7.56	6.50	5.25	3.81	1.39
7	15		67.8	10.17	13.74	12.69	10.77	8.84	6.89	4.69	2.62
	75		**70.3**	**52.71**		**63.16**	**53.04**	**44.06**	**33.98**	**23.83**	**12.35**