Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2
Gold Level G1

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 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
66	56	45	35	27	19

1. (*a*) Explain what you understand by a census.

(1)

Each cooker produced at GT Engineering is stamped with a unique serial number. GT Engineering produces cookers in batches of 2000. Before selling them, they test a random sample of 5 to see what electric current overload they will take before breaking down.

(*b*) Give one reason, other than to save time and cost, why a sample is taken rather than a census.

(1)

(*c*) Suggest a suitable sampling frame from which to obtain this sample.

(1)

(*d*) Identify the sampling units.

(1)

---

2. David claims that the weather forecasts produced by local radio are no better than those achieved by tossing a fair coin and predicting rain if a head is obtained or no rain if a tail is obtained. He records the weather for 30 randomly selected days. The local radio forecast is correct on 21 of these days.

Test David's claim at the 5% level of significance.

State your hypotheses clearly.

(7)

---

3. (*a*) State two conditions under which a Poisson distribution is a suitable model to use in statistical work.

(2)

The number of cars passing an observation point in a 10 minute interval is modelled by a Poisson distribution with mean 1.

(*b*) Find the probability that in a randomly chosen 60 minute period there will be

(i) exactly 4 cars passing the observation point,

(ii) at least 5 cars passing the observation point.

(5)

The number of other vehicles, other than cars, passing the observation point in a 60 minute interval is modelled by a Poisson distribution with mean 12.

(*c*) Find the probability that exactly 1 vehicle, of any type, passes the observation point in a 10 minute period.

(4)

---

4. The continuous random variable X is uniformly distributed over the interval $[-4, 6]$.

(*a*) Write down the mean of X.

(1)

(*b*) Find $P(X \leq 2.4)$.

(2)

(*c*) Find $P(-3 < X - 5 < 3)$.

(2)

The continuous random variable Y is uniformly distributed over the interval $[a, 4a]$.

(*d*) Use integration to show that $E(Y^2) = 7a^2$.

(4)

(*e*) Find $\text{Var}(Y)$.

(2)

(*f*) Given that $P(X < \frac{8}{3}) = P(Y < \frac{8}{3})$, find the value of a.

(3)

---

5. A web server is visited on weekdays, at a rate of 7 visits per minute. In a random one minute on a Saturday the web server is visited 10 times.

(*a*) (i) Test, at the 10% level of significance, whether or not there is evidence that the rate of visits is greater on a Saturday than on weekdays. State your hypotheses clearly.

(ii) State the minimum number of visits required to obtain a significant result.

(7)

(*b*) State an assumption that has been made about the visits to the server.

(1)

In a random two minute period on a Saturday the web server is visited 20 times.

(*c*) Using a suitable approximation, test at the 10% level of significance, whether or not the rate of visits is greater on a Saturday.

(6)

---

6. A bag contains a large number of coins. It contains only 1p and 2p coins in the ratio 1:3.

(*a*) Find the mean μ and the variance σ^2 of the values of this population of coins.

(3)

A random sample of size 3 is taken from the bag.

(*b*) List all the possible samples.

(2)

(*c*) Find the sampling distribution of the mean value of the samples.

(6)

---

7. The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{6}x & 0 < x \leq 3 \\ 2 - \frac{1}{2}x & 3 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

(*a*) Sketch the probability density function of X.
(3)

(*b*) Find the mode of X.
(1)

(*c*) Specify fully the cumulative distribution function of X.
(7)

(*d*) Using your answer to part (*c*), find the median of X.
(3)

TOTAL FOR PAPER: 75 MARKS

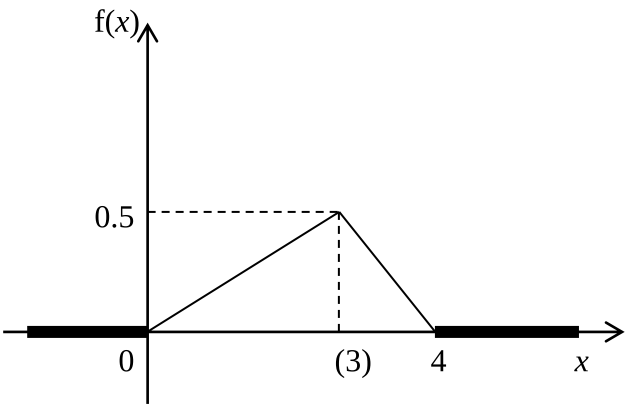
END

Question Number	Scheme	Marks
1.		
(a)	A census is when <u>every member</u> of the <u>population</u> is investigated.	**B1**
(b)	There would be no cookers left to sell.	**B1**
(c)	A list of the unique identification numbers of the cookers.	**B1**
(d)	A cooker	**B1**
		(4)
2.	$H_0 : p = 0.5$	**B1**
	$H_1 : p > 0.5$	**B1**
	X~B(30,0.5) Using correct Bin	**M1**
	P($X \geq$ 21) = 1 - P($X \leq$20) or P($X \leq$19) = 0.9506 P($X \geq$20) = 0.0494	**M1**
	= 1 - 0.9786	
	= 0.0214 CR $X \geq 20$	**A1**
	so significant/reject H_0 /in Critical region	**M1 dep**
	Evidence to suggest **<u>David's claim is incorrect</u>** or The weather **<u>forecast</u>** produced by the local **<u>radio</u>** is better than those achieved by **<u>tossing/flipping a coin</u>**	**A1** **(7)**
		(7 marks)
3.		
(a)	<u>Events</u> occur at a constant rate. any two of the 3 <u>Events</u> occur independently or randomly. <u>Events</u> occur singly.	**B1 B1** **(2)**
(b)	Let X be the random variable the number of cars passing the observation point.	
(i)	Po(6)	**B1**
	P(X ≤ 4) – P(X ≤ 3) = 0.2851 – 0.1512 or $\frac{e^{-6}6^4}{4!}$	**M1**
	= 0.1339	**A1**
(ii)	1 – P(X ≤ 4) = 1 – 0.2851 or $1-e^{-6}\left(\frac{6^4}{4!}+\frac{6^3}{3!}+\frac{6^2}{2!}+\frac{6^4}{1!}+1\right)$	**M1**
	= 0.7149	**A1** **(5)**
(c)	P (0 car and 1 others) + P (1 cars and 0 other)	**B1**
	= $e^{-1} \times 2e^{-2} + 1e^{-1} \times e^{-2}$ = 0.3679 × 0.2707 + 0.3674 × 0.1353 = 0.0996 + 0.0498 = 0.149	**M1 A1** **A1** **(4)**

Question Number	Scheme	Marks
4. (a)	Mean = 1	B1 **(1)**
(b)	$\mathrm{P}(X \le 2.4) = \left(2.4 - -4\right) \times \frac{1}{10}$	M1
	$= 0.64$ or $\frac{16}{25}$	A1
		(2)
(c)	$\mathrm{P}(-3 < X - 5 < 3) = \mathrm{P}(2 < X < 6)$	M1
	$= 0.4$	A1
		(2)
(d)	$\int_a^{4a} \frac{y^2}{4a - a} \mathrm{d}y = \left[\frac{y^3}{9a}\right]_a^{4a}$	M1 M1 dep
	$= \frac{64a^3 - a^3}{9a}$	A1
	$= 7a^2$ *AG	A1cso
		(4)
(e)	$\mathrm{Var}(Y) = \frac{1}{12}(4a - a)^2$ or $\mathrm{Var}(Y) = 7a^2 - \left(\frac{5}{2}a\right)^2$	M1
	$= \frac{3}{4}a^2$	A1cso
		(2)
(f)	$\frac{2}{3} = \frac{1}{3a}\left(\frac{8}{3} - a\right)$	M1 A1
	$a = \frac{8}{9}$	A1
		(3)
		Total 14

Question Number	Scheme	Marks
5. (a) (i)	$H_0 : \lambda = 7$ $\quad$ $H_1 : \lambda > 7$	B1
	X = number of visits. $X \sim Po(7)$	B1
	$P(X \geq 10) = 1 - P(X \leq 9)$ $\quad$ $1 - P(X \leq 10) = 0.0985$	M1
	$= 0.1695$ $\quad$ $1 - P(X \leq 9) = 0.1695$	
	CR $X \geq 11$	A1
	$0.1695 > 0.10$, $\quad$ CR $X \geq 11$	
	Not significant or it is not in the critical region or do not reject H_0	M1
	The rate of visits on a Saturday is not greater/ is unchanged	A1 no ft
(ii)	$X = 11$	B1
		(7)
(b)	(The visits occur) randomly/ independently or singly or constant rate	B1
		(1)
(c)	$[H_0 : \lambda = 7$ $\quad$ $H_1 : \lambda > 7$ $\quad$ (or $H_0 : \lambda = 14$ $\quad$ $H_1 : \lambda > 14)]$	
	X ~N;(14,14)	B1;B1
	$P(X \geq 20) = P\left(z \geq \frac{19.5-14}{\sqrt{14}}\right)$ $\quad$ +/- 0.5, stand	M1 M1
	$= P(z \geq 1.47)$	
	$= 0.0708$ $\quad$ or $z = 1.2816$	A1dep both M
	$0.0708 < 0.10$ therefore significant. The rate of visits is greater on a Saturday	A1dep 2nd M
		(6)

Question Number	Scheme	Marks
6. (a)	<table><tr><td>x</td><td>1p</td><td>2p</td></tr><tr><td>$\mathrm{P}(X=x)$</td><td>$\frac{1}{4}$</td><td>$\frac{3}{4}$</td></tr></table>	
	$\mu = 1\times\frac{1}{4}+2\times\frac{3}{4}=\frac{7}{4}$ or $1\frac{3}{4}$ or 1.75	B1
	$\sigma^2 \quad = 1^2\times\frac{1}{4}+2^2\times\frac{3}{4}-\left(\frac{7}{4}\right)^2$	M1
	$=\frac{3}{16}$ or 0.1875	A1
		(3)
(b)	(1,1,1) , (1,1,2) any order, (1,2,2) any order, (2,2,2)	B1
	(1,2,1) (2,1,1) (2,1,2) (2,2,1) all 8 cases considered. May be implied by 3 * (1,1,2) and 3*(1,2,2)	B1
		(2)
(c)	<table><tr><td>$\bar{x}$</td><td>1</td><td>$\frac{4}{3}$</td><td>$\frac{5}{3}$</td><td>2</td></tr><tr><td>$\mathrm{P}(\bar{X}=\bar{x})$</td><td>$\frac{1}{4}\times\frac{1}{4}\times\frac{1}{4}=\frac{1}{64}$</td><td>$3\times\frac{1}{4}\times\frac{1}{4}\times\frac{3}{4}=\frac{9}{64}$</td><td>$3\times\frac{1}{4}\times\frac{3}{4}\times\frac{3}{4}=\frac{27}{64}$</td><td>$\frac{3}{4}\times\frac{3}{4}\times\frac{3}{4}=\frac{27}{64}$</td></tr></table>	B1 M1 A1 M1 A1A1
		(6)
		Total [11]

Question Number	Scheme		Marks
7 (a)		(0), 4, 0.5 0 may be implied by start at y axis	B1
		both patio must be straight	B1 B1
			(3)
(b)	Mode is $x = 3$		B1 (1)
(c)	$F(x) = \int_0^x \frac{1}{6}t\ dt$ (for $0 \le x \le 3$)	ignore limits for M	M1
	$= \frac{1}{12}x^2$	must use limit of 0	A1
	$F(x) = \int_3^x 2 - \frac{1}{2}t\ dt; + \int_0^3 \frac{1}{6}t\ dt$ (for $3 < x \le 4$)	need limit of 3 and variable upper limit; need limit 0 and 3	M1; M1
	$= 2x - \frac{1}{4}x^2 - 3$		A1
	$F(x)\begin{cases} 0 & x < 0 \\ \frac{1}{12}x^2 & 0 \le x \le 3 \\ 2x - \frac{1}{4}x^2 - 3 & 3 < x \le 4 \\ 1 & x > 4 \end{cases}$	middle pair ends	B1 ft B1
			(7)
(d)	F(m) = 0.5	either eq	M1
	$\frac{1}{12}x^2 = 0.5$	eq for their $0 \le x \le 3$	A1ft
	$x = \sqrt{6} = 2.45$	$\sqrt{6}$ or awrt 2.45	A1
			(3) Total 14

Examiner reports

Question 1

Nearly all candidates achieved at least one of the available marks but it was disappointing that there were not more attaining full marks.

(a) Too many candidates referred to the national census rather than a general definition. Some felt an enumeration was adequate and others failed to recognise that EVERY member had to be investigated.

(b) A failure to put the question in context and consider the consequences of testing every item meant that some candidates scored 0 in this part of the question. A few candidates did not read the question carefully and used cheap and quick as their reasons why a census should not be used when the question specifically said give a reason "other than to save time and cost".

(c) Many candidates mentioned a list; database or register and so attained the available mark. However, some did not seem to differentiate between the population and the sampling frame.

(d) Most candidates were able to identify the sampling units correctly, although those who had not scored in part (c) tended to say: "the sample of 5 cookers" in part (d).

Question 2

Responses to this question generally reflected candidates' understanding with a high percentage gaining 5 of the 7 available marks. A common error in giving the hypotheses was to write the alternative hypothesis as $p < 0.5$ or $p \neq 0.5$. Occasionally letters other than 'p' were used. The majority of candidates used the correct Binomial and successfully calculated the probability 0.0214. The most common error was to use the incorrect statement $\mathrm{P}(X \geq 21) = 1 - \mathrm{P}(X \leq 21)$ or calculate $\mathrm{P}(X = 21)$. It is pleasing to see that a higher proportion of candidates are able to use statistical tables accurately.

Having calculated a correct probability of 0.0214 and stated that the null hypothesis should be rejected, a number of candidates showed that they had not read the question carefully by stating incorrectly that 'David's claim is supported/correct' Many candidates do not use the full context of the question in their final statement. For the final statement, if the statement saying 'David is not correct' was not used, the candidates needed to include the words 'forecast', 'radio' and 'tossing/flipping a coin'.

Question 3

Most candidates were able to attempt part (b) successfully as these were fairly standard calculations. However, when required to apply Poisson probabilities to a problem it was only the better candidates who attained any marks in part (c).

(a) A sizeable proportion of candidates, whilst having learnt the conditions for a Poisson distribution, failed to realise that this applied to the <u>events</u> occurring. There were references to 'trials' and 'things' in some solutions offered.

b(i) Most candidates recognised Po(6) and were able to answer this successfully, either from the tables or by calculation. Common errors were using incorrect values from the table or calculating the exact value incorrectly.

(ii) Although many candidates attained full marks for this part, some were unable to express ‘at least 5’ correctly as an inequality and used $P(X \leq 5)$.

(b) Many of the successful candidates used a Po(3) to give a correct solution. Of the rest most candidates failed to realise that there were two ways for exactly one vehicle to pass the point and so only performed one calculation. This was often for P(1 car) and P(1 other vehicle), which were then added together. However there were some candidates who gained the correct answer via this method.

Question 4

There were many exemplary responses to this question with candidates getting full marks. Errors in parts (d) and (e) often reflected some candidates' lack of competence in the manipulation of algebraic fractions.

Errors in part (a) were down to a few candidates giving the mean as 5 or 2.

Responses to part (b) showed that the majority of candidates were able to find the required probability using a continuous uniform distribution over the given range. Some candidates found $P(X > 2.4)$ and did not proceed to find the correct solution using $1 – P(X > 2.4)$. Also, a small number of candidates tried to find the probability using a discrete uniform distribution.

Part (c) challenged many candidates and the ability to interpret $P(–3 < X – 5 < 3)$ was varied. An answer of 0.6 was often seen which was found from working such as $P(–3< X <3)$, ignoring $X – 5$, or from $P(2 < X < 8)$.

Many excellent responses to part (d) were seen but equally it was evident that this question proved very challenging for a high proportion of candidates by the number of attempts (including pages of crossed-out working) at finding the required solution. The majority of candidates followed the instruction to ‘**use integration** to show...’ and gained at least one mark for writing $\int \frac{y^2}{4a-a}\,dy$ or the equivalent. Candidates who used a as the variable were more likely to lose marks through errors made in treating the constant $3a$ in the denominator as a variable and cancelling this with the expression in the numerator, either before or after integration. This, in turn, created problems for candidates when substituting the values for the limits of integration. Working was also seen on a number of occasions where candidates lost the final two marks when substituting the limits and calculating $(4a)^3 = 4a^3$.

Candidates who used $E(Y^2)$ $E(Y)^2$ did not score any marks for this part of the question. Part (e) was also generally accessible to the majority of candidates. Candidates who used $\frac{(b-a)^2}{12}$ were less likely to lose marks through errors than those who used $E(Y^2)E(Y)^2$. A common error was where candidates found the value of the mean but then forgot to square it before subtracting from $E(Y^2)$.

In part (f) a high proportion of candidates successfully found $P\left(X < \frac{8}{3}\right) = \frac{2}{3}$ and solved $P\left(Y < \frac{8}{3}\right) = \frac{2}{3}$ to get an answer $a = \square\frac{8}{9}$. For a large minority of candidates finding $P\left(Y < \frac{8}{3}\right)$ proved difficult, and in some cases impossible. This was evident in the number of attempts shown. A number of candidates lost the final mark through errors in handling algebraic fractions, e.g it was common to see working $\frac{8}{9a}-\frac{1}{3}=\frac{2}{3}$ followed by $\frac{8}{9a}=1$ or $\frac{8}{9}a=1$, giving a final answer of $\frac{9}{8}$

Question 5

In Part (a) there are a sizeable number of candidates who are not using the correct symbols in defining their hypotheses although the majority of candidates recognised Po(7).

For candidates who attempted a critical region there were still a number who struggled to define it correctly for a number of reasons:

- Looking at the wrong tail and concluding $X \leq 3$.
- Incorrect use of > sign when concluding 11 - not appreciating that this means ≥12 for a discrete variable.
- Not knowing how to use probabilities to define the region correctly and concluding 10 or 12 instead of 11.

The candidates who opted to calculate the probability were generally more successful.

A minority still try to calculate a probability to compare with 0.9. This proved to be the most difficult route with the majority of students unable to calculate the probability or critical region correctly. We must once again advise that this is not the recommended way to do this question.

There are still a significant number who failed to give an answer in context although fewer than in previous sessions.

Giving the minimum number of visits needed to obtain a significant result proved challenging to some and it was noticeable that many did not use their working from part (a) or see the connection between the answer for (i) and (ii) and there were also number of candidates who did not recognise inconsistencies in their answers.

A number of candidates simply missed answering part (b) but those who did usually scored well.

There were many excellent responses in part (c) with a high proportion of candidates showing competence in using a Normal approximation, finding the mean and variance and realising that a continuity correction was needed. Marks were lost, however, for not including 20, and for not writing the conclusion in context in terms of the **rate** of visits being **greater**. Some candidates attempted to find a critical value for X using methods from S3 but failing to use 1.2816.

There were a number of candidates who calculated P(X=20) in error.

Question 6

A high proportion of candidates attempted the first two parts of this question successfully, with the majority of candidates getting at least one mark for part (b). Those less successful in part (a) either misread the question and ended up with a denominator of 3 for the probabilities or confused formulae for calculating the mean and variance and used, for example, $\sum \frac{xp(x)}{n}$ for the mean or used $E(X^2)$ for σ^2. The solution to part (c) proved beyond the capability of a minority of candidates but, for the majority, many exemplary answers were evident, reflecting sound preparation on this topic. Candidates who found all 8 cases in (b) usually gained four marks in part (c) for calculating the probabilities. For a small percentage of those candidates, calculating the means was difficult and hence completing the table correctly was not possible. A few candidates tried unsuccessfully to use the binomial to answer part (c).

Question 7

This question has been a good discriminator. The majority of candidates attempted this question with varying degrees of success. In part (a) there were many good sketches with clear labelling but many lost a mark through not marking the patios clearly. In part (b) the most common mistake is to give the value of f(x) i.e. ½. Part (c) was a problem for a majority of candidates. It was evident that many candidates were not competent in finding the CDF of a function given in two parts. Finding F(x) for $0 \le x \le 3$ was reasonably well answered, but quite often candidates did not use the limits or simply wrote down the answers without showing any working.

Candidates were less successful in finding F(x) for $3 \le x \le 4$, with few using limits correctly and many not taking into consideration the answer to the first part. Candidates who used the alternative method, using '+c' and F(0) = 0 and F(4) = 1 were generally more successful in getting the correct F(x). Responses to part (d) would seem to reflect a lack of understanding of what the median is. Candidates quoted F(x) = 0.5 and the proceeded to put F(x) for $3 \le x \le 4 = 0.5$ and solve. It was rare to find evidence of candidates checking which part to use before setting up an equation. Many candidates solved F(x) = 0.5 for both parts and then not said which answer was the median. Another common error was adding F(x) for $0 \le x \le 3$ and F(x) for $3 \le x \le 4$ and then solving.

Statistics for S2 Practice Paper Gold 1

Qu	Max Score	Modal score	Mean %	ALL	A*	A	B	C	D	E	U
				Mean average scored by candidates achieving grade:							
1	4		63.0	2.52		2.80	2.12	1.42	1.30	0.86	0.36
2	7		68.7	4.81	5.63	5.23	4.04	3.18	2.62	2.30	0.58
3	11		66.0	7.26		7.34	6.38	5.45	4.68	4.04	2.18
4	14	14	67.0	9.42	12.47	11.36	8.54	6.72	4.69	3.44	2.14
6	14		68.1	9.53		11.64	8.37	6.40	3.65	2.65	0.53
7	11		63.2	6.95		8.33	5.11	3.55	2.71	2.06	0.27
8	14		65.4	9.16		11.63	9.77	8.10	6.65	5.00	2.43
	75		**66.2**	**49.65**		**58.33**	**44.33**	**34.82**	**26.30**	**20.35**	**8.49**