Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2
Silver Level S4

Time: 1 hour 30 minutes

Materials required for examination papers
Mathematical Formulae (Green)

Items included with question papers
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
68	59	50	41	33	25

1. Jean regularly takes a break from work to go to the post office. The amount of time Jean waits in the queue to be served at the post office has a continuous uniform distribution between 0 and 10 minutes.

(*a*) Find the mean and variance of the time Jean spends in the post office queue.

(3)

(*b*) Find the probability that Jean does not have to wait more than 2 minutes.

(2)

Jean visits the post office 5 times.

(*c*) Find the probability that she never has to wait more than 2 minutes.

(2)

Jean is in the queue when she receives a message that she must return to work for an urgent meeting. She can only wait in the queue for a further 3 minutes.

Given that Jean has already been queuing for 5 minutes,

(*d*) find the probability that she must leave the post office queue without being served.

(3)

---

2. A traffic officer monitors the rate at which vehicles pass a fixed point on a motorway. When the rate exceeds 36 vehicles per minute he must switch on some speed restrictions to improve traffic flow.

(*a*) Suggest a suitable model to describe the number of vehicles passing the fixed point in a 15 s interval.

(1)

The traffic officer records 12 vehicles passing the fixed point in a 15 s interval.

(*b*) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not the traffic officer has sufficient evidence to switch on the speed restrictions.

(6)

(*c*) Using a 5% level of significance, determine the smallest number of vehicles the traffic officer must observe in a 10 s interval in order to have sufficient evidence to switch on the speed restrictions.

(3)

---

3. A single observation x is to be taken from a Binomial distribution $B(20, p)$.

This observation is used to test $H_0 : p = 0.3$ against $H_1 : p \neq 0.3$.

(*a*) Using a 5% level of significance, find the critical region for this test. The probability of rejecting either tail should be as close as possible to 2.5%.

(3)

(*b*) State the actual significance level of this test.

(2)

The actual value of x obtained is 3.

(*c*) State a conclusion that can be drawn based on this value, giving a reason for your answer.

(2)

---

4. A continuous random variable X is uniformly distributed over the interval $[b, 4b]$ where b is a constant.

(*a*) Write down $E(X)$.

(1)

(*b*) Use integration to show that $\text{Var}(X) = \dfrac{3b^2}{4}$.

(3)

(*c*) Find $\text{Var}(3 - 2X)$.

(2)

Given that $b = 1$, find

(*d*) the cumulative distribution function of X, $F(x)$, for all values of x,

(2)

(*e*) the median of X.

(1)

---

5. The continuous random variable X has a cumulative distribution function

$$\mathrm{F}(x) = \begin{cases} 0, & x < 1, \\ \dfrac{x^3}{10} + \dfrac{3x^2}{10} + ax + b, & 1 \le x \le 2, \\ 1, & x > 2, \end{cases}$$

where a and b are constants.

(*a*) Find the value of a and the value of b.

(4)

(*b*) Show that $\mathrm{f}(x) = \dfrac{3}{10}(x^2 + 2x - 2), \quad 1 \le x \le 2.$

(1)

(*c*) Use integration to find $\mathrm{E}(X)$.

(4)

(*d*) Show that the lower quartile of X lies between 1.425 and 1.435.

(3)

---

6. (*a*) Explain what you understand by a hypothesis.

(1)

(*b*) Explain what you understand by a critical region.

(2)

Mrs George claims that 45% of voters would vote for her.

In an opinion poll of 20 randomly selected voters it was found that 5 would vote for her.

(*c*) Test at the 5% level of significance whether or not the opinion poll provides evidence to support Mrs George's claim.

(4)

In a second opinion poll of n randomly selected people it was found that no one would vote for Mrs George.

(*d*) Using a 1% level of significance, find the smallest value of n for which the hypothesis $\mathrm{H}_0 : p = 0.45$ will be rejected in favour of $\mathrm{H}_1 : p < 0.45$.

(3)

---

7. As part of a selection procedure for a company, applicants have to answer all 20 questions of a multiple choice test. If an applicant chooses answers at random the probability of choosing a correct answer is 0.2 and the number of correct answers is represented by the random variable X.

(*a*) Suggest a suitable distribution for X.

(2)

Each applicant gains 4 points for each correct answer but loses 1 point for each incorrect answer. The random variable S represents the final score, in points, for an applicant who chooses answers to this test at random.

(*b*) Show that $S = 5X - 20$.

(2)

(*c*) Find $\mathrm{E}(S)$ and $\mathrm{Var}(S)$.

(4)

An applicant who achieves a score of at least 20 points is invited to take part in the final stage of the selection process.

(*d*) Find $\mathrm{P}(S \geq 20)$.

(4)

Cameron is taking the final stage of the selection process which is a multiple choice test consisting of 100 questions. He has been preparing for this test and believes that his chance of answering each question correctly is 0.4.

(*e*) Using a suitable approximation, estimate the probability that Cameron answers more than half of the questions correctly.

(5)

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)	E(X) = 5	B1
	Var(X) = $\frac{1}{12}(10-0)^2$ or attempt to use $\int \frac{x^2}{10}dx - \mu^2$	M1
	$= \frac{100}{12} = \frac{25}{3} = 8\frac{1}{3} = 8.\dot{3}$ awrt 8.33	A1 (3)
(b)	P($X \le 2$) = $(2-0)\times\frac{1}{10} = \frac{1}{5}$ or $\frac{2}{10}$ or 0.2	M1 A1 (2)
(c)	$\left(\frac{1}{5}\right)^5$ = 0.00032 or $\frac{1}{3125}$ or 3.2×10^{-4} o.e.	M1 A1 (2)
(d)	P(X $\ge$ 8) or P(X > 8)	M1
	P($X \ge 8 \mid X \ge 5$) = $\frac{P(X \ge 8)}{P(X \ge 5)}$	M1
	$= \frac{2/10}{5/10}$	
	$= \frac{2}{5}$	A1 (3)
		(10 marks)

2. (a)	Poisson	B1 (1)
(b)	$H_0 : \mu = 9$ (or $\lambda = 36$) $H_1 : \mu > 9$ (or $\lambda > 36$)	B1 B1
	X ~Po(9) and P($X \ge 12$) = 1 - P($X \le 11$) or P($X \le 14$) = 0.9585, P($X \ge 15$) = 0.0415	M1
	= 1-0.8030 = <u>0.197</u> <u>CR X $\ge$ 15</u>	A1
	(0.197 > 0.05)	
	so not significant/ accept H_0/ Not in CR	M1d
	he does not have evidence to switch on the <u>speed restrictions</u> (o.e)	A1ft (6)
(c)	Let Y = the number of vehicles in 10 s then Y ~Po(6)	B1
	Tables: P($Y \le 10$) = 0.9574 so P($Y \ge 11$) = 0.0426	M1
	so needs <u>11</u> vehicles	A1 (3)
		10

Question Number	Scheme	Marks
3 (a)	$X \sim B(20, 0.3)$	M1
	$P(X \le 2) = 0.0355$	
	$P(X \ge 11) = 1 - 0.9829 = 0.0171$	
	Critical region is $(X \le 2) \cup (X \ge 11)$	A1 A1 (3)
(b)	Significance level = 0.0355 + 0.0171, = 0.0526 or 5.26%	M1 A1 (2)
(c)	Insufficient evidence to reject H_0 **Or** sufficient evidence to accept H_0 /not significant	B1 ft
	$x = 3$ (or the value) is not in the critical region or 0.1071> 0.025	B1 ft (2)
	Do not allow inconsistent comments	

4(a)	$E(X) = \frac{5b}{2}$	B1 (1)
(b)	$Var(X) = E(X^2) - (E(X))^2$	
	$= \int_b^{4b} \frac{x^2}{3b} dx - (5b/2)^2$	M1
	$= \left[\frac{x^3}{9b}\right]_b^{4b} - \frac{25b^2}{4}$	M1d
	$= \frac{63b^3}{9b} - \frac{25b^2}{4}$	
	$= \frac{3b^2}{4}$	A1cso (3)
(c)	$Var(3 - 2X) = 4Var(X)$	M1
	$= 3b^2$	A1 (2)
(d)	$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{3} & 1 \le x \le 4 \\ 1 & x > 4 \end{cases}$	B1B1 (2)
(e)	$\frac{x-1}{3} = 0.5$ so $x = 2.5$	B1 (1) **[9]**

Question Number	Scheme	Marks
5(a)	F(1) = 0, $\frac{4}{10}+a+b=0$	M1
	$a=-\frac{3}{5}$ or $b=\frac{1}{5}$	A1
	F(2) = 1, $2+2a+b=1$	M1
	Solving gives $a=-\frac{3}{5}, b=\frac{1}{5}$	A1
	Alt	(4)
	F(2) - F(1) = 1, $2+2a+b-\frac{4}{10}-a-b=1$	M1
	$a=-\frac{3}{5}$	A1
	F(2) =1 **or** F(1) = 0	
	$2-\frac{6}{5}+b=1$ **or** $\frac{4}{10}-\frac{3}{5}+b=0$	M1
	$b=\frac{1}{5}$	A1
		(4)
(b)	Differentiating cdf gives f(x) = $\frac{3}{10}x^2+\frac{6}{10}x+a$, $1\le x\le 2$	
	$=\frac{3}{10}(x^2+2x-2)$	B1 cso
		(1)
(c)	$\mathrm{E}(X)=\int_1^2 \frac{3}{10}(x^3+2x^2-2x)\mathrm{d}x$	M1
	$=\frac{3}{10}\left[\frac{1}{4}x^4+\frac{2}{3}x^3-x^2\right]_1^2$	M1d A1
	$=\frac{13}{8}$	A1
		(4)
(d)	F(1.425) = 0.24355, F(1.435) = 0.25227	M1A1
	0.25 lies between F(1.425)and F(1.435) hence result.	A1
		(3)
		[12]

Question Number	Scheme	Marks
6(a)	A statement concerning a **population parameter**	B1
(b)	A critical region is the range / set of values / answers **or** a test statistic **or** region/area **or** values (where the test is significant)	B1
	that would lead to the rejection of H0 / acceptance of H_1	B1
		(3)
(c)	$H_0 : p = 0.45$ $H_1 : p < 0.45$ (or $p \neq 0.45$)	
	$X \sim B(20, 0.45)$	M1
	$P(X \leq 5) = 0.0553$ CR $X \leq 4$	A1
	Accept H_0. Not significant. 5 does not lie in the Critical region.	M1d
	There is no evidence that the proportion who voted for Mrs George is not 45% or there is evidence to support Mrs George's claim	A1cso
		(4)
(d)	B(8, 0.45): P(0) = 0.0084	M1
	B(7, 0.45): P(0) = 0.0152	A1
	Hence smallest value of n is **8**	B1
		(3)
	Alternative	
	$(0.55)^n < 0.01$	M1
	$n\log 0.55 < \log 0.01$	
	$n > 7.7\ldots$	A1
	Hence smallest value of n is **8**	B1cso
		Total 10

Question Number	Scheme	Marks
7.		
(a)	$X \sim \text{B}(20, 0.2)$	M1 A1
		(2)
(b)	$S = 4X - 1(20 - X)$	M1
	$S = 5X - 20$	A1cso
		(2)
(c)	$\text{E}(X) = 4, \quad \text{Var}(X) = 3.2$	B1, B1
	$\text{E}(S) = 5\times4-20=0, \quad \text{Var}(S) = 5^2\,\text{Var}(X) = 80$	M1 A1
		(4)
(d)	$S \geq 20$ implies $5X - 20 \geq 20$	M1
	[So $5X \geq 40$] $\quad X \geq 8$	A1
	$\text{P}(S \geq 20) = \text{P}(X \geq 8) = 1 - \text{P}(X \leq 7)$	M1
	$= 1 - 0.9679 =$ **0.0321**	A1
		(4)
(e)	[Let C = no. Cameron gets correct. $C \sim \text{B}(100, 0.4)$] $\quad Y \sim \text{N}(40, \sqrt{24}^{\,2})$	M1A1
	$\text{P}(C > 50) \simeq \text{P}(Y > 50.5)$	
	$= \text{P}\left(Z > \dfrac{50.5-40}{\sqrt{24}}\right)$	M1 M1
	$= \text{P}(Z > 2.14...) = 1 - 0.9838 = 0.0162$ or $0.016044..$ (awrt **0.016**)	A1
	N.B. exact Bin (0.01676...) Poisson approx (0.0526...)	**(5)**
		[17]

Examiner reports

Question 1

Part (a) was answered well. A minority of candidates had difficulty with parts (b) and (c). In (b) they generally looked at $P(X > 2)$ rather than $P(X \leq 2)$.

In part (c) $\frac{1}{5}$ x 5 or $\left(\frac{4}{5}\right)^5$ was used even when part (b) was correct.

The work on part (d) was disappointing. Candidates did not understand the concept of conditional probability.

Only the very best candidates are likely to have got this correct. Those that did used the $U[0,5]$ or $U\ [5,10]$ route.

Question 2

There was a very variable response to question 2, with many candidates producing "textbook answers", whilst many others failing to recognise a Poisson distribution in part (a), offered either (or sometimes both) a binomial or a normal model.

The latter candidates either stopped at part (a) or pursued their chosen model to little effect.

In part (b) the vast majority successfully opted for a 1-tailed alternative hypothesis, although some did insist on using the parameter p. The value of $P(X \geq 12)$ or the CR was usually found correctly and most candidates were able to make a successful comparison, thereby leading to a well expressed contextual conclusion. Some candidates whose alternative hypothesis suggested a 2-tailed test, still opted to perform a 1-tailed test.

Question 3

Part (a) of this question was poorly done. Candidates would appear unfamiliar with the standard mathematical notation for a Critical Region. Thus $11 \leq X \leq 2$ made its usual appearances, along with $c_1 = 2$ and $P(X \leq 2)$.

In part (b) candidates knew what was expected of them although many with incorrect critical regions were happy to give a probability greater than 1 for the critical region.

Part (c) was well answered. A few candidates did contradict themselves by saying it was "significant" and "there is no evidence to reject H_0" so losing the first mark.

Question 4

In part (a) a large majority of candidates were able to obtain E(X) = $\frac{5b}{2}$, although E(X) = $\frac{4b-b}{2}$ = $\frac{3b}{2}$ was occasionally seen.

Despite the clear instruction in part (b) to 'use integration' to show that Var(X) = $\frac{3b^2}{4}$, a significant minority quoted and used the formula $\frac{(b-a)^2}{12}$, scoring no marks.

Of those who did attempt the integration, there was a mixed response with some candidates providing a perfect solution, whilst others contained at least one of the following common errors: an incorrect expression for f(x) was used; sometimes $[\mathrm{E}(X)]^2$ was not subtracted from $\mathrm{E}(X^2)$; b was sometimes used for the variable of integration as well as a constant. This was condoned unless candidates cancelled $\int_b^{4b} \frac{b^2}{3b}\mathrm{d}x$ to $\int_b^{4b} \frac{b}{3}\mathrm{d}x$ prior to integration, which led to a forfeiture of all the 3 available marks.

Deducing the value of Var (3 – 2X) is a routine calculation on S1, but it proved too demanding for a significant minority of candidates in part (c).

In part (d) the distribution function was correctly obtained by a majority of candidates, with the two most common errors being: forgetting to subtract F(1), yielding $\frac{x}{3}$ on the middle line; carelessness with the inequalities on the 1st and/or 3rd lines, e.g. $x < 0$ or $x > 1$, $x > 5$.

The median was generally given correctly as $\frac{5b}{2}$, usually using the distribution function obtained in part (b). Those who obtained various incorrect answers may have benefited from an appeal to the symmetry of the continuous uniform distribution.

Question 5

Again, many exemplary responses to this question with a high percentage of candidates gaining full marks.

In finding values for a and b, candidates used a number of different methods, e.g. finding two linear equations, F(1) = 0 and F(2) = 1 and solving, or using F(2) – F(1) = 1 which gave the value of a and candidates then used F(1) = 0 and F(2) = 1 to find the value of b. Less successful candidates put the value of a in F(2) – F(1) = 1 again and often came up with a value b = 0 or 1.

Candidates who struggled to find values for a and b sometimes used an alternative method i.e. $\int_1^x \frac{3}{10}(x^2+2x-2)$, which they saw in part (b) to get $\frac{x^3}{10}$ + $\frac{3x^2}{10} - \frac{3x}{5}$ + $\frac{1}{5}$ and used the coefficient of x to give the value for a and the constant for b. No marks were awarded in such cases as the equations but candidates were not penalised for using these values in other parts of the question

Part (b) was generally well answered. Candidates who were able to obtain values for a and b in part (a) often used the expression given for f(x) and 'derived' a value for a = -0.6 which they then used in F(x) before differentiating and then factorising successfully.

A high percentage of candidates answered part (c) well. The majority of candidates used xf(x), with the f(x) given in part (b), and successfully integrated to get the correct answer. A small percentage of candidates lost marks through integrating f(x) or forgetting to multiply through by $\frac{3}{10}$ to get the final answer.

In part (d), the majority of candidates used the method of finding values for F(1.425) and F(1.435), making reference to $F(Q_1) = 0.25$ and making a statement that '0.25 lies between F(1.425) and F(1.435)'. Candidates were less successful in finding the value of the lower quartile when solving F(x) = 0.25 as the majority were unable to solve the cubic equation found. Marks were also lost for quoting incorrect values for F(1.425) and/or F(1.435) without showing substitution. Common incorrect final statements given were 'the lower quartile lies between F(1.425) and F(1.435)' or 'the lower quartile lies between these values' without stating which values. A small number of candidates referred to the median.

Question 6

Apart from part (a), this question was well answered and responses reflected good preparation and understanding of hypothesis tests.

Only a few candidates were able to discuss a hypothesis in terms of a population parameter in part(a). Candidates' errors included discussing hypotheses and critical values. Part(b) was generally well answered although some candidates wrote comments such as 'it is an area where a hypothesis could be rejected' without identifying which of the two hypotheses was being referred to. References to 'original' and 'new' hypotheses rather than H_0 and H_1 were also seen. Correct terminology is important.

In part(c) many candidates achieved full marks but there were some common errors. These included absence or incorrectly stated hypotheses, finding $P(X = 5)$ rather than $P(X \leq 5)$, comparing 0.553 to 0.05 if using a two-tailed test or with 0.25 when using a one-tailed test, conflicting non-contextual conclusions and incorrect (or no) contextual conclusion, some including double negatives.

Part (d) of the question was tackled well by a large number of candidates, with some able to complete it accurately with little or no working. Those that investigated $P(X = 0)$ for various values of n nearly always attained a correct solution and most of those using logs also gained full marks, the main error being forgetting to reverse the inequality sign when dividing by log 0.55.

Question 7

This question allowed candidates to score some marks and a variety of solutions were seen. Only the best candidates were able to score full marks and part (b) caused the major problem.

Part (a) was done well by the majority of candidates and errors included either leaving out the 20 and 0.2 or stating a Poisson instead of binomial.

Part (b) was either done well or poorly. A few candidates left this part out whilst others showed that $S = 5X$-20 for a particular value and not for the general case.

In part (c) there was some confusion between X and S. Some candidates could only state $\mathrm{E}(X) = 4$ and $\mathrm{Var}(X) = 3.2$. Others assumed S to be uniform.

Those candidates who were able to answer the question often scored full marks but the common error for these candidates was to incorrectly calculate Var (S) (often giving an answer of – 0.4).

In part (d) many candidates realised that they had to solve an inequality for X and usually scored full marks, with only a minority making the error of using $\mathrm{P}(X \geq 8)$ as $1 - \mathrm{P}(X \leq 8)$. Candidates that failed to realise that they needed to solve the inequality in X assumed here that S was normal or at least that they could approximate to the normal.

Part (e) was well done by large numbers of candidates and even those who obtained few marks elsewhere in the question often gave a completely correct solution here. It appeared that candidates have been well trained in answering normal approximation questions. Some answers gave the wrong variance, 40, assuming that the mean and the variance were the same, as with the Poisson. There were a surprising number of candidates who did not show the details of their standardisation which would have been advisable as marks could not be awarded if they had an incorrect value or the variance or mean.

Statistics for S2 Practice Paper Silver 4

				Mean average scored by candidates achieving grade:							
Qu	**Max Score**	**Modal score**	**Mean %**	**ALL**	**A***	**A**	**B**	**C**	**D**	**E**	**U**
1	10		65.2	6.52		7.60	6.59	6.17	5.55	4.85	3.67
2	10		68.5	6.85	9.22	8.53	7.27	5.82	4.18	2.65	1.19
3	7		63.9	4.47		5.21	3.58	2.62	1.48	1.58	0.58
4	9	9	67.0	6.03	8.20	7.56	6.34	5.21	4.08	2.98	1.65
5	12	12	68.2	8.18	11.44	10.69	8.87	6.91	5.21	3.38	1.74
6	10	9	65.0	6.53	8.04	7.44	5.66	4.27	3.25	2.39	1.18
7	17		67.8	11.52	14.91	12.84	10.08	9.25	7.73	6.46	3.72
	75		**66.8**	**50.10**		**59.87**	**48.39**	**40.25**	**31.48**	**24.29**	**13.73**