Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Bronze Level B4

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 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
73	66	58	50	42	34

1. A manufacturer produces sweets of length L mm where L has a continuous uniform distribution with range [15, 30].

(*a*) Find the probability that a randomly selected sweet has length greater than 24 mm.

(2)

These sweets are randomly packed in bags of 20 sweets.

(*b*) Find the probability that a randomly selected bag will contain at least 8 sweets with length greater than 24 mm.

(3)

(*c*) Find the probability that 2 randomly selected bags will both contain at least 8 sweets with length greater than 24 mm.

(2)

---

2. The probability of a bolt being faulty is 0.3. Find the probability that in a random sample of 20 bolts there are

(*a*) exactly 2 faulty bolts,

(2)

(*b*) more than 3 faulty bolts.

(2)

These bolts are sold in bags of 20. John buys 10 bags.

(*c*) Find the probability that exactly 6 of these bags contain more than 3 faulty bolts.

(3)

---

3. The continuous random variable X is uniformly distributed over the interval [–1, 3].
Find

(*a*) $\mathrm{E}(X)$

(1)

(*b*) $\mathrm{Var}(X)$

(2)

(*c*) $\mathrm{E}(X^2)$

(2)

(*d*) $\mathrm{P}(X < 1.4)$

(1)

A total of 40 observations of X are made.

(*e*) Find the probability that at least 10 of these observations are negative.

(5)

---

4. (*a*) Write down the two conditions needed to approximate the binomial distribution by the Poisson distribution.

(2)

A machine which manufactures bolts is known to produce 3% defective bolts. The machine breaks down and a new machine is installed. A random sample of 200 bolts is taken from those produced by the new machine and 12 bolts are defective.

(*b*) Using a suitable approximation, test at the 5% level of significance whether or not the proportion of defective bolts is higher with the new machine than with the old machine. State your hypotheses clearly.

(7)

---

5. Each cell of a certain animal contains 11 000 genes. It is known that each gene has a probability 0.0 005 of being damaged.

A cell is chosen at random.

(*a*) Suggest a suitable model for the distribution of the number of damaged genes in the cell.

(2)

(*b*) Find the mean and variance of the number of damaged genes in the cell.

(2)

(*c*) Using a suitable approximation, find the probability that there are at most 2 damaged genes in the cell.

(4)

---

6. A factory produces components of which 1% are defective. The components are packed in boxes of 10. A box is selected at random.

(*a*) Find the probability that the box contains exactly one defective component.

(2)

(*b*) Find the probability that there are at least 2 defective components in the box.

(3)

(*c*) Using a suitable approximation, find the probability that a batch of 250 components contains between 1 and 4 (inclusive) defective components.

(4)

---

7. A shopkeeper knows, from past records, that 15% of customers buy an item from the display next to the till. After a refurbishment of the shop, he takes a random sample of 30 customers and finds that only 1 customer has bought an item from the display next to the till.

(*a*) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not there has been a change in the proportion of customers buying an item from the display next to the till.

(6)

During the refurbishment a new sandwich display was installed. Before the refurbishment 20% of customers bought sandwiches. The shopkeeper claims that the proportion of customers buying sandwiches has now increased. He selects a random sample of 120 customers and finds that 31 of them have bought sandwiches.

(*b*) Using a suitable approximation and stating your hypotheses clearly, test the shopkeeper's claim. Use a 10% level of significance.

(8)

---

8. A telesales operator is selling a magazine. Each day he chooses a number of people to telephone. The probability that each person he telephones buys the magazine is 0.1.

(*a*) Suggest a suitable distribution to model the number of people who buy the magazine from the telesales operator each day.

(1)

(*b*) On Monday, the telesales operator telephones 10 people. Find the probability that he sells at least 4 magazines.

(3)

(*c*) Calculate the least number of people he needs to telephone on Tuesday, so that the probability of selling at least 1 magazine, on that day, is greater than 0.95.

(3)

A call centre also sells the magazine. The probability that a telephone call made by the call centre sells a magazine is 0.05. The call centre telephones 100 people every hour.

(*d*) Using a suitable approximation, find the probability that more than 10 people telephoned by the call centre buy a magazine in a randomly chosen hour.

(3)

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1(a)	$P(L>24) = \frac{1}{15} \times 6$	M1
	$= \frac{2}{5}$ or 0.4 oe	A1
		(2)
(b)	Let X represent the number of sweets with $L > 24$	
	$X \sim B(20, 0.4)$	M1
	$P(X \geq 8) = 1 - P(X \leq 7)$	M1dep
	$= 1 - 0.4159$	
	$= 0.5841$ awrt 0.584	A1
		(3)
(c)	$P(\text{both } X \geq 8) = (0.5841)^2$	M1
	$= 0.341\ldots$	A1 ft
		(2)
		Total 7

2.		
(a)	Let X be the random variable the number of faulty bolts	
	$P(X \leq 2) - P(X \leq 1) = 0.0355 - 0.0076$ or $(0.3)^2 (0.7)^{18} \frac{20!}{18!2!}$	**M1**
	$= 0.0279$ $= 0.0278$	**A1**
		(2)
(b)	$1 - P(X \leq 3) = 1 - 0.1071$	**M1**
	$= 0.8929$	**A1**
		(2)
(c)	$\frac{10!}{4!6!}(0.8929)^6 (0.1071)^4 = 0.0140.$	**M1A1√A1**
		(3)

Question Number	Scheme	Marks
3.		
(a)	$\mathrm{E}(X) = \frac{3-1}{2} = 1$	B1 cao
		(1)
(b)	$\mathrm{Var}(X) = \frac{(3+1)^2}{12} = \frac{4}{3}$ oe	M1A1
		(2)
(c)	$\mathrm{E}(X^2) = \frac{4}{3} + 1, \; = \frac{7}{3}$ oe	M1,A1
		(2)
(d)	P(X<1.4)=0.6	B1 cao
		(1)
(e)	$\mathrm{P}\left(X < 0\right) = 0.25$	B1
	Y is number of values less than 0	
	$Y \sim \mathrm{Bin}\left(40, 0.25\right)$	M1A1
	$\mathrm{P}(Y \geq 10) = 1 - \mathrm{P}(Y \leq 9)$	M1
	$= 1 - 0.4395 = 0.5605$	A1
		(5)
		[11]

4.		
(a)	n – large/high/big/ n >50	B1
	p – small/close to 0 / p < 0.2	B1
		(2)
(b)	$\mathrm{H}_0 : p = 0.03 \quad \mathrm{H}_1 : p > 0.03$	B1,B1
	Po(6)	B1
	$\mathrm{P}(X \geq 12) = 1 - \mathrm{P}(X \leq 11)$ or $\mathrm{P}(X \leq 10) = 0.9574$	M1
	$= 1 - 0.9799$ $\quad \mathrm{P}(X \geq 11) = 0.0426$	
	$= 0.0201$ $\quad$ CR $X \geq 11$	A1
	(0.0201 < 0.05)	
	Reject H_0 or Significant or 12 lies in the Critical region.	M1 dep.
	There is evidence that the proportion of defective bolts has increased.	A1 ft
		(7)
		Total 9

Question Number	Scheme	Marks
5. (a)	$X \sim B(\ 11000,\ 0.0005)$	M1 A1 (2)
(b)	$E(X) = 11000 \times 0.0005 = 5.5$	B1
	$\text{Var}\ (X) = 11000 \times 0.0005 \times (1 - 0.0005)$ $= 5.49725$	B1 (2)
(c)	$X \sim \text{Po}\ (5.5)$	M1 A1
	$P(\ X \leq 2) = 0.0884$	M1 A1 (4)
		(8 marks)

Question Number	Scheme	Marks
6(a)	X represents the number of defective components.	
	$P\ (X = 1) = \left(0.99\right)^{9}\left(0.01\right)\times 10\ = 0.0914$	M1A1
		(2)
(b)	$P\ (X \geq 2) = 1 - P(\ X \leq 1)$ $= 1 - (p)^{10} - \text{(a)}$ $= 0.0043$	M1 A1√ A1
		(3)
(c)	$X \sim \text{Po}(2.5)$	B1B1
	$P(1 \leq X \leq 4) = P(\ X \leq 4) - P(\ X = 0)$	M1
	$= 0.8912 - 0.0821$	
	$= 0.809$	A1
		(4)

Question Number	Scheme	Marks
7. (a)	$\text{H}_0 : p = 0.15 \quad \text{H}_1 : p \neq 0.15$	B1 B1
	$X \sim \text{B}(30, 0.15)$	M1
	$\text{P}(X \leq 1) = 0.0480$ or CR: $X = 0$	A1
	$(0.0480 > 0.025)$	
	not a significant result or do not reject H_0 or not in CR	M1
	there is no evidence of a change in the proportion of customers buying an item from the display.	A1ft
		(6)
(b)	$\text{H}_0 : p = 0.2 \quad \text{H}_1 : p > 0.2$	B1
	Let S = the number who buy sandwiches, $S \sim \text{B}(120, 0.2)$,	
	$S \approx W \sim \text{N}\left(24, \sqrt{19.2}^2\right)$	M1 A1
	$\text{P}(S \geq 31) = \text{P}(W \geq 30.5)$	M1
	$= \text{P}\left(Z > \dfrac{30.5 - 24}{\sqrt{19.2}}\right)$ or $\dfrac{x - 0.5 - 24}{\sqrt{19.2}} = 1.2816$	M1
	$[= \text{P}(Z > 1.48..)]$	
	$= 1 - 0.9306$	M1
	$= 0.0694$ $\quad x = 30.1$	A1
	< 0.10 so a significant result, there is evidence that more customers are purchasing sandwiches or the shopkeepers claim is correct.	B1ft
		(8)
		14

8 (a)	Distribution $X \sim \text{B}(n, 0.1)$	B1
		(1)
(b)	$Y \sim \text{B}(10, 0.1)$	B1
	$\text{P}(Y \geq 4) = 1 - \text{P}(Y \leq 3)$	M1
	$= 1 - 0.9872$	
	$= 0.0128$	A1
		(3)
(c)		
	$0.9^n < 0.05$ or $1 - (0.9)^n > 0.95$	M1
	$n > 28.4$	A1
	$n = 29$	A1
	alternative	
	$\text{B}(28, 0.1)$: $\text{P}(0) = 0.0523$	M1
	$\text{B}(29, 0.1)$: $\text{P}(0) = 0.0471$	A1
	$n = 29$	A1cao
		(3)
(d)	$C \sim \text{Po}(5)$	B1
	$\text{P}(C > 10) = 1 - \text{P}(C \leq 10)$	M1
	$= 1 - 0.9863$	
	$= 0.0137$	A1
		(3)
		[10]

Examiner reports

Question 1

This question was accessible to the majority of candidates, with many gaining full marks. Most were familiar with the continuous uniform distribution and were able to find P(L > 24) in part (a) A small minority of candidates used a discrete uniform distribution and calculated P($L \geq 25$) giving a probability of $\frac{1}{3}$. Part (b) was generally well answered and candidates used binomial tables for B(20, 0.4) with few errors. The most common error was to write and then evaluate P($X \geq 8$) = 1 – P($X \leq 8$). Candidates who got an answer of $\frac{1}{3}$ for part (a) struggled with using B(20, $\frac{1}{3}$) in part (b) and resorted to finding P(X = 8). A small minority used Poisson or normal distributions, e.g. Po(8), N(20, 0.4) or N(40, 0.4). Those who gained few marks in parts (a) and (b) were often able to gain obtain full marks for part (c) for finding the square of their answer to (b). The most common errors in part (c) were to double the answer in part (b), rather than square it or use the distribution B(40, 0.4) and evaluate $\mathrm{P}(X \leq 8)$.

Question 2

This question was well answered by the majority of candidates.

(a) This proved relatively straightforward for most, with the occasional response finding P($X \leq 2$) or using the binomial formula with x = 3.

(b) The most common error was interpreting ‘more than 3’ as 1 - P($X \leq 2$).

(c) Answers to this part reflected good preparation on this topic with a high proportion of successful responses. Candidates who lost marks on this question did so because their response to part (b) was incorrect or they used n = 20. A few candidates gave the answer as P(X = 6) = (answer to part (b))^6.

Question 3

This question was a good source of marks for many candidates. In parts (a) and (b) the candidates who chose to use the formula generally did so successfully. The few who attempted integration to obtain a solution did so with variable success. In part (c) a few candidates who attempted to use the formula $\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \left(\mathrm{E}(X)\right)^2$ were unable to correctly rearrange it to obtain $\mathrm{E}\left(X^2\right) = \mathrm{Var}(X) + \left(\mathrm{E}(X)\right)^2$.

The most common error was to get $\mathrm{E}\left(X^2\right) = \mathrm{Var}(X) - \left(\mathrm{E}(X)\right)^2$.

A small number of candidates used an alternative method, starting from ‘first principles’: $\mathrm{E}\left(X^2\right) = \int_{-1}^{3} \frac{1}{4} x^2 \mathrm{d}x = \left[\frac{x^3}{12}\right]_{-1}^{3} = \frac{27}{12} - \frac{-1}{12} = \frac{7}{3}$. Most of these candidates were successful, although a final answer of 13/6 was obtained by a few candidates who failed to deal successfully with the two negatives. Part (e) was very well done by a large majority of candidates: a clear and concise method was provided together with fully detailed working leading to the correct answer.

Question 4

This question was accessible to the majority of candidates. Whilst many candidates knew "the two conditions needed to approximate the binomial distribution by the Poisson distribution this knowledge was by no means universal. Some candidates appear to have tried to memorise the conditions for:

- modelling a situation using a Binomial distribution
- modelling a situation using a Poisson distribution
- approximating the binomial distribution with the Poisson distribution
- approximating the binomial distribution with the Normal distribution

but failed to remember which set of conditions applies in which situation.

The response to part (b) was very good, with many candidates gaining full marks. The most common error made was to state the hypotheses as $\mathrm{H}_0 : \lambda = 6$ and $\mathrm{H}_1 : \lambda > 6$. The question clearly states that we are testing for a 'proportion', so that the null hypotheses should be $\mathrm{H}_0 : p = 0.03$ and $\mathrm{H}_1 : p > 0.03$. The majority of candidates used the correct Poisson distribution and successfully calculated the probability 0.0201.

It was common to see candidates using the incorrect statement P($X \geq 12$) = 1 – P($X \leq 12$) or calculating 1 – P($X \leq 10$) and writing CR: $X \geq 10$. The candidates who tried to find the critical region were more likely to make an error and it is recommended that the probability route is used.

It is pleasing to see that most candidates finished their answers well with a clear conclusion using the context written in the question.

Question 5

This question was generally answered well. A few candidates put the Poisson for (a) and then used Variance = Mean to got 5.5 for the variance. Some candidates rounded incorrectly giving an answer of 5.49 for the variance.

Part (c) was generally answered correctly although a minority of candidates used the normal approximation – most used 2.5 in their standardisation and so got 1 mark out of the 4.

Question 6

This question was attempted by most candidates with a good degree of success for those who were competent in using the Binomial formula. A number of candidates had difficulty writing 1% as a decimal and used 0.1 in error.

In part (b) the most common errors were to see 'at least 2' translated as P(X >2) or to write P($X \geq 2$) as 1 - P($X \leq 2$). Many final accuracy marks were lost as a result of inadequate rounding in both parts (a) and (b).

In part (c) Po(2.5) ensured many scored at least 2 marks but P($X \leq 4$) – P($X \leq 1$) was a common error.
A normal approximation was seen but not quite as often as in previous years.

Question 7

This appeared to be generally a fairly straightforward question and there were many fully correct and clear solutions. In part (a) a number of candidates chose the wrong $H_1 : p < 0.15$. There are two main methods for conducting this significance test. It was noticeable that the candidates who chose the ‘Critical Region’ approach were generally less successful than those who evaluated the probability of obtaining an outcome ‘as bad or worse’ than that observed. It is advised that using the Critical Region method for hypothesis tests should be discouraged. The main errors generally involved comparing the probability found with an incorrect value. It was not uncommon to see the correct probability of 0.0480 compared with 0.05, instead of with 0.025 when a two tail test had been indicated by the hypotheses. Others incorrectly found $P(X = 1)$. Part (b) appeared to be a lot easier than part (a) and many were able to accurately find the p value and gave good conclusions in context, Some candidates approximated to the Poisson and then to the normal so having the wrong variance for their distribution. Whilst most candidates attempted the continuity correction a substantial number wrongly used 31.5 instead of 30.5. A very small minority of the candidates attempted to use a ‘Critical Region’ approach however these were mostly incorrect and gained few marks and were not as successful as those who calculated a probability.

Question 8

A significant proportion of candidates obtained full marks on parts (a), (b) and (d). The main exceptions were to refer to the Poisson distribution featured in part (a) and use it in part (b). Attempts to use the Normal distribution appeared occasionally in part (d).

Part (c) proved to be a challenge for many candidates. A significant number failed to make a serious attempt at solving the problem. The next most commonly appearing error was made by those candidates who believed that the cumulative Binomial table should be used and gave an answer of n = 30, the closest number to the true answer that appears in the table.

Many candidates adopted a formal approach: $0.9^n < 0.05$. On the other hand, there were candidates who adopted a less rigorous approach, writing $0.9^n = 0.05$. They were then able to write $n = 28.4$, and conclude, on the grounds of common sense, that the answer must be 29. It is pleasing to report that some candidates were able to handle the inequality signs correctly (reversing the inequality when dividing by a negative). However, some did not: they arrived (as the result of incorrect work) at $n < 28.4$. Many of these candidates, unfortunately, also used ‘common sense’ to conclude that the answer must be 29, despite the fact that an answer of 29 is not consistent with the statement $n < 28.4$. In this case the final mark could not be awarded.

Statistics for S2 Practice Paper Bronze 4

Qu	Max Score	Modal score	Mean %	Mean average scored by candidates achieving grade: ALL	A*	A	B	C	D	E	U
1	7		83.9	5.87	6.85	6.73	6.39	5.71	4.65	3.30	1.72
2	7		82.9	5.80		6.53	5.66	4.71	3.74	2.87	1.46
3	11		80.5	8.86	10.27	9.94	8.44	7.24	5.47	4.89	2.80
4	9		76.2	6.86	8.35	7.99	7.22	6.45	5.37	3.95	2.11
5	8		81.8	6.54		7.23	6.68	6.19	5.62	5.23	3.85
6	9		81.0	7.29		8.09	7.00	6.23	5.63	4.28	1.67
7	14		72.1	10.10	12.71	11.98	10.77	9.34	7.22	4.74	2.28
8	10	10	71.8	7.18	9.16	8.44	7.39	6.54	5.76	4.81	3.23
	75		**78.0**	**58.50**		**66.93**	**59.55**	**52.41**	**43.46**	**34.07**	**19.12**