6684/01

Edexcel GCE

Statistics S2
Silver Level S2

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
70	62	54	46	38	29

1. A disease occurs in 3% of a population.

   (*a*) State any assumptions that are required to model the number of people with the disease in a random sample of size n as a binomial distribution.

   (2)

   (*b*) Using this model, find the probability of exactly 2 people having the disease in a random sample of 10 people.

   (3)

   (*c*) Find the mean and variance of the number of people with the disease in a random sample of 100 people.

   (2)

   A doctor tests a random sample of 100 patients for the disease. He decides to offer all patients a vaccination to protect them from the disease if more than 5 of the sample have the disease.

   (*d*) Using a suitable approximation, find the probability that the doctor will offer all patients a vaccination.

   (3)

---

2. An effect of a certain disease is that a small number of the red blood cells are deformed. Emily has this disease and the deformed blood cells occur randomly at a rate of 2.5 per ml of her blood. Following a course of treatment, a random sample of 2 ml of Emily's blood is found to contain only 1 deformed red blood cell.

   Stating your hypotheses clearly and using a 5% level of significance, test whether or not there has been a decrease in the number of deformed red blood cells in Emily's blood.

   (6)

---

3. The probability of a telesales representative making a sale on a customer call is 0.15.

   Find the probability that

   (*a*) no sales are made in 10 calls,

   (2)

   (*b*) more than 3 sales are made in 20 calls.

   (2)

   Representatives are required to achieve a mean of at least 5 sales each day.

   (*c*) Find the least number of calls each day a representative should make to achieve this requirement.

   (2)

   (*d*) Calculate the least number of calls that need to be made by a representative for the probability of at least 1 sale to exceed 0.95.

   (3)

---

4. The length of a telephone call made to a company is denoted by the continuous random variable T. It is modelled by the probability density function

$$f(t) = \begin{cases} kt, & 0 \le t \le 10 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that the value of k is $\dfrac{1}{50}$.

(3)

(b) Find $P(T > 6)$.

(2)

(c) Calculate an exact value for $E(T)$ and for $Var(T)$.

(5)

(d) Write down the mode of the distribution of T.

(1)

It is suggested that the probability density function, $f(t)$, is not a good model for T.

(e) Sketch the graph of a more suitable probability density function for T.

(1)

3

5. A continuous random variable X has the probability density function f(x) shown in Figure 1.
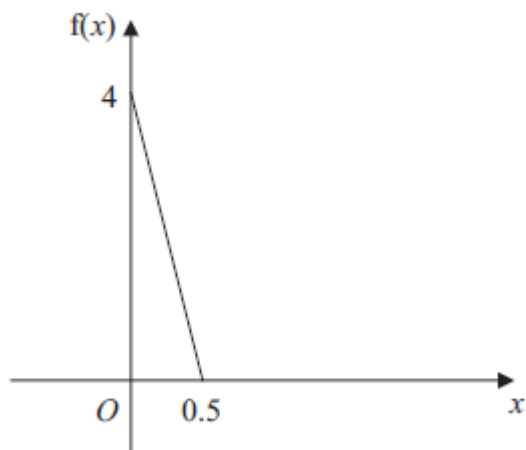

Figure 1

(a) Show that f(x) = 4 − 8x for $0 \leq x \leq 0.5$ and specify f(x) for all real values of x.

(4)

(b) Find the cumulative distribution function F(x).

(4)

(c) Find the median of X.

(3)

(d) Write down the mode of X.

(1)

(e) State, with a reason, the skewness of X.

(1)

_____

6. A company claims that a quarter of the bolts sent to them are faulty. To test this claim the number of faulty bolts in a random sample of 50 is recorded.

(a) Give two reasons why a binomial distribution may be a suitable model for the number of faulty bolts in the sample.

(2)

(b) Using a 5% significance level, find the critical region for a two-tailed test of the hypothesis that the probability of a bolt being faulty is $\frac{1}{4}$. The probability of rejection in either tail should be as close as possible to 0.025.

(3)

(c) Find the actual significance level of this test.

(2)

In the sample of 50 the actual number of faulty bolts was 8.

(d) Comment on the company's claim in the light of this value. Justify your answer.

(2)

The machine making the bolts was reset and another sample of 50 bolts was taken. Only 5 were found to be faulty.

(e) Test at the 1% level of significance whether or not the probability of a faulty bolt has decreased. State your hypotheses clearly.

(6)

_____

7. (a) Explain briefly what you understand by

(i) a critical region of a test statistic,

(ii) the level of significance of a hypothesis test.

(2)

(b) An estate agent has been selling houses at a rate of 8 per month. She believes that the rate of sales will decrease in the next month.

(i) Using a 5% level of significance, find the critical region for a one tailed test of the hypothesis that the rate of sales will decrease from 8 per month.

(ii) Write down the actual significance level of the test in part (b)(i).

(3)

The estate agent is surprised to find that she actually sold 13 houses in the next month. She now claims that this is evidence of an increase in the rate of sales per month.

(c) Test the estate agent's claim at the 5% level of significance. State your hypotheses clearly.

(5)

_____

TOTAL FOR PAPER: 75 MARKS

END

Question number	Scheme	Marks
1. **(a)**	Occurrences of the disease are independent B1 The probability of catching the disease remains constant.	**B1** **B1** **(2)**
(b)	$X \sim \text{Bin}(10,0.03)$ $P(X = 2) = \dfrac{10 \times 9}{2}\,(0.03)^2\,(0.97)^8 = 0.0317$	**B1** **M1 A1** **(3)**
(c)	$E(X) = 100 \times 0.03 = 3$ $\text{Var}(X) = 100 \times 0.03 \times 0.97 = 2.91$	**B1cao** **B1cao** **(2)**
(d)	$\lambda = 100 \times 0.03 = 3$ $Y \sim \text{Po}(3)$ $P(Y > 5) = 1 - P(Y \le 5)$ $\quad\quad = 1 - 0.9161$ $\quad\quad = 0.0839$	**B1 (use of)** **dM1** **A1** **(3)** **[10]**
2.	$H_0: \lambda = 2.5$ (or $\lambda = 5$) $\quad$ $H1: \lambda < 2.5$ (or $\lambda < 5$) $\quad\quad\quad\quad \lambda$ or μ	B1B1
	$X \sim \text{Po}(5)$	M1
	$P(X \le 1) = 0.0404$ $\quad\quad$ or $\quad$ CR $X \le 1$	A1
	[0.0404<0.05] this is significant $\quad$ or $\quad$ reject H_0 $\quad$ or $\quad$ it is in the critical region	M1
	There is evidence of a <u>decrease</u> in the (mean) <u>number/rate</u> of <u>deformed blood cells</u>	A1 $\quad$ (6) **(6 marks)**
3. (a)	$P(X = 0) = 0.85^{10}$ or from tables	M1
	$= 0.1969$ $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ awrt 0.197	A1 $\quad$ (2)
(b)	$P(X > 3) = 1 - P(X \le 3)$	M1
	$= 1 - 0.6477$ $= 0.3523$ $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ awrt 0.352	A1 $\quad$ (2)
(c)	$n \times 0.15 = 5$	M1
	$n = 33$ or 34	A1 $\quad$ (2)
(d)	$1 - P(X = 0) > 0.95$	M1
	$1 - (0.85)^n > 0.95.$	A1
	$0.85^n < 0.05$	
	$n > 18.4$	
	$n = 19$	A1 $\quad$ (3) **(9 marks)**

Question number	Scheme	Marks
4 **(a)**	$\int_0^{10} kt\,\mathrm{d}t = 1$ or Area of triangle $= 1$	M1
	$\left[\dfrac{kt^2}{2}\right]_0^{10} = 1$ or $10 \times 0.5 \times 10k = 1$ or linear equation in k	M1
	$50k = 1$	
	$k = \frac{1}{50}$ cso	A1
		(3)
(b)	$\int_6^{10} kt\,dt = \left[\dfrac{kt^2}{2}\right]_6^{10}$	M1
	$= \frac{16}{25}$	A1
		(2)
(c)	$\mathrm{E}(T) = \int_0^{10} kt^2\,dt = \left[\dfrac{kt^3}{3}\right]_0^{10}$	M1
	$= 6\frac{2}{3}$	A1
	$\mathrm{Var}\,(T) = \int_0^{10} kt^3\,dt - \left(6\dfrac{2}{3}\right)^2 = \left[\dfrac{kt^4}{4}\right]_0^{10} ; - \left(6\dfrac{2}{3}\right)^2$	M1;M1dep
	$= 50 - \left(6\frac{2}{3}\right)^2$	
	$= 5\frac{5}{9}$	A1
		(5)
(d)	10	B1
		(1)
(e)		B1

| | | (1) |

Question number	Scheme	Marks
5.		
(a)	$m = -\dfrac{4}{0.5} = -8$	**M1** **A1cso**
	$f(x) = 4 - 8x$ (*)	**B1**
	$f(x) = \begin{cases} -8x+4 & 0 \le x \le 0.5 \\ 0 & \textit{otherwise} \end{cases}$	**B1** **(4)**
(b)	$F(x) = \displaystyle\int_0^x (-8x+4)\,dx$	**M1**
	$\quad = \left[4x^2 + 4x\right]_0^x$	**M1**
	$F(x) = \begin{cases} 0 & x < 0 \\ -4x^2 + 4x & 0 \le x \le 0.5 \\ 1 & x > 0.5 \end{cases}$	**A1 B1** **(4)**
(c)	$-4x^2 + 4x = 0.5$	**M1**
	$x = \dfrac{1}{4}(2 - \sqrt{2}) = 0.146$	**M1 A1** **(3)**
(d)	$x = 0$	**B1** **(1)**
(e)	Positive Skew as mode < median	**B1ft** **(1)** **[13]**

Question number	Scheme	Marks
6. **(a)**	2 outcomes/faulty or not faulty/success or fail	B1
	A constant probability	B1
	Independence	
	Fixed number of trials (fixed n)	(2)
(b)	$X \sim B(50, 0.25)$	M1
	$P(X \leq 6) = 0.0194$	
	$P(X \leq 7) = 0.0453$	
	$P(X \geq 18) = 0.0551$	
	$P(X \geq 19) = 0.0287$	
	CR $X \leq 6$ and $X \geq 19$	A1 A1 (3)
(c)	$0.0194 + 0.0287 = 0.0481$	M1A1 (2)
(d)	8(It) is not in the Critical region or 8(It) is not significant or $0.0916 > 0.025$;	M1;
	There is evidence that the probability of a faulty bolt is 0.25 or the company's claim is correct.	A1ft (2)
(e)	H0 : $p = 0.25$ H1 : $p < 0.25$	B1B1
	$P(X \leq 5) = 0.0070$ or CR $X \leq 5$	M1A1
	$0.007 < 0.01$,	
	5 is in the critical region, reject H0, significant.	M1
	There is evidence that the probability of faulty bolts has decreased	A1ft 6) [15]

Question number	Scheme	Marks
7. (a) (i)	The **range of values/region/area/set of values** of the test statistic that would lead you to **reject H_0**	B1
(a) (ii)	The probability of incorrectly rejecting H_0 or Probability of rejecting H_0 when H_0 is true	B1 (2)
(b) (i)	$X \sim Po(8)$ $P(X \le 4) = 0.0996$ $P(X \le 3) = 0.0424$ Critical region [0,3]	M1 A1
(b) (ii)	awrt 0.0424	B1 (3)

(c)	$H_0 : \lambda = 8$ (or $\mu = 8$)
	 $H_1 : \lambda > 8$ (or $\mu > 8$)
	 $P(X \ge 13) = 1 - P(X \le 12)$ or $P(X \le 13) = 0.9658$
	                     or $P(X \ge 14) = 0.0342$
	       $= 1 - 0.9362$
	       $= 0.0638$ CR $X \ge 14$
	 so insufficient evidence to reject H_0 /not significant/ not in critical region
	 There in insufficient evidence of an increase/change in the **rate/number** of sales per month **or** the estate **agents** claim is incorrect
		B1
		
		 M1
		
		 A1
		 M1 dep
		 A1
		 (5)

		(10 marks)

Examiner reports

Question 1

This question proved to be a very good start to the paper for a large majority of candidates. In part (a) the assumptions were written "in context". However, there are still too many scripts where there was no mention of context at all. Many simply wrote a list of reasons as to why a Poisson distribution should be used rather than stating the assumptions that had been made.

In parts (b) and (c) fully correct solutions were seen in a large majority of candidates. However, there were a small number who failed to spot the change in the value of the parameter n from part (b) to part (c).

In part (d) most candidates provided a clear and accurate solution. However, there were a few candidates who used a Normal approximation, which is clearly inappropriate in this situation since "n is large and p is small" applies in this case and therefore indicates the use of a Poisson approximation. A small but significant number of candidates used a Binomial distribution despite the question requesting that a suitable approximation be used. A common error from those using a Poisson approximation was the use of $P(X \leq 4)$ instead of $P(X \leq 5)$.

Question 2

Candidates seemed better prepared for this type of question than in previous years. Marks were often lost for not using λ or μ in the hypotheses and for not putting the conclusion into context. A significant minority of candidates found $P(X=1)$ instead of $P(X \leq 1)$ but only a few candidates chose the critical region route.

Question 3

Part (a) was answered very well and most candidates secured both marks. There were the usual arithmetic slips leading to expressions like $3x > 20$ or $x > 5$ and there were a few candidates who thought that division by 5 meant the inequality should be reversed.

In part (b) most produced a quadratic equation with 3 terms and proceeded to solve and the correct critical values were usually obtained although 2 and 6 or −6 and 2 were sometimes seen. Some stopped at this stage and made no attempt to identify the appropriate regions. There were a number of sketches seen and these usually helped candidates to write down the correct inequalities but some lost the final mark for writing their answer as $-2 > x > 6$ or "$x < -2$ and $x > 6$".

Question 4

Part (a), with its 'answer given', produced fewer problems than similar questions in previous papers. Most candidates were able to obtain the required value of k.

Part (b) was generally well done. There were a wide variety of methods used such as finding the area of a trapezium, others found the area of the triangle and subtracted from 1. Others obtained F(x). The most common fault was the use of incorrect limits, 7 was often seen as the lower limit.

Part (c) was a good source of marks for a majority of candidates. A few lost marks as a result of not writing their answers as an exact number. However, many provided answers as both exact fractions and as approximated decimals. The most common error was to find $E(T^2)$, call it $Var(T)$ and then stop.

Part (d) was not popular. Of those who attempted it, there were some long-winded methods involving calculus and ultimately incorrect answers. The most successful candidates did a quick sketch of the p.d.f. to find the mode.

There were a few good sketches in part (e) however; there were all sorts of alternatives. Many just gave a sketch of the original p.d.f.

Question 5

The majority of candidates were able to attempt all parts of this question. However, part (a), proved to be a challenge to many. It was common to find candidates verifying rather than showing that $y = 4 - 8x$, by substituting in either value in each pair of co-ordinates to get the other or showing that $\int_0^{0.5} (4 - 8x)\, dx = 1$ and then stating 'f(x) $= 4 - 8x$ for $0 \le x \le 0.5$'. For a minority finding the gradient of the line also proved challenging, with a variety of methods seen. Often seen was $m = \dfrac{4}{0.5} = 8$ using a diagram as an aid, with only the more observant candidates adding a note to explain why it must be negative. In some cases, candidates gave exemplary responses to the first part of part (a) but did not then proceed to specify f(x) fully, hence losing two marks. Part (b) was generally well answered by the majority of candidates. In part (c) the most common errors were to find F(0.5) or to solve f(x) = 0.5. The common incorrect modes given in part (d) were 4 or 0.5.

The majority of candidates were able to follow through their answers to part (c) and (d) to give the correct direction and reason for the skewness of X. A few candidates also calculated the mean, usually correctly, which was unnecessary.

Question 6

Part (a) was well answered as no context was required.

In part (b) candidates identified the correct distribution and with much of the working being correct. However although the lower limit for the critical region was identified the upper limit was often incorrect. It is disappointing to note that many candidates are still losing marks when they clearly understand the topic thoroughly and all their work is correct except for the notation in the final answer. It cannot be overstressed that $P(X \leq 6)$ is not acceptable notation for a critical region. Others gave the critical region as $6 \leq X \leq 19$.

In part (c) the majority of candidates knew what to do and just lost the accuracy mark because of errors from part (b) carried forward.

Part (d) tested the understanding of what a critical region actually is, with candidates correctly noting that 8 was outside the critical region but then failing to make the correct deduction from it. Some were clearly conditioned to associate a claim with the alternative hypothesis rather than the null hypothesis. A substantial number of responses where candidates were confident with the language of double-negatives wrote "8 is not in the critical region so there is insufficient evidence to disprove the company's claim". Other candidates did not write this, but clearly understood when they said, more simply "the company is correct".

Part (e) was generally well done with correct deductions being made and the contextual statement being made. A few worked out $P(X = 5)$ rather than $P(X \leq 5)$

Question 7

Some candidates were familiar with the term 'critical region', but for all but a tiny number, 'significance level' was uncharted territory.

Part (b) was answered well by the majority of candidates. The main error was the use of the incorrect notation for a critical region, i.e. P($X \leq 3$).

There were many good solutions to part (c) that demonstrated correct identification of the situation "as bad or worse than that observed", accurate calculation of the relevant probability and a clear conclusion in context..
A small number of candidates found P($X = 13$), clearly failing to grasp the central concept of an event "as bad or worse than that observed". P($X > 14$) also featured in some scripts, bearing no relation to either this latter concept or any attempt to find the critical region.

Statistics for S2 Practice Paper Silver 2

Qu	Max Score	Modal score	Mean %	**Mean average scored by candidates achieving grade:**							
				ALL	A*	A	B	C	D	E	U
1	10		81.1	8.11	8.86	8.44	7.69	7.19	5.92	5.16	3.23
2	6		73.7	4.42		5.46	4.77	3.93	2.96	2.03	0.96
3	9		72.6	6.53	7.37	6.81	5.73	4.87	4.16	3.30	2.21
4	12		76.3	9.16		10.42	8.81	8.27	7.71	6.53	2.61
5	13		76.8	9.99	11.88	11.31	9.35	8.47	6.38	4.79	2.61
6	15		70.9	10.64	13.87	12.97	11.22	9.38	6.78	4.51	2.25
7	10		69.5	6.95	8.26	7.56	6.37	4.85	3.59	2.48	0.81
	75		**74.4**	**55.80**		**62.97**	**53.94**	**46.96**	**37.50**	**28.80**	**14.68**