Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Silver Level S1

Time: 1 hour 30 minutes

Materials required for examination papers	**Items included with question papers**
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
72	65	56	47	38	29

1. A botanist is studying the distribution of daisies in a field. The field is divided into a number of equal sized squares. The mean number of daisies per square is assumed to be 3. The daisies are distributed randomly throughout the field.

Find the probability that, in a randomly chosen square there will be

(*a*) more than 2 daisies,
(3)

(*b*) either 5 or 6 daisies.
(2)

The botanist decides to count the number of daisies, x, in each of 80 randomly selected squares within the field. The results are summarised below

$$\sum x = 295 \qquad \sum x^2 = 1386$$

(*c*) Calculate the mean and the variance of the number of daisies per square for the 80 squares. Give your answers to 2 decimal places.
(3)

(*d*) Explain how the answers from part (*c*) support the choice of a Poisson distribution as a model.
(1)

(*e*) Using your mean from part (*c*), estimate the probability that exactly 4 daisies will be found in a randomly selected square.
(2)

---

2. In a large college 58% of students are female and 42% are male. A random sample of 100 students is chosen from the college. Using a suitable approximation find the probability that more than half the sample are female.
(7)

---

3. An engineering company manufactures an electronic component. At the end of the manufacturing process, each component is checked to see if it is faulty. Faulty components are detected at a rate of 1.5 per hour.

(*a*) Suggest a suitable model for the number of faulty components detected per hour.
(1)

(*b*) Describe, in the context of this question, two assumptions you have made in part (*a*) for this model to be suitable.
(2)

(*c*) Find the probability of 2 faulty components being detected in a 1 hour period.
(2)

(*d*) Find the probability of at least one faulty component being detected in a 3 hour period.
(3)

---

4. Richard regularly travels to work on a ferry. Over a long period of time, Richard has found that the ferry is late on average 2 times every week. The company buys a new ferry to improve the service. In the 4-week period after the new ferry is launched, Richard finds the ferry is late 3 times and claims the service has improved. Assuming that the number of times the ferry is late has a Poisson distribution, test Richard's claim at the 5% level of significance. State your hypotheses clearly.

(6)

---

5. Sue throws a fair coin 15 times and records the number of times it shows a head.

(*a*) State the distribution to model the number of times the coin shows a head.

(2)

Find the probability that Sue records

(*b*) exactly 8 heads,

(2)

(*c*) at least 4 heads.

(2)

Sue has a different coin which she believes is biased in favour of heads. She throws the coin 15 times and obtains 13 heads.

(*d*) Test Sue's belief at the 1% level of significance. State your hypotheses clearly.

(6)

---

6. A random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{2}, & 0 \le x < 1, \\ x - \frac{1}{2}, & 1 \le x \le k, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a positive constant.

(*a*) Sketch the graph of f(x).

(2)

(*b*) Show that $k = \frac{1}{2}(1 + \sqrt{5})$.

(4)

(*c*) Define fully the cumulative distribution function F(x).

(6)

(*d*) Find P($0.5 < X < 1.5$).

(2)

(*e*) Write down the median of X and the mode of X.

(2)

(*f*) Describe the skewness of the distribution of X. Give a reason for your answer.

(2)

---

7. A random variable X has probability density function given by

$$f(x) = \begin{cases} -\frac{2}{9}x + \frac{8}{9}, & 1 \le x \le 4 \\ 0, & \text{otherwise.} \end{cases}$$

(*a*) Show that the cumulative distribution function F(x) can be written in the form $ax^2 + bx + c$, for $1 \le x \le 4$ where a, b and c are constants.

(3)

(*b*) Define fully the cumulative distribution function F(x).

(2)

(*c*) Show that the upper quartile of X is 2.5 and find the lower quartile.

(6)

Given that the median of X is 1.88,

(*d*) describe the skewness of the distribution. Give a reason for your answer.

(2)

TOTAL FOR PAPER: 75 MARKS

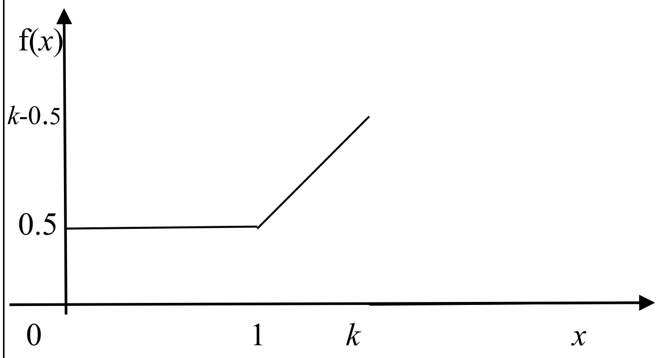
END

Question Number	Scheme	Marks
1	The random variable X is the number of daisies in a square. Poisson(3)	B1
(a)	$1 - P(X \le 2) = 1 - 0.4232$ $\quad 1 - e^{-3}(1 + 3 + \frac{3^2}{2!})$	M1
	$= 0.5768$	A1
		(3)
(b)	$P(X \le 6) - P(X \le 4) = 0.9665 - 0.8153$ $\quad e^{-3}\left(\frac{3^5}{5!} + \frac{3^6}{6!}\right)$	M1
	$= 0.1512$	A1
		(2)
(c)	$\mu = 3.69$	B1
	$\text{Var}(X) = \frac{1386}{80} - \left(\frac{295}{80}\right)^2$	M1
	$= 3.73/3.72/3.71$ $\quad$ accept $s^2 = 3.77$	A1
		(3)
(d)	For a Poisson model , Mean = Variance ; For these data $3.69 \approx 3.73$ $\Rightarrow$ Poisson model	B1
		(1)
(e)	$\frac{e^{-3.6875}3.6875^4}{4!} = 0.193$ $\quad$ allow their mean or var	M1
	Awrt 0.193 or 0.194	A1 ft
		(2)

2.	$X \sim B(100, 0.58)$ $Y \sim N(58, 24.36)$	B1 B1 B1
	$[P(X > 50) = P(X \ge 51)]$	
	using 50.5 or 51.5 or 49.5 or 48.5	M1
	$= P\left(z \ge \pm\left(\frac{50.5 - 58}{\sqrt{24.36}}\right)\right)$ $\quad$ standardising 50.5, 51, 51.5 ,48.5,49, 49.5 and their μ and σ for M1	M1
	$= P(z \ge -1.52\ldots)$	A1
	$= 0.9357$	A1 (7)
		(7 marks)

Question Number	Scheme		Marks
3 (a)	$X \sim \text{Po}(1.5)$	need Po and 1.5	**B1** **(1)**
(b)	<u>Faulty</u> components occur at a constant rate. <u>Faulty</u> components occur independently or randomly. <u>Faulty</u> components occur singly.	any two of the 3 only need faulty once	**B1** **B1** **(2)**
(c)	$P(X=2) = P(X \le 2) - P(X \le 1)$ or $\dfrac{e^{-1.5}(1.5)^2}{2}$		**M1**
	$= 0.8088 - 0.5578$		
	$= 0.251$	awrt 0.251	**A1** **(2)**
(d)	$X \sim \text{Po}(4.5)$	4.5 may be implied	**B1**
	$P(X \ge 1) = 1 - P(X=0)$ $= 1 - e^{-4.5}$ $= 1 - 0.0111$		**M1**
	$= 0.9889$	awrt 0.989	**A1** **(3)**
			Total 8
4.	$H_0: \lambda = 8$ or $\mu = 2$ $\quad H_1: \lambda < 8$ or $\mu < 2$		**B1 B1**
	Under H0 , $X \sim \text{Po}(8)$		**M1**
	$P(X \le 3) = 0.0424$ $\quad$ CR $X \le 3$		**A1**
	$0.0424 < 0.05$, Reject H_0. Richard's claim is supported.		**M1A1ft** **[6]**

Question Number	Scheme	Marks
5. **(a)**	$X \sim \mathrm{B}(15, 0.5)$	B1 B1 (2)
(b)	$\mathrm{P}(X = 8) = \mathrm{P}(X \le 8) - \mathrm{P}(X \le 7)$ or $\left(\frac{15!}{8!7!}(p)^8(1-p)^7\right)$	M1
	$= 0.6964 - 0.5$	
	$= 0.1964$ awrt 0.196	A1 (2)
(c)	$\mathrm{P}(X \ge 4) = 1 - \mathrm{P}(X \le 3)$	M1
	$= 1 - 0.0176$	
	$= 0.9824$	A1 (2)
(d)	$H_0 : p = 0.5$	B1
	$H_1 : p > 0.5$	B1
	$X \sim \mathrm{B}(15, 0.5)$	
	$\mathrm{P}(X \ge 13) = 1 - \mathrm{P}(X \le 12)$ | $[\mathrm{P}(X \ge 12) = 1 - 0.9824 = 0.0176]$ att $\mathrm{P}(X \ge 13)$	M1
	$= 1 - 0.9963$ | $\mathrm{P}(X \ge 13) = 1 - 0.9963 = 0.0037$	
	$= 0.0037$ | CR $X \ge 13$ awrt 0.0037/ CR $X \ge 13$	A1
	$0.0037 < 0.01$ | $13 \ge 13$	
	Reject H_0 or it is significant or a correct statement in context from their values	M1
	There is sufficient evidence at the 1% significance level that the coin is <u>biased in favour of heads</u> *or* There is evidence that Sue's belief is correct	A1 (6)
		(12 marks)

Question Number	Scheme	Marks
6. (a)		B1 B1 (2)
(b)	$\int_1^k \left(x-\frac{1}{2}\right)dx = \frac{1}{2}$	M1
	$\left[\frac{1}{2}x^2 - \frac{1}{2}x\right]_1^k = \frac{1}{2}$	
	$k^2 - k - 1 = 0$ o.e.	A1
	$k = \frac{1}{2}\left(1+\sqrt{5}\right)$	M1A1 cso (4)
(c)	$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x, & 0 \le x < 1 \\ \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{2}, & 1 \le x \le k \\ 1, & x > k \end{cases}$	B1 M1A1A1B1 B1 1st and last (6)
	Note: Working for the M1A1A1 $\int_1^k x - \frac{1}{2}dx + C = \frac{1}{2}x^2 - \frac{1}{2}x\ ; + \frac{1}{2}$	(M1A1;A1)
(d)	$P(0.5 < X < 1.5) = F(1.5) - F(0.5)$ $= 0.875 - 0.25$ $= 0.625$	M1 A1 (2)
(e)	Median is $x = 1$	B1
	Mode is $x = k$ or $\frac{1}{2}(1+\sqrt{5})$ or awrt1.62	B1 (2)
(f)	Negative skew Median<mode or from graph more values are to the right.	B1 B1d (2) **(18 marks)**

Question Number	Scheme	Marks
7 (a)	$\mathrm{F}(x_0) = \int_1^x -\frac{2}{9}x + \frac{8}{9}\,\mathrm{d}x = \left[-\frac{1}{9}x^2 + \frac{8}{9}x\right]_1^x$	M1A1
	$= \left[-\frac{1}{9}x^2 + \frac{8}{9}x\right] - \left[-\frac{1}{9} + \frac{8}{9}\right]$	
	$= -\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9}$	A1
		(3)
(b)	$\mathrm{F}(x) = \begin{cases} 0 & x < 1 \\ -\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9} & 1 \le x \le 4 \\ 1 & x > 4 \end{cases}$	B1B1√
		(2)
(c)	$\mathrm{F}(x) = 0.75$; or $\mathrm{F}(2.5) = -\frac{1}{9}\times 2.5^2 + \frac{8}{9}\times 2.5 - \frac{7}{9}$	M1;
	$-\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9} = 0.75$	
	$4x^2 - 32^x + 55 = 0$	
	$-x^2 + 8x - 13.75 = 0$	
	$x = 2.5$ $= 0.75$ cso	A1
	and $\mathrm{F}(x) = 0.25$	
	$-\frac{1}{9}x^2 + \frac{8}{9}x - \frac{7}{9} = 0.25$	M1
	$-x^2 + 8x - 7 = 2.25$	
	$-x^2 + 8x - 9.25 = 0$ quadratic 3 terms =0	M1 dep
	$x = \dfrac{-8 \pm \sqrt{8^2 - 4\times -1 \times -9.25}}{2\times -1}$	M1 dep
	$x = 1.40$	A1
		(6)
(d)	$Q_3 - Q_2 > Q_2 - Q_1$	M1
	Or mode = 1 and mode < median	
	Or mean = 2 and median < mode	
	Sketch of pdf here or be referred to if in a different part of the question	
	Box plot with Q_1, Q_2, Q_3 values marked on	
	Positive skew	A1
		(2)

Examiner reports

Question 1

This question proved to be a good start to the paper for a majority of the candidates. There were many responses seen which earned full marks.

The most common errors in parts (a) and (b) concerned the routine manipulation of inequalities. In part (a) $1 - \mathrm{P}(X \le 1)$ was often seen and in (b), while most candidates agreed that $\mathrm{P}(5 \le X \le 6)$ was the required probability, with many then choosing the standard technique of $\mathrm{P}(X \le 6) - \mathrm{P}(X \le 4)$, there were candidates who proceeded with a variety of methods. Incorrect expressions such as $\mathrm{P}(X \le 6) - \mathrm{P}(X \ge 4)$ were seen not infrequently. A correct but inefficient method which was commonly used included: $\mathrm{P}(X = 5) + \mathrm{P}(X = 6) = \left(\mathrm{P}(X \le 6) - \mathrm{P}(X \le 5)\right) + \left(\mathrm{P}(X \le 5) - \mathrm{P}(X \le 4)\right)$

Part (c) was poorly answered. There were a significant minority of candidates who obtained a ‘correct’ answer for the mean in part (c), but who nevertheless lost the mark because their answer was not written, as instructed, correct to 2 decimal places. Many candidates were unable to calculate the variance. There were a variety of incorrect formulae used.

The general response to (d) was good, although many candidates simply gave the response that is appropriate for a more frequent type of question on the Poisson distribution requiring comment: (“singly/independently/randomly/constant rate”).

Part (e) was particularly well done. Even the minority who struggled, or even omitted, some of the earlier parts of the question were able to gain both marks in part (e).

Question 2

Candidates did this well on the whole, with the main error being the absence of a continuity correction (which was penalised appropriately.) The question worked well although the 42% mentioned in the question caused some confusion as many took this to be the percentage of students that were female rather than 58%. Very few candidates were unable to make a reasonable attempt at a Normal approximation.

Question 3

This question was quite well done. Most candidates were aware of the conditions for a Poisson distribution but many lost marks as they did not answer the question in context, although the examination question actually specified this. A few did not realise that independent and random were the same condition. Cumulative probability tables were well used in part (c) and there were many accurate solutions using the Poisson formula. Most candidates used a mean of 4.5 in part (d) and there were many accurate results.

Question 4

Many candidates gained full marks in this question. In particular, it is to be noted that most candidates had few problems with either the hypotheses or the conclusion. A sizeable minority of candidates used > instead of < in H_1. The most common error was to use $\mathrm{P}(X = 3)$ instead of $\mathrm{P}(X \le 3)$. There were also a number of candidates who failed to place their conclusion in context.

Question 5

Most candidates recognised that this was a Binomial for part (a) but quite a few did not define it completely by putting in the 15 and 0.5. Parts (b) and (c) were answered well .Most candidates are getting much better at the layout of their solutions and it was good to see that candidates are identifying one-tail tests, although some candidates did not get the hypotheses completely correct; omitting p altogether or using λ was not uncommon. The working for finding 0.0037 or a critical region ≥ 13 was often done well. Those who used a critical region method required more working and seemed to make more errors. It is recommended that this method is not used in Hypothesis testing at this level. There was the usual confusion in interpreting the test. In particular candidates often did not interpret correctly in context. 'The coin is biased' was a common inadequate answer.

Question 6

Most of the sketches seen in part (a) were of the correct shape. However, a disappointingly low number were awarded the second mark. Labelling the coordinates of end points and joins are essential when graphs are either in sections or else do not extend indefinitely in any direction.

In part (b) the most common method was to use integration and use the fact that total probability must be equal to one. However, there were a few candidates who wrote $\int_1^k \left(x - \frac{1}{2}\right) dx = 1$, forgetting about the probability distribution for $0 \le X \le 1$.

The second most common method was to calculate the areas of the simple shapes involved: either a rectangle and a parallelogram or else a rectangle and a triangle. This required some fiddly details, but it was pleasing to note that many fully correct responses were seen using this method.

Not all candidates were able to complete part (b) successfully. Some arrived at the correct equation $x^2 - x - 1 = 0$ and then stopped, as if they had never seen a quadratic equation before. It would appear that some candidates were a little too hasty when their next (and only) line following $x^2 - x - 1 = 0$ was $\frac{1}{2}\left(1+\sqrt{5}\right)$, which is the answer given in the question. Such candidates need to be reminded of the 'Advice to Candidates' on the front of the examination paper: "You should show sufficient working to make your methods clear to the Examiner. Answers without working may not gain full credit."

Some candidates made extra work for themselves by omitting the simple expedient of 'clearing fractions first'. The equation $x^2 - x - 1 = 0$ is much easier to deal with than the same equation $\frac{1}{2}x^2 - \frac{1}{2}x - \frac{1}{2} = 0$. For instance, there were candidates who obtained a correct expression $\dfrac{\frac{1}{2} \pm \sqrt{1.25}}{1}$ who were then unable to relate this to the required answer.

There were many fully correct answers to part (c) but there was also the usual significant number of candidates who failed to consider the constant ½ which needs to be added to $\int_1^x \left(t - \frac{1}{2}\right) dt$. The constant can be obtained in various ways, the simplest being to look at the sketch in part (a). By way of contrast, there were a few candidates who managed to obtain the correct value of the constant by the hardest possible route in this case: solving $F\left(\frac{1}{2}\left(1+\sqrt{5}\right)\right)=1$.

The candidates who were correct in (c) were generally also correct in (d). The most common error was to use $\left(\frac{1}{2}\times1.5^2-\frac{1}{2}\times1.5+0.5\right)-\left(\frac{1}{2}\times0.5^2-\frac{1}{2}\times0.5+0.5\right)$ where the 0.5 has been substituted into the wrong part of F(x).
There was a variety of ways of solving part (d). It is a matter of some concern that not all candidates made their method clear. This particularly affected those candidates whose answers in (c) were incorrect: they could have gained some marks if some explanation of their method had been provided.

Disappointingly few candidates responded to the hint "write down" in part (e). The most effective, and simplest, methods were visual. Half the area, as already noted in part (c), is to the left of $X = 1$. Furthermore, the highest point on the graph occurs when $X = k$. However, there were many correct answers obtained by other methods, for example, by solving the quadratic equation $\frac{1}{2}m^2-\frac{1}{2}m+\frac{1}{2}=0.5$.

Question 7

This question proved challenging in parts to some candidates but was attempted in full by many, with a high degree of success.

In part (a) most candidates were aware that they needed to integrate the given function and did so successfully, including the fractions. Problems generally arose in the use of the correct limits. It was common to see candidates use limits of 0 or 1 and 4 rather than using a variable upper limit. Several candidates chose to use a constant c rather than limits but often did not proceed to use F(4) =1 or F(1) = 0 to find the value of c. A large number of candidates who got the correct answer went on to multiply their expression by 9.

In (b) F(x) was defined well – candidates seem to be more aware of the need for the 0 and 1 and there were a limited number who had the wrong ranges for these.

The majority of correct answers in (c) were found by solving the quadratic rather than by the easier method of substituting 2.5 into the equation. Many of those who used the quadratic formula used complicated coefficients. Most went on to correctly find Q_1

There is still a great deal of confusion in the minds of some candidates over skewness with a number writing reasons such as $Q_1<Q_2<Q_3$. There was a tendency to write wordy explanations rather than the succinct $Q_3-Q_2>Q_2-Q_1$. This gained the marks but many candidates were unable to express themselves clearly.

There is still confusion between positive and negative skewness with a few candidates doing correct calculations but concluding it was negative.

A few candidates calculated the mean or mode and used mean > median > mode. These gained full marks if correctly found but used precious time doing unnecessary calculations.

Statistics for S2 Practice Paper Silver 1

Qu	Max Score	Modal score	Mean %	Mean average scored by candidates achieving grade: ALL	A*	A	B	C	D	E	U
1	11		80.0	8.80		9.68	7.69	6.88	6.43	4.70	2.49
2	7		79.0	5.53		6.48	5.88	5.28	4.31	3.24	1.43
3	8		77.9	6.23		7.19	6.66	6.07	5.30	4.38	2.31
4	6		80.8	4.85	5.67	5.41	4.44	3.78	2.11	1.19	0.32
5	12		80.6	9.67		11.19	10.22	9.22	7.74	6.00	3.68
6	18		71.7	12.90	16.11	15.18	12.61	8.96	7.23	5.48	1.97
7	13		70.0	9.10		11.56	8.28	6.30	5.04	3.30	0.96
	75		**76.1**	**57.08**		**66.69**	**55.78**	**46.49**	**38.16**	**28.29**	**13.16**