Paper Reference(s)

6684/01

Edexcel GCE

Statistics S2

Advanced Level

Wednesday 9 June 2010 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Pink)

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

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S2), the paper reference (6684), your surname, other name and signature.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has 7 questions.
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 not gain full credit.

1. Explain what you understand by

(*a*) a population,

(1)

(*b*) a statistic.

(1)

A researcher took a sample of 100 voters from a certain town and asked them who they would vote for in an election. The proportion who said they would vote for Dr Smith was 35%.

(*c*) State the population and the statistic in this case.

(2)

(*d*) Explain what you understand by the sampling distribution of this statistic.

(1)

---

2. Bhim and Joe play each other at badminton and for each game, independently of all others, the probability that Bhim loses is 0.2.

Find the probability that, in 9 games, Bhim loses

(*a*) exactly 3 of the games,

(3)

(*b*) fewer than half of the games.

(2)

Bhim attends coaching sessions for 2 months. After completing the coaching, the probability that he loses each game, independently of all others, is 0.05.

Bhim and Joe agree to play a further 60 games.

(*c*) Calculate the mean and variance for the number of these 60 games that Bhim loses.

(2)

(*d*) Using a suitable approximation calculate the probability that Bhim loses more than 4 games.

(3)

---

3. A rectangle has a perimeter of 20 cm. The length, X cm, of one side of this rectangle is uniformly distributed between 1 cm and 7 cm.

Find the probability that the length of the longer side of the rectangle is more than 6 cm long.

(5)

---

4. The lifetime, X, in tens of hours, of a battery has a cumulative distribution function $F(x)$ given by

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{4}{9}(x^2 + 2x - 3) & 1 \le x \le 1.5 \\ 1 & x > 1.5 \end{cases}$$

(a) Find the median of X, giving your answer to 3 significant figures.

(3)

(b) Find, in full, the probability density function of the random variable X.

(3)

(c) Find $P(X \geq 1.2)$

(2)

A camping lantern runs on 4 batteries, all of which must be working. Four new batteries are put into the lantern.

(d) Find the probability that the lantern will still be working after 12 hours.

(2)

---

5. A company has a large number of regular users logging onto its website. On average 4 users every hour fail to connect to the company's website at their first attempt.

(a) Explain why the Poisson distribution may be a suitable model in this case.

(1)

Find the probability that, in a randomly chosen **2 hour** period,

(b) (i) all users connect at their first attempt,

(ii) at least 4 users fail to connect at their first attempt.

(5)

The company suffered from a virus infecting its computer system. During this infection it was found that the number of users failing to connect at their first attempt, over a 12 hour period, was 60.

(c) Using a suitable approximation, test whether or not the mean number of users per hour who failed to connect at their first attempt had increased. Use a 5% level of significance and state your hypotheses clearly.

(9)

---

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. The random variable Y has probability density function f(y) given by

$$\text{f}(y) = \begin{cases} ky(a - y) & 0 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

where k and a are positive constants.

(a) (i) Explain why $a \geq 3$.

(ii) Show that $k = \dfrac{2}{9(a-2)}$.

(6)

Given that E(Y) = 1.75,

(b) show that $a = 4$ and write down the value of k.

(6)

For these values of a and k,

(c) sketch the probability density function,

(2)

(d) write down the mode of Y.

(1)

TOTAL FOR PAPER: 75 MARKS

END

edexcel

June 2010
Statistics S2 6684
Mark Scheme

Question Number		Scheme	Marks
Q1	(a)	A population is collection of all items	B1 (1)
	(b)	(A random variable) that is a function of the sample which contains no unknown quantities/parameters.	B1 (1)
	(c)	The voters in the town	B1
		Percentage/proportion voting for Dr Smith	B1 (2)
	(d)	Probability Distribution of those voting for Dr Smith from all possible samples (of size 100)	B1 (1)
			[5]

Notes

(a) **B1** – collection/group **all** items – need to have /imply all eg entire/complete/every

(b) **B1** – needs function/calculation(o.e.) of the sample/random variables/observations **and** no unknown quantities/parameters(o.e.) NB do not allow unknown variables

e.g. "A calculation based solely on observations from a given sample." B1
"A calculation based only on known data from a sample" B1
"A calculation based on known observations from a sample" B0

Solely/only imply no unknown quantities

(c) **B1** – Voters

Do not allow 100 voters.

B1 – percentage/ proportion voting (for Dr Smith)
the **number** of people voting (for Dr Smith)
Allow 35% of people voting (for Dr Smith)
Allow 35 people voting (for Dr Smith)
Do **not** allow 35% or 35 alone

(d) **B1** – answers must include all three of these features
(i) All possible samples,
(ii) their associated probabilities,
(iii) context of voting for Dr Smith.

e.g "It is all possible values of the percentage and their associated probabilities." B0 no context

Question Number	Scheme	Marks
Q2 (a)	Let X be the random variable the number of games Bhim loses. $X \sim B(9, 0.2)$	B1
	$P(X \leq 3) - P(X \leq 2) = 0.9144 - 0.7382$ or $(0.2)^3(0.8)^6 \frac{9!}{3!6!}$	M1
	$= 0.1762$ $= 0.1762$ awrt 0.176	A1 (3)
(b)	$P(X \leq 4) = 0.9804$ awrt 0.98	M1A1 (2)
(c)	Mean = 3 variance $= 2.85, \frac{57}{20}$	B1 B1 (2)
(d)	Po(3) poisson	M1
	$P(X > 4) = 1 - P(X \leq 4)$	M1
	$= 1 - 0.8153$	
	$= 0.1847$	A1 (3)
		[10]

Notes

(a) B1 – writing or use of B(9, 0.2)

M1 for writing/ using $P(X \leq 3) - P(X \leq 2)$ or $(p)^3(1-p)^6 \frac{9!}{3!6!}$

A1 awrt 0.176

(b) **M1** for writing or using $P(X \leq 4)$
A1 awrt 0.98

(c) **B1** 3
B1 2.85, or exact equivalent

(d) **M1** for using Poisson
M1 for writing or using $1 - P(X \leq 4)$ NB $P(X \leq 4)$ is 0.7254 Po(3.5) and 0.8912 Po(2.5)
A1 awrt 0.185

Special case :Use of Po(1.8) in (a) and (b)

(a) can get B1 M1 A0 – B1 if written B(9, 0.2), M1 for $\frac{e^{-1.8}1.8^3}{3!}$ **or** awrt to 0.161

If B(9, 0.2) is not seen then the only mark available for using Poisson is M1.
(b) can get M1 A0 - M1 for writing or using $P(X \leq 4)$ or may be implied by awrt 0.964

Use of Normal in (d)
Can get M0 M1 A0.- for M1 they must write $1 - P(X \leq 4)$ or get awrt 0.187

Question Number	Scheme			Marks
Q3	Method 1	Method 2	Method 3	
	P $(X > 6) = \frac{1}{6}$	P$(4 < X < 6) = \frac{1}{3}$	P $(X > 6) = \frac{1}{6}$	B1 M1
	P$(X < 4) = \frac{1}{2}$		Y~U[3,9] P $(Y > 6) = \frac{1}{2}$	A1
	total = $\frac{1}{6}+\frac{1}{2}= \frac{2}{3}$	$1 - \frac{1}{3}= \frac{2}{3}$	total = $\frac{1}{6}+\frac{1}{2}= \frac{2}{3}$	M1dep B A1 (5) **[5]**
	Notes **Methods 1 and 2** B1 for 6 and 4 (allow if seen on a diagram on x-axis) **M1** for P$(X > 6)$ or P$(6 < X < 7)$; or P$(X < 4)$ or P$(1 < X < 4)$; or P$(4 < X < 6)$ Allow $\leq$ *and* $\geq$ signs **A1** $\frac{1}{6}; or \frac{1}{2}; \frac{1}{3}$ must match the probability statement **M1** for adding their " P$(X > 6)$" and their "P$(X < 4)$" or 1 - their "P$(4 < X < 6)$" dep on getting first B mark **A1** cao $\frac{2}{3}$ **Method 3 *Y*~U[3, 9]** **B1** for 6 with U[1,7]and 6 with U[3,9] **M1** for P$(X > 6)$ or P$(6 < X < 7)$ or P$(6 < Y < 9)$ **A1** $\frac{1}{6}; or \frac{1}{2};$ must match the probability statement **M1** for adding their " P$(X > 6)$" and their "P$(Y >6)$" dep on getting first B mark **A1** cao $\frac{2}{3}$			

Question Number	Scheme	Marks
Q4 (a)	$\frac{4}{9}\left(m^2+2m-3\right)=0.5$	M1
	$m^2 + 2m - 4.125 = 0$	
	$m = \frac{-2\pm\sqrt{4+16.5}}{2}$	M1
	m =1.26, -3.264	
	(median =) 1.26	A1 (3)
(b)	Differentiating $\frac{\mathrm{d}\left(\frac{4}{9}\left(x^2+2x-3\right)\right)}{\mathrm{d}x}=\frac{4}{9}\left(2x+2\right)$	M1 A1
	$\mathrm{f}(x)=\begin{cases}\frac{8}{9}(x+1) & 1\le x\le 1.5\\ 0 & \text{otherwise}\end{cases}$	B1ft (3)
(c)	P($X \geq$ 1.2) = 1 – F(1.2)	M1
	= 1 – 0.3733	
	$=\frac{47}{75}$, 0.6267 awrt	A1 (2)
	0.627	
(d)	$(0.6267)^4$= 0.154 awrt 0.154 or 0.155	M1 A1 (2)
		[10]

Notes

(a) **M1** putting F(x) = 0.5
M1 using correct quadratic formula. If use calc need to get 1.26 (384...)
A1 cao 1.26 must reject the other root.
If they use Trial and improvement they have to get the correct answer to gain the second M mark.

(b) **M1** attempt to differentiate. At least one $x^n \rightarrow x^{n-1}$
A1 correct differentiation
B1 must have both parts- follow through their $\mathrm{F}'(x)$ Condone <

(c) **M1** finding/writing 1 – F(1.2) may use/write $\int_{1.2}^{1.5}\frac{8}{9}(x+1)\mathrm{d}x$ or 1 - $\int_{1}^{1.2}\frac{8}{9}(x+1)\mathrm{d}x$
or $\int_{1.2}^{1.5}$"their f(x)" dx . Condone missing dx

A1 awrt 0.627

(d) **M1** (c)^4 If expressions are not given you need to check the calculation is correct to 2sf.
A1 awrt 0.154 or 0.155

Question Number	Scheme	Marks
Q5 (a)	Connecting occurs at random/independently, singly or at a constant rate	B1 (1)
(b)	Po (8)	B1
(i)	$P(X = 0) = 0.0003$	M1A1
(ii)	$P(X \geq 4) = 1 - P(X \leq 3)$ $= 1 - 0.0424$ $= 0.9576$	M1 A1 (5)
(c)	$H_0: \lambda = 4$ (48) $H_1: \lambda > 4$ (48)	B1
	N(48,48)	M1 A1
	Method 1 $P(X \geq 59.5) = P\left(Z \geq \frac{59.5-48}{\sqrt{48}}\right)$ $= P(Z \geq 1.66)$ $= 1 - 0.9515$ $= 0.0485$	Method 2 $\frac{x-0.5-48}{\sqrt{48}} = 1.6449$ $x = 59.9$
		M1 M1 A1 A1
	$0.0485 < 0.05$ Reject H_0. Significant. 60 lies in the Critical region The number of failed connections at the first attempt has increased.	M1 A1 ft (9) **[15]**

Notes

(a) **B1** Any one of randomly/independently/singly/constant rate. Must have context of connection/logging on/fail

(b) **B1** Writing or using Po(8) in (i) or (ii)

(i) **M1** for writing or finding $P(X = 0)$
A1 awrt 0.0003

(ii) **M1** for writing or finding $1 - P(X \leq 3)$
A1 awrt 0.958

(c) **B1** both hypotheses correct. Must use λ or μ
M1 identifying normal
A1 using or seeing mean and variance of 48
These first two marks may be given if the following are seen in the standardisation formula : 48 and $\sqrt{48}$ or awrt 6.93
M1 for attempting a continuity correction (Method 1: 60 $\pm$ 0.5 / Method 2: $x \pm 0.5$)
M1 for standardising using their mean and their standard deviation and using either Method 1 [59.5, 60 or 60.5. accept $\pm$ z.] Method 2 [($x \pm 0.5$) and equal to a $\pm z$ value)
A1 correct z value awrt ± 1.66 **or** $\pm\frac{59.5-48}{\sqrt{48}}$, or $\frac{x-0.5-48}{\sqrt{48}} = 1.6449$
A1 awrt 3 sig fig in range 0.0484 – 0.0485, awrt 59.9
M1 for "reject H_0" or "significant" maybe implied by "correct contextual comment"
If one tail hypotheses given follow through "their prob" and 0.05 , $p < 0.5$
If two tail hypotheses given follow through "their prob" with 0.025, $p < 0.5$
If one tail hypotheses given follow through "their prob" and 0.95 , $p > 0.5$
If two tail hypotheses given follow through "their prob" with 0.975, $p > 0.5$
If no H_1 given they get M0
A1 ft correct contextual statement followed through from their prob and H_1. need the words number of failed connections/log ons has increased o.e.
Allow "there are more failed connections"
NB A correct contextual statement **alone** followed through from their prob and H_1 gets M1 A1

Question Number		Scheme	Marks
Q6	(a)	2 outcomes/faulty or not faulty/success or fail A constant probability Independence Fixed number of trials (fixed n)	B1 B1 (2)
	(b)	$X \sim B(50,0.25)$ $P(X \leq 6) = 0.0194$ $P(X \leq 7) = 0.0453$ $P(X \geq 18) = 0.0551$ $P(X \geq 19) = 0.0287$ CR $X \leq 6$ and $X \geq 19$	M1 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 $0.0916 > 0.025$; There is evidence that the probability of a faulty bolt is 0.25 or the company's claim is correct.	M1; A1ft (2)
	(e)	$H_0 : p = 0.25$ $H_1 : p < 0.25$ $P(X \leq 5) = 0.0070$ or CR $X \leq 5$ $0.007 < 0.01$, 5 is in the critical region, reject H_0, significant. There is evidence that the probability of faulty bolts has decreased	B1B1 M1A1 M1 A1ft 6) **[15]**

Notes

(a) **B1 B1** one mark for each of any of the four statements. Give first B1 if only one correct statement given. No context needed.

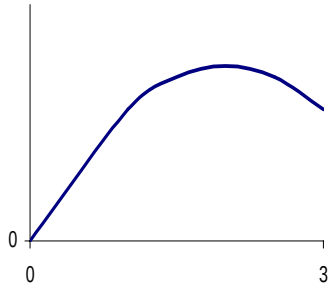
(b) **M1** for writing or using B(50,0.25) also may be implied by both CR being correct. Condone use of P in critical region for the method mark.
A1 $(X) \leq 6$ o.e. [0,6] DO NOT accept $P(X \leq 6)$
A1 $(X) \geq 19$ o.e. [19,50] DO NOT accept $P(X \geq 19)$

(c) **M1** Adding two probabilities for two tails. Both probabilities must be less than 0.5
A1 awrt 0.0481

(d) **M1** one of the given statements followed through from their CR.
A1 contextual comment followed through from their CR.
NB A correct contextual comment **alone** followed through from their CR.will get M1 A1

(e) **B1** for H_0 must use p or π (pi)
B1 for H_1 must use p or π (pi)
M1 for finding or writing $P(X \leq 5)$ or attempting to find a critical region or a correct critical region
A1 awrt 0.007/CR $X \leq 5$
M1 correct statement using their Probability and 0.01 if one tail test
or a correct statement using their Probability and 0.005 if two tail test.
The 0.01 or 0.005 needn't be explicitly seen but implied by correct statement compatible with their H_1. If no H_1 given M0
A1 correct contextual statement follow through from their prob and H_1. Need faulty bolts and decreased.
NB A correct contextual statement **alone** followed through from their prob and H_1 get M1 A1

Question Number	Scheme		Marks
Q7 (ai)	$f(y) \geq 0$ or $f(3) \geq 0$		M1
	$ky(a-y) \geq 0$ or $3k(a-3) \geq 0$ or $(a-y) \geq 0$ or $(a-3) \geq 0$		
	$a \geq 3$		A1 cso
(ii)	$\int_0^3 k(ay-y^2)dy = 1$	integration	M1
	$\left[k\left(\frac{ay^2}{2}-\frac{y^3}{3}\right)\right]_0^3 = 1$	answer correct	A1
	$k\left(\frac{9a}{2}-9\right)=1$	answer = 1	M1
	$k\left[\frac{9a-18}{2}\right]=1$		
	$k=\frac{2}{9(a-2)}$ *		A1 cso 6)
(b)	$\int_0^3 k(ay^2-y^3)\mathrm{dy}=1.75$	Int $\int xf(x)$	M1
		Correct integration	A1
	$\left[k\left(\frac{ay^3}{3}-\frac{y^4}{4}\right)\right]_0^3 = 1.75$	$\int xf(x) = 1.75$ and limits 0,3	M1dep
	$k\left(9a-\frac{81}{4}\right)=1.75$		
	$2\left(9a-\frac{81}{4}\right)=15.75(a-2)$	subst k	M1dep
	$2.25a = -31.5+\frac{81}{2}$		
	a = 4 *		A1cso
	$k=\frac{1}{9}$		B1 (6)

Question Number	Scheme	Marks
(c)		B1 B1 (2)
(d)	mode = 2	B1 (1)
		[15]

Notes

(a) (i) **M1** for putting f(y) $\geq$ 0 or f(3) $\geq$ 0 or $ky(a-y)\geq 0$ or $3k(a-3)\geq 0$ or $(a-y)\geq 0$ or $(a-3)\geq 0$ or state in words the probability can not be negative o.e.

A1 need one of $ky(a-y)\geq 0$ or $3k(a-3)\geq 0$ or $(a-y)\geq 0$ or $(a-3)\geq 0$ **and** $a\geq 3$

(ii) **M1** attempting to integrate (at least one $y^n \rightarrow y^{n+1}$) (ignore limits)

A1 Correct integration. Limits not needed. And equals 1 not needed.

M1 dependent on the previous M being awarded. Putting equal to 1 and have the correct limits. Limits do not need to be substituted.

A1 cso

(b) **M1** for attempting to find $\int yf(y)\,dy$ (at least one $y^n \rightarrow y^{n+1}$) (ignore limits)

A1 correct Integration

M1 $\int yf(y) = 1.75$ and limits 0,3 dependent on previous M being awarded

M1 subst in for k. dependent on previous M being awarded

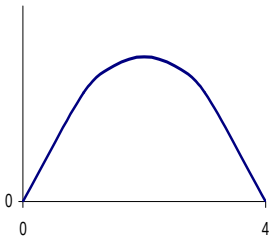
A1 cso 4

B1 cao 1/9

(c) **B1** correct shape. No straight lines. No need for patios.

B1 completely correct graph. Needs to go through origin and the curve ends at 3.

Special case: If draw full parabola from 0 to 4 get B1 B0 Allow full marks if the portion between x = 3 and x = 4 is dotted and the rest of the curve solid.



(d) **B1** cao 2