edexcel
advancing learning, changing lives

Mark Scheme (Final)
Summer 2009

GCE

GCE Statistics S2 (6684/01)

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

edexcel

June2009
6684 Statistics S2
Mark Scheme

Question Number	Scheme	Marks
Q1 (a)	$[X \sim \mathrm{B}(30, 0.15)]$	
	$\mathrm{P}(X \leq 6)$, = 0.8474 awrt 0.847	M1, A1 (2)
(b)	$Y \sim \mathrm{B}(60, 0.15) \approx \mathrm{Po}(9)$ for using Po(9)	B1
	$\mathrm{P}(Y \leq 12)$, = 0.8758 awrt 0.876	M1, A1 (3)
		(5)
	[N.B. normal approximation gives 0.897, exact binomial gives 0.894]	
(a)	M1 for a correct probability statement $\mathrm{P}(X \leq 6)$ or $\mathrm{P}(X < 7)$ or $\mathrm{P}(X=0) + \mathrm{P}(X=1) + \mathrm{P}(X=2) + \mathrm{P}(X=4) + \mathrm{P}(X=5) + \mathrm{P}(X=6)$. (may be implied by long calculation) Correct answer gets M1 A1. allow 84.74%	
(b)	B1 may be implied by using Po(9). Common incorrect answer which implies this is 0.9261 M1 for a correct probability statement $\mathrm{P}(X \leq 12)$ or $\mathrm{P}(X < 13)$ or $\mathrm{P}(X=0) + \mathrm{P}(X=1) + \ldots + \mathrm{P}(X=12)$ (may be implied by long calculation) and attempt to evaluate this probability using their Poisson distribution. Condone $\mathrm{P}(X \leq 13) = 0.8758$ for B1 M1 A1 Correct answer gets B1 M1 A1 Use of normal or exact binomial get B0 M0 A0	

Question Number	Scheme	Marks
Q2	$H_0: \lambda = 2.5$ (or $\lambda = 5$) H1: $\lambda < 2.5$ (or $\lambda < 5$) λ or μ $X \sim$ Po(5) $P(X \leq 1) = 0.0404$ or CR $X \leq 1$ [0.0404<0.05] this is significant or reject H_0 or it is in the critical region There is evidence of a <u>decrease</u> in the (mean) <u>number/rate</u> of <u>deformed blood cells</u>	B1B1 M1 A1 M1 A1 (6) **(6)**
	1st B1 for H_0 must use lambda or mu; 5 or 2.5. 2nd B1 for H_1 must use lambda or mu; 5 or 2.5 1st M1 for use of Po(5) may be implied by probability(must be used not just seen) eg. P ($X = 1$) = 0.0404 - … would score M1 A0 1st A1 for 0.0404 seen or correct CR 2nd M1 for a correct statement (this may be contextual) comparing their probability and 0.05 (or comparing 1 with their critical region). Do not allow conflicting statements. 2nd A1 is not a follow through. Need the word decrease, number or rate and deformed blood cells for contextual mark. If they have used $\neq$ in H_1 they could get B1 B0 M1 A1 M1A0 mark as above except they gain the 1st A1 for $P(X \leq 1) = 0.0404$ or CR $X \leq 0$ 2nd M1 for a correct statement (this may be contextual) comparing their probability and 0.025 (or comparing 1 with their critical region) They may compare with 0.95 (one tail method) or 0.975 (one tail method) Probability is 0.9596.	

Question Number	Scheme	Marks
Q3 (a)	*A statistic* is a function of $X_1, X_2, ...X_n$ that does not contain any unknown parameters	B1 B1 (2)
(b)	The <u>probability</u> distribution of Y or the distribution of all possible values of Y (o.e.)	B1 (1)
(c)	Identify (ii) as not a statistic Since <u>it contains</u> unknown parameters μ and σ.	B1 dB1 (2)
		(5)
(a)	**NB If you want to give one mark for their answer give the first B1g never award B0 B1** Some suggested other wording is as follows but you need to use your judgment. It may be neither statement is quite there but you feel it is worth a generous B1 B1 e.g. is a function of the sample or the data / is a quantity calculated from the sample or the data / is a random variable calculated from the sample or the data B1 e.g. does not contain any unknown parameters/quantities contains only known parameters/quantities <u>only</u> contains values of the sample Y is a function of $X_1, X_2, ...X_n$ that does not contain any unknown parameters B1B1 is a function of the values of a sample with no unknowns B1B1 is a function of the sample values B1B0 is a function of all the data values B1B0 A random variable calculated from the sample B1B0 A random variable consisting of any function B0B0 A function of a value of the sample B1B0 A function of the sample which contains no other values/ parameters B1B0	
(b)	Examples of other acceptable wording All possible values of the statistic together with their associated probabilities	
(c)	1st B1 for selecting only (ii) 2nd B1 for a reason. This is dependent upon the first B1. Need to mention at least one of mu (mean) or sigma (standard deviation or variance) or unknown parameters. Examples since it contains mu B1 since it contains sigma B1 since it contains unknown parameters/quantities B1 since it contains unknowns B0	

Question Number	Scheme	Marks
Q4 (a)	$X \sim$ B(20, 0.3)	M1
	$P(X \leq 2) = 0.0355$	A1
	$P(X \leq 9) = 0.9520$ so $P(X \geq 10) = 0.0480$	A1
	Therefore the critical region is $\{X \leq 2\} \cup \{X \geq 10\}$	A1A1 (5)
(b)	0.0355 + 0.0480 = 0.0835 awrt (0.083 or 0.084)	B1 (1)
(c)	11 is in the critical region	B1ft
	there is evidence of a <u>change/ increase</u> in the <u>proportion/number</u> of <u>customers buying single tins</u>	B1ft (2)
		(8)
(a)	M1 for B(20,0.3) seen or used 1st A1 for 0.0355 2nd A1 for 0.048 3rd A1 for $(X) \leq 2$ or $(X) < 3$ or [0,2] They get **A0** if they write $P(X \leq 2 / X < 3)$ 4th A1 $(X) \geq 10$ or $(X) > 9$ or [10,20] They get **A0** if they write $P(X \geq 10 / X > 9)$ **10** $\leq X \leq 2$ etc is accepted To describe the critical regions they can use any letter or no letter at all. It does not have to be X.	
(b)	B1 correct answer only	
(c)	1st B1 for a correct statement about 11 and their critical region. 2nd B1 for a correct comment in context consistent with their CR and the value 11 Alternative solution 1st B0 $P(X \geq 11) = 1 - 0.9829 = 0.0171$ since no comment about the critical region 2nd B1 a correct contextual statement.	

Question Number	Scheme	Marks
Q5 (a)	X = the number of errors in 2000 words so $X \sim$Po(6) P($X \geq 4$) = 1 - P($X \leq 3$) = 1 - 0.1512 = 0.8488 awrt 0.849	B1 M1 A1 (3)
(b)	Y = the number of errors in 8000 words. $Y \sim$ Po(24) so use a Normal approx $Y \approx\sim$ N(24, $\sqrt{24}^2$) Require P($Y \leq 20$) = P$\left(Z < \frac{20.5-24}{\sqrt{24}}\right)$ = P(Z < - 0.714…) = 1 – 0.7611 = 0.2389 awrt (0.237~0.239) [N.B. Exact Po gives 0.242 and no $\pm$ 0.5 gives 0.207]	M1 A1 M1 M1 A1 M1 A1 (7) **(10)**

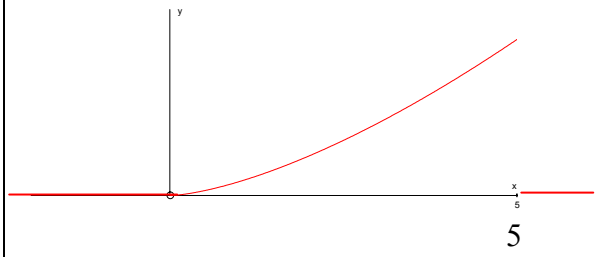
(a) B1 for seeing or using Po(6)
M1 for 1 - P($X \leq 3$) or 1 – [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]
A1 awrt 0.849

SC If B(2000, 0.003) is used and leads to awrt 0.849 allow B0 M1 A1
If no distribution indicated awrt 0.8488 scores B1M1A1 but any other awrt 0.849 scores B0M1A1

(b) 1st M1 for identifying the normal approximation
1st A1 for [mean = 24] **and** [sd = $\sqrt{24}$ or var = 24]

These first two marks may be given if the following are seen in the standardisation formula : 24
$\sqrt{24}$ or awrt 4.90

2nd M1 for attempting a continuity correction (20/ 28 $\pm$ 0.5 is acceptable)
3rd M1 for standardising using their mean and their standard deviation.
2nd A1 correct z value awrt $\pm$0.71 **or** this may be awarded if see $\frac{20.5-24}{\sqrt{24}}$ or $\frac{27.5-24}{\sqrt{24}}$
4th M1 for 1 - a probability from tables (must have an answer of < 0.5)
3rd A1 answer awrt 3 sig fig in range 0.237 – 0.239

Question Number	Scheme	Marks
Q6 (a)	$P(A>3)=\frac{2}{5}=0.4$	B1 (1)
(b)	$(0.4)^3$,$=0.064$ or $\frac{8}{125}$	M1, A1 (2)
(c)	$f(y)=\frac{d}{dy}(F(y)=\begin{cases}\frac{3y^2}{125} & 0\le y\le 5\\ 0 & otherwise\end{cases}$	M1A1 (2)
(d)	 Shape of curve and start at (0,0) Point (5, 0) labelled and curve between 0 and 5 and pdf ≥ 0	B1 B1 (2) B1 (1)
(e)	Mode = 5	M1M1A1 (3)
(f)	$E(Y)=\int_0^5\left(\frac{3y^3}{125}\right)dy=\left[\frac{3y^4}{500}\right]_0^5=\frac{15}{4}$ or 3.75	M1A1 (2)
(g)	$P(Y>3)=\begin{cases}\int_3^5\frac{3y^2}{125}\,dy\\ \text{or } 1-F(3)\end{cases}=1-\frac{27}{125}=\frac{98}{125}=0.784$	**(13)**
(a)	B1 correct answer only(cao). Do not ignore subsequent working	
(b)	M1 for cubing their answer to part (a) A1 cao	
(c)	M1 for attempt to differentiate the cdf. They must decrease the power by 1 A1 fully correct answer including 0 otherwise. Condone < signs	
(d)	B1 for shape. Must curve the correct way and start at (0,0). No need for y = 0 (patios) lines B1 for point (5,0) labelled and pdf only existing between 0 and 5, may have y=0 (patios) for other values	
(e)	B1 cao	
(f)	1st M1 for attempt to integrate their $yf(y)$ $y^n \to y^{n+1}$. 2nd M1 for attempt to use correct limits A1 cao	
(g)	M1 for attempt to find $P(Y>3)$. e.g. writing $\int_3^5$ *their* $f(y)$ must have correct limits or writing 1 – F(3)	

Question Number	Scheme	Marks
Q7 (a)	$E(X) = 2$ (by symmetry)	B1 (1)
(b)	$0 \leq x < 2$, gradient $= \frac{\frac{1}{2}}{2} = \frac{1}{4}$ and equation is $y = \frac{1}{4}x$ so $a = \frac{1}{4}$	B1
	$b - \frac{1}{4}x$ passes through (4, 0) so $b = 1$	B1 (2)
(c)	$E(X^2) = \int_0^2 \left(\frac{1}{4}x^3\right) dx + \int_2^4 \left(x^2 - \frac{1}{4}x^3\right) dx$	M1M1
	$= \left[\frac{x^4}{16}\right]_0^2 + \left[\frac{x^3}{3} - \frac{x^4}{16}\right]_2^4$	A1
	$= 1 + \frac{64-8}{3} - \frac{256-16}{16}$ = $4\frac{2}{3}$ or $\frac{14}{3}$	M1A1
	$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{14}{3} - 2^2$, $= \frac{2}{3}$ (so $\sigma = \sqrt{\frac{2}{3}} = 0.816$) (*)	M1 A1cso (7)
(d)	$P(X \leq q) = \int_0^q \frac{1}{4}x \, dx = \frac{1}{4}$, $\frac{q^2}{2} = 1$ so $q = \sqrt{2} = 1.414$ awrt 1.41	M1A1,A1 (3)
(e)	2- σ = 1.184 so 2 - σ, 2 + σ is wider than IQR, therefore greater than 0.5	M1,A1 (2) **(15)**

(a) B1 cao

(b) B1 for value of a. B1 for value of b

(c) 1st M1 for attempt at $\int ax^3$ using their a. For attempt they need x^4. Ignore limits.

2nd M1 for attempt at $\int bx^2 - ax^3$ use their a and b. For attempt need to have either x^3 or x^4. Ignore limits

1st A1 correct integration for both parts

3rd M1 for use of the correct limits on each part

2nd A1 for either getting 1 and $3\frac{2}{3}$ or awrt 3.67 somewhere or $4\frac{2}{3}$ or awrt 4.67

4th M1 for use of $E(X^2) - [E(X)]^2$ must add both parts for E (X^2) and only have subtracted the mean^2 once. You must see this working

(d) 3rd A1 $\sigma = \sqrt{\frac{2}{3}}$ or $\sqrt{0.66667}$ or better with no incorrect working seen.

M1 for attempting to find LQ, integral of either part of f(x) with their 'a' and 'b' = 0.25

Or their F(x) = 0.25 i.e. $\frac{ax^2}{2} = 0.25$ or $bx - \frac{ax^2}{2} + 4a - 2b = 0.25$ with their a and b

If they add both parts of their F(x), then they will get M0.

1st A1 for a correct equation/expression using their 'a'

(e) 2nd A1 for $\sqrt{2}$ or awrt 1.41

M1 for a reason based on their quartiles

- Possible reasons are P(2 - σ<X< 2 + σ)= 0.6498 allow awrt 0.65
- 1.184 < LQ(1.414)

A1 for correct answer > 0.5

NB you must check the reason and award the method mark. A correct answer without a correct reason gets M0 A0

Question Number	Scheme	Marks
Q8 (a)	$X \sim \text{Po}(2)$ $\quad \text{P}(X=4) = \dfrac{\text{e}^{-2}\times 2^4}{4!} = 0.0902$ awrt 0.09	M1 A1 (2)
(b)	$Y \sim \text{Po}(8)$ $\text{P}(Y>10) = 1 - \text{P}(Y \leq 10) = 1 - 0.8159 = 0.18411\ldots$ awrt 0.184	B1 M1A1 (3)
(c)	F = no. of faults in a piece of cloth of length x $\quad F \sim \text{Po}(x \times \frac{2}{15})$ $\text{e}^{-\frac{2x}{15}} = 0.80$ $\text{e}^{-\frac{2}{15}\times 1.65} = 0.8025\ldots, \quad \text{e}^{-\frac{2}{15}\times 1.75} = 0.791\ldots$ These values are either side of 0.80 therefore x = 1.7 to 2 sf	M1A1 M1 A1 (4)
(d)	Expected number with no faults $= 1200\times 0.8 = 960$ Expected number with some faults $= 1200\times 0.2 = 240$ So expected profit $= 960\times 0.60 - 240\times 1.50, \quad = £216$	M1 A1 M1, A1 (4) **(13)**
(a)	M1 for use of Po(2) may be implied A1 awrt 0.09	
(b)	B1 for Po(8) seen or used M1 for 1 - P($Y \leq 10$) oe A1 awrt 0.184	
(c)	1st M1 for forming a suitable Poisson distribution of the form $\text{e}^{-\lambda} = 0.8$ 1st A1 for use of lambda as $\dfrac{2x}{15}$ (this may appear after taking logs) 2nd M1 for attempt to consider a range of values that will prove 1.7 is correct **OR** for use of logs to show lambda = … 2nd A1 correct solution only. Either get 1.7 from using logs or stating values either side	
S.C	for $\text{e}^{-\frac{2}{15}\times 1.7} = 0.797\ldots \approx 0.80 \quad \therefore x = 1.7$ to 2 sf allow 2nd M1A0	
(d)	1st M1 for one of the following 1200 p or 1200 (1 – p) where p = 0.8 or 2/15. 1st A1 for both expected values being correct or two correct expressions. 2nd M1 for an attempt to find expected profit, must consider with and without faults 2nd A1 correct answer only.	