edexcel
advancing learning, changing lives

Mark Scheme (Results)

Summer 2009

GCE

GCE Mathematics (6691/01)

A PEARSON COMPANY

June 2009
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
Q1 (a)	Randomly select a number between 00 and 499 (001 and 500) select every 500th person	B1 B1 (2)
(bi)	Quota **Advantage:** Representative sample can be achieved (with small sample size) Cheap (costs kept to a minimum) **not** "quick" Administration relatively easy **Disadvantage** Not possible to estimate sampling errors (due to lack of randomness) Not a random process Judgment of interviewer can affect choice of sample – bias Non-response not recorded Difficulties of defining controls e.g. social class	B1 B1 (2)
(bii)	Systematic **Advantage:** Simple or easy to use **not** "quick" or "cheap" or "efficient" It is suitable for large samples (not populations) **Disadvantage** Only random if the ordered list is (truly) random Requires a list of the population or must assign a number to each member of the pop.	B1 B1 (2) **[6]**
(a)	1st B1 for idea of using random numbers to select the first from1 - 500 (o.e.) 2nd B1 for selecting every 500th (name on the list) If they are clearly trying to carry out stratified sample then score B0B0	
(b)	Score B1 for any one line	
(i)	1st B1 for Quota advantage 2nd B1 for Quota disadvantage	
(ii)	3rd B1 for Systematic Advantage 4th B1 for Systematic Disadvantage	

Question Number	Scheme	Marks
Q2 (a)	Limits are $20.1 \pm 1.96 \times 0.5$ **<u>(19.1, 21.1)</u>**	M1 B1 A1cso (3)
(b)	98 % confidence limits are $20.1 \pm 2.3263 \times \frac{0.5}{\sqrt{10}}$ **<u>(19.7, 20.5)</u>**	 M1 B1 A1A1 (4)
(c)	The growers claim is not correct Since 19.5 does not lie in the interval (19.7, 20.5)	B1 dB1 (2) **[9]**
(a)	M1 for $20.1 \pm z \times 0.5$. Need 20.1 and 0.5 in correct places with no $\sqrt{10}$ B1 for z = 1.96 (or better) A1 for awrt 19.1 <u>and</u> awrt 21.1 **but must have scored both M1 and B1** [Correct answer only scores 3/3]	
(b)	M1 for $20.1 \pm z \times \frac{0.5}{\sqrt{10}}$, need to see 20.1, 0.5 and $\sqrt{10}$ in correct places B1 for z = 2.3263 (or better) 1st A1 for awrt 19.7 2nd A1 for awrt 20.5 [Correct answer only scores M1B0A1A1]	
(c)	1st B1 for rejection of the claim. Accept "unlikely" or "not correct" 2nd dB1 Dependent on scoring 1st B1 in this part for rejecting grower's claim for an argument that supports this. Allow comment on <u>their</u> 98% CI from (b)	

Question Number	Scheme	Marks
Q3 (a)		
	| | A | B | C | D | E | F | G | H | I | J | | BMI | 1 | 6 | 3 | 8 | 4 | 5 | 7 | 2 | 9 | 10 | | or | 10 | 5 | 8 | 3 | 7 | 6 | 4 | 9 | 2 | 1 | | Finishing position | 3 | 5 | 1 | 9 | 6 | 4 | 10 | 2 | 7 | 8 | | d^2 | 4 | 1 | 4 | 1 | 4 | 1 | 9 | 0 | 4 | 4 |	M1
	$\sum d^2 = 32$ (298)	M1
	$r_s = 1 - \frac{6\times 32}{10\times 99}$	M1 A1ft
	$= 0.80606\ldots$ (-0.80606) accept $\pm\frac{133}{165}$ **awrt $\pm$ 0.806**	A1 (5)
(b)	$H_0 : \rho = 0$, $H_1 : \rho > 0$,	B1 B1
	Critical value is $(\pm)0.5636$	B1
	(0.806 > 0.5636 therefore) in critical region/ reject H_0	M1
	The lower the BMI the higher the position in the race./ support for doctors belief	A1ft (5)
(c)	The position is already ranked OR Position is not Normally distributed	B1 (1) **[11]**
(a)	1st M1 for attempt to rank BMI scores	
	2nd M1 for attempt at $\sum d^2$ (<u>must</u> be using ranks)	
	3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.	No ranking can score 3rd M1 only
	1st A1ft for a correct expression. ft their $\sum d^2$ but only if all 3 Ms are scored	
	2nd A1 awrt $\pm$ 0.806 (but sign must be compatible with their $\sum d^2$)	
(b)	2nd B1 for $\rho > 0$ (or <0 but must be one tail and consistent with their ranking)	
	3rd B1 for critical value that is compatible with their H_1. If one-tail must be $\pm$ 0.5636 if two-tail must be $\pm$ 0.6485 [Condone wrong sign]	No H_1 assume one-tail for 3rd B1
	M1 for a correct statement relating their r_s with their cv. e.g. "reject H_0", "in critical region", "significant result" May be implied by a correct comment	
	A1ft for correct comment in context. Must mention low/high BMI and race/fitness <u>or</u> doctor's belief. Comment should be <u>one</u>-tailed. Allow positive <u>correlation</u> between… but <u>NOT</u> …positive <u>relationship</u>…	
(c)	B1 for a correct and relevant comment either based on the fact that the data was originally partially ordered <u>or</u> on the underlying normal assumption "Quicker" or "easier" score B0	

Question Number	Scheme	Marks
Q4	$X \sim \text{N}(55, 3^2)$ therefore $\overline{X} \sim \text{N}(55, \frac{9}{8})$	B1 B1
	$\text{P}(\overline{X} > 57) = \text{P}\left(Z > \frac{57-55}{\sqrt{\frac{9}{8}}}\right) \quad = \text{P}(Z > 1.8856\ldots)$	M1
	$= 1 - 0.9706$	M1
	$= 0.0294$ **<u>0.0294~0.0297</u>**	A1
		[5]
	1st B1 for $\overline{X}$ ~ normal and $\mu = 55$, may be implied but must be $\overline{X}$ 2nd B1 for $\text{Var}(\overline{X})$ or st. dev of $\overline{X}$ e.g. $\overline{X} \sim \text{N}(55, \frac{9}{8})$ or $\overline{X} \sim \text{N}\left(55, \left(\frac{3}{\sqrt{8}}\right)^2\right)$ for B1B1 Condone use of X if they clearly mean $\overline{X}$ so $X \sim \text{N}\left(55, \frac{9}{8}\right)$ is OK for B1B1 1st M1 for an attempt to standardize with 57 and mean of 55 and their st. dev. $\neq 3$ 2nd M1 for 1 - tables value. Must be trying to find a probability < 0.5 A1 for answers in the range 0.0294~0.0297	
ALT	$\sum_{1}^{8} X_i \sim \text{N}\left(8 \times 55, 8 \times 3^2\right)$ 1st B1 for $\sum X$ ~ normal and mean = 8×55 2nd B1 for variance = 8×3^2 1st M1 for attempt to standardise with 57×8, mean of 55×8 and their st dev $\neq 3$	

<table>
<tr><th>Question Number</th><th>Scheme</th><th>Marks</th></tr>
<tr><td>Q5 (a)

(b)

(c)</td><td>$\lambda = \frac{0\times40+1\times33+2\times14+3\times8+4\times5}{100} = 1.05$

Using Expected frequency = 100 × P(X = x) = 100 × $\frac{e^{-1.05}1.05^x}{x!}$ gives
r = 36.743 awrt 36.743 or 36.744
s = 19.290 19.29 or awrt 19.290

H_0 : Poisson distribution is a suitable model
H_1 : Poisson distribution is not a suitable model

<table>
<tr><th>Number of goals</th><th>Frequency</th><th>Expected frequency</th><th></th></tr>
<tr><td>0</td><td>40</td><td>34.994</td><td></td></tr>
<tr><td>1</td><td>33</td><td>36.743</td><td></td></tr>
<tr><td>2</td><td>14</td><td>19.290</td><td></td></tr>
<tr><td>3</td><td>8</td><td>6.752</td><td rowspan="2">8.972443</td></tr>
<tr><td>≥ 4</td><td>5</td><td>2.221</td></tr>
</table>

$\upsilon = 4 - 1 - 1 = 2$
CR : $\chi_2^2(0.05) > 5.991$
$\sum\frac{(O-E)^2}{E} = \frac{(40-34.9937)^2}{34.9937} + \ldots + \frac{(13-8.972443)^2}{8.972443}$
[=0.7161…+0.3813…+1.4508…+1.80789..]
= 4.356. (ans in range 4.2 – 4.4)
Not in critical region
Number of goals scored can follow a Poisson distribution / managers claim is justified</td><td>M1 A1 (2)

M1
A1
A1 (3)

B1

M1

B1ft
B1

M1

A1

A1 ft (7)

[12]</td></tr>
<tr><td>(a)

(b)

(c)</td><td>M1 for an attempt to find the mean- at least 2 terms on numerator seen
Correct answer only will score both marks

M1 for use of correct formula (ft their mean). 1st A1 for r , 2nd A1 for s (19.29 OK)

1st B1 Must have both hypotheses and mention Poisson at least once
inclusion of their value for mean in hypotheses is B0 but condone in conclusion
1st M1 for an attempt to pool ≥ 4
2nd B1ft for n -1-1 = 2 i.e realising that they must subtract 2 from their n
3rd B1 for 5.991 only
2nd M1 for an attempt at the test statistic, at least 2 correct expressions/values (to 3sf)
1st A1 for answers in the range 4.2~4.4
2nd A1 for correct comment in context based on their test statistic and their cv that mentions goals or manager. Dependent on 2nd M1
Condone mention of Po(1.05) in conclusion
Score A0 for inconsistencies e.g. “significant” followed by “manager’s claim is justified”</td><td></td></tr>
</table>

Question Number	Scheme	Marks
Q6 (a)	μ_U ~ mean length of upper shore limpets, μ_L ~ mean length of lower shore limpets $H_0 : \mu_u = \mu_L$ $H_1 : \mu_u < \mu_L$ both	B1
	s.e. = $\sqrt{\frac{0.42^2}{120} + \frac{0.67^2}{150}}$ = 0.0668	M1 A1
	$z = \frac{5.05 - 4.97}{0.0668} = (\pm)1.1975$ awrt $\pm$ **<u>1.20</u>**	dM1 A1
	Critical region is $z \geq 1.6449$, or probability = awrt (0.115 or 0.116) $z = \pm\ 1.6449$	B1
	($1.1975 < 1.6449$) therefore not in critical region / accept H_0/not significant (or $P(Z \geq 1.1975) = 0.1151,\ 0.1151 > 0.05$ or z not in critical region)	M1
	There is no evidence that the limpets on the upper shore are shorter than the limpets on the lower shore.	A1 (8)
(b)	Assume the populations or variables are independent Standard deviation of sample = standard deviation of population [Mention of <u>Central Limit Theorem</u> does <u>NOT</u> score the mark]	B1 B1 (2) **[10]**
(a)	1st B1 If μ_1, μ_2 used then it must be clear which refers to upper shore. Accept sensible choice of letters such as *u* and *l*. 1st M1 Condone minor slips e.g. $\frac{0.67^2}{120}$ or $\frac{0.67}{150} + \frac{0.42^2}{120}$ etc i.e. swapped *n* or one sd and one variance but M0 for $\sqrt{\frac{0.67}{150} + \frac{0.42}{120}}$ 1st A1 can be scored for a fully correct expression. May be implied by awrt 1.20 2nd dM1 is dependent upon the 1st M1 but can ft their se value if this mark is scored. 2nd A1 for awrt ($\pm$) 1.20 3rd M1 for a correct statement based on their *z* value and their cv. No cv is M0A0 If using probability they must compare their *p* (<0.5) with 0.05 (o.e) so can allow 0.884< 0.95 to score this 3rd M1 mark. May be implied by their contextual statement and M1A0 is possible. 3rd A1 for a correct comment to accept null hypothesis that mentions <u>length</u> of <u>limpets</u> on the two <u>shores</u>.	
(b)	1st B1 for one correct statement. Accept "samples are independent" 2nd B1 for both statements	

Question Number	Scheme	Marks
Q7 (a)	Estimate of Mean $= \dfrac{600.9}{5} = 120.18$	M1A1
	Estimate of Variance $= \frac{1}{4}\left\{ 72216.31 - \dfrac{600.9^2}{5} \right\}$ or $\dfrac{0.148}{4} = 0.037$	M1 A1ft A1 (5)
(b)	P(-0.05 < $\mu - \hat{\mu}$ < 0.05) = 0.90 or $P(-0.05 < \overline{X} - \mu < 0.05) = 0.90$ [$\leq$ is OK]	B1
	$\dfrac{0.05}{\frac{0.2}{\sqrt{n}}} = 1.6449$	M1 A1
	$n = \dfrac{1.6449^2 \times 0.2^2}{0.05^2}$	dM1
	$n = 43.29\ldots$	A1
	$n = 44$	A1 (6) **[11]**
(a)	1st M1 for an attempt at $\sum x$ (accept 600 to 1sf) 1st A1 for $\dfrac{600.9}{5}$ = awrt 120 or awrt 120.2. No working give M1A1 for awrt 120.2 2nd M1 for the use of a correct formula including a reasonable attempt at $\sum x^2$ (Accept 70 000 to 1sf) or $\sum\left(x - \overline{x}\right)^2 = 0.15$ (to 2 dp) 2nd A1ft for a correct expression with correct $\sum x^2$ but can ft their <u>mean</u> (for expression - no need to check values if it is incorrect) 3rd A1 for 0.037 Correct answer with no working scores 3/3 for variance	
(b)	B1 for a correct probability statement <u>or</u> "width of 90% CI = $0.05 \times 2 = 0.1$" 1st M1 for $\dfrac{0.05}{\frac{0.2}{\sqrt{n}}}$ = z value <u>or</u> $2 \times \dfrac{0.2}{\sqrt{n}} \times z = 0.1$ Condone 0.5 instead of 0.05 <u>or</u> missing 2 <u>or</u> 0.05 for 0.1 for M1 1st A1 for a correct equation including 1.6449 2nd dM1 Dependent upon 1st M1 for rearranging to get $n = \ldots$ Must see "squaring" 2nd A1 for n = awrt 43.3 3rd A1 for rounding up to get $n = 44$ Using e.g.1.645 instead of 1.6449 can score all the marks except the 1st A1	1st B1 may be implied by 1st A1 scored or correct equation.

Question Number	Scheme	Marks
Q8 (a)	$E(4X - 3Y) = 4E(X) - 3E(Y)$	M1
	$= 4\times 30 - 3\times 20$	
	$= 60$	A1 (2)
(b)	$Var(4X - 3Y) = 16\ Var(X) + 9\ Var(Y)$ 16 or 9; adding	M1; M1
	$= 16 \times 9 + 9 \times 4$	
	$= 180$	A1 (3)
(c)	$E(B) = 80$	B1
	$Var(B) = 16$	B1
	$E(B - A) = 20$ E(B)-E(A)	M1
	$Var(B - A) = 196$ ft on 180 and 16	A1ft
	$P(B - A > 0) = P\left(Z > \frac{-20}{\sqrt{196}}\right) = [P(Z > -1.428...)]$ stand. using their mean and var	dM1
	$= 0.923 \ldots$ awrt 0.923 – 0.924	A1 (6) **[11]**
(a)	M1 for correct use of $E(aX + bY)$ formula	
(b)	1st M1 for $16Var(X)$ or $9Var(Y)$ 2nd M1 for adding variances Key points are the 16, 9 and +. Allow slip e.g using $Var(X)=4$ etc to score Ms	
(c)	1st M1 for attempting $B - A$ and $E(B - A)$ or $A - B$ and $E(A - B)$ This mark may be implied by an attempt at a correct probability e.g. $P\left(Z > \frac{0-(80-60)}{\sqrt{180+16}}\right)$. To be implied we must see the "0" 1st A1ft for $Var(B - A)$ can ft their $Var(A) = 180$ and their $Var(B) = 16$ 2nd dM1 Dependent upon the 1st M1 in part (c). for attempting a correct probability i.e. $P(B-A>0)$ or $P(A-B < 0)$ and standardising with their mean and variance. They must standardise properly with the 0 to score this mark 2nd A1 for awrt 0.923 ~ 0.924	