Mark Scheme 4767
June 2005

GENERAL INSTRUCTIONS

Marks in the mark scheme are explicitly designated as **M**, **A**, **B**, **E** or **G**.

M marks ("method") are for an attempt to use a correct method (not merely for stating the method).

A marks ("accuracy") are for accurate answers and can only be earned if corresponding **M** mark(s) have been earned. Candidates are expected to give answers to a sensible level of accuracy in the context of the problem in hand. The level of accuracy quoted in the mark scheme will sometimes deliberately be greater than is required, when this facilitates marking.

B marks are independent of all others. They are usually awarded for a single correct answer. Typically they are available for correct quotation of points such as 1.96 from tables.

E marks ("explanation") are for explanation and/or interpretation. These will frequently be sub divisible depending on the thoroughness of the candidate's answer.

G marks ("graph") are for completing a graph or diagram correctly.

- Insert part marks in **right-hand** margin in line with the mark scheme. For fully correct parts tick the answer. For partially complete parts indicate clearly in the body of the script where the marks have been gained or lost, in line with the mark scheme.

- Please indicate incorrect working by ringing or underlining as appropriate.

- Insert total in **right-hand** margin, ringed, at end of question, in line with the mark scheme.

- Numerical answers which are not exact should be given to at least the accuracy shown. Approximate answers to a greater accuracy *may* be condoned.

- Probabilities should be given as fractions, decimals or percentages.

- FOLLOW-THROUGH MARKING SHOULD NORMALLY BE USED WHEREVER POSSIBLE. There will, however, be an occasional designation of '**c.a.o.**' for "correct answer only".

- Full credit MUST be given when correct alternative methods of solution are used. If errors occur in such methods, the marks awarded should correspond as nearly as possible to equivalent work using the method in the mark scheme.

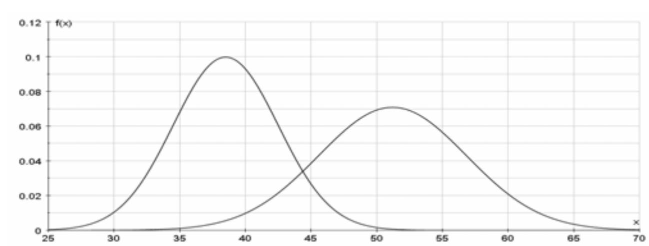
- The following notation should be used where applicable:

Question 1

(i)	Uniform average rate of occurrence; Successive arrivals are independent. Suitable arguments for/against each assumption: Eg Rate of occurrence could vary depending on the weather (any reasonable suggestion)	E1,E1 for suitable assumptions E1, E1 must be in context	**4**
(ii)	Mean $= \dfrac{\Sigma xf}{n} = \dfrac{39+40+36+32+15}{100} = \dfrac{162}{100} = 1.62$ Variance $= \dfrac{1}{n-1}\left(\Sigma fx^2 - n\bar{x}^2\right)$ $= \dfrac{1}{99}\left(430 - 100\times 1.62^2\right) = 1.69$ (to 2 d.p.)	B1 for mean ***NB answer given*** M1 for calculation A1	**3**
(iii)	Yes, since mean is close to variance	B1FT	**1**
(iv)	$\mathrm{P}(X=2) = \mathrm{e}^{-1.62}\dfrac{1.62^2}{2!}$ $= 0.260$ (3 s.f.) *Either:* Thus the expected number of 2's is 26 which is reasonably close to the observed value of 20. *Or*: This probability compares reasonably well with the relative frequency 0.2	M1 for probability calc. M0 for tables unless interpolated A1 B1 for expectation of 26 or r.f. of 0.2 E1	**4**
(v)	$\lambda = 5\times 1.62 = 8.1$ Using tables: $\mathrm{P}(X\geq 10) = 1 - \mathrm{P}(X \leq 9)$ $= 1 - 0.7041 = 0.2959$	B1FT for mean (SOI) M1 for probability from using tables to find $1 - \mathrm{P}(X \leq 9)$ A1 FT	**3**
(vi)	Mean no. of items in 1 hour $= 360 \times 1.62 = 583.2$ Using Normal approx. to the Poisson, $X \sim \mathrm{N}(583.2, 583.2)$: $\mathrm{P}(X \leq 550.5) = \mathrm{P}\left(Z \leq \dfrac{550.5 - 583.2}{\sqrt{583.2}}\right)$ $= \mathrm{P}(Z \leq -1.354) = 1 - \Phi(1.354) = 1 - 0.9121$	B1 for Normal approx. with correct parameters (SOI) B1 for continuity corr. M1 for probability	**4**

	$= 0.0879$ (3 s.f.)	using correct tail A1 CAO, (but FT wrong or omitted CC)	
			19

Question 2

(i)	$X \sim \mathrm{N}(38.5, 16)$ $\mathrm{P}(X > 45) = \mathrm{P}\left(Z > \dfrac{45 - 38.5}{4}\right)$ $= \mathrm{P}(Z > 1.625)$ $= 1 - \Phi(1.625) = 1 - 0.9479$ $= 0.0521$ (3 s.f.) *or* 0.052 (to 2 s.f.)	M1 for standardizing A1 for 1.625 M1 for prob. with tables and correct tail A1 CAO (min 2 s.f.)	**4**
(ii)	From tables $\Phi^{-1}(0.90) = 1.282$ $\dfrac{x - 38.5}{4} = -1.282$ $x = 38.5 - 1.282 \times 4 = 33.37$ So 33.4 should be quoted	B1 for 1.282 seen M1 for equation in x and negative z-value A1 CAO	**3**
(iii)	$Y \sim \mathrm{N}(51.2, \sigma^2)$ From tables $\Phi^{-1}(0.75) = 0.6745$ $\dfrac{55 - 51.2}{\sigma} = 0.6745$ $3.8 = 0.6745\,\sigma$ $\sigma = 5.63$	B1 for 0.6745 seen M1 for equation in σ with z-value A1 ***NB answer given***	**3**
(iv)		G1 for shape G1 for means, shown explicitly or by scale G1 for lower max height in diesel G1 for higher variance in diesel	**4**
(v)	$\mathrm{P}(\text{Diesel} > 45) = \mathrm{P}\left(Z > \dfrac{45 - 51.2}{5.63}\right)$	M1 for prob. calc. for diesel	

	$= \text{P}(Z > -1.101) = \Phi(1.101) = 0.8646$ P(At least one over 45) = 1 – P(Both less than 45) = 1 - (1 - 0.0521) x (1 - 0.8646) = 1 - 0.9479 x 0.1354 = 0.8717 NB allow correct alternatives based on: P(D over, P under)+P(D under, P over)+ P(both over) *or* P(D over) + P(P over) – P(both over)	M1 for correct structure M1*dep* for correct probabilities A1 CAO (2 s.f. min)	**4**
			18

Question 3

(i)	$\bar{x} = 4.5,\ \ \bar{y} = 26.85$ $b = \dfrac{Sxy}{Sxx} = \dfrac{983.6 - 36 \times 214.8/8}{204 - 36^2/8} = \dfrac{17}{42} = 0.405$ OR $b = \dfrac{983.6/8 - 4.5 \times 26.85}{204/8 - 4.5^2} = \dfrac{2.125}{5.25} = 0.405$ hence least squares regression line is: $y - \bar{y} = b(x - \bar{x})$ $\Rightarrow\ y - 26.85 = 0.405(x - 4.5)$ $\Rightarrow\ y = 0.405x + 25.03$	B1 for $\bar{x}$ and $\bar{y}$ used (SOI) M1 for attempt at gradient (b) A1 for 0.405 **cao** M1 *indep* for equation of line A1FT for complete equation	**5**
(ii)	$x = 4 \Rightarrow$ predicted $y = 0.405 \times 4 + 25.03 = 26.65$ Residual = 27.5 – 26.65 = 0.85	M1 for prediction A1FT for ± 0.85 B1FT for sign (+)	**3**
(iii)	The new equation would be preferable, since the equation in part (i) is influenced by the unrepresentative point (4,27.5)	B1 E1	**2**
(iv)	H_0: $\rho = 0$; H_1: $\rho > 0$ where ρ represents the population correlation coefficient Critical value at 5% level is 0.3783 Since 0.209 < 0.3783, there is not sufficient evidence to reject H_0, i.e. there is not sufficient evidence to conclude that there is any correlation between cycling and swimming times.	B1 for H_0 and H_1 B1 for defining ρ B1 for 0.3783 M1 for comparison leading to conclusion A1*dep on cv* for conclusion in words	**5**

		in context	
(v)	Underlying distribution must be bivariate normal. The distribution of points on the scatter diagram should be approximately elliptical.	B1 E1	**2**
			17

Question 4

(a) (i)	H_0: $\mu = 166500$; H_1: $\mu > 166500$ Where μ denotes the mean selling price in pounds of the population of houses on the large estate	B1 for both correct B1 for definition of μ	**2**
(ii)	n = 6, $\Sigma x = 1018500$, $\bar{x} = £169750$ Test statistic = $\dfrac{169750 - 166500}{14200/\sqrt{6}} = \dfrac{3250}{5797}$ $= 0.5606$ 5% level 1 tailed critical value of z = 1.645 0.5606 < 1.645 so not significant. There is insufficient evidence to reject H_0 It is reasonable to conclude that houses on this estate are not more expensive than in the rest of the suburbs.	B1CAO M1 must include √6 A1FT B1 for 1.645 M1 for comparison leading to a conclusion A1 for conclusion in words in context	**6**

<table>
<tr>
<td>(b)</td>
<td>
H_0: no association between customer and drink types;
H_1: some association between customer and drink types;

<table>
<tr><td colspan="2" rowspan="2">Observed</td><td colspan="2">Type of drink</td><td rowspan="2">Row totals</td></tr>
<tr><td>Alcoholic</td><td>Soft drinks</td></tr>
<tr><td rowspan="3">Type of customer</td><td>Business</td><td>54</td><td>63</td><td>117</td></tr>
<tr><td>Tourist</td><td>95</td><td>41</td><td>136</td></tr>
<tr><td>Local</td><td>71</td><td>76</td><td>147</td></tr>
<tr><td colspan="2">Column totals</td><td>220</td><td>180</td><td>400</td></tr>
</table>

<table>
<tr><td colspan="2" rowspan="2">Expected</td><td colspan="2">Type of drink</td><td rowspan="2">Row totals</td></tr>
<tr><td>Alcoholic</td><td>Soft drinks</td></tr>
<tr><td rowspan="3">Type of customer</td><td>Business</td><td>64.35</td><td>52.65</td><td>117</td></tr>
<tr><td>Tourist</td><td>74.80</td><td>61.20</td><td>136</td></tr>
<tr><td>Local</td><td>80.85</td><td>66.15</td><td>147</td></tr>
<tr><td colspan="2">Column totals</td><td>220</td><td>180</td><td>400</td></tr>
</table>

<table>
<tr><td colspan="2" rowspan="2">Chi squared contribution</td><td colspan="2">Type of drink</td><td rowspan="2">Row totals</td></tr>
<tr><td>Alcoholic</td><td>Soft drinks</td></tr>
<tr><td rowspan="3">Type of customer</td><td>Business</td><td>1.665</td><td>2.035</td><td>3.699</td></tr>
<tr><td>Tourist</td><td>5.455</td><td>6.667</td><td>12.122</td></tr>
<tr><td>Local</td><td>1.200</td><td>1.467</td><td>2.667</td></tr>
</table>

$X^2 = 18.49$

Refer to χ_2^2
Critical value at 5% level = 5.991
Result is significant
There is some association between customer type and type of drink.
NB if H_0 H_1 reversed, or ‘correlation’ mentioned, do not award first B1or final B1 or final E1
</td>
<td>
B1

M1 A1 for expected values (to 2dp)

M1 for valid attempt at $(O-E)^2/E$

M1dep for summation

A1CAO for X^2

B1 for 2 deg of f
B1 CAO for cv
B1dep on cv
E1
</td>
<td>6

4</td>
</tr>
<tr><td></td><td></td><td></td><td>18</td></tr>
</table>