RECOGNISING ACHIEVEMENT

GCE

# Mathematics (MEI)

Advanced GCE

Unit **4768:** Statistics 3

# Mark Scheme for June 2011

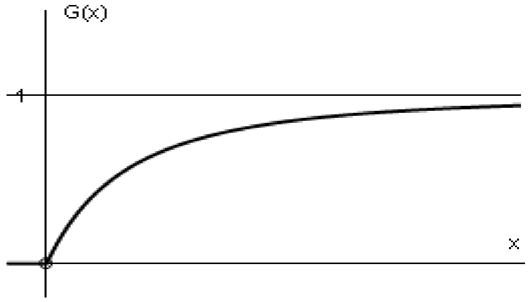
| Q1 | | | | |
|---|---|---|---|---|
| (i) | $t$ test might be used because<br>• population variance is unknown<br>• background population is Normal | E1<br>E1 | Allow "sample is small" as an alternative. | **2** |
| (ii) | $H_0: \mu = 15.3$<br>$H_1: \mu < 15.3$ | B1 | Both hypotheses. Hypotheses in words only must include "population". Do NOT allow " $\bar{X} = \ldots$ " or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean. | |
| | where $\mu$ is the mean of Gerry's times. | B1 | For adequate verbal definition. Allow absence of "population" if correct notation $\mu$ is used. | |
| | $\bar{x} = 14.987$ $\quad s_{n-1} = 0.4567(5)$ | B1 | $s_n = 0.4333$ but do NOT allow this here or in construction of test statistic, but FT from there. | |
| | Test statistic is $\dfrac{14.987 - 15.3}{\frac{0.45675}{\sqrt{10}}}$ | M1 | Allow c's $\bar{x}$ and/or $s_{n-1}$.<br>Allow alternative: 15.3 + (c's –1.833) $\times \dfrac{0.45675}{\sqrt{10}}$ (= 15.035) for subsequent comparison with $\bar{x}$ .<br>(Or $\bar{x}$ – (c's –1.833) $\times \dfrac{0.45675}{\sqrt{10}}$ (= 15.252) for comparison with 15.3.) | |
| | = –2.167(0). | A1 | c.a.o. but ft from here in any case if wrong.<br>Use of $\mu - \bar{x}$ scores M1A0, but ft. | |
| | Refer to $t_9$.<br>Single-tailed 5% point is –1.833. | M1<br>A1 | No ft from here if wrong.<br>Must be minus 1.833 unless absolute values are being compared. No ft from here if wrong.<br>P($t$ < –2.167(0)) = 0.0292. | |
| | Significant.<br>Seems that Gerry's times have been reduced on average. | A1<br>A1 | ft only c's test statistic.<br>ft only c's test statistic. Conclusion in context to include "average" o.e. | **9** |
| (iii) | A 5% significance level means that the probability of rejecting $H_0$ given that it is true is 0.05.<br>Decreasing the significance level would make it less likely that a true $H_0$ would be rejected.<br>Evidence for rejecting $H_0$ would need to be stronger. | E1<br>E1<br>E1 | Or equivalent. Allow answers that relate to the context of the question. | **3** |
| (iv) | CI is given by 14.987 ± | M1 | ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0.<br>Recovery to $t_9$ is OK. | |
| | 2·262<br>$\times \dfrac{0.45675}{\sqrt{10}}$ | B1<br>M1 | | |
| | = 14.987 ± 0.3267= (14.66(0), 15.31(3)) | A1 | c.a.o. Must be expressed as an interval. | **4** |
| | | | | **18** |

Q2

(i)

| No. particles | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Obs fr | 4 | 7 | 10 | 20 | 17 | |
| Prob'y | 0.0150 | 0.0630 | 0.1322 | 0.1852 | 0.1944 | |
| Expfr | 1.50 | 6.30 | 13.22 | 18.52 | 19.44 | |
| Contrib to $X^2$ | (4.1667) | (0.0778) | 0.7843 | 0.1183 | 0.3063 | |
| Combined | 11<br>7.80<br>1.3128 | | | | | |

| | | | | |
|---|---|---|---|---|
| | | M1 | Probs correct to 3d.p. or better. | |
| | | M1 | × 100 for expected frequencies. | |
| | | A1 | All correct. | |
| | $X^2 = 1.3128 + 0.7843 + 0.1183 + 0.3063 + 0.1083 +$ | M1 | Merge first 2 cells. | |
| | $0.1813 + 0.6676 + 0.4056$ | M1 | Calculation of $X^2$. | |
| | $= 3.884(5)$ | A1 | c.a.o. (For ungrouped cells $X^2 = 6.816$.) | |
| | $H_0$: The Poisson model fits the data. | B1 | Ignore any reference to the parameter. | |
| | $H_1$: The Poisson model does not fit the data. | B1 | Do not accept "data fit model" oe. | |
| | Refer to $\chi^2_6$. | M1 | Allow correct df (= cells – 2) from wrongly grouped table and ft. Otherwise, no ft if wrong. | |
| | Upper 10% point is 10.64. | A1 | No ft from here if wrong. ($\chi^2_7 = 12.02$)<br>$P(X^2 > 3.884) = 0.6924$. | |
| | Not significant. | A1 | ft only c's test statistic. | |
| | Evidence suggests that the model fits the data. | A1 | ft only c's test statistic. Do not accept "data fit model" oe. | **12** |
| (ii) | $H_0$: $m = 15$ $H_1$: $m > 15$ | B1 | Both. Accept hypotheses in words. | |
| | where $m$ is the population median diameter( in μm). | B1 | Adequate definition of $m$ to include "population". | |
| | Given $W_- = 53$ ($\therefore$ $W_+ = 157$) | | | |
| | Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 20$. | M1 | No ft from here if wrong. | |
| | Lower 5% point is 60 (or upper is 150 if $W_+$ used). | A1 | i.e. a 1-tail test. No ft from here if wrong. | |
| | Result is significant. | A1 | ft only c's test statistic. | |
| | Evidence suggests that the median diameter appears to be more than 15 μm. | A1 | ft only c's test statistic. Conclusion in context to include "average" o.e. | **6** |
| | | | | **18** |

| Q3 | | | | |
|---|---|---|---|---|
| (i)<br>(A) |  | M1<br><br>A1<br>A1 | Increasing curve, through (0, 0), in first quadrant only.<br>Asymptotic behaviour.<br>Asymptote labelled; condone absence of axis labels. | **3** |
| (B) | For the UQ G($u$) = 0.75<br>$\therefore \left(1+\frac{u}{200}\right)^{-2} = \frac{1}{4} \quad \therefore u = 200$<br>For the LQ G($l$) = 0.25<br>$\therefore \left(1+\frac{l}{200}\right)^{-2} = \frac{3}{4} \quad \therefore l = 200\left(\frac{2}{\sqrt{3}}-1\right) = 30.94...$<br>$\therefore$IQR = 200 – 30.94 = 169(.06)<br>For an outlier $x$ > UQ +1.5 × IQR = 200 + 1.5 × 169 =453(.58) ≈ 454 (nearest hour) | M1<br>A1<br><br>A1<br>M1<br>M1<br>E1 | Use of G($x$) for either quartile.<br>c.a.o.<br><br>c.a.o.<br>UQ – LQ<br>UQ +1.5 × IQR.<br>Answer given; must be obtained genuinely. | **6** |
| (ii)<br>(A) | $F(x) = \int_0^x \frac{1}{200} e^{\frac{-t}{200}} dt$<br>$= \left[ -e^{\frac{-t}{200}} \right]_0^x = \left(-e^{\frac{-x}{200}}\right) - \left(-e^{\frac{-0}{200}}\right) = 1 - e^{\frac{-x}{200}}$ | M1<br><br>A1<br>E1 | Correct integral, including limits (which may be implied subsequently).<br>Correctly integrated.<br>Limits used. Answer given; must be shown convincingly. Condone the omission of x < 0 part.<br>Allow use of "+ $c$" with F(0) = 0. | **3** |
| (B) | P($X$ > 50) = 1 – F(50)<br>$= e^{\frac{-50}{200}} = e^{-0.25}$ | M1<br><br>E1 | Use of 1 – F($x$)<br><br>Answer given: must be convincing. (= 0.7788(0)) | **2** |
| (C) | $P(X > 400) = e^{\frac{-400}{200}} = 0.1353(35)$<br>$P(X > 450) = e^{\frac{-450}{200}} = 0.1053(99)$<br>$P(X > 450 \mid X > 400) = \frac{P(X > 450)}{P(X > 400)}$<br>$= \frac{e^{\frac{-450}{200}}}{e^{\frac{-400}{200}}} = e^{\frac{-50}{200}} = e^{-0.25} \; (= 0.7788)$ | B1<br>B1<br>M1<br><br>A1 | Accept any form.<br>Accept any form.<br>Conditional probability. Not P($X$ > 50) × P($X$ > 400) unless clearly justified.<br><br>Accept division of decimals, 3dp or better.<br>Accept a.w.r.t. 0.778 or 0.779. | **4** |
| | | | | **18** |

| | | | | |
|---|---|---|---|---|
| Q4 | $C \sim N(10,\ 0.4^2)$, $D \sim N(35,\ 3.5^2)$<br>When a candidate's answers suggest that (s)he appears to have neglected to use the difference columns of the Normal distribution tables penalise the first occurrence only. | | | |
| (i) | $P(C < 9.5) = P\left(Z < \frac{9.5-10}{0.4} = -1.25\right)$<br>$= 1 - 0{\cdot}8944 = 0.1056$ | M1<br>A1<br>A1 | For standardising. Award once, here or elsewhere.<br>c.a.o. | 3 |
| (ii) | $D - S = D - (C_1 + C_2 + C_3 + C_4) \sim N(-5,$<br>$\sigma^2 = 3.5^2 + (0.4^2 + 0.4^2 + 0.4^2 + 0.4^2) = 12.89)$<br>Want $P(D > S) = P(D - S > 0)$<br>$= 1 - \Phi\left(\frac{0-(-5)}{3.59} = 1.39(27)\right)$<br>$= 1 - 0.9182 = 0.0818$ | B1<br>B1<br>M1<br>A1 | Mean. Accept +5 for $S - D$.<br>Variance. Accept sd (= 3.590…).<br>Formulation of requirement. Accept S – D < 0. This mark could be awarded in (iii) if not earned here.<br>c.a.o. | 4 |
| (iii) | *New* $(D - S) = (D \times 1.3) - (C_1 + \ldots + C_5) \sim N(-4.5,$<br>$\sigma^2 = (3.5^2 \times 1.3^2) + (0.4^2 + \ldots + 0.4^2) = 21.5025)$<br>Again want $P(D > S) = P(D - S > 0)$<br>$= 1 - \Phi\left(\frac{0-(-4.5)}{4.637} = 0.9704\right)$<br>$= 1 - 0.8341 = 0.1659$ | B1<br>M1<br>A1<br>A1 | Mean. Accept +4.5 for $S - D$.<br>Correct use of $\times 1.3^2$ for variance.<br>c.a.o. Accept sd (= 4.637…)<br>Or S – D < 0. M1 for formulation in (ii) available here.<br>c.a.o. | 4 |
| (iv) | CI is given by $9.73 \pm$<br>1.96<br>$\times \frac{0.4}{\sqrt{12}}$<br>$= 9.73 \pm 0.2263 = (9.50(37),\ 9.95(63))$<br>Since 10 lies above this interval, it seems that the cheeses are underweight.<br>In repeated sampling,<br>95% of all confidence intervals<br>constructed in this way will contain the true mean. | M1<br>B1<br>M1<br>A1<br>E1<br>E1<br>E1 | <br>1.96 seen.<br><br>c.a.o. Must be expressed as an interval.<br>Ft c's interval. | 7 |
| | | | | **18** |