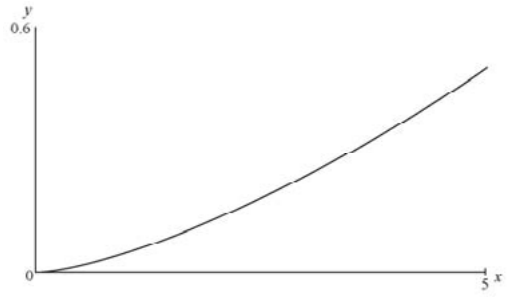
**Paper 4E: Further Statistics 2 Mark Scheme**

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **1(a)** | $\mathrm{P}(A > 3) = \frac{2}{5}$ | B1 | 1.1b |
| | $\left(\frac{2}{5}\right)^3 = \frac{8}{125}$ | M1<br>A1 | 1.1a<br>1.1b |
| | | **(3)** | |
| **(b)** | $\mathrm{f}(y) = \frac{3y^2}{125}$ | M1 | 2.1 |
| | $\mathrm{E}(Y) = \int_0^5 \frac{3y^3}{125}\,\mathrm{d}y$<br>$= \left[\frac{3y^4}{500}\right]_0^5 \qquad \left[= \frac{15}{4}\right]$ | M1 | 1.1b |
| | $\mathrm{Var}(Y) = \int_0^5 \left(\frac{3y^4}{125}\right)\mathrm{d}y \qquad \left(\frac{15}{4}\right)^2$ | M1 | 1.1b |
| | $= 0.9375$* | A1*cso | 1.1b |
| | | **(4)** | |
| **(c)** | Mode = 5 | B1 | 1.2 |
| |  Or reason based on $\frac{\mathrm{df}(y)}{\mathrm{d}y} > 0$ | B1 | 2.4 |
| | | **(2)** | |
| **(d)** | From a sketch or mode > mean therefore it has negative skew | B1ft | 2.4 |
| | | **(1)** | |
| **(e)** | $\frac{(2k)^3}{125} - \frac{k^3}{125} = 0.189$ | M1 | 3.1a |
| | $\frac{7k^3}{125} = 0.189$ | A1 | 1.1b |
| | $k = 1.5$ | A1 | 1.1b |
| | | **(3)** | |
| | | | **(13 marks)** |

**Question 1 notes:**

**(a)**

**B1:** $\frac{2}{5}$ o.e. may be implied by a correct answer

**M1:** $\left("\text{their}\left(\frac{2}{5}\right)"\right)^3$ may be implied by a correct answer

**A1:** $\frac{8}{125}$ o.e.

**(b)**

**M1:** Realising that firstly need to find pdf f($y$) and attempt to differentiate F($y$)

**M1:** Continuing the argument with an attempt to integrate $y \times$ "their f($y$)" $y^n \to y^{n+1}$

**M1:** Integrating $y^2 \times$ "their f($y$)" - ["their E($Y$)"]$^2$ $y^n \to y^{n+1}$

**A1*:** Complete correct solution no errors

**(c)**

**B1:** 5 only

**B1:** Explain their reason by either an accurate sketch or $\frac{\mathrm{d}f(y)}{\mathrm{d}y} > 0$ therefore an increasing function o.e.

**(d)**

**B1ft:** Explaining the reason for their answer. Follow through their part(b) or mean from(d) and mode from(c). A correct sketch of "their f($y$)" – may be seen anywhere in question or ft their mean and mode plus a correct conclusion

**NB:** Watch for gaming. A student who writes both negative skew with a reason and positive skew with a reason. Please send these to your Team Leader

**(e)**

**M1:** Attempting to translate the problem into an equation using $2k$ and $k$. Allow if the brackets are missing e.g. $\frac{2k^3}{125} - \frac{k^3}{125}$. No need for the 0.189

**A1:** A correct equation in any form

**A1:** A correct answer only

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **2(a)** | $H_0 : \rho = 0, H_1 : \rho > 0$ | B1 | 2.5 |
| | Critical value at 1% level is 0.8929 | B1 | 1.1b |
| | $r_s < 0.8929$ so not significant evidence to reject $H_0$ | M1 | 2.1 |
| | The researcher's claim is not correct (at 1% level)<br>**or** insufficient evidence for researcher's claim<br>**or** there is insufficient evidence that water gets deeper further from inner bank<br>**or** no (positive) correlation between depth of water and distance from inner bank | A1ft | 2.2b |
| | | **(4)** | |
| **(b)(i)** | The **ranks will remain the same** therefore there will be **no change** to the spearman's rank correlation coefficient | **B1** | 2.4 |
| **(ii)** | Spearman's rank correlation coefficient will **increase** since | **B1** | 2.2a |
| | The **ranks are the same** for both distance and depth therefore ***d* = 0** however, ***n* has increased or** the new position follows the pattern that large *b* is associated with large *s* and so $r_s$ will increase | **B1** | 2.4 |
| | | **(3)** | |
| **(c)** | The mean of the tied ranks is given to each… | B1 | 2.4 |
| | … then use PMCC | B1 | 2.4 |
| | | **(2)** | |
| | | | **(9 marks)** |

**Notes:**

**(a)**
**B1:** Both hypotheses correct written using the notation $\rho$
**B1:** awrt 0.893
**M1:** Drawing a correct inference using their answer to part(a) and their CV
**A1ft:** Drawing a correct inference in context using their answer to part(a) and their CV

**(b)(i)**
**B1:** Stating **no change** and an explanation including **ranks remain unchanged** o.e. and **no change o.e.**

**(b)(ii)**
**B1:** Interpreted the outcome of adding a point as **increased** oe
**B1:** Explaining why. Need to mention the **ranks are the same for both oe** and ***n* has increased oe**

**(c)**
**B1:** Explaining that the mean of the values for the tied ranks is given to both values
**B1:** Explaining that the PMCC must be used

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **3(a)** | 95% CI for $\mu$ uses $t$ value of **2.064** | B1 | 3.3 |
| | | | |
| | $\frac{\hat{\sigma}}{\sqrt{25}}\times"2.064"=\frac{1}{2}(2.232-1.128)$ or $\frac{1}{2}(2.232+1.128)+"2.064"\times\frac{\hat{\sigma}}{\sqrt{25}}=2.232$ (oe) | M1 | 2.1 |
| | $\hat{\sigma}=\frac{2.76}{"2.064"}$ or 1.3372… | M1 | 1.1b |
| | $\hat{\sigma}^2=1.788...$ [=1.79 (3sf)] * | A1*cso | 1.1b |
| | | **(4)** | |
| **(b)** | $12.401, < \frac{24\times1.79}{\sigma^2} <, 39.364$ | B1<br>M1 | 1.1b<br>1.1a |
| | **1.09** $< \sigma^2 <$ **3.46** | A1 | 1.1b |
| | | **(3)** | |
| | | | **(7 marks)** |

**Notes:**

**(a)**

**B1:** Realising that the *t*-distribution must be used as a model and finding the correct value awrt 2.06

**M1:** Using the correct formula with a *t*-value, $\frac{\hat{\sigma}}{\sqrt{25}}\times"t\text{ value}"=\frac{1}{2}(2.232-1.128)$ or

$\frac{1}{2}(2.232+1.128)+"t\text{ value}"\times\frac{\hat{\sigma}}{\sqrt{25}}=2.232$ **or**

$\frac{1}{2}(2.232+1.128)-"t\text{ value}"\times\frac{\hat{\sigma}}{\sqrt{25}}=1.128$

**M1:** Rearranging one of these formula accurately to find a value of $\hat{\sigma}$

**A1cso*:** A correct solution only using awrt 1.79

**(b)**

**B1:** awrt 12.4 or 39.4 May be implied by a correct confidence interval

**M1:** $\frac{24\times1.79}{\sigma^2}$ May be implied by a correct confidence interval

**A1:** awrt 1.09 and awrt 3.46

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **4(a)** | $H_0$: $\sigma_G^2 = \sigma_B^2$, $H_1$: $\sigma_G^2 \neq \sigma_B^2$, | B1 | 2.5 |
| | $s_B^2 = \frac{1}{6}(56130 - 7 \times 88.9^2) = \frac{807.53}{6} = 134.6$ | M1<br>A1 | 2.1<br>1.1b |
| | $s_G^2 = \frac{1}{7}(55746 - 8 \times 83.1^2) = \frac{501.12}{7} = 71.58$ | A1 | 1.1b |
| | $\frac{s_B^2}{s_G^2} = 1.880\ldots$ | M1 | 3.4 |
| | Critical value $F_{6,7} = 3.87$ | B1 | 1.1b |
| | Not significant, variances can be treated as the same | A1 ft | 2.2b |
| | | **(7)** | |
| **(b)** | $H_0$: $\mu_B = \mu_G$, $H_1$: $\mu_B > \mu_G$ | B1 | 2.5 |
| | Pooled estimate of variance $s^2 = \frac{6\times134.6 + 7\times71.58}{13} = 100.6653\ldots$ | M1 | 3.1b |
| | Test statistic $t = \frac{88.9 - 83.1}{s\sqrt{\frac{1}{7} + \frac{1}{8}}}$ = awrt 1.12 | M1<br>A1 | 1.1b<br>1.1b |
| | Critical value $t_{13}(5\%) = 1.771$ | B1 | 1.1b |
| | Insufficient evidence to support mother's claim | A1 ft | 2.2b |
| | | **(6)** | |
| | | | **(13 marks)** |

**Notes:**

**(a)**

**B1:** Both hypotheses correct using the notation $\sigma^2$. Allow $\sigma$ rather than $\sigma^2$

**M1:** Using a correct Method for either $s_B^2$ or $s_G^2$ May be implied by a correct value

**A1:** awrt 135

**A1:** awrt 71.6

**M1:** Using the F-distribution as the model e.g. $\frac{s_B^2}{s_G^2}$

**B1:** awrt 3.87

**A1ft:** Drawing a correct inference following through their CV and value for $\frac{s_B^2}{s_G^2}$

**(b)**

**B1:** Both hypotheses correct using the notation $\mu$

**M1:** For realising the need to find the pooled estimate for the test require from a correct interpretation of the question

**M1:** Correct method for test statistic $t = \frac{88.9 - 83.1}{\text{"their } s\text{"}\sqrt{\frac{1}{7} + \frac{1}{8}}}$ May be implied by a correct awrt 1.12

**A1:** awrt 1.12

**B1:** awrt 1.77

**A1ft:** Drawing a correct inference following through their CV and test statistic

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **5(a)** | Let $X = L - 4S$ then $E(X) = 19.6 - 4 \times 4.8$ | M1 | 2.3 |
| | $= 0.4$ | A1 | 1.1b |
| | $\text{Var}(X) = \text{Var}(L) + 4^2\text{Var}(S) = 0.6^2 + 16 \times 0.3^2$ | M1 | 2.1 |
| | $= 1.8$ | A1 | 1.1b |
| | $P(X > 0) = [P(Z > \frac{0-0.4}{\sqrt{1.8}} = -0.298\ldots\ldots)]$ | M1 | 2.1 |
| | $= 0.617202...$ awrt **0.617** | A1 | 1.1b |
| | | **(6)** | |
| **(b)** | $T = S_1 + S_2 + S_3 + S_4$ (May be implied by 0.36) | M1 | 3.3 |
| | $T \sim N(19.2, 0.36)$ $E(T) = 19.2$ | B1 | 1.1b |
| | $\text{Var}(T) = 0.36$ **or** $0.6^2$ | A1 | 1.1b |
| | | **(3)** | |
| **(c)** | Let $Y = L - T$ $E(Y) = E(L) - E(T) = [0.4]$ | M1 | 3.3 |
| | $\text{Var}(Y) = \text{Var}(L) + \text{Var}(T) = [0.72]$ | M1 | 1.1b |
| | Require $P(-0.2 < Y < 0.2)$ | M1 | 3.1a |
| | $= 0.16708\ldots$ awrt **0.167** | A1 | 1.1b |
| | | **(4)** | |
| | | | **(13 marks)** |

**Notes:**

**(a)**

**M1:** Selecting and using an appropriate model i.e. $\pm(L - 4S)$. May be implied by 0.4

**A1:** 0.4 oe

**M1:** For realising the need to use $\text{Var}(L) + 4^2\text{Var}(S)$. Allow use of 0.6 for $\text{Var}(L)$ instead of $0.6^2$ and/or 0.3 for $\text{Var}(S)$ instead of $0.3^2$ may be implied by 1.8

**A1:** 1.8 only

**M1:** For realising $P(X > 0)$ is required and an attempt to find it e.g. $\frac{0-0.4}{\sqrt{\text{"their Var}(X)\text{"}}}$ but do not allow a negative $\text{Var}(X)$

**A1:** awrt 0.617

**(b)**

**M1:** Selecting and using an appropriate model ie $S_1 + S_2 + S_3 + S_4$ : may be implied by 0.36

**B1:** 19.2 only

**A1:** 0.36

**(c)**

**M1:** Setting up and using the model $Y = L - T$. May be implied by $E(Y) = E(L) - E(T)$

**M1:** Using $\text{Var}(Y) = \text{Var}(L) + \text{Var}(T)$

**M1:** Dealing with the modulus and realising they need to find $P(-0.2 < Y < 0.2)$

**A1:** awrt 0.167

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AOs</th></tr>
<tr><td rowspan="4">6(a)</td><td>$\left[ b = \frac{S_{xm}}{S_{xx}} = -0.0277576 \right]$</td><td>M1</td><td>3.3</td></tr>
<tr><td>$[a = \bar{m} - b\bar{x} = 1.278 + 0.0277576 \times 8.5 = 1.5139]$</td><td></td><td></td></tr>
<tr><td>$m = 1.5139 - 0.02775\ldots x$</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td>(2)</td><td></td></tr>
<tr><td rowspan="3">(b)</td><td>$\text{RSS} = 0.12756 - \frac{(-2.29)^2}{82.5}$</td><td>M1</td><td>1.1b</td></tr>
<tr><td>$= 0.06399$*</td><td>A1*</td><td>1.1b</td></tr>
<tr><td></td><td>(2)</td><td></td></tr>
<tr><td rowspan="3">(c)</td><td>
<table>
<tr><th>$x$</th><th>$m$</th><th>$m = a + bx$</th><th>$\varepsilon$</th></tr>
<tr><td>4</td><td>1.50</td><td>1.4029</td><td>+0.0971</td></tr>
<tr><td>5</td><td>1.20</td><td>1.3752</td><td>– 0.1752</td></tr>
<tr><td>6</td><td>1.40</td><td>1.3474</td><td>+0.0526</td></tr>
<tr><td>7</td><td>1.40</td><td>1.3196</td><td>+0.0804</td></tr>
<tr><td>8</td><td>1.23</td><td>1.2919</td><td>– 0.0619</td></tr>
<tr><td>9</td><td>1.30</td><td>1.2641</td><td>+0.0359</td></tr>
<tr><td>10</td><td>1.20</td><td>1.2364</td><td>– 0.0364</td></tr>
<tr><td>11</td><td>1.15</td><td>1.2086</td><td>– 0.0586</td></tr>
<tr><td>12</td><td>1.25</td><td>1.1808</td><td>+0.0692</td></tr>
<tr><td>13</td><td>1.15</td><td>1.1531</td><td>– 0.0031</td></tr>
</table>
</td><td>M1<br>A1</td><td>3.4<br>1.1b</td></tr>
<tr><td></td><td>(2)</td><td></td></tr>
<tr><td>(d)</td><td>The point (5, 1.2) is an outlier</td><td>B1ft</td><td>2.2b</td></tr>
<tr><td></td><td></td><td>(1)</td><td></td></tr>
<tr><td>(e)(i)</td><td>It is a valid piece of data so should be used<br>or<br>It does not follow the pattern according to the residuals so may contain an error making the result invalid so should be removed</td><td>B1</td><td>2.4</td></tr>
<tr><td rowspan="2">(ii)</td><td>$a = \bar{m} - b\bar{x} = 1.28667 + 0.03765 \times 8.88889 = 1.6213$</td><td>M1</td><td>3.3</td></tr>
<tr><td>$m = 1.6213 - 0.03765x$</td><td>A1</td><td>1.1b</td></tr>
<tr><td rowspan="2">(iii)</td><td>$m = 1.6213 - 0.03765 \times 15$</td><td></td><td></td></tr>
<tr><td>$= 1.056$ or awrt 1.06</td><td>B1ft</td><td>3.4</td></tr>
<tr><td rowspan="2">(iv)</td><td>The model is only reliable if the values are limited to those in the given range so probably not reliable</td><td>B1</td><td>3.5b</td></tr>
<tr><td></td><td>(5)</td><td></td></tr>
<tr><td colspan="4">(12 marks)</td></tr>
</table>

| **Question 6 notes:** | |
|---|---|
| **6(a)** | |
| **M1:** | Realising the need to use $b = \frac{S_{xm}}{S_{xx}}$ and $a = \bar{m} - b\bar{x}$ |
| **A1:** | $m$ = awrt 1.51) – (awrt 0.0278) $x$. Award M1A1 for correct equation |
| **(b)** | |
| **M1:** | Using $S_{mm} - \frac{(S_{xm})^2}{S_{xx}}$ |
| **A1*:** | awrt 0.064 |
| **(c)** | |
| **M1:** | Using the model in part (a) i.e. $m$ – ("1.5139" – "0.02775"$x$) implied by a correct value |
| **A1:** | All correct.<br>Award M1A1 for a list of correct residuals |
| **(d)** | |
| **B1:** | Inferring from the residuals that the outlier is (5, 1.2) ft their residuals. |
| **(e)(i)** | |
| **B1:** | Explaining why the outlier should be removed or not. |
| **(ii)** | |
| **M1:** | Removing the outlier and refining the model by finding a new regression line. |
| **A1:** | $m$ = (awrt 1.62) – (awrt 0.0377)$x$ |
| **(iii)** | |
| **B1ft:** | using their model in e(i) with $x$ = 15. awrt 1.06 or ft their e(ii) |
| **(iv)** | |
| **B1:** | Realising the limitations of the model by stating it is <u>not reliable</u> and giving the reason why i.e. extrapolation/out of range o.e. |

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **7(a)** | $S_{xx} = \sum(10s)^2 - \dfrac{\left(\sum 10s\right)^2}{10}$ | M1 | 2.1 |
| | $2658.9 = 100\sum(s)^2 - \dfrac{100\left(\sum s\right)^2}{10}$ | M1 | 1.1b |
| | $2658.9 = 100\, S_{ss}$ | | |
| | $S_{ss} = 26.589$ * | A1*cso | 1.1b |
| | | **(3)** | |
| **(b)** | $64 = \sum_1^{10} 10(d_i - 9)$ | M1 | 3.1a |
| | $64 = 10\sum_1^{10} d_i - 900$ | | |
| | $\sum_1^{10} d_i = 96.4$ | A1 | 1.1b |
| | $S_{dd} = 1081.74 - \dfrac{("96.4")^2}{10}$ | M1 | 1.1b |
| | $= 152.444$ | | |
| | $r = 0.935$ | A1ft | 1.1b |
| | | **(4)** | |
| **(c)** | Linear correlation is significant but scatter diagram suggests a non-linear relationship between the level of serum magnesium, and the level of the disease protein | B1 | 3.5a |
| | | **(1)** | |
| | | | **(8 marks)** |

**Notes:**

**(a)**

**M1:** Attempting to use $S_{xx} = \sum x^2 - \dfrac{\left(\sum x\right)^2}{10}$ with $x = 10s$

**M1:** Substituting in 2658.9 and dealing with the 10 correctly

**A1*:** cso A complete solution with no errors leading to 26.589 only

**(b)**

**M1:** Realising that either $64 = \sum_1^{10} 10(d_i - 9)$ or $64 = 10\sum_1^{10} d_i - 900$ o.e. must be used. May be implied by seeing 96.4

**A1:** 96.4 only

**M1:** Attempting to use $S_{dd} = \sum d^2 - \dfrac{\left(\sum d\right)^2}{10}$ may be implied by 0.935

**A1ft:** awrt 0.935 ft "their 96.4"

**(c)**

**B1:** A correct comment comparing their value of $r$ and the scatter diagram in context