# Mark Scheme (Results)

## January 2022

Pearson Edexcel International A Level
In Statistics S3 (WST03) Paper 01

**Edexcel and BTEC Qualifications**

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

**Pearson: helping people progress, everywhere**

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

January 2022
Question Paper Log Number P72405A
Publications Code WST03_01_2201_MS

**EDEXCEL IAL MATHEMATICS**
**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.

**General Instructions for Marking**

1. The total number of marks for the paper is 75.

2. The Edexcel Mathematics mark schemes use the following types of marks:

   - **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
   - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
   - **B** marks are unconditional accuracy marks (independent of M marks)
   - Marks should not be subdivided.

3. Abbreviations

   These are some of the traditional marking abbreviations that will appear in the mark schemes.

   - bod – benefit of doubt
   - ft – follow through
   - the symbol  will be used for correct ft

- cao – correct answer only
- cso - correct solution only. There must be no errors in this part of the question to obtain this mark
- isw – ignore subsequent working
- awrt – answers which round to
- SC: special case
- oe – or equivalent (and appropriate)
- dep – dependent
- indep – independent
- dp decimal places
- sf significant figures
- ∗ The answer is printed on the paper
- ⌐ ¬ The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

6. If a candidate makes more than one attempt at any question:
   - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
   - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

7. Ignore wrong working or incorrect statements following a correct answer

**Special notes for marking Statistics exams (for AAs only)**

- Any correct method should gain credit. If you cannot see how to apply the mark scheme but believe the method to be correct then please send to review.
- For method marks, we generally allow or condone a slip or transcription error if these are seen in an expression. We do not, however, condone or allow these errors in accuracy marks.

| Question Number | Scheme | Marks |
|---|---|---|
| 1 (a) | $\overline{x} = 11.42$ | B1 |
| | $s^2 = \dfrac{1310.464 - 10 \times 11.42^2}{9}$ | M1 |
| | $= 0.7$ | A1 |
| | | (3) |
| (b) | $z$ value for 95% CI is 1.96 | B1 |
| | $'11.42' \pm 1.96 \times \dfrac{0.8}{\sqrt{10}}$ | M1 |
| | (10.924…, 11.915…) awrt (10.92, 11.92) | A1 A1 |
| | | (4) |
| (c) | $Y \sim N(\text{“}11.92\text{”}, 0.8^2)$ | M1 |
| | $P(Y < 10.5) = P\left(Z < \dfrac{10.5 - "11.92"}{0.8}\right) [= P(Z < -1.775]$ | M1 |
| | $= 0.03837\ldots$ awrt 0.038 | A1 |
| | | (3) |

| | Notes | | Total 10 |
|---|---|---|---|
| 1(a) | **B1** | for 11.42 cao | |
| | **M1** | for use of $s^2 = \dfrac{\sum x^2 - n\overline{x}^2}{n-1}$ | |
| | **A1** | for 0.7 cao | |
| (b) | **B1** | for writing or using 1.96 (or better from calculator 1.9599…) | |
| | **M1** | For use of $\overline{x} \pm z$ value $\times \dfrac{\sigma}{\sqrt{n}}$ ft their $z$ value, $1 < \lvert z \rvert < 2$ and their 11.42 | |
| | **A1** | for awrt 10.9 or awrt 11.9 | |
| | **A1** | for awrt 10.92 and awrt 11.92 | |
| (c) | **M1** | for identifying the normal distribution with the upper confidence interval value as mean and 0.8 as standard deviation (may be seen in standardisation) | |
| | **M1** | for standardising with 10.5, their mean (which must be in their confidence interval (including limits) from part (b)) and standard deviation = 0.8 | |
| | **A1** | awrt 0.038 (tables = 0.0375) | |

| Question Number | Scheme | Marks |
|---|---|---|
| 2(a) | $H_0 : \mu_{year7} = \mu_{year8}$ $H_1 : \mu_{year7} \neq \mu_{year8}$ | B1 |
| | $SE = \sqrt{\frac{38}{240} + \frac{42}{240}}$ | M1 |
| | $z = \frac{103 - 101}{SE}$ | M1 |
| | $= (\pm)3.464....$ $\left(2\sqrt{3}\right)$ awrt $(\pm)3.46$ | A1 |
| | $Z_{critical} = 2.5758$ | B1 |
| | In CR/Significant/Reject $H_0$ | M1 |
| | There is sufficient evidence to suggest that the regional education officer's claim is not correct/ There is a difference between the mean scores of the two year groups. | A1 |
| | | (7) |
| (b) | CLT allows us to use sample means (oe) being normally distributed | B1 |
| | | (1) |

| | Notes | Total 8 |
|---|---|---|
| (a) | **B1** | both hypotheses correct. Allow equivalent rearrangements. Must be in terms of $\mu$<br>If using e.g. $\mu_A = \mu_B$ $A$ and $B$ must be clearly identified with year groups |
| | **M1** | for use of SE with 38 and 42 (may be implied by SE = awrt 0.577) |
| | **M1** | for a correct standardisation expression using 103, 101 (in either order) and SE = awrt.0577<br>or ft their stated SE<br>or if not stated (i.e. only seen in standardisation) only allow $\sqrt{\frac{38^2}{240} + \frac{42^2}{240}}$ or $\sqrt{\frac{\sqrt{38}}{240} + \frac{\sqrt{42}}{240}}$ |
| | **A1** | awrt 3.46 or awrt –3.46<br>allow $p$ value of awrt 0.000266 |
| | **B1** | \|CV\| = 2.5758 or better (seen) |
| | **M1** | a correct statement linking their test statistic and their CV – need not be contextual but do not allow contradicting non contextual comments. |
| | **A1** | do not allow a ft conclusion here.<br>a correct contextual statement (dependent on 2nd M1) which must be consistent with their test statistics and CV and which also must reject $H_0$ . It must mention the officer or mean scores. |
| (b) | **B1** | a correct explanation which must mention sample means oe (population means are normally distributed is B0)<br>ignore extraneous non-contradictory comments |

| Question Number | Scheme | Marks |
|---|---|---|
| 3 (a) | $\left[ r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \right] \frac{15.1608}{\sqrt{6.90181 \times 45.304}}$ | M1 |
| | $= 0.8573\ldots$ awrt 0.857 | A1 |
| | | (2) |
| (b) | $H_0 : \rho = 0$ , $H_1 : \rho > 0$ | B1 |
| | Critical value 5% = 0.5494 | B1 |
| | Significant evidence to suggest that there is a positive correlation between MR and BMI | B1 |
| | | (3) |
| (c) | MR and BMI measurements are normally (or bivariate normal) distributed | B1 |
| | | (1) |
| (d) | Ranks for MR: 9 10 6 7 8 4 5 1 2 3 | B1 |
| | $\sum d^2 = 1+9+9+1+4+1+16+9+9+1 [= 60]$ | M1 |
| | $r_s = 1 - \frac{6(60)}{10(99)}$ | M1 |
| | $= 0.6363$ awrt $(\pm) 0.636$ | A1 |
| | | (4) |
| (e) | $[ H_0 : \rho = 0$ , $H_1 : \rho \neq 0 ]$ | |
| | Critical value 0.6485 | B1 |
| | There is insufficient evidence of a correlation between MR and DPA | B1 |
| | | (2) |

| | Notes | Total 12 |
|---|---|---|
| (a) | **M1** | for use of $\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ |
| | **A1** | awrt 0.857 |
| (b) | **B1** | both hypotheses correct. Must be in terms of $\rho$. Must be attached to $H_0$ and $H_1$ Do not allow hypotheses in words on their own. |
| | **B1** | critical value of 0.5494 |
| | **B1** | correct conclusion rejecting $H_0$ which must mention positive correlation, MR and BMI which must be consistent with their CV and their $r$, with \|their CV\| < 1 and \|their $r$\| < 1 |
| (c) | **B1** | correct assumption referring to MR and BMI needing to be normally distributed |
| (d) | **B1** | attempt to rank MR (at least four correct) (may be implied by correct $\sum d^2$ or correct answer) allow reverse ranks for MR: 2 1 5 4 3 7 6 10 9 8 |
| | **M1** | for finding the difference between each of the ranks and evaluating $\sum d^2$ (implied by $\sum d^2 = 60$ or for reverse ranks $\sum d^2 = 270$) |
| | **M1** | using $1 - \frac{6\sum d^2}{10(99)}$ with their $\sum d^2$ |
| | **A1** | awrt $(\pm) 0.636$ |
| (e) | **B1** | critical value of 0.6485 (or –0.6485 if $r_s < 0$) |
| | **B1** | correct conclusion which is not rejecting $H_0$, which must mention MR and DPA which must be consistent with their CV and their $r_s$, with \|their CV\| < 1 and \|their $r_s$ \| < 1 |
| | | |

| Question Number | Scheme | Marks |
|---|---|---|
| 4(a) | Non random sampling/description of non random sampling oe | B1 |
| | from (different groups of the) population until each quota has been met | B1 |
| | | (2) |
| (b) | $H_0$ : Subject enjoyed the most and group are independent<br>$H_1$ : Subject enjoyed the most and group are not independent | B1 |
| | Expected: Maths / Physics / Chemistry / Total<br>Group *A*: 21.06 / 8.97 / 8.97 / (39)<br>Group *B*: 32.94 / 14.03 / 14.03 / (61)<br>Total: (54) / (23) / (23) / (100) | M1 |
| | Observed / Expected / $\frac{(O-E)^2}{E}$ / $\frac{O^2}{E}$<br>16 / 21.06 / 1.215745… / 12.15575…<br>10 / 8.97 / 0.118272… / 11.14827…<br>13 / 8.97 / 1.81058… / 18.84058…<br>38 / 32.94 / 0.77728… / 43.83728…<br>13 / 14.03 / 0.075617… / 12.04562…<br>10 / 14.03 / 1.157584… / 7.127584…<br>Totals / 5.155… / 105.155… | dM1 |
| | $\left[X^2 =\right]\sum\frac{(O-E)^2}{E}$ or $\sum\frac{O^2}{E} - 100$ | dM1 |
| | = 5.155… awrt 5.16 or awrt 5.15 | A1 |
| | $\nu = (3-1)(2-1) = 2$ | B1 |
| | $\chi^2_2(0.05) = 5.991$ | B1ft |
| | [not in CR/not significant/Do not reject $H_o$] There is not sufficient evidence to suggest that subject enjoyed and group are not independent | A1 |
| | | (8) |
| (c)(i) | No change (as the test is still the same) | B1 |
| (ii) | No change (as $\nu = 2$ still) | B1 |
| (iii) | Test statistics would double (= 10.310…) (as all observed and expected values are doubled.) | B1 |
| (iv) | Conclusion is the opposite (There is sufficient evidence to suggest that subject enjoyed and group are not independent) as test statistic is now greater than the critical value (10.31 > 5.991) | B1 |
| | | (4) |

| | **Notes** | **Total 14** |
|---|---|---|
| (a) | **B1** | for a correct statements referring to non-random sampling or a description of a non-random method for selecting participants e.g. choosing people as they leave the school . Do not allow labelling or numbering |
| | **B1** | for a correct statement referring to selection from different groups until quota is filled |
| (b) | **B1** | both hypotheses correct. Must mention "Subject" and "group" at least once. (may be written in terms of association) |
| | **M1** | Some attempt at $\frac{(\text{Row Total})\times(\text{Column Total})}{\text{Grand Total}}$ Can be implied by at least one correct $E_i$ to 1 dp |
| | **dM1** | dependent on 1st M1 for at least 2 correct terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their $E_i$ Accept 2 sf accuracy |

| | | |
|---|---|---|
| | **dM1** | dependent on 2nd M1 for applying $\sum \frac{(O-E)^2}{E}$ or $\sum \frac{O^2}{E} - 100$ |
| | **A1** | awrt 5.16 |
| | **SC** | If no expected frequencies shown, then an answer of awrt 5.16 scores M0M0M1A1 |
| | **B1** | $\nu = 2$ may be implied by a correct critical value of 5.991 |
| | **B1ft** | 5.991 allow ft from their stated degrees of freedom (may see 3.841, 7.815, 9.488, 11.070) |
| | **A1** | dependent on 3rd M1 and 3rd B1. A correct contextualised conclusion which is not rejecting $H_o$ Must mention subject and group. Contradictory statements score A0 e.g. "significant, do not reject $H_o$" If no hypotheses or hypotheses wrong way round do not award. |
| (c)(i) | **B1** | a correct statement |
| (ii) | **B1** | a correct statement |
| (iii) | **B1** | a correct statement which must state that the test statistic doubles |
| (iv) | **B1** | a correct statement with correct reasoning |

| Qu. No. | Scheme | Marks |
|---|---|---|
| 5 (a) | Let $T$ = total time taken | |
| | $T \sim \text{N}(41+81+57, 5.2^2+4.2^2+6.6^2)$ [So $T \sim \text{N}(179, 88.24)$ ] | M1 A1 |
| | $\text{P}(T>180) = \text{P}\left(Z > \frac{180-179}{\sqrt{88.24}}\right)$ | M1 |
| | $=1-0.5438=0.4562$ (calculator gives 0.4576…) awrt 0.456 to 0.458 | M1 A1 |
| | | (5) |
| (b) | Let $Y$ = difference between run and swim or Let $D = R - S - 20$ | |
| | $Y \sim \text{N}(16, 70.6)$ or $D \sim \text{N}(-4, 70.6)$ | B1 |
| | $\text{P}(Y>20) = \text{P}\left(Z > \frac{20-16}{\sqrt{70.6}}\right)$ or $\text{P}(D>0) = \text{P}\left(Z > \frac{0-(-4)}{\sqrt{70.6}}\right)$ | M1 |
| | $=1-0.6844=0.3156$ (calculator gives 0.3170…) awrt 0.316/0.317 | M1 A1 |
| | | (4) |
| (c) | $\text{P}(T>t)=0.95 \Rightarrow \text{P}\left(Z > \frac{t-179}{\sqrt{88.24}}\right)=0.95 \Rightarrow \frac{t-179}{\sqrt{88.24}} = -1.6449$ | M1 B1 |
| | $t = 163.548...$ awrt 164 | A1 |
| | | (3) |
| (d) | Let $X$ = the number of times greater than 3 hours in 6 attempts | |
| | $X \sim \text{B}(6, "0.456")$ | B1ft |
| | $\text{P}(X \geqslant 1) = 1 - \text{P}(X=0) = 1 - "0.5438"^6$ $\text{P}(X \geqslant 1) = 1 - \text{P}(X=0) = 1 - "0.5438"^6$ | M1 |
| | = 0.9741… (using the calculator value gives 0.9745…) awrt 0.974/0.975 | A1 |
| | | (3) |
| (e) | eg The times for each event are not now likely to be independent | M1 |
| | Jane is correct / calculation is not valid | A1 (2) |
| | **Notes** | **Total 17** |

| | | |
|---|---|---|
| (a) | **M1** | for setting up a normal distribution with a mean 41 + 81 + 57 (= 179) |
| | **A1** | for a correct expression of variance implied by (variance =) 88.24 or for s.d. = awrt 9.39 |
| | **M1** | for standardising with 180, their mean and their standard deviation |
| | **M1** | use of $1-p$ with $0.5 < p < 1$ |
| | **A1** | awrt 0.456 to 0.458 |
| (b) | **B1** | For $\text{N}(\pm 16, 70.6)$ or $\text{N}(\pm 4, 70.6)$ May be seen in a calculation |
| | **M1** | for standardisation with $\pm 20$ or 0, their mean and their s.d.(their var must be > 0) must be compatible e.g. – 20 with –16 |
| | **M1** | use of $1-p$ with $0.5 < p < 1$ |
| | **A1** | awrt 0.316/0.317 |
| (c) | **M1** | for standardising using their mean and standard deviation = $z$ value $1 < \lvert z \rvert < 2$ |
| | **B1** | for correct $z$ value $\pm 1.6449$ or better. Must have compatible sign with standardisation |
| | **A1** | awrt 164 |
| (d) | **B1ft** | for writing or using B(6,'0.4562') ft their answer to part (a) to 3sf |
| | **M1** | use of $\text{P}(X \geqslant 1) = 1 - \text{P}(X=0)$ [$=1-(1-\textit{their}(a))^6$ ] allow $\text{P}(X \geqslant 1) = \text{P}(X=1) + \text{P}(X=2) + ... + \text{P}(X=6)$ |
| | **A1** | awrt 0.974/0.975 |
| (e) | **M1** | Reference to the events no longer being independent (he might get tired after each event or events now follow consecutively)/ calculation does not include time between events |
| | **A1** | Correct conclusion (Jane is correct) with corresponding reason |

| Qu. No. | Scheme | Marks |
|---|---|---|
| 6(a) | $\mathrm{P}(S<303.5)=\mathrm{P}\left(Z<\dfrac{303.5-310}{4}\right)$ or $\mathrm{P}(S>315.5)=\mathrm{P}\left(Z>\dfrac{315.5-310}{4}\right)$ | M1 |
| | $=0.05208$ or $0.084565\ldots$ awrt 0.052 or awrt 0.084/0.085 | A1 |
| | So $a=5.2$ or $b=8.5$ awrt 5.2 or awrt 8.4/8.5 | A1 |
| | e.g. $b=100-10.6-16.3-19.6-18.4-13.6-7.8-$ '5.2' | M1 |
| | Both $a=5.2$ and $b=8.5$ awrt 5.2/5.3 and awrt 8.4/8.5 | A1 |
| | | (5) |
| (b) | $H_0$: The normal distribution N(310, 16) is a suitable model/The data are consistent with the model.<br>$H_1$: The normal distribution N(310, 16) is not a suitable model/The data are not consistent with the model. | B1 |
| | $\left[X^2=\right]\sum\dfrac{(O-E)^2}{E}=\dfrac{(5-'5.2')^2}{'5.2'}+\dfrac{(4-'8.5')^2}{'8.5'}+9.71$ | M1 M1 |
| | $=12.10\ldots$ awrt 12.0 to 12.1 | A1 |
| | $\nu=7$ | B1 |
| | $\chi^2_7(0.05)=14.067$ | B1ft |
| | [not in the CR/not significant/Do not reject $H_0$] There is not sufficient evidence to suggest that N[(310, 16)] is not a suitable model/The model is suitable/The data are consistent with the model | A1 |
| | | (7) |
| (c) | $\nu=8-3=5$ / two parameters estimated so additional degrees of freedom subtracted | M1 |
| | Therefore the critical value is reduced/now 11.070 | A1 |
| | | (2) |

| | Notes | Total 14 |
|---|---|---|
| (a) | **M1** | for standardising with 303.5 or 315.5, 310 and 4 |
| | **A1** | awrt 0.052 or awrt 0.084/0.085 |
| | **A1** | either correct value |
| | **M1** | a complete method to find the second missing value |
| | **A1** | both correct values |
| (b) | **B1** | both hypotheses correct. If mentioning normal, must mention N(310, 16) at least once. |
| | **M1** | for either $\dfrac{(5-'5.2')^2}{'5.2'}$ or $\dfrac{(4-'8.5')^2}{'8.5'}$ |
| | **M1** | for a complete method to find $\sum\dfrac{(O-E)^2}{E}$ e.g. 9.71 + 2 additional terms<br>this mark is independent of the 1st M1 |
| | **A1** | allow awrt 12.0 to 12.1 |
| | **B1** | $\nu=7$ This mark can be implied by a correct critical value of 14.067 |
| | **B1ft** | 14.067 (may see 5.991, 7.815, 9.488, 11.070, 12.592) |
| | **A1** | dependent on 2nd M1 a correct conclusion which states that the model is suitable and must be consistent with their $X^2$ value and their $\chi^2$ critical value.<br>If no hypotheses or hypotheses wrong way round do not award. |
| (c) | **M1** | a statement that implies 2 additional degrees of freedom are subtracted |
| | **A1** | a correct conclusion from correct reasoning. |