

Mr Ronald Maths and Statistics
GCSE (9–1) Statistics – Paper 2 (Higher Tier)

Mark Scheme

Predicted paper for the examination on 12 June 2026

Marking abbreviations

- **M** – method mark, awarded for a correct method (may not be explicit if the answer is correct).
- **A** – accuracy mark, dependent on a relevant method mark being earned.
- **B** – independent mark, awarded for a correct result or statement without needing method.
- *ft* – follow through; *oe* – or equivalent; *awrt* – answer which rounds to; *dep* – dependent.

| Q | Answer / working | Marks | Notes |
|------|--|-------------|---|
| 1(a) | Any two valid criticisms, e.g. the response boxes overlap (£20,000 and £40,000 each appear in two boxes); there is no box for incomes above £60,000 (boxes not exhaustive); no box for £0 / “prefer not to say”; salary is sensitive so people may not answer honestly. | BB | B1 for each distinct valid criticism (max 2). |
| 1(b) | A suitable question with non-overlapping, exhaustive response boxes, e.g. “What is your annual salary?” with boxes £0–£19,999, £20,000–£39,999, £40,000–£59,999, £60,000 or more. | BB | B1 non-overlapping & exhaustive boxes; B1 suitable wording (incl. a time period). |
| 1(c) | Tick Discrete and Quantitative (these two only). | B1 | Both required, and no other box ticked, for the mark. |
| 1(d) | Secondary; because the data was collected / published by someone else (not by the researcher herself). | BB | B1 secondary; B1 valid reason. |
| 2(a) | Completed table: Boys Bus = 30; Girls Walk = 24; Girls Total = 50; Bus Total = 44; Car Total = 34. | B3 | B3 all correct. B2 for ≥ 4 correct entries; B1 for ≥ 2 correct entries. |
| 2(b) | $\frac{18}{70}$ | MA | M1 for $\frac{18}{70}$ <i>oe</i> (boys who walk \div total boys; sample space is the 70 boys); A1 for $\frac{9}{35}$ <i>oe</i> (<i>awrt</i> 0.26). |
| 3(a) | Opportunity (convenience) sampling | B1 | |
| 3(b) | Only employees who arrive early on that one Monday are asked, so workers on later shifts / absent that day / arriving after the first 40 are excluded \Rightarrow sample not representative. | BB | B1 identifies an excluded group; B1 links to bias / not representative. |
| 3(c) | $\frac{200}{800} \times 40 = 10$ | MMA | M1 $\frac{200}{800}$ <i>oe</i> (0.25); M1 (<i>dep</i>) $\times 40$; A1 10 . |
| 4(a) | Midpoints 5, 15, 30, 55; $\sum fx = 50 + 300 + 450 + 275 = 1075$; $1075 \div 50$ | MMMA | M1 midpoints (at least 3 correct); M1 $\sum f \times$ midpoint; M1 (<i>dep</i>) $\div 50$; A1 21.5 (minutes). |

| Q | Answer / working | Marks | Notes |
|------|---|-------------|--|
| 4(b) | The midpoint of each class is used, so exact data values are not known once grouped. | B1 | oe. |
| 4(c) | $\frac{n}{2} = 25$; cumulative frequency 10, 30, ... so 25th value lies in the second class. | MA | M1 finds the 25th/26th position; A1 $10 < t \leq 20$. |
| 4(d) | A frequency (a count of customers) cannot be negative. | B1 | oe. |
| 5(a) | $44 - 28 = 16$ | B1 | |
| 5(b) | $1.5 \times 16 = 24$; lower boundary $28 - 24 = 4$, upper boundary $44 + 24 = 68$; $70 > 68$. | MMAA | M1 $1.5 \times \text{IQR}$; M1 both boundaries (4 and 68); A1 compares 70 with 68 (and 12 with 4); A1 (dep) 70 is an outlier (12 is not). |
| 5(c) | It should be kept / retained in the data set. | B1 | Do <i>not</i> accept “delete / remove it”. |
| 5(d) | $\frac{3(38.2 - 35)}{14.5} = \frac{9.6}{14.5}$ | MMA | M1 use of $\frac{3(\text{mean} - \text{median})}{\text{sd}}$; M1 correct substitution; A1 0.66 (awrt 0.66). Positive skew. |
| 6(a) | $\frac{72 - 60}{8}$ | MA | M1 substitution (Spanish); A1 1.5 . |
| 6(b) | $\frac{80 - 70}{12}$ | MA | M1 substitution (French); A1 0.83 (awrt; accept $\frac{5}{6}$). |
| 6(c) | Spanish score (1.5) > French score (0.83); Maria is more standard deviations above the mean in Spanish; so she performed better in the Spanish test relative to her class. | BBB | B1 states Spanish $z >$ French z (ft their values); B1 interpretation (further above the mean / more sd above average); B1 conclusion: Spanish, in context. |
| 7(a) | Frequency densities: $\frac{30}{30} = 1.0$, $\frac{40}{20} = 2.0$, $\frac{25}{10} = 2.5$, $\frac{30}{20} = 1.5$, $\frac{10}{20} = 0.5$. Histogram with bars of these heights and correct widths. | B3 | B3 all 5 bars correct (height = frequency density, correct widths, no gaps). B2 for 3 or 4 correct bars; B1 for 1 or 2 correct bars. Accept correct FD values seen in the table for B1 . |
| 7(b) | Lower class $0 < s \leq 30$ contributes 30. Part of $30 < s \leq 50$: $\frac{45 - 30}{50 - 30} \times 40 = \frac{15}{20} \times 40 = 30$ (or area 15×2.0). Total $30 + 30$. | MMA | M1 $\frac{15}{20} \times 40$ oe (fraction of the $30 < s \leq 50$ class, or area $15 \times$ their FD); M1 adds the whole $0 < s \leq 30$ class (+30); A1 60 . |

| Q | Answer / working | Marks | Notes |
|-------|--|------------|---|
| 7(c) | The exact scores of individual students within a class are not known (the data is grouped); we assume the scores are spread evenly (uniformly) across each class interval. | BB | B1 estimate because data is grouped / exact values unknown; B1 states the assumption of an even/uniform spread within the class. |
| 8(a) | $\frac{1380}{1200} \times 100$ | MA | M1 $\frac{1380}{1200} \times 100$ oe; A1 115 . |
| 8(b) | The price increased by 15% between 2020 and 2023. | B1 | ft (their index –100)%. Must state an <i>increase</i> . |
| 8(c) | $\frac{125}{100} \times 1200$ (oe 1.25×1200) | MA | M1 $\frac{125}{100} \times 1200$ oe; A1 £1500 . |
| 8(d) | A weighted index takes account of how much is spent on (the relative importance of) the different items, so it better reflects typical overall spending. | B1 | Allow: items do not all contribute equally / reflects amounts bought. |
| 9(a) | 0.85 is close to +1: strong (positive) agreement between the judges – they ranked the competitors in a very similar order. | BB | B1 strong/positive; B1 in context (judges agree). |
| 9(b) | –0.92 is close to –1: strong negative <i>linear</i> correlation. | BB | B1 strong negative; B1 states <i>linear</i> . |
| 10(a) | $n(A \cup B) = 100 - 35 = 65$, so intersection = $58 + 44 - 65 = 37$. Then A only = 21, B only = 7, outside = 35. | BBB | B1 intersection = 37 (or $n(A \cup B) = 65$ seen); B1 A only = 21 and B only = 7; B1 35 outside and all regions sum to 100. |
| 10(b) | $\frac{21 + 7 + 37}{100} = \frac{65}{100}$ | MA | M1 their union $\div 100$; A1 0.65 (oe $\frac{13}{20}$, 65%). |
| 10(c) | $\frac{58}{100}$ | B1 | 0.58 (accept $\frac{29}{50}$, 58%). |
| 10(d) | $P(\text{fever} \text{headache}) = \frac{37}{58} = 0.638$; $P(\text{fever} \text{no headache}) = \frac{7}{42} = 0.16\dot{6}$; $0.638 \div 0.16\dot{6}$ | MMA | M1 both conditional probabilities (ft their Venn); M1 (dep) divides them; A1 awrt 3.83 . |
| 11 | 20 spins is a small sample, so the result could easily be due to chance – it does not show the spinner is <i>definitely</i> biased. Maria should spin the spinner many more times (increase the number of trials) before drawing a conclusion. | BB | B1 small sample / could be chance / cannot be certain it is biased; B1 spin many more times for a more reliable conclusion. |
| 12 | Nigeria (wide base, narrowing sharply): higher birth rate and lower life expectancy. United Kingdom (narrow, even base, larger proportion of older people): lower birth rate and higher life expectancy. | BBB | B1 birth-rate comparison (Nigeria higher); B1 life-expectancy comparison (UK higher); B1 clear reference to the shape of the pyramids supporting a comparison. |

| Q | Answer / working | Marks | Notes |
|------------------------|--|-----------------|--|
| 13(a) | $\bar{x} = \frac{1 + 2 + 4 + 4 + 5 + 6 + 6 + 8}{8} = \frac{36}{8} = 4.5; \bar{y} = \frac{42 + 36 + 64 + 46 + 58 + 74 + 56 + 80}{8} = \frac{456}{8} = 57.$ | MA | M1 a correct method for \bar{x} or \bar{y} ($\Sigma \div 8$); A1 mean point (4.5, 57) (both coordinates). |
| 13(b) | Mean point (4.5, 57) plotted; a single ruled line of best fit drawn through it. | BB | B1 mean point plotted at (4.5, 57); B1 a sensible ruled straight line of best fit passing through the mean point (positive gradient, good fit). |
| 13(c) | Using two points on the line, e.g. gradient $\approx \frac{57 - 30}{4.5 - 0} = 6$ and intercept ≈ 30 , giving $y = 30 + 6x$. | MA | M1 a correct method for the gradient of <i>their</i> line (rise \div run from two points on the line); A1 a correct equation, a in range 24–36 and b in range 5–7, e.g. $y = 30 + 6x$ (ft their line). |
| 13(d) | 12 hours is outside the range of the data (1–8 hours), so this is extrapolation and the trend may not continue; also the equation would predict a score above 100%, which is impossible. | B1 | B1 any one valid reason (extrapolation / outside data range / score would exceed 100%). |
| Total for paper | | 80 marks | |