InkMaths

ANSWER BOOK

Representing and interpreting data

Statistics · Year 12-13

7 questions · 21 marks

NAVY EDITION

inkphysics.co.uk/maths/ · youtube.com/@InkPhysics

SECTION A · Warm-up

A1 Density = frequency/width = 30/5 = 6. M1 for frequency divided by width, A1 for 6. In a histogram the area of a bar carries the frequency, so its height has to be frequency per unit of width. That is why bars over classes of different widths can only be compared by area. A wide class spreads its frequency thinly and draws a low bar. [2]

A2 Frequency = density × width = 1.5 × 20 = 30. M1 for density times width, A1 for 30. A low wide bar can hold exactly as much frequency as a tall narrow one. The height on its own tells you only how concentrated the data are. [2]

SECTION B · Standard

B1 The first bar's density is 40/20 = 2, so the second bar's density is 2 × 1.6 = 3.2 and its frequency is 3.2 × 5 = 16. M1 for the first bar's density, M1 for scaling it by 1.6, A1 for 16. Heights compare densities and never frequencies. The second bar is the taller of the two yet holds well under half the frequency, because it is only a quarter as wide. Answering 64 by scaling the frequency directly is the trap. [3]

B2 IQR = 36 - 20 = 16, so 1.5 × IQR = 24 and the fences sit at 20 - 24 = -4 and 36 + 24 = 60. The value 62 lies beyond the upper fence, so it is an outlier. The value 58 does not, so it is large but within bounds. M1 for 1.5 × IQR, A1 for both fences, B1 for the two verdicts. State both fences and then compare, because the fence is what justifies the verdict. The rule exists so that the same decision is reached every time, rather than by eye. [3]

B3 Positive skew. The lower half of the middle 50% is squashed and the upper half stretched, so values above the median straggle further than values below it. B1 for positive skew, B1 for what it says about the data. Most observations sit low with a long tail of high values, which is the classic shape of an income distribution. [2]

B4 Route B is typically two minutes quicker, since its median is 28 minutes against route A's 30. Route B is far less predictable, though. Its middle half spans 18 minutes against route A's 6. A commuter who has to arrive on time would sensibly take route A and trade the two minutes for the consistency. B1 for comparing the medians, B1 for comparing the spread and B1 for the recommendation, all in minutes and all naming the routes. [3]

SECTION C · Stretch

C1 [6]

Turn every density back into a frequency first, using frequency = density × width: 1.2 × 10 = 12, 2.4 × 15 = 36, 1.0 × 15 = 15 and 0.35 × 20 = 7. These total 70, which confirms the readings. For the count below 20, take the whole first class and the part of the second class up to 20, which is 10 of its 15 units of width: 12 + (10/15) × 36 = 12 + 24 = 36 observations. For the median, the position is 70/2 = 35. The first class carries 12, so the median lies 23 into the second class: 10 + (23/36) × 15 = 19.6. M1 for converting the densities, A1 for the four frequencies, M1 for the proportion of the second class below 20, A1 for 36 observations, M1 for interpolating for the median, A1 for 19.6. Both parts assume the data are spread evenly within each class, which is the standard assumption behind interpolation and worth stating. As a check, 36 of the 70 observations fall below 20 and that is just over half, so a median a little under 20 is exactly right.