# Mark scheme and fully worked solutions

Shadow Paper 2A Statistical Techniques

## Question 1

### 1(a)

Planned numbers:

| Type | Number |
|---|---|
| 1-bed flat | 14 |
| 2-bed flat | 16 |
| 2-bed house | 20 |
| 3-bed house | 18 |
| 4-bed house | 7 |
| 5-bed house | 5 |

Flats: $14 + 16 = 30$. Houses: $20 + 18 + 7 + 5 = 50$.

Ratio houses : flats

$$50 : 30 = \frac{50}{10} : \frac{30}{10} = 5 : 3$$

**Answer:** $5 : 3$.

### 1(b)

Cycle spaces:

$$\begin{aligned} \text{1-bed flats: } & 14 \times 1 = 14 \\ \text{2-bed flats: } & 16 \times 2 = 32 \\ \text{2-bed houses: } & 20 \times 2 = 40 \\ \text{3-bed houses: } & 18 \times 3 = 54 \\ \text{4-bed houses: } & 7 \times 3 = 21 \\ \text{5-bed houses: } & 5 \times 4 = 20 \end{aligned}$$

Total minimum required:

$$14 + 32 + 40 + 54 + 21 + 20 = 181$$

Planned spaces: 204.

Since $204 > 181$, the plans meet the minimum requirement (and allow visitor spaces).

**Answer:** Yes, the plans meet the minimum requirement.

### 1(c)

Fixed properties (not changing):

$$14\ (\text{1-bed flats}) + 18\ (\text{3-bed houses}) + 7\ (\text{4-bed}) + 5\ (\text{5-bed}) = 44$$

Total properties must be 80, so remaining:

$$80 - 44 = 36$$

Let

$$x = \text{number of 2-bed flats}, \quad y = \text{number of 2-bed houses}.$$

Then

$$x + y = 36.$$

Affordable housing types: 1-bed flats, 2-bed flats, 3-bed houses.
Number of affordable properties:

$$14\ (\text{1-bed}) + x\ (\text{2-bed flats}) + 18\ (\text{3-bed}) = 32 + x.$$

At least 25% of 80 must be affordable:

$$0.25 \times 80 = 20.$$

We have $32 + x \geq 32$, which is already greater than 20 for any $x \geq 0$. So the affordable condition is always satisfied.
To maximise profit, we want as many 2-bed houses as possible, so take $x$ as small as possible:

$$x = 0, \quad y = 36.$$

**Answer:** 2-bed flats $= 0$, 2-bed houses $= 36$.

## Question 2

### 2(a)

Any two sensible improvements, for example:

- Add a clear title and axis labels (e.g. “Percentage of electricity generated by fuel type”).
- Show exact percentages or add data labels to make reading values easier.
- Use clearer colours or patterns and a well-explained key.
- Separate bars for each fuel type instead of stacked bars.

**Answer:** e.g. “Add axis labels and a title; show exact percentages on the bars.”

## 2(b)

From the Preliminary Material (England 2024):

$$\text{Fossil fuels} = 92\,480 \text{ GWh}, \quad \text{Renewables} = 89\,740 \text{ GWh}.$$

**Morning Herald:**

$$\frac{\text{Renewables}}{\text{Fossil}} = \frac{89\,740}{92\,480}$$

Compute:

$$92\,480 - 89\,740 = 2\,740, \quad \frac{89\,740}{92\,480} \approx 0.970$$

So renewables are about 97% of fossil fuels, which is more than 80%.

**Conclusion:** The Morning Herald's claim is valid.

Within renewables:

$$\text{Wind} = 39\,820, \quad \text{Other renewables} = 89\,740 - 39\,820 = 49\,920.$$

Ratio wind : other renewables:

$$39\,820 : 49\,920.$$

Divide both by 10:

$$3\,982 : 4\,992.$$

Approximate:

$$\frac{3\,982}{4\,992} \approx 0.798 \approx \frac{4}{5}.$$

So wind : other renewables $\approx 4 : 5$.

**Conclusion:** The Daily Energy Review's claim is also reasonable.

**Answer:** Both claims are valid; renewables are about 97% of fossil fuels, and wind : other renewables is about $4 : 5$.

## 2(c)

$$4\,300 \text{ GWh} = 4\,300 \times 1\,000\,000 = 4.3 \times 10^9 \text{ kWh}.$$

Price per kWh: $15.2\text{p} = £0.152$.

$$\text{Value} = 4.3 \times 10^9 \times 0.152.$$

Compute:

$$4.3 \times 0.152 = 0.6536.$$

So

$$\text{Value} = 0.6536 \times 10^9 = £653.6 \text{ million}.$$

This is greater than £600 million.

**Answer:** About £654 million; Anna is correct.

### 2(d)

From Chart 1, Wales renewable share in 2024 is $35.4\%$.

$$\text{Total generation} = \frac{\text{Renewables}}{\text{Share}} = \frac{7\,900}{0.354}.$$

Compute:

$$\frac{7\,900}{0.354} \approx 22\,316.38 \text{ GWh}.$$

Rounded suitably:

$$\text{Total} \approx 22\,300 \text{ GWh}.$$

**Answer:** Approximately $22\,300$ GWh.

### 2(e)

Bobby has added the four national percentages and divided by 4. The UK average should be a *weighted* mean, using the total electricity generated in each nation. Because the nations generate different amounts, a simple average of the four percentages does not give the UK figure.

**Answer:** He should use a weighted mean based on total generation in each country, not a simple average.

## Question 3

Mean of the standardised normal distribution: $0$. Standard deviation of the standardised normal distribution: $1$.

**Answer:** Match "Mean" to 0 and "Standard deviation" to 1.

## Question 4

IQ $\sim N(\mu, 400)$, so $\sigma = 20$. Sample size $n = 25$, sample mean $\bar{x} = 106$.

### 4(a)

For a 90% confidence interval, the critical value is $z = 1.64$.

**Answer:** $1.64$.

### 4(b)

Standard error:

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = \frac{20}{5} = 4.$$

Margin of error:

$$1.64 \times 4 = 6.56.$$

Confidence interval:

$$106 - 6.56 = 99.44, \quad 106 + 6.56 = 112.56.$$

**Answer:** $99.4 < \mu < 112.6$ (or equivalent rounding).

### 4(c)

Claim: mean IQ is 115. 115 is greater than 112.6, so it lies outside the 90% confidence interval. Therefore the claim is not supported by this sample.

**Answer:** The claim is not supported because 115 is outside the confidence interval.

## Question 5

### 5(a)

150 scores arranged row-wise, 15 per row.

- 12th score: row 1, column 12 $\Rightarrow$ 1.
- 89th score: 5 full rows $= 75$ scores, so 89th is in row 6, position $89 - 75 = 14$. Row 6, column 14 is 2.
- 143rd score: 9 full rows $= 135$ scores, so 143rd is in row 10, position $143 - 135 = 8$. Row 10, column 8 is 5.

**Answer:** $12 \to 1,\ 89 \to 2,\ 143 \to 5.$

### 5(b)

Carly uses the largest sample size ($n = 30$), so her estimate is based on more data and is likely to be closest to the true mean.

**Answer:** Carly, because she uses the largest sample.

### 5(c)

Sample from 5(a): scores $1, 2, 5$.

Mean of this sample:

$$\bar{x}_{(a)} = \frac{1+2+5}{3} = \frac{8}{3} \approx 2.67.$$

Total of scores in each sample:

$$\begin{aligned} \text{Ali: } & 10 \times 3.7 = 37.0 \\ \text{Becky: } & 10 \times 4.1 = 41.0 \\ \text{Carly: } & 30 \times 3.9 = 117.0 \\ \text{Our sample: } & 3 \times 2.67 \approx 8.0 \end{aligned}$$

Total scores:

$$37 + 41 + 117 + 8 = 203.$$

Total number of scores:

$$10 + 10 + 30 + 3 = 53.$$

Combined point estimate:

$$\bar{x} = \frac{203}{53} \approx 3.830.$$

**Answer:** Approximately $3.83$.

### 5(d)

For a fair 6-sided dice, the expected mean is:

$$\frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$$

Our estimate is about 3.83, which is higher than 3.5. However, this could be due to sampling variation. A reasonable conclusion is that there is not enough evidence from these samples alone to be certain the dice is unfair.

**Answer:** The estimate (3.83) is higher than 3.5, but this may be due to sampling; we cannot be sure the dice is unfair.

## Question 6

### 6(a)

Perfect negative correlation: $-1$.

**Answer:** $-1$.

### 6(b)

As time increases, distance generally increases. The points suggest a strong positive correlation.
**Answer:** Strong positive correlation.

## Question 7

Heights $X$ are normally distributed with mean 1.62 and standard deviation 0.28.

### 7(a)

$$X \sim N(1.62,\ 0.28^2).$$

**Answer:** $X \sim N(1.62,\ 0.28^2)$.

### 7(b)

$$P(X < 2.0) = P\left(Z < \frac{2.0 - 1.62}{0.28}\right) = P\left(Z < \frac{0.38}{0.28}\right) = P(Z < 1.357\ldots).$$

Using tables, $P(Z < 1.36) \approx 0.913$.
**Answer:** Approximately $0.91$.

### 7(c)

$$P(X < 1.4) = P\left(Z < \frac{1.4 - 1.62}{0.28}\right) = P\left(Z < \frac{-0.22}{0.28}\right) = P(Z < -0.7857\ldots).$$

From tables, $P(Z < 0.79) \approx 0.785$.

$$P(Z < -0.79) = 1 - 0.785 = 0.215.$$

**Answer:** Approximately $0.21$.

### 7(d)

60% have height $> k \Rightarrow$ 40% have height $< k$.

$$P(X < k) = 0.40.$$

From tables, $P(Z < -0.25) \approx 0.401$, so $z \approx -0.25$.

$$\frac{k - 1.62}{0.28} = -0.25 \Rightarrow k - 1.62 = -0.25 \times 0.28 = -0.07 \Rightarrow k = 1.62 - 0.07 = 1.55.$$

**Answer:** $k \approx 1.55$ m.

# Question 8

Data:

| Student | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $I$ | $J$ | $K$ | $L$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 24 | 82 | 49 | 73 | 38 | 59 | 78 | 29 | 31 | 42 | 66 | abs |
| $y$ | 42 | 78 | 44 | 80 | 72 | 55 | 64 | abs | 36 | 45 | 69 | 52 |

## 8(a)

On the scatter graph, student E $(38, 72)$ lies away from the general trend (high $y$ for relatively low $x$).
**Answer:** Student E.

## 8(b)

Use students A, B, C, D, F, G, I, J, K (exclude E and missing H, L). There are $n = 9$ points.
Compute sums:

$$\sum x = 24 + 82 + 49 + 73 + 59 + 78 + 31 + 42 + 66 = 504,$$
$$\sum y = 42 + 78 + 44 + 80 + 55 + 64 + 36 + 45 + 69 = 513.$$

$$\begin{aligned}\sum x^2 &= 24^2 + 82^2 + 49^2 + 73^2 + 59^2 + 78^2 + 31^2 + 42^2 + 66^2 \\ &= 576 + 6\,724 + 2\,401 + 5\,329 + 3\,481 + 6\,084 + 961 + 1\,764 + 4\,356 \\ &= 31\,676.\end{aligned}$$

$$\begin{aligned}\sum xy &= 24 \cdot 42 + 82 \cdot 78 + 49 \cdot 44 + 73 \cdot 80 + 59 \cdot 55 + 78 \cdot 64 + 31 \cdot 36 + 42 \cdot 45 + 66 \cdot 69 \\ &= 1\,008 + 6\,396 + 2\,156 + 5\,840 + 3\,245 + 4\,992 + 1\,116 + 1\,890 + 4\,554 \\ &= 31\,197.\end{aligned}$$

Slope:

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} = \frac{9 \times 31\,197 - 504 \times 513}{9 \times 31\,676 - 504^2}.$$

Compute numerator:

$$9 \times 31\,197 = 280\,773, \quad 504 \times 513 = 258\,552,$$

$$\text{Numerator} = 280\,773 - 258\,552 = 22\,221.$$

Denominator:

$$9 \times 31\,676 = 285\,084, \quad 504^2 = 254\,016,$$

$$\text{Denominator} = 285\,084 - 254\,016 = 31\,068.$$

$$b = \frac{22\,221}{31\,068} \approx 0.716.$$

Intercept:

$$a = \frac{\sum y - b\sum x}{n} = \frac{513 - 0.716 \times 504}{9}.$$

Compute:

$$0.716 \times 504 \approx 361.664, \quad 513 - 361.664 = 151.336,$$

$$a = \frac{151.336}{9} \approx 16.815.$$

So the regression line is approximately:

$$y = 16.8 + 0.72x \quad \text{(rounded)}.$$

**Answer:** $y \approx 16.8 + 0.72x$.

## 8(c)

Test 2 mark for H ($x = 29$):

$$y_H = 16.8 + 0.72 \times 29.$$

$$0.72 \times 29 = 20.88, \quad y_H \approx 16.8 + 20.9 = 37.7 \approx 38.$$

Test 1 mark for L ($y = 52$):

$$52 = 16.8 + 0.72x \Rightarrow 52 - 16.8 = 0.72x \Rightarrow 35.2 = 0.72x \Rightarrow x = \frac{35.2}{0.72} \approx 48.9 \approx 49.$$

**Answers:** H's test 2 mark $\approx 38$; L's test 1 mark $\approx 49$.

## 8(d)

Totals $t = x + y$ including outlier and estimated marks:

| Student | $t$ |
|---|---|
| $A$ | $24 + 42 = 66$ |
| $B$ | $82 + 78 = 160$ |
| $C$ | $49 + 44 = 93$ |
| $D$ | $73 + 80 = 153$ |
| $E$ | $38 + 72 = 110$ |
| $F$ | $59 + 55 = 114$ |
| $G$ | $78 + 64 = 142$ |
| $H$ | $29 + 38 = 67$ |
| $I$ | $31 + 36 = 67$ |
| $J$ | $42 + 45 = 87$ |
| $K$ | $66 + 69 = 135$ |
| $L$ | $49 + 52 = 101$ |

Sum of totals:

$$\begin{aligned} 66 + 160 &= 226 \\ 226 + 93 &= 319 \\ 319 + 153 &= 472 \\ 472 + 110 &= 582 \\ 582 + 114 &= 696 \\ 696 + 142 &= 838 \\ 838 + 67 &= 905 \\ 905 + 67 &= 972 \\ 972 + 87 &= 1\,059 \\ 1\,059 + 135 &= 1\,194 \\ 1\,194 + 101 &= 1\,295. \end{aligned}$$

So

$$\sum t = 1\,295.$$

Mean:

$$m = \frac{1\,295}{12} \approx 107.9167 \approx 107.9.$$

Grade 3 boundaries: $0.8m \leq t < 1.1m$.

$$0.8m = 0.8 \times 107.9167 \approx 86.333, \quad 1.1m = 1.1 \times 107.9167 \approx 118.708.$$

So Grade 3 if $86.3 \leq t < 118.7$.

Check each student:

| Student | $t$ | Grade 3? |
|---|---|---|
| $A$ | 66 | $< 86.3 \Rightarrow$ No |
| $B$ | 160 | $> 118.7 \Rightarrow$ No |
| $C$ | 93 | $\in [86.3, 118.7) \Rightarrow$ Yes |
| $D$ | 153 | $> 118.7 \Rightarrow$ No |
| $E$ | 110 | $\in [86.3, 118.7) \Rightarrow$ Yes |
| $F$ | 114 | $\in [86.3, 118.7) \Rightarrow$ Yes |
| $G$ | 142 | $> 118.7 \Rightarrow$ No |
| $H$ | 67 | $< 86.3 \Rightarrow$ No |
| $I$ | 67 | $< 86.3 \Rightarrow$ No |
| $J$ | 87 | $\in [86.3, 118.7) \Rightarrow$ Yes |
| $K$ | 135 | $> 118.7 \Rightarrow$ No |
| $L$ | 101 | $\in [86.3, 118.7) \Rightarrow$ Yes |

**Answer:** Students C, E, F, J and L are awarded Grade 3.

# Suggested grade boundaries for this 60-mark paper

These are *suggested* teacher boundaries (not official AQA):

| **Grade** | **Marks** |
|---|---|
| A | 48–60 |
| B | 40–47 |
| C | 32–39 |
| D | 24–31 |
| E | 16–23 |
| U | 0–15 |

**Note:** You can adjust these to fit your cohort or to align with past series.

END OF MARK SCHEME