Question 1

Worked Solution

Part (a): Mean

Table 1 lists windspeed values 6, 7, 8, 9, 11, 12, 13, 14, 16 with frequencies 2, 3, 2, 2, 3, 1, 2, 1, 2. The 13 values recorded as n/a are excluded (data cleaning — non-numeric entries cannot be used in calculations).

This leaves $n = 18$ valid data values. Enter the (x, f) pairs into your calculator in statistics mode to obtain the mean directly.

Mean = 10.2 knots (awrt 10.2)

Part (b): Standard deviation

Using the same calculator entries from part (a), read off σ_x directly.

Standard deviation = 3.17 knots (awrt 3.17). Units: **knots** (or kn).

Part (c): Which month had mean 11.57?

From knowledge of the large data set, wind speeds at Camborne tend to be highest in the autumn months. Mean 11.57 is the largest in Table 2, indicating a particularly windy month.

October — it is windier in autumn / it is the month associated with the hurricane season / it is the latest (most autumnal) month in the data set.

Part (d)(i): Meaning of * on box plots

The $*$ symbol represents **outliers** — individual data values that lie more than $1.5 \times$ IQR beyond the nearer quartile.

Part (d)(ii): Which month corresponds to box plot Y?

Box plot Y has a low median (around 6–8, consistent with a lowish mean) but a large range/IQR and an outlier extending well above 20, indicating high spread. From Table 2, month B has mean 8.26 and the largest standard deviation (3.89), consistent with a distribution that has a relatively low centre but high variability.

Box plot Y most likely corresponds to month **B**: the low median is consistent with a mean of 8.26, and the large spread/outlier is consistent with the largest standard deviation in Table 2 (3.89).

Question 2

Worked Solution

Part (a): Sampling technique

Systematic sampling

Part (b): Why the process may not generate a sample of size 20

In the large data set, some days have missing (gap) entries because data was not recorded. If a selected position corresponds to a missing entry, that value cannot be included, reducing the sample below 20.

Part (c): Standard deviation from summary statistics

Given $n = 20$, $\sum t = 374$, $\sum t^2 = 7600$.

Since only summary statistics (not raw values) are available, use the formula directly.

Mean:

$$\bar{t} = \frac{374}{20} = 18.7$$

Standard deviation:

$$\sigma_t = \sqrt{\frac{\sum t^2}{n} - \bar{t}^2} = \sqrt{\frac{7600}{20} - (18.7)^2} = \sqrt{380 - 349.69} = \sqrt{30.31} = 5.505\ldots$$

Standard deviation $= 5.51$ knots (awrt 5.51) [Accept $s_t = 5.65$ using $n - 1$]

Question 3

Worked Solution

Part (a): Why stratified random sampling cannot be used

It is not possible to construct a sampling frame for the fish in the lake — the fish cannot be individually identified or listed, so random selection within each stratum (species) is not possible.

Part (b): How to take a representative sample of 160

Total estimated population: $450 + 300 + 850 = 1600$ fish.

Sample sizes proportional to population:

$$\begin{aligned} \text{Mirror carp: } & \tfrac{450}{1600} \times 160 = 45 \\ \text{Leather carp: } & \tfrac{300}{1600} \times 160 = 30 \\ \text{Common carp: } & \tfrac{850}{1600} \times 160 = 85 \end{aligned}$$

Use **quota sampling**: fish until 45 mirror carp, 30 leather carp, and 85 common carp have been caught. Return (ignore) any fish caught once its quota is full.

Part (c): Standard deviation from summary statistics

Given $\sum fm = 692$, $\sum fm^2 = 3053$, $n = 160$.

Mean:

$$\bar{m} = \frac{692}{160} = 4.325$$

Standard deviation:

$$\sigma = \sqrt{\frac{\sum fm^2}{n} - \bar{m}^2} = \sqrt{\frac{3053}{160} - \left(\frac{692}{160}\right)^2} = \sqrt{19.081 - 18.706} = \sqrt{0.3751} = 0.6125\ldots$$

Estimated standard deviation $= 0.613$ kg (awrt 0.613)

Part (d): Effect of correcting the transposed figures

(i) Correcting 2.3 to 3.2:

The value 2.3 lies in the class $2 \leq w < 3.5$ (midpoint 2.75). The corrected value 3.2 also lies in the same class $2 \leq w < 3.5$ (midpoint 2.75). The class midpoint used in the estimate is unchanged.

(i) **No effect**: both 2.3 and 3.2 fall in the class $2 \leq w < 3.5$, so the midpoint used and the frequency distribution are unchanged.

(ii) Correcting 4.6 to 6.4:

The value 4.6 is in the class $4.5 \leq w < 5$ (midpoint 4.75), but 6.4 moves to the class $5 \leq w < 6$ (midpoint 5.5). The corrected value is now further from the mean (≈ 4.3), since $6.4 - 4.6 \approx 1.8 \approx 3\sigma$.

(ii) **Standard deviation increases**: 6.4 falls in a different (higher) class than 4.6, placing a data value further from the mean and increasing the estimated spread.

Question 4

Worked Solution

Part (a): Why Joshua needs to clean the data

The large data set records trace rainfall ($0 < r \le 0.05$ mm) as the text entry "`tr`". This is non-numeric and must be converted to a number before any calculation can be performed.

Part (b): Upper quartile by linear interpolation

Total $n = 121 + 10 + 24 + 12 + 17 = 184$.

Q_3 position: $\frac{3\times184}{4} = 138$th value.

Cumulative frequencies: $r < 0.5$: 121; $r < 1.0$: 131; $r < 5.0$: 155.

The 138th value falls in the class $1.0 \le r < 5.0$ (24 values, starting from cumulative position 131):

$$Q_3 = 1.0 + \frac{138 - 131}{24} \times (5.0 - 1.0) = 1.0 + \frac{7}{24} \times 4 = 1.0 + 1.167 = 2.17 \text{ mm}$$

$Q_3 \approx 2.17$ mm (awrt 2.17)

Part (c): Standard deviation from summary statistics

Given $\sum fx = 539.75$, $\sum fx^2 = 7704.1875$, $n = 184$.

$$\sigma = \sqrt{\frac{7704.1875}{184} - \left(\frac{539.75}{184}\right)^2} = \sqrt{41.871 - (2.934)^2} = \sqrt{41.871 - 8.608} = \sqrt{33.263} = 5.768\ldots$$

Standard deviation ≈ 5.77 mm (awrt 5.77)

Part (d)(i): Assumption when using class midpoints

It assumes that the data values are **uniformly distributed within each class** — that the midpoint is representative of the typical value in that class.

Part (d)(ii): Why this assumption does not hold here

In the large data set, most values in the first class $0 \le r < 0.5$ are recorded as `tr` (trace) or exactly 0, meaning the values are **concentrated near 0** rather than uniformly spread up to 0.5. The midpoint 0.25 overestimates the typical value in this class.

Part (d)(iii): Is the actual mean larger or smaller than the estimate?

The first class (frequency 121, by far the largest) has values concentrated at or near

0, which is below the midpoint 0.25 used in the estimate. The estimate therefore overstates the contribution from this class.

The actual mean is likely **smaller** than the estimate, because the majority of values in the largest class ($0 \le r < 0.5$) are at or near 0, below the midpoint 0.25 assumed.

Question 5

Worked Solution

Part (a): Units of x

Hectopascals (hPa) [also accept millibars / mb]

Part (b): Mean Daily Mean Pressure from coded data

Coding: $y = x - 1010$, so $x = y + 1010$ and $\bar{x} = \bar{y} + 1010$.

$$\bar{y} = \frac{\sum y}{n} = \frac{214}{30} = 7.1\overline{3}$$
$$\bar{x} = 7.1\overline{3} + 1010 = 1017.13\ldots$$

Mean Daily Mean Pressure = 1017 hPa (awrt 1017)

Part (c): Standard deviation from coded summary statistics

Adding a constant does not affect the standard deviation, so $\sigma_x = \sigma_y$.

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{5912}{30} - (7.1\overline{3})^2} = \sqrt{197.07 - 50.88} = \sqrt{146.19} = 12.09\ldots$$

Standard deviation = 12.1 hPa (awrt 12.1)

Part (d): Cardinal Wind Directions

The pressures (1029, 1028, 1028) are all well above 1013 hPa, confirming **high pressure**. Under high pressure in the UK, winds circulate **clockwise**. Wind direction is the direction the wind blows *from*.

The three locations from north to south are: **Leuchars** (Scotland), **Heathrow** (London), **Hurn** (Dorset). Heathrow has the highest pressure (1029), so the high-pressure centre is closest to Heathrow. For clockwise circulation around a high centre near the south:

- **Heathrow** (near centre, south): wind from **NE**
- **Hurn** (south of centre, on the southern edge): wind from **E**
- **Leuchars** (north of centre): wind from **W**

Heathrow: **NE**, Hurn: **E**, Leuchars: **W**

High pressure ⇒ clockwise circulation. Locations north to south: Leuchars, Heathrow, Hurn. Wind direction is the direction wind blows from.

Question 6

Worked Solution

Frequency distribution for exam times t (minutes), $n = 200$.

Classes (using continuous boundaries): 10.5–20.5 ($f = 62$, $m = 15.5$); 20.5–25.5 ($f = 88$, $m = 23$); 25.5–30.5 ($f = 16$, $m = 28$); 30.5–35.5 ($f = 13$, $m = 33$); 35.5–45.5 ($f = 11$, $m = 40.5$); 45.5–60.5 ($f = 10$, $m = 53$).

Part (a): Mean and standard deviation

Enter the (m, f) pairs into your calculator in statistics mode. The given value $\sum ft^2 = 134{,}281.25$ can be used to verify. Read off $\bar{t}$ and σ_t directly.

Mean $\bar{t} = 24.2$ minutes (awrt 24.2)
Standard deviation $\sigma_t = 9.29$ minutes (awrt 9.29; accept $s_t = 9.32$)

Check: $\bar{t} = 4837.5/200 = 24.1875$; $\sigma = \sqrt{134281.25/200 - 24.1875^2} = \sqrt{86.37} = 9.294$

Part (b): Median by linear interpolation

The median is the 100th/101st value (average position 100.5). Cumulative: 62 (up to 20.5); 150 (up to 25.5).

The 100.5th value lies in the class 20.5–25.5 (88 values, starting from cumulative 62):

$$Q_2 = 20.5 + \frac{100.5 - 62}{88} \times 5 = 20.5 + \frac{38.5}{88} \times 5 = 20.5 + 2.19 = 22.7 \text{ min}$$

Median ≈ 22.7 minutes (awrt 22.7)

Part (c): Show lower quartile ≈ 18.6

Q_1 position: 50th/51st values (average 50.5). Cumulative through 20.5 is 62, so both lie in the first class 10.5–20.5 (62 values, starting from cumulative 0):

$$Q_1 = 10.5 + \frac{50.5 - 0}{62} \times 10 = 10.5 + \frac{50.5}{62} \times 10 = 10.5 + 8.145 = 18.6 \text{ min} \quad \checkmark$$

$Q_1 = 10.5 + \frac{50.5}{62} \times 10 = 18.6$ minutes (to 3 s.f.) ✓

Part (d): IQR

Q_3 position: 150th/151st values. The 150th value is exactly the last in the class 20.5–25.5 (cumulative 150), so $Q_3 = 25.5$.

$$\text{IQR} = Q_3 - Q_1 = 25.5 - 18.6 = 6.9 \text{ min}$$

IQR $= 6.9$ minutes

Part (e): Why mean and SD are not most appropriate

The data is **skewed** (the distribution has a long tail), so the median and IQR are more appropriate summary statistics than the mean and standard deviation.

Part (f): Effect of subtracting 5 minutes from all times

Subtracting a constant c from every value: shifts measures of *location* down by c, but leaves measures of *spread* unchanged.

Mean: decreases by 5 (to ≈ 19.2 min). **Standard deviation: unchanged.**

Median: decreases by 5 (to ≈ 17.7 min). Q_1**:** decreases by 5 (to ≈ 13.6 min).

IQR: unchanged — both quartiles decrease by 5, so their difference is the same.

Mean **decreases by 5** min; standard deviation **unchanged**.
Median **decreases by 5** min; Q_1 **decreases by 5** min; IQR **unchanged**.

Question 7

Worked Solution

Coding: $y = 1.4x - 20$, with $\bar{y} = 60.8$ and $\sigma_y = 6.60$.

Mean of x:

From $y = 1.4x - 20$, applying to the mean: $\bar{y} = 1.4\bar{x} - 20$, so:

$$\bar{x} = \frac{\bar{y} + 20}{1.4} = \frac{60.8 + 20}{1.4} = \frac{80.8}{1.4} = 57.714\ldots$$

Standard deviation of x:

Adding or subtracting a constant does not change the standard deviation. Multiplying by a constant scales it. Since $\sigma_y = 1.4\,\sigma_x$:

$$\sigma_x = \frac{\sigma_y}{1.4} = \frac{6.60}{1.4} = 4.714\ldots$$

Mean of $x = 57.7$ (awrt 57.7)
Standard deviation of $x = 4.71$ (awrt 4.71)

Question 8

Worked Solution

Keith's 7 daily values: 2.8, 5.6, 2.3, 9.4, 0.0, 0.5, 1.8. Jenny's 21 days: $\sum x = 84.6$.

Part (a): Mean over all 28 days

Add Keith's 7 values into your calculator to obtain their sum, then combine with Jenny's total.

Keith's sum: $2.8 + 5.6 + 2.3 + 9.4 + 0.0 + 0.5 + 1.8 = 22.4$

Combined total: $22.4 + 84.6 = 107$

$$\bar{x} = \frac{107}{28} = 3.821\ldots \approx 3.82 \text{ mm}$$

Mean $= \dfrac{107}{28} \approx 3.82$ mm (awrt 3.8)

Part (b): Effect of correcting 9.4 → 4.9 and 0.5 → 5.0

Correction 1: sum changes by $4.9 - 9.4 = -4.5$.

Correction 2: sum changes by $5.0 - 0.5 = +4.5$.

Net change to the sum $= -4.5 + 4.5 = 0$. The total is still 107.

No effect on the mean. One correction reduces the sum by 4.5 and the other increases it by exactly 4.5, leaving the total (and therefore the mean) unchanged.

Question 9

Worked Solution

Part (a): How $0 < r \leq 0.05$ is recorded

Recorded as "tr" (trace).

Part (b): Mean and standard deviation for Camborne, August 2015

Given $n = 31$, $\sum r = 174.9$, $\sum r^2 = 3523.283$.

Since only summary statistics are given, use the formulas directly.

(i) Mean:

$$\mu = \frac{\sum r}{n} = \frac{174.9}{31} = 5.642\ldots$$

(ii) Standard deviation:

$$\sigma_r = \sqrt{\frac{\sum r^2}{n} - \mu^2} = \sqrt{\frac{3523.283}{31} - (5.642)^2} = \sqrt{113.654 - 31.832} = \sqrt{81.822} = 9.046\ldots$$

Mean $= 5.64$ mm (awrt 5.64)
Standard deviation $= 9.05$ mm (awrt 9.05; accept $s = 9.19$)

Part (c): Does this support Dian's belief?

Dian's belief: mean August rainfall is *less* in the South than in the North.

Leuchars is in Scotland (North); Camborne is in Cornwall (South).

Mean for Camborne (South) $= 5.64$ mm; mean for Leuchars (North) $= 1.72$ mm.

The North (Leuchars) has a *smaller* mean than the South (Camborne) — this is the opposite of Dian's belief.

This does **not** support Dian's belief. Leuchars (North) has a mean of 1.72 mm which is *less than* Camborne (South) at 5.64 mm, so mean August rainfall appears greater in the South than the North.

Part (d): Why B(14, 0.27) may not be a reasonable model

The probability $p = 0.27$ of a day with no rain is **unlikely to be constant** across a 14-day summer event (it may vary by day or weather conditions). Additionally, consecutive days are unlikely to be **independent** — a dry day makes an adjacent dry day more (or less) likely depending on weather patterns.

End of Worked Solutions