A Level
Large Data Set

Practice Questions

Question	Scheme
1	
(a)	Mean rainfall = $\frac{0.025(7)+1.5(4)+3(3)}{7+4+3}=\underline{1.083...}$ mm
	Standard deviation = $\sqrt{\left(\frac{0.025^2(7)+1.5^2(4)+3^2(3)}{7+4+3}\right)-(1.083...)^2}=\underline{1.182...}$ mm
(b)	Rainfalls between (about) $0 \le r \le 2.26$ mm are within one standard deviation of the mean
(c/i)	Not suitable because her sample only consisted of 14 days from one location and from one month
(c/ii)	e.g. Use more data from more **UK** locations and months **Must reference UK locations**

Question	Scheme
2	
(a)	e.g. Cloud cover **Accept 'Daily mean windspeed on the Beaufort scale'**
(b)	• Generate (some) two digit random numbers • Enumerate the data points. For each random number chosen, select the corresponding data point on the enumerated list. • If the random number does not correspond to a data point (due to gaps or being out of range), ignore it and choose another. If a random number is repeated, ignore it and choose another • Continue in this way until 15 data points are chosen
(c)	Not reliable since he only used 15 data points, which is unlikely to be a good representation of the weather in Lemming in 2015
(d)	The large data set only contains data for the months May-October and not the whole year

Question	Scheme
3	
(a/i)	The large data set contains data for the months May-October and there are 184 days between (1^{st}) May and (31^{st}) October.
(a/ii)	e.g. The large data set contains gaps
(b)	(Starting from 1^{st} May), each day the total amount of rainfall in Leuchars in 2015 decreases by 0.0027 mm
(c)	$x = 3 \Rightarrow T = 16.551 - 0.0027(3) = \underline{16.5429}$
(d)	*Idea that* The daily mean rainfall in Leuchars (in 2015) does not decrease at a steady rate, but fluctuates **IGNORE** references to 'extrapolation' – the question asks for discussion about the unreliability for any day in Leuchars in 2015, not just those outside of the data range

Question	Scheme
4	
(a)	Simple random sampling
(b)	e.g. easier/quicker/etc. to process / analyse/etc. the data since the large data set has a lot of data points
(c)	4
(d)	*Idea that* The large data set does not contain information on cloud cover for Beijing

Question	Scheme
5	
(a)	Temperature outliers are $T < 5$ and $T > 37.8$ Pressure outliers are $p < 993$ and $p > 1025$ But all values of T are between 9.7 and 27.2, so there are no temperature outliers and all values of p are between 994 and 1017, so there are no pressure outliers
(b)	Negative **Ignore quantifiers e.g. 'weak, strong'**
(c)	For every 1 $^{\circ}C$ increase in the temperature (in 2015 in Beijing), the pressure decreases by 0.71
(d)	$p = -0.71(8.5) + 1022 = \underline{1015.965}$ hPa
(e)	Unreliable because the large data set only contains data for May-October, and so December is outside of the data range (used to produce the regression line) **Accept 'extrapolation' but they must make reference to the fact that the LDS only contains data for May-October**
(f)	Use a greater number/sample size of days from Beijing **No marks for reference to 'Asia'/'other places', since Beijing is the only place in Asia the LDS has data for**
(g)	*Idea that* he should consider data from other places in Asia, but Beijing is the only place in Asia that the LDS has data for

Question	Scheme
6	
	$8.45 \times 1.15 = \frac{194.35}{n}$ $\Rightarrow n = \frac{194.35}{8.45 \times 1.15} = 20$ so the size of Zain's sample is 20 **[Here it has been used that 1knot = 1.15 mph]**