

3. Ben is studying the Daily Total Rainfall, x mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the following table.

x	0	0.1–0.5	0.6–1.0	1.1–1.9	2.0–4.0	4.1–6.9	7.0–12.0	12.1–20.9	21.0–32.0	tr
Frequency	55	18	18	21	17	9	9	6	2	29

(a) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.

(2)

(a) A value needs to be assigned for tr - trace readings.
As "trace" means 0 to 0.05, 0.025 could be used for these.

Using all 184 of these values, Ben estimates $\sum x = 390$ and $\sum x^2 = 4336$

(b) Calculate estimates for

- the mean Daily Total Rainfall,
- the standard deviation of the Daily Total Rainfall.

(3)

(b)(i) $\bar{x} = \frac{\sum x}{n} = \frac{390}{184} = 2.1195... = 2.12 \text{ mm (3sf)}$

(ii) Variance, $\sigma_x^2 = \text{mean of squares} - \text{square of mean}$
 $= \frac{\sum x^2}{n} - \bar{x}^2$
 $= \frac{4336}{184} - \left(\frac{390}{184}\right)^2 = 19.07...$

Standard Deviation, $\sigma_x = \sqrt{19.07...} = 4.367...$
 $= 4.37 \text{ mm (3sf)}$

Ben suggests using the statistic calculated in part (b)(i) to estimate the annual mean Daily Total Rainfall in Leeming for 1987

(c) Using your knowledge of the large data set,

- give a reason why these data would not be suitable,
- state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

(2)

(c)(i) The large data set only contains readings from May to October, which is not representative of a whole year with changing seasons.

(ii) The missing months - November to April - are the winter when we would expect more rain.
So, the estimate in (b)(i) would underestimate the annual mean.