Exercise walk-through

2.7
Statistics – Numerical Measures of Variability

GAC Mathematics

Statistics: collecting and displaying data

- A **population** is everyone you want to describe; a **sample** is who you actually measure.
- To generalise, the **sample** must be **representative**: chosen so that every part of the **population** has a fair chance of being included.
- **Qualitative data** are categories; **quantitative data** measure amounts, either **discrete** (counted) or **continuous** (measured).
- **Bias** enters through who you ask, who answers and how the question is worded. The *Literary Digest* poll had all three problems: its lists favoured the wealthy, and the 24% who replied were a self-selected group.

$$D_1 = \frac{20}{10} = 2 \qquad D_2 = \frac{20}{20} = 1$$

Statistics: measures of variability

- The **range** is the largest value minus the smallest, so one extreme value can change it a lot.
- The **interquartile range (IQR)** is $Q_3 - Q_1$, the spread of the middle half. A **box plot** shows the minimum, quartiles, **median** and maximum.
- The **standard deviation** measures spread using squared distances from the **mean**. For a whole **population**, $\sigma = \sqrt{\sum (x-\mu)^2/n}$; for a sample estimate, $s = \sqrt{\sum (x-\bar{x})^2/(n-1)}$. It has the same units as the data.
- Two data sets can share a mean and be completely different: a city with every day at 20 °C and one swinging between 0 °C and 40 °C have the same mean temperature.

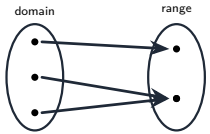
$$\text{range} = 12 - 2 = 10 \qquad IQR = 8 - 4 = 4$$

Sources for this lesson

- GAC010 Mathematics II course outline (sequences, finance, probability and statistics).
- *Labour Law of the People's Republic of China* (1994), Article 44, on overtime pay.
- National Bureau of Statistics of China, *Statistical Communiqué on the 2023 National Economic and Social Development* (per capita disposable income, **mean** and **median**).

Exercises 2.1·2.2

2.1 Find the range of 3, 5, 8 and 12. [2]

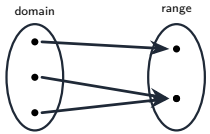


2.2 A five-number summary is minimum 4, first quartile 7, median 9, third quartile 13, maximum 20. Find its IQR and range. [2]



Exercises 2.1·2.2

2.1 Find the range of 3, 5, 8 and 12. [2]



2.1

Subtract the minimum from the maximum.

$$R = x_{max} - x_{min} = 12 - 3 = 9$$

Range 9.

2.2 A five-number summary is minimum 4, first quartile 7, median 9, third quartile 13, maximum 20. Find its IQR and range. [2]



2.2

$IQR = Q_3 - Q_1 = 13 - 7 = 6$. Range $R = 20 - 4 = 16$.

Exercise 2.3

2.3 Using the stated half-list convention, find the median, first quartile and third quartile of 2, 4, 6, 8, 10, 12, 14, 16. [3]



Exercise 2.3

2.3 Using the stated half-list convention, find the median, first quartile and third quartile of 2, 4, 6, 8, 10, 12, 14, 16. [3]



2.3

Median is the mean of the fourth and fifth values: $m = (8 + 10)/2 = 9$. Lower-half median $Q_1 = (4 + 6)/2 = 5$. Upper-half median $Q_3 = (12 + 14)/2 = 13$.

Exercise 2.4

2.4 The complete population is 2, 4, 6. Find its mean, population variance and population standard deviation, giving the latter to 3 significant figures. [3]



Exercise 2.4

2.4 The complete population is 2, 4, 6. Find its mean, population variance and population standard deviation, giving the latter to 3 significant figures. [3]



2.4

Mean $\mu = (2 + 4 + 6)/3 = 4$.

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n} = \frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3} = \frac{8}{3}$$

Population variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{8/3} \approx 1.63$$

Population standard deviation.

Exercise 3.1

3.1 Two complete populations are A: 4, 5, 6 and B: 1, 5, 9. Find each mean and population standard deviation, to 3 significant figures. Compare their spread without claiming a cause. [6]



Exercise 3.1

3.1 Two complete populations are A: 4, 5, 6 and B: 1, 5, 9. Find each mean and population standard deviation, to 3 significant figures. Compare their spread without claiming a cause. [6]



3.1

Both means are 5. Use population division by n, not n minus 1.

$$\sigma_A = \sqrt{\frac{(4-5)^2 + (5-5)^2 + (6-5)^2}{3}} = \sqrt{2/3} \approx 0.816$$

A standard deviation.

$$\sigma_B = \sqrt{\frac{(1-5)^2 + (5-5)^2 + (9-5)^2}{3}} = \sqrt{32/3} \approx 3.27$$

B standard deviation. B is more spread around the shared mean. The summaries do not identify why the populations differ.

Exercise 3.2

3.2 The ordered data are 2, 4, 6, 8, 10, 12, 14, 16. (a) Draw a minimum-to-maximum box plot using the stated quartile convention. (b) Find the range and IQR. (c) Replace 16 by 100 and find the new range and IQR. Explain why these two spread measures respond differently and whether the plot alone can establish the cause of the extreme value. [8]



Exercise 3.2

3.2 The ordered data are 2, 4, 6, 8, 10, 12, 14, 16. (a) Draw a minimum-to-maximum box plot using the stated quartile convention. (b) Find the range and IQR. (c) Replace 16 by 100 and find the new range and IQR. Explain why these two spread measures respond differently and whether the plot alone can establish the cause of the extreme value. [8]



3.2

(a) Five-number summary is 2, 5, 9, 13, 16. Correct quartiles and median earn B1; a box from 5 to 13 with median 9 earns B1; whiskers to 2 and 16 on a labelled scale earn B1.

(b) Range $R = 16 - 2 = 14$. IQR $= 13 - 5 = 8$.

(c) New range $R' = 100 - 2 = 98$. Quartiles stay 5 and 13, so IQR remains 8. The range uses the changed extreme directly, while the middle-half endpoints stay fixed; neither the plot nor these summaries establishes the cause of 100. Do not delete that value without checking its origin.