Data

KEY TERMS

Learn these words first. EQAO questions often test whether you understand what the math language means.

Population
The entire group you want information about.
Ex: all Grade 9 students

Sample
A smaller group selected from the population.
Ex: 100 students surveyed

Census
Data collected from every member of a population.
Ex: asking every student

Primary Data
Data you collect yourself for a specific purpose.
Ex: your own survey

Secondary Data
Data collected by someone else that you use.
Ex: Statistics Canada table

Variable
A characteristic that can take different values.
Ex: height, score, age

One-variable Data
Data involving one measured characteristic.
Ex: test scores only

Two-variable Data
Paired data involving two characteristics for each case.
Ex: hours studied + score

Mean
The arithmetic average: total divided by number of values.
Ex: (4+6+8)/3 = 6

Median
The middle value after the data are ordered.
Ex: 2, 4, 9 -> median 4

Quartile
A value that divides ordered data into four sections.
Ex: Q1, Q2, Q3

Interquartile Range (IQR)
The spread of the middle 50% of the data: Q3 - Q1.
Ex: 18 - 10 = 8

Box Plot
A graph showing minimum, Q1, median, Q3 and maximum.
Ex: five-number summary

Scatter Plot
A graph of paired numerical data shown as points.
Ex: height vs. arm span

Correlation
The direction and strength of a relationship between two variables.
Ex: positive, negative, none

Line of Best Fit
A line that models the overall trend of a scatter plot.
Ex: used for prediction

Outlier
A value far from the overall pattern of the data.
Ex: one unusually high point

Extrapolation
Using a model to predict beyond the observed data range.
Ex: riskier prediction

KEY CONCEPTS & DEFINITIONS

Representative Samples
A useful sample should resemble the population. Biased sampling can produce misleading conclusions.

Ask: who was included and who was left out?

Centre vs. Spread
Centre describes a typical value; spread describes how varied the data are. Both matter when comparing groups.

Median + IQR is a common pair.

Correlation Is Not Causation
Two variables moving together does not prove that one causes the other. A third factor may be involved.

Relationship does not automatically mean cause.

Models Are Approximations
A line of best fit summarizes a trend, but real points do not have to lie exactly on the line.

Predictions are estimates, not guarantees.

Data

FORMULAS & RULES

MEAN

mean = sum of values / number of values

Add every value, then divide by how many values there are.

INTERQUARTILE RANGE

IQR = Q3 - Q1

Measures the spread of the middle 50%.

RANGE

range = maximum - minimum

A quick measure of total spread.

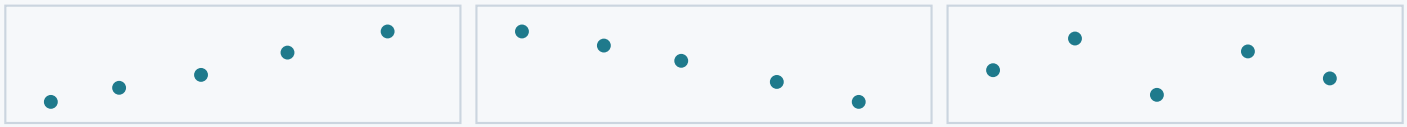
MODEL PREDICTION

use y = mx + b after fitting a line

Substitute an x-value to estimate the matching y-value.

VISUALS & FAST FACTS

VISUAL MEMORY AID: CORRELATION



Positive | Negative | None

FAST FACTS: Order data before finding median or quartiles. | An outlier can strongly affect the mean. | Extrapolation is less reliable than interpolation.

THREE WORKED EXAMPLES

1

Mean and Median

Find the mean and median of 3, 5, 7, 7, 13.

- Mean = (3+5+7+7+13) / 5 = 35/5 = 7.
- The ordered middle value is 7.
- Both measures of centre are 7.

ANSWER

mean = 7; median = 7

2

IQR

For Q1 = 12 and Q3 = 21, find the IQR.

- Use IQR = Q3 - Q1.
- IQR = 21 - 12.
- IQR = 9.

ANSWER

9

3

Use a Trend

A model is y = 4x + 52. Predict y when x = 6.

- Substitute x = 6.
- y = 4(6) + 52.
- y = 24 + 52 = 76.

ANSWER

76