

Data Management

Ontario MTH1W · types of data, graphs, mean, median and mode, quartiles and box plots, scatter plots, lines of best fit and modelling

vertexmath.ca
Worksheet, answers &
this cheat sheet



MOST TESTED Find the five-number summary, the IQR and any outliers of 4, 6, 8, 9, 10, 12, 14, 16, 31. (card 5)

Which average?

Symmetric data, no outliers

Mean

Outliers or skewed data

Median

Categories, or the most common value

Mode

1 Types of data

USE Primary: you collect it; **secondary:** someone else did. A **sample** is part of the population.

labels or words **categorical**

counted: 0, 1, 2, ... **discrete**

measured: any value **continuous**

TRAP Time rounded to whole minutes is still **continuous**: it is measured.

2 Graphs

$$\text{sector angle} = \frac{\text{part}}{\text{total}} \times 360^\circ$$

Which graph?

compare categories **bar graph**

parts of one whole **circle graph**

continuous data in intervals **histogram (bars touch)**

two numerical variables **scatter plot**

EXAMPLE 9 of 24 students:
 $9/24 \times 360^\circ = 135^\circ$.

TRAP A 15% sector is
 $0.15 \times 360^\circ = 54^\circ$, not 15° .

3 Mean, median, mode

$$\text{mean} = \frac{\text{sum of the values}}{\text{number of values}}$$

$$\text{range} = \text{max} - \text{min}$$

USE Median: order first. Odd count: the middle value; even: the mean of the two middle values. **Mode:** the most frequent value.

EXAMPLE 3, 8, 5, 8, 6 ordered:
3, 5, 6, 8, 8. Mean 6, median 6, mode 8, range 5.

TRAP Values 0, 1, 2 with frequencies 3, 5, 2: mean = $(0 + 5 + 4)/10 = 0.9$, not $(0 + 1 + 2)/3 = 1$.

4 Using the mean

$$\text{total} = \text{mean} \times \text{count}$$

USE An outlier pulls the mean; the median barely moves. Combine groups by adding totals.

EXAMPLE Wages \$15, \$16, \$16, \$18, \$60 an hour: mean \$25, median \$16.

TRAP 24 students average 65 and 16 average 80: $(24 \cdot 65 + 16 \cdot 80)/40 = 71$, not 72.5.

5 Quartiles and IQR

$$\text{IQR} = Q_3 - Q_1$$

$$\text{outlier} < Q_1 - 1.5 \times \text{IQR}$$

$$\text{outlier} > Q_3 + 1.5 \times \text{IQR}$$

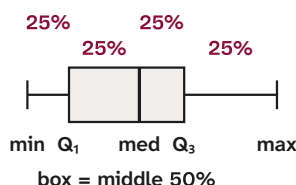
USE Order first. Q_1, Q_3 : medians of the lower and upper halves. Odd count: the median is in **neither** half.

EXAMPLE 4, 6, 8, 9, 10, 12, 14, 16, 31:
 $Q_1 = 7$, median 10, $Q_3 = 15$, IQR = 8;
 $15 + 12 = 27$, so 31 is an outlier.

TRAP 1, 3, 5, 7, 9: $Q_1 = 2$ (median of 1, 3), not 3 (median of 1, 3, 5).

6 Box plots

USE Draw from the five-number summary. Compare medians for the centre, IQRs for the spread.



TRAP A long whisker, 30 to 55, still holds only about a quarter of the data.

7 Correlation

USE Strength: how closely the points lie to a line.

points rise left to right **positive**

points fall left to right **negative**

no clear pattern **no correlation**

TRAP Steep is not strong: points exactly on $y = 0.2x + 5$ are a **perfect** correlation.

8 Line of best fit

$$m = \frac{\text{rise}}{\text{run}} = \frac{y_2 - y_1}{x_2 - x_1}$$

$$y = mx + b$$

USE Follow the trend. Use two points **on the line**; give m and b with units.

EXAMPLE Through (2, 30) and (8, 12):
 $m = -18/6 = -3$; $30 = -6 + b$, so
 $y = -3x + 36$.

TRAP (2, 4) to (8, 7): $m = 3/6 = 0.5$, not $6/3 = 2$.

TRAP Given $y = 21$: solve
 $21 = -3x + 36$, so $x = 5$; not
 $-3(21) + 36$.

9 Predictions

USE Correlation is not cause: a third variable can drive both. Critique: axis not from 0? Biased sample?

x inside the data's range **interpolation**

x below or above it **extrapolation (risky)**

TRAP Axis from 50: bars 52 and 56 look 1 to 3; 56 is only about 8% more.