Key focus

This objective is the symbolic follow-on to analysing scatter plots (TARSOG): once a linear trend and a line of best fit have been identified visually, students form the line's equation ($y = mx+c$) and use it to predict values. The equation may be given directly, built from a stated gradient and y-intercept, or found from two points on the line. The second half of the objective is about judging predictions: distinguishing interpolation (within the data range) from extrapolation (beyond it), and discussing how correlation strength and sample size affect confidence in a prediction.

Suggested teaching sequence

1. Revisit the line of best fit from TARSOG scatter plots, then introduce its equation form $y = mx + c$: m is the gradient (rate of change of y per unit of x), c is the y-intercept (the value of y when $x = 0$).
2. Practise substituting a known x into a given equation to predict y, and reading off the gradient and y-intercept directly from an equation.
3. Model finding the gradient from two points on the line: $m = \dfrac{y_2 - y_1}{x_2 - x_1}$, then finding c by substituting one point back into $y = mx + c$.
4. Apply the method to real contexts (study time vs score, advertising spend vs sales, age vs reaction time) so predictions have concrete meaning.
5. Introduce interpolation (predicting within the range of x-values used to build the model) vs extrapolation (predicting outside that range), and why extrapolation is generally less reliable – the linear trend is only known to hold where there was data.
6. Discuss what affects confidence in a prediction: how closely points cluster around the line (correlation strength), the number of data points used, and how far a prediction sits from the known data range.
7. Practise working backwards: given a target y-value, solve the equation for x.
8. Discuss (without requiring calculation) that a line of best fit is a model, not a guarantee – real data scatters around it, so predictions are estimates.

Core examples to model

- Direct substitution: $y = 6x + 40$ (hours studied vs score); predict y at $x = 7$: $y = 6(7) + 40 = 82$.
- Gradient from two points: line through $(2, 14)$ and $(6, 26)$: $m = \dfrac{26 - 14}{6 - 2} = \dfrac{12}{4} = 3$. With y-intercept 8: $y = 3x + 8$.
- Interpolation vs extrapolation: data collected for $x = 0$ to 10; predicting at $x = 7$ is interpolation (reliable); predicting at $x = 25$ is extrapolation (less reliable, no evidence the trend continues).
- Working backwards: $y = -2x + 300$; find x when $y = 200$: $200 = -2x + 300 \Rightarrow x = 50$.

- Confidence: a strong correlation (points close to the line) supports more confident predictions than a weak correlation (points widely scattered), and a line built from many data points is more stable than one from very few.

Watch for

- Substituting into the wrong side of the equation, or mixing up which variable is x and which is y in a word context.
- Treating an extrapolated prediction as just as reliable as an interpolated one – the model is only supported by data within the range it was built from.
- Forgetting to state which point was used to find c after calculating the gradient from two points (either point works, but one must be substituted back in).
- Assuming a prediction from the line of best fit will be exactly correct – real data varies around the line; the prediction is an estimate.
- Confusing correlation strength (how closely points sit around the line) with the size of the gradient – a steep line can still have a weak correlation, and a shallow line can have a strong one.
- Ignoring context when interpreting the y-intercept – e.g. a negative or zero x may not be a meaningful real-world value.

Checks for understanding

- Can students substitute a value into $y = mx + c$ to predict y, and rearrange to find x from a given y?
- Can they state the gradient and y-intercept from an equation, or write an equation given the gradient and y-intercept?
- Can they find the gradient and equation of a line of best fit from two points on it?
- Can they distinguish interpolation from extrapolation, and explain why extrapolation is generally less reliable?
- Can they explain how correlation strength and sample size affect confidence in a prediction?
- Can they explain that a line of best fit is a model whose predictions are estimates, not guaranteed exact values?