Residuals, Checking the Fit

Residual = actual minus predicted: the vertical gap from a point to the line. Residual plot: random scatter means a line fits; a pattern means curved data. Use the scatter plot, r, and the residual plot together.

1 THE VOCABULARY
Words we use

Residual actual – predicted, for one point.
Positive residual point above the line; the model guessed too low.
Negative residual point below the line; the model guessed too high.
Residual plot the residuals against x, with a flat line at 0.
Random scatter, pattern a line fits; or the data is curved.

- Watch out for these.
- Subtracting predicted minus actual. It is actual minus predicted.
- Trusting a big r by itself.
- Reading a U-shaped residual plot as fine because the residuals are small.
- Mixing up above and below: positive is above.

2 THE METHOD
Check a fit

1. PREDICT put each x into the model.
2. SUBTRACT actual – predicted, for every point.
3. SIGNS positive: above, too low. Negative: below, too high.
4. PLOT residuals against x, with a line at 0.
5. DECIDE random scatter: keep the line. Pattern: try another family.

WHY A PATTERN MEANS THE WRONG FAMILY
If the data really followed a line, the misses would be random. A U or a hill is the curve of the data showing through.

3 TUITION
R(t) = 326t + 6440

t = 4: predicted 326(4) + 6440 = 7,744	put t in
actual 7,874	from the data
residual 7,874 − 7,744 = +130	above the line; the model guessed too low
t = 8: 8,928 − 9,048 = −120	below the line; too high
all eleven: no pattern	a line is a good model

THE GNATS
A line gives r of about 0.98, but residuals +5.8, −3.8, −6.4, +1.0, +3.4 make a U. The data curves: a line is the wrong family.

SAY IT OUT LOUD
Actual minus predicted. Random scatter, keep the line. A pattern, try another family.