LEARN

Deborah owns a smoothie cart, and she tracked daily high temperatures alongside how many smoothies she sold. When she plotted the data, the points formed a clear upward trend: Hotter days meant more sales. To turn that trend into something she could actually use for planning, she needed a single equation that best fit all those data points. That equation is a **least squares regression line**, and it lets her plug in a temperature and get a predicted sales number.

The chart below shows the daily high temperatures and smoothie sales Deborah recorded over several weeks; this is the actual data her regression equation was built from.

No alt text for image below. CPE/PM

The smoothie cart's least squares regression model is:

$y = 5.2x - 140$

As a reminder: x represents the input value (the daily high temperature, in °F), and y represents the predicted output value (the number of smoothies sold).

This equation has two key parts: the **constant term** and the **rate of change**. Expand each section below to examine the algebra.

The Constant Term (y-intercept $= -140$) +

The constant term is the value of y when $x = 0$.

Substitute $x = 0$:

$y = 5.2(0) - 140$

$y = 0 - 140$

$y = -140$

Selling **−140 smoothies** is physically impossible. This does not represent a real prediction at 0°F. The constant term anchors the regression line at the correct position on the graph; it is a mathematical artifact, not a meaningful forecast. Always check whether the y-intercept makes sense in context before interpreting it literally.

The Rate of Change (slope $= 5.2$) +

The slope tells you how much y changes for every 1-unit increase in x.

Verify at two consecutive temperatures:

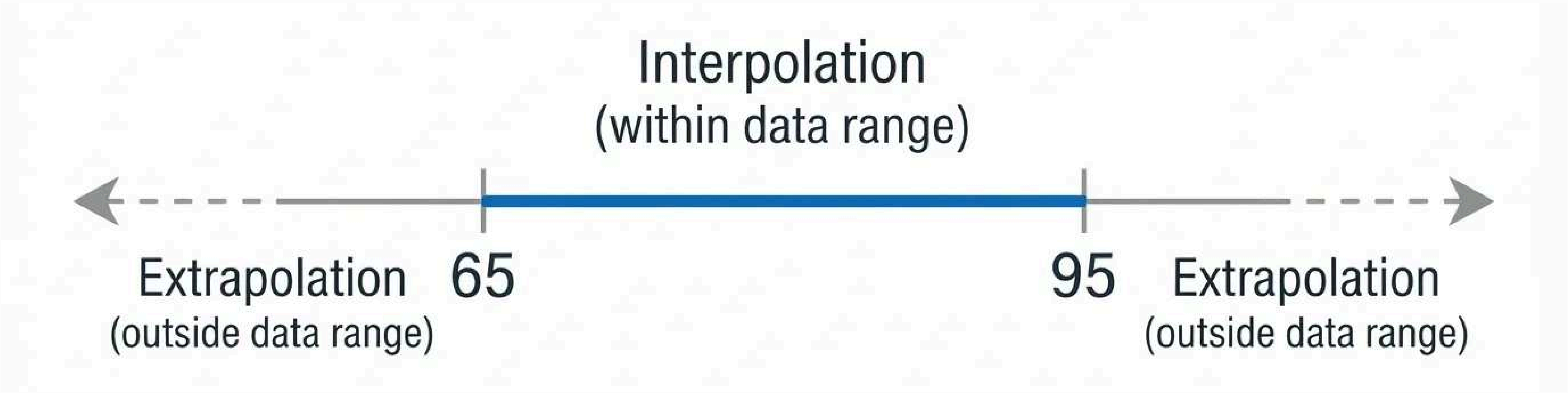
$y(80) = 5.2(80) - 140 = 416 - 140 = 276$

$y(81) = 5.2(81) - 140 = 421.2 - 140 = 281.2$

Difference $= 281.2 - 276 = 5.2$ ✓

Interpretation: For every additional 1°F in temperature, the model predicts **5.2 more smoothies** sold. Always state slope with units (in this case, *smoothies per degree Fahrenheit*).

Not every prediction from a regression model is equally trustworthy. Whether an x-value falls inside or outside the range of the original data changes how much you can trust the result. Let's look at both cases: interpolation and extrapolation.



Interpolation is a prediction made using an x-value that falls *within* the range of the original data (the smoothie cart's data spans $65^\circ F$–$95^\circ F$). Because the model was built using data in this range, interpolated predictions are generally reliable.

Worked example: $x = 85^\circ F$

$y = 5.2(85) - 140$

$y = 442 - 140$

$y = 302$ smoothies

85°F falls inside the data range, so this is an interpolation.

Try it yourself. Use the model $y = 5.2x - 140$ to predict smoothie sales when $x = 90^\circ F$. Explain what y is, and show your substitution steps.

Type answer

Extrapolation is making a prediction using an x-value that falls *outside* the range of the original data. The algebra still works, but the real-world meaning may not.

Worked example: $x = 112^\circ F$

$y = 5.2(112) - 140$

$y = 582.4 - 140$

$y = 442.4$ smoothies

The original data was collected over a normal range: roughly $65^\circ F$ to $95^\circ F$. At $112^\circ F$, parks issue heat advisories, people stay indoors, and a linear trend no longer holds. **The model keeps going. Reality does not.**

Try it yourself. A forecaster predicts a $105^\circ F$ day. Use the model $y = 5.2x - 140$ to predict smoothie sales at $x = 105^\circ F$, and then state whether this is an interpolation or an extrapolation and why.

Type answer

Interpolation vs. Extrapolation

Interpolation: prediction within the data range → Generally reliable

Extrapolation: prediction outside the data range → Use with caution

The further outside the data you predict, the less you can trust the model.

THINK

Use the regression model $y = 5.2x - 140$ to predict smoothie sales when $x = 72^\circ F$. Show the substitution steps and final answer.

Type answer

A heat wave is forecast to reach 112°F. The regression model gives $y = 5.2(112) - 140 = 442.4$ smoothies. Which statement **best** explains why this prediction is unreliable?

Select one

1. The equation contains a calculation error: 5.2×112 does not equal 582.4.

   The arithmetic is actually correct: $5.2 \times 112 = 582.4$, and $582.4 - 140 = 442.4$. The problem isn't the calculation; it's what the number means.

2. 112°F is outside the original data range, making this an extrapolation. The model has no data to support a prediction at that temperature.

   Extrapolation extends a model beyond the data it was built on. At 112°F, real-world behavior (heat advisories, park closures, changes in foot traffic) may make the linear trend completely invalid.

3. The slope of 5.2 is too steep for high temperatures, so it automatically produces inflated predictions above 100°F.

   The slope doesn't change based on temperature range; it's a fixed value in the model. The real issue is that the model was never tested at 112°F, not that the slope behaves differently there.

4. Regression models can only be trusted within 10°F of the average temperature in the data set.

   There is no fixed-degree rule like this. The concern is whether the prediction falls outside the data range entirely, which is what extrapolation describes.

Tracking Plant Growth

A researcher measures a sunflower seedling's height once a week. She collects data from **week 0 to week 25** and fits a regression model:

$y = 3.5x + 4$,

where x is the number of weeks since planting, and y is the predicted height in centimeters.

Which statement correctly interprets **both** the slope and the y-intercept of $y = 3.5x + 4$ in this context?

Select one

1. For every additional week, the plant's predicted height increases by 3.5 cm; at week 0, the model predicts a starting height of 4 cm.

   The slope (3.5) is the weekly growth rate, and the y-intercept (4) is the predicted height at week 0.

2. For every additional week, the plant's predicted height increases by 4 cm; at week 0, the model predicts a starting height of 3.5 cm.

   The values are swapped: 3.5 is the rate of change (per week), and 4 is the starting height at week 0, not the other way around.

3. The plant grows to a maximum height of 3.5 cm; 4 is the number of weeks it takes to reach that height.

   This describes neither value correctly. 3.5 is a weekly growth rate, not a maximum height, and 4 is a starting height, not the number of weeks.

4. The plant's total height after 25 weeks is 3.5 cm, and 4 is an unrelated rounding constant.

   Neither value works this way. 3.5 is the slope (rate of change per week), and 4 is the y-intercept (starting height at week 0).

Predict a future value. Use the model to predict the seedling's height at **week 20**. Show the substitution and calculate the result.

Type answer

Discuss extrapolation limits. The researcher only collected data from week 0 to week 25. If someone used this model to predict the plant's height at **week 200**, the equation gives $y = 3.5(200) + 4 = 704$ cm (about 23 feet tall). Is this a reasonable prediction? Explain why or why not, using the idea of extrapolation.

Type answer