Summary – Relations and Models

Two linked quantities make a relation. We describe it, then model it so we can predict.

Relations and variables

A **relation** is a set of ordered pairs that links two quantities. In the salary data, each pair is (years of education, average salary), such as (9, 16 322).

The **independent variable** is the one that is controlled or, when nothing is controlled, the one we predict *from*; it comes first in each ordered pair and goes on the horizontal axis. The **dependent variable** depends on it and goes on the vertical axis. Nobody controlled years of education; we chose it as independent because we predict salary from it. The same choice set the axes for thumb and palm.

Four Ways to Describe a Relation: Education and Salary

A relation can be described in words, as a table of values, as an equation, and as a graph.

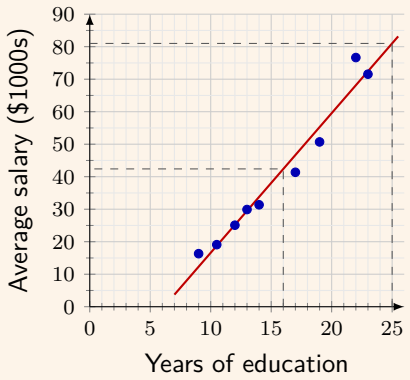
As a table of values.

Years of education	Average salary ($1000s)
9.0	16.3
10.5	19.1
12.0	25.1
13.0	29.9
14.0	31.4
17.0	41.4
19.0	50.7
22.0	76.7
23.0	71.5

As a graph. A **scatter plot**: one point per pair.

In words. Describe the plot in three ways. **Direction: positive** if one quantity increases as the other increases; **negative** if one decreases as the other increases. **Strength: strong**, **moderate**, or **weak**, by how closely the points follow a pattern. **Shape: linear** if the points follow a line; **non-linear** if they follow a curve. Salary against years of education is positive, strong, and roughly linear: about $4000 more for each extra year.

A **line of best fit** follows the trend, with about as many points above it as below. When the points bend, we draw a **curve of best fit** instead.



Data: Augusta State University, 2002.

As an equation. Not yet: the salaries are measurements, so no rule gives them exactly. A line of best fit is a model, and a model can have an equation, but the equations in this thread come later.

Interpolate and extrapolate

For 16 years of education, my line of best fit gives about $42 000. Sixteen years lies inside the data (9 to 23 years), so this is an **interpolation**: a prediction between known values. For 25 years it gives about $81 000. Twenty-five years lies beyond the data, so this is an **extrapolation**: a prediction outside the known values. I trust the interpolation more: an extrapolation assumes the trend carries on beyond the data.

When a model is not justified

A sportswriter claimed that Roy Halladay pitched poorly in hot weather. His Game Scores plotted against the temperature form a wide cloud with no clear trend: hot days and cool days both have high and low scores. A line through that cloud would be nearly flat and predict almost nothing, so the data do not support the claim. A model is only as good as the relation behind it, so check for a trend before drawing a line or curve.

The question for this thread

The palm data bend, so a line will not fit them. To predict from a curve, we need an equation that describes it, the fourth way to describe a relation. The rest of this thread builds one: the quadratic relation.