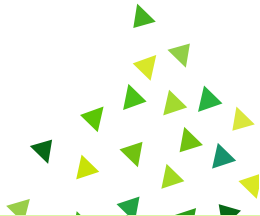


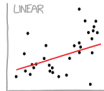


Saturday Morning Statistics #28

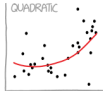
June 11th, 2022



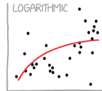
CURVE-FITTING METHODS AND THE MESSAGES THEY SEND



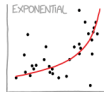
"HEY, I DID A REGRESSION!"



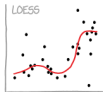
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH!"



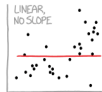
"LOOK, IT'S TAPERING OFF!"



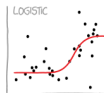
"LOOK, IT'S GROWING UNCONTROLLABLY!"



"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE!"



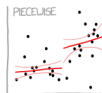
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO!"



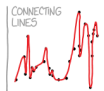
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH!"



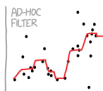
"LISTEN, SCIENCE IS HARD, BUT I'M A SERIOUS PERSON DOING MY BEST."



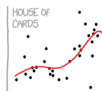
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND!"



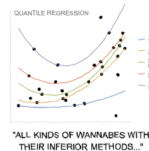
"I CLICKED 'SMOOTH LINES' IN EXCEL."



"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"



"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE- WAIT NO NO DON'T EXTEND IT AAAAAA!"

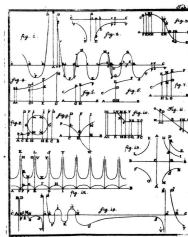
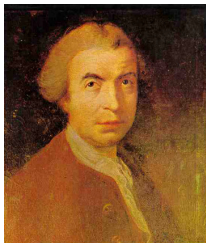


Source: XKCD 2048 as amended by Anton Antonov for a 2019 talk at an R-user meeting in Boston.
<http://www.econ.uiuc.edu/~roger/research/rq/rq.html>

A Brief History in Statistics: Least Absolute Deviation

Least Absolute Deviation (LAD) Model

- A special case of quantile regression where $q=0.5$.
- Proposed in 1760 by Roger Joseph Boscovich (1711 - 1787).
 - ▶ Preceded the least squares method (1805) by fifty years.



Quantile Regression

A standard regression estimates the mean of the conditional distribution, for example

$$\hat{y} = \beta_0 + \beta_x \hat{x}.$$

A **quantile regression** is a method for estimating conditional quantiles, such as the **median**. The conditional τ th quantile is assumed to be a linear function of the explanatory variables:

$$Q_{Y|X}(\tau) = X\beta_\tau.$$

Question and Hypothesis

Questions of the day.

- Is the relationship between β -**caryophyllene** and **humulene** constant across the conditional distribution?
- Is the relationship between β -**pinene** and **D-limonene** constant across the conditional distribution?



Thank you for coming.

Insights of the Day

- Uncertainty is unavoidable.
- There are many useful (conditional) statistics within our grasp.