

Computing Relationships

Linear Regression is a way of computing the **line of best fit**, which minimizes the sum of vertical distances of all scatter plot points from the line. Calculating the slope and intercept of this line is a task best left to computing or statistical software.

- **Slope** provides us with the easiest summary to grasp: it's how much we predict the y-variable (response variable) will increase or decrease for each unit that the x-variable (explanatory variable) increases.
- **Correlation is not causation!** Correlation only suggests that two column variables are related, but does not tell us if one causes the other. For example, hot days are correlated with people running their air conditioners, but air conditioners do not cause hot days!
- **Sample size matters!** The number of data values is also relevant. We'd be more convinced of a positive relationship in general between cat age and time to adoption if a correlation of +0.57 were based on 50 cats instead of 5.