Linear models

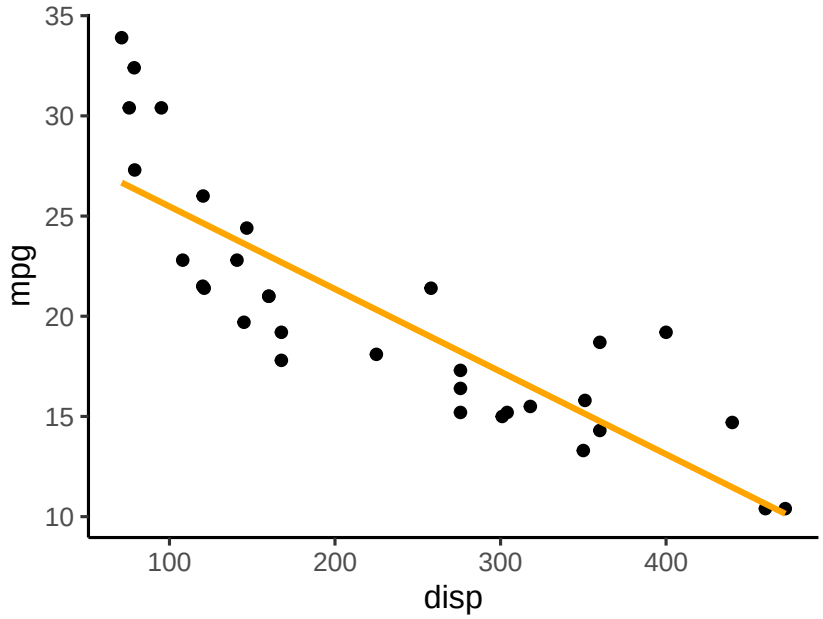
How do they work?

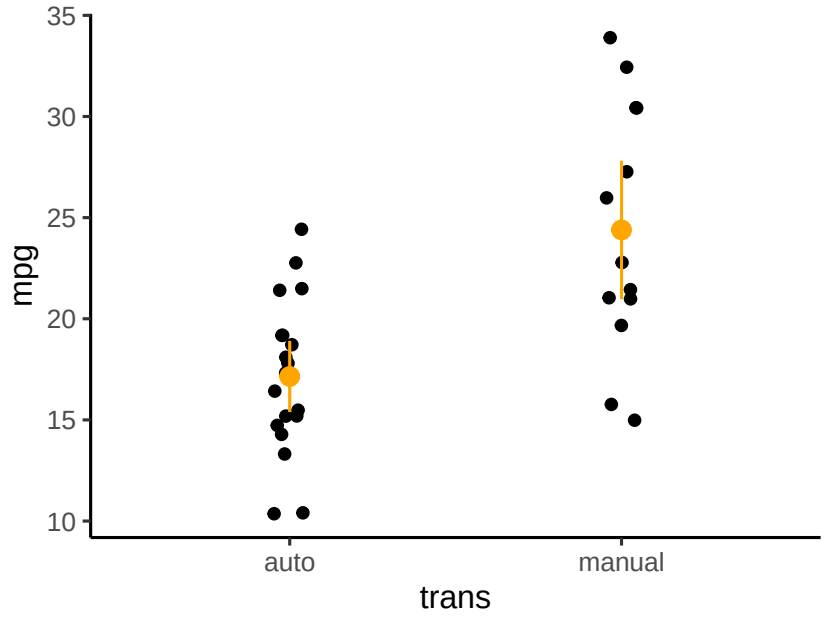
Samuel Robinson, Ph.D.

October 1, 2020

Motivation

- *I have some bivariate data (2 things measured per row), and I want to know if they're related to each other*
- *I have 2+ groups of data, and I want to know whether the means are different*





Model terminology

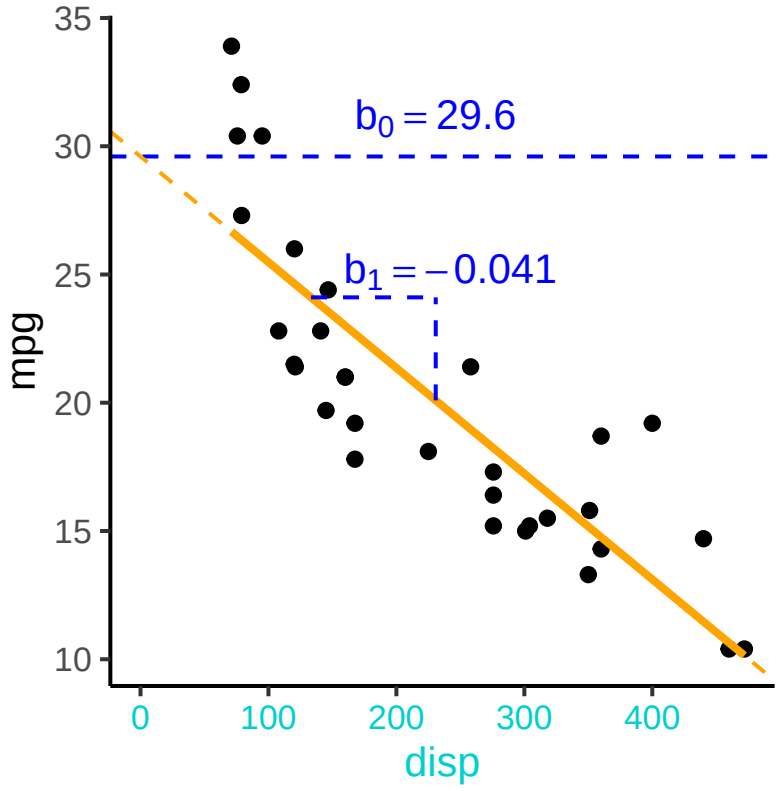
- All linear models take the form:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

- y is the thing you're interested in predicting
- $\hat{y}$ is the *predicted value* of y
- $x_1 ... x_i$ are *predictors* of y
- $b_1 ... b_i$ are *coefficients* for each predictor x_i
- b_0 is the *intercept*, a coefficient that doesn't depend on predictors
- $y \sim Normal(\hat{y}, \sigma)$ means:
  - "y follows a Normal distribution with mean $\hat{y}$ and SD σ"

This may look terrifying, but let's use a simple example:

Example



$$\hat{mpg} = b_0 + b_1 disp$$

$$mpg \sim Normal(\hat{mpg}, \sigma)$$

- *mpg* is the thing you're interested in predicting
- $\hat{mpg}$ is the *predicted value* of *mpg*
- *disp* is the *predictor* of *mpg*
- b_0 is the *intercept*, b_1 is the *coefficient* for *disp*
- $mpg \sim Normal(\hat{mpg}, \sigma)$ means:
  - "*mpg* follows a Normal distribution with mean $\hat{mpg}$ and SD σ"
- σ isn't displayed on the figure. Where is it?

How do I get R to fit this model?

`lm` is one of the main functions used for linear modeling:

A challenger approaches!

- Simulate your own data with *3* levels, rather than 2
- Use `lm` to fit a model to the data you just simulated

Why do we call them "linear models"?

To answer this, we need a brief review of matrix algebra

- This is a matrix:

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

- This is a vector

$$b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

- Multiplying them requires a *transposition* (flipping along main diagonal, denoted by $^\top$)

$$A \times b = Ab^\top = 1 \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \times \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 3 \times \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 30 \\ 36 \\ 42 \end{bmatrix}$$

Test slide

Stuff here

More stuff here