Regression I

EDS 232
Mateo Robbins
Winter 2023

Linear Regression

Assumes the data is produced by a line:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- β_0 is the intercept
- β_1 is the slope

Estimation

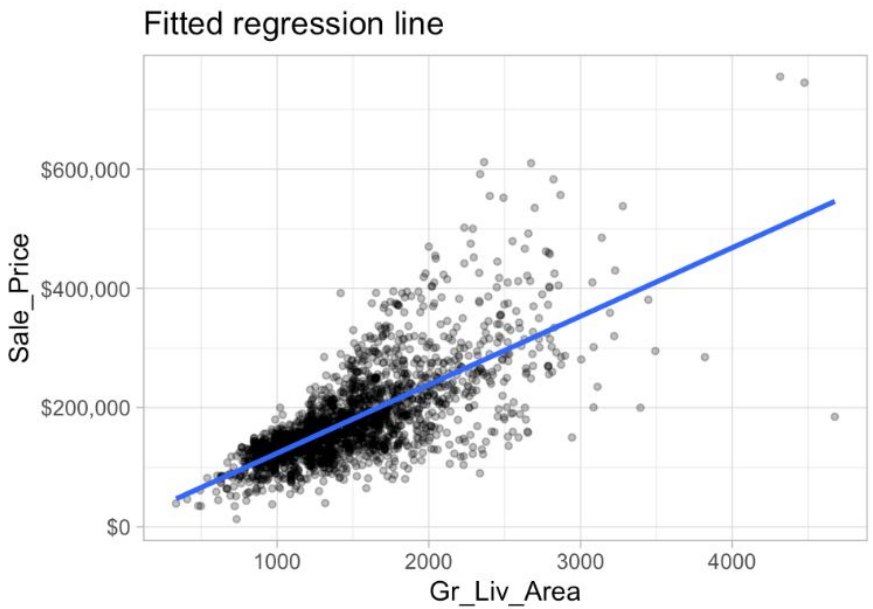
We want estimates of the coefficients that give us the “best fitting” line.

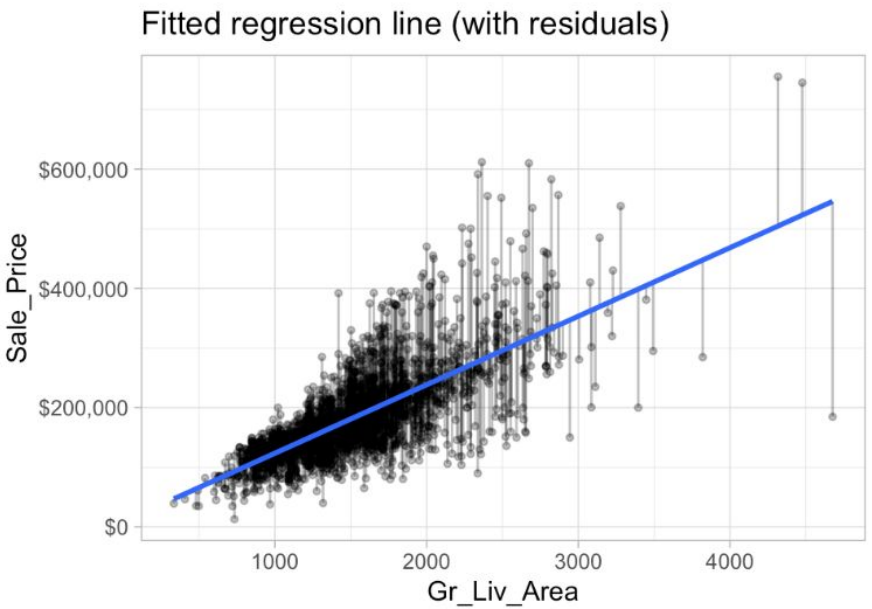
$$RSS\left(\beta_0, \beta_1\right) = \sum_{i=1}^{n} \left[Y_i - \left(\beta_0 + \beta_1 X_i\right)\right]^2 = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_i\right)^2$$

Estimation

We want estimates of the coefficients that give us the “best fitting” line.

- Most common approach uses method of least squares (LS) estimation
- This type of regression is “ordinary least squares” (OLS)regression


Fitted regression line
$600,000
$400,000
$200,000
$0
Sale_Price
1000
2000
3000
4000
Gr_Liv_Area


Fitted regression line (with residuals)
$600,000
$400,000
$200,000
$0
Sale_Price
1000
2000
3000
4000
Gr_Liv_Area

A general note: Caret vs. TidyModels

The caret package:

- one-stop solution for machine learning in R

- used by the authors of our textbook

- well-established (more resources available) and widely used

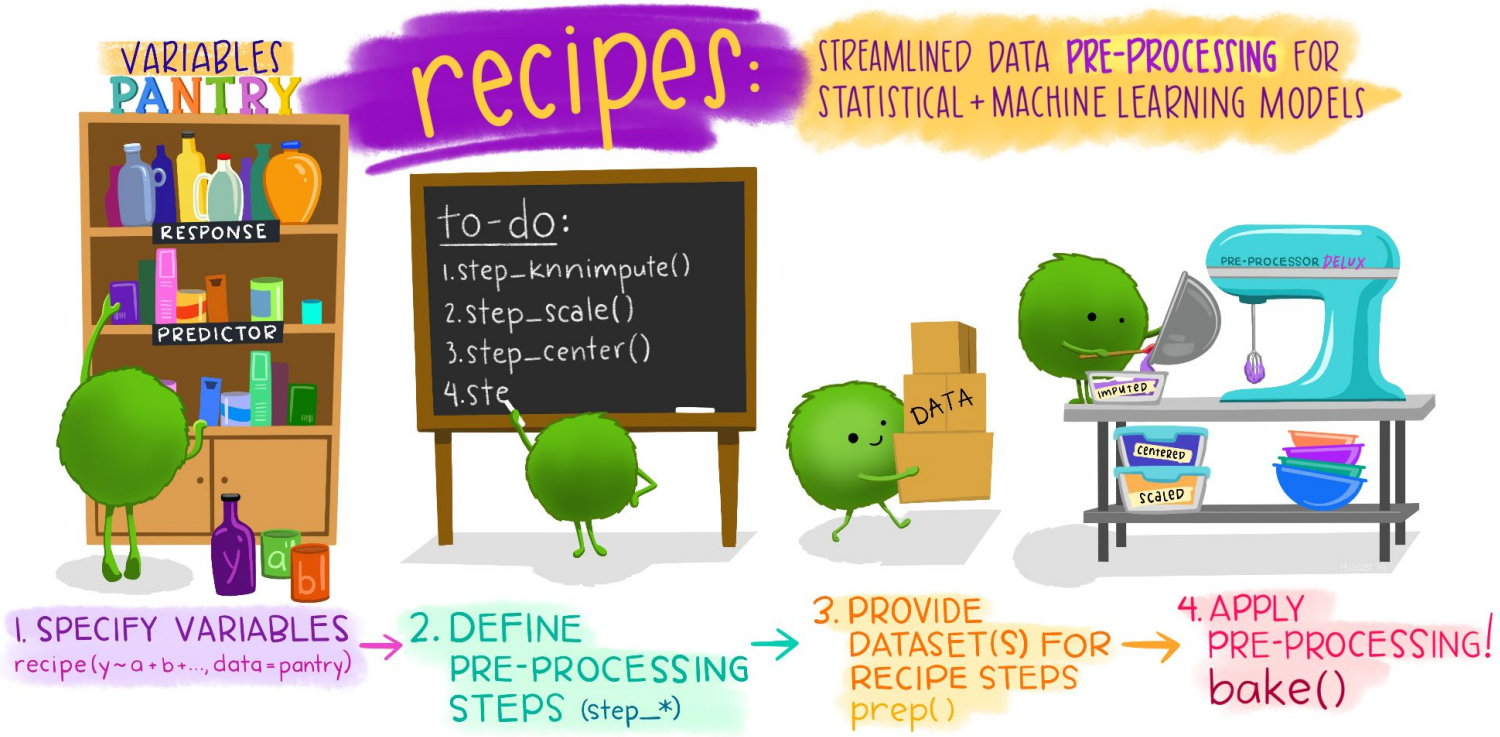
Tidymodels meta-package:

- More modern option, based on Tidyverse principles

- New and growing

- A collection of ML-based packages. With in Tidy models, parsnip is the equivalent of caret (get it?)

What's cooking?



No quiz next week

We are jumping in today...