What is regression? The goal of a regression analysis is to create a model which predicts the value of one variable from other variables.

Types of Regression

- Linear Regression (4.3): predict the outcome/dependent variable y from a single predictor/independent variable x as a linear relationship. Assumes $x, y \in \mathbb{R}$.

$$y = ax + b + e$$

- MultiLinear Regression (4.4): predict the outcome/dependent variable y from multiple predictors/independent variables $x_1, x_2, \ldots x_n$ as a linear relationship. Assumes $x_1, x_2, \ldots x_n, y \in \mathbb{R}$.

$$y = a_1x_1 + a_2x_2 + \ldots a_nx_n + b + e$$

- Models with Interactions (4.5)

$$y = a_1x_1 + a_2x_2 + a_3x_1x_2 + b + e$$

- Variations on the above: linear and non-linear transformations (5.5) effectively apply functions from reals to reals ($f : \mathbb{R} \to \mathbb{R}$) to one or more of the variables and then conduct the regression analysis

- What if x_i, y do not range over $\mathbb{R}$?
  - if $x_1, x_2, \ldots x_n \in \{0, 1\}$ then they are categorical predictors (4.3.6)
  - if $y \in \{0, 1\}$ then the outcome is categorical (Logistic Regression, Chapter 6)
  - if $y \in \mathbb{N}$ then the outcome is another kind of categorical variable (Poisson Regression)

So what? Many of the linguistic studies end up having a categorical or ordinal variable. One example would be modeling acceptability and/or grammaticality as a binary category {1,0}. For people who think acceptability should be modelled as probabilities, that would be [0,1] or maybe (0,1). In this case, we should transform it using "logit" and "probit" functions to get a standard regression problem.