* Econometrics – the use of statistical methods applied towards economics
  + We always start with economic reasoning
* Regression Analysis
  + Dependent variable – y – is influenced by the explanatory variable
    - This is the variable we want to analyze
    - y
    - Alternative names for this variable:
      * response variable
      * target variable
      * output variable
  + Explanatory variable – x – the variable we use the information on to predict or describe changes in the dependent variable
    - Alternative names for this variable:
      * Predictor variable
      * Independent variable
      * Input variables
* To develop a regression model we include an error term which captures the stochastic nature of the relationship (non-deterministic). It is impossible to include all x’s that can help explain y
* Y(x1,x2,x3,x4,…)
  + The simplest function to analyze is a linear function so we say it is a linear function
* Simple linear regression model
  + - Β – is unknown and is the parameter. Once you estimate it then you can call it B0, B1.
    - B is the estimate. Beta is the parameter
    - This is missing the error term, since this is a stochastic relationship
      * Epsilon is the beginning and end of econometrics
      * Epsilon – the error term since this is a random relationship
* Multiple regression model:
* 𝑦 = 𝛽0 + 𝛽1𝑥1 + 𝛽2𝑥2 + ⋯ + 𝛽𝑘𝑥𝑘 + 𝜀
* 𝛽0, 𝛽1, 𝛽2, … , 𝛽𝑘 are unknown parameters
  + We use the sample data to obtain b0,b1,b2,…,bk

There are two important perspectives for econometrics:

* Predictive modeling perspective that captures conditional expectation: E(y|x1,x2,x3,…)
  + If your trying to predict then you are interested in this
* Casual estimation perspective that captures the partial effect, 𝜕y/𝜕𝑥j
  + If the policy is important then you are interested in this

Both of these perspectives are tremendously important:

* If you want to make a prediction the former is relevant
* If you want to make a decision the latter is relevant

# Simple Linear Regression Model:

* A simple linear regression model uses only one explanatory variable
* Consider y = Beta0+B1 + E

A graph of a simple linear relationship

Description automatically generated with medium confidence

# The multiple regression model:

* Sample data: n observations of y1, x1,x2,x3,…xk
* Use of the sample data to obtain b0,b1,b2 which are estimates of Beta 1, 2, 3…

# The estimated regression:

* Residual: e = y-y^
* Ordinary Least Squares (OLS) chooses the estimated regression equation by minimizing the error (residual) sum of squares, SSE
  + Gives an equation “closest” to the data

# Dummy Variables

* Explanatory variables can be numerical or categorical
  + Numerical – income
  + Gender – numerical
  + Loan default (yes or no) – categorical
* Dummy variable (d)– assumes 1 for one of the categories and 0 for the other
* Categorical variables cannot be used in their original form – that is, in a non-numerical format. We can convert a categorical variable into a dummy (indicator) variable