What is Linear Regression?

- Technique used for the modeling and analysis of numerical data
- Exploits the relationship between two or more variables so that we can gain information about one of them through knowing values of the other
- Regression can be used for prediction, estimation, hypothesis testing, and modeling causal relationships

Regression Lingo $Y = X1 + X2 + X3$

Y	=	X1 + X2 + X3
Dependent Variable		Independent Variable
Outcome Variable		Predictor Variable
Response Variable		Explanatory Variable

Why Linear Regression?

- Develop basic concepts of linear regression from a probabilistic framework
- Estimating parameters and hypothesis testing with linear models
- Linear regression in R

Suppose we want to model the dependent variable Y in terms of three predictors, X1, X2, X3
$Y = f(X1, X2, X3)$

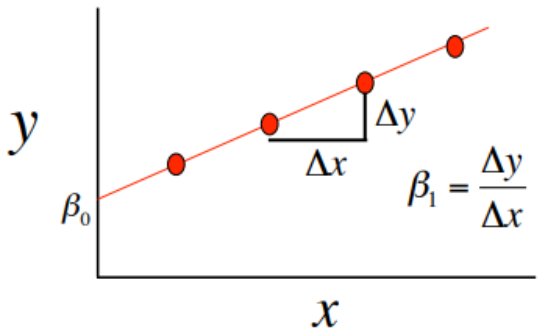
Typically will not have enough data to try and directly estimate f

Therefore, we usually have to assume that it has some restricted form, such as linear
$Y = X1 + X2 + X3$

Linear Regression is a Probabilistic Model

Much of mathematics is devoted to studying variables that are deterministically related to one another.

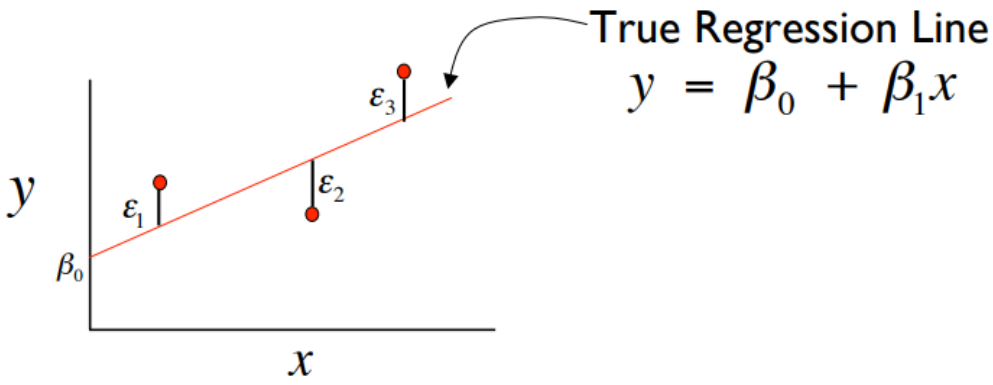
$$y = \beta_0 + \beta_1 x$$



But we're interested in understanding the relationship between variables related in a nondeterministic fashion

- ## A Linear Probabilistic Model

Definition: There exists parameters β_0, β_1 and such that for any fixed value of the independent variable x, the dependent variable is related to x through the model equation
• ε is assumed to be N $(0, \sigma^2)$



Implications

- The expected value of Y is a linear function of X, but for fixed x, the variable Y differs from its expected value by a random amount
- Formally, let x* denote a particular value of the independent variable x, then our linear probabilistic model says:

  E (Y | x*) = $\mu_{Y|x*}$ = mean value of Y when x is x *
  V (Y | x*) = $\sigma^2{}_{Y|x*}$ = variance of Y when x is x *