# STAT 425 Note 01

Wenxiao Yang

9/1/2021

## Introduction to Regression Analysis

### Regression Analysis

It is a "tool" used to examine the relationship between a *Dependent Variable* or *Response* $Y$ , and one (or more) *Independent Variables* or *Regressors* or *Predictors* $X_1, X_2, ..., X_p$.

## Simple Linear Regression

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

$\beta_0$ is the *intercept*; $\beta_1$ is the *slope*. One Response $\mathcal{Y}$; One Predictor $\mathcal{X}$ The data come in pairs:

$$\begin{array}{cc} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{array}$$

$Y$ is a RANDOM VARIABLE that has a distribution for every level of the independent variable.

### Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where the **intercept** $\beta_0$, the **slope** $\beta_1$, and the **error variance** $\sigma^2$ are the *model parameters*.

The **errors** $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are assumed to

– have **mean zero**: $E(\varepsilon_i) = 0$

– be **uncorrelated**: $Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$

– be **homoscedastic**: $Var(\varepsilon_i) = \sigma^2$ does not depend on $i$.