

# Linear Regression Model

Suppose that we have two variables

1.  $Y$  – the dependent variable (response variable)
2.  $X$  – the independent variable (explanatory variable, factor)

$X$  may or may not be a random variable .

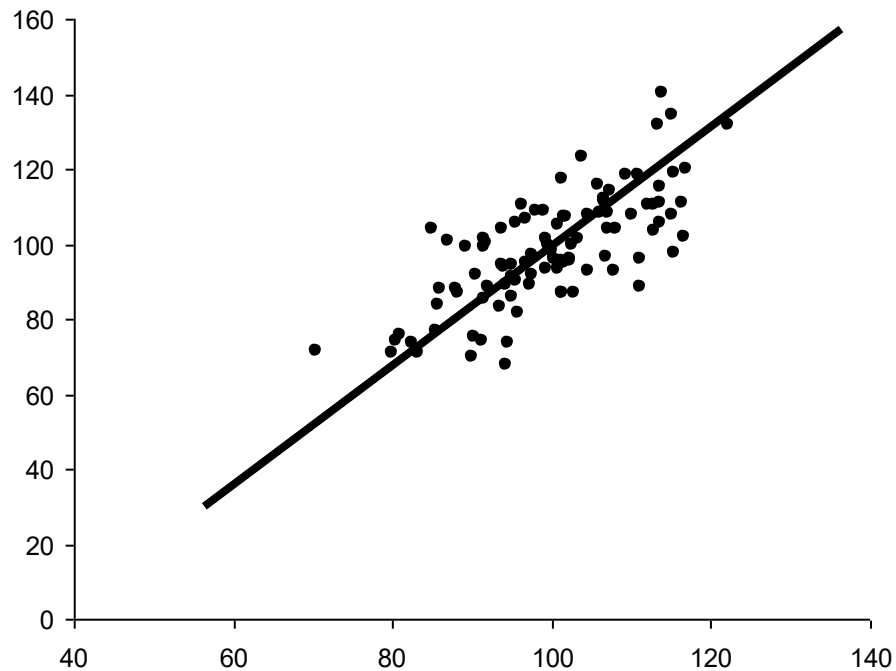
The dependent variable,  $Y$ , is assumed to be a random variable. The distribution of  $Y$  is dependent on  $X$

Assume that we have collected data on two variables  $X$  and  $Y$ . Let

$$(X_1, Y_1) (X_2, Y_2) (X_3, Y_3) \dots (X_n, Y_n)$$

denote the pairs of measurements on the two variables  $X$  and  $Y$  for  $n$  cases in a sample (or population)

- When data is correlated it falls roughly about a straight line.



# Regression Line

At a given value of  $x$ , the equation:

$$E(Y|x) = \beta_0 + \beta_1 x$$

Predicts a single value of the response variable

But... we should not expect all subjects at that value of  $x$  to have the same value of  $y$

Variability occurs in the  $y$  values!

# Simple Linear Regression Model

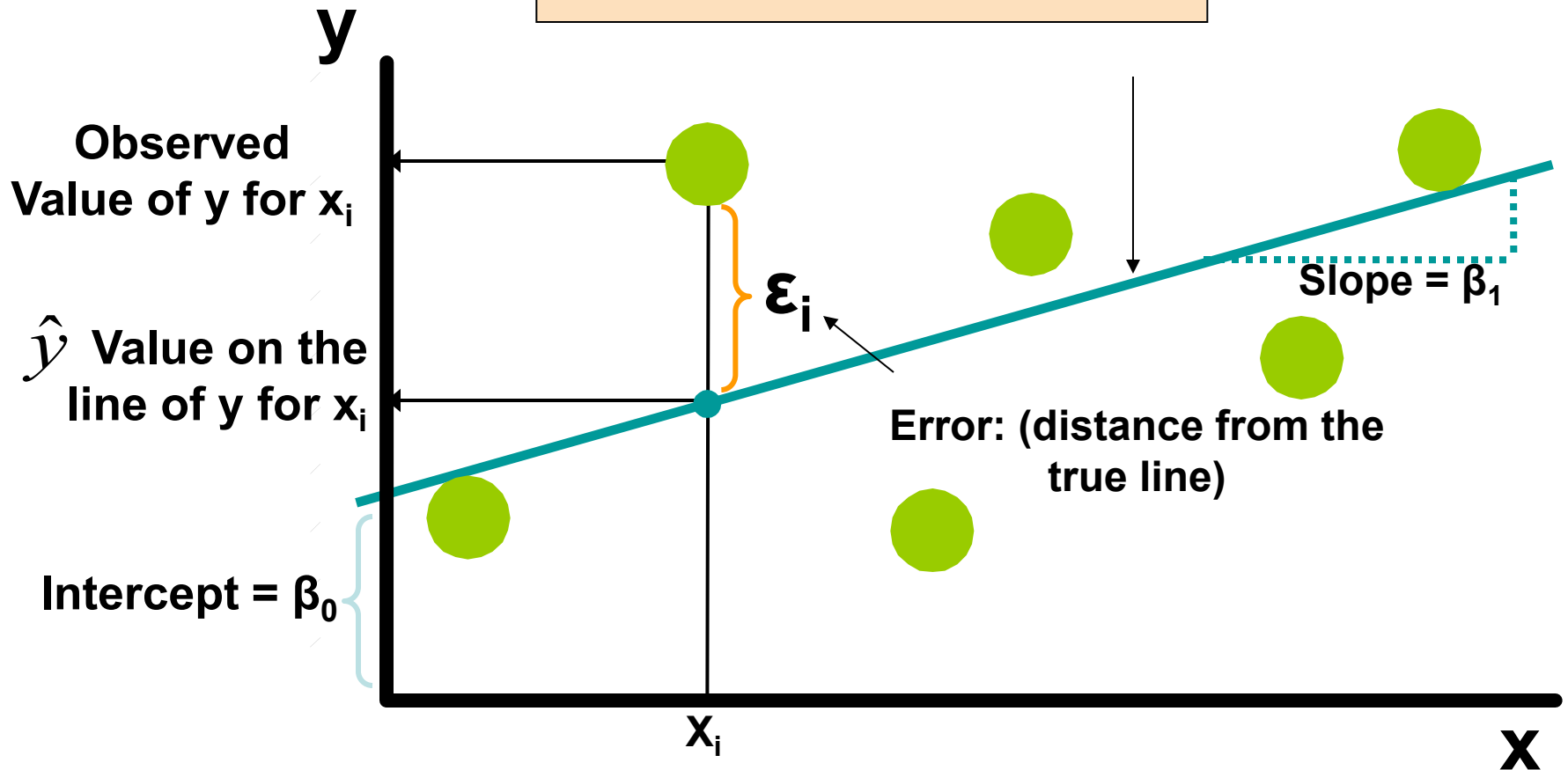
- $y = \beta_0 + \beta_1 x + \varepsilon$ 
  - $x$ : regressor variable
  - $y$ : response variable
  - $\beta_0$ : the intercept, unknown
  - $\beta_1$ : the slope, unknown
  - $\varepsilon$  : error with  $E(\varepsilon) = 0$  and  $\text{Var}(\varepsilon) = \sigma^2$  (unknown)
- The errors are uncorrelated.

# Linear Regression Model

- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ 
  - $x_i$ : regressor variable
  - $y_i$ : response variable
  - $\beta_0$ : the intercept, unknown
  - $\beta_1$ : the slope, unknown
  - $\varepsilon_i \sim N(0, \sigma^2)$  i.i.d
- The errors are uncorrelated.

# Simple Linear Regression Model

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



# Least Squares Estimation

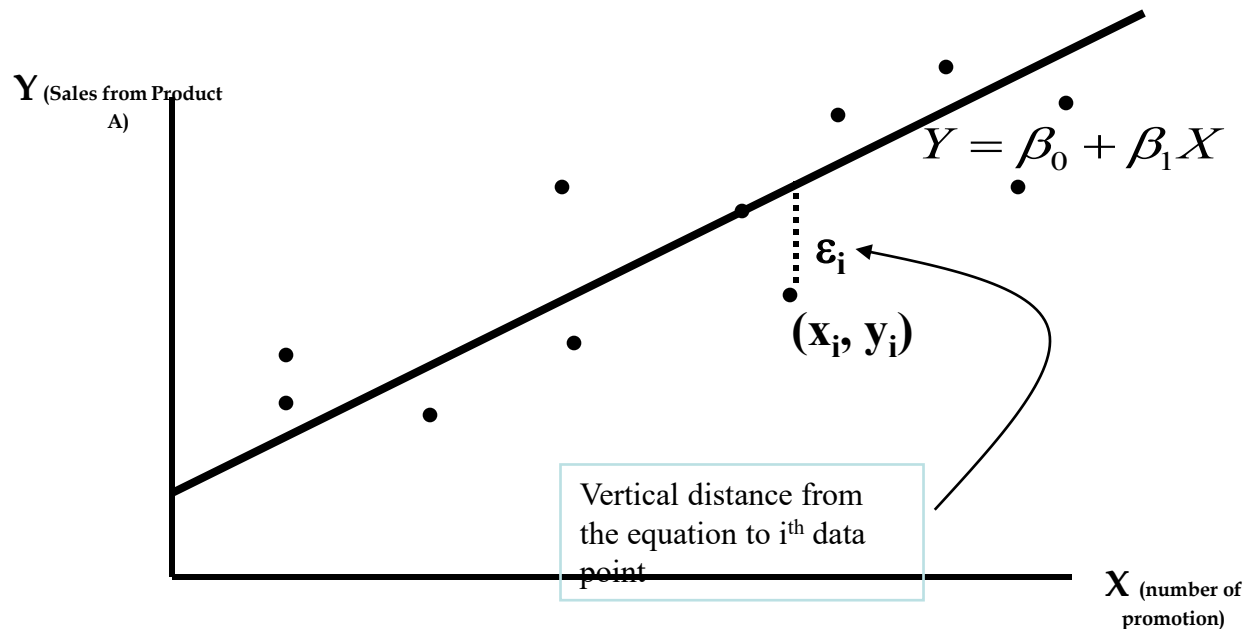
How to find best values for the parameters?  
Minimize distance between observed and  
corresponding value on the line

$$\min_{(\beta_0, \beta_1)} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\min \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

# Least Squares Estimation

We choose (estimate) the values for  $\beta_0$  and  $\beta_1$  so that the sum of the squared (vertical) distances from the equation to each data point is minimized.



$$S(\beta_0, \beta_1) = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2.$$

- $$-\frac{\partial S}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$-\frac{\partial S}{\partial \beta_1} \Big|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

- Least-squares normal equations:

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \end{aligned}$$

- The least-squares estimator:

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \end{cases}$$

$$- \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$- S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\begin{aligned} - S_{xy} &= \sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} \\ &= \sum_{i=1}^n y_i (x_i - \bar{x}) = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \end{aligned}$$

# Fitting linear regression

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

# Fitting linear regression

$$\hat{\beta}_1 = \frac{\frac{\sum_{i=1}^n X_i Y_i}{N} - \bar{X} \bar{Y}}{\frac{\sum_{i=1}^n X_i^2}{N} - \bar{X}^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

## Maximum Likelihood Estimation of the parameters

the intercept  $\beta_0$

the slope  $\beta_1$

the standard deviation  $\sigma$

Recall the joint density of  $y_1, y_2, \dots, y_n$  is:

$$\begin{aligned} f(y_1, y_2, \dots, y_n | \beta_0, \beta_1, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}} \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2} \\ &= L(\sigma, \beta_0, \beta_1) \\ &= \text{the Likelihood function} \end{aligned}$$

the log Likelihood function

$$l(\sigma, \beta_0, \beta_1) = \log L(\sigma, \beta_0, \beta_1)$$
$$= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

To find the maximum Likelihood estimates of  $\beta_0$ ,  $\beta_1$  and  $\sigma$  we need to solve the equations:

$$\frac{\partial l}{\partial \beta_0} = 0 \quad \frac{\partial l}{\partial \beta_1} = 0 \quad \frac{\partial l}{\partial \sigma^2} = 0$$

$$\frac{\partial l}{\partial \beta_0} = 0 \quad \text{becomes} \quad \frac{\partial}{\partial \beta_0} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 = 0$$

$$\frac{\partial l}{\partial \beta_1} = 0 \quad \text{becomes} \quad \frac{\partial}{\partial \beta_1} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 = 0$$

These are the same equations for the least squares line which have solution:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

The third equation:

$$\frac{\partial l}{\partial \sigma^2} = 0$$

$$\frac{\partial}{\partial \sigma^2} l(\beta_0, \beta_1, \sigma^2) = \frac{\partial}{\partial \sigma^2} \left( -\frac{n}{2} \log \sigma^2 - \frac{\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right)$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left( y_i - \left( \hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right)^2$$

# Sufficient Statistics

$$\begin{aligned} L(\sigma, \beta_0, \beta_1) &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - [\beta_0 + \beta_1 x_i])^2} \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2y_i[\beta_0 + \beta_1 x_i] + [\beta_0 + \beta_1 x_i]^2)} \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^n y_i^2 - 2\beta_0 \sum_{i=1}^n y_i - 2\beta_1 \sum_{i=1}^n y_i x_i + n\beta_0^2 + 2\beta_0\beta_1 \sum_{i=1}^n x_i + \beta_1^2 \sum_{i=1}^n x_i^2 \right)} \end{aligned}$$

# Sufficient Statistics

$$\begin{aligned}
 L(\sigma, \beta_0, \beta_1) &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - [\beta_0 + \beta_1 x_i])^2} \\
 &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2y_i[\beta_0 + \beta_1 x_i] + [\beta_0 + \beta_1 x_i]^2)} \\
 &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{n\beta_0^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^n y_i^2 - 2\beta_0 \sum_{i=1}^n y_i - 2\beta_1 \sum_{i=1}^n y_i x_i + n\beta_0^2 + 2\beta_0 \beta_1 \sum_{i=1}^n x_i + \beta_1^2 \sum_{i=1}^n x_i^2 \right)} \\
 &= h(x, y) c(\theta) \exp \left( \sum_{j=1}^k w_j(\theta) t_j(x, y) \right)
 \end{aligned}$$

$$T_1 = \sum_{i=1}^n y_i^2 \quad T_2 = \sum_{i=1}^n y_i \quad T_3 = \sum_{i=1}^n y_i x_i \quad T_4 = \sum_{i=1}^n x_i \quad T_5 = \sum_{i=1}^n x_i^2$$

# Sufficient Statistics

$$L(\beta_0, \beta_1) =$$
$$= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^n y_i^2 \right)} e^{-\frac{1}{2\sigma^2} \left( -2\beta_0 \sum_{i=1}^n y_i - 2\beta_1 \sum_{i=1}^n y_i x_i + n\beta_0^2 + 2\beta_0\beta_1 \sum_{i=1}^n x_i + \beta_1^2 \sum_{i=1}^n x_i^2 \right)}$$

$$T_1 = \sum_{i=1}^n y_i \quad T_2 = \sum_{i=1}^n y_i x_i \quad T_3 = \sum_{i=1}^n x_i \quad T_4 = \sum_{i=1}^n x_i^2$$

Thus all three estimators are functions of the set of complete sufficient statistics.

If they are also asymptotically unbiased then they are Uniform Minimum Variance Unbiased (UMVU) estimators (using the Lehmann-Scheffe theorem)

# Fisher Information Matrix

$$\begin{aligned}\frac{\partial}{\partial \beta_0} l(\beta_0 \beta_1, \sigma^2) &= \frac{\partial}{\partial \beta_0} \left( - \frac{\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) \\ \frac{\partial^2}{\partial \beta_0^2} l(\beta_0 \beta_1, \sigma^2) &= -\frac{n}{\sigma^2} \\ \frac{\partial}{\partial \beta_1} l(\beta_0 \beta_1, \sigma^2) &= \frac{\partial}{\partial \beta_1} \left( - \frac{\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_i)) \\ \frac{\partial^2}{\partial \beta_1^2} l(\beta_0 \beta_1, \sigma^2) &= -\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 \\ \frac{\partial^2}{\partial \beta_1 \partial \beta_0} l(\beta_0 \beta_1, \sigma^2) &= -\frac{1}{\sigma^2} \sum_{i=1}^n x_i\end{aligned}$$

# Fisher Information Matrix

$$\begin{aligned}\frac{\partial}{\partial \sigma^2} l(\beta_0 \beta_1, \sigma^2) &= \frac{\partial}{\partial \sigma^2} \left( -\frac{n}{2} \log \sigma^2 - \frac{\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right) = \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \\ \frac{\partial^2}{\partial \sigma^2} l(\beta_0 \beta_1, \sigma^2) &= +\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \\ \frac{\partial^2}{\partial \sigma^2 \partial \beta_0} l(\beta_0 \beta_1, \sigma^2) &= -\frac{1}{\sigma^4} 2 \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) \\ \frac{\partial^2}{\partial \sigma^2 \partial \beta_1} l(\beta_0 \beta_1, \sigma^2) &= -\frac{1}{\sigma^4} 2 \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_i))\end{aligned}$$