SFWR TECH 4DA3
Multiple Regression

Instructor: Dr. Jeff Fortuna, B. Eng, M. Eng, PhD, (Electrical Engineering),
McMaster University

email: fortunjj@mcmaster.ca

Simple Linear Regression Model

Probabilistic Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

where y_i = a value of the dependent **random** variable, y

x_i = a value of the independent **random** variable, x

β_0 = the y-intercept of the regression line

β_1 = the slope of the regression line

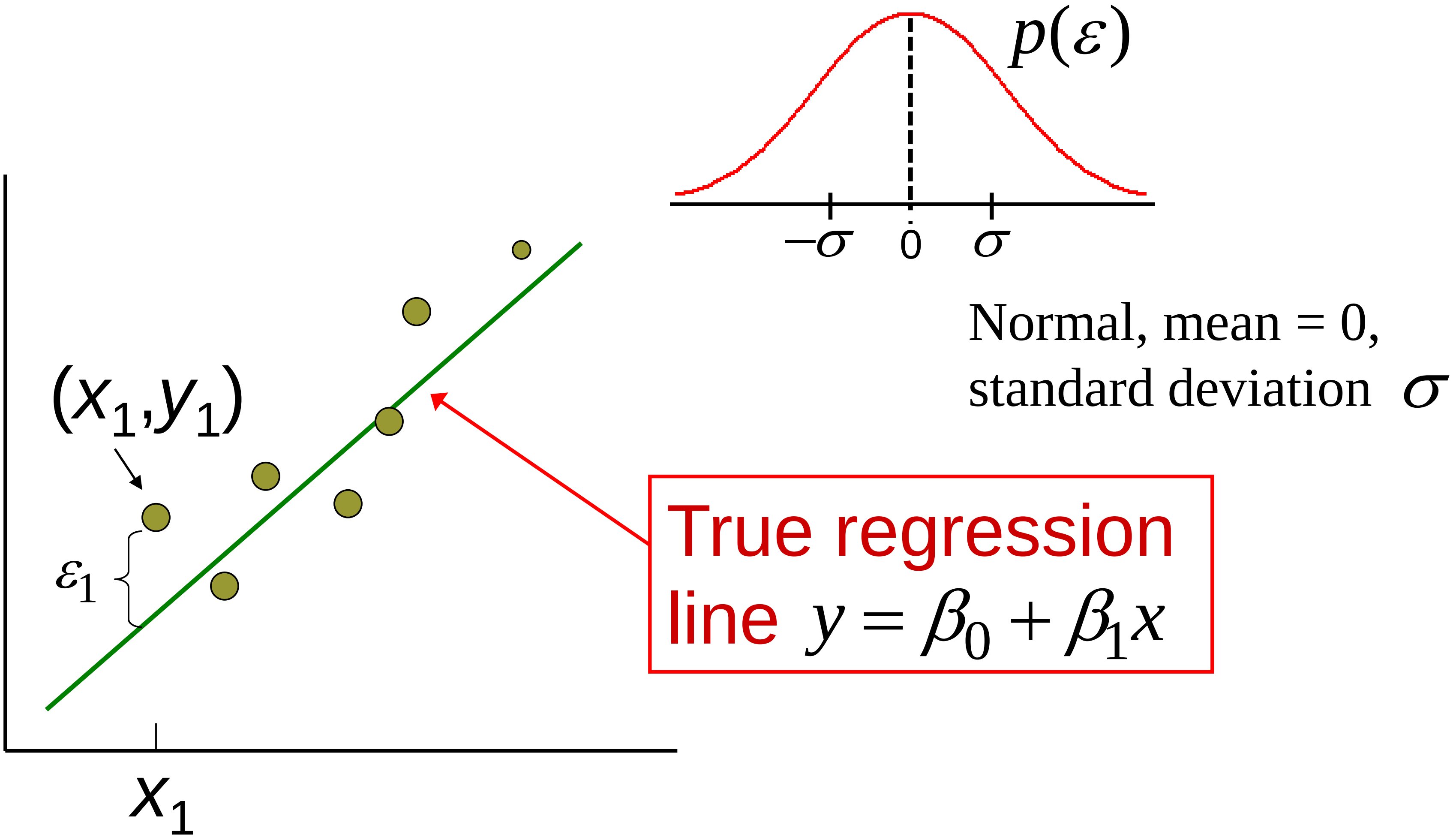
ε_i = random error, the residual

Deterministic Model:

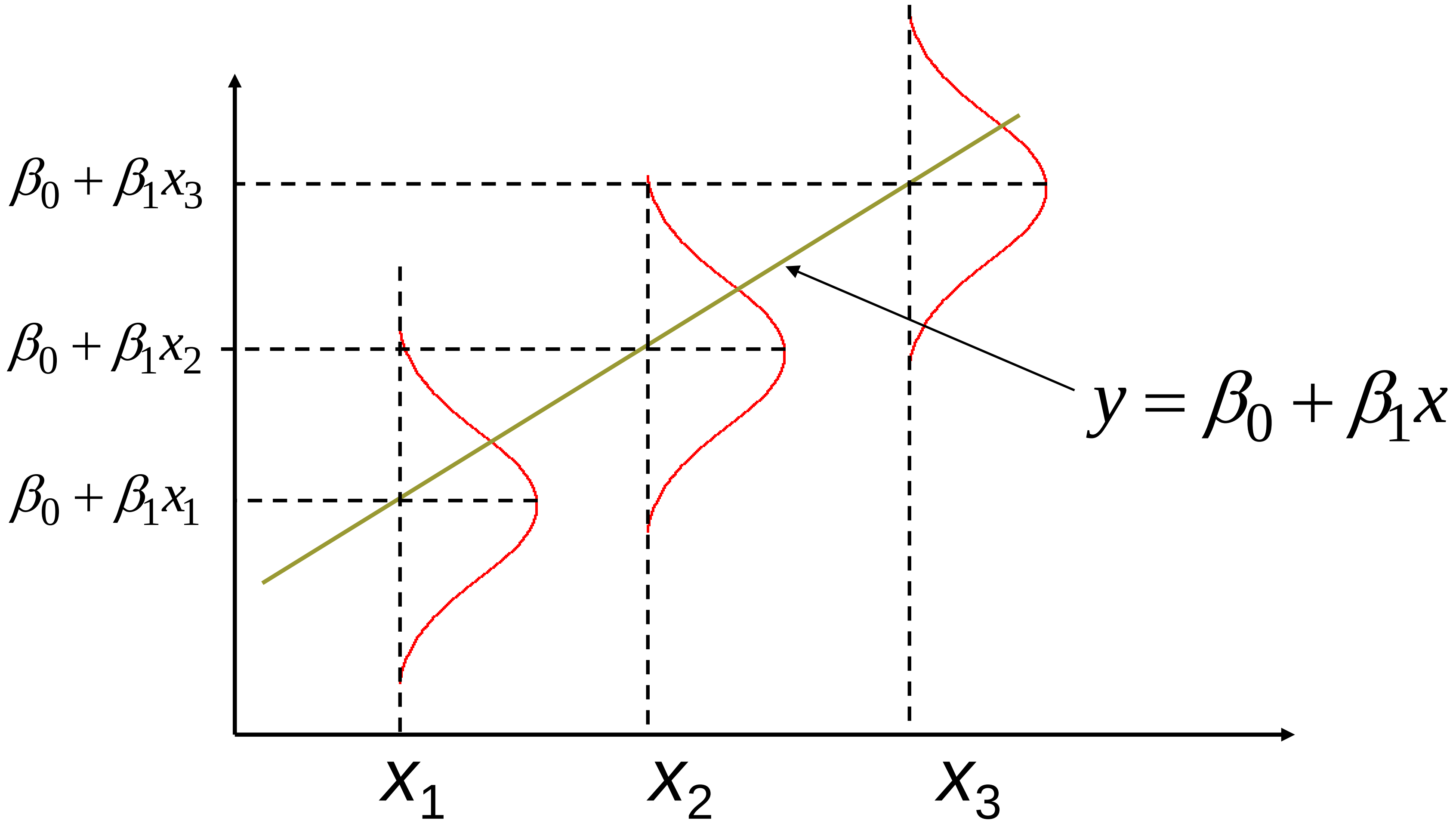
$\hat{y}_i = b_0 + b_1 x_i$ where $b_0 \approx \beta_0,\ b_1 \approx \beta_1$

and $\hat{y}_i$ is the **predicted** value of y in contrast to the actual value of y.

Linear Regression Model



Distribution of *Y* for Different Values of *x*



The Multiple Regression Model

- **Probabilistic Model**

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

where y_i = a value of the dependent random variable, y

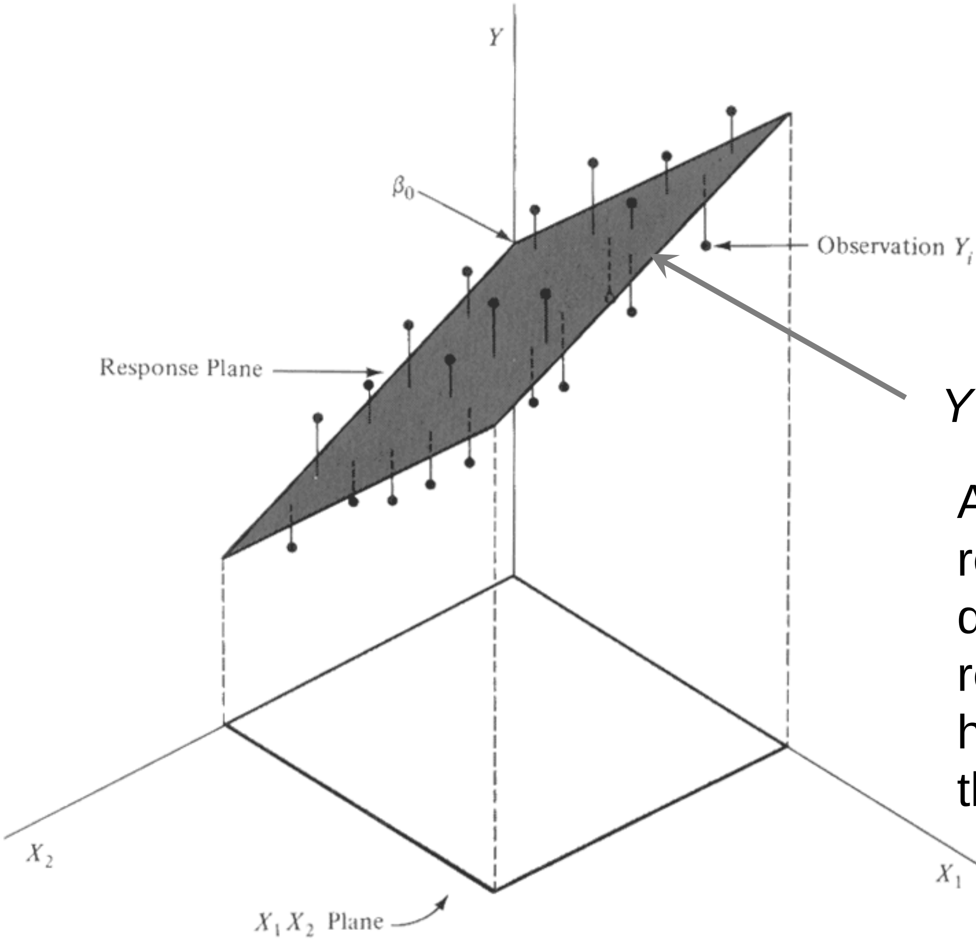
β_0 = the y-intercept

$x_{1i}, x_{2i}, \ldots, x_{ki}$ = individual values of the independent random variables, $x_1, x_2, \ldots, x_k$

$\beta_1, \beta_2, \ldots, \beta_k$ = the partial regression coefficients for the independent variables, $x_1, x_2, \ldots, x_k$

ε_i = random error, the residual

Multiple Regression



$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

Although we can't draw multiple regression in more than 2 dimensions, we can calculate the regression model in arbitrarily high dimensions. The geometry of the output is called a **hyperplane**

Vector / Matrix form

- The multiple regression model in vector / matrix form is very simple:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$\mathbf{Y}$ is $n \times 1$, $\mathbf{Z}$ is $n \times (p+1)$, $\boldsymbol{\beta}$ is $(p+1) \times 1$ and $\boldsymbol{\varepsilon}$ is $n \times 1$

Note: the first column of $\mathbf{Z}$ is a column of 1's:

$$\mathbf{Z} = \begin{bmatrix} 1 & \\ 1 & \\ \vdots & \mathbf{X} \\ 1 & \end{bmatrix}$$

That way, the first value of $\boldsymbol{\beta}$, β_0 is the y intercept for the plane and the rest of the $\boldsymbol{\beta}$ values are the slopes for each dimension.

Assumptions about the model

$Cov(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I}$ is an $n \times n$ covariance matrix

for the random errors.

Then,

$E(\mathbf{Y}) = \mathbf{Z}\boldsymbol{\beta}$ and $Cov(\mathbf{Y}) = \sigma^2\mathbf{I}$

where I is the $n \times n$ identity matrix:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Estimating the vector β

A reasonable way to estimate β is to chose the value of β that minimuzes the sum of squared residuals:

$$\min_{\beta}\left[(\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta})\right]$$

The solution (without proof) is:

$$\hat{\boldsymbol{\beta}}=(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}$$

In other words, the solution depends directly on the covariance of the data (**Z**) and the covariance of the data (**Z**) with the observations (**Y**)

- This problem can be framed in the context of ANOVA but I am not going to do that
- All I am doing here is showing the prediction **Y** from data **X**

Simple Example Using Excel…

- Done Live…

Predicting Values

- Assuming that we have found β, given a previously unseen x we can predict y by:

$$\mathbf{Y} = \mathbf{Z\beta}$$

where Z is a single row of data with a 1 as the first value

and β is a column vector of coefficients from the model.

$$Y = \begin{bmatrix} 1 & 6 & 8 \end{bmatrix} \begin{bmatrix} 5.12 \\ 3.86 \\ 1.95 \end{bmatrix}$$

$$Y = 5.12 + (6 * 3.86) + (8 * 1.95) = 43.88$$