MULTILINEAR REGRESSION

$\mathbb{R} = (-\infty, +\infty)$ = SET OF REAL NUMBERS

GIVEN: $\{(x_i, y_i) \mid x_i \in \mathbb{R}^p \,;\, y_i \in \mathbb{R} \,,\, 1 \le i \le n\}$

GOAL: FIT THE DATA USING A LEAST SQUARES APPROACH

OBSERVATIONS

- MULTILINEAR REGRESSION IS A STATISTICAL TECHNIQUE THAT USES EXPLANATORY VARIABLES TO PREDICT THE OUTCOME OF A RESPONSE VARIABLE.
- THAT IS, A CONTINUOUS DEPENDENT VARIABLE IS DEPENDENT ON TWO OR MORE INDEPENDENT VARIABLES.
- MULTILINEAR REGRESSION IS A GENERALIZATION OF SIMPLE LINEAR REGRESSION ANALYSIS.
- A LINEAR RELATIONSHIP IS ASSUMED BETWEEN THE DEPENDENT VARIABLE AND INDEPENDENT VARIABLES.

ANALYSIS

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i \quad ; \quad 1 \le i \le n$$

$$e_i = \text{ERROR TERM} \quad ; \quad 1 \le i \le n$$

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \quad ; \quad 1 \le i \le n$$

$$y^T = [y_1 \ y_2 \ \cdots \ y_m]$$

$$e^T = [e_1 \ e_2 \ \cdots \ e_m] = \text{ERROR VECTOR}$$

$$\beta^T = [\beta_0 \ \beta_1 \ \cdots \ \beta_p] = \text{PARAMETER VECTOR}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix}$$

$$Y = X\beta + e$$

$$E = \sum_{i=1}^{m} e_i^2 = e^T e \quad \text{IS MINIMIZED}$$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y} \quad \text{IF } (X^T X)^{-1} \text{ EXISTS}$$

SIMILAR TO THE LINEAR REGRESSION ANALYSIS (MATRIX FORMULATION)

NOTE THAT X IS AN $m \times (p+1)$ MATRIX

$\therefore$ $X^T X$ IS A $(p+1) \times (p+1)$ MATRIX

$p = 1$ RESULTS IN THE SIMPLE LINEAR REGRESSION MODEL.