



## §1. Introduction

Regression Analysis: A statistical technique for exploring and modeling the relationship between variables.

Ex: Sir Francis Galton - <Regression towards mediocrity in hereditary stature>

Relations: Functional and Statistical

Functional:  $Y = f(x)$

Statistical:  $Y = B_0 + B_1 x + \epsilon$ , where  $\epsilon$  represents error.

Relationship: in the form of a model, connecting the response and one or more predictor variable.

Ex: Weight and height, age and education, temperature and pressure ...

Graph: Scatterplot. 1<sup>st</sup> step in the investigation

Ex1: Inheritance of trait from generation to generation (Heights)

mheight (y)      dheight (x)

$$mheight = 0,4445 \cdot dheight + 34,1167$$

Ex2: Relationship between atmosphere pressure (pres) and the boiling point of water (bp). (Forbes)

Ex3: Predict Snowfall (ftcollinssnow)

late season (Late) vs early season (Early)  
(y) (x)

Mean Function:  $E(Y|X=x)$ , if there is a linear function  $Y = B_0 + B_1 x$ ,

$$E(Y|X=x) = B_0 + B_1 x$$

Variance Function:  $\text{Var}(Y|X=x)$  frequent assumption in fitting linear regression

$$\text{Var}(Y|X=x) = \sigma^2.$$

Ex 4: Summary Graph. (Anscombe)

$$X_1 - X_4, Y_1 - Y_4$$

Ex 5: Scattermatrix (fuel 2001).

Simple Linear Regression:  $Y = \beta_0 + \beta_1 x + \epsilon$

Multiple linear Regression:  $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$ ,  $p$ -dimension

Type of Variables:

Predictor: Quantitative, qualitative, mixture of quantitative and qualitative, quantitative or mix

Response: Quantitative, Quantitative, Quantitative, Qualitative

Technique: regression, ANOVA, ANCOVA, GLM (logistic regression)

\* Linear does not describe the relationship. If the parameters enter the equation linearly, we call it linear

$$Y = \beta_0 + \beta_1 \log x + \epsilon \quad \text{linear}$$

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \quad \text{linear}$$

$$Y = \beta_0 + \beta_1 x^{\beta_2} + \epsilon \quad \text{non-linear}$$

The observed relationship may be due to

① Causation, ② Common response, ③ Confounding

①  $X$  = food eaten amount,  $Y$  = hunger level

②  $X$  = # of TV sets,  $Y$  = life expectancy,  $Z$  = income

③  $X$  = poor eating habits,  $Y$  = child's weight,  $Z$  = mother's weight