Chapter 5
A Statistical Modeling for Policy Making: A Role of Regression Analysis

D. K. Ghosh
Saurashtra University, India

ABSTRACT

The concept of regression was first introduced by Sir Francis Galton in the context of inheritance of stature. Karl Pearson found the regression of son's height on father's height to be a biometrical fact. Soon the use of regression technique became too common for a variety of problems. The relationship between variables, if it exists, may be linear or curvilinear. In this chapter, a detailed discussion of linear regression is carried out. The regression analysis is an important statistical tool to find the relationship between different variables that can also be applied to the field of governance.

1. INTRODUCTION

The concept of regression was first introduced by Sir Francis Galton in context of inheritance of stature in human being (Bulmer, 2003). Similarly Karl Pearson (1857-1936) found the regression of son's height on father's height for biometrical fact. But soon the use of regression technique became too common for a variety of problems. The relationship between variables, if it exists, may be linear or curvilinear. In this chapter a detailed discussion of linear regression is carried out.

DOI: 10.4018/978-1-4666-5146-3.ch005

In the Government, semi-Government, self Finance, Private Offices etc, it is very much require knowing the work efficiency of the employee. There may be possibility of variation in the rates of work efficiency because of different age of the employee, different weight of employee, different types of habit to complete the work etc. Henceforth it is crucial for the Heads of the Institution of Public Governance to know relation between age, weight of the employee and time taken in completion of same work given to different employee. It will help the office to know the best employee for the particular work. For this purpose, the concept of regression analysis is required. It is understandable that, in the modern form of governance, efficiency is the key player. Thus the issues of public interests have to be addressed efficiently and statistics is an important tool to analyze and understand data (Croxton, F.E. & Cowden, J. 1949).

Linear relationship between two variables is represented by a straight line which is known as a regression line. The line of average relationship is another name for a regression line. In the study of linear relationship between two variables Y and X, suppose the variable Y is such that it depends on X, then we call it the regression line of Y on X. if X depends on Y, it is called the regression of X on Y. To find out the regression line, the observations (x_i, y_i) on the variables X and Y are necessarily taken in pairs, on the units which may be people, animal, plots, spare parts, plants or any other thing. For example, If we take height of ten Father then weight of ten Sons of the corresponding Father is to be considered, otherwise it will not give any relation. If a pair is formed of a father's height and eldest son's height at a particular age, in all cases this has to be maintained and not a combination of different sons with the different father should form a pair of observations. Generally the studies are based on samples of size n, and hence n pairs of sample observations can be written as $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$.

In experimental sciences the correlation between two variables, there are situations when it is necessary to estimate or predict the value of another character (variable say Y) from the knowledge of the other character (variable say X) such as to estimate height when weight is known. This is possible when the two are linearly correlated. The former variable (Y i.e. weight) to be estimated is called dependent variable and the latter (X i.e. Height) which is known, is called the independent variable. This is done by finding another constant called regression coefficient (b).

Regression means change in the measurement of a variable character, on the positive or negative side, beyond the mean. Regression coefficient is a measure of the change in one dependent (Y) character with one unit change in the independent character (X). It is denoted by letter 'b' which indicates the relative change (Yc) in one variable (Y) from the mean ($\bar{Y}$) for one unit of move, deviation or change (X) in another variable (X) from the mean ($\bar{X}$) when both are correlated (Dielman E.