Chapter 3
Uncovering Data for Decision Making With Critical Statistical Analysis

Soma Roychowdhury
Indian Institute of Social Welfare and Business Management, India

Debasis Bhattacharya
Visva-Bharati University, India

ABSTRACT

Every field of study generates a huge amount of data. The volume of data generated leads to information overload, and the ability to make sense of all these data is becoming increasingly important. This requires a good understanding of the data to be analyzed and different statistical techniques to be used in that context. On the basis of the issues important to the data set as well as other practical considerations, it is necessary to select appropriate methods to apply to the problem under study. This work focuses on different issues arising in the context of data analysis which need attention like understanding classifications of data, magnitude of errors in measurement, missing observations in the data set, outlier observations and their influences on the conclusion derived from the data, non-normal data, meta analysis, etc. In the process of discussion some examples have been included to illustrate how critical a data analysis procedure could be in order to make a meaningful decision from a data set.

DOI: 10.4018/978-1-5225-4947-5.ch003

1. INTRODUCTION

Data are lifeblood of every science, and it is through data we can have the scientific understanding of the events happening around us. All fields of study gather and store data, e.g., human resource department of an organization stores employee data on all relevant parameters, retail companies gather information on sales, insurance companies gather information on claims, meteorological department measures and collects data related to weather conditions, department of research and development of pharmaceutical companies generate data through clinical trials at various phases etc., with the purpose of making optimum decisions based on the information collected. The process of taking the raw data and converting them into meaningful information necessary to make decisions is the focus of this study.

Prior to starting any data analysis, the data should be characterized, cleaned, transformed (if necessary) and partitioned into appropriate form for further processing. Preparation of data is one of the most time-consuming parts of any data analysis exercise. How the data have been collected and prepared is critical to the confidence with which decisions can be made. The data should be reliable, and it should relate to the target population. The quality of the results obtained through data analysis depends on the quality of the data. By reliability of a measurement we mean that the measurement taken on the same individual or object for time after time produces the same result. A data set with reliable measurements constitutes a reliable data set. A valid measurement is one that actually measures what it claims to measure. If we try to measure job satisfaction using an IQ test, we would not get a valid measure of job satisfaction.

One should always exercise caution in data analysis. Let us consider the following two studies, where one may reach at a misleading conclusion if he or she does not have proper understanding of data handling.

A study was conducted to see how the time spent (in hours per week) on watching television programmes affect the academic performance of students as reflected in their grade point averages (GPA) out of 4 (see Table 1).

The correlation between television watched (hrs/week) and GPA comes out to be -0.02, which is negligible, almost zero, as if there is no relation between time

Table 1.

Hrs/Week	2.5	3.25	4	5.25	5.5	5.75	6.5	7	7.5	8	9.25
GPA	3.1	3.2	3	2.6	2.8	3.1	2.4	2.5	2.9	2.6	2.4
Hrs/Week	9.5	10.5	11	12	13	14	15	15.5	18	19.1	
GPA	3.1	3.6	3.4	2.6	3.3	2.1	3.8	2.1	3.9	2	