

Chapter 1

Analyzing data

1.1 INTRODUCTION

Imagine being asked to analyze the results of a market survey in which 500 individuals respond to about 40 questions. The questions deal with personal characteristics of the respondents (age, sex, marital status, level of education, income, etc.), the respondents' expenditures on selected products and services, and the number, type, and frequency of purchase of various newspapers and magazines. The survey results can be visualized as arranged in the form of a table, the rows of which correspond to respondents and the columns to the questions. Though such a survey is by no means large, as market surveys go, clearly it would be very difficult to draw any meaningful conclusions simply by glancing at the long list of figures.

However, by suitable reduction—very much as common sense and the techniques that follow describe—it is possible to make comparisons, measure relationships, display graphics, identify trends, and so on. It may be possible, for example, to form a “profile” of the “typical” reader of a sports magazine, to find out if there is any relationship between age and readership of newspapers, to measure the relationship between income and travel expenditures, or to determine trends in magazine readership. This analysis could be of use not only to publishers, but also to advertisers and firms having to decide in which publications a certain product will be advertised.

Techniques and methods by which large sets of data can be summarized and analyzed form the subject of this chapter.

1.2 DISTRIBUTIONS

Almost every type of analysis begins with a set of observations on one or more variables or attributes. This set of observations forms the raw material for statistical analysis. When the number of observations is large, it may be difficult to consider them in the raw form in which they were obtained—hence the need for various measures that reduce the data in a meaningful way.