

PART A

BASICS

Statistics is the science of collecting, summarizing, presenting and interpreting data, and of using them to estimate the magnitude of associations and test hypotheses. It has a central role in medical investigations. Not only does it provide a way of organizing information on a wider and more formal basis than relying on the exchange of anecdotes and personal experience, it takes into account the intrinsic variation inherent in most biological processes. For example, not only does blood pressure differ from person to person, but in the same person it also varies from day to day and from hour to hour. It is the interpretation of data in the presence of such variability that lies at the heart of statistics. Thus, in investigating morbidity associated with a particular stressful occupation, statistical methods would be needed to assess whether an observed average blood pressure above that of the general population could simply be due to chance variations or whether it represents a real indication of an occupational health risk.

Variability can also arise *unpredictably* (randomly) within a population. Individuals do not all react in the same way to a given stimulus. Thus, although smoking and heavy drinking are in general bad for the health, we may hear of a heavy smoker and drinker living to healthy old age, whereas a non-smoking teetotaler may die young. As another example, consider the evaluation of a new vaccine. Individuals vary both in their responsiveness to vaccines and in their susceptibility and exposure to disease. Not only will some people who are unvaccinated escape infection, but also a number of those who are vaccinated may contract the disease. What can be concluded if the proportion of people free from the disease is greater among the vaccinated group than among the unvaccinated? How effective is the vaccine? Could the apparent effect just be due to chance? Or, was there some bias in the way people were selected for vaccination, for example were they of different ages or social class, such that their baseline risk of contracting the disease was already lower than those selected into the non-vaccinated group? The methods of statistical analysis are used to address the first two of these questions, while the choice of an appropriate design should exclude the third. This example illustrates that the usefulness of statistics is not confined to the analysis of results. It also has a role to play in the design and conduct of a study.

In this first part of the book we cover the basics needed to understand data and commence formal statistical analysis. In Chapter 1 we describe how to use the book to locate the statistical methods needed in different situations, and to progress from basic techniques and concepts to more sophisticated analyses.

Before commencing an analysis it is essential to gain an understanding of the data. Therefore, in Chapter 2 we focus on defining the data, explaining the concepts of populations and samples, the structure of a dataset and the different types of variables that it may contain, while in Chapter 3 we outline techniques for displaying and tabulating data.