CHAPTER 1
MEDICAL AND ECONOMIC DATA STATISTICAL MODELLING

Nur Arina Bazilah Kamisan and Noraslinda Mohamed Ismail

1.1 INTRODUCTION

Statistics is a field that illustrates data more scientifically and properly. It identifies the data and by using statistics, one can analyse, illustrate, model, forecast and present their data technically. There are many techniques and models that can be applied to data. Each technique and model have different attributes which make them unique and could produce different results for the same data. Thus, for every statistical model, one could apply it based on their preferences. In this book, some applications of statistical models in medical and economic fields will be discussed.

The book will provide a few methods for analysing medical data in statistics. As a result of the study, a thorough evaluation of the aspects of the scenarios will be conducted, offering a deeper understanding of the causes or effects. Aside from that, statistical analysis could be utilised to estimate how long a medical case or disease will last. In economics, statistical applications are often used to anticipate the future growth of the economy and societal challenges. Because society and economics are intricately interwoven, understanding the economic issue is crucial. A healthy economy will aid a country's survival and ensure the well-being of its people.

Demonstrating a few applications of statistics within these two scopes will illustrate and explain how statistics may be used to forecast or analyse data to describe or demonstrate the data.

1.2 STATISTICAL MODELLING

Statistical modelling is the process of applying statistical analysis to a dataset. A statistical model is a mathematical representation or mathematical model of observable data. When data analysts use a variety of statistical models for the data they are analysing, they may strategically analyse and understand the data. Rather than trawling through raw data, this technique enables them to identify relationships between variables, forecast future sets of data, and show that data in a form that non-analysts and stakeholders can understand and apply.

There are numerous statistical models to choose from, and a competent data analyst should be well-versed in all of them. One should be able to determine not just which model is most suited to the data they are dealing with in each case, but also which model is best suited to the issue at hand. Knowing how different statistical models work and how they use data might help you choose which data is most relevant to the question you are trying to answer.

Data is rarely ready for analysis in its raw form. To ensure that your analysis is valid and viable, the data must first be cleaned up. Typical components of this cleaning include organising the obtained data and removing "poor or incomplete data" from the sample (Boulton & Hammersley, 2006). This point of view is supported by active label cleaning, which seeks to correct any biases by improving the training dataset's quality and preserving as many samples as possible. This is particularly crucial in safety-critical areas like healthcare, where model robustness must be validated on clean labelling (Bernhardt et al., 2022).

A solid understanding of statistical modelling may help you communicate more successfully with others since you will be better equipped to draw conclusions and, as a result, build better data visualisations, which are excellent for communicating complex ideas to

non-analysts. Simultaneously, having a solid understanding of how these models work on the backend will allow you to provide and explain more specific information as needed.

In this book, the application of statistical models such as Weibull proportional hazard model, Cox regression model, Bayesian logistic regression model, multiple logistic regression, multiple linear regression, Kaplan-Meier method, autoregressive (AR) model, autoregressive integrated moving average (ARIMA) model, seasonal ARIMA (SARIMA) model, and fuzzy time series model will be demonstrated with data from medical and economics area.

1.3 STATISTICAL MODELLING APPLICATION

There are many types of data that can be used for statistical modelling. Different types of models serve different purposes of focus. In this book, there are two types of data that will be discussed: the medical data and the economic data. These two data sets are among the most used statistical models.

1.3.1 Statistical Modelling in Medical Field

Understanding the complexities of health needs research that examines several elements and their interrelationships. This includes clinical and laboratory data, as well as additional variables such as risk factors and socioeconomic indicators are included (Dopson & Fitzgerald, 2005). Statistical modelling is utilised to incorporate several variables into a mathematical equation that shows the variables' relativities. Prediction and explanation are two of the most common modelling purposes.

Mistakes in model selection might lead to prejudice and misunderstanding. One of the most common flaws in a prediction model is the use of statistical variable selection methods to find causes. Regardless of how perfect the case design and measurements are, the unverifiable assumption of no unmeasured confounding of exposure effect is required for causal inference from observational data, whether confounding