The appreciation of statistical thoughts

Zhao Yujie

Taylor & Francis
Taylor & Francis Group

BOOK REVIEW

The appreciation of statistical thoughts, Jinglong Wang, Beijing, Science Press, 2018, ¥28 (paperback),
ISBN: 9787030519474

The Appreciation of Statistical Thoughts is written by Professor Jinglong Wang from East China Normal University. Motivated by extensive research interests along with long-term teaching experience, Professor Wang wrote this book to share important ideas in statistics. In general, Professor Wang views statistics as a combination of science and art. It is easy for us to regard statistics as a science subject due to its mathematical formula. Yet, it can also be seen as an art subject, because the applications of statistics depend lots on decision-makers' judgements, such as the choice of significance level. Therefore, in the application of statistics, it is not omnipotent for us to directly copy statistical formulas. Instead, man-induced flexibility and innovation are required for good statistical decisions, which makes statistics behave similar to art. After introducing the main ideas, Professor Wang follows some important concepts in each chapter, including induction and deduction, exactness and asymptotic inference, quantity and quality, correlation and causality, data description and modelling, sample and population, randomised controlled trial and blinded-experiment, as well as many high-level statistical thoughts.

This book is designed for a broad audience. For beginners, it provides them with a whole picture of statistics and a lot of real-life examples to explain obscure theories. For researchers and data scientists, this book helps guiding their work from a high-level insight. The key characteristic of this book is that it contains many examples to help readers understand abstract statistical ideas through practical applications.

This book consists of nine chapters.

Chapter 1 is an introduction of statistics – statistics is a combination of science and art for the collection and analysis of data. Firstly it introduces the Second World War bomber reinforcement story showing that solving statistical problems sometimes requires converse thinking and imagination. Then an example of marital status and death rate is used to illustrate that, attention ought to be paid for correct usages of statistical methods for the elimination of stereotype and misunderstanding. For this example, without correct statistical methods, one may jump into the conclusion that the death rate of the married is higher than those single. Finally, it gives two examples – the first one is to study the relation between temperature and the number of laid eggs to show the power of regression, and the second one is to study the frequency of the amount of insurance compensation to show the importance of heavy-tailed distributions. These two examples tell us that there is no best, only better method when you solve the statistical problems.

Chapter 2 distinguishes the difference between induction and deduction, and introduces the concepts of statistical inference, non-parametric statistics, robustness and data quality. The example of president selection is used through this chapter to show the efficiency of good sampling in statistical inference.

Chapter 3 discusses the difference of precision in statistics and mathematics, where the concepts of hypothesis test and confidence interval are introduced. To demonstrate that the statistical inference introduces intervals with uncertainty, various examples, such as the survey of US president and UK parliamentary elections, are used to provide readers with physical understanding towards confidence intervals in statistics.

Chapter 4 develops the techniques to transform quantitative problems into qualitative analysis in statistics, which requires a well-designed index system with properly assigned weights. To apply the theoretical methodology to practical problems, the author uses the CPI (Consumer Price Index) dataset in HK during 1994–1995 to show the criteria of good indices and methods to choose weights.

Chapter 5 introduces another pair of important concepts in statistics, correlation and causality. The first part of this chapter describes three well-known correlation coefficients with their formulas, assumptions and implications, namely Pearson coefficient, Spearman coefficient and Kendall coefficient. The remaining of this chapter briefly discusses the relation between correlation and causality.

The difference between set (mathematics) and population (statistics) is discussed in Chapter 6, where it first gives the mathematical definitions of set, and then introduces the concepts of population in statistics along with widely-used population distributions, such as normal distribution, Weibull distribution and so on.

Chapter 7 discusses modelling techniques in statistics. The author first explains the reason for the popularity of statistical modelling, and then provides a concrete modelling example by utilising descriptive statistics and statistical inference together with model diagnosis.

Chapter 8 talks about three important terminologies in statistics, namely the definition of retrospective survey, prospective survey, and randomised controlled blinded experiment.

Chapter 9 discusses the critical attitude towards statistical inference. The author suggests that we should not only pay attention to the location of the data, but should also take the variation of the data into account. For scenarios when the sample mean is not reliable, the author provides three alternative methods. Besides, the author encourages us to be critical towards the inference made by statisticians. Questions should always be raised towards the statistical inference, including the data quality, sample size, as well as analysis methods. Finally, the author suggests being consciousness of risk prevention to small probability events.

As a conclusion, *The Appreciation of Statistical Thoughts* is a well-written book, which can not only be used as tutorial for beginners, but also serves as a reference book for data scientists seeking high-level statistical insights. Besides, the various cases studied in this book works perfectly for applying the statistical theories into real-life applications. Finally, the story-telling writing style, based on the author's years of teaching experience, makes it interesting enough even as a bedtime reading material.

Zhao Yujie
Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA, USA
✉ yzhao471@gatech.edu

Check for updates