Beyond Statistical Methods – Compendium of Statistical Methods for Researchers

Ondřej Vozár[1] | *Czech Statistical Office, Prague, Czech Republic*

The book serves as a compendium of statistical methods for researchers in a very broad sense. In addition to an overview and explanation of the broad scope of statistical methods it also provides exposition to the principles of empirical research, research plans, ethical considerations, meta-analysis and a practical guide on how to write or read statistical reports.

The first part deals with principles of empirical research, how to formulate and verify hypotheses, write a proposal of research project and also ethics of the research. Ideas on population, sampling, types of variables, measurement errors, various research plans (census, sample survey, experiment, observational study) are very clearly explained. Also practical considerations of data collection, processing and dealing with missing values are given. Statisticians have to learn these issues mostly by trial and error in their practice.

The second part provides a deep but very readable introduction to descriptive statistics, statistical graphs, analysis of outlying values, transformation of the variables, probability, common probability distributions, central limit theorem and sample distributions of the relative frequency, mean and variance.

The third part on statistical induction and statistical methods forms the core of the book. The author combines clear notation in formulae, graphs and solved real life numerical examples to explain the ideas behind the methods. The reader is always informed on possible traps, misinterpretations and ways how to fix them. Author aptly uses decision trees and block diagrams for cases when it is necessary to make choice of the method or decide on the interpretation of the data.

The fifth chapter provides a lucid explanation of point estimates, interpretation of confidence intervals, steps of hypotheses testing, characteristics of the test (power and significance level), simultaneous hypothesis testing and ways how to avoid misinterpretation of the null hypothesis. These vital issues

---
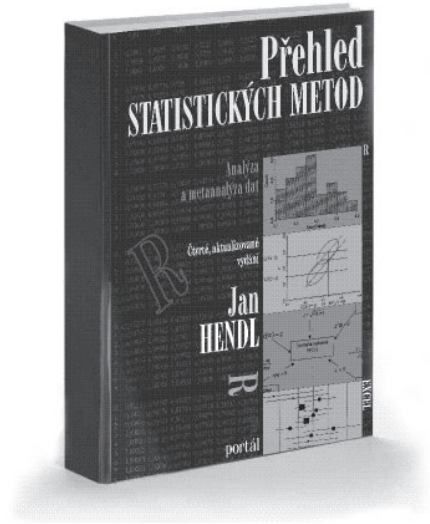
[1]   Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: ondrej.vozar@czso.cz.

are not usually explained in such a detail in the most of available textbooks. The sixth chapter discusses standard situations in the hypotheses testing (testing mean or variance in one sample, two samples or in paired data). Then non-parametric tests on mean and normality are presented.

The seventh chapter deals with analysis of bivariate data including graphs, correlation and regression analysis. In the subsequent chapters standard methods of categorical analysis, analysis of variance and multivariate regression are presented. Reader is kept informed on possible traps (Simpson's paradox, randomization) in application of the methods.

The eleventh chapter guides statisticians in how to impartially assess the size of effects in a statistical study. For example it means to find which value of the sample correlation coefficient means a high effect with given significance level, power of the test and sample size. The twelfth chapter provides guidelines on the choice of the appropriate statistical method based on clear classification. The second part explains common misinterpretations of the results of hypotheses testing and how to avoid or corrects them. Bayesian approach to the statistical induction and computer intensive methods (bootstrap, jacknife, crossvalidation) are explained.

The thirteenth chapter explains multivariate statistical methods including logistic regression analysis, generalized linear models, regression trees, survival analysis, cluster analysis, principal component analysis, factor analysis, discrimination analysis, multivariate scaling, multivariate contingency tables etc. The scope of the methods and the part on structural equations models using latent variables is unique in the Czech statistical literature.

The fourteenth chapter is the first systematic treatise on meta-analysis in the Czech statistical literature. It explains all the phases of this kind of study, presents relevant statistical methods, gives a guide on how to write a research report and discusses its pros and cons.

The fifteenth chapter gives clear guidelines on the structure of a final report and corresponding ethical considerations.

The last chapter gives an overview of the statistical packages, criteria of the method selection and a brief introduction to the language R ("lingua franca" of the statisticians). The only weak point is that the overview of the packages was a bit outdated even in 2012.

Annexes contain the model structure of a final report, statistical tables, basic definitions and formulae of matrix algebra, description of the basic commands in R language, etc.

As mentioned above the book is unique in its scope, as it provides the only up-to-date presentation of some topics in the Czech language. The approach to the presentation combining formulae, graphs, decision trees and solved numerical examples enables understanding of the complex issues for a broad audience from various backgrounds. In my view this book can serve as a bridge between statisticians and researchers, but also between the authors of the scientific papers and reports and their readers.

Having said that I can recommend this book as a reference on the statistical methods and empirical research for professional statisticians, researchers and students in general.