SKILLS NEEDED FOR MODERN DAY STATISTICIANS

Andrej Blejec
National Institute of Biology, Ljubljana, Slovenia
andrej.blejec@nib.si

Probably we agree that being a statistician is an interesting but complex profession. Statistics is not an isolated field but is always connected with some sicientific or societal content. Statisticians need to interact with professionals and general public, interested in diverse fields of human endevaur. To be successful in this interaction, statisticians have to have certain skills. Some are gained and developed during their studies, yet some are adapted later in their career. What skills, other then knowledge of probability and statistical theory, can be listed?

Statistics has tools to solve a range of problems involving certain amount of uncertainty. Typically people are not good in reasoning with uncertainty. To interact with other people, communication skills are needed. For successful use of statistics, we have to adapt and understand basic concepts of the application field.

In recent years data are produced with unprecedented speed and quantity. This calls for extensive use of information technology, especially computational and visualization tools on the web. Needed computational skills adapt to contemporary and future uses of statistics in practice. In the past hundred years we faced several, what we might call, technological revolutions. Mechanical calculators were replaced by digital mainframe computers, openig completely new ways to do satistics. With little statistical software available, computer programing skills were needed to be a statistician. With the advent of statistical packages, proficiency in use of one or two statistical packages, usually the one that was available at particular university computer center, was taught. Data were organized in formats suitable for specific packages, preferably in the plain text tabular form. Then the computing power appeared on our tables, with less limited options for statistical software. With desktop computing, data analysis software diversified and thus selection of prefered statistical software skills become difficult. On the other hand, user friendly computing availability enabled non-statisticians to analyze their own data. In the past decade or two, modern statistical computation is based on the S language. Open source and free R language become lingua franca for statistical computing. There is little doubt that proficiency in R language is what modern day statisticians need. Nowadays R is used both for development of modern statistical methods and data analysis in diverse fields of applications. R can support a rich set of modern statistical graphics and visualizations. With growing community of users and developers, it is safe to predict that there will be place for R on the computer of every statistician.

Currently we face tremendous changes in availability of information. More and more data are publicly available. Demand for open access to public data, and growing quantities of corporate data, lead to the phenomenon of Big data. Big data are not only large in numbers of observations, variables or both, but also unstructured. Large quantity and dimensionality of data can be tackled by bigger computer, distributed computation, and better algorithms. Unstructured, non-tabular, data that were not collected for the purpose of analysis and are usually a by-product of social and economic processes are another problem. Data need to be structured and often collected from a plethora of sources, requiring additional programming information technology skills as well as proficiency in data analysis. New internet-based tools for analysis and visualization enable “data scientists” to quickly produce appealing data summaries, suitable for the fast changing environment, where data are often short lived and results quickly become obsolete. For better communication of result we must emphasize visualization, with dynamic and interactive graphics.

To be competitive with all others who want to analyze data, we have to be flexible, responsive and productive, using statistical know-how as the competitive advantage.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org