Panels Session: Statistics beyond 2020: challenges and risks

by Maurizio Vichi,
Sapienza University of Rome

Statistical Science is radically changing, since the last 40 years, under the influence of the capillary diffusion in the world of internet and the new technologies such as computers and smartphones. I will discuss this changes with the viewpoint of the methodologist, underling at least a challenge and a risk in these changes.

How Statistical Science was 40 years ago, and it is now, regarding the following five main topics?

1. *Data usage*: we passed from the use of few data to the data deluge. From small samples of observations treated according to classical inferential statistics to large samples, and more recently to big data -coming from internet and from the globalization of the information- and analysed by multivariate statistical methods, data mining and data analytics methods.
Now, classical media, such as Radio, TV and Journals do not have any more the monopole of information since each user of internet is potentially a source of information.
Challenges & Rewards. The data deluge allows a better description of reality with many details and Statistical Science acquires the power to describe more neutrally the reality. In fact, the description of phenomena based on many open data, does not allow someone to impose his/her viewpoint, how it could be in case only few data partially selected are used. The abundance of data reduces the idea that Statistics can be used to impose the “distorted” perspective of the statistician or the viewpoint of the third person that has commissioned the statistical analysis.
Risks. With the data deluge: demagogues, populists or illiterate statistical users, with their own unchecked data, can provide their false “truth” of what is going on across society. In general, there is an increase of quantitative “alternative facts”, frequently “fake news”, coming from alternative sources of information. How to stop “fake news”? The nice IFLA infographics includes also “ask the expert” as a step to stop distorted information. The expert is certainly the professional statistician who can certify if the data have been produced according to the statistical principles, and have a sufficient quality for the purpose with which they have to be used.

2. *Statistical phenomena description* (e.g., complex phenomena such as the monitoring of the progress of nations): we passed from the use of few variables (e.g., GDP and few economic indicators considered as golden variables for the description of the progress of the nations), to the use of many indicators (e.g., Sustained Development Goals (i.e., 17 goals, 169 targets and 240 indicators)). Thus, from a partial to a comprehensive description of the phenomena;
Challenges. The comprehensive phenomena description by using many variables calls for the study of new statistical models that can deal with the increasing complexity of the data. Therefore, new statistical methods have to be studied in order to model complexity and uncertainty linked to reality.
Risks. There are only possible risks linked to the difficulties of communicating, in a simple way, complex phenomena. However, citizens of the 21st century need to have skills to

understand and analyse reality; thus, statistics must become a fundamental competence for the citizens, acquired by starting from school.

3*. Data collection*: it was manual and now is mainly electronic; frequently with an automatic web-collection (e.g., from internet of things, i.e., data automatically generated by machineries); thus, the data collection was mainly based on structured information (from questionnaires), now on, it will be more based on unstructured information, i.e., not organized in a predefined manner, including books, journals, documents, images, text such as emails, tweets, etc..
Challenges & Rewards. Statistics and especially official statistics can become faster in describing the reality; the cost of collecting data will be drastically reduced; the use of administrative data collected for different aims and reused for statistical purposes will become more relevant.
Risks. With the data automatic collection commercial companies can become sellers of statistical information with low quality (in order to facilitate sales reducing prices); thus, these companies will compete with National Statistical Institutes that for reasons of high quality will not be able to deliver information so quickly.

4. *Methodologies*: we passed from theoretical inferential statistics to intensive computer based inferential statistics and data analytics, including: simulation, statistical learning, bootstrap and resampling, multivariate data analysis, and analytics for social media;
Challenges. With the new methodologies, there are enormous developments of statistical algorithms and methodologies unthinkable few years ago; for example the development of algorithms for statistical data driven decisions to allow better decisions and for creating artificial intelligence.
Risks. Artificial intelligence will deprive millions jobs. There is a need to reconvert workers to skill that enhance the capacity to elaborate data. An additional risk regards the automatic decisions. Are automatic decisions always ethical?

5. *Competences and knowledges of statistics*: in the past a figure of the theoretical statistician was privileged; now the data scientist, i.e., the modern statistician with informatics skills and data communication skill is preferred. Together with mathematical and probability competences the computer science becomes a fundamental basic knowledge for the modern statistician. However, also the soft skill are relevant because the statistician generally has to work with a team.
Challenges. The inclusion of the new interdisciplinary competences is not optional; otherwise, statistics will lose the ability to describe clearly reality; there is a need of statistics professionalism to combat the improper use of data by illiterate statistical users, demagogues and populist, and reduce quantitative “fake news”.
Risks. The evolution of technologies and the rapid changes of the 21-century society produce a rapid deterioration of the competences. Universities must adopt a different education scheme: the five years education of the Bologna process, i.e., first degree no shorter than 3 years and master no shorter than 2 is not sufficient. There is a need to re-call the graduated people after a period (e.g., 10 years) for a cycle of retraining on the new methodologies and technologies in strict connection with enterprises that have to contribute to render education in statistics closer to what it is needed in the job market. The case of the EMOS, the European Master in Official Statistics organized by EUROSTAT and 22 European Universities is a clear example of optimal collaboration necessary to reduce the gap between the theory and practice.