Data, Data Science, and Official Statistics

Yves Tillé [1]

Abstract

In this article, we begin by discussing the terminology and methodologies of statistics and data science, focusing on the complexities of data production and validation. We trace the historical evolution from exhaustive censuses to modern sampling and the use of administrative sources and big data. The challenges of defining statistical units, handling partial data and ensuring accurate extrapolation are explored. The paper argues that many modern methods in data science are rooted in classical statistics. It also discusses current issues in official statistics, such as declining survey response rates, data protection and the integration of diverse data sources. Finally, it highlights future challenges, including better use of administrative data, error correction, technological innovation and the need to develop a skilled workforce.

---

[1] Yves Tillé (yves.tille@unine.ch), University of Neuchatel, Switzerland.