

ASA issues statement on role of statistics in data science

October 1 2015

In a [policy statement](#) issued today, the American Statistical Association (ASA) stated statistics is "foundational to data science"—along with database management and distributed and parallel systems—and its use in this emerging field empowers researchers to extract knowledge and obtain better results from Big Data and other analytics projects.

The statement also encourages "maximum and multifaceted collaboration" between statisticians and data scientists to maximize the full potential of Big Data and data science.

"Through this statement, the ASA and its membership acknowledge that data science encompasses more than statistics, but at the same time also recognize that statistical science plays a critical role in the fast-growing field," said ASA President David R. Morganstein, who is director of the statistical staff for Westat, Inc. "It is our hope the statement will reinforce the relationship of statistics to data science and further foster mutually collaborative relationships among all key contributors in data science."

The ASA statement acknowledges the lack of consensus on what constitutes data science, but notes the following essential role of each of the three computer science and statistics professional communities that are foundational to the field:

- Database Management, which enables transformation, conglomeration, and organization of data resources

- Statistics and Machine Learning, which convert data into knowledge
- Distributed and Parallel Systems, which provide the computational infrastructure to carry out data analysis

"At its most fundamental level, we view data science as a mutually beneficial collaboration among these three professional communities, complemented with significant interactions with numerous related disciplines," says the ASA statement.

It continues by elaborating on the key role of statistics in the data science field: "Framing questions statistically allows researchers to leverage data resources to extract knowledge and obtain better answers. The central dogma of statistical inference, that there is a component of randomness in data, enables researchers to formulate questions in terms of underlying processes and to quantify uncertainty in their answers. A statistical framework allows researchers to distinguish between causation and correlation and thus to identify interventions that will cause changes in outcomes. It also allows them to establish methods for prediction and estimation, to quantify their degree of certainty, and to do all of this using algorithms that exhibit predictable and reproducible behavior. In this way, [statistical methods](#) aim to focus attention on findings that can be reproduced by other researchers with different data resources. Simply put, statistical methods allow researchers to accumulate knowledge."

The statement also calls on the ASA membership to expand the cooperative relationships already in place among data science practitioners: "For statisticians to help meet the statistical challenges faced by data scientists requires a sustained and substantial collaborative effort with researchers with expertise in data organization and in the flow and distribution of computation. Statisticians must engage them, learn from them, teach them and work with them. Engagement must occur at all levels: with individuals, groups of researchers, academic

departments and the [data [science](#)] profession as a whole."

New problem-solving strategies are needed to develop "soup-to-nuts" pipelines that start with managing raw data and end with user-friendly efficient implementations of principled statistical methods and the communication of substantive results. Engendering these next-generation strategies will be fostered from the ground up in [data science](#) and statistics programs at colleges and universities across the country, explains the statement.

"Statistical education and training must continue to evolve—the next generation of statistical professionals needs a broader skill set and must be more able to engage with database and distributed systems experts. While capacity is increasing within existing and innovative new degree programs, more is needed to meet the massive expected demand. The next generation must include more researchers with skills that cross the traditional boundaries of statistics, databases and distributed systems; there will be an ever-increasing demand for such 'multi-lingual' experts," concludes the statement.

Provided by American Statistical Association

Citation: ASA issues statement on role of statistics in data science (2015, October 1) retrieved 15 August 2024 from <https://phys.org/news/2015-10-asa-issues-statement-role-statistics.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--