Why Statistics?

POPULAR MEDIA AND SCIENCE PUBLICATIONS SOUND THE DRUM: "BIG DATA" WILL DRIVE OUR future, from translating genomic information into new therapies, to harnessing the Web to untangle complex social interactions, to detecting infectious disease outbreaks. Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances. This field will become ever more critical as academia, businesses, and governments rely increasingly on data-driven decisions, expanding the demand for statistics expertise.

The melding of science and statistics has often propelled major breakthroughs. Last year's Nobel Prize in Physics was awarded for the discovery of the accelerating expansion of the universe. That discovery was facilitated by sophisticated statistical methods, establishing that the finding was not an artifact of imprecise measurement or miscalculations. Statistical methods also allowed the trial demonstrating that zidovudine reduces the risk of HIV transmission from infected pregnant women to their infants to be stopped early, benefiting countless children. Statistical principles have been the foundation for field trials that have improved agricultural quality and for the randomized clinical trial, the gold standard for comparing treatments and the backbone of the drug regulatory system.

Statistics often informs policy development. For example, in the United States, billions of dollars are allocated to school districts based on county-specific estimates of income and poverty, derived by combining data using statistical methods. In evaluating pollutants, statistical modeling isolates true associations with illnesses and deaths. Big Data payoffs can be enormous, but there are many pitfalls. Take the promise of personalized medicine: Achieving this goal will require the integration of vast landscapes of genomic, clinical, and related data from legions of patients. The potential for false discovery looms large. New statistical methods will be needed to address some of these issues. Similar challenges arise from the haystacks of information on social network, time-use, economic, and other activities that can be mined to benefit science, business, and society. Close collaboration with statisticians is the best way to ensure that critical issues are identified and solutions found.

A dramatic increase in the number of statisticians is required to fill the nation's needs for expertise in data science. A 2011 report by a private consulting firm projected a necessary increase of nearly 200,000 professionals (a 50% increase) by 2018.* Graduates specializing in statistics are equipped with skills that allow them to pursue diverse careers, and there has been a surge in applications for graduate education in these fields. But available places are limited; for example, the ratio of qualified applicants to slots in the Johns Hopkins Biostatistics program exceeds 10 to 1. Resources must be found to expand the number and size of graduate programs.

Online
sciencemag.org
Listen to the author podcast at http://scim.ag/ed_6077

No amount of statistical intervention can circumvent flawed subject-matter models or salvage valid conclusions from poorly designed studies, and even sound statistical analysis may fail to yield straightforward answers. The future demands that scientists, policy-makers, and the public be able to interpret increasingly complex information and recognize both the benefits and pitfalls of statistical analysis. It is a good sign that the new U.S. Common Core K-12 Mathematics Standards† introduce statistics as a key component in precollege education, requiring that students be skilled in describing data, developing statistical models, making inferences, and evaluating the consequences of decisions. Embedding statistics in science and society will pave the route to a data-informed future, and statisticians must lead this charge.

– Marie Davidian and Thomas A. Louis

10.1126/science.1218685

Marie Davidian is William Neal Reynolds Professor of Statistics at North Carolina State University, president-elect of the American Statistical Association, and executive editor and former editor of *Biometrics*. E-mail: davidian@ncsu.edu.

Thomas A. Louis is professor of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, retiring chair of the AAAS Statistics Section (U), former president of the International Biometric Society, and former editor of *Journal of the American Statistical Association* and *Biometrics*. E-mail: tlouis@jhsph.edu.

*www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_Data_The_Next_frontier_for_innovation.
†www.corestandards.org.