Statistics within business in the era of big data

Gareth M. James

Department of Data Sciences and Operations, University of Southern California, United States

ABSTRACT

The last decade has seen a dramatic increase in the availability of business data. Here I discuss some of the corresponding opportunities and challenges for business related statistical applications and the role that statisticians can play within a business school.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Over the last century the field of statistics has made significant contributions to efforts aimed at developing rational and objective interpretations of data, and information in general. Furthermore, there can be little doubt that the prominence and stature of the field has grown considerably over the last 20 years. This growth has been primarily driven by developments outside statistics, in particular technological advances that have both created vast quantities of new and complicated types of data, and correspondingly provided us with the mechanism for its analysis.

The late nineties, when I graduated, was a good time to be a statistician. The first wave of new data, the sequenced human genome, had just arrived. Suddenly biologists were inundated with enormous quantities of data, of a type that had never been studied before. This opened up the possibility to address significant new research questions and statisticians made major contributions towards addressing these problems. This continued a strong and productive tradition of statisticians working closely with the biological and medical sciences.

However, when I graduated I was not headed to a statistics or biostatistics department, but instead to a statistics group within the Marshall School of Business here at USC. For various reasons, the relationship between statistics and the business disciplines has historically been more strained. At that time, while elsewhere the field was flourishing, statistics departments within business schools were closing or merging with other departments, and the number of statistics faculty was on the decline.

Ten years ago represented the lowest point for statistics within Marshall and likely business schools in general. Several statistics faculty retired and there was a serious discussion about simply closing the group. Fortunately, our dean made a farsighted decision and elected to rebuild statistics within Marshall. Although at the time not an obvious call, in retrospect this turned out to be fantastic strategic decision. We have recruited extremely strong faculty almost every year since, started a successful Ph.D. program, and increased the number of students we teach in our electives by several hundred percent. Our department, which used to comfortably fit on one floor, is now overflowing on two floors, and the statistics group has become one of the largest, and strongest, business groups in the country. Marshall is now a very friendly environment to be a statistician.

This expansion was of course fueled by the second wave of new data, this time from industry and government. Whether it was Netflix offering one million dollars to build a better movie recommendation system, Amazon or Facebook performing A/B testing on millions of customers, Uber trying to set prices for both drivers and riders in a complicated two-sided dynamic market, Waze optimizing their traffic recommendations, Target predicting which customers were pregnant based

on purchase history or countless other examples, many companies suddenly had an urgent need for statistical and related expertise. Businesses soon discovered that their survival depended on their ability to efficiently process these vast quantities of data. This in turn drove employers to demand stronger quantitative skills in the students they hired. A decade ago I launched a new MBA elective class, from which my "An Introduction to Statistical Learning" book eventually evolved. My colleagues roundly predicted that the class enrollment would provide a real world operationalization of the "measure zero" concept and worked to make the prophecy self-fulfilling by scheduling it at 8 am, but it ended up being well attended. At the time several students reported obtaining internships, and even job offers, because they represented a tiny minority of MBA students who could confidently discuss concepts such as Boosting or Support Vector Machines during an interview.

Fast forward a decade and the pendulum has swung in the opposite direction. Now, in many industries, it has become a necessary condition for our students to be able to converse fluently on such topics. As a result, despite a vast increase in our elective options, students are struggling to get into the statistics and related classes that they need. At the same time there is strong demand for our Ph.D. students, partly driven by the need for statisticians to teach all these new classes. In addition, more business school faculty are recognizing that new statistical methodologies are advantageous, and in many cases essential, to the analysis of these new types of data.

Working as a statistician within a business school is a perfect match to my interests. I am proud to represent both areas but there are also some obvious tensions that I encounter on a regular basis. I believe that statistical applications in the biological and medical sciences are extremely important. However, that topic has, and will continue to be, well covered. Hence, my goal here is to highlight some of the challenges and opportunities that statistics, and statisticians, face within business and business schools. In the following sections I discuss implications for research as well as for Ph.D., masters and undergraduate programs.

2. Research

Likely, the biggest challenge in developing a harmonious relationship between statistics and business schools revolves around the impact of statistical research on business applications. There appears to be clear agreement that business students need an exposure to statistical concepts, a sentiment which has only grown with the recent data analytics revolution. On the research front a number of my Marshall colleagues, especially in the Marketing area, have been very open to exploring new statistical ideas. However, for a variety of reasons, many business faculty, while implementing classical statistical methods in their own research, see little benefit in adopting more modern and innovative statistical approaches. The issue is prevalent and a key cause for many of the research tensions. Statisticians often also contribute to the problem by displaying little interest in educating their colleagues about the potential impact of these new methodologies.

In an attempt to address this issue, a colleague of mine and I are in the process of organizing a workshop titled "Business Research Applications Needing Data Science" (BRANDS). The emphasis of the workshop will be educating participants on the connections among analytics, data science and business research problems, with the goal of stimulating future collaborations. Hence, BRANDS will bring together researchers with expertise in data science, faculty working on important business research problems, and external industry leaders with the goal of generating cross-fertilization of knowledge across the areas. Providing faculty with exposure to real world problems and needs from industry has the potential to greatly inspire innovative research ideas and developments that will further advance the frontiers in both academia and industry.

I believe that business schools could benefit from exploring the successful "dry labs" model favored by biostatisticians, consisting of a senior faculty member working with junior faculty, postdocs, and Ph.D. students. A key goal of the BRANDS workshop is that it will ultimately lead to a BRANDS lab, where statistical faculty will interact with general business faculty and industry leaders in a cooperative and interactive environment. We believe that such a lab has the potential to help bridge the gap between statistics and general business research.

Another issue faced by business schools is that many of them are relatively isolated within their universities. Statisticians provide an opportunity to increase interactions across schools. In particular, the fields of statistics and computer science have grown closer over the last few decades as computational methods have become the norm within statistics. Several of my colleagues have joint research projects, and grants, with computer science faculty, which opens a conduit for interactions between the business and engineering schools.

Finally, perhaps somewhat surprisingly, business schools themselves often struggle to form productive connections with industry and, as a result, are concerned about the relevance of their research. Here too statisticians can help. In the last couple of years several faculty from my department have taken semester long sabbaticals within tech companies and returned to Marshall, not only with strong connections to these organizations, but also a much clearer understanding of the main issues they face, which in turn positively influences their research and teaching. Companies appear to understand the potential value of these interactions and are actively seeking them out. For instance, Uber runs symposiums bringing together academics from economics, statistics, computer science and related fields to meet with their data scientists to discuss potential joint research topics. Such events provide a perfect opportunity to bring relevance, as well as rigor, into our research. Despite some cultural differences, statistics faculty provide significant opportunities for new and innovative business research, provided successful mechanisms can be created to ensure productive interactions.