

JANUARY 26, 2012, 10:30 AM

What Are the Odds That Stats Would Be This Popular?

By **QUENTIN HARDY**

“Most of my life I went to parties and heard a little groan when people heard what I did,” says Robert Tibshirani, a statistics professor at Stanford University. “Now they’re all excited to meet me.”

It’s not because of a new after-shave. Arcane statistical analysis, the business of making sense of our growing data mountains, has become high tech’s hottest calling. There are billions of bytes generated daily, not just from the Internet but also from sciences like genetics and astronomy. Companies like Google and Facebook, as well as product marketers, risk analysts, spies, natural philosophers and gamblers are all scouring the info, desperate to find a new angle on what makes us and the world tick. Computing has become cheap and available enough to process any number of formulas.

What no one has are enough people to figure out the valuable patterns that lie inside the data.

At North Carolina State, an advanced analytics program lasting 10 months has, since its founding in 2006, placed over 90 percent of its students annually. The average graduate’s starting salary for an entry-level job is \$73,000. Its current class of 40 students had 185 applicants, and next year’s applications are already twice that. In 2009, Harvard awarded four undergraduate degrees in statistics. Two graduates went into finance, one to political polling and one became a substitute teacher. There were nine graduates in 2010, 13 last year. They headed into Google, biosciences and Wall Street, as well as Stanford’s literature department. Globally, LinkedIn recently found that from 2009 to 2011 the new jobs with titles related to “analytics” and “data science” grew by 53 percent.

Stanford’s Department of Statistics, both renowned and near so many Internet and bioscience companies, is at the center of the boom. It received 800 résumés for next year’s 60 graduate positions, twice the number of applications it had three years ago. Graduates head to business school at a starting salary of \$150,000 or more, or to Facebook for about \$130,000.

“We won’t look at many people without a 3.8 average, and for the advanced program, an advanced math graduate exam in the top 10 percent,” says Trevor Hastie, the professor overseeing admissions. Statistics, he notes, dictate that many in the top 10 percent have placed equally, with perfect scores.

Mr. Hastie and Mr. Tibshirani are the authors of “The Elements of Statistical Learning,” a top textbook in the field. Their star course, called “Modern Applied Statistics: Learning,” started a decade ago with 30 students. Its current enrollment just closed off at 190. “We try to give them long and difficult homework assignments,” Mr. Hastie says. “Nothing works.”

(F.Y.I., most of us would fail the course. This week's homework involves building a sequence of nearest neighbor classifiers around its tuning parameter K , choosing K using prediction accuracy estimated through tenfold cross validation. In other words, how do you pick the best way to estimate whether an e-mail is spam?)

The two men also teach a two-day course for businesspeople called "Statistical Learning and Data Mining" that costs \$1,450 and attracts a broad range of data-laden people. "We had two guys from Hong Kong who taught a course in horse race prediction," Mr. Tibshirani says. "One of them came back and told us they're making \$10 million a year by modeling the last-minute betting."

Both men also consult for a number of well-known and emerging companies, but they worry about how many of their students will be drawn away from academics to the private sector. The boom in scientific data may be greatest in academia, particularly genetics.

"The data are going to be very wide, with lots of genes, and deep, with billions of variations, once we're able to decode individual human genes for \$1,000," Mr. Tibshirani says. "It will call for new statistical techniques, and new kinds of computation." His own recent work, on the ability to spot false discoveries, is aimed at the genetics problem. Another major life's work, his son, just graduated from Stanford and is teaching statistics at Carnegie Mellon.

About half of the Stanford stat professors have joint appointments with other departments, including economics, human biology and environmental science. "Statistics is unusual," Mr. Hastie notes. "It's a service field to other disciplines. It doesn't rely on its own work. It needs others."