
2009

Review of *Super Crunchers* by Ian Ayers

Eric Gaze

Alfred University, gaze@alfred.edu

Follow this and additional works at: <https://digitalcommons.usf.edu/numeracy>



Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Gaze, Eric. "Review of *Super Crunchers* by Ian Ayers." *Numeracy* 2, Iss. 2 (2009): Article 8. DOI: <http://dx.doi.org/10.5038/1936-4660.2.2.8>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

Review of *Super Crunchers* by Ian Ayers

Abstract

Ayers, I. *Super Crunchers: Why Thinking-by-Numbers Is the New Way to be Smart*. (Bantam Dell Publishing Group, 2007). 272 pp. Hard cover \$25 ISBN 978-0-553-80540-6.

Super Crunchers tells the story of how analyzing data is changing the ways in which decisions are made. We in the National Numeracy Network make a case for the importance of quantitative literacy by referring to how much quantitative information is now available to each of us: “a world awash in numbers.” Ian Ayres zeroes in on the people who are making a living crunching all of these data. From the seemingly innocuous (how wines are rated, and the scouting of baseball players) to the life impacting (diagnosis of disease, and parole of inmates), this book paints a vivid portrayal of how data analysis is impacting decision making at every level in our society. The use of simple regression models and randomized trials is calling into question who the “experts” of the twenty-first century will be, and why thinking-by-numbers really is the new way to be smart.

Keywords

data-analysis, regression, random trials

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Introduction

Super Crunchers begins with two compelling examples. Orley Ashenfelter set out to determine what factors create a great wine, and Bill James wanted to isolate the key attributes of a great baseball player. Ashenfelter's regression equation:

$$\text{Wine Quality} = 12.145 + 0.00117 \text{ winter rainfall} + 0.0614 \text{ average growing season temperature} - 0.00368 \text{ harvest rainfall},$$

seems almost comical with its combination of simplistic inputs—rain and temperature—and ultra-precise numerical coefficients to five decimal places. Yet this equation turns out to be a better predictor of wine quality than the experts! James created entirely new statistics to measure the worth of a baseball player:

$$\text{Runs Created} = (\text{Hits} + \text{Walks}) \times \text{Total Bases} / (\text{At Bats} + \text{Walks}).$$

Both men arrived at their formulas not by walking in the vineyards or sitting in the bleachers, but by sitting in their offices crunching data. They were not influenced by what name was on the label of the wine bottle or what a player looked like. They cared only about the cold hard facts captured in the reams of data that were entered into their computers and processed as trillions of zeroes and ones (binary digits or “bits”). And both men so outperformed the traditional experts that neither the wine nor the baseball industries could ignore them.

Corporate Data Mining

Ian Ayres confronts a central tension throughout his book that is highlighted by the two introductory examples: super crunchers versus traditional experts. He says (p. 12):

Super crunchers are not just invading and displacing traditional experts; they're changing our lives. They're not just changing the way decisions are made; they're changing the decisions themselves.

He begins with exploring the ways in which corporations are utilizing the data available to them to change the way they interact with their customers. Airlines and online booksellers are now tracking individual customers and making recommendations based on individual preferences. Casinos and auction sites are computing your price points to determine just how much money they can get out of you. The Internet is a critical factor here: surfing the Web and using credit cards do not allow us to shop anonymously; we leave in our wake a sea of data,

flotsam that is trolled continuously by companies with their own best interest in mind.

How much data are we talking about? Google is said to have 5 petabytes of storage capacity. A petabyte is 1,000 terabytes or 1 quadrillion bits (1 kb = 1,000 bits), so Google can crunch up to 5,000 terabytes of data! To put this in perspective, the entire Library of Congress is only about 100 terabytes as of May 2009. Organizing all this information has led to the creation of entirely new companies called database aggregators. ChoicePoint was founded in 1997, and now generates over \$1 billion a year in revenue from packaging and selling data. Axiom is even bigger, managing over 20 billion customer records (about 850 terabytes of raw data) culled from public census records, tax data, and corporate information supplied by its clients. Axiom knows more about you than your mother does and is constantly assigning every person to one of its 70 “lifestyle” segments. Corporations are willing to pay handsomely for such detailed information about your habits.

Consumer Data Mining

There are also many companies that are crunching to give consumers an edge. Oren Etzioni’s Farecaster.com will provide you with the best buys for airline tickets and even predict whether the price will be going up or down. You can now shop for the best insurance rates and find out exactly what you should be paying for your next car. Zillow.com crunches over 67 million home prices to aid both buyers and sellers. These datasets tend to use the regression techniques similar to Orley’s wine quality approach. They crunch the data to determine the weightings of causal factors involved in home prices or insurance rates. This is even being applied to evidence-based medicine, with doctors now being able to Google symptoms to diagnose illnesses. In 2004 Don Berwick announced his “100,000 lives Campaign.” He claimed that if hospitals implemented six changes in care he identified as being preventable errors, 100,000 lives could be saved in the next 18 months. On June 14, 2006, the campaign announced an estimated 122, 342 deaths had been prevented.

Random Trials

Don Berwick’s campaign highlights another important statistical technique central to the super crunching revolution, the random trial. Regression techniques date back to James Galton in 1877. The randomized trial was not introduced until 1925 by Ronald Fisher, the father of modern statistics. This is such a simple idea that it is amazing it took so long for anyone to try it. If you want to determine the effect of using leeches to cure illness, for example, you simply randomly assign

one group of sick people to get bled by leeches and the other group to act as a control. The randomness of the assignments is what controls for all other differences between the two groups, allowing you to measure the effect of the leeches. It is important to note that we are able to measure an effect with no a priori reason why that effect happens. This is probably what prevented this technique from arising sooner. Aristotelian science dictates we first understand the cause and then develop an intervention to treat that cause. Random trials could care less about causal factors; they are interested only in results, the effect! In fact, when Ignaz Semmelweis first suggested in 1840 that doctors wash their hands before entering the maternity ward at Vienna General Hospital (especially after leaving the autopsy room) he was ridiculed because he did not have a good reason why. Even when he presented data showing a reduction in childbirth fever death rates from 12% to 2% he still could not convince doctors to wash their hands. Old habits die hard, and intuitive expertise is not always correct.

With today's computing power, anyone can run random trials. Companies like Offermatica.com allow you to test different Web page versions in real time. Google's Adwords program let Ian Ayres test whether "The End of Intuition" was a better title for his book. He found that "Super Crunchers" generated 63% more click-throughs on his ads. The World Health Organization now runs randomized trials around the globe studying the efficacy of various health programs. Randomized trials in Africa have conclusively shown that circumcision drastically lowers the chance of AIDS infection. The Oportunidades program in Mexico has shown that paying mothers to attend health screenings and keep their kids in school has a significant impact on children's health and schooling. India was able to study the effects of having female tribal chiefs. Of course, corporations can also utilize randomized trials to find out how best to get our business. Capital One has led the way in these experiments to discover that a 4.9% rate for 6 months is better received than 7.9% for 12 months. Credit Indemnity made the disturbing discovery that a picture of a smiling woman on the envelope has the same effect as a 4.5 percentage-point drop in interest rate!

Conclusion

Decision making is a complicated process. Ellen Peters in her article, "Numeracy and the Perception and Communication of Risk" (Peters 2008), talks about the dual process modes in decision making. We process information in both a deliberative mode and an experiential mode (p. 2):

The deliberative mode is conscious, analytical, reason-based, verbal and relatively slow. Processing in the experiential mode, on the other hand, is intuitive, automatic, associative and fast.

Ian Ayres in his book, *Super Crunchers*, is showing that super crunching is the deliberative mode on steroids, and that traditional experts all too often rely on the experiential mode (Ayers, p. 62):

One thing that we have seen over and over is that decision makers overestimate the powers of their own intuition.

It is fun to snicker at wine snobs and baseball scouts being knocked down a peg or two, but it is entirely different when someone says a computer program is going to start making our own decisions for us. It is this tension that Ian Ayres is balancing and which he resolves by saying (p. 18):

This book will not try to bury intuition or experiential expertise as norms of decision making, but will show how intuition and experience are evolving to interact with data-based decision making.

Super Crunchers is an important book highlighting how the playing field of decision making is changing before our very eyes. Everyone needs to know that information is being used to manipulate decisions at all levels in our personal and civic lives. Ian Ayres does a great job making this a fun, entertaining book to read while also providing a detailed analysis of how super crunchers go about their business. I highly recommend *Super Crunchers* to everyone!

Reference

Peters, Ellen. 2008. Numeracy and the perception and communication of risk. In *Strategies for risk communication: Evolution, evidence, experience*, ed. W. T. Tucker, S. Ferson, A. M. Finkel, and D. Slavin. *Annals of the New York Academy of Sciences*, 1128, p. 1–7. New York: The New York Academy of Sciences.