UNIVERSITY OF TORONTO

DATA SCIENCES INSTITUTE

INTRODUCTION TO ANDREW GELMAN

ROHAN ALEXANDER, 17 SEPTEMBER 2021

ANDREW GELMAN

ACADEMIC APPOINTMENTS

- **Higgins Professor of Statistics, Columbia University.**
- **Professor, Department of Political Science, Columbia University.**

> variation. Randomness is one way to supply variation, and it's one way to model variation, but it's not necessary. Nor is it necessary to have "true" randomness (of the dice-throwing or urn-sampling variety) in order to have a useful probability model.

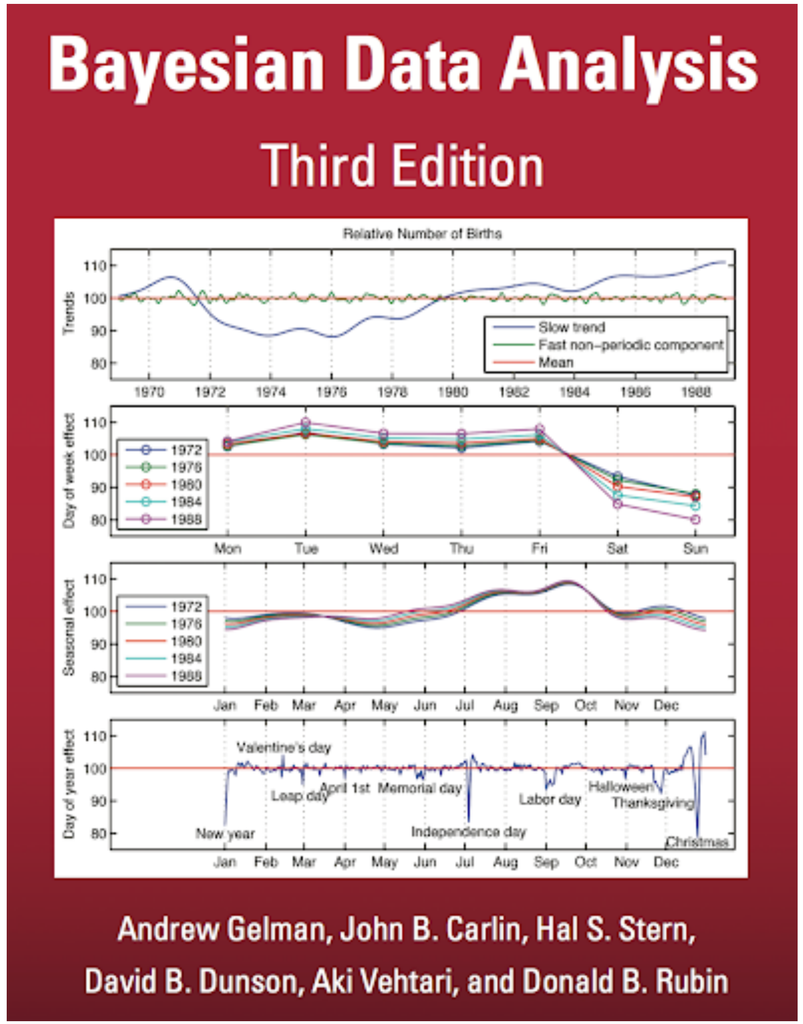
For my money, the #1 neglected topic in statistics is **measurement**.

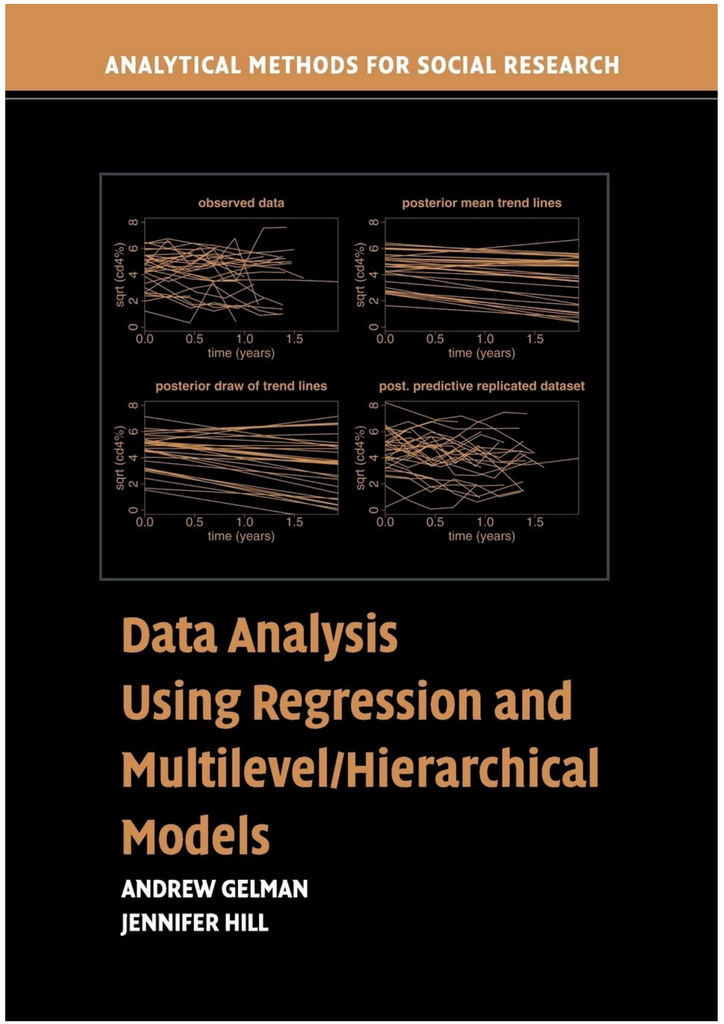
In most statistics texts that I've seen, there's a lot on data analysis and some stuff on data collection—sampling, random assignment, and so forth—but nothing at all on measurement. Nothing on reliability and validity but, even more than that, nothing on the *concept* of

https://statmodeling.stat.columbia.edu/2015/04/28/whats-important-thing-statistics-thats-not-textbooks/

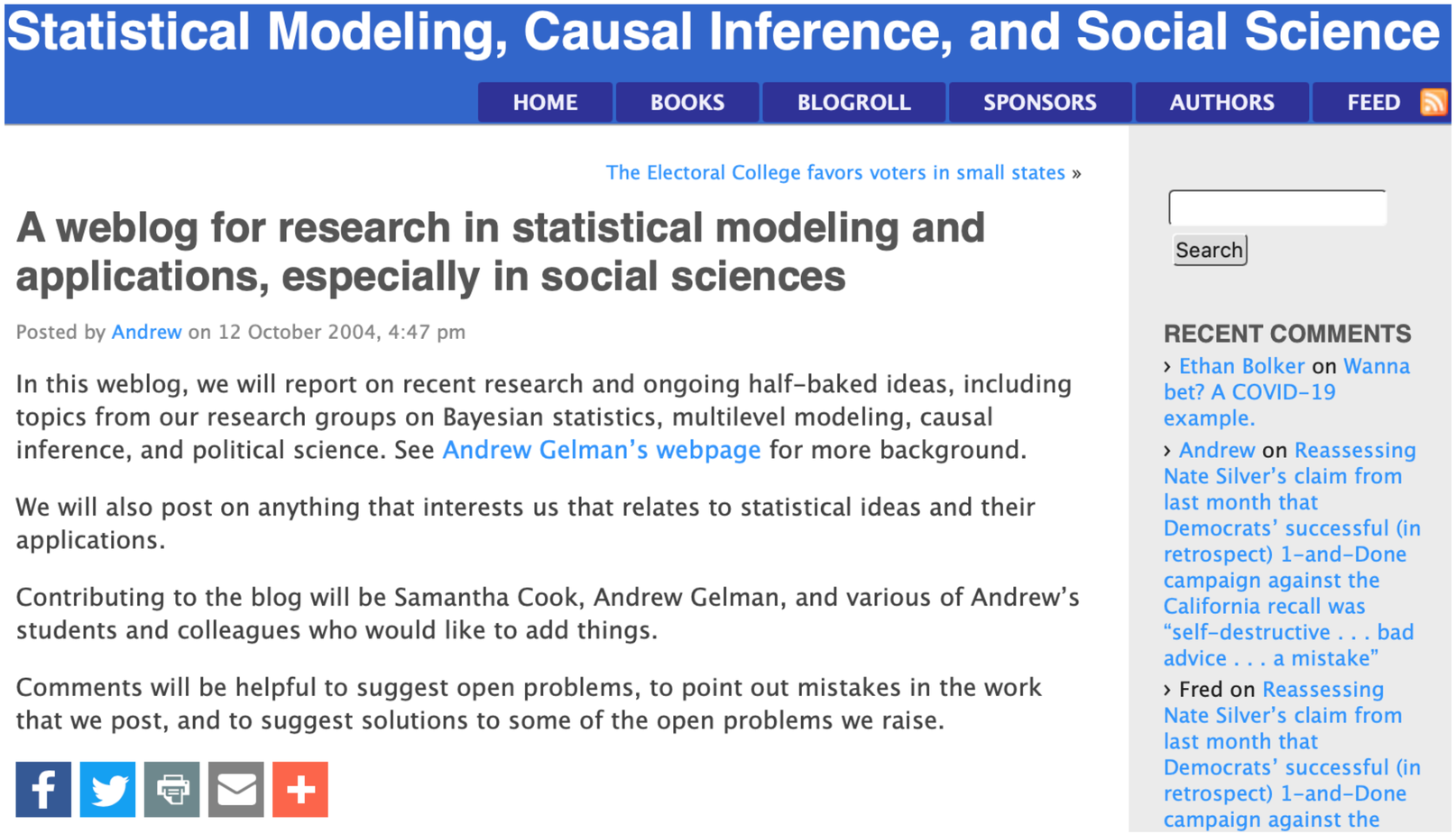
ANDREW H. WALKER // GETTY IMAGES

BOOKS


Bayesian Data Analysis
Third Edition
Relative Number of Births
Trends
Slow trend
Fast non−periodic component
Mean
Day of week effect
Seasonal effect
Day of year effect
Valentine's day
Leap day
April 1st
Memorial day
Labor day
Halloween
Thanksgiving
New year
Independence day
Christmas
Andrew Gelman, John B. Carlin, Hal S. Stern,
David B. Dunson, Aki Vehtari, and Donald B. Rubin


ANALYTICAL METHODS FOR SOCIAL RESEARCH
observed data
posterior mean trend lines
posterior draw of trend lines
post. predictive replicated dataset
sqrt (cd4%)
time (years)
Data Analysis
Using Regression and
Multilevel/Hierarchical
Models
ANDREW GELMAN
JENNIFER HILL

BLOG



Statistical Modeling, Causal Inference, and Social Science

HOME | BOOKS | BLOGROLL | SPONSORS | AUTHORS | FEED

The Electoral College favors voters in small states »

A weblog for research in statistical modeling and applications, especially in social sciences

Posted by Andrew on 12 October 2004, 4:47 pm

In this weblog, we will report on recent research and ongoing half-baked ideas, including topics from our research groups on Bayesian statistics, multilevel modeling, causal inference, and political science. See Andrew Gelman's webpage for more background.

We will also post on anything that interests us that relates to statistical ideas and their applications.

Contributing to the blog will be Samantha Cook, Andrew Gelman, and various of Andrew's students and colleagues who would like to add things.

Comments will be helpful to suggest open problems, to point out mistakes in the work that we post, and to suggest solutions to some of the open problems we raise.

Search

RECENT COMMENTS

› Ethan Bolker on Wanna bet? A COVID-19 example.

› Andrew on Reassessing Nate Silver's claim from last month that Democrats' successful (in retrospect) 1-and-Done campaign against the California recall was "self-destructive . . . bad advice . . . a mistake"

› Fred on Reassessing Nate Silver's claim from last month that Democrats' successful (in retrospect) 1-and-Done campaign against the

REPRODUCIBILITY

What has happened down here is the winds have changed

Posted by Andrew on 21 September 2016, 9:03 am

Someone sent me this article by psychology professor Susan Fiske, scheduled to appear in the APS Observer, a magazine of the Association for Psychological Science. The article made me a little bit sad, and I was inclined to just keep my response short and sweet, but then it seemed worth the trouble to give some context.

I'll first share the article with you, then give my take on what I see as the larger issues. The title and headings of this post allude to the fact that the replication crisis has redrawn the topography of science, especially in social psychology, and I can see that to people such as Fiske who'd adapted to the earlier lay of the land, these changes can feel catastrophic.

https://statmodeling.stat.columbia.edu/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/

First, second, and third order bias corrections (also, my ugly R code for the mortality-rate graphs!)

Posted by Andrew on 18 November 2015, 9:39 am

As an applied statistician, I don't do a lot of heavy math. I did prove a true theorem once (with the help of some collaborators), but that was nearly twenty years ago. Most of the time I walk along pretty familiar paths, just hoping that other people will do the mathematical work necessary for me to fit my models (for example, taking care of all the intricacies of implementing differential equation models in Stan, or developing the mathematical tools necessary to derive algorithms to sample from difficult distributions).

Every once in awhile, though, I'm reminded that a baseline level of mathematical

https://statmodeling.stat.columbia.edu/2015/11/18/first-second-and-third-order-bias-corrections-also-my-ugly-r-code-for-the-mortality-rate-graphs/

PAPERS

SOME RECENT HIGHLIGHTS

- Gelman et al, 2020, ‘Bayesian workflow’.
- Kennedy, Khanna, Simpson, Gelman, 2020, ‘Using sex and gender in survey adjustment’.
- Lerner and Gelman, 2020, ‘Build Your Own Statistics Course For Students In A Non-Quantitative Field’.
- Gelman and Carpenter, 2020, ‘Bayesian analysis of tests with unknown specificity and sensitivity’.
- Heidemanns, Gelman, and Morris, 2020, ‘An updated dynamic Bayesian forecasting model for the 2020 election’.
- ....

International Journal of Forecasting 31 (2015) 980–991

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting elections with non-representative polls

Wei Wang [a,*], David Rothschild [b], Sharad Goel [b], Andrew Gelman [a,c]

[a] Department of Statistics, Columbia University, New York, NY, USA
[b] Microsoft Research, New York, NY, USA
[c] Department of Political Science, Columbia University, New York, NY, USA

ARTICLE INFO

Keywords:
Non-representative polling
Multilevel regression and poststratification
Election forecasting

ABSTRACT

Election forecasts have traditionally been based on representative polls, in which randomly sampled individuals are asked who they intend to vote for. While representative polling has historically proven to be quite effective, it comes at considerable costs of time and money. Moreover, as response rates have declined over the past several decades, the statistical benefits of representative sampling have diminished. In this paper, we show that, with proper statistical adjustment, non-representative polls can be used to generate accurate election forecasts, and that this can often be achieved faster and at a lesser expense than traditional survey methods. We demonstrate this approach by creating forecasts from a novel and highly non-representative survey dataset: a series of daily voter intention polls for the 2012 presidential election conducted on the Xbox gaming platform. After adjusting the Xbox responses via multilevel regression and poststratification, we obtain estimates which are in line with the forecasts from leading poll analysts, which were based on aggregating hundreds of traditional polls conducted during the election cycle. We conclude by arguing that non-representative polling shows promise not only for election forecasting, but also for measuring public opinion on a broad range of social, economic and cultural issues.
© 2014 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

At the heart of modern opinion polling is representative sampling, built around the idea that every individual in a particular target population, such as registered or likely US voters, has the same probability of being sampled. From address-based, in-home interview sampling in the 1930s to random digit dialing after the growth of landlines and cellphones, leading polling organizations have put immense efforts into obtaining representative samples.

The wide-scale adoption of representative polling can be traced largely back to a pivotal polling mishap in the 1936 US presidential election campaign. During that campaign, the popular magazine *Literary Digest* conducted a mail-in survey that attracted over two million responses, a huge sample even by modern standards. However, the magazine incorrectly predicted a landslide victory for Republican candidate Alf Landon over the incumbent Franklin Roosevelt. In actual fact, Roosevelt won the election decisively, carrying every state except for Maine and Vermont. As pollsters and academics have pointed out since, the magazine's pool of respondents was highly biased: it consisted mostly of auto and telephone owners, as well as the magazine's own subscribers, which underrepresented Roosevelt's core constituencies (Squire, 1988). During that same campaign, various pioneering

* Corresponding author.
E-mail addresses: ww2243@columbia.edu (W. Wang), davidmr@microsoft.com (D. Rothschild), sharadg@microsoft.com (S. Goel), gelman@stat.columbia.edu (A. Gelman).

http://dx.doi.org/10.1016/j.ijforecast.2014.06.001
0169-2070/© 2014 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

STAN

- Stan is a probabilistic programming language.
- It is named after Stanislaw Ulam.
- Widely used:
  - >2K stars on GitHub.
  - >4K citations.
  - Large ecosystem has developed around it.

IMPACT ON OTHERS

"I LEARNT STATISTICS FROM HIS BOOKS, HIS XBOX PAPER GAVE MY CAREER PURPOSE, I COPY-PASTE HIS CODE MOST DAYS, HIS BLOG PROVIDED ME WITH A COMMUNITY WHEN I WAS INTELLECTUALLY ISOLATED, AND FROM HIS EXAMPLE I LEARNT THE TYPE OF ACADEMIC I WANTED TO BE."

ROHAN ALEXANDER

IMPACT ON OTHERS

"HIS BLOG MADE ME REALISE THAT THERE IS A PLACE FOR PEOPLE LIKE ME, THAT IS, PEOPLE WHO WANT TO DO GOOD APPLIED STATS ON INTERESTING SOCIAL PROBLEMS. THIS AND READING GELMAN HILL INFLUENCED MY DECISION TO GO TO GRAD SCHOOL."
MONICA ALEXANDER

ANDREW PROVIDED A VOICE THAT CUT THROUGH THE OFTEN CONFUSING AND RULE BASED INTERPRETATION OF STATISTICS TO INSTEAD PROVIDE AN INTERPRETATION THAT IS REASON BASED. THIS REPRESENTED A TURNING POINT FOR ME IN THE WAY I UNDERSTOOD EXISTING STATISTICAL PROBLEMS AND RATIONALISED ABOUT NEW CHALLENGES.
LAUREN KENNEDY

"HIS CLASS (AND WRITING) SHOWED ME THE CONNECTIONS BETWEEN STATISTICS, SCIENCE, AND PHILOSOPHY AND GAVE ME A LANGUAGE THROUGH WHICH TO EXPRESS MY FRUSTRATIONS WITH THE WAYS THAT DATA IS TORTURED. I PARTICULARLY APPRECIATED HIS EMPHASIS THAT "STATS IS HARD" AND IT'S OK TO MAKE MISTAKES AS LONG AS YOU LEARN FROM THEM!"
JAMES DOSS-GOLLIN

"WORKING IN GOVERNMENT AND POLITICS, I CHANNEL ANDY, BOTH AS A STATISTICIAN AND AS A TEAMMATE. I QUOTE INSIGHTS FROM HIS BOOKS, PAPERS, AND BLOG POSTS. I LEARN FROM HIS HUMOR, OPENNESS ABOUT BEING CONFUSED, AND COMMITMENT TO DOING THE RIGHT THING."
SHIRA MITCHELL

"ANDREW'S MY OBI-WAN-KENOBI."
BOB CARPENTER

"ANDREW'S BLOG MADE ME AWARE OF THE SHORTCOMINGS IN MY TRAINING AND RESEARCH, AND GAVE ME MOTIVATION TO IMPROVE. IT ALSO GAVE ME THE MOTIVATION TO GET OUT OF PROJECTS THAT COULD BECOME POSTS IN THE ZOMBIES CATEGORY."
MARTA KOŁCZYŃSKA

"HERE'S SOMETHING THAT I REALLY APPRECIATE ABOUT ANDREW AND ISN'T COMMON ENOUGH IN ACADEMIA: IF YOU DO GOOD WORK ANDREW WILL WANT TO WORK WITH YOU (EVEN HIRE YOU) REGARDLESS OF YOUR CREDENTIALS."
JONAH S GABRY

"MY CONCEPTION OF WHAT GOOD RESEARCH WAS - WHAT KIND OF RESEARCH I WANTED TO BE DOING - COMPLETELY CHANGED WHEN I READ ANDREW'S PAPER ON THE GARDEN OF FORKING PATHS. IT WAS ONE OF THE MAIN REASONS I DECIDED TO GO BACK TO SCHOOL FOR A DEGREE IN STATISTICS, AND IT'S DEEPLY INFLUENCED THE WAY I TRY TO DO RESEARCH AND THE RESEARCH PRACTICES I ADVOCATE FOR AT MY WORKPLACE. SINCE THAT PAPER CAME OUT IN 2013 I'VE BEEN A DEDICATED READER OF HIS BLOG, AND HIS POSTS HAVE MADE ME FUNDAMENTALLY RETHINK THE WAY I DO RESEARCH AN EMBARRASSING NUMBER OF TIMES - AND THEY CONTINUE TO MOTIVATE ME TO TRY TO DO CAREFUL AND THOUGHTFUL WORK."
ISAAC MADDOW-ZIMET