


What are the most important statistical ideas of the past 50 years?

Amina Menaa

 Email: amina.m24@outlook.com



What are the most important statistical ideas of the past 50 years?*

Andrew Gelman[†] and Aki Vehtari[‡]

3 June 2021

Abstract

We review the most important statistical ideas of the past half century, which we categorize as: counterfactual causal inference, bootstrapping and simulation-based inference, overparameterized models and regularization, Bayesian multilevel models, generic computation algorithms, adaptive decision analysis, robust inference, and exploratory data analysis. We discuss key contributions in these subfields, how they relate to modern computing and big data, and how they might be developed and extended in future decades. The goal of this article is to provoke thought and discussion regarding the larger themes of research in statistics and data science.

Contents of the presentation



01

**The most
important
statistical ideas of
the past 50 years**

02

**What these ideas
have in common
and how they differ**

03

**What will be the
important
statistical ideas of
the next few
decades**



01

The most important statistical ideas of the past 50 years



Counterfactual causal Inference

Idea: bridging the gap between naïve causal interpretation of observed data and the recognition that correlation does not simply imply causation

Bootstrapping and simulation-based inference

Idea: substitution of mathematical analysis with computational methods; considering prediction and resampling as fundamental principles from which we can derive statistical operations

Overparameterized models and regularization

Idea: getting the flexibility of a nonparametric/highly parameterized approach while avoiding the overfitting problem

Bayesian multilevel models

Idea: multilevel models have parameters varying by group allowing them to adapt to multiple structured settings

Generic computation algorithms

Idea: allowing decoupling of the development of the models so that changing the model did not require changes to the algorithm implementation to simplify process of applying statistical models to data

Adaptive decision analysis

Idea: the application of decision theory and statistical methods to make informed decisions in dynamic and changing environments

Robust inference

Idea: ability to use models even when they have assumptions that are not true

Exploratory data analysis

Idea: moving the field of statistics away from theorem-proving and toward a more open and healthier perspective of learning from the data

02

**What these ideas
have in common
and how they differ**



01

Ideas lead to methods and workflows

Each of the ideas was not so much a method for solving an existing problem but a new way of thinking about statistics

02

Advances in computing

Statisticians no longer limited to simple models with analytic solutions and simple closed-form algorithms

03

Big data

Facilitate the use of big data

04

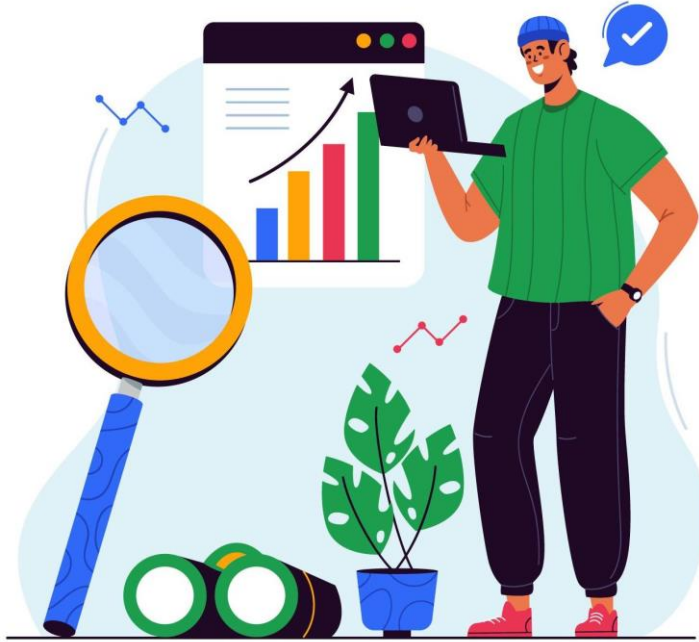
Connections and interactions among these ideas

Example: regularized overparameterized models can be optimized using machine learning meta-algorithms that yield inferences that are considered robust

05

Links to other new and useful developments in statistics

Emphasizes the interconnections between various statistical models, methods, applications, and principles.



03

**What will be the
important statistical
ideas of the next few
decades**



Looking backward

- In considering the most important statistical ideas of the past 50 years, it would also make sense to reflect upon the most important statistical ideas of the previous centuries.
- The point of asking what are the most important statistical ideas is not so much to answer the question but to stimulate discussion of what it means for a statistical idea to be important.

Looking forward



01

What will come next?

- Progress on existing combinations of methods
- Progress on experimental design and sampling
- Advances in computation

02

What are the biggest challenges and opportunities facing statisticians?

- Big data
- Messy data
- Complicated questions

03

Can we anticipate what new areas might arise, about which statisticians should become aware?

Statistics should continue to be open to Ideas, general theoretical frameworks as well as specific models and methods, coming from other fields.

Any questions?