



**Samuel Neaman Institute**

for National Policy Research

The Israel Statistics Association  
and Data Science Association  
الرابطة الإحصائية  
للإحصاء وعلم البيانات  
האיגוד הישראלי  
לסטטיסטיקה  
ומדע הנתונים

# On the foundations of applied statistics

A hybrid ideas exchange forum via zoom

11/4/2024

10:00-17:00



**Prof Edna Schechtman  
1948-2022**

Dedication

# Background

# what is the foundation of applied statistics



ChatGPT

The foundation of applied statistics lies in the principles of statistical theory and methodology, which are applied to real-world problems and data sets across various fields. Here are some key components of the foundation of applied statistics:

**Probability Theory:** Probability theory provides the mathematical framework for quantifying uncertainty and randomness. It underlies many statistical methods and concepts, such as probability distributions, random variables, and sampling theory.

**Descriptive Statistics:** Descriptive statistics involves summarizing and presenting data in a meaningful way, often through measures such as means, medians, standard deviations, and graphical representations like histograms and scatter plots.

**Inferential Statistics:** Inferential statistics allows researchers to draw conclusions or make inferences about a population based on sample data. It includes methods such as hypothesis testing, confidence intervals, and regression analysis.

**Experimental Design:** Experimental design focuses on designing studies and experiments to efficiently collect data and draw valid conclusions. It involves considerations such as randomization, replication, and control of experimental variables.

**Regression Analysis:** Regression analysis explores the relationship between variables, typically with the goal of predicting or explaining one variable based on others. Linear regression is a fundamental technique, but there are also more complex methods like logistic regression and time series analysis.

**Multivariate Analysis:** Multivariate analysis deals with datasets containing multiple variables, examining patterns, dependencies, and relationships among them. Techniques include principal component analysis, factor analysis, and cluster analysis.

**Statistical Computing and Software:** Applied statisticians often utilize statistical software packages for data analysis, visualization, and modeling. Proficiency in programming languages like R, Python, and SAS is essential for conducting statistical analyses efficiently.

**Ethics and Reproducibility:** Ethical considerations are crucial in applied statistics, including issues related to data privacy, bias, and transparency. Reproducibility, ensuring that analyses can be independently verified and replicated, is also a fundamental principle.

Overall, the foundation of applied statistics encompasses a diverse set of principles, methods, and practices that enable statisticians to analyze data effectively and derive meaningful insights to inform decision-making in various fields.



# The importance of foundations

---

Nancy Reid  
University of Toronto

June 8 2023



Sir David Cox (1924-2022)

## ... What are the foundations of statistics?

- probability, analysis, applied mathematics modelling
- Bayes, Neyman, Fisher approaches to inference
- nature of uncertainty epistemic, empirical
- nature of induction belief functions, inferential models
- interpretation of  $p$ -values, confidence regions, credibility intervals, likelihood ratios
- role of sufficiency, ancillarity, conditioning, asymptotic theory
- sparsity, causality, assumption-free/lean inference, stability, prediction, decisions



WILEY SERIES IN PROBABILITY AND STATISTICS

# Data Analysis

What Can Be Learned From  
the Past 50 Years



Peter J. Huber

WILEY

But still, the best way to convey the principles, rather than the mere techniques of data analysis, and to prepare the general mental framework, appears to be through anecdotes and case studies, and I shall try to walk this way. There are more than enough textbooks and articles explaining specific statistical techniques. There are not enough texts concerned with issues of overall strategy and tactics, with pitfalls, and with statistical methods (mostly graphical) geared toward providing insight rather than quantifiable results. So I shall concentrate on those, to the detriment of the coverage of specific techniques. My principal aim is to distill the most important lessons I have learned from half a century of involvement with data analysis, in the hope to lay the groundwork for a future theory. Indeed, originally I had been tempted to give this book the ambitious programmatic title: *Prolegomena to the Theory and Practice of Data Analysis*.

*Prolegomena to Any Future Metaphysics That Will Be Able to Present Itself as a Science* ([German](#): *Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können*) is a book by the German philosopher [Immanuel Kant](#), published in 1783, two years after the first edition of his [Critique of Pure Reason](#)





# A Note on the Theory of Applied Statistics

9 Pages • Posted: 6 Nov 2012 • Last revised: 24 Aug 2018

[Ron S. Kenett](#)

KPA Ltd.; The Samuel Neaman Institute, Technion; University of Turin and College Carlo Alberto

Date Written: November 5, 2012

## Abstract

This note is a sketch of what could be the basis for a theory of applied statistics. Such a theory is needed to help statistics become more relevant, with significant impact and innovative developments. To achieve this goal, statisticians need to get involved in new activities within the research, application and pedagogical areas. In particular social networks and new data structures require new statistical models. Eliciting the components of a theory of applied statistics is a prerequisite to a review of the training and education of statisticians who want to have an impact on the organizations they work in and society in general.

**Keywords:** applied statistics, practical statistical efficiency (PSE), information quality (InfoQ)

**JEL Classification:** A10, A20, B40, C10, C42, C44, C90, M10

**Suggested Citation:**



[https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID3236539\\_code1304617.pdf?abstractid=2171179&mirid=1](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3236539_code1304617.pdf?abstractid=2171179&mirid=1)

### **A Note on the Theory of Applied Statistics**

Version 18

Ron S. Kenett  
ron@kpa-group.com

*"Much fine work in statistics involves minimal mathematics; some bad work in statistics gets by because of its apparent mathematical content."*  
David Cox (1981). Theory and general principle in statistics, JRSS(A), 144, pp. 289-297.

Kenett, Ron S., A Note on the Theory of Applied Statistics (November 5, 2012). Available at SSRN:  
<https://ssrn.com/abstract=2171179> or <http://dx.doi.org/10.2139/ssrn.2171179>

# Data analytic understanding of statistics

Christian Hennig<sup>1</sup>

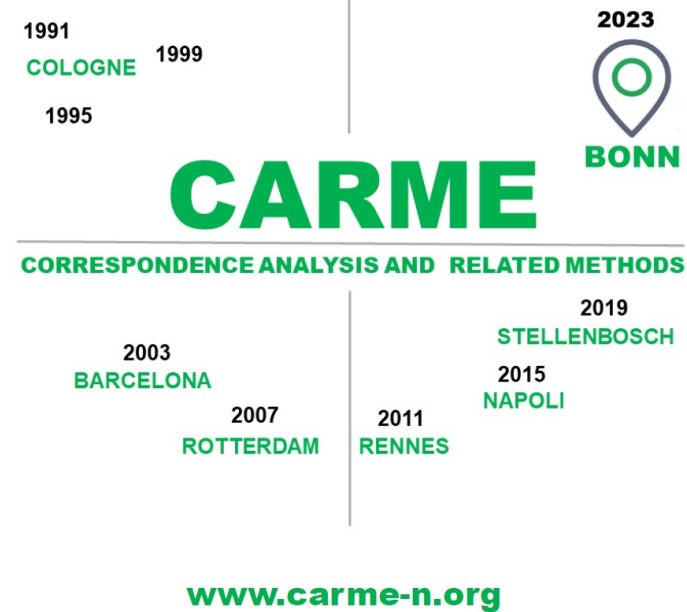
<sup>1</sup>University of Bologna, Italy; christian.hennig@unibo.it

The title refers to the pursuit of understanding how the methods of statistics and data analysis process and transform the data, and what this implies. Statistics is predominantly taught and interpreted based on probability models, but, as pioneers of data analysis such as Tukey, Gower, and Benzécri knew, there is far more to the understanding of statistical methods than their performance assuming certain models.

This includes some rather elementary considerations with wide-ranging consequences that are rarely taught or discussed, such as the different effects of different standardization techniques when aggregating variables, or what a hypothesis test actually does in case the assumed model is not true (pretty much always, that is).

My presentation will make an appeal to look at our methods from the angle of a direct interpretation of what they do to the data. I will also discuss how this is philosophically different from setting up probability models and trying to infer "truths" assuming them, and what the role of models can be for the data analytic understanding of statistics if not assumed to be true.

*Keywords:* foundations of statistics, data science, probability models, robustness, data visualization





Advanced

<https://ijhpr.biomedcentral.com/articles/10.1186/s13584-022-00531-y>

Save

Email

› Isr J Health Policy Res. 2022 Apr 20;11(1):22. doi: 10.1186/s13584-022-00531-y.

# The role of statisticians in the response to COVID-19 in Israel: a holistic point of view

Itai Dattner <sup>1</sup>, Reuven Gal <sup>2</sup>, Yair Goldberg <sup>3</sup>, Inbal Goldshtein <sup>4</sup>, Amit Huppert <sup>5</sup>,  
Ron S Kenett <sup>2 6 7</sup>, Orly Manor <sup>8</sup>, Danny Pfeffermann <sup>8 9 10</sup>, Edna Schechtman <sup>11</sup>,  
Clelia di Serio <sup>12</sup>, David M Steinberg <sup>13</sup>

# Helping reviewers assess statistical analysis: A case study from analytic methods

Ron S. Kenett✉, Bernard G. Francq

First published: 16 June 2022 | <https://doi.org/10.1002/ansa.202000159>

☰ SECTIONS



PDF



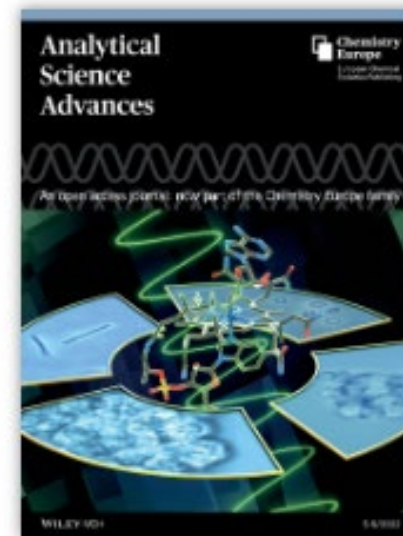
TOOLS



SHARE

## Abstract

Analytic methods development, like many other disciplines, relies on experimentation and data analysis. Determining the contribution of a paper or report on a study incorporating data analysis is typically left to the reviewer's experience and good sense, without reliance on structured guidelines. This is amplified by the growing role of machine learning driven analysis, where results are based on computer intensive algorithm applications. The evaluation of a predictive model where cross validation was used to fit its parameters adds challenges to the evaluation of regression models, where the estimates can be easily reproduced. This lack of structure to support reviews increases uncertainty and variability in reviews. In this paper, aspects of statistical assessment are considered. We provide checklists for reviewers of applied statistics work with a focus on analytic method development. The checklist covers six aspects relevant to a review of statistical analysis, namely: (1) study design, (2) algorithmic and inferential methods in frequentism analysis, (3) Bayesian methods in Bayesian analysis (if relevant), (4) selective inference aspects, (5) severe testing properties and (6) presentation of findings. We provide a brief overview of these elements providing references for a more elaborate treatment. The robustness analysis of an analytical method is used to illustrate how an improvement can be achieved in response to questions in the checklist. The paper is aimed at both engineers and seasoned researchers.

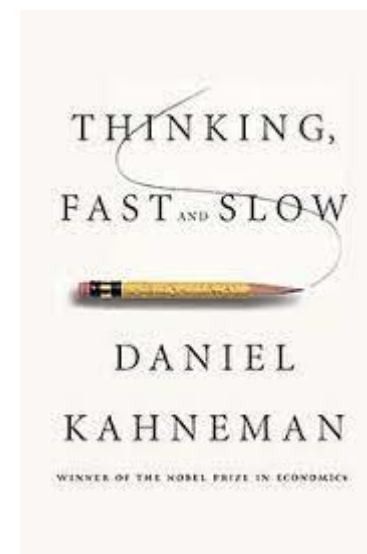


[Volume 3, Issue 5-6](#)

June 2022

Pages 212-222

<https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/ansa.202000159>



<https://fs.blog/before-you-make-that-big-decision/>



# Reviewing of Applied Research with an Industry 4.0 Perspective

11 Pages • Posted: 11 Jun 2020 • Last revised: 24 Jan 2022

Ron S. Kenett

KPA Ltd.; The Samuel Neaman Institute, Technion; University of Turin and College Carlo Alberto

Date Written: May 2, 2020

## Abstract

This note provides a perspective on questions to be used in reviewing applied research using analytics and statistical analysis. The paper emphasizes applied research related to Industry 4.0. It expands on a paper providing similar guidelines with a general focus on information quality (Kenett and Shmueli, 2016a). The note is methodological. It covers aspects of study design, algorithmic and inferential methods in frequentism analysis, Bayesian methods in Bayesian analysis, selective inference aspects, severe testing properties and presentation of findings. Information quality is based on good responses to questions about a specific report such as what is the goal of the analysis, is the data resolution adequate for the stated or implicit goal, how is data from different sources integrated, has a generalization claim been made, on what basis?

In reviewing studies done in Industry 4.0 topics, one finds data collected actively or passively, models developed with empirical methods, first principles or as hybrid models. Industry, as opposed to science, is less concerned with reproducibility of results, but it should. The industrial cycle provides short term opportunities to try out new products of new process set ups and, based on the results, determine follow up actions. Deriving misleading conclusions can be however very costly and time consuming.


# SSRN

[https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID4015629\\_code1304617.pdf?abstractid=3591808&mirid=1](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4015629_code1304617.pdf?abstractid=3591808&mirid=1)





## Generalizing research findings for enhanced reproducibility: an approach based on verbal alternative representations

Ron S. Kenett<sup>1,2,3</sup>  · Abraham Rubinstein<sup>1</sup>

Received: 22 August 2020 / Accepted: 9 February 2021  
© Akadémiai Kiadó, Budapest, Hungary 2021

### Abstract

Research aims at generating research claims. The paper introduces a "border of meaning", abbreviated BOM, as a mode of representation of research findings that supplements statistical tests. The suggested approach was originally developed in a pedagogical context of promoting conceptual understanding in education. Here we aim at helping better understand research claims stated in a scientific paper. Considering new approaches to the presentation of findings, has an impact on the reproducibility of research. The BOM approach is demonstrated using examples from clinical research and translational medicine. Specifically, we map research findings into a list that delineates a demarcation line between alternative representations of the research claims, some with meaning equivalence and some with surface similarity. Such a mapping can be statistically evaluated by sin type error tests. Our main message is that findings should be presented and generalized with a BOM.

**Keywords** Reproducible research · Replicability · Repeatability · Generalizability of findings · Information quality · Boundary of meaning · Study design · S-type errors



[https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID4009802\\_code1304617.pdf?abstractid=3035070&mirid=1](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4009802_code1304617.pdf?abstractid=3035070&mirid=1)

<https://link.springer.com/article/10.1007/s11192-021-03914-1>

[https://econpapers.repec.org/article/sprscient/v\\_3a126\\_3ay\\_3a2021\\_3ai\\_3a5\\_3ad\\_3a10.1007\\_5fs11192-021-03914-1.htm](https://econpapers.repec.org/article/sprscient/v_3a126_3ay_3a2021_3ai_3a5_3ad_3a10.1007_5fs11192-021-03914-1.htm)

# On representation of findings from clinical research

THE JOURNAL OF PEDIATRICS • www.jpeds.com

ORIGINAL  
ARTICLES

## A Structured Gradual Exposure Protocol to Baked and Heated Milk in the Treatment of Milk Allergy

Adi Efron, MD<sup>1</sup>, Yuri Zeldin, MD<sup>2,3</sup>, Leora Gotesdyner, MSc<sup>1</sup>, Tali Stauber, MD<sup>1,4,5</sup>, Ramit Maoz Segal, MD<sup>5</sup>, Inga Binson, MD<sup>3</sup>, Mira Dinkin, MD<sup>3</sup>, Larisa Dinkowitz, MD<sup>3</sup>, Danit Shahar, BSN<sup>4,5</sup>, Michal Deutch, BSN<sup>5</sup>, Mazal Yaron, BSN<sup>3</sup>, Ayelet Nevet, BSN<sup>3</sup>, Avner Reshef, MD<sup>5</sup>, Nancy Agmon-Levin, MD<sup>1,5</sup>, Ron S. Kenett, PhD<sup>6</sup>, and Mona I. Kidon, MD<sup>1,3,4,5</sup>

**Objective** To evaluate the efficacy and safety of a structured gradual exposure protocol (SGEP) with extensively heated and baked milk in promoting allergy resolution in children with cow milk allergy (CMA).

**Study design** In a case control study, children with CMA aged 1-4 years who were treated with SGEP including extensively heated and baked milk, were compared with children treated with strict avoidance. Data were collected from medical records and from validated telephone questionnaires. Data analysis was performed using a nonparametric Kaplan-Meier and proportional hazard Cox regression model, after evaluation of the adequacy of the case control matching.

**Results** There were 43 children with milk allergy—26 (62%) males with a mean age at intervention of 21 months (range, 12-47 months)—who were treated with SGEP and followed to a mean age of 40 months (range, 20-82 months). The median age at resolution of CMA was compared with a matched group of 67 children treated with strict avoidance at least until 4 years of age or followed until earlier resolution, with a mean age at follow-up of 71 months (range, 11-176 months). The median estimated age at CMA resolution in the SGEP group was 36 months (95% CI, 34.5-49.7) compared with 98 months (95% CI, 82.4-114.1) in controls ( $P < .001$ ). At last follow-up, 86% of treated children were tolerant to unheated milk proteins vs 52% of controls ( $P = .003$ ).

**Conclusion** A structured protocol with extensively heated and baked milk seems to promote faster resolution of CMA. (*J Pediatr* 2018;■■■:■■■-■■■).





**Table III. Boundary of meaning statements**

**BOM**

Target statement	Meaning equivalence findings	Surface similarity findings
Finding 1: The quality of life of patients and families affected with a food allergy to staple foods (milk, egg, sesame, peanut) is impaired	<p>Food allergy in children impacts negatively on the day-to-day activities of the whole family</p> <p>The incidence of accidental exposures to allergenic foods in preschool children is high</p> <p>The currently recommended management of food allergy in children is patient education, strict avoidance, and carrying an epinephrine autoinjector</p>	<p>Educating patients on strict avoidance and carrying an epinephrine autoinjector is completely effective in avoiding accidental exposures in preschool children</p>
Finding 2: All children suspected of an allergic reaction to foods should be referred to a center that includes appropriate facilities, medical, and support staff experienced in the diagnosis and treatment of children with food allergies as early as possible	<p>The diagnosis of food allergy in children should be performed soon after the suspected event</p> <p>There are no age limitations on the performance of diagnostic allergy tests, such as SPTs or observed food challenges, provided these are performed by well trained and experienced medical teams</p>	<p>Recommending strict avoidance of suspected allergenic foods is the best treatment for all young food allergic children</p> <p>Laboratory test such as sIgE to food can accurately diagnose food allergy in children</p>
Finding 3: The natural history of CM allergy in children is still favorable as in most—it seems to resolve with time	<p>The median age at resolution of CMA (by which time 50% of children have resolved their allergies) is between 6 and 8 years</p> <p>Children with CMA and a positive family history of atopy, an initial anaphylactic reaction, recurrent wheezing or moderate/severe atopic dermatitis are less likely to resolve their CMA</p>	<p>Food allergy in children resolves in the first years of life</p> <p>Avoidance of allergenic foods is beneficial in preventing food allergy in children</p>
Finding 4: A majority of children with IgE mediated CMA are capable of consuming certain amounts of EHBM proteins	<p>Some children with CMA can develop immediate, life-threatening reactions to the ingestion of EHBM</p> <p>A minority of children with CMA are allergic also to heat denatured milk products. These are the most severely affected and least likely to resolve their allergies</p>	<p>Families of children with IgE-mediated CMA should be encouraged to try baked milk at home</p> <p>All forms of heated and baked milk are similarly safe</p>
Finding 5: In preschool children with CMA capable of ingesting EHBM safely, SGEP seems to promote earlier resolution	<p>The median age at CMA resolution of preschool children, capable of ingesting EHBM safely and treated with SGEP including EHBM, seems to be significantly lower than in children treated with avoidance</p> <p>Most preschool children capable of ingesting EHBM safely and treated with SGEP including EHBM will be able to tolerate milk in their regular diet before entering school</p>	<p>Preschool children capable of ingesting EHBM safely and treated with SGEP including EHBM are developing true long-term tolerance to milk</p> <p>EHBM is not a form of oral immunotherapy in food allergic children and therefore the follow-up recommended for these children is similar to patients with natural resolution of CMA (none)</p>
Finding 6: A protocol of SGEP including EHBM, seems safe in children <4 years of age	<p>A protocol of SGEP, including EHBM, performed by medical teams trained and experienced in the treatment of food allergy in children is safe</p>	<p>All children with IgE-mediated CMA should be treated with an SGEP with EHBM</p>



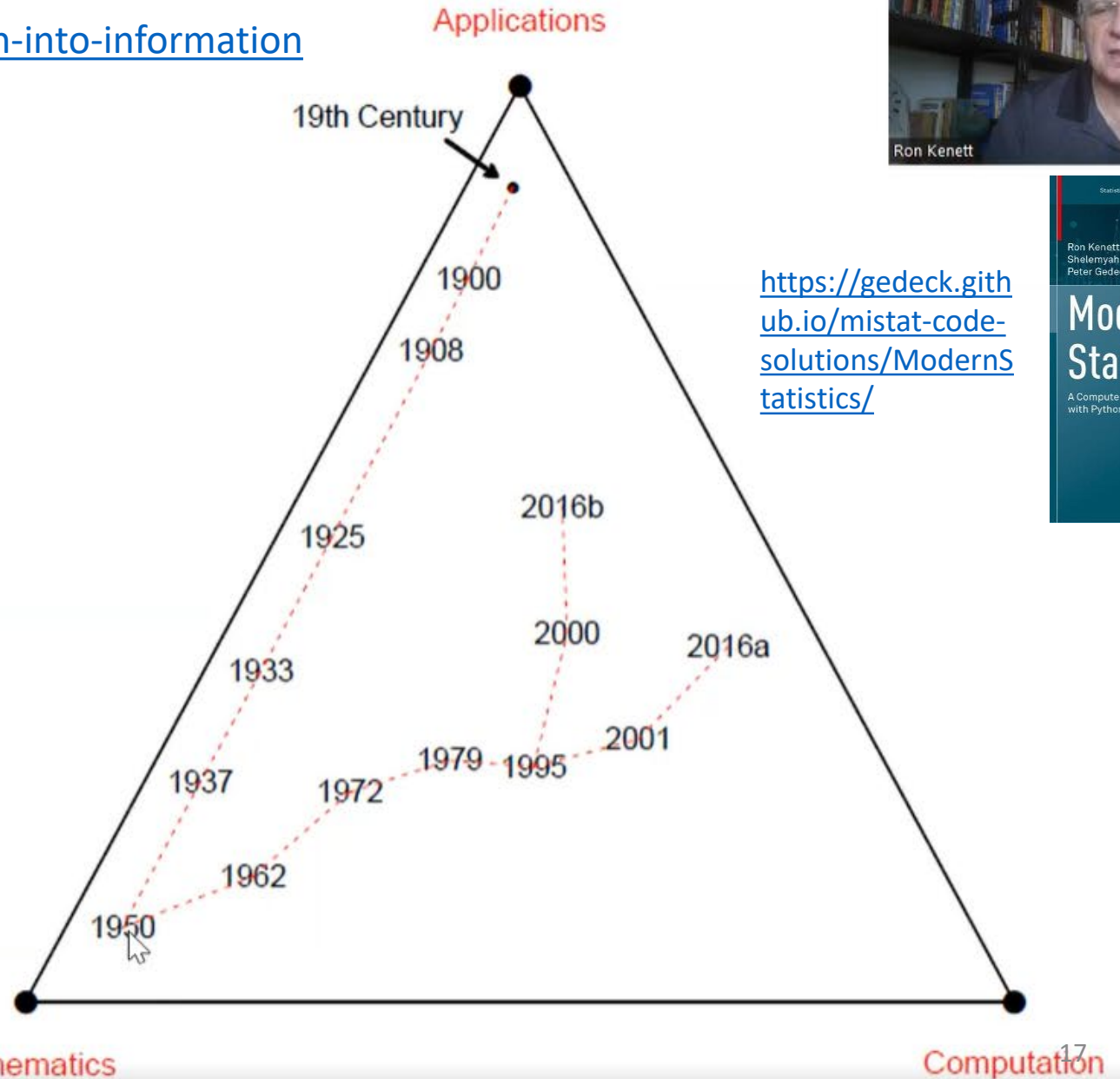
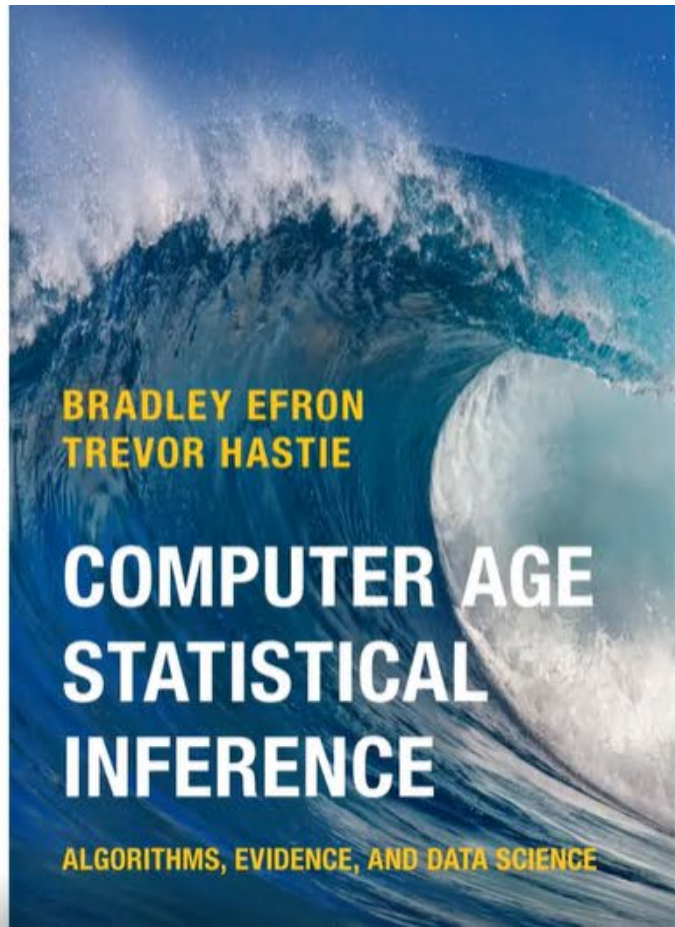
<https://user-images.githubusercontent.com/8720575/180794703-c6f05f40-eefd-4e1a-93f9-42cb78e6a6b4.mp4>

<https://shepherd.com/best-books/how-numbers-turn-into-information>

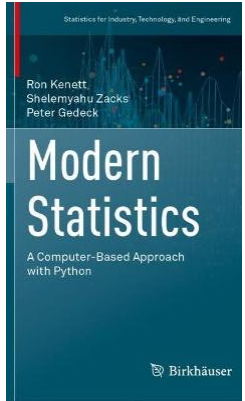
## The best books on how numbers turn into information: the challenge of analytics



By Ron S. Kenett  



<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>



Special issue editors:  
Ron Kenett and David Steinberg

<https://www.statisticsviews.com/call-for-papers-asmbi-special-issue-on-foundations-of-applied-statistics/>



STATS & DATA SCIENCE VIEWS

ARTICLES TOPICS NEWS EVENTS LATEST RESEARCH

SEARCH



# Call for Papers: ASMBI Special Issue on Foundations of Applied Statistics

13th February 2024

Applied Stochastic  
Models in Business  
and Industry

The official journal of  
the International Society for  
Business and Industrial Statistics



We are pleased to announce a “Foundations of Applied Statistics” Special Issue of the journal **Applied Stochastic Models in Business and Industry** (ASMBI) following a conference on this topic held on April 11th 2024 at the Samuel Neaman Institute, Technion, Israel, see <https://www.neaman.org.il/en/On-the-foundations-of-applied-statistics>

Statistics has evolved in the past three centuries with different emphases on problem driven methods, mathematical methods and computational analytics. It has witnessed heated discussions contrasting frequency based inference with Bayesian methods. Other schisms exist between predictive models and explanation of existing data, what Deming called analytic and enumerative studies. Significant contributors to this evolution include Leo Breiman, John Tukey and David Cox. All these authors wrote papers on the foundations of statistics as both a theoretical and an applied discipline. This special issue is about the foundations of applied statistics.

The aim of this Special Issue is to attract high-quality, innovative, and original works related but not

Deadline for  
submissions

31 January  
2025