History of Statistics Paper VI

Yingke(Kiki) He

July 15th 2025

Week 9

The Rise of Causal Thinking in Statistics

I. Introduction

For much of the history of statistics, practitioners have warned against confusing correlation with causation. This cautionary principle has become a standard refrain in classrooms and research papers, emphasizing that statistical associations do not imply a causal relationship. While this message serves as a useful warning, it has often been used as a substitute for methodological clarity about what causation actually means or how it might be discovered. As Judea Pearl argues, this mindset has stalled progress by encouraging statisticians to avoid causal questions rather than to develop tools for answering them (Pearl & Mackenzie, 2018, p. 27).

Pearl challenges the traditional view that causality lies outside the domain of statistics. Instead, he offers a framework in which cause and effect can be formally analyzed, using mathematical models and logic. In Chapter 2 of The Book of Why, he revisits the story of James Lind's 1747 scurvy experiment to show how early experiments revealed the potential for causal inference long before the development of modern statistics. He then introduces key concepts such as interventions, counterfactuals, and the ladder of causation, arguing that statistical science must expand beyond passive observation to include active reasoning about causes and consequences.

Pearl's critique of correlation-based thinking and his proposal for a structured approach to causal inference show the importance of experimental design, the conceptual progression from association to intervention, and the role of counterfactuals in scientific explanation. It highlights a growing recognition that statistics must not only describe the world but also provide a way to reason about how the world could change.

II. The Problem with Correlation

Statistical practice has long been dominated by tools that measure association. Correlation coefficients, regression models, and contingency tables allow analysts to summarize relationships between variables, but they stop short of identifying whether one variable causes another. This limitation has led to a widespread and cautious refrain: correlation does not imply causation. While this warning is important, Pearl argues that it has too often served as an intellectual stop sign, discouraging serious attempts to answer causal questions (Pearl & Mackenzie, 2018, p. 27).

The core issue is that correlation describes what tends to occur together, but not why it happens. For example, data might show a strong correlation between ice cream sales and drowning deaths. However, the true explanation lies in a third factor, such as hot weather, which affects both. Without a model of the underlying mechanisms, correlation alone cannot tell us whether changing one variable would change the other. This distinction is critical for real-world decisions in fields like public health, policy, and economics, where the goal is not just to observe patterns but to understand how actions lead to consequences.

Pearl emphasizes that relying solely on observed associations often leads to faulty reasoning. Without a causal framework, researchers may misinterpret spurious correlations as evidence of causation or overlook the influence of hidden variables. Moreover, even carefully controlled regression models can be misleading when they do not account for the direction of influence or the presence of feedback loops.

The central problem is not with correlation itself, which remains a useful descriptive tool, but with treating it as sufficient for causal understanding. Pearl calls for a shift away from passive observation and toward active inquiry, where analysts ask what would happen under different conditions. This requires models that go beyond statistical fit and instead encode causal assumptions explicitly.

In this view, correlation is just the starting point. To reason about interventions and counterfactuals, analysts must adopt tools that allow them to move from data to mechanism.

Pearl's framework offers a way forward by connecting statistical analysis to the logic of cause and effect, something that traditional methods have largely ignored.

III. James Lind and the Emergence of Experimental Design

In 1747, British naval surgeon James Lind conducted what is now recognized as one of the earliest clinical experiments. Onboard a ship bothered by scurvy, Lind selected twelve sailors suffering from similar symptoms and divided them into six pairs. Each pair received a different treatment, ranging from cider to seawater to citrus fruits. The pair given lemons and oranges recovered rapidly, while the others did not. Though Lind did not use the terminology of modern science, he effectively carried out a controlled trial that demonstrated a causal relationship between citrus intake and the cure for scurvy (Pearl & Mackenzie, 2018, pp. 34–35).

Pearl highlights this case to show that experimental reasoning about causation predates the formal development of statistical tools. Lind did not rely on correlation. Instead, he directly manipulated conditions and observed their effects. His study introduced two foundational ideas of experimental design: the importance of controlled comparison and the isolation of specific interventions. These principles remain central to the design of randomized controlled trials today.

Despite the clear causal implications of Lind's experiment, his findings were not immediately adopted by the British Navy. The institutional acceptance of citrus as a treatment came decades later. This delay illustrates how evidence alone is not always enough. Scientific recognition of causality requires a shared framework for interpreting results, and that framework did not yet exist. Lind’s study was treated as anecdotal or suggestive, rather than definitive. Pearl uses this example to critique how statistics developed in the centuries that followed. While correlation-based techniques gained sophistication, the logic of controlled experimentation was often undervalued or separated from statistical inference. As a result, the kind of thinking that allowed Lind to discover a cure for scurvy remained on the margins of statistical education and practice.

Lind’s experiment shows that causal knowledge is not inherently mysterious or unreachable. It can be revealed through deliberate action and thoughtful comparison. Pearl argues that acknowledging this is essential to building a modern framework for causal inference, one that unites empirical data with structured reasoning about interventions.

IV. The Ladder of Causation

One of the most influential ideas introduced by Pearl is the concept of the ladder of causation, a framework that organizes different types of reasoning into three distinct levels: association, intervention, and counterfactuals. According to Pearl, most traditional statistical tools operate only on the first level. They identify patterns and correlations in data, which allows us to answer questions such as, “What is the probability of Y given X?” This is useful for prediction, but it does not address how Y would change if X were actively altered (Pearl & Mackenzie, 2018, pp. 28–29).

The second level of the ladder is intervention. Here, we shift from observing to doing. Instead of asking what tends to happen when X and Y occur together, we ask what happens when we actively set X to a specific value. This difference may seem subtle, but it is essential for making policy decisions and for evaluating the effects of treatments or actions. At this level, we are interested in the outcomes of hypothetical interventions. The formal tool Pearl introduces to handle such reasoning is the do operator, which allows analysts to model the effect of forcing a variable to take a specific value, rather than just conditioning on its observed value.

The third and highest level of the ladder is counterfactual reasoning. This involves asking questions such as, “What would have happened if a different decision had been made?” Counterfactuals allow us to assess the effect of an action that was not actually taken, using both observational data and assumptions encoded in a causal model. These questions are crucial in contexts like legal reasoning, retrospective medical analysis, and ethical decision-making. Counterfactuals require a richer set of assumptions and a deeper model of the data-generating process.

Pearl emphasizes that each level of the ladder builds on the one below it, but also introduces qualitatively different reasoning. You cannot answer counterfactual questions using only correlational data, and you cannot infer the effects of interventions without a model that includes causal structure. Traditional statistics, which tends to remain on the association level, lacks the tools to ascend this ladder.

By formalizing these levels, Pearl offers a new way of understanding what kinds of questions we can answer with data and what additional information we need to answer them well. The ladder of causation encourages statisticians and scientists to move beyond surface-level patterns and to think carefully about mechanisms, actions, and possibilities that extend beyond the observed world.

V. Causal Models and the Do-Calculus

To move from observing associations to reasoning about interventions and counterfactuals, Pearl argues that we must adopt formal causal models. These models go beyond fitting equations to data. They encode assumptions about how variables influence one another, often represented as directed graphs, where arrows depict causal relationships. Such models allow us to simulate changes, isolate pathways, and test alternative scenarios in a structured way (Pearl & Mackenzie, 2018, pp. 32–33).

At the center of Pearl's framework is the do-calculus, a set of mathematical rules for manipulating expressions that involve interventions. These rules allow us to derive the probability of an outcome given an intervention, based on the structure of the causal graph and observed data. The do-calculus helps answer questions of the form, "What is the effect of doing X?" rather than "What is the association between X and Y?"

The distinction between conditioning and intervening is essential. Conditioning on a variable involves observing its value and analyzing correlations. Intervening means actively changing the variable and studying the consequences. For instance, observing that patients who take a certain drug recover more quickly does not tell us whether the drug caused the recovery. There may be confounding factors, such as healthier patients being more likely to choose the drug. The do-calculus provides a way to adjust for such confounders and recover the causal effect.

Causal models also support counterfactual reasoning, which is the highest level on the ladder of causation. Using structural equations and assumptions about how variables relate to one another, analysts can answer questions like, "Would this patient have recovered if they had not taken the drug?" These questions require imagining alternative realities and depend heavily on the structure of the model.

Pearl emphasizes that causal models are not merely optional additions to statistical practice. They are necessary tools for answering questions that matter in science, medicine, law, and policy. Without them, we remain confined to surface-level correlations and cannot evaluate the consequences of actions we might take or avoid. Importantly, Pearl does not claim that causal models eliminate uncertainty. Rather, they make our assumptions transparent and subject to critique. By showing how conclusions depend on the structure of the model, they invite scrutiny and refinement. This openness is a strength, not a weakness, and it distinguishes causal inference from more opaque forms of statistical analysis.

VI. What I've Learned

Before reading this chapter, I had thought of causality as something external to statistics, something that belonged more to philosophy or experimental science than to statistics. What stood out most in Pearl and Mackenzie's account was how clearly they demonstrate that causality can be approached with the same precision and formality as other areas of statistical reasoning if we are willing to make our assumptions explicit and use appropriate tools.

The story of James Lind's scurvy experiment helped me understand that causal reasoning has always been part of scientific inquiry, even if it was not formalized mathematically. Lind's ability to identify a treatment that worked, simply by assigning conditions and comparing outcomes, shows that causality does not require complicated machinery. It requires comparison and a willingness to act on hypotheses.

Pearl's ladder of causation helped clarify for me why many statistical methods fail when it comes to answering real-world questions. It is not enough to observe patterns because in order to make decisions, we need to reason about interventions and imagine different outcomes. This was especially useful in thinking about counterfactuals. I now see that being able to answer “what if” questions is not only possible, but essential in areas like medicine, education, and social policy.

I also liked how Pearl frames causal models not as perfect representations of reality, but as structured ways of making our assumptions testable. That idea resonates with how I think about all models, as tools for thought, not as final truths. The do-calculus and structural graphs provide a language to ask better questions and to understand when and how we can answer them. Most importantly, I learned that causal inference is not an extension of statistics, but a foundation that should inform it. Without a causal framework, we are limited to describing what is. With one, we can reason about what could be.

Week 10

How Computing and Prediction Transformed Statistical Thinking

I. Introduction

In recent decades, statistics has undergone a transformation driven by the growth of data, the rise of machine learning, and the increasing importance of computing. While classical statistics focused on inference, theory, and carefully constructed models, modern data work often prioritizes prediction, scalability, and computational efficiency. These changes have sparked debate about what statistics is, how it should be practiced, and whether the field is being eclipsed or redefined by data science.

Leo Breiman’s influential paper framed this debate by identifying a fundamental divide between two cultures in statistical modeling: one focused on stochastic models and inference, the other on algorithmic prediction and real-world performance (Breiman, 2001). Around the same time, practitioners like David Hand and David Donoho argued that data science was not a new discipline but a continuation of statistics shaped by computing and practical needs (Hand, 2015; Donoho, 2017). Meanwhile, Gelman and Vehtari (2021) cataloged major statistical advances over the past 50 years, many of which grew out of a shift toward simulation, computation, and flexible modeling.

This paper examines how the culture of statistics has evolved through the influence of computing and prediction. It explores the tensions between interpretability and performance, the integration of ideas from computer science, and the need to reconnect academic statistics with real-world applications. The goal is not to resolve the debate between traditions but to understand how these changes reflect deeper shifts in how uncertainty is modeled and managed in practice.

II. Breiman's Two Cultures

Leo Breiman's 2001 paper introduced a powerful critique of traditional statistical thinking by dividing the field into two cultures. The first, which he called the data modeling culture, assumes that data are generated by a specific stochastic process. Statisticians working in this culture focus on estimating parameters and testing hypotheses using models such as linear regression or logistic regression. These models emphasize interpretability and theoretical understanding but often make unrealistic assumptions about the structure of the data (Breiman, 2001, p. 200).

The second, which Breiman called the algorithmic modeling culture, approaches the data as a black box and prioritizes predictive performance. Instead of assuming a fixed model, algorithmic approaches use flexible tools such as decision trees, neural networks, and random forests. The emphasis is on minimizing prediction error and validating models empirically, often through cross-validation on test data.

Breiman argued that this culture was more aligned with real-world applications, where the primary concern is accurate forecasting rather than inference about a predefined model. Breiman's central claim was that the dominance of the data modeling culture had limited the field of statistics. He argued that its focus on idealized models often led to poor performance in complex or high-dimensional settings, and that statisticians had failed to engage with tools developed outside of their discipline, especially in machine learning. In contrast, algorithmic models had achieved strong results in practical domains such as speech recognition, handwriting classification, and medical diagnosis.

Key innovations such as bagging and boosting, which are methods for improving predictive accuracy by combining multiple models and his work on random forests showed that it was possible to achieve strong predictive performance without relying on assumptions about linearity or normality. He acknowledged that these models were often less interpretable, but he argued that their predictive value justified their use. Breiman concluded that statisticians should expand their focus beyond inference and model testing. He called for a shift in training and mindset, urging the field to embrace the diversity of tools available and to evaluate models based on their ability to solve meaningful problems. His call continues to resonate in debates about the future of statistics and its relationship to data science.

III. The Genesis of Data Science

The rise of data science has prompted questions about whether it represents a new discipline or a continuation of statistics transformed by computing. David Hand (2015) argues that many of the core ideas in data science were already present in statistics long before the term "data science" became popular. What changed was not the goals of the field, but the tools available to achieve them. The development of powerful computers allowed statisticians to tackle problems that had previously been too complex or computationally expensive.

For example, methods like bootstrapping and Bayesian computation, once considered impractical, became routine with the growth of computing power. Similarly, ideas from computer science, including decision trees and support vector machines, were initially met with skepticism by statisticians but later became integral to applied data analysis. These developments blurred the boundaries between disciplines and led to a convergence of methods focused on solving practical problems with large, complex datasets.

Hand emphasizes that data science is not just about working with big data or using machine learning algorithms. It also includes issues such as data preprocessing, visualization, interpretation, and uncertainty. These aspects were always part of statistical work, but computing made them more central and more scalable. As data volumes grew, the challenges of cleaning, storing, and exploring data became inseparable from modeling itself.

There is a risk in treating data science as something entirely new and separate from statistics. Hand warns that this attitude can lead to the reinvention of existing techniques under new names, often without a full understanding of their limitations or assumptions. Rather than creating divides, he advocates for collaboration between statisticians and computer scientists, and for a recognition that both communities contribute to a shared goal: learning from data in the presence of uncertainty.

What emerges is a picture of data science as a continuation of statistical thinking, enriched by computing and broadened by interdisciplinary applications. The shift is not only in tools, but also in scope: moving from abstract theory to real-world systems, from static datasets to streaming environments, and from isolated analysis to integrated data workflows.

IV. Reclaiming Data Science for Statistics

David Donoho (2017) takes the argument further by examining what data scientists actually do and showing that many of their tasks were always part of statistical practice. He identifies six areas of activity in data science: data preparation, data representation and transformation, computing with data, modeling, visualization, and scientific communication. Traditional statistics, he argues, focused almost exclusively on modeling and inference, while neglecting the other essential components of data analysis. This narrow focus created a gap between what was taught in statistics departments and what was needed in practice.

Donoho points out that in industry and applied research, most of the effort goes into understanding, cleaning, and transforming data, as well as building pipelines for computation and interpretation. These tasks are rarely emphasized in academic statistics, where proof-based courses and parametric inference dominate the curriculum. As a result, many students trained in statistics are unprepared for the work expected of data scientists. Donoho does not suggest abandoning modeling or inference, but instead expanding the field to fully embrace the broader workflow of modern data analysis.

A key insight from Donoho is that the definition of statistical success has shifted. Rather than focusing on whether a model is theoretically optimal, practitioners care more about whether the analysis yields reliable, interpretable, and reproducible results. This shift reflects the growing importance of computation and transparency, especially when working with large or complex data. As computing became central to data analysis, so too did the skills and practices needed to manage it effectively.

Donoho also warns against the tendency of data science to define itself in opposition to statistics. He argues that many of the tools now associated with data science (such as resampling, robust modeling, or high-dimensional methods) have their origins in statistical research. Instead of drawing sharp boundaries, he encourages a more inclusive view of the discipline, one that sees data science as the modern expression of statistics adapted to new tools and challenges. By expanding the scope of what statistics includes, Donoho calls for a more complete vision of the field, one that matches how data are actually used in the world. This vision does not abandon rigor, but complements it with flexibility, communication, and computation.

V. Transformative Statistical Ideas

Andrew Gelman and Aki Vehtari (2021) identify several statistical developments that have fundamentally reshaped the field over the past 50 years. These innovations reflect a broader shift in statistical practice, one that places greater emphasis on flexibility, computation, and practical engagement with complex data. Many of the ideas they highlight grew out of the challenges described by Breiman, Hand, and Donoho, and offer responses to the limitations of traditional models.

Among the most significant developments is the formalization of causal inference. New frameworks using counterfactuals, potential outcomes, and graphical models have clarified what assumptions are necessary to identify causal effects from observational data. These tools allow analysts to move beyond correlation and reason about interventions in a systematic way.

Another major shift is the widespread use of simulation-based methods, particularly bootstrapping and Bayesian inference via Markov chain Monte Carlo. These approaches make it possible to estimate uncertainty and evaluate models without relying on closed-form solutions. Gelman and Vehtari emphasize that computation has become as essential to statistics as theory, enabling new types of modeling that were previously impractical.

The rise of overparameterized models, including neural networks and other machine learning techniques, has challenged older beliefs about the dangers of complexity. In many applications, highly flexible models can outperform simpler ones if regularization is used to prevent overfitting. This has led to new thinking about the trade-off between bias and variance and about the role of predictive accuracy in evaluating models.

Another key development is the growth of Bayesian multilevel modeling, which allows for partial pooling of information across groups and more structured representations of uncertainty. These models are well suited to hierarchical data and small sample problems, and they integrate naturally with modern computational tools.

Gelman and Vehtari also draw attention to exploratory data analysis, emphasizing the importance of visualization and model checking. Rather than treating analysis as a linear process, they advocate for an iterative approach in which models are continuously evaluated and revised based on data and context. Together, these ideas represent a transformation in how statistics is practiced. They illustrate a shift from focusing on narrow parametric forms to embracing uncertainty, complexity, and iteration. The goal is not only to build models, but to understand the data-generating process and to support decisions in the presence of uncertainty.

VI. What I’ve Learned

The readings for this week challenged my understanding of what statistics is and what it can become. I thought of statistics primarily as a collection of formulas and models used to draw conclusions from data. These papers showed me that the field is much broader, shaped by evolving technologies, practical demands, and philosophical debates about what it means to learn from data.

Breiman’s distinction between modeling cultures helped me see how different goals (interpretability versus prediction) can lead to entirely different approaches. His argument for algorithmic modeling made me reconsider the value of models that do not explain much but still perform well. At the same time, I understood that the emphasis on prediction does not replace the need for careful reasoning about assumptions and uncertainty.

Hand and Donoho emphasized how computing changed the nature of statistical work. They made it clear that many of the tasks central to modern data science like cleaning data, building workflows, and managing computation, were always part of applied statistics, even if they were not always taught that way. Their call to recognize these contributions as part of the discipline, rather than outside it, connected with my experience of working with real datasets. The ideas from Gelman and Vehtari helped me appreciate the role of flexibility and iteration in statistical thinking. I especially liked the idea that good models are not always the simplest ones, and that modern tools like regularization, bootstrapping, and Bayesian methods allow us to engage with complexity in principled ways. Their emphasis on model checking and exploratory analysis reminded me that statistics is not just about producing results, but about making sure those results are meaningful and robust.

Overall, this week helped me connect the different aspects of statistics as it is taught in theory and the ways it is used in practice. It showed me that computation is not a threat to statistics, but an opportunity to expand its scope and relevance. More importantly, it reminded me that the core purpose of statistics remains the same: to make sense of uncertainty and support better decisions in a complex world.

References

Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science, 16(3), 199–231. https://doi.org/10.1214/ss/1009213726

Donoho, D. L. (2017). 50 years of data science. Journal of Computational and Graphical Statistics, 26(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

Gelman, A., & Vehtari, A. (2021). What are the most important statistical ideas of the past 50 years? Journal of the American Statistical Association, 116(536), 2087–2097. https://doi.org/10.1080/01621459.2021.1938081

Hand, D. J. (2015). Statistics and computing: The genesis of data science. Statistics and Computing, 25, 1039–1049. https://doi.org/10.1007/s11222-015-9565-6

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (Chapter 2). Basic Books.