History of Statistics Paper III

Yingke(Kiki) He

June 15th 2025

Week 5: Early 1900s

Regression, Correlation, and the Question of Causality in Early 20th-Century Statistics

I. Introduction

By the early 20th century, statistical tools like correlation and regression were becoming widely used in both the natural and social sciences. They offered a way to summarize patterns in data and quantify relationships between variables, giving the impression of scientific precision. Yet behind this technical polish was a deeper tension: What exactly did these associations mean? Could they be used to infer causation, or were they simply descriptive tools?

Figures like Karl Pearson and George Udny Yule helped formalize these methods, introducing formulas and diagrams that became central to the emerging discipline of statistics (Stigler, 1986, Ch. 10). Their work marked the rise of what David Freedman later called the "correlational era," where mathematical associations were emphasized without a clear framework for interpreting causality (Freedman, 1999). Meanwhile, Francis Ysidro was attempting to build a more rigorous theory of statistical estimation using ideas like likelihood and characteristic functions, though his work was often too mathematically abstract to guide applied research at the time (Stigler, 1986, Ch. 9).

This backdrop sets the stage for a major shift in the goals and methods of statistical science. As statistical tools were increasingly used to justify public policy and scientific claims, the stakes grew higher and so did the need to distinguish association from causation. Freedman (1999) emphasizes that many of the persistent misinterpretations of regression and correlation, especially in the social sciences, stem from this foundational ambiguity. Revisiting the origins of these methods shows how much of modern inference depends not only on technical refinement, but on philosophical clarity and thoughtful study design.

Francis Ysidro Edgeworth (1845 – 1926), about 1892

Figure 1. Francis Ysidro Edgeworth (1845–1926), a pioneer in theoretical statistics whose mathematical depth often limited the accessibility of his ideas. Photograph ca. 1892.

II. Edgeworth's Theoretical Contributions and Limitations

Francis Ysidro Edgeworth occupies a curious place in the history of statistics. He was one of the first to seriously consider the mathematical foundations of statistical inference, anticipating techniques like likelihood-based estimation and asymptotic distribution theory. His work explored the use of characteristic functions, Taylor series expansions, and approximate solutions to inverse problems: ideas that would become central to modern statistical theory (Stigler, 1986, Ch. 9). Yet for all his insight, Edgeworth's influence during his lifetime remained limited.

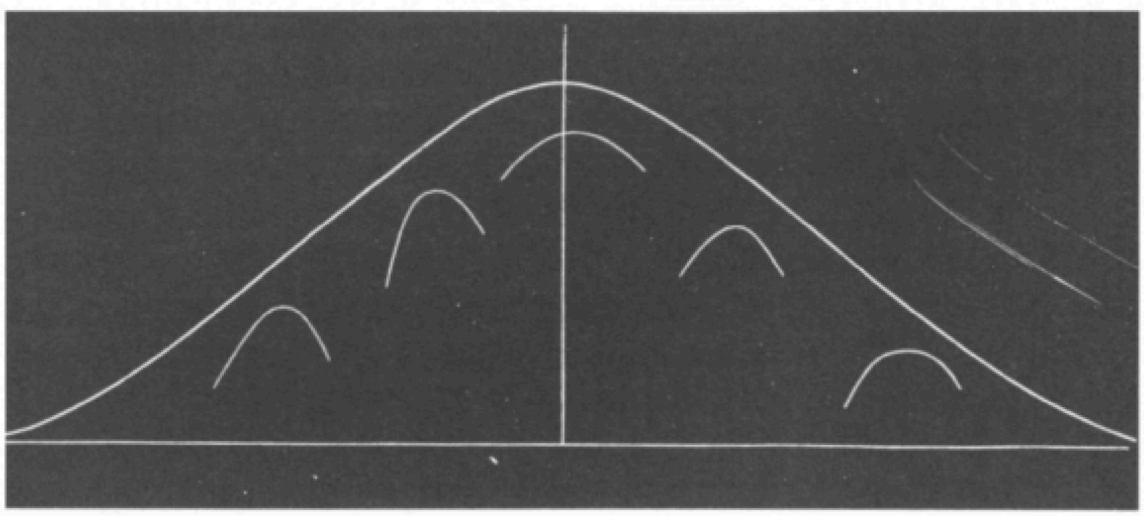

Figure 9.1. Edgeworth's 1885 illustration of Galton's 1875 insight. The smaller curves represent the distributions of the sizes of pears from specific parts of a garden (e.g., near a particular wall), and the large curve gives the distribution for the whole garden, thus showing the normal curve as a mixture of normal components. (From Edgeworth, 1885b.)

Figure 2. Edgeworth's 1885 visual representation of Galton's insight that aggregate distributions (e.g., of pear sizes) could emerge from the mixture of several local distributions. Source: Edgeworth (1885).

Part of the challenge was Edgeworth's style. His writing was dense, often meandering, and lacked concrete examples that would have made his ideas more accessible. While he contributed important ideas, such as early discussions of maximum likelihood and confidence intervals in spirit, which were typically buried in long, speculative treatises (Stigler, 1986, pp. 296–307). Few of his contemporaries were equipped or motivated to work through the complexity.

Still, Edgeworth deserves credit for recognizing that probability theory alone was not enough: statistics required a theory of inference, one that could handle uncertainty and draw conclusions from incomplete data. As Sheynin (2018, Ch. 12) notes, Edgeworth worked on generalizing the

Central Limit Theorem and explored the logic behind combining observations, prefiguring later work by Fisher and Neyman.

Even though Edgeworth lacked the pedagogical clarity or institutional reach of Pearson, his work laid a foundation for later formalism. His career is a reminder that theoretical breakthroughs do not always arrive neatly packaged and that the development of usable statistical methods often requires others to translate abstract insight into concrete tools.

III. Pearson and Yule: Building Empirical Tools

Karl Pearson (1857–1936), in 1890

George Udny Yule (1871–1951)

I hope that you flourish in Probabilities.

—From a letter from Edgeworth to Pearson.
11 September 1893

Figure 3. Karl Pearson (left) and George Udny Yule (right), who transformed statistical tools like correlation and regression into widely used empirical methods.

While Edgeworth worked to establish a theoretical base for statistical inference, Karl Pearson and George Udny Yule focused on building and systematizing statistical methods that could be directly applied to real data. Pearson, in particular, was instrumental in formalizing tools such as the correlation coefficient, regression analysis, and the chi-squared test, methods that became staples of empirical research across disciplines (Stigler, 1986, Ch. 10).

Pearson's approach was rooted in biology and eugenics, where he used statistical tools to study heredity and variation in human populations. His work extended Galton's interest in measuring traits to an entire statistical program, complete with standardized distributions and empirical curve-fitting methods. As MacKenzie (1981) emphasizes, this was not a politically neutral project. Pearson's biometric school was deeply tied to a belief in improving society through selective breeding and population control, and statistical methods were used to lend authority to those aims.

Yule, a student of Pearson, helped refine and expand these tools. He worked on partial correlation, time series analysis, and the theory of spurious correlation, an early recognition that mathematical associations could arise from structural artifacts rather than causal connections (Stigler, 1986, pp. 336–341). Yule's efforts helped extend Pearsonian statistics beyond biology into economics and demography, broadening the reach of statistical modeling.

However, as Freedman (1999) warns, the growing popularity of regression and correlation also brought misinterpretations. These methods, originally developed as tools for exploring associations, were increasingly used to draw causal conclusions, often without the experimental controls or theoretical justifications necessary for valid inference. The shift from description to explanation happened quickly, and not always carefully.

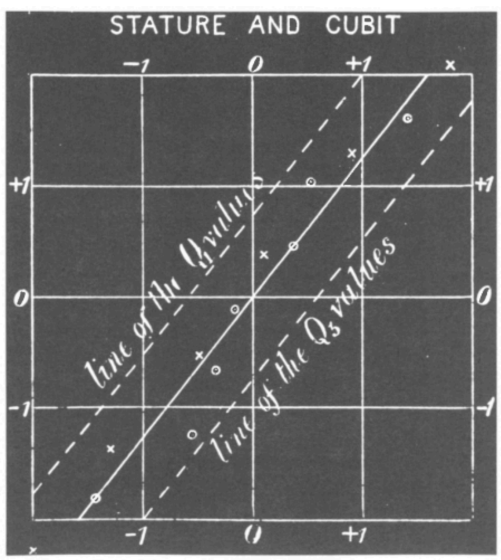

Figure 4. Galton's 1888 graphical determination of the correlation coefficient using median plots and visual estimation. This influenced Pearson's later formal development of correlation.

Together, Pearson and Yule helped make statistics usable and visible, institutionalizing it through journals like Biometrika and university laboratories. But their legacy also includes the widespread assumption, which are still common today, that significant associations imply meaningful causation. This assumption, as later critics and reformers would show, can be deeply misleading if not paired with careful study design and critical thinking.

IV. Freedman's Caution: Association ≠ Causation

As statistical methods like regression and correlation became standard tools for data analysis, they also began to take on interpretive roles they weren't designed for. In his 1999 paper, David Freedman argues that this shift from using models to describe patterns, to using them to infer causality introduced a persistent problem in applied statistics: mistaking association for cause and effect (Freedman, 1999).

Freedman critiques the assumption that a statistically significant regression coefficient implies a meaningful causal relationship. He shows that without careful attention to study design, ignored variable bias, and model assumptions, the results of regression models can be easily misinterpreted.

This concern is especially important in the social sciences, where researchers often rely on observational data and have limited control over the mechanisms that generate the data. This critique applies directly to the work of Pearson and Yule. While their methods provided a way to explore and quantify relationships in data, they did not and were not originally intended to support strong causal claims. Yet over time, users of these tools began to treat statistical significance as evidence of explanation, a pattern that continues in many fields today. Yule's own recognition of spurious correlation was an early warning about the dangers of overinterpreting regression results in systems where unobserved confounding could be present (Stigler, 1986, Ch. 10).

Freedman's historical account reminds us that the misuse of statistical models is not a recent problem. From the very beginning, there was a gap between what statistical tools could do and how they were actually used. His work invites a return to the principles of careful inference: understanding the data generating process, considering alternative explanations, and emphasizing the limits of what models alone can tell us.

V. Between Theory and Application: Tensions in Early Statistical Practice

The contrast between Edgeworth's abstract formulations and Pearson and Yule's empirical emphasis highlights a broader tension that defined early 20th century statistics: Should the discipline prioritize theoretical rigor or practical usability? This question shaped not only the methods that emerged, but also the institutions, journals, and political causes with which statisticians became aligned.

Edgeworth envisioned a mathematically unified theory of inference, drawing on probability and asymptotic logic. Yet his work remained largely inaccessible due to its complexity and lack of clear application (Stigler, 1986, Ch. 9). In contrast, Pearson and Yule built a hands-on statistical system that could be taught, published, and used in fields ranging from biology to economics, even if that meant leaving some theoretical questions unresolved (MacKenzie, 1981, Ch. 2).

This tension between abstract theory and applied modeling also extended to questions about the role of statistics in public life. As MacKenzie shows, statistical methods were not just tools for scientific inquiry; they were embedded in broader ideological projects. For Pearson, this meant supporting eugenics with data and modeling; for others, it meant using statistics to inform social reform or economic planning. In both cases, the authority of statistical models came to rest not only on their internal logic, but on their institutional backing and rhetorical power.

Freedman (1999) helps frame this legacy by pointing out that many of today's debates about causal inference: for instance, whether regression can stand in for experimental evidence, tracing back to these early unresolved tensions. The tools developed during this period were foundational, but their interpretation was never straightforward. They were designed in one context and often deployed in another, creating a persistent risk of misuse or overconfidence.

In retrospect, the period spanning Edgeworth to Yule represents a transitional moment in the history of statistics: a time when the language of models was maturing, but the logic of inference was still being negotiated. That negotiation would continue and become more formalized in the decades to follow, especially with the contributions of Fisher, Neyman, and Pearson's next generation.

VI. What I've Learned

What stood out most in this week's readings to me personally was how early statisticians built tools that remain foundational today: correlation, regression, and chi-squared testing without always agreeing on what those tools meant or how they should be used. I had often thought of these methods as neutral and well-defined, but learning about the historical debates between Edgeworth, Pearson, and Yule helped me see that statistics has always involved both mathematical reasoning and philosophical judgment.

Freedman's idea about treating association as causation is very interestingt. In the past, I've used regression models in the classroom under the assumption that significant coefficients imply meaningful effects. But this week helped clarify that such interpretations depend heavily on study design, assumptions about omitted variables, and whether the data reflect the kind of process we aim to understand. It also reminded me that the language of causality often get into our analysis even when we don't intend it to.

I was also shocked by how statistical practice is connected to institutions and ideologies. MacKenzie's account of Pearson's biometric school emphasized that statistics didn't grow in a vacuum: it grew through eugenics, academic rivalry, and the needs of the state. That context shaped which problems were seen as worth solving, which tools were developed, and how statistical knowledge was framed as scientific authority.

Overall, this week deepened my understanding of both the power and limitations of early statistical methods. It showed me that understanding where tools come from (and the assumptions behind them) is as important as knowing how to apply them. The history of statistics isn't just technical; it's also about judgment, values, and the consequences of simplification.

Week 6: Early 1900s Continue

1933 and the Foundations of Modern Inference: Competing Frameworks and the Rise of Mathematical Statistics

I. Introduction

By the early 1930s, statistics had matured from a loosely connected collection of empirical tools into a discipline that is on the edge of a formal consolidation. Techniques like correlation, regression, and chi-squared testing were well established, but the foundational questions of inference like how to draw reliable conclusions from data remained contested. The year 1933 has often been identified as a pivotal moment in this transition. It was the year when Andrei Kolmogorov introduced a rigorous axiomatic framework for probability theory, and when the competing philosophies of R. A. Fisher and the Neyman Pearson school began to solidify into distinct, sometimes oppositional approaches to statistical inference (Stigler, 1996; Sheynin, 2018, Chs. 14–15).

In this period, two major frameworks for inference emerged. Fisher advocated for likelihood-based reasoning and experimental design grounded in inductive inference from a single study (Sheynin, 2018, Ch. 14; Lehmann, 2011). Meanwhile, Neyman and Pearson proposed a more rule-based, frequentist logic centered on long-run error rates and hypothesis testing as a decision-making tool (Sheynin, 2018, Ch. 15; Stigler, 1996). These philosophical divergences would define much of statistical thinking for the next century.

At the same time, Kolmogorov's axiomatization of probability redefined the mathematical keystone of the field. Building on measure theory, Kolmogorov created a structure that could accommodate both continuous and discrete systems, unifying diverse applications under a single formal language (von Plato, 1994; Sheynin, 2018, Ch. 15). His work marked the beginning of probability as a rigorous mathematical discipline, and implicitly marginalized competing interpretations like the subjectivist theory of Bruno de Finetti.

These developments did not occur without any signs. As MacKenzie (1981) and Lehmann (2007) remind us, statistics evolved in close relation to institutions, ideology, and community. Textbooks, journals, and training programs played essential roles in shaping which ideas took hold. Even reflections like Stigler's "Seven Pillars" (2016) show that the consolidation of statistical wisdom was as much about framing as it was about discovery. Understanding this moment reveals the philosophical assumptions, historical events, and institutional pressures that still shape statistical practice today.

II. The Year 1933: Stigler's Framing

In his essay The History of Statistics in 1933, Stephen Stigler identifies that year as a critical turning point which is "a moment of intellectual crystallization" in the development of modern statistics (Stigler, 1996). It was a time when key foundational ideas came into sharp focus, even as deep divisions emerged between competing schools of thought. Most notably, it marked the emergence of two parallel achievements: Kolmogorov's axiomatic foundation for probability and the formalization of Neyman Pearson decision theory. Though different in scope and style, both contributed to a new scope in the field.

Stigler highlights how Jerzy Neyman and Egon Pearson's formulation of hypothesis testing established a structured framework for statistical decisions based on pre-specified error rates and clearly defined hypotheses. This logic, centered on Type I and Type II errors and the concept of test power, stood in contrast to Fisher's more flexible and inductive style. Neyman and Pearson sought to impose a frequentist discipline on inference, treating statistics as a tool for decision-making under uncertainty rather than inductive reasoning from evidence.

At the same time, Andrei Kolmogorov's 1933 monograph, Foundations of the Theory of Probability, gave probability a rigorous mathematical home within measure theory. His axioms allowed probability to be treated like any other branch of mathematics, making it possible to unify the analysis of discrete, continuous, and stochastic processes under one general framework (Sheynin, 2018, Ch. 15; von Plato, 1994).

Stigler also emphasizes the role of institutional growth in reinforcing these conceptual shifts. By the 1930s, statistics had begun to acquire disciplinary autonomy: new journals (e.g., the Supplement to the Journal of the Royal Statistical Society), university departments, and research labs provided the infrastructure to train statisticians and apply new methods. These settings not only facilitated innovation but also shaped which ideas were taken seriously. As Lehmann (2007) later reflects, academic communities especially at Berkeley and in Britain became key engines in the propagation of new statistical paradigms.

Thus, 1933 was not just a year of technical advance. It was a moment of consolidation, when the intellectual and institutional architecture of modern statistics began to resemble what we recognize today: a field with internal debates, competing frameworks, and growing professional identity.

III. Competing Philosophies: Fisher vs. Neyman Pearson

The intellectual divide between R. A. Fisher and the Neyman Pearson school represents one of the most important debates in the history of statistics. Although all parties aimed to formalize statistical inference, their philosophies diverged in how they understood evidence, uncertainty, and the role of probability. Fisher believed statistical inference should center on likelihood and inductive reasoning. For him, probability measured the strength of evidence within a given dataset, not long-run error rates. He introduced the idea of the null hypothesis, the p-value, and maximum likelihood estimation, emphasizing their use in drawing conclusions from a single experiment (Sheynin, 2018, Ch. 14).

Fisher also encourage the design of experiments, including innovations like randomization and blocking, which allowed researchers to control for bias and enhance internal validity. His view was that good design minimized the need for elaborate inference; the data, properly collected, should speak for themselves.

In contrast, Neyman and Pearson emphasized decision rules based on long-run behavior. They introduced formal definitions of Type I and Type II errors, the power of a test, and the idea that inference should guide decisions rather than beliefs. In their framework, statistical procedures were judged by how often they erred when applied repeatedly under fixed conditions (Sheynin, 2018, Ch. 15). This perspective treated statistical inference more like quality control than scientific discovery, less concerned with interpreting a particular result and more with establishing procedures that performed well over repeated use.

The philosophical argument was clear. While Fisher aimed to learn from data, Neyman and Pearson were concerned with managing risk in a formalized and replicable way. Stigler (1996) notes that even their shared terminology like "hypothesis testing" faced substantial conceptual differences. For Fisher, rejecting the null hypothesis was a measure of surprise, not a decision to act. For Neyman and Pearson, it was precisely that: a rule-based decision within a structured framework.

This division would shape not only methodological debates but also statistical education and applied practice for decades. As Lehmann (2011) recounts, even students trained in both paradigms often lean toward one, depending on whether they valued interpretability or rigor, flexibility or control. Though both frameworks are still used today, their unresolved tensions continue to spark debates about how to reason statistically in fields ranging from medicine to economics.

IV. Probability Formalized: Kolmogorov and the Axiomatic Approach

While Fisher and Neyman Pearson debated how to conduct inference, Andrei Kolmogorov addressed a deeper problem: how to define probability itself. In 1933, he published Grundbegriffe der Wahrscheinlichkeitsrechnung (Foundations of the Theory of Probability), which introduced an axiomatic framework grounded in measure theory. This work, later recognized as one of the most important mathematical contributions to modern statistics, gave probability a rigor and generality that earlier formulations lacked (Sheynin, 2018, Ch. 15; von Plato, 1994).

Kolmogorov’s approach defined probability as a non-negative, countably additive measure on a set of events, normalized so that the total probability is one. These axioms allowed probability to be treated like geometry or algebra—a formal system that could be applied to both discrete and continuous problems. His framework unified concepts such as random variables, expectation, and conditional probability, providing the basis for later developments in stochastic processes, limit theorems, and statistical modeling.

What made Kolmogorov’s theory so powerful was its generality. By using tools from Lebesgue integration and abstract measure spaces, he could capture phenomena ranging from coin flips to Brownian motion. Von Plato (1994) emphasizes that this formalization emerged not only from pure mathematics but from needs in statistical physics, where concepts like continuous probability distributions were central. Kolmogorov’s work offered a consistent language for both theoretical and applied problems.

However, Kolmogorov’s axioms also marginalized alternative interpretations. In particular, the subjective probability model proposed by Bruno de Finetti, in which probabilities represent degrees of belief, grounded in personal coherence rather than frequencies was marginalized in the mainstream. De Finetti’s work on exchangeability and Bayesian logic would only gain influence much later, despite its internal coherence and practical appeal (Sheynin, 2018, Ch. 15).

Kolmogorov’s axiomatic system did not resolve all philosophical debates. It provided clarity on how to manipulate probabilities, but not on what they mean. Still, it became the dominant foundation for probability theory in the 20th century, influencing everything from statistical inference to quantum mechanics. His clear and structured approach helped turn probability from a loose collection of rules into a well-developed part of mathematics.

V. Broader Context: Institutions, Culture, and Memory

While 1933 marked a turning point in the intellectual formalization of statistics, these theoretical advances did not emerge in isolation. The field's development was deeply shaped by its institutional settings, social values, and the personal relationships among its leading figures. These broader forces played a critical role in determining which ideas became popular, which were left behind, and how statistical knowledge was framed and transmitted.

Donald MacKenzie (1981) underscores that the growth of statistics in Britain, particularly from the late 19th to early 20th century, was linked to broader ideological currents. For example, the rise of Pearson's biometric school was not only a scientific project but a political one, which linked to eugenics, imperial governance, and the belief that human traits could be measured, ranked, and controlled. Statistical tools gained credibility in part because they aligned with the administrative and reformist agendas of the time. The success of any method, MacKenzie argues, depended not only on its mathematical validity but also on who used it, where it was taught, and what institutions endorsed it.

This perspective helps explain why the debates between Fisher and the Neyman Pearson school became so central. Beyond technical disagreements, their frameworks were supported by rival institutions and professional networks. Lehmann's memoir (2007) highlights how personality, mentorship, and academic culture shaped the diffusion of these ideas—particularly within elite departments like Berkeley and UCL. The acceptance of Neyman Pearson methods in applied science, for example, owed as much to their institutional visibility and reproducibility as to their theoretical merit.

Meanwhile, Stigler's *Seven Pillars* (2016) looks back and outlines the key ideas that shape modern statistical thinking. From aggregation and likelihood to experimental design and residual analysis, His pillars are not just about theory—they also reflect important values, like keeping things simple, replication, and data-centered reasoning. Notably, he highlights regression not just as a mathematical discovery but as a solution to a biological puzzle (the stability of hereditary traits), linking the rise of statistical tools to specific scientific problems of the time.

Together, these sources remind us that statistics is not merely a collection of formulas. It is a product of its history, shaped by social conditions, institutional power, and the evolving demands of science and governance. Recognizing this context helps us better understand why certain methods became dominant, and why foundational tensions like those between frequentist and Bayesian reasoning still persist today.

VI. What I've Learned

This week's readings showed that modern statistics was shaped not only by new ideas, but also by different views on what statistical inference should be. I had previously learned about Fisher and Neyman Pearson in terms of their technical differences, p-values versus Type I/II errors, but reading their ideas in historical context made their disagreements seem more important. Fisher’s emphasis on likelihood and single-experiment reasoning now looks less like a methodological preference and more like a different philosophy of science, one centered on evidence and discovery.In contrast, Neyman and Pearson’s method of hypothesis testing shows their focus on clear rules and controlling the chance of errors.

I also better understand how foundational these differences were to the later development of statistics. Learning how Kolmogorov used measure theory to define probability showed me how math added precision to the field, but also made me realize that this precision came with downsides, like pushing aside other views such as de Finetti’s subjectivism. The review of Creating Modern Probability by von Plato (1994) reinforced this, showing that modern probability was as much a response to the needs of physics as to the internal logic of statistics.

Stigler’s Seven Pillars (2016) helped me frame all of this conceptually. It clarified why core ideas like aggregation and regression still matter, and how they connect to much older problems in science, like variation in heredity. It was also helpful to see how institutional dynamics shaped the spread of ideas through departments, journals, and personal networks, as described by Lehmann (2007) and MacKenzie (1981). These readings reminded me that, like any science, statistics is shaped by the people who use it and the places where it's developed.

In the end, I gained a better understanding of statistics, not just as a set of tools, but as a field that has grown through debates and changing ideas. The push and pull between flexibility and rules, between drawing conclusions and making decisions, and between theory and practice are still relevant today. Knowing where these tensions came from helps me better recognize the assumptions I make when choosing a method or interpreting results.

References

Freedman, D. A. (1999). From association to causation: Some remarks on the history of statistics. *Statistical Science, 14*(3), 243–258. https://doi.org/10.1214/ss/1009212800

Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (N. Morrison, Trans., 2nd ed.). Chelsea Publishing. (Original work published 1933)

Speed, T. (2008). Review of Reminiscences of a Statistician: The Company I Kept by Erich L. Lehmann. International Statistical Review, 76(1), 153–154. https://doi.org/10.1111/j.1751-5823.2007.00039.x

Lehmann, E. L. (2011). *Fisher, Neyman, and the creation of classical statistics*. Springer.

MacKenzie, D. (1981). *Statistics in Britain, 1865–1930: The social construction of scientific knowledge*. Edinburgh University Press.

Sheynin, O. (2018). *Theory of probability: A historical essay* (arXiv:1802.09966). https://arxiv.org/abs/1802.09966

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.

Stigler, S. M. (1996). The history of statistics in 1933. *Statistical Science, 11*(1), 70–86. https://doi.org/10.1214/ss/1032280213

Krashniak, A., & Lamm, E. (2017). Was regression to the mean really the solution to Darwin's problem with heredity? Essay Review of Stigler, Stephen M. 2016. The Seven Pillars of Statistical Wisdom. Cambridge, Massachusetts: Harvard University Press. *Biology & Philosophy*, *32*, 749-758.

Shafer, G. (1998). Review of *Creating modern probability: Its mathematics, physics, and philosophy in historical perspective* by J. von Plato. *The Annals of Probability, 26*(1), 416–424.