

# Biostatistics and the philosophy of scientific inquiry

AJ Smit

2024-03-28

## Table of contents

<b>The Classical View: Karl Popper, Falsification, and Certainty</b>	<b>1</b>
<b>Traditional Biostatistics and Beyond</b>	<b>4</b>
Parametric Statistics . . . . .	4
Non-Parametric Statistics . . . . .	4
Multivariate Statistics . . . . .	5
Bayesian Statistics . . . . .	5
Machine Learning and Artificial Intelligence . . . . .	5
Computational Models and Simulations . . . . .	5
Qualitative Analysis . . . . .	6
Phylogenetic Analysis . . . . .	6
<b>Challenges and Alternatives: Bayesian Approaches and Inductive Reasoning</b>	<b>7</b>
<b>Reconciling Perspectives: Hypothesis Testing as a Tool in the Scientific Toolbox</b>	<b>7</b>

Hypothesis testing plays a central role in scientific inquiry, yet its prominence and exact interpretation are subject to debate, even within university scientific departments. In this essay/lecture I will explore contrasting views on hypothesis testing and its function within the process of scientific knowledge generation. I will also discuss the implications of these perspectives for the practice of biostatistics and the philosophy of scientific inquiry.

## The Classical View: Karl Popper, Falsification, and Certainty

Karl Popper (1902-1994) was a central figure in 20th-century philosophy of science. His emphasis on empirically-based falsifiability revolutionised how we understand the nature of science

and conduct the process of scientific inquiry.

Popper challenged the prevailing **verificationist paradigm** (and the attendant **inductivist approach** that stemmed from the accumulation of empirical evidence), held by the logical positivists, that science aimed to prove theories through empirical evidence. Instead, he argued that a theory is scientific only if it can, in principle, be proven false. Scientific theories can never be definitively verified, but can only withstand attempts at refutation with the weight of current evidence. This is an inherently skeptical perspective that emphasises the provisional nature of knowledge and encourages continuous re-evaluation.

### **i Verificationist and Inductivist Approaches**

**Verificationism:** Verificationism is closely associated with the logical positivist movement prominent in the early 20th century. Logical positivists argued that for a statement to be meaningful, it must be either empirically verifiable or analytically true (true by definition). According to verificationists, a scientific statement is only meaningful if it can be verified through observation or experiment. In this view, the purpose of scientific inquiry is to accumulate observations that support or verify hypotheses and theories.

Verificationists believe that through the accumulation of positive observations they can confirm the validity of a theory. This approach is often linked with the principle of induction.

**Inductivism:** Inductivists take an approach to science that emphasises induction as the primary method for developing knowledge. Induction involves making generalisations based on a number of specific observations or experiments. For example, observing that the sun rises in the east every morning leads to the general principle that the sun always rises in the east.

In inductivism, scientific knowledge is built up from a base of empirical observations. Scientists observe phenomena, detect patterns, formulate general laws from these observations, and then infer broader theories that can explain the laws. The more observations that align with a theory, the stronger the inductive support for that theory. Inductive reasoning is, therefore, the process of moving from particular instances to general conclusions.

Hypothesis testing, as practiced today, has at its heart Popper's philosophy of falsification. The null hypothesis ( $H_0$ )—representing the default position of no effect or difference—is the unambiguously-stated falsifiable claim. The alternative hypothesis ( $H_1$ ) proposes the existence of an equally precisely-stated a measurable effect. We design measurable experiments or structured field sampling campaigns specifically to gather evidence that could potentially refute the null hypothesis. If the effect revealed by the data contradicts the null hypothesis, we reject it in favour of the alternative. This process is iterative, with hypotheses (and eventually also theories) gaining support as they withstand repeated attempts at falsification.

While Popper's framework provides a philosophical foundation, the specific statistical tools

used in hypothesis testing stem from the work of statisticians like Ronald Fisher, John Tukey, Jerzy Neyman, and Egon Pearson. Their development of null hypothesis significance testing (also called inferential statistics) created the widely used statistical tools for evaluating scientific claims—that is, to decide between the null and alternative hypotheses.

### **i The Fathers of Modern Inferential Statistics**

**Ronald Fisher (1890–1962):** Fisher is often referred to as the father of modern statistics. His contributions are vast, including the development of Analysis of Variance (ANOVA), Fisher’s exact test, and the concept of maximum likelihood estimation. Fisher’s work in the early 20th century, particularly in the 1920s and 1930s, laid the groundwork for much of the statistical hypothesis testing and experimental design used today.

**Egon Pearson (1895–1980):** Egon Pearson is best known for his collaboration with Jerzy Neyman. Together, they developed the Neyman-Pearson lemma in the 1930s, which established the foundation for hypothesis testing. Their work introduced the concepts of Type I and Type II errors, and the power of a test, which are critical to the field of statistical decision theory.

**Jerzy Neyman (1894–1981):** While Neyman’s collaboration with Pearson was pivotal, his contributions to the practice of statistics extended beyond. Notably, in the 1930s and onwards, Neyman developed confidence intervals and furthered the formal mathematical basis of statistical hypothesis testing. His work, alongside Pearson, has been instrumental in defining the procedures for statistical inference.

**John Tukey (1915–2000):** Tukey’s influence became pronounced in the mid-20th century. He introduced the box plot, the Fast Fourier Transform (FFT) algorithm, and coined the term ‘bit’. Tukey was instrumental in developing exploratory data analysis (EDA), which emphasizes the importance of exploring data before making assumptions or hypotheses. His work in the 1960s and 1970s on robust statistics and data analysis has had a lasting impact on the field.

Popper’s emphasis on critical rationalism supports this iterative search for knowledge. We propose hypotheses, rigorously test them, and adjust our understanding as evidence either supports or challenges existing explanations. This ongoing process of conjecture and refutation leads to a more accurate and nuanced understanding of the world.

This view places hypothesis as the foundation of deductive reasoning and supports the idea that science cannot confirm absolute truths but only advances by successively eliminating falsehoods. Falsification’s strengths lie in emphasising rigour, scepticism, and the continuous evolution of scientific knowledge.

Popper’s principles of making bold conjectures and seeking refutations are widely applied, albeit through diverse methodologies that may not always involve traditional statistical hypothesis testing of the kind taught in biostatistics courses such as this one. Such disciplines

leverage a range of alternative approaches to challenge and expand existing knowledge, demonstrating the adaptability and breadth of scientific inquiry.

## Traditional Biostatistics and Beyond

Scientific research is rich and varied and encompasses disciplines that range from the rigorously quantitative to the deeply qualitative. Within this spectrum, the choice of statistical methods—be it parametric, non-parametric, Bayesian, or alternatives beyond traditional statistical frameworks—plays a crucial role in how data are analysed and interpreted.

In the context of hypothesis testing, the choice of statistical test is guided by the nature of the data and the research question. For example, a *t*-test is used to compare the means of two groups, while a *chi*-square test is used to assess the association between two categorical variables. These tests are designed to evaluate specific hypotheses and are based on assumptions that must be met for the results to be valid. In traditional biostatistics courses such as this one, students are taught to select the appropriate test based on the research question and data type, and to interpret the results in the context of the study design. Typically, such courses focus on parametric and non-parametric inferential statistical methods. Let's quickly look at what these methods are and offer some alternative approaches.

### Parametric Statistics

Parametric statistics rely on assumptions about the distribution of the data, typically assuming a normal distribution. This approach is powerful when those assumptions hold, as it allows for the derivation of exact statistical properties and more precise estimates of parameters. In fields like ecological and biological sciences, where experiments are designed to measure specific outcomes under controlled conditions, parametric tests can offer important insights. However, the reliance on distributional assumptions limits the applicability of parametric methods in fields dealing with more complex, non-normally distributed, or heteroscedastic data, such as socio-ecological studies or aspects of oceanography influenced by myriad interacting factors.

### Non-Parametric Statistics

Non-parametric statistics, by contrast, do not assume a specific distribution for the data. This flexibility makes non-parametric methods invaluable in fields like taxonomy and systematics or certain areas of oceanography, where data may not meet the stringent requirements of parametric tests. Non-parametric methods can analyse ranks or medians, accommodating data that are skewed, ordinal, or on a nominal scale. However, this flexibility often comes at the cost of statistical power and the inability to make more detailed inferences about parameters, limiting their utility in cases where parametric assumptions are met.

## **Multivariate Statistics**

Multivariate methods in ecological studies include a suite of statistical approaches designed to deal with complex data where multiple variables or responses are measured simultaneously. These methods are suited to the complexities of ecological research, where interactions between species, environmental variables, and temporal changes may influence outcomes in a non-linear and interdependent manner. Unlike univariate analyses (parametric and non-parametric statistics), which consider a single response variable, multivariate analyses enable us to examine patterns and relationships across multiple dimensions of the ecosystem. They offer a more integrated understanding of ecological dynamics.

## **Bayesian Statistics**

Bayesian statistics offer a different paradigm by incorporating prior knowledge into the analysis, updating our current thinking in light of new evidence. This approach is useful in complex, data-sparse fields like deep-ocean exploration or predictive modelling in socio-ecological research, where prior expert knowledge can be quantitatively integrated with sparse or uncertain data. Bayesian methods facilitate a more nuanced understanding of uncertainty and allow for direct probability statements about parameters. However, the need for specifying prior distributions and the computational complexity of Bayesian analysis can pose challenges, especially in fields where prior information is subjective or hard to quantify. Bayesian methods are sometimes suggested to be more intuitive and flexible than traditional frequentist methods—placing them outside of Popper’s falsification paradigm—but they require careful consideration of prior information and model assumptions.

## **Machine Learning and Artificial Intelligence**

Machine Learning (ML) and Artificial Intelligence (AI) methods are revolutionising the landscape of hypothesis-driven research across almost all scientific disciplines. These computer-intensive methods offer sophisticated tools to analyse large volumes of data, identify patterns, and make predictions. They are increasingly enhancing our ability to test hypotheses and derive insights. The application of ML and AI in hypothesis-driven research marks a significant shift towards data-driven science, where computational models and algorithms play a central role in uncovering the complexities of natural and social phenomena.

## **Computational Models and Simulations**

Widely used in oceanography and molecular research, these models enable the exploration of complex systems and the making of testable predictions without relying directly on statistical tests of data. For example, oceanographers use comparative and longitudinal studies to

understand oceanic processes over time and space. By comparing historical data with current observations, they can infer changes in ocean chemistry, temperature, and currents. These data then feed into simulation models that are used to predict future states of ocean systems under various scenarios, such as climate change. These models are continuously refined based on new empirical data. They enable the exploration of complex systems and the making of testable predictions without relying directly on statistical tests of data.

## **Qualitative Analysis**

Socio-ecological research frequently employs qualitative research approaches such as thematic analysis, narrative analysis, case studies, ethnography, stakeholder (or ‘actors’ as socially-minded people call them) interviews, and participatory methods to understand the complex interactions between people and their environments. These researchers often try to arrive at deep insights into patterns and processes that elude quantification. Such qualitative approaches allow for deep dives into specific contexts, generating insightful data sets that can be used to develop and test our thinking about social and ecological dynamics. Scenario analysis and systems modelling provide tools to explore future outcomes based on current knowledge, offering a form of hypothesis testing that is grounded in theoretical frameworks and empirical data. These approaches are essential for understanding context-specific dynamics and can inform the development of a more generalised understanding of people’s relation with nature.

## **Phylogenetic Analysis**

In taxonomy and systematics, phylogenetic methods use genetic data to reconstruct the evolutionary history of species. These methods, while informed by statistical principles, rely on algorithms and computational methods to infer relationships, rather than traditional hypothesis testing. There may also be a requirement for morphological data to explore the relationships among species. Phylogenetic trees, constructed based on similarities and differences in genetic markers or morphometric features, offer testable predictions about the evolutionary history of life forms. These predictions can be challenged and refined as new species are discovered or as more sophisticated analytical techniques become available. Phylogenetic methods use genetic data to reconstruct the evolutionary history of species. These methods, while informed by statistical principles, rely on algorithms and computational methods to infer relationships, rather than traditional hypothesis testing. Bayesian inference is a common tool in phylogenetic studies; it uses probabilistic models to estimate the most likely evolutionary tree given the data.

## Challenges and Alternatives: Bayesian Approaches and Inductive Reasoning

The dominance of falsification-based hypothesis testing has been challenged. One critique stems from the Duhem-Quine thesis, which argues that real-life experiments don't isolate a single hypothesis. Instead, they test a package of theories and assumptions simultaneously. Therefore, a failed experimental prediction may not necessarily falsify the main hypothesis but could indicate a problem somewhere else within the interconnected web of assumptions.

Furthermore, Bayesian approaches to probability suggest that hypothesis testing should focus on updating the degree of belief in a hypothesis based on evidence rather than outright rejection. Bayesian methods incorporate prior knowledge, offering a framework to quantify how new data shifts the likelihood of a hypothesis being true. This approach arguably better reflects the incremental nature of scientific knowledge building.

Additionally, some philosophers of science argue that hypothesis testing shouldn't be the sole focus when evaluating scientific ideas. Inductive reasoning, pattern identification, exploratory research, and model building also play crucial roles in expanding our understanding of the natural world.

## Reconciling Perspectives: Hypothesis Testing as a Tool in the Scientific Toolbox

Rather than viewing these perspectives as mutually exclusive, it's valuable to see them as complementary tools within the broader scientific process. In many scientific fields, rigorous hypothesis testing and its emphasis on falsifiability remain essential to ensure reliable and objective knowledge. However, it's crucial to acknowledge the following:

- **Context is Key:** The appropriateness of different approaches depends on the research question, the maturity of the field, and available data.
- **Statistics are Interpretative:** Statistical significance does not equate to real-world importance; researchers must critically interpret the meaning of their findings.
- **Science is Iterative:** Knowledge progresses through a cycle of hypothesis generation, testing, refinement, and occasionally, the development of entirely new paradigms.

Hypothesis testing in the traditional scientific method often relies on statistical methods, especially in fields like biology and medicine. However, scientific rigour can be maintained without heavy reliance on traditional statistical hypothesis testing through various other philosophies and paradigms:

1. **Mechanistic Understanding:** In fields like molecular biology, the focus is often on understanding the mechanisms and pathways at a cellular or molecular level. These studies may use detailed observations, experiments, and models to develop a mechanistic understanding of phenomena without extensive statistical tests.
2. **Exploratory Data Analysis (EDA):** Instead of starting with a hypothesis, we may explore data without preconceived notions to uncover patterns, trends, and relationships. This can lead to the generation of new hypotheses that can be tested more rigorously later.
3. **Simulation and Modelling:** Disciplines such as oceanography and climate science often use computational models to understand complex systems. These models are based on physical laws and are validated against empirical observations, sometimes without traditional hypothesis tests.
4. **Systematics and Taxonomy:** These fields rely on the comparative method and often use phylogenetic analysis. They may not always use hypothesis testing in the statistical sense but use morphological, genetic, or biochemical characteristics to infer relationships and classifications.
5. **Qualitative Research:** In social sciences, qualitative research methods such as interviews, focus groups, and ethnography aim to gather an in-depth understanding of human behaviour and the reasons that govern such behaviour or views. They seek to understand ‘how’ and ‘why’ things happen rather than just ‘what’ happens.
6. **Philosophy of Falsification:** Popper’s falsification is a philosophical approach where scientific hypotheses are subjected to tests where they could potentially be proven false. Instead of confirming a hypothesis, we try to falsify it, and hypotheses that withstand attempts at falsification are considered robust.
7. **Evidence-based Practice:** Particularly in fields like healthcare and policy-making, decisions are made based on a preponderance of evidence from various studies, which may include a mix of quantitative and qualitative research, expert opinion, and systematic reviews.
8. **Bayesian Inference:** This statistical approach incorporates prior knowledge in addition to current experimental data. It can provide a more nuanced understanding of the likelihood of a hypothesis being true.
9. **Consilience:** This approach involves the integration of evidence from multiple disciplines to draw a comprehensive conclusion. It is based on the unity of knowledge and can be particularly useful in socio-ecological studies where different fields intersect.
10. **Critical Realism:** A philosophical approach used in social sciences that combines a realistic view of the world with a recognition of the importance of understanding the subjective experiences and beliefs of individuals involved.



These alternative methods emphasise the importance of context, understanding the underlying mechanisms, the use of models and simulations, the collection of qualitative data, or the integration of various sources of evidence. They all can provide scientific rigour and confidence in findings, often complementing or serving as valuable alternatives to traditional statistical hypothesis testing.