Biostatistics and the philosophy of scientific inquiry

AJ Smit

2024-03-28

On this page

The Many Faces of Scientific Inquiry: Hypothesis Testing and Beyond

Today, we're going to delve into the concept of hypothesis testing. We'll look at its foundations, its uses in various disciplines, and the ongoing debates about its role in the grander scheme

of scientific knowledge creation. To start, let's go back to a key figure in the philosophy of science...

Karl Popper, Falsification, and a Search for Certainty

Karl Popper (1902-1994) was an influential philosopher of science. His ideas revolutionised how we think about scientific inquiry. He challenged the prevailing view of his time, *verificationism*, which held that science seeks to *prove* theories true through evidence.

> **ℹ Verificationist and Inductivist Approaches**
>
> **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** emphasises developing knowledge through making generalisations based on a number of specific observations or experiments. This process is called indiction. 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. We 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.

Popper took a different approach. He argued that true scientific hypotheses are inherently *falsifiable*. They must be stated in a way that could, in principle, be proven wrong. While we can never *definitively* prove an hypothesis false, but we can gain support for alternative outcomes that cause us to not accept the null statement. This emphasises scepticism, rigour, and an understanding that scientific knowledge is always evolving as weaker arguments are discarded in the face of mounting new evidence.

Hypothesis testing, as it us used today, has Popper's philosophy at its heart. Think about it: The null hypothesis (H0), the sceptical default statement of 'no effect', is a clearly stated falsifiable claim. The alternative (H1), the outcome we're often more interested in, offers

a precise, opposing prediction. Experiments are designed precisely to *try* and disprove the null.

Of course, the statistical tools we use *within* this framework owe much to statisticians like Fisher, Neyman, Pearson, and Tukey. Their work gave us the means to decide, with a level of calculated confidence, whether the data does indeed contradict our null hypothesis.

> **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 emphasises 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, Critical Rationalism, and the Iterative Scientific Process

Popper's ideas tie into the broader philosophy of *critical rationalism.* Science is a process of conjecture and refutation. We propose hypotheses, test them mercilessly, and adjust our understanding based on the evidence. This constant cycle of refinement is what moves knowledge forward. This philosophy of asserted conjecture and attempts at deliberate refutation applies across disciplines, even those that may not use statistical hypothesis testing in the traditional sense. It's about a mindset of challenging our ideas and seeking ways to improve or even overturn them.

> **ℹ Critical Realism**
>
> Critical realism is a philosophical approach to understanding the nature of reality and the methods through which we can come to know it. It was developed primarily by the philosopher Roy Bhaskar in the 1970s. Critical realism posits that the world exists independently of our perceptions, theories, and constructions, embodying a realist ontology. However, it also acknowledges the complexity of this reality, suggesting that our understanding of it is always mediated by social, cultural, and linguistic factors, which aligns with a constructivist epistemology.

Hypothesis Testing in Traditional Biostatistics... and Beyond

Scientific research takes many forms, from the quantitative to the qualitative. Our choice of analytical methods – parametric, non-parametric, Bayesian, or even approaches outside statistics entirely – depends heavily on the data and our research question.

In biostatistics courses like BCB744, we focus on traditional hypothesis tests. A *t*-test to compare means, a *chi*-square to look for associations... these give us a *toolkit* when our data and questions fit a certain mould. But it's just one of many toolkits. Let's briefly explore some others and discuss where they might be a better fit.

Parametric Statistics

Parametric statistics rely on certain distributional assumptions—most commonly, the normal distribution—and serve as a foundation in quantitative analysis across many scientific disciplines. They facilitate the extraction of precise statistical properties and parameter estimates, making them particularly powerful for testing hypotheses and deriving inferences in controlled experimental setups or data derived from systematic, structured sampling campaigns. In cases where data do not conform to the assumed distributions, we must employ data transformations or leverage non-parametric methods that do not require specific distributional assumptions.

Non-Parametric Statistics

While parametric statistics offer significant advantages in terms of precision and power, their applicability across various scientific disciplines necessitates a thorough understanding of their assumptions and limitations. Non-parametric statistics are inherently more flexible and do not depend on restrictive assumptions about the nature of our data—as such, they can accommodate non-normal data, skewed data, or data measured on an ordinal scale. They focus instead on ranks or medians rather than mean values, and provide a means to conduct robust set of statistical inference tests on a far wider range of data types.

There are a few trade-offs we need to know about when opting for non-rarametric approaches. This includes the potential loss in statistical power and the nuances of interpreting rank-based or median-based results as opposed to mean values. Nevertheless, non-parametric statistics are a critical component of our toolbox across disciplines such as taxonomy, systematics, organismal biology, ecology, socio-ecological studies, and Earth sciences.

Multivariate Analyses

Ecologists consider questions about the complex interactions between the biotic and abiotic world. Often they work across multiple spatial and temporal scales. Multivariate analyses such as cluster analysis and ordination are powerful exploratory tools. They untangle ecological datasets to discern patterns and relationships among multiple variables—be it species abundance across different habitats, environmental gradients, or the dynamical properties of ecosystems. Cluster analysis groups similar entities based on their characteristics and reveals natural groupings within the data. Ordination, on the other hand, reduces multidimensional space, making it easier to visualise and interpret complex ecological relationships. In contrast to the more traditional parametric and non-parametric statistics, which often focus on testing hypotheses about the relationships between variables, multivariate analyses provide a more overarching view. They allow us to uncover hidden structures and gradients in the data without *a priori* hypotheses. While parametric methods hinge on assumptions about data distribution and non-parametric methods offer flexibility in handling data that don't meet these assumptions, multivariate analyses surpass these by focusing on the ecosystem's interconnectedness and the patterns emerging from these connections.

Bayesian Methods

Bayesian methods offer a distinct perspective within the statistical toolbox, allowing us to formally incorporate prior knowledge into our data analysis. Unlike traditional frequentist statistics, which focus solely on the observed data, Bayesian approaches let us blend in existing beliefs or information and then update those beliefs as new evidence comes in. This emphasis on continuously refining our understanding, rather than just finding a single best-fit hypothesis given the data, makes Bayesian methods powerful in scientific fields where we have substantial background knowledge but still need to carefully quantify uncertainty. Bayesian methods are particularly useful in fields like ecology or phyologenetics, where prior knowledge about species interactions or relatiosnhips, environmental conditions, or ecosystem dynamics can be leveraged to make more informed inferences. They also provide a natural framework for decision-making under uncertainty, allowing us to quantify the risks and benefits of different courses of action.

The downside of Bayesian analyses is that they can be computationally intensive, especially when dealing with complex models or large datasets. They also require careful consideration

of the prior distributions, which can introduce subjectivity into the analysis. But as computational resources continue to expand and methodologies evolve, Bayesian approaches are likely to play an increasingly prominent role in advancing our understanding of the natural world.

Taxonomy, Systematics, and Phylogenetics

Phylogenetic analysis, while grounded in data and statistical principles, operates within a distinct framework from traditional parametric, non-parametric, or multivariate methods. Its primary goal is to infer evolutionary relationships and patterns of change, rather than classical hypothesis testing. Phylogenetics explicitly models the interconnectedness of evolutionary lineages, a stark contrast to the assumption of independent data points often found in other statistical approaches. While multivariate analyses help examine complex interactions among multiple variables, phylogenetics focuses on how those variables (traits or genes) have evolved across a branching, tree-like structure. Bayesian statistics offer a powerful tool within phylogenetics, aiding in the estimation of probabilities for different evolutionary histories. Yet, the core of phylogenetics lies in specialised algorithms and models designed to reconstruct these evolutionary narratives.

Phylogenetic analysis is deeply intertwined with systematics and taxonomy, disciplines that seek to understand and classify the diversity of life. Systematics broadly encompasses the study of organismal relationships, while taxonomy focuses on the practice of naming and classification. Phylogenetics serves as a powerful tool within the systematics toolbox, using data to infer evolutionary patterns and inform classification decisions. While statistical methods like parametric and non-parametric tests are used in systematics and taxonomy (e.g., for comparing morphological traits), much of the analytical toolkit centers on techniques specifically designed for evolutionary data. These techniques include methods for building phylogenetic trees, assessing congruence between different data sources (like genes and morphology), and interpreting patterns of diversification over time.

Artificial Intelligence and Machine Learning

Models and Simulations

Qualitative Analysis

Phylogenetic Analysis

- **Non-Parametric Statistics:** Free us from strict distributional assumptions, great in areas like taxonomy where data might violate parametric rules.
- **Multivariate Analyses:** Let us tackle ecological complexity where multiple factors interweave with messy, non-linear outcomes.

- **Bayesian Statistics:** Update our beliefs based on evidence, valuable where prior knowledge exists and data is uncertain.
- **AI and Machine Learning:** Data-driven patterns and prediction,a powerful addition to the hypothesis-testing arsenal.
- **Models and Simulations:** Allow us to explore complex systems and make predictions, vital in fields like oceanography.
- **Qualitative Analysis:** Socio-ecological studies benefit from in-depth exploration of human attitudes and actions, where quantification may not tell the full story.
- **Phylogenetic Analysis:** Data-driven exploration of evolutionary relationships, less about statistical tests and more about algorithms and inference.

Challenges to the Dominance of Falsification

Not everyone agrees that Popper-style hypothesis testing should be the pinnacle of science. The Duhem-Quine thesis reminds us that experiments rarely test a hypothesis in isolation – a failure could mean many things, not just that our main idea is wrong.

Bayesian approaches focus less on outright rejection, and more on how evidence shifts the probability of a hypothesis being true.This aligns well with that notion of iterative, evolving knowledge.

Importantly, some argue that hypothesis testing shouldn't be our sole focus.Inductive reasoning, spotting patterns, and exploratory research are also vital in building our understanding of the world.

A Toolkit, Not a Dogma

Rather than getting stuck in philosophical battles, it's wiser to see these approaches as complementary.Rigorous hypothesis testing is essential in many fields. But here's the key:

- **Context Matters:** The right approach depends on your question, your data, and your discipline's norms.
- **Meaning, Not Just Numbers:** Statistics help us decide, but it's up to *us* to understand the real-world implications of what the data says.
- **Science Evolves:** Knowledge progresses through testing, refinement, and occasionally, whole new paradigms when the old models truly fail.

Whether you're using traditional statistics, phylogenetic trees, or in-depth interviews, the core of good science is the same: asking questions, seeking evidence in a structured way, and being open to the possibility that you might be wrong. That's true scientific thinking.