Why Statistics

Conor Redington

Contents

- What I see the goal of Science

- The importance of rigour.

- What is statistics and how does it align the above.

- The tension between the practical and the abstract.

What is the 'case' is a powerful idea. The 'case' is physical reality, at some moment in time or space. It is absolute reality, a platonic ideal we deduce from out sensory experience.

0.1 Blind men and the elephant

My intuitive feeling for what is the 'case' is best summarised by the parable of the blind men and the elephant. Several blind men encounter an elephant for the first time and attempt to understand what it is by touching different parts of its body.

Each blind man touches a different part of the elephant—the trunk, the tusks, the ears, the legs, and the tail—and based on their limited experience, they come to vastly different conclusions about what an elephant is. One might describe the elephant as like a snake (feeling the trunk), another might say it's like a spear (feeling the tusks), while another might say it's like a fan (feeling the ears).

It's my philosophy that an objective reality exists and that a scientists job is to explore that reality, to generate new knowledge about it.

0.2 Rigour

One of my biggest inspirations is Carl Sagan. I read his work, the 'Demon Haunted World', to explore the supernatural, to become 'excited' about reality. I thought, 'there must be more to reality than this boring science'. Anyone whose read the book will see the irony in my motivation. Sagan dismantles notions of the supernatural but maintains the validity of the claims without the scientific method. The subtle is 'Science as a candle in the dark'. This method is used to keep us from fooling ourselves. Sagan pushes for humility but rigour in how we should teach science.

This rigour is of vital importance. The failing of intuitive explanations are widespread in popular science books. While stoking the imagination is a great service, I also think these books do a disservice in expectation of application for the reader. Reading about black holes and using terms familiar to our physical ideas can feel like understanding but leave an emptiness and inability to actually say anything useful about a black hole or explain it to others.

In April of 2023 I started to self-study John Tsitsiklis' course on probability from MIT opencourseware. Tsitsiklis' approach was to provide rigour as well as general practical application. What fascinated me was that when intuitive explanations didn't work I would just try and learn the formulas, to get a sense of the notation and symbols. This was the mathematical object that described the phenomenon. No matter the 'intuition' you came up with to describe it, you could never truly say more or less than the formula itself. Having this source of truth to grapple with made me start to like to seemingly opaque mathematics. I began to explore the formula, like a painting for what was and was not allowed in my explanation to myself.

Take for instance the idea of a random variable. This is a real valued function from the outcome space to the real number line. You might feel, 'well thats no explanation'. But that is what it is. The number that faces up on the roll of a die is a random variable, the outcome of a coin toss (provided we label heads one and tails zero) is a random variable. If I had begun with seemingly disparate phenomenon, or worse yet, never stated the mathematical object some would gain the 'proper' intuition while others would be left by the wayside.

0.3   What is Statistics?

Statistics is the applied philosophy of science. It is a technology to relate pieces of data which are 'feedback' from reality. Statistics is the application of our understanding of how to extract information from data and our understanding of uncertainty.

The raw material of statistics is data. Data is information from the world, or reality. We might further define it as *evidence* noting the definition of evidence as:

facts, information, documents, etc. that give reason to believe that something is true

This data takes the form of numbers. There is a lack of ambiguity in numbers. They only have one property. Numbers come directly from the phenomena being studied whereas words come from the brain. Data are natures evidence, seen through the lens of the measuring instrument.

0.4   Maintaining practicality with a healthy dose of the philosophical.

I cannot but question things. At the same time, one needs to pursue things. The methods of statistics are well known and some are mechanistic in their use. I want to learn these mechanisms while maintaining abstract notions about the fundamental task that statistics is concerned with.