

BIOS6611 Homework1 Question2

Toward Evidence-Based Medical Statistics. 1: The p-value Fallacy.

a. The meaning of inductive vs. deductive reasoning as presented in the paper

Both deduction and induction belong to the inferential reasoning functions. Deductive reasoning is based on a knowledge or hypothesis (which we already know what is the natural truth) and we predict the possible outcome, to whether accept or reject the hypothesis. That means if the premises are correct, we cannot get a wrong result. In this way, deductive reasoning is regarded as “objective”, but deduction cannot expand our knowledge to the new situations or new things. Statistically, Neyman and Pearson’s hypothesis test is the classical way to check the truth of deductive reasoning.

Inductive reasoning is the reversed way of deduction. Induction goes from what we have observed toward what is supposed to be the natural truth (the hypothesis). That is the reason why evidence and judgement of the observation are more essential for induction. Every new piece of information and new observation will expand our knowledge toward new hypotheses. In this way, induction fits better for our general understanding of science. However, we can never be absolutely sure about the natural truth, which is also the problem of induction. From statistical point of view, Bayes’ Theorem is the almost perfect way form inductive reasoning analysis.

In the paper, the author used a vivid analogy of medical practice to the scientific reasoning. Given a textbook listing every kind of illness, so the medical student, who memorize every detail of an illness, can deductively figure it out the signs and symptoms. However, the inductive reasoning is more like that a physician bases on the symptoms, observation, and possibility of a disease occurrence to find out what kind of disease is the patient suffering.

b. What the author means by the “p-value fallacy”

The definition and practical meaning of p-value are preposterous. P-value is a combination of Neyman and Pearson’s hypothesis test and Fisher’s p-index on the strength of evidence. However, this combination is not coherent.

1. As two different type frequentist’s inferential test. Neyman and Pearson hypothesis test focuses on the long-run of the perspective, which error-based and deductive. In this way, the researchers would not care about the correctness for single value or experiment, the individual is anonymous and interchangeable less important. Fisher’s p-value is specifically used to test the strength of evidence on the type I error rate, which focuses on the short-run of the perspective. In this way, Fisher’s p-value pay more attention to the robustness of individual value, so called the evidence-based and inductive. Fisher’s value requires the evidence must be identifiable and unique. However, we are asking the p-value to work out both the long-run and the short-run, anonymous as well as identifiable, deductive as well as inductive. That is why p-value theory is so paradoxical that to be called p-value fallacy.

2. Another way to think about this problem is that the result from the p-value test could be deceptive. It is hard for researchers to calculate the correct p-value. Sometimes it all depends on how we design research and explain the results. There are so many examples listed in the paper,

about the six-patient trial and the student ranking. In my opinion, this is also a fallacy, that a statistical method should have taken to secure the natural truth, now can amplify the fallacy into a conceptual error. This is pretty ironical that a statistical method could cause a profound influence on the process of science and the nature of science truth, as mention in the paper.

c. Confidence intervals vs. p-values as presented in the paper

Although there are a lot of misleading perspectives for p-value, such as the incoherence of combined methods, no guide for the size of p-value, confusion of p-value and alpha value, and paradoxical inference philosophy, there are certain necessities mentioned on the existence of p-value. Most important, the p-value method is misleading our understanding of scientific judgement and prejudice. We are ignoring the evidence-based experiment but focus on the data. When we discuss the results, most of time, it is talking about the p-value without mentioning of prior research, background knowledge. P-value has become an “unintentional brand of tyranny”. It is urgent for use to find a better way.

The confidence intervals and the allotments of common sense can show more effects on the compatible with data. It could potentially surrogate for the hypothesis. The confidence intervals can measure and evaluate the strength evidence but depends on too much on the external evidence. It does not offer any mechanism to unite external evidence until there is extra relation. Thus, as mentioned in the paper “the confidence intervals are a step in the right direction, but they are not a solution for the frequentist methods. We are expecting the solutions such as likelihood or Bayesian methods in the next paper.