

## Preface

The implied reader of our book is a significance tester, the keeper of numerical things. We want to persuade you of one claim: that William Sealy Gosset (1876–1937)—aka “Student” of Student’s *t*-test—was right and that his difficult friend, Ronald A. Fisher, though a genius, was wrong. Fit is not the same thing as importance. Statistical significance is not the same thing as scientific finding.  $R^2$ , *t*-statistic, *p*-value, *F*-test, and all the more sophisticated versions of them in time series and the most advanced statistics are misleading at best.

No working scientist today knows much about Gosset, a brewer of Guinness stout and the inventor of a good deal of modern statistics. The scruffy little Gosset, with his tall leather boots and a rucksack on his back, is the heroic underdog in our story. Gosset, we claim, was a great scientist. He took an economic approach to the logic of uncertainty. For over two decades he quietly tried to educate Fisher. But Fisher, our flawed villain, erased from Gosset’s inventions the consciously economic element. We want to bring it back.

We lament what could have been in the statistical sciences if only Fisher had cared to understand the full import of Gosset’s insights. Or if only Egon Pearson had had the forceful personality of his father, Karl. Or if only Gosset had been a professor and not a businessman and had been positioned therefore to offset the intellectual capital of Fisher.

But we don’t consider the great if mistaken Fisher and his intellectual descendants our enemies. We have learned a great deal from Fisher and his followers, and still do, as many have. We hope you, oh significance tester, will read the book optimistically—with a sense of how “real” significance can transform your science. Biometricians who study AIDS and economists who study growth policy in poor countries are causing damage with

a broken statistical instrument. But wait: consider the progress we can make if we fix the instrument.

Can so many scientists have been wrong over the eighty years since 1925? Unhappily, yes. The mainstream in science, as any scientist will tell you, is often wrong. Otherwise, come to think of it, science would be complete. Few scientists would make that claim, or would want to. Statistical significance is surely not the only error in modern science, although it has been, as we will show, an exceptionally damaging one. Scientists are often tardy in fixing basic flaws in their sciences despite the presence of better alternatives. Think of the half century it took American geologists to recognize the truth of drifting continents, a theory proposed in 1915 by—of all eminently ignorable people—a German meteorologist. Scientists, after all, are human. What Nietzsche called the “twilight of the idols,” the fear of losing a powerful symbol or god or technology, haunts us all.

In statistical fields such as economics, psychology, sociology, and medicine the idol is the test of significance. The alternative, Gossetian way is a uniformly more powerful test, but it has been largely ignored. Unlike the Fisherian idol, Gosset’s approach is a rational guide for decision making and easy to understand. But it has been resisted now for eighty years.

Our book also addresses implied readers outside the statistical fields themselves such as intellectual historians and philosophers of science. The history and philosophy of applied statistics took a wrong turn in the 1920s, too. In an admittedly sketchy way—Ziliak himself is working on a book centered on Gosset—we explore the philosophy and tell the history here. We found that the recent historians of statistics, whom we honor in other matters, have not gotten around to Gosset. The historiography of “significance” is still being importantly shaped by R. A. Fisher himself four decades beyond the grave. It is known among sophisticates that Fisher took pains to historicize his prejudices about statistical methods. Yet his history gave little credit to other people and none to those who in the 1930s developed a decision-theoretic alternative to the Fisherian routine. Since the 1940s most statistical theorists, particularly at the advanced level, have not mentioned Gosset. With the notable exception of Donald MacKenzie, a sociologist and historian of science, scholars have seldom examined Gosset’s published works. And it appears that no one besides the ever-careful Egon S. Pearson (1895–1980) has looked very far into the Gosset archives—and that was in 1937–39 for the purpose of an obituary.

The evidence on the Gosset-Fisher relationship that Ziliak found in the archives is startling. In brief, Gosset got scooped. Fisher’s victory over

Gosset has been so successful and yet so invisible that a 2006 publication on *anti*-Fisherian statistics makes the usual mistake, effectively equating Fisher's approach with Gosset's (Howson and Urbach 2006, 133). In truth it was Gosset, in 1905, not Neyman, in 1938, who gave "the first emphasis of the behavioralistic outlook in statistics" (Savage 1954, 159).

Only slowly did we realize how widespread the standard error had become in sciences other than our home field of economics. Some time passed before we systematically looked into them. Thus the broader intervention here. We couldn't examine every science or subfield. And additional work remains of course to be done, on significance and other problems of testing and estimation. Some readers, for example, have asked us to wade in on the dual problems of specification error and causality. We reply that we agree—these are important issues—but we couldn't do justice to them here.

But we think the methodological overlaps in education and psychology, economics and sociology, agriculture and biology, pharmacology and epidemiology are sufficiently large, and the inheritance in them of Fisherian methods sufficiently deep, that our book can shed some light on all the *t*-testing sciences. We were alarmed and dismayed to discover, for example, that supreme courts in the United States, state and federal, have begun to decide cases on the basis of Fisher's arbitrary test. The law itself is distorted by Fisher. Time to speak up.

We invite a general and nontechnical reader to the discussion, too. If he starts at the beginning and reads through chapter 3 he will get the main point—that oomph, the difference a treatment makes, dominates precision. The extended but simple "diet pill example" in chapter 3 will equip him with the essential logic and with the replies he'll need to stay in the conversation. Chapter 17 through to the end of the book provides our brief history of the problem and a sketch of a solution.

Readers may find it strange that two historical economists have intruded on the theory, history, philosophy, sociology, and practice of hypothesis testing in the sciences. We are not professional statisticians and are only amateur historians and philosophers of science. Yet economically concerned people have played a role in the logic, philosophy, and dissemination of testing, estimation, and error analysis in all of the sciences from Mill through Friedman to Heckman. Gosset himself, we've noted, was a businessman and the inventor of an economic approach to uncertainty. Keynes wrote *A Treatise on Probability* (1921), an important if somewhat neglected book on the history and foundations of probability theory.

Advanced empirical economics, which we've endured, taught, and written about for years, has become an exercise in hypothesis testing, and is broken. We're saying here that the brokenness extends to many other quantitative sciences—though notably—we could say significantly—not much to physics and chemistry and geology. We don't claim to understand fully the sciences we survey. But we do understand their unhappy statistical rhetoric. It needs to change.