Journal of Economic Perspectives—Volume 22, Number 4—Fall 2008—Pages 199–216

Retrospectives
Guinnessometrics: The Economic Foundation of "Student's" t

Stephen T. Ziliak

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please write to Joseph Persky, c/o *Journal of Economic Perspectives,* Department of Economics (M/C 144), University of Illinois at Chicago, 601 South Morgan Street, Room 2103, Chicago, Illinois 60607-7121.

Introduction

In economics and other sciences, "statistical significance" is by custom, habit, and education a necessary and sufficient condition for proving an empirical result (Ziliak and McCloskey, 2008; McCloskey and Ziliak, 1996). The canonical routine is to calculate what's called a t-statistic and then to compare its estimated value against a theoretically expected value of it, which is found in "Student's" t table. A result yielding a t-value greater than or equal to about 2.0 is said to be "statistically significant at the 95 percent level." Alternatively, a regression coefficient is said to be "statistically significantly different from the null, $p \leq .05$." Canonically speaking, if a coefficient clears the 95 percent hurdle, it warrants additional scientific attention. If not, not.

The first presentation of "Student's"test of significance came a century ago, in "The Probable Error of a Mean" (1908b), published by an anonymous "Student." The author's commercial employer required that his identity be shielded from competitors, but we have known for some decades that the article was written by William Sealy Gosset (1876–1937), whose entire career was spent

■ *Stephen T. Ziliak is Professor of Economics, Roosevelt University, Chicago, Illinois. His e-mail address is ⟨sziliak@roosevelt.edu⟩.*

at Guinness's brewery in Dublin, where Gosset was a master brewer and experimental scientist (E. S. Pearson, 1937). Perhaps surprisingly, the ingenious "Student" did not give a hoot for a single finding of "statistical" significance, even at the 95 percent level of significance as established by his own tables. Beginning in 1904, "Student," who was a businessman besides a scientist, took an *economic* approach to the logic of uncertainty, arguing finally that statistical significance is "nearly valueless" in itself (Gosset, 1937, quoted in E. S. Pearson, 1939, p. 244).

Though "Student's" revolutionary article was published in Francis Galton's and Karl Pearson's *Biometrika,* it was at first ignored. Problems of small sample inference were rarely experienced or heeded by members of the then-dominant Pearson school of statistics. Not, that is, until Ronald A. Fisher (1890–1962), the great mathematical geneticist and statistician, picked up "Student's" *t* and made it central to biometrics, economics, medicine, and other sciences.

Ronald Fisher, who is considered by some to be the greatest biologist since Charles Darwin, is also known as a great champion of "Student's" test of significance. Economists, like most other scientists, learned about "Student's" test from Fisher (Savage, 1971, pp. 441–442), and in fact eminent statisticians have long considered Fisher to be the chief architect and visionary of modern statistics—small sample and other—period (for example, Hotelling, 1951; Kendall, 1978; Efron, 1998). Fisher's *Statistical Methods for Research Workers* (1925c), the book which first introduced "Student's" methods to a world audience, remains extraordinarily influential in shaping how scientists use and interpret "Student's" test.

Yet against "Student's" wishes and periodic warnings, it was this same extraordinary Fisher, "Student's" younger friend and colleague, who invented and campaigned for the 5 percent rule of statistical significance. Today, Fisher's preferred interpretation of "Student's" test is customary if not enforced in most sciences, journals, and even courts of law.[1] Meantime, most books on method continue to give little or no space at all to "Student" himself. Yet real economic and philosophical differences between "Student's" theory of inference and Fisher's reformulation of it were scarcely acknowledged by Fisher. He did not discuss or even mention "Student's" rather fundamental objections to Fisher's use and interpretation of *t*.

At the centenary mark of "Student's" path-breaking article, a first-hand study of "Student," his circumstances of discovery, and his actual approach to *t*, might be instructive. "It will be seen then that the difference between Prof. Fisher and myself," "Student" prophesied, "is not a matter of mathematics—heaven forbid—but of opinion" (Student, 1938, p. 367).

[1] Parts of medicine and psychology have long enforced a more stringent, 1 percent rule; other sciences, such as parts of economics and sociology, have customarily accepted a looser, 10 percent rule (Ziliak and McCloskey, 2008, chaps. 5, 6, 11, and 14).

William S. Gosset ("Student")

Source: Annals of Eugenics (1939/9), frontispiece, Blackwell Publishing. Reprinted with permission.

"Student"—A Porter Brewer's Tale

"Student" is the pseudonym used in 19 of 21 published articles by William Sealy Gosset, who was a chemist, brewer, inventor, and self-trained statistician, agronomer, and designer of experiments (Student, 1942). William Gosset was born in 1876 in Canterbury, England. He was cradled in the gentry and educated at Winchester School and Oxford University, New College.[2] The great unknown of statistical science worked his entire adult life—1899 to 1937—as an experimental brewer for one employer: Arthur Guinness, Son & Company, Ltd., Dublin, St. James's Gate. Gosset was a master brewer and rose in fact to the top of the top of the brewing industry: Head Brewer of Guinness. By all accounts it was a happy and comfortable life for the Canterbury lad who was widely admired and loved. A friend from childhood recalled that Gosset possessed "an immovable foundation of niceness" (McMullen, 1939, p. 208). Friends and colleagues such as Karl Pearson, Ronald Fisher, Egon S. Pearson, Florence David, Udny Yule, and others wholeheartedly agreed. Said Yule, "He is a very pleasant chap. Not at all the autocrat of the *t* table" (quoted in Kendall, 1952, p. 159).

[2] Guinness Archives, GDB/C004.06/0001.04 (File: "William Sealy Gosset, Memoranda and correspondence regarding Gosset's recruitment to Junior Brewer").

Gosset was in 1899 an energetic—if slightly loony—23 year-old gentleman scientist.[3] He possessed a wickedly fertile imagination and more energy and focus than a St. Bernard in a snowstorm. An obsessive observer, counter, cyclist, and cricket nut, the self-styled brewer had a sizzle for invention, experiment, and the great outdoors. Guinness was giving its brewery a radical makeover, and the nice chemist was one of the men it brought in to help.

The look-touch-and-sniff approach of eighteenth-century craft brewers had satisfied the Guinness bottom line since its founding in 1759. No longer. The extent of Guinness's market was in a sense already large—in fact, the largest in the world. By 1914, annual production at Gosset's brewery would surpass 2 million hogsheads or 0.84 billion pints.[4] But until the 1890s, the extent of the market was limited by the guildsmen's division of labor (Dennison and MacDonagh, 1998, p. 23, 38; O'Grada, 1995, pp. 304–5).

The Guinness future was in "scientific brewing"—large-scale, industrially controlled brewing—wherein all factors of production, from barley breeding to taste testing, are controlled, improved, and confirmed by experimental science. A degree from either Oxford or Cambridge in a natural science was a minimum requirement for a Guinness brewer in the new era (Dennison and MacDonagh, 1998, chap. 6), whereas family ties had secured employment in the past. Danish and German breweries were transforming similarly—perhaps even a little ahead of Guinness's imposing pace. For discussion of the early history of Guinness, and for industry background, see Lynch and Vaizey (1960), Mathias (1959), and Bamforth (2002).

But Guinness took brewing science to another level—the economic level. Edward Cecil Guinness (1847–1927) was the extraordinary capitalist here. Himself a master brewer and one-time sheriff of Dublin, the first Lord Iveagh went on to become a legendary entrepreneur, philanthropist, chief executive officer, and long-time Chairman of the Board of Guinness (Dennison and MacDonagh, 1998, pp. 5–15). His great brewery was unique in a number of ways, and especially—for the future development of statistics—by vesting selectively chosen brewer-scientists, such as Gosset, with economic authority. With scientific brewers in managerial position, the theory was, experiments could shine a light on the bottom line, and the bottom line, a light on the experiments.

On August 4, 1899, C. D. La Touche, then managing director, recorded in a note that "Mr Gosset" graduated Winchester as "Scholar of New College [Oxford University], [earning a] 1st Class in Mathematical Moderations, Trinity Term 1897, and 1st Class in Chemistry, July 1899. He is short-sighted and wears spectacles," added La Touche. "Seems generally speaking suitable."[5]

[3] "Loony" is a trait he never outgrew: see, for example, how Student (1927 [1942], p. 145, footnote) uses two kangaroos and a platypus to explain the meaning of kurtosis.

[4] Guinness Archives, GDB C004.06/0016 (File: "Comparative Statement and Summary of Financial Operations"); author calculations.

[5] C. D. La Touche, August 4, 1899, in GDB/C004.06/0001.04 (File: "William Sealy Gosset, Memoranda and correspondence regarding Gosset's recruitment to Junior Brewer").

From 1899 to 1906, Gosset was Apprentice Brewer, mostly in the "Experimental Brewery," a miniature brewery close to the Main.[6] In 1904, he began to tackle the problem of making an inference from small samples of malt and hops, two of the crucial inputs to the beer. With Board endorsement, he spent in 1906–1907 a sabbatical year at Karl Pearson's Biometrics Laboratory, University College London, where his general logic of small samples—and his groundbreaking article on "The Probable Error of a Mean"—fermented (Student, 1908b; E. S. Pearson 1990, pp. 47–48). Gosset married in 1906 a national field hockey player and coach, Marjory Surtees Phillpotts, with whom he sired three children.

He signed his published articles and notes with a pseudonym, which was standard business practice. Search his published works for references to beer and odds are you'll fail. Why he chose to be known as "Student" is not known for sure but here is a clue: besides being a humble man, there is stamped on the cover of his early notebook a manufacturer's imprint, "The *Student's* Science Notebook, Eason and Son, Ltd., Dublin and Belfast."[7]

In 1907, Gosset returned to Dublin as Head Experimental Brewer, a position he held through calm and turbulent times until 1935. (He volunteered to fight in the First World War, but, like Fisher, his application was denied by reason of short-sightedness.) In the early 1920s, Gosset became Head of the Statistics Department he established at Guinness. His first task: to estimate the effect of their first-ever advertising campaign on beer sales in Scotland. In 1935, Gosset was promoted to Head Brewer in charge of the new plant at London, Park Royal (now closed), and in September, 1937, he was appointed Head Brewer of all Guinness.[8] Pounding out up to 100 million gallons of Guinness annually, Gosset introduced the quantitative side of scientific brewing, and with it, a storehouse of statistical and experimental theory and tools.

On Guinness's clock, the anonymous "Student" invented or inspired solution concepts that today represent about a dozen different research programs in statistics, econometrics, agronomy, decision science, brewing, and industrial quality control.[9]

Gosset died of a heart attack, his third in close succession, in Beaconsfield, England, October 16, 1937.

[6] Guinness Archives, GDB/C004.06/0001.04 (File: "William Sealy Gosset, Recruitment papers").

[7] Folder 282, "Student's Haemocytometer Paper on Yeast-cell Counts," 1905–1907, in the Pearson Papers, UCL. Dennison and MacDonagh (1998, p. 90, footnote 9) suggest the "Student" name originates with the managing director La Touche, but the Guinness historians show no evidence for it, and apparently they were not aware of "Student's" eponymous notebook, located in London.

[8] Guinness Archives, GDB/C004.09/0004.14 (File: "Research Papers from Second Volume of the Brewery History—Brewers and Directors"); Letter of W. S. Gosset to E. Somerfield, October 1, 1937, GDB/BR01/0964 (File: "Correspondence").

[9] Harold Hotelling (1930, p. 189), a vice president of the American Statistical Association and a teacher of many leading economists and econometricians, wrote: "I have heard guesses in this country, identifying 'Student' with Egon S. Pearson and the Prince of Wales."

The Economic Origins of "Student's" t

Porter is a name given to a dark and bitter beer with a good head. In the nineteenth and early twentieth century, porter was alternatively called "stout" or "stout porter." Stout is the name used now to describe a beer such as Guinness that is bitter on the up-take (bitterness being a function of both the quantity and quality of hops added per barrel of malt) and yet smooth, mellow, and slightly smoky on the finish. Black-ruby tint arises like the smoky finish from roasted barley or "malt"—the distinguishing ingredient of stout.

Brewing "experimentally" introduced challenges and trade-offs. For example, in Gosset's time, Guinness stout was a completely natural and unpasteurized beer. In keg, cask, or bottle, the life of a natural beer is numbered in days. Yet Guinness's beer was shipped worldwide, on an increasingly large scale. Hop is a natural and effective preservative, but it is bitter, and brings bacteria and pests. A heavily hopped beer, such as "Foreign Extra Stout," would last longer than the light hopped, but it would continue to condition on the ocean voyage, becoming increasingly bitter. This was a trade-off to be estimated. In 1911, Guinness carried out an international taste test by shipping bottles of Foreign Extra Stout to agents located in Rio de Janiero, Auckland, San Francisco, and other cities. The object was to "ascertain whether any distinctions could be drawn as regards behaviour, . . . acidity in bottle, or flavour" (Case, 1911, p. 284).

Another challenge was that Guinness, a wholesale dealer only, pursued an unusual pricing strategy. The company tried to hold constant the nominal product price, and many years it did (measured in Dublin prices).[10] At the same time, between 1887 and 1914, output more than *doubled*. Plant size expanded too and with it, the capital/output ratio. So the crucial question facing scientific brewers was: how can experimental science advance economies of scale in brewing? And how can inferential statistics help to control and improve product, while at the same time, help to reduce average total costs for the firm?

Take the choice of hops, for example. In 1898, Guinness used 4.72 million imperial pounds of the fruitful yellow cone. The traditional method of choosing hops based on looks or fragrance wasn't efficient or reliable on this large scale. But was it any more reliable to take small samples out of a larger quantity of hops, test them for certain key characteristics, and then draw an inference about the general quality of the whole lot?

In 1898, Thomas B. Case, Guinness's first scientific brewer, led a team to address this question. Case and his team felt that the key characteristic was the degree of soft resins in the hops. Thus, Case analyzed the average percentage of soft and hard resins found in samples of 50 grams taken from several seasons of American and Kent hops (Case, 1898, p. 47). He compared his samples with those of a cooperating scientist named Briant. For example, Case examined 11 samples

[10] From 1887 to the First World War, the average price of Guinness stayed between 2.6 and 2.8 British pounds per hogshead (one hogshead = 52 U.S. gallons): "Comparative Statement and Summary of Financial Operations," GDB C004.06/0016, Guinness Archives; author calculations.

of hops (n = 11) from Kent in 1897, finding on average 8.1 percent soft resins content. Briant examined 14 samples drawn from the same lot, finding 8.4 percent soft resins—a difference of 0.3 percentage points. The mean difference between their two samples of "American, 1895" was even higher, at 0.7 percent (soft resins) and 1.0 percent (hard).

How should these results be interpreted? Case wrote: "We could not . . . support the conclusion that there are no differences between pockets of the same lot." He worried about "defects" in his procedure, especially the "difficulty of sampling." The more fundamental difficulty, however, was a lack of knowledge about inferential statistics, period. Lacking a theory of how to draw inferences from small samples, Case had no basis for evaluating whether observed differences represented random error from the samples or actual differences in the population.

Other experiments were conducted on barley yield, malt extract, and kiln-drying. In 1899, Gosset was recruited by Case, and by 1904 the Canterbury lad was determined to solve the inference problem. Economic considerations tended to leave the brewers with small samples, Gosset noticed, time and again. In barley yield experiments, for example, it was common to test new varieties and treatments on plots which were 2–4 acres in size. Experimental plots could be further subdivided, a practice which Gosset encouraged and labeled "the principle of maximum contiguity" (Student, 1942, p. 95). The constraint on samples available for analysis is in any case understandable given the opportunity cost of the commercial farmers who where commissioned by Guinness to run the experiments, not to mention the cost of the Guinness scientists' own time.[11] Mixing new beers and malts took time, too, placing a limit on the size of those samples. As Gosset (1962, Letter no. 1, Sept. 15, 1915, emphasis added) wrote to Ronald Fisher in 1915: "Experiments naturally required a solution of the mean/S.D. problem and the Experimental Brewery which concerns such things as the connection between the analysis of malt or hops, and the behavior of the beer, *and which takes a day to each unit of the experiment, thus limiting the numbers.*"

In early November, 1904, Gosset discussed his first breakthrough in an internal report entitled "The Application of the 'Law of Error' to the Work of the Brewery" (Gosset, 1904 *Laboratory Report,* Nov. 3, 1904, p. 3). Gosset (p. 3) wrote:

> Results are only valuable when the amount by which they probably differ from the truth is so small as to be insignificant for the purposes of the experiment. What the odds should be depends—
> 1. On the degree of accuracy which the nature of the experiment allows, and
> 2. On the importance of the issues at stake.

Two features of Gosset's report are especially worth highlighting here. First, he suggested that judgments about "significant" differences were not a purely probabilistic exercise: they depend on the "importance of the issues at stake." Second,

[11] Student (1931c, p. 1342); reprinted in Student (1942, p. 150) and Beaven (1947, p. 164).

Gosset underscored a positive correlation in the normal distribution curve between "the square root of the number of observations" and the level of statistical significance. Other things equal, he wrote, "the greater the number of observations of which means are taken [the larger the sample size], the smaller the [probable or standard] error" (p. 5). "And the curve which represents their frequency of error," he illustrated, "becomes taller and narrower" (p. 7).

Since its discovery in the early nineteenth century, tables of the normal probability curve had been created for large samples; Stigler (1986, 1999) offers a useful early history of the normal distribution. The relation between sample size and "significance" was rarely explored. For example, while looking at biometric samples with up to thousands of observations, Karl Pearson declared that a result departing by more than three standard deviations is "definitely significant."[12] Yet Gosset, a self-trained statistician, found that at such large samples, nearly everything is *statistically* "significant"—though not, in Gosset's terms, economically or scientifically "important." Regardless, Gosset didn't have the luxury of large samples. One of his earliest experiments employed a sample size of 2 (Gosset, 1904, p. 7) and in fact in "The Probable Error of a Mean" he calculated a *t* statistic for n = 2 (Student, 1908b, p. 23).

Gosset's analysis focused on malt extract, which was measured in "degrees saccharine" per barrel of 168 lbs. malt.[13] At the time, an extract in the neighborhood of 133° gave the targeted level of alcohol content for Guinness's beer. A higher extract affected the life of the beer, and also the alcohol content—which in turn affected the excise tax paid on alcoholic beverages. In Gosset's view, ±.5 was a difference or error in malt extract level which Guinness and its customers could swallow. "It might be maintained," he said, that malt extract "should be [estimated] within .5 of the true result with a probability of 10 to 1" (p. 7). Using mean differences of extract values with samples drawn from the Main and Experimental breweries, he then calculated the odds of observing the stipulated accuracy for small and large numbers of observations (p. 7). He found:

Odds in favour of smaller error than .5 [are:]	
2 observations	4:1
3 "	7:1
4 "	12:1
5 "	19:1
82 "	practically infinite

Thus, Gosset (p. 8) concluded, "In order to get the accuracy we require [that is, 10 to 1 odds with .5 accuracy], we must, therefore, take the mean of [at least] four determinations."

[12] *K. P. Lectures Volume I* [Gosset's Classroom Notebook], p. 13, 1906–7, in the Pearson Papers, Gosset file, UCL.

[13] The formula is: Malt extract = ([Specific gravity of the wort] − 1000) x 4.67. See Alan Jackson (with the assistance of W. S. Gosset), "The Relationship between Laboratory and Brewery Extracts, Introduction and Part I," *Laboratory Report,* Vol. 7, Oct. 25, 1906, p. 2.

The report, showing how to achieve quality control of output through small sample estimation, was instantly hailed by the Board. But Gosset himself wasn't convinced: "We have been met with the difficulty," he cautioned, "that none of our books mentions the odds, which are conveniently accepted as being sufficient to establish any conclusion." He said (p. 12), "It might be of assistance to us to consult some mathematical physicist on the matter." Perhaps, Gosset thought, there is a conventional rule about how to set the level of significance, and a mathematical physicist might be able to tell him. Board Endorsement No. 62, signed March 9, 1905, explains: "Mr. Case will make arrangements for Mr. Gosset to have an interview with Prof. Karl Pearson."[14]

Professor Pearson, who was the leading figure in the era before Fisher, was willing to meet Mr. Gosset at Pearson's summer home, in July 1905. His only request was that Gosset write him a letter in advance, detailing his question (for a discussion of Pearson's life, Porter (2004) is a fine contemporary starting point). In the letter, Gosset actually answers the very question he poses. Gosset's letter of 1905 to Karl Pearson (quoted in E. S. Pearson 1939, pp. 215–216; first italics in original) invents an economic approach to the logic of uncertainty:

> *My original question and its modified form.* When I first reported on the subject [in the 1904 internal memo], I thought that perhaps there might be some degree of probability which is conventionally treated as sufficient in such work as ours and I advised that some outside authority in mathematics should be consulted as to what certainty is required to aim at in large scale work. However it would appear that in such work as ours *the degree of certainty to be aimed at* must depend on the *pecuniary advantage to be gained by following the result of the experiment, compared with the increased cost of the new method, if any, and the cost of each experiment.* This is one of the points on which I should like advice.

Notice that in Gosset's view, setting the "degree of probability" to be "treated as sufficient" is not to be made "conventionally" or by "some outside authority in mathematics." Instead, the "degree of certainty to be aimed at," Gosset wrote, depends on the opportunity cost of following a result as if true, added to the opportunity cost of conducting the experiment itself. Gosset never deviated from this central position.[15]

The July 1905 meeting took place. Pearson apparently did not answer the "modified" question, but Gosset spent the next year working on his logic of small samples while on sabbatical at Pearson's laboratory.

As early as 1904 Gosset began to use regression analysis, a technique he learned by reading Merriman's *Method of Least Squares* (1884) and G. B. Airy's *The*

[14] Also see: W. S. Gosset, "The Pearson Co-efficient of Correlation," *Laboratory Report,* No. 2, Vol. III, August, 30, 1905, in the Guinness Archives.

[15] See, for example, Student (1923, p. 271, paragraph one) and Student (1931c, p. 1342, paragraph one) reprinted in Student (1942, p. 90 and p. 150).

Theory of Errors of Observation (1861).[16] In 1908 he used regression to revisit Case's 1898 "hops input" versus "life of beer" question. In a fantastic analytical leap, Gosset (1908, p. 145)—assisted only by slide rule and a mechanical calculator—estimated parabolas of the form

$L = A + BH^2$, where
L = life [of beer] in days
A = life in days of no-hopped beer depending on conditions
H = lbs. of hops, and
B = a constant [slope parameter] depending on the hops and on the conditions.

"No-hop brewings" (A) could survive between "12.2 and 16.7 days" he found after numerous repetitions of the experiment under same and different conditions, whereas "hopped" (with H) could live for a month or beyond.

To Gosset's persistently economic approach, the effect sizes he estimated with "life" regressions were only the beginning. A few months later he noted the "advantage to replace a vague character like increase in life [the dependent variable] . . . by a single definite value which can be directly converted into £ *s. d.*"—pounds, shillings, and pence (Gosset, 1909, *Laboratory Report*, p. 211). From partial correlations that he had calculated on the percentage of soft resins and "[brewing] value," he argued, "We can find an equation giving the probable value for any given percentage of soft resins. The equation is $V = 2.82 + 10.78S$, where V is the per cent Value compared with the 'standard' hop, and S is the soft resins measure from 9 per cent . . . . [E]ach one per cent of soft resins makes a difference of 10.78 per cent. in the value of the hops. This," he said, "at the average [1909] price of hops, represents about 8*s.* per [hundredweight]." At a 1909 input of 6.79 million pounds of hops, Gosset discovered a big economic difference.

"The probable error of the prediction is large," Gosset however cautioned, "being about 6.6 per cent." But the noisy resins variable did not stop Gosset from making a judgment about resins' *economic* importance to brewing value. "Of this [probable error] some 3.2 per cent. is due to errors of analysis and sampling," he said, "leaving [a residual experimental error] due to brewing errors and other factors not included in the analysis" (p. 212). But with the new if imperfect and noisy method of making inferences from small samples, Gosset was able to reject about one-third of the "standard" hops that unscientific methods had previously commended (pp. 212–13). Again the Board cheered.

What Gosset did later for the bottom line in barley is hopped up by several orders of magnitude (Ziliak, 2007). In variety trials designed and co-conducted with E. S. Beaven (1857–1941), Gosset biometrically proved the commercial value of three varieties that would eventually be grown on "well over five million acres of [English] land" (Beaven, 1947, p. xiv). The Gosset-enhanced barleys were used to brew Guinness, but also to make breakfast cereals and feed livestock.

[16] Gosset (1904, *Laboratory Report*, op. cit. p. 4, footnote).

Precision matters, as in soft resins and malt extract. But high statistical significance ranked low in "Student's" ordering. "Student" was looking for practical guidelines on how to improve and maintain beer quality and production without raising the costs. He would not have dreamed of stopping an experiment on grounds that a result reached or failed to reach an arbitrary level of statistical significance.

Additional examples of Gosset's disregard for the 5 percent rule are easy to cite. A few more shall suffice. In 1927, when Fisher was actively promoting the 5 percent rule, Gosset considered "p=.13" a "fairly good fit" (Student, 1927 [1942], p. 147).

On May 18, 1929, Gosset wrote a letter to Egon S. Pearson (1895–1980), repeating to the son the same message that had been delivered to the father (as reprinted in Pearson 1939, p. 244; italics added).

> I fancy you give me credit for being a more systematic cove than I really am ["Student" said] in the matter of limits of significance. What would actually happen would be that I should make out P_t (normal) and say to myself "that would be about 50:1; pretty good but as it may not be normal we'd best not be too certain". . . and whether one would be content with that or would require further work would *depend on the importance of the conclusion and the difficulty of obtaining suitable experience.*

In 1937, Gosset again wrote to Egon (who was by then the editor of *Biometrika* and also the chair of University College of London's brilliant Statistics Department), this time using no uncertain terms (as quoted in E. S. Pearson 1939, p. 244; emphasis in original):

> Obviously the important thing . . . is to have a low real error, not to have a [statistically] "significant" result at a particular station. The latter seems to me to be nearly valueless in itself . . . . Experiments at a single station [that is, tests of statistical significance on a single set of data] *are* almost valueless . . . . You want to be able to say not only "We have significant evidence that if farmers in general do this they will make money by it," but also "we have found it so in nineteen cases out of twenty and we are finding out why it doesn't work in the twentieth." To do that you have to be as sure as possible which is the 20th—your real error must be small.

In short, "Student" saw statistical significance at any level as being "nearly valueless" in itself.[17]

[17] Few realize that Neyman–Pearson "power" can also be traced to "Student," though Egon Pearson himself credited him. "Student" intuited the idea of power in two letters of May 1926 to Egon Pearson (as told in E. S. Pearson, 1966, pp. 4–11; Pearson 1939, p. 242; Letter no. 1, p. 1, Green Box, Pearson Papers, Egon Collection, UCL).

Fisher's 5 Percent Rule

"Student" was Fisher's older friend and behind-the-scenes mentor. Fisher was in 1912 a student at Cambridge when he inquired in a letter about a mistake in Gosset's "The Probable Error of a Mean" (Box, 1978, pp. 70–1). He proved that Gosset's test statistic should be divided by what he would later call "degrees of freedom" ($n - 1$) not by total sample size (n) (Student, 1925, pp. 105–6). Gosset was grateful, and he asked Karl Pearson to publish the young man's correction in *Biometrika.* Although Gosset and Fisher would not meet in person until 1922, a friendship developed in prolific correspondence, with over 150 scientific letters surviving, mostly from Gosset to Fisher (Gosset, 1962).

Fisher developed his own philosophy of "Student's" methods while claiming to teach the original. For example, in *Statistical Methods,* Fisher (1925c [1941], p. 42, italics supplied) wrote: "The value for which $P = .05$, or 1 in 20, is 1.96 or nearly 2; *it is convenient to take this point as a limit in judging* whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation [said Fisher] *are thus formally regarded as significant.*"

Fisher was a rhetorical magician: notice how his 5 percent rule evolved in consecutive sentences from a rule of convenience to a formal regard. Look again at Fisher's "formal" .05 rule and recall Gosset's beer-life regression, which had a probable error of .066.[18] Following Fisher's rule at Guinness would have brought reduced profit and quality to the firm and customers.

Fisher took a hard line on the 5 percent rule. For example, he wrote (1926, p. 504, italics supplied):

> It is convenient to draw the line at about the level at which we can say: "*Either there is something in the treatment, or a coincidence has occurred* such as does not occur more than once in twenty trials."... Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and *ignore entirely all results which fail to reach this level.*

In 1935, Fisher (1935 [1960], p. 13, italics supplied) declared in his other classic book of methods, *The Design of Experiments*:

> *It is usual and convenient for experimenters to take 5 per cent. as a standard* level of significance, in the sense that they are prepared to *ignore all results which fail to reach this standard,* and, by this means, *to eliminate from further discussion* the greater part of the fluctuations which chance causes have introduced into their experimental results.

[18] The probable error is .6745 times the standard error. Thus by today's measure, the odds of observing Gosset's hops coefficients were less still—far below the 19 to 1 odds required by the rigid 5 percent rule. The odds were good enough for Guinness to justify a gamble on a new hops strategy.

Fisher took from the brewer only what he needed. He readily acknowledged (1935, in Savage, 1971a, pp. 471–2, italics supplied) that testing a null hypothesis did not necessarily provide much information about underlying true values. Said he:

> *A null hypothesis may, indeed, contain arbitrary elements,* and in more complicated cases often does so: as, for example, if it should assert that the death-rates of two groups of animals are equal, without specifying what these death-rates actually are. *In such cases it is evidently the equality rather than any particular values of the death-rates that the experiment is designed to test, and possibly to disprove.*

"Student" did not admire Fisher's null hypothesis test procedure. "Student" focused on estimation, prior probability, alternative hypotheses, power, questions of "how much" (for example, Student, 1926 [1942], p. 126). "Student" himself was a philosophical Bayesian, Jeffreys noticed (1939 [1961], pp. 379–81), and took in his laboratory work a Bayesian and decision-theoretic approach (for example, Gosset, 1909 *Laboratory Report,* pp. 205–6).

Fisher tended to take the opposite approach (Zabell, 1989). He sought retreat from the consequences of being incorrect in judgment, and indeed from measuring or evaluating practical consequences in any currency at all. Said Fisher (1955a, p. 75, italics supplied):

> *Finally, in inductive inference we introduce no cost functions for faulty judgments . . . In fact, scientific research is not geared to maximize the profits* of any particular organization, but is rather an attempt to improve public knowledge *undertaken as an act of faith . . . We make no attempt to evaluate these consequences, and do not assume that they are capable of evaluation in any currency.*

Contrast again "Student," working at the then-largest for-profit brewery in the world: "the degree of certainty to be aimed at must depend on the *pecuniary advantage to be gained by following the result of the experiment, compared with the increased cost of the new method, if any, and the cost of each experiment.*"

Reasons for widespread acceptance of Fisher's 5 percent philosophy are too complex to disentangle here. In Ziliak and McCloskey (2008, chap. 20–23), my coauthor and I make an attempt to do so.[19] But the fact that Fisher's way has been widely accepted is obvious in tables reporting "significant" results in economics and other sciences.

[19] For another example, consider the transfer in copyright holding on *t* itself, 1908 to 1938. To Fisher's 1922 request to "quote" "Student's" revised table, "Student" replied: "As to 'quoting' the table in *Biometrika* it depends just what you mean by quoting . . . I don't think, if I were Editor, that I would allow much more than a reference!" Fisher finally did "quote" "Student's" complete tables, many times over. But ever since Fisher and Yates (1938), which became the major reference book, most researchers are not made aware of "Student's" earlier copyrights on the tables (Student, 1908b, 1914, 1917, 1925). Fisher and Yates (1938, pp. 41–43) did not mention "Student's" copyrights and neither have volumes of textbooks ever since (Ziliak, 2008; Ziliak and McCloskey, 2008, chap. 22).

The Great Experimentalist

At the center of Fisher's statistical approach is a desire for a rule-based philosophy of science. At the center of Gosset's statistical approach is an economic logic and experimentalism aimed at learning about the actual magnitudes holding farm and brewing economies together—guided by the profit motive. Over the years, various eminent statisticians have warned against a too-literal interpretation of Fisher's preferred rules. For example, Arnold Zellner (1984, p. 277), a past-president of the American Statistical Association, wrote: "The rationale for the 5% 'accept–reject syndrome' which afflicts econometrics and other areas requires immediate attention." James O. Berger, in his Fisher Memorial Lecture (2003, p. 4), wrote: "The harm from the common misinterpretation of p = 0.05 as an error probability is apparent." Disputes over how rigidly or loosely to apply a rule of significance seem likely to continue.

But it would be tragic to rescue "Student" from obscurity, only then to pigeonhole him as a simple economic counterpoint to Fisher's 5 percent rule.

William Sealy Gosset is more fairly summarized as a great experimentalist. The difference is academically significant with theoreticians outranking experimentalists even in Gosset's day. For example, Fisher (1922a, 1923, 1925a, 1925b) and Hotelling (1931) are conventionally considered the "real" foundation of *t*.[20]

But Gosset's proof in 1908 reflected the experimentalist's values, which only now are coming into fashion. In Gosset's "Original Small Sample Notebook" (1906–7), the proof of *t* was initially fairly well established by his seminal small sample simulation—what we today call a Monte Carlo simulation—rather than exclusively by abstract, mathematical deduction (Student, 1908b, pp. 13–19; folder 283 ("Original Small Sample Notebook," Pearson Papers, UCL). To simulate the accuracy of his theoretical *t*-distribution, "Student" in 1907 randomly selected 750 samples of $n = 4$ from a $n = 3{,}000$ study of height and finger length correlations by Macdonnell (1901). He describes his Monte Carlo in "The Probable Error of a Mean" article (as discussed by Geweke, 2005, and Hanley, Julien, and Moodie, 2008).

Gosset downplayed his own mathematical ability. In a letter to Karl Pearson he confessed (as quoted in E. S. Pearson, 1968, p. 446), "I don't feel at home in more than three dimensions." He told Fisher (in Gosset, 1962, Letter no. 2), "My mathematics stopped at Maths. Mods. at Oxford [the number theorist G. H. Hardy was a classmate and pal], consequently I have no facility therein." But as Fisher

[20] Student (1908b, p. 19) defined "Student's" z (as he originally called it) as:

$$z = (X' - \mu)/S,$$

where $X_i, \ldots, X_n$ are i.i.d. with normal distribution $N(\mu, \sigma^2)$, $X' = \Sigma (X_i)/n$, and $S^2 = \Sigma (X_i - X')^2/n$, and where n is total sample size. His original table begins at $n = 4$ and ends at $n = 10$ (p. 19). In the September of 1922, Gosset and Fisher agreed to call Gosset's revised and expanded table "t" (Student, 1925, pp. 105–6), which is "Student's" z reduced by $(n-1)$, "degrees of freedom." Said his editor Karl Pearson in a letter: it made little difference whether the standard error was divided by n or the rigorously correct $(n-1)$ "because only naughty brewers take n so small that the difference is not of the order of the probable error!"

(1950, p. 4.330) pointed out, Gosset derived the theoretical Poisson distribution unaware while counting yeast cells, and he was first to illustrate its value in a small sample estimation context (Student, 1907). Evidently a lack of elegance with abstract proofs did not stop the experimentalist from advancing knowledge.

Indeed, Gosset developed his own error-minimizing "balanced" design of experiments, two decades before the publication of Fisher's *Design of Experiments* (Fisher, 1935). Randomization of design, Gosset found, led to larger posterior errors. A "balanced" design—which considers the sources of error from both economic and statistical points of view—tends to lower the real error. An example is when a field has a systematic bias in the soil (such as a "non-linear fertility slope"). If such a bias can be identified and controlled for a priori, but is not, one is likely to go wrong (Gosset, 1936; Neyman and Pearson, 1938). Gosset argued that both randomization and significance fail to supply a rational plan. By 1914, Gosset had enough confidence in his procedures to mail an outline of a textbook for Karl Pearson to review.

Sadly, Gosset never found time to complete his textbook—and then he lost the only surviving copy of it when the Gossets moved back to London, in 1934 (Pearson, 1990, p. 19). An introduction to his seminal economic approach to the design and evaluation of experiments are Student (1911, 1923, 1926, 1931a, 1931c, 1938), Gosset (1936), and Beaven (1947, chap. 26–27, throughout).

Thus, it would seem a short list of Gosset's contributions includes: small sample distribution theory (1904–1932); "Student's" *t* table and test of significance (1906–08, 1922–26, 1931–32); modern Monte Carlo analysis (1907–08); the efficient design of experiments (1908–1937); the economic approach to the logic of uncertainty ("net pecuniary value" and "real error" substituting for a 5 percent or other rule of statistical significance, 1904–1937); and alternative hypotheses and "power" (1926, in two letters he wrote to Egon Pearson). To Fisher's delight, Gosset gave an experimental answer to the "minimum number of genes" question of Darwinian evolution (1933–34, in Student, 1942, pp. 181–91); he pioneered industrial quality control, using regression analysis on repeated experiments to set control standards on a variety of beer inputs, long before W. Edwards Deming introduced statistics and industrial quality control to manufacturers in Japan (early 1900s to early 1920s); and he solved a problem concerning secular change in time series data with an original difference-in-difference method (1908–9, 1914). The list is rather incomplete, as it fails to mention for example the contributions Gosset made to the wealth of Guinness shareholders and indeed to the on-going enjoyment of beer drinkers everywhere.

■ *In the Guinness Archives, I've been ably assisted by Eibhlin Roche, Deirdre Flood, and Clare Hackett (Diageo, St. James's Gate); by Gillian Furlong and Steven Wright at University College London, Special Collections Library; and by staff at the Eckhart Library, University of Chicago, and the Museum of English Rural Life, University of Reading. James Hines, Andrei Shleifer, and especially Timothy Taylor provided helpful suggestions for the final manuscript. For conversations and email correspondence I've too many to thank by name but*

I must acknowledge Deirdre McCloskey, Jane (Gosset) Galbraith, Arnold Zellner, Jim Ziliak, and faculty members at the 2006 holiday party for the University College Dublin Department of Economics. Participants in an August 2007 seminar held at the National University of Ireland, Galway, Department of Economics, were brilliant and encouraging. This paper was delivered there and in a special session on the centennial of "Student's" t *which was held at the 24th meeting of the International Biometric Society, University College Dublin, July 16, 2008. Any errors are my own.*

References

Primary Sources

(detailed citations in text)

Beaven Collection. The Museum of English Rural Life, University of Reading.

Case, Thomas B. 1898. "Hops, Season 1897." *Laboratory Report,* Vol. 1, Aug. 31, Arthur Guinness, Son & Co., Ltd., pp. 47–57.

Case, Thomas B. 1911. "Comparison of Intermediate and Ordinary F.E.S. [Foreign Extra Stout] Under Trade Conditions." *Laboratory Report,* Vol. 12, Dec. 13th, Arthur Guinness, Son & Co., Ltd., Board Endorsement No. 2868, pp. 284–307.

Gosset, William S. [Student]. 1904. "The Application of the 'Law of Error' to the Work of the Brewery." *Laboratory Report,* Vol. 8, Nov. 3, Arthur Guinness, Son & Co., Ltd., pp. 3–16 and unnumbered appendix.

Gosset, William S. 1908. "The Present Position of Our Knowledge of the Connection between Life and Hops in the Experimental Brewery." *Laboratory Report,* Vol. 10, Oct. 13, Arthur Guinness, Son & Co., Ltd., pp. 137–150.

Gosset, William S. 1909. "The Brewing of the Experimental Hop Farm Hops, 1907 Crop (Part II), Together with a Note on the Present Method of Hop Analysis." *Laboratory Report,* Vol. 10, Jan. 19, Arthur Guinness, Son & Co., Ltd., pp. 202–220.

Gosset, William S. 1962. *Letters of William Sealy Gosset to R. A. Fisher.* Vols. 1–5, Eckhart Library, University of Chicago. Private circulation.

Gosset Collection. Guinness Archives, Dublin (Diageo).

Guinness Archives. Especially the fifteen volumes of *Laboratory Reports,* 1898–1912. (Dublin (Diageo).

Pearson Papers (Karl Pearson, Egon Pearson, Gosset, Fisher, Neyman files). University College London (UCL), Special Collections.

Secondary Sources

Airy, George B. 1861. *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.* London: MacMillan and Co.

Bamforth, Charles W. 2002. *Standards of Brewing: A Practical Approach to Consistency and Excellence.* Boulder, CO: Brewers Publications.

Beaven, Edwin S. 1947. *Barley: 50 Years of Observation and Experiment.* London: Duckworth.

Berger, James O. 2003. "Could Fisher, Jeffreys and Neyman Have Agreed on Testing?" *Statistical Science,* 18(1): 1–32.

Boland, Philip J. 1984. "A Biographical Glimpse of William Sealy Gosset." *The American Statistician,* August, 38(3): 179–83.

Box, Joan Fisher. 1978. *R. A. Fisher: The Life of a Scientist.* New York: Wiley & Sons.

Dennison, S. R., and Oliver MacDonagh. 1998. *Guinness 1886–1939.* Cork: Cork University Press.

Efron, Bradley. 1998. "R. A. Fisher in the 21st Century." *Statistical Science,* 13(2): 95–122.

Fisher, Ronald A. 1915. "Frequency Distribution of the Values of the Correlation Coefficients in Samples from an Indefinately Large Population." *Biometrika,* 10(4): 507–21

Fisher, Ronald A. 1921. "On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample." *Metron,* 1(4): 3–32.

Fisher, Ronald A. 1922a. "On the Mathematical Foundations of Theoretical Statistics." *Philospophical Transactions of the Royal Soocity of London,* A, 222: 309–368.

Fisher, Ronald A. 1922b. "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P." *Journal of the Royal Statistical Society,* 85(Jan.): 87–94.

Fisher, Ronald A. 1923. "Note on Dr Burnside's recent paper on errors of observation." *Proceedings of the Cambridge Philosophical Society,* 21: 655–58.

Fisher, Ronald A. 1925a. "Applications of 'Student's' Distribution." *Metron,* 5(3): 90–104.

Fisher, Ronald A. 1925b. "Expansion of 'Student's' Integral in Powers of n^{-1}." *Metron* 5(3): 110–120.

Fisher, Ronald A. 1925c [1941]. *Statistical Methods for Research Workers.* New York: G. E. Stechart and Co. Eighth of thirteen editions. (Originally published Edinburgh: Oliver and Boyd.)

Fisher, Ronald A. 1926. "Arrangement of Field Experiments." *Journal of Ministry of Agriculture,* 33(9): 503–13.

Fisher, Ronald A. 1935. *The Design of Experiments.* Edinburgh: Oliver & Boyd. (Reprinted in eight editions and at least four languages.)

Fisher, Ronald A. 1939. " 'Student'." *Annals of Eugenics,* no. 9, pp, 1–9.

Fisher, Ronald A. 1950. *Contributions to Mathematical Statistics,* ed. Walter A. Shewhart. New York: Wiley.

Fisher, Ronald A. 1955. "Statistical Methods and Scientific Induction." *Journal of the Royal Statistical Society, Series B (Methodological),* 17(1): 69–78.

Fisher, Ronald A. 1956 [1959]. *Statistical Methods and Scientific Inference.* 2nd ed. New York: Hafner.

Fisher, Ronald A., and Frank Yates. 1938 [1963]. *Statistical Tables for Biological, Agricultural and Medical Research.* 6th ed. Edinburgh: Oliver and Boyd.

Geweke, John. 2005. *Contemporary Bayesian Econometrics and Statistics.* New York: Wiley.

Gosset, William S. [Student]. 1936. "Co-Operation in Large-Scale Experiments." *Supplement to the Journal of the Royal Statistical Society,* 3(2): 115–36.

Hanley, James A., Marilyse Julien, and Erica E. M. Moodie. 2008. "Student's *z, t,* and *s:* What if Gosset had *R?*" *The American Statistician,* February, 62(1): 1–6.

Hotelling, Harold. 1930. "British Statistics and Statisticians Today." *Journal of the American Statistical Association,* June, 25(170): 186–90.

Hotelling, Harold. 1931. "The Generalization of Student's Ratio." *Annals of Mathematical Statistics,* 2(3): 360–78.

Hotelling, Harold. 1951. "The Impact of R. A. Fisher on Statistics." *Journal of the American Statistical Association,* March, 46 (243): 35–46.

Jeffreys, Harold. 1939 [1961]. *Theory of Probability.* 3rd revised edition. London: Oxford University Press.

Kendall, Maurice G. 1952. "George Udny Yule, 1871–1951." *Journal of the Royal Statistical Society,* 115A: 156–61.

Kendall, Maurice G. 1978. "The History of Statistical Method." In *International Encyclopedia of Statistics,* Vol. 2., ed. W. Kruskal and J. Tanur, 1093–110. Free Press.

Kruskal, William H., and Judith M. Tanur. 1978. *International Encyclopedia of Statistics,* Vols. 1–2. New York: The Free Press.

Lynch, Patrick, and John Vaizey. 1960. *Guinness's Brewery in the Irish Economy, 1758–1876.* London: Cambridge University Press.

Macdonnell, William R. 1901. "On Criminal Anthropometry and the Identification of Criminals." *Biometrika,* 1(2): 177–227.

Matthias, Peter. 1959. *The Brewing Industry in England, 1700–1830.* London: Cambridge University Press.

McCloskey, Deirdre N., and Stephen T. Ziliak. 1996. "The Standard Error of Regressions." *Journal of Economic Literature,* 34(1): 97–114.

McMullen, Launce. 1939. " 'Student' as a Man." *Biometrika,* 30(3/4): 205–10.

Merriman, Mansfield. 1884. *A Text-book on the Method of Least Squares.* New York: John Wiley & Sons.

Neyman, Jerzy. 1961. "Silver Jubilee of My Dispute with Fisher." *Journal of Operations Research,* Society of Japan, 3: 145–54.

Neyman, Jerzy, and E. S. Pearson. 1938. "Note on Some Points on 'Student's' Paper on 'Comparison Between Balanced and Random Arrangements of Field Plots.' " *Biometrika,* 29(3/4): 379–88.

O'Grada, Cormac. 1995. *Ireland: A New Economic History.* Oxford: Oxford University Press.

Pearson, Egon S. 1937. "Mr. W. S. Gosset: 'Student.' " *Nature,* November 14, 140(3550): 838.

Pearson, Egon S. 1939. " 'Student' as Statistician." *Biometrika* 30 (3/4): 210–50.

Pearson, Egon S. 1966. "The Neyman–Pearson Story: 1926–34." In *Research Papers in Statistics: Festschrift for J. Neyman,* ed. F. N. David, 1–24.

Pearson, Egon S. 1968. "Studies in the History of Probability and Statistics. XX: Some Early Correspondence between W. S. Gosset, R. A. Fisher, and Karl Pearson, with Notes and Comments." *Biometrika,* 55(3): 445–57.

Pearson, Egon S. 1990. [posthumous]. *'Student': A Statistical Biography of William Sealy Gosset.* Edited and augmented by R. L. Plackett, with the assistance of G. A. Barnard. Oxford: Clarendon Press.

Pearson, Karl. 1914. *Tables for Statisticians and*

Biometricians, Part I. London: Biometrics Laboratory, University College London.

Porter, Theodore M. 2004. *Karl Pearson: The Scientific Life in a Statistical Age.* Princeton: Princeton University Press.

Savage, Leonard J. 1971 [1976] [posthumous]. "On Re-Reading R. A. Fisher." *Annals of Statistics,* 4(3): 441–500.

Stigler, Stephen. 1986. *The History of Statistics.* Cambridge: Harvard University Press.

Stigler, Stephen. 1999. *Statistics on the Table.* Cambridge: Harvard University Press.

Student. [W. S. Gosset]. 1907. "On the Error of Counting with a Haemacytometer." *Biometrika,* 5(3): 35160.

Student. 1908a. "The Probable Error of a Correlation Coefficient." *Biometrika* 6(2/3): 300–310.

Student. 1908b. "The Probable Error of a Mean." *Biometrika,* 6(1): pp. 1–24.

Student. 1911. "Appendix to Mercer and Hall's Paper on 'The Experimental Error of Field Trials.' " *Journal of Agricultural Science,* 9: 128–131.

Student. 1914. (See Pearson, Karl, 1914.)

Student. 1917. "Tables for Estimating the Probability that the Mean of a Unique Sample of Observations Lies Between $-\infty$ and Any Given Distance of the Mean of the Population from which the Sample is Drawn."*Biometrika,* 11(4): 414–17.

Student. 1923. "On Testing Varieties of Cereals." *Biometrika,* 15(3/4): 271–93.

Student. 1925. "New Tables for Testing the Significance of Observations." *Metron,* 5(3): 105–8.

Student. 1926 [1942]. "Mathematics and Agronomy." *Journal of the American Society of Agronomy,* 18(8): 703–791. (Reprinted in 1942 in *Student's Collected Papers,* ed. E. S. Pearson and J. Wishart, 121–34.)

Student. 1927 [1942]. "Errors of Routine Analysis." *Biometrika,* 19(1/2): 151–64. (Reprinted in 1942 in *Student's Collected Papers,* ed. E. S. Pearson and J. Wishart.)

Student. 1931a. "The Lanarkshire Milk Experiment." *Biometrika,* 23(3/4): 398–406.

Student. 1931b. "On the 'z' Test." *Biometrika,* 23(3/4): 407–8.

Student. 1931c. "Yield Trials." *Bailliere's Encyclopedia of Scientific Agriculture,* pp. 1342–60. (Reprinted in 1942 in *"Student's" Collected Papers,* pp. 150–168.)

Student. 1938 [posthumous]. "Comparison between Balanced and Random Arrangements of Field Plots." *Biometrika,* 29(3/4): 363–78.

Student. 1942 [posthumous]. *"Student's" Collected Papers.* Ed. E. S. Pearson and John Wishart. London: Biometrika Office.

Yates, Frank. 1951. "The Influence of *Statistical Methods for Research Workers* on the Development of the Science of Statistics." *Journal of the American Statistical Association,* 46(253): 19–34.

Yates, Frank, and Kenneth Mather. 1963. "Ronald Aylmer Fisher." *Biographical Memoirs of Fellows of the Royal Society of London,* 9: 91–129.

Zabell, Sandy. 1989. "R. A. Fisher on the History of Inverse Probability." *Statistical Science,* 4(3): 247–63.

Zellner, Arnold. 1984. *Basic Issues in Econometrics.* Chicago: University of Chicago Press.

Ziliak, Stephen T. 2007. *"Student's" Methods: Science before Fisher.* Book in progress. Roosevelt University, Department of Economics.

Ziliak, Stephen T. 2008. "R. A. Fisher and the Copyright History of 'Student's' *t*," Available at author's website: http://faculty.roosevelt.edu/Ziliak (click on "The Great Skew").

Ziliak, Stephen T., and Deirdre N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives.* Ann Arbor: University of Michigan Press.