History of Statistics 7. The t-distribution. W S GOSSET (1876-1937)

Perhaps the most commonly used statistical method is the ***t*-test**. We have William S. Gosset to thank for that. Gosset discovered the formula for the *t*-curve used to get *p*-values in studies that compare means based on small samples. He did this primarily by a careful study of data patterns. At a later date, a younger man, the great mathematical theorist, R. A. Fisher, proved rigorously that Gosset's findings were correct. Fisher was able then to generalize Gosset's original work and present a unified system for comparing means of several groups.

W S Gosset aka "Student" (1908)

The turn of the 20th century saw tremendous growth in statistical methodology, mostly the work of mathematicians in England. The development of the *t*-test was part of that growth. Three prominent names are linked in this history: **William Sealy Gosset** (who wrote under the pseudonym, "**Student**," so that readers would not know he was a scientist at Guinness Brewery); **Karl Pearson** (older, established leading professor of Statistics in London and editor of the prestigious statistical journal *Biometrika*); and **R.A. Fisher** (young brilliant mathematician, soon to be the greatest mathematical statistician of his time).

Gosset graduated with honors in both mathematics and chemistry from Oxford University. Shortly after, in 1899, when he was 23 years old, he was hired by Guinness Brewery in Ireland as a brewer because he was a *chemist*. It was almost unheard of for a commercial business to hire someone who was "just" a mathematician. It was a big deal to be appointed as a brewer at Guinness. In 1886 they had gone public and in the reorganization was moving away from the older style of apprenticeship for brewers to making brewing more "scientific," and professional chemists were highly valued. Guinness hired a half dozen young bright chemists, one of whom was Gosset, into senior management positions.

Guinness was then the world's largest brewery. They needed farmers to grow consistent and superior crops of barley, and skilled "maltsters" to turn the barley into malt. The whole process for the final product needed to be tightly controlled. These young brewers designed many field and lab experiments to determine the best barley, best hops, best temperatures for brewing, etc.

> **They began to accumulate data and, at once, they ran into difficulties because their measurements varied. The effects they were looking for were not usually clearcut or consistent, as they had expected, and they had no way of judging whether the differences they found were effects of treatment or accident. Two difficulties were confounded: the variation was high and the observations were few.**
>
> **The young research brewers worked well together- some were very close friends. Each seemed to fit into his own role in brewery affairs. And to them it seemed natural to take their numerical problems to Gosset. He had done some mathematics at Oxford and seemed less scared of mathematics than they were.** [Fisher Box, p48]

For Gosset the fundamental new statistical problem was that with such small samples it was unclear how well the standard deviation, ***s***, of the sample reflected the population standard deviation. For example, one farmer might grow four plots of barley but find quite different results in the four plots, while another farmer might grow four plots of a different barley and also get quite different results in his four plots. That made it very difficult to tell if there was any significant difference between the two barleys. There was no way to assign probabilities accurately to observed differences, to distinguish "real" differences from random variation. The contrast between the two types of barley might have been clearer if each farmer had been able to plant more plots, but that was financially not practical.

A more precise statistical analysis depended on knowing the probability distribution of *s* based on small samples, but this distribution was not known. The organization that was doing the best research in statistical theory then was Karl Pearson's Biometric Laboratory at the University of London. Gosset was given leave to leave Ireland and spend the academic year 1906-7 in Pearson's lab to work on this "small sample" issue.

During the sabbatical, Gosset, with Pearson's help, solved the distribution problem and wrote **"The Probable Error of the Mean,"** the paper that introduced the *t*-test. It was published the following year in Pearson's journal, *Biometrika*.

> **There are … experiments, however, which cannot easily be repeated very often; in such cases it is sometimes necessary to judge of the certainty of the results from a very small sample, which itself affords the only indication of the variability. Some chemical, many biological, and most agricultural and large scale experiments belong to this class, which has hitherto been almost outside the range of statistical enquiry.**
>
> **Again, although it is well known that the method of using the normal curve is only trustworthy when the sample is "large," no one has yet told us very clearly where the limit between "l large " and " small " samples is to be drawn.**
>
> **The aim of the present paper is to determine the point at which we may use the tables of the probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few.** (Student, 1908, page 2)

Gosset's paper involves a good deal of mathematics. For example, here is his formula for the probability distribution of the z statistic.

… we have as an equation giving the distribution of *z*

$$y = \frac{\sqrt{\frac{n}{2\pi}}}{\sigma} \frac{\int_0^\infty s^{n-1} e^{-\frac{ns^2(1+z^2)}{2\sigma^2}} ds}{\int_0^\infty s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds}.$$

Using this formula, he produced a table similar to the ones we use today, expecting researchers to use the table, not the formula, to find areas under probability curves. (Now the formula is built into the software on calculators and computers, so most researchers today have no idea what the formula for a *t*-curve looks like.)

Gossett had worked out a good bit of his paper by matching known distribution formulas to patterns he saw in raw data. But he was aware that he was unable to arrive at his formulas from any general principle. "The law of formation … appears to be a simple one, but I have not seen my way to a general proof."

In any case, Gosset included worked out examples using his table. Here is the first published example of a *t*-test. He took a set of published medical data for his illustration.

> **As an instance of the kind of use which may be made of the tables, I take the following figures from a table by A. R. Cushny and A. R. Peebles in the Journal of Physiology for 1904, showing the different effects of the optical isomers of hyoscyamine hydrobromide in producing sleep.** (page 20)

There were only 10 patients in the study, a "small" sample. Student's *t*-test showed that the mean difference in hours slept, 1.58 hours, was large enough to support the idea that medicine 2 was superior to medicine 1. His conclusion sounds very much like one we might see today.

Additional hours' sleep gained by the use of hyoscyamine hydrobromide.

Patient	1 (Dextro-)	2 (Laevo-)	Difference (2–1)
1.	+ ·7	+ 1·9	+ 1·2
2.	− 1·6	+ ·8	+ 2·4
3.	− ·2	+ 1·1	+ 1·3
4.	− 1·2	+ ·1	+ 1·3
5.	− 1	− ·1	0
6.	+ 3·4	+ 4·4	+ 1·0
7.	+ 3·7	+ 5·5	+ 1·8
8.	+ ·8	+ 1·6	+ ·8
9.	0	+ 4·6	+ 4·6
10.	+ 2·0	+ 3·4	+ 1·4
	Mean + ·75	Mean + 2·33	Mean + 1·58
	S. D. 1·70	S. D. 1·90	S. D. 1·17

> **The mean value of this series is + 1.58 while the S.D. is 1.17, the mean value being + 1.35 times the S.D. From the table the probability is .9985 or the odds are about 666 to 1 that 2 is the better soporific. The low value of the S.D. is probably due to the different drugs reacting similarly on the same patient, so that there is correlation between the results. Of course odds of this kind make it almost certain that 2 is the better soporific, and in practical life such a high probability is in most matters considered as a certainty.** (page 21)

Remarkably, but as frequently happens in scientific research, Gosset's ground-breaking paper of 1908 was pretty much ignored. Then, in 1912, R A Fisher, a brilliant graduate student in mathematics at Cambridge University, on the suggestion of his tutor, sent Gosset a paper he was writing about probability frequency curves. Gosset thought he found an error in Fisher's work and wrote back. Fisher, in turn, responded, pointing out an error in Gosset's note.

This exchange of letters was the start of a long, friendly correspondence between the two men, and they became good friends. It seems likely that Gossett helped Fisher get a job at Britain's most prestigious agricultural research center, where he worked for about 13 years, and from which he revolutionized statistics. The two men did not meet in person until 1922, with Gosset in Ireland and Fisher in England. It may seem now that it would be easy for a researcher in Ireland and one in London to meet, but WWI lasted from 1914 to 1918, and the Irish War of Independence against England raged from 1919 to 1921.

In writing back to Gosset, Fisher showed how, by using *n*-dimensional geometry, he had re-derived the z-curve formula Gosset had shown in his 1908 paper, thereby providing the rigorous proof that Gosset had been unable to supply. The *n*-dimensional approach proof by this hot-shot mathematician was over Gosset's head. He sent the letter on to Karl Pearson,

> **Dear Pearson,**
> **I am enclosing a letter which gives a proof of my formulae for the frequency distribution of z (=xs), where x is the distance of the mean of n observations from the general mean and s is the S.D. of the n observations. Would you mind looking at it for me; I don't feel at home in more than three dimensions even if I could understand it otherwise.** (E S Pearson, p 445)

Pearson, in fact, didn't pay much attention to it. He wasn't particularly interested in studies with small samples. And so, for the next thirteen years, until 1925, the *t*-test essentially was ignored by almost everyone except Fisher. But for Fisher it became the building block for almost all the kinds of analyses statisticians still use, especially the generalization of Gosset's test to the broader methods called Analysis of Variance and Multiple Regression.

In 1925 Fisher wrote what became the most influential textbook of statistical methods ever published, *Statistical Methods for Research Methods*. It went through 14 editions, from 1925 to 1970. Here is a quote from the Introduction that honors "Student."

> **It is equally fortunate that the distribution of t, first established by "Student" in 1908, in his study of the probable error of the mean, should be applicable, not only to the case there treated, but to the more complex, but even more frequently needed problem of the comparison of two mean values. It further provides an exact solution of the sampling errors of the enormously wide class of statistics known as regression coefficients. (page 17)**

With the publication of Fisher's book, "Student" finally became well known.

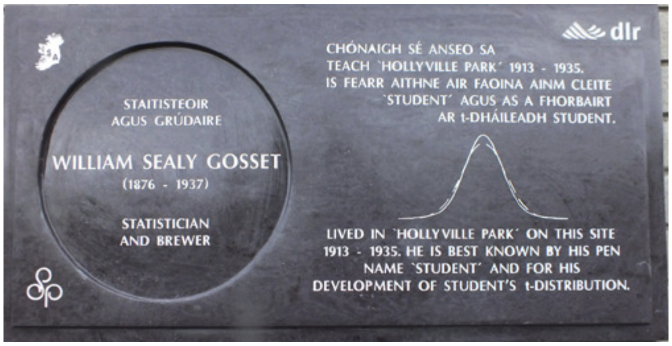
Thorough his whole career Gosset was known as a good-humored, modest man. Here's a characteristic anecdote.

> **Once, in 1937, a young statistician who went to consult him, said pompously: "On behalf of fellow statisticians, I would like to thank you for all that you have done for the advancement of statistics." to which Gosset replied: "Oh, that's nothing – Fisher would have discovered it all anyway ."** (Cunliffe, page 4)

Gosset remained at Guinness his whole career and continued statistical research, not only in brewery issues, but also in agricultural experimental design, and genetics. He moved to London in 1935 as master brewer at a new Guinness brewery. The move let him attend meetings of the Royal Statistical Society, and in 1936 he presented a paper at one of their meetings – the first and only time he wrote under his real name. He died in 1937. In an appreciation of his accomplishments Fisher wrote,

> **THE untimely death of W. S. Gosset, at the age of 61, in October 1937, has taken one of the most original minds in contemporary science. Without being a professed mathematician, he first published, in 1908, a fundamentally new approach to the classical problem of the theory of errors, the consequences of which are still only gradually coming to be appreciated in the many fields of work to which it is applicable.** (Fisher, page 1)

Here's plaque in Dublin that honors him.



Comment on the sources

E. S. Pearson is the son of Karl Pearson. E.S. Pearson was a statistician who developed many of the basic concepts of hypothesis testing. Joan Fisher Box is the daughter of R A Fisher. She married George Box, another important statistician.

Sources

Box, Joan Fisher, "Guinness, Gosset, Fisher, and Small Samples," *Statistical Science* 1987, Vol. 2, No. 1, 45-52
Cunliffe, Stella, Interaction, *Journal of the Royal Statistical Society*. Series A (General), Vol. 139, No. 1, 1976
Fisher, R. A., Student*, Annals of Eugenics*, 1939
Fisher, R A, *Statistical Methods for Research Workers*, 14th edition 1970
Pearson, E S, "Studies in the History of Probability and Statistics. XX Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson", *Biometrika* 1968
Stigler, Stephen. *The Seven Pillars of Statistical Wisdom*, Harvard, 2016
Student. Mathematics and Agronomy, *Journal of the American Agronomy Society*, 1926. As quoted in Tankard.
Student. The Probable Error of the Mean, *Biometrika*, 1908
Tankard, James W. *The Statistical Pioneers*, Schenkman Publishing 1984
Image of Gosset. https://en.wikipedia.org/wiki/William_Sealy_Gosset
Plaque image: http://blogs.roosevelt.edu/sziliak/w-s-gosset-aka-student