WILLIAM SEALY GOSSET

b. 13 June 1876
d. 16 October 1937

Summary. Better known by his pseudonym, "Student," Gosset's name is associated with the discovery of the *t*-distribution and its use. He had a profound effect on the practice of statistics in industry and agriculture.

Introduction

William Sealy Gosset was born in Canterbury, England. He received a degree from Oxford University in Chemistry and went to work as a "brewer" at Arthur Guinness Son and Co. Ltd. in Dublin, Ireland in 1899. He died in Beaconsfield, England at the age of 61, still in the employ of Guinness.

By the circumstances of his work, Gosset was led early in his career at Guinness to examine the relationship between the raw materials for beer and the finished product, and this activity naturally led him to learn the tools of statistical analysis. In 1905, Gosset sought out the advice of Karl Pearson (q.v.) and subsequently spent the better part of a year, in 1906–1907, in Pearson's Biometric Laboratory at University College London, where he worked on small sample statistics problems. Gosset then produced a pair of papers that were published in *Biometrika* in 1908, under the nom de plume, "Student." The first of these derived what we now know as "Student"s' *t*-distribution, and the second dealt with the small sample distribution of Pearson's correlation coefficient. These contributions placed Gosset among the great men of the newly emerging field of statistical methodology. In fact, the *t*-test based on his 1908 paper is perhaps the single most widely used statistical tool in applications.

C. C. Heyde et al. (eds.), *Statisticians of the Centuries*

In the years that followed, Gosset worked on a variety of statistical problems in agriculture, including experimentation. He was in active correspondence with the leading English statisticians of his day, including Karl Pearson, Egon Pearson (q.v.), and R. A. Fisher (q.v.). Gosset's correspondence with Fisher dealt with highly varied topics and was, as Plackett and Barnard note, "interspersed with friendly advice on both sides." In his later years, he had a number of public disagreements with Fisher over the role of randomisation in experimentation. Gosset was a strong advocate of experimental control, a point that came through quite vividly in his proposal in connection with the Lanarkshire milk experiment in "Student" (1931), although in this paper he was also critical of an evaluation of the study carried out by Bartlett and Fisher (1931). In particular, Gosset was enamoured of the use of systematic experimental plans and opposed the use of randomisation. This controversy led Gosset to prepare his final paper ('Student," 1937) published a few months after his death.

In the next section, we comment on some of the technical details of Gosset's seminal 1908 contributions. For further details on Gosset's life and contributions, see Plackett and Barnard (1990). Gosset's writings are collected in "Student" (1942).

Gosset on the Mean and the Correlation Coefficient

Small Sample Theory of the Mean

In 1908, Gosset's work at the Guinness brewery led him to publish the results that would become associated with his name in future generations. In an article entitled "The probable error of a mean" ("Student," 1908a), he established the sampling distributions of s^2 and s for an independent and identically distributed sample of size n from a normal population. He then showed that the mean and the standard deviation of such a sample are uncorrelated and derived what we now know as the "Student" t-distribution.

At time of publication, the importance of these results was not fully recognised. The focus among most contemporary statisticians was on large-sample theory and Gosset's emphasis on small samples, arising from his work at the brewery, set him somewhat apart. In fact, it was not until Fisher generalised the "Student" t-distribution that it came into widespread use outside of the Guinness brewery itself.

Aside from the derivations mentioned above, there are a number of interesting features in the 1908 manuscript. First there is the break from the tradition of the Biometric School, which used the same symbol for both the population parameter and the sample statistic. In Gosset's paper, he uses s^2 for the sample variance and σ^2 for the population variance. Work with large samples had obscured the need for this distinction, which became clearer when the focus was shifted to small samples (E.S. Pearson, 1939).

Another aspect of this paper worthy of note is Gosset's use of a sampling experiment to help empirically solve the problem at hand, instead of finding an analytic solution. The essence of the simulation was the following—using data on the height and left middle finger measurements of 3000 criminals, he generated 750 random samples of size 4. Gosset then calculated the means, standard deviations and correlation coefficient of each sample as well as t-statistics. He plotted the empirical distributions of the latter and compared them to the theoretical ones he had derived. Using χ^2 tests for goodness of fit, he deemed the results to be satisfactory. In connection with this empirical study, Gosset noted that ". . . if the distribution is approximately normal, our theory gives us a satisfactory measure of the certainty to be derived from a small sample . . . " ("Student," 1908a, p. 19). Furthermore, "[i]f the distribution is not normal, the mean and the standard deviation of a sample will be positively correlated, so that although both will have greater variability, yet they will tend to counteract each other, a mean deviating largely from the general mean tending to be divided by a larger standard deviation. Consequently, I believe that the table given . . . below may be used in estimating the degree of certainty arrived at by the mean of a few experiments, in the case of most laboratory or biological work where the distributions are as a rule . . . sufficiently nearly normal" (*ibid*). Gosset's intuition that the t-test would be robust against small departures from normality, while not proven here, would later be verified (E.S. Pearson, 1929; Geary (q.v.), 1936, 1947).

This paper implicitly takes an inverse probability approach, although there is no discussion of prior distributions. We encounter, for instance, statements such as "Thus, to take the tables for samples of 6, the probability of the mean of the population lying between $-\infty$ and once the standard deviation of the sample is 0.9622 . . . " ("Student," 1908a, p. 20). Jeffreys (q.v.) (1937) was later to observe exactly how Gosset's derivation coincided with his own, which was based on inverse probability. Interestingly, a treatment of the small sample theory of the mean from the inverse probability perspective had appeared earlier, in a paper by Edgeworth (q.v.) (1883), who also derived the t-distribution. Edgeworth's derivation, however, was heavily reliant on the form of the prior distribution for μ and σ, which he assumed had the form $C\sigma^{-2}d\mu\, d\sigma$. Gosset appears to have been unaware of this contribution of Edgeworth. Welch (1958) provides a thorough discussion of Edgeworth's 1883 paper and its connection to "Student"s' own work (see also Stigler, 1978). More recently, Pfanzagl and Sheynin (1996) present evidence of an even earlier derivation, in 1876, of a generalization of the t-distribution by the German mathematician Jakob Lüroth, also from what we would call a Bayesian perspective.

The Correlation Coefficient

In addition to the famous article establishing the t-distribution, there appeared in 1908 another effort by "Student," this time dealing with the correlation coefficient ("Student," 1908b). Gosset did not actually succeed in deriving a sampling distribution for r, but relied instead on the empirical method used in the first paper to establish properties of the distribution of r assuming that $\rho = 0$. We note that here the inverse probability approach is much more obviously stated than in Gosset's article dealing with the sample mean; in particular, he suggests various "priors" for ρ, the population correlation. Since he was not able to give a concrete expression for the sampling distribution $f(r|\rho)\,dr$, Gosset was also unable to write down the posterior distribution for ρ. Subsequently, in his first major contribution to mathematical statistics, R. A. Fisher (1915) derived the sampling distribution of r, using a geometrical argument, and this work led to the famous Fisher-Gosset correspondence. In this same paper, Fisher also established that for the case of normal data under consideration, the mean and standard deviation are not only uncorrelated, but independent.

Gosset on Experimental Design

R. A. Fisher's correspondence with Gosset began in 1912, when Fisher sent Gosset a copy of his paper applying maximum likelihood (as it would later come to be known) to estimate the mean and variance of a normal population. They did not meet until a decade later, however, when Gosset visited Rothamsted and presented Fisher with a copy of his statistical tables. They continued to correspond on a variety of topics and, in 1923, there was an exchange of letters between the two on Fisher's work with Mackenzie on the design of experiments, in which Gosset advocated the use of systematic field arrangements, in essence rejecting Fisher's proposal for randomisation. Their disagreement on the use of randomisation continued in private correspondence (see various excerpts in Plackett and Barnard, 1990, Chapter 5) and could hardly be discerned in Gosset's only public criticism of Fisher, in the context of a published comment on the infamous Lanarkshire milk experiment ("Student," 1931).

In 1936, however, the debate became public during the discussion of a paper read before the Royal Statistical Society on "co-operation in large-scale experiments." Gosset led off the discussion by extolling the virtues of Beaven's half-drill strip systematic design, and Fisher, who spoke next, expressed his opposition to such systematic designs. This was followed by a paper by Fisher and Barbacki criticising "the supposed precision of systematic designs' and an exchange of letters between Fisher and Gosset in *Nature*.

At the time of his death, Gosset was working on a detailed response to Fisher in which he once again put forth his support of systematic experimentation and expressed doubts about the role of randomisation. After so many years, they had not resolved their differences on this fundamental statistical issue. The paper appeared posthumously in 1938, and when he read it, Fisher observed in a letter to Harold Jeffreys (Bennett (1990), pp. 271–272):

> So far as I can judge, "Student" and I would have differed quite inappreciably on randomisation if we had seen enough of each other to know exactly what the other meant, and if he had not felt in duty bound, not only to extol the merits, but also to deny the defects of Beaven's half drill strip system.
> . . . I fancy also that Gosset never realised that a fertility gradient when, as in my experience is not very frequent, it is important enough to bother about, can easily be eliminated from a randomised experiment. It is, I think, my fault that I have not made this clear earlier, but until the last two years I had really thought that "Student" accepted all that I had put forward on behalf of randomisation.

— Stephen E. Fienberg and Nicole Lazar

Bibliography

Bartlett, S. and Fisher, R.A. (1931). Pasteurized and raw milk. *Nature*, **127**, 591–592.

Bennett, J.H. (1990). *Statistical Inference and Analysis. Selected Correspondence of R.A. Fisher.* Oxford, pp. 271–272.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507–521.

Geary, R. C. (1936). The distribution of "Student"s' ratio for non-normal samples. *Journal of the Royal Statistical Society Supplement* (superseded by *Series B*), **3**, 178–184.

Geary, R. C. (1947). Testing for normality. *Biometrika*, **34**, 209–242.

Jeffreys, H. (1937). On the relation between direct and inverse methods in statistics. *Proceedings of the Royal Society, Series A*, **160**, 325–348.

Pearson, E. S. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biometrika*, **21**, 259–286.

Pearson, E. S. (1939). "Student" as statistician. *Biometrika*, **30**, 210–250.

Pearson, E. (1990). *"Student": A Statistical Biography of William Sealy Gosset.* (R.L. Plackett and G.A. Barnard, eds.) Clarendon Press, Oxford.

Pfanzagl, J. and Sheynin, O. (1996). Studies in the history of probability and statistics XLIV: A forerunner of the t-distribution. *Biometrika*, **83**, 891–898.

Stigler, S. (1978). Francis Ysidro Edgeworth, Statistician (with Discussion). *Journal of the Royal Statistical Society*, **141**, 287–322.

"Student" (1908a). The probable error of a mean. *Biometrika*, **6**, 1–25.

"Student" (1908b). Probable error of a correlation coefficient. *Biometrika*, **6**, 302–310.

"Student" (1931). The Lanarkshire milk experiment. *Biometrika*, **23**, 398–406.

"Student" (1937). Random and balanced arrangements. *Biometrika*, **29**, 363–379.

"Student" (1942). *"Student" 's Collected Papers.* (ed. by E.S. Pearson and J. Wishart), with a forward by L. McMullen. Biometrika Office, University College.

Welch B. L. (1958). "Student" and small sample theory. *Journal of the American Statistical Association*, **53**, 777–788.