

# William Sealy Gosset, the Student

Michele Nichelatti

Oncology and Haematology Department, Niguarda Ca' Granda Hospital, Milan, Italy

William Sealy Gosset was born in Canterbury, UK, on June 13, 1876. He was born into a Huguenot family that had left France at end of the XVII century when King Louis XIV issued the Edict of Fontainebleau. His father, Frédéric, was a colonel and engineer in the British Army, while his mother, Agnès Sealy Vidal, was a housewife.

William attended Winchester College, a famous English public school, where he displayed an aptitude for mathematics and natural sciences. He graduated from New College, Oxford, in 1899 with a first-class degree in chemistry. In the October of the same year, he began what was to prove a lifelong career at Arthur Guinness & Son, a Dublin brewery. Immediately, Gosset set to work applying a careful scientific analysis method to his study of the company's production processes, from barley production to yeast fermentation, trying to find ways of improving the quality of the final product. At the time, the only fully developed theory of estimation was based on large samples, while nothing had been done to develop a small-sample estimation theory. However, small samples were what Gosset's work at the brewery involved, and he therefore had to develop an entirely new theory by himself. The Guinness business revolved around beer and agriculture, and the company also had a special interest in applied research. There should thus have been nothing unusual in a Guinness researcher publishing a scientific paper. However, some years previously, after an employee had written a paper containing classified information on production processes, Guinness had forbidden all its employees from publishing papers on any topic whatsoever. As a result, Gosset had to write under the pen name of Student. In any case, this company regulation had profound effects on Gosset's future scientific production.

Gosset took a leave of absence in 1906 and 1907 to attend lectures and tutorials (in particular, regarding correlation, non-normal distributions) at the Biomet-

ric Laboratory headed by Karl Pearson at University College London. In London, Gosset began writing his most important paper, which was to be published in 1908 in the journal *Biometrika* under the pen name of Student and contained his most renowned result: "Student's *t*-test" or the *t* distribution. This distribution, obtained for the study of small samples, has an approximately normal shape, with a larger (i.e., flatter) spread than the normal curve. However, it becomes closer and closer to normal distribution as sample size increases. Gosset's discovery was the relationship between sample size and spread of the *t* distribution, given by the degrees of freedom of the sample (i.e.,  $n-1$ ,  $n$  being the sample size). He therefore computed the tables of the *t* distribution for different degrees of freedom: as with the standard normal distribution, it became possible to obtain a given  $(1 - \alpha)$  100% confidence interval by means of a critical value  $t_{\alpha/2, n-1}$ , corresponding to a point beyond which the integral of the curve is  $\alpha/2$ . Thus, since  $se(\bar{x}) = s / \sqrt{n}$ , the result obtained for a small sample of size  $n$  is

$$1 - \alpha = \Pr(\bar{x} - t_{\alpha/2, n-1} se(\bar{x}) \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} se(\bar{x})),$$

which is exactly what is obtained using the *z* distribution for large samples, and from which we can be  $(1 - \alpha)$ 100% confident that the population mean lies in the interval  $\bar{x} \pm t_{\alpha/2, n-1} se(\bar{x})$ . The first statistician really to appreciate the importance of Gosset's discovery was Sir Ronald Fisher, who had a lifelong admiration for Gosset and called him "the Faraday of statistics" on account of his ability to discover a general principle and immediately apply it to practical ends. Fisher was also the first to apply the *t* distribution to the theory of linear regression. Pearson, on the other hand, did not appear to grasp fully the importance of Gosset's work, probably because the model (designed for small sample sizes) met the needs of people dealing more with quality control

processes than with those of pure biostatistics, even though Pearson did help Gosset with aspects of his paper.

After returning to Guinness in Dublin, Gosset's research was focused mainly on techniques of experimental design, particularly with regard to barley cultivation. He studied new plant varieties that were less sensitive to changes in soil chemistry and to changes in the climate and seasons. All but one of the papers he wrote were published under his pen name, again largely to protect the discoveries in quality control from competitors. In 1935, Gosset became head brewer and moved to London to work in a new fac-

tory owned by Guinness. Only two years later, on 16 October, 1937, he died suddenly, in Beaconsfield, of a heart attack.

Gosset was a modest man who never flaunted his achievements. He would apparently dismiss praise with the comment that "Fisher would have discovered it all anyway".

## References

1. Pearson ES, Bernard GA. Student: Statistical Biography of William Sealy Gosset. Clarendon Press, Oxford, 1990.