From:
Coladarci, T. (2010). Student's *t* statistic. In I. B. Weiner & W. E. Craighead (Eds.), *Corsini Encyclopedia of Psychology* (4th ed., p. 1723). Hoboken, NJ: Wiley.

# STUDENT'S *t* STATISTIC

At the dawn of the twentieth century, a young scientist faced the limitations of existing statistical theory when applied to research involving small samples. He worked for the Guinness brewery in Dublin, and his investigations of the quality of raw ingredients, the nature of the production process, and the resulting quality of the end product—beer—often entailed decidedly small samples of 10 or fewer experiments. The scientist was William Sealy Gosset (1876–1937). He is better known by his pseudonym "Student," which, for proprietary reasons, his employer required him to use when publishing.

Consider a question frequently asked by Gosset: What is the probability that the population mean falls within a given proximity to the sample mean? There is sampling error associated with the sample mean, of course. But Gosset's concern was that there is sampling error associated with the sample standard deviation as well. Moreover, Gosset cautioned, "As we decrease the number of experiments, the value of the standard deviation found from the sample of experiments becomes itself subject to an increasing error, until judgments reached in this way may become altogether misleading" (Student, 1908, p. 2). To evaluate a mean when sample size is small, one therefore should not apply the standard normal distribution. In short, Gosset needed something that did not yet exist: a test statistic that took into account the greater sampling error associated with small samples.

Undeterred, Gosset began by deriving the distribution of $z = \left(\bar{X} - \mu\right)/s$ and, in turn, tabulating probability values for *z* for various sample sizes (Student, 1908, p. 19). Gosset thus introduced the notion of a family of distributions—a different distribution for each sample of size *n*. Through the profound influence of R. A. Fisher, Gosset's friend and collaborator, *z* was later modified to take the form of the now-familiar one-sample *t* ratio:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}},$$

where $\bar{X}$ is the sample mean, μ is the mean specified in the null hypothesis, *s* is the sample standard deviation, and *n* is the number of observations on which $\bar{X}$ is based.

The one-sample *t* ratio is distributed with *n* – 1 degrees of freedom (*df*)—again, a family of distributions comprising a different distribution for each sample of size *n*. The *t* distribution is identical to the standard normal distribution when $df = \infty$, whereas there is more lift in the tails of *t* when the sample size is small. For example, $t = \pm 1.96$ captures the middle 95% of the area under the curve when $df = \infty$. To capture this same area when $df = 9$, however, one must move farther out in the tails to $t = \pm 2.26$, and farther still to $t = \pm 4.30$ when $df = 2$. This increasingly greater lift in the tails of *t* reflects the statistic's increasingly greater sampling variability as sample size decreases. It is in this way that *t* gives more accurate probabilities, in comparison with the standard normal distribution, when inferences are made from small samples.

Student's *t* also is enlisted for the two-sample case, such as the difference between means of independent samples:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{\sum_{i=1}^{n_1}(X_i - \bar{X}_1)^2 + \sum_{i=1}^{n_2}(X_i - \bar{X}_2)^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

which is distributed with $n - 2$ degrees of freedom. Student's *t* is frequently used in application to correlation and regression coefficients as well.

One cannot overstate the importance of Gosset's contribution to statistical theory and its practical applications. Nor can one overstate the grace and humility with which this gentle man carried out his work and collaborated with others. There is the oft-told account of an admirer who approached Gosset in his later years, saying, "On behalf of fellow statisticians, I would like to thank you for all that you have done for the advancement of statistics." "Oh, that's nothing," Gosset replied. "Fisher would have discovered it all anyway."

# Reference

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

# Suggesting Readings

Boland, P. J. (1984). A biographical glimpse of William Sealy Gosset. *American Statistician, 38,* 179–183.

Box, J. F. (1987). Guinness, Gosset, Fisher, and small samples. *Statistical Science, 2*(1), 45–52.

Eisenhart, C. (1979). On the transition from "Student's" *z* to "Student's" *t*. *American Statistician, 33,* 6–10.

Tankard, J. W. (1984). W. S. Gosset and the *t*-test. In *The statistical pioneers*. Cambridge, MA: Schenkman.

THEODORE COLADARCI

*University of Maine*

*See also*: Parametric Statistical Tests