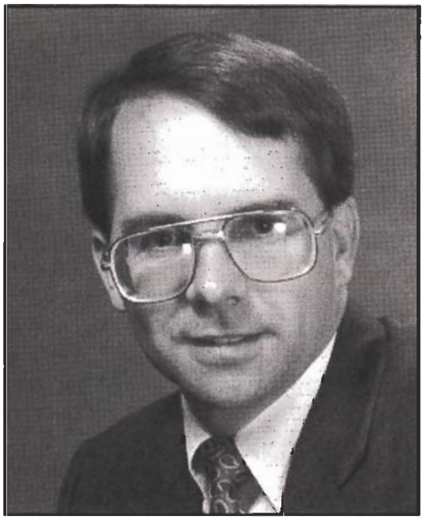
The "Student" of the Student's *t*-Test

By Brian R. Page



Brian R. Page *is a systems engineer for Network Imaging Systems (449 Pleasant Hill Road, Suite 202, Lilburn, GA 30247). He received an M.A. in the history of science from Virginia Tech in 1980 and maintains an avocational interest in the history of the physical sciences. He last wrote for* The Physics Teacher *in November 1991 on the subject of rocket pioneer Robert Goddard.*

A major theme in physics over the past few centuries has been the quest for greater accuracy in measurements. Consider the positional astronomy of Tycho Brahe in the sixteenth century. Johannes Kepler used Tycho's extremely precise measurements of planetary motions to detect an eight-arc-minute anomaly in the orbit of Mars. This breakthrough finally provided a quantitative observational basis to the Copernican heliocentric theory. Whereas Nicolaus Copernicus possessed no information not available to ancient observers, Kepler's achievement was a direct result of improved accuracy.

In recent times the need for increasingly accurate time keeping has resulted in the routine worldwide availability over shortwave radio of a transmitted time signal accurate to 0.00001 seconds. Similarly, nearly all physical constants, from pi to the Hubble parameter describing the expansion of the universe, are further refined with each passing year.

This steady improvement in measurements of all sorts has resulted in part from improved instrumentation. However, as instruments improved, experimenters ran smack into the fact that there is no such thing as a perfect measurement. Any measurement along a continuous scale, as opposed to counts of discrete objects, is subject to a variety of influences. The continuous scale itself, by definition, contains an infinite number of points. Increasing accuracy merely reduces the uncertainty of the value.

The need to understand variations in measurements led to important concepts in statistics such as bias, mean, and standard deviation. The practice in statistics concerned with the meaning of measurement is known as the analysis of variance. This field formally developed only in the last century, beginning with a significant insight by a little-known chemist working in Ireland and writing under the pseudonym "Student." In this article we will look at Student's *t*-test of statistical hypothesis and learn something about the man who articulated a relationship fundamental to every experimenter. Before turning to the life of Student, let's review some basic statistics.

Statistics

Statistics has an unfortunate reputation as being difficult, mysterious, and able to "prove" any proposition. Admittedly, advanced statistics can be quite complicated. However, the basic principles are rather straightforward and can be of enormous benefit to experimenters at even the earliest stage of their education. Let's begin with average.

Taking an arithmetic mean from a number of measurements is almost second nature. The mean, or average, is simply the sum of all measurements divided by the number of measurements. In most work, where tolerances are not of critical importance and it is not necessary to actually calculate error margins, the mean suffices to characterize a physical property. This is especially true when it can be easily seen that the individual measurements do not display a significant range.

Doubts about the value of the mean begin to creep into the picture when a series of measurements shows a good scatter. A series of observations of the brightness of a star might differ by a magnitude or more. These differences could reflect inexperienced observers, poor eyesight, defective optics, or a number of other sources of error. Can we simply accept the average without considering some of these problems? Can we report the average but qualify it with some value reflecting the quality of the data?

Consider a student opinion poll on what kind of music to play during a class party. When asked to rate two types of music on a scale from 1 to 10, the average was close to five. Does this mean that the students really did not care? Of course not! Music is a topic about which students hold very strong opinions. When rock and roll was pitted against country music, approximately half of the students rated country a 10 and rock and roll 1. The students favoring rock and roll reported the opposite ranking. Although the average was five, we can clearly see that this metric did not tell the whole story. We need some way to show that although the average was in the middle of our range, the individual measurements varied considerably.

Here is a more practical example: let's assign two teams of students the task of measuring the temperature of a sample of water. They are directed to make ten individual measurements using an ordinary laboratory thermometer calibrated in Celsius. We will call the teams Alpha and Zed. Each student must estimate the temperature to one-tenth of a degree. The results are shown in Table I.

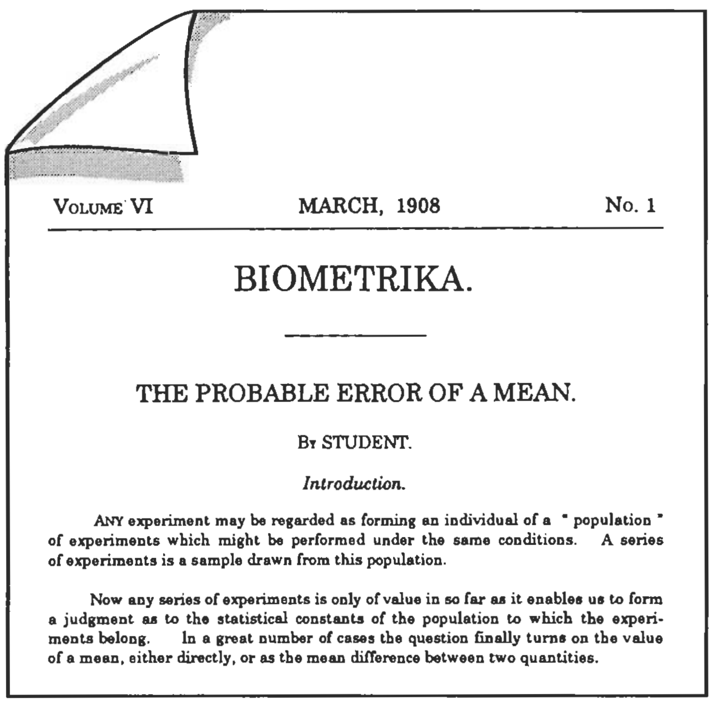
At first glance, we can see that the average temperatures determined by the two teams are remarkably close. Now let's look more closely at the individual measurements and see if we should place more confidence in one team or the other. To do this, we will turn to the standard deviation, which will eventually take us into the heart of Student's t-test.

Standard deviation is the key to understanding the diversity of measurements that go into an average. It is useful in determining the intrinsic error inherent in any series of measurements, as in the water temperature example. It is also useful in identifying any extremes that might exist in a set of data, as in the survey of music preferences. With a thorough understanding of standard deviation, a mere average by itself will never again seem adequate.

Calculation of a standard deviation from a collection of measurements is a five-step process.[1] Standard deviation is equal to

$$s = \sqrt{\frac{\Sigma\,(dev)^2}{n-1}} \qquad (1)$$

The numerator, $\Sigma\,(dev)^2$, is determined by subtracting each measurement from the mean, squaring the result, and summing the squares. In the denominator $(n-1)$, n is the number of samples or measurements of a given population. Table II shows the same measurements as Table I, this time with additional calculations. Along with each measurement of temperature, I have now included the difference with the mean, the square of this difference, and the quotient of the difference divided by the standard deviation. For the moment, ignore this last column.



VOLUME VI — MARCH, 1908 — No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between two quantities.

Beginning of the landmark paper by Student. Illustration courtesy of University College London.

Referring to the data for Team Alpha, we can easily verify the table. The sum of the squares is 0.204. This amount, divided by 9 (which is one less than the number of measurements) is approximately equal to 0.0227, the square root of which is 0.15. Compare the standard deviations between the two teams. Although the averages differ by only two-hundredths of a degree Celsius, it is apparent that Team Alpha had less scatter in their measurements. Does this mean that Team Alpha did a better job? Does it mean that their measurements are closer to the true, but unknown water temperature? Not necessarily. If all considerations of bias, instrumental error, and faulty technique were equal between the two teams, then we might say that Team Alpha's result is closer to the truth. However, we may simply have discovered that the members of Team Alpha shared their measurements as they made them, thus influencing subsequent measurements. The value of an experiment cannot be determined solely through statistical calculations. An ex-

	Team Alpha Degrees C	Team Zed Degrees C
1	14.3	14.2
2	14.4	14.3
3	14.4	14.4
4	14.5	14.4
5	14.5	14.5
6	14.6	14.6
7	14.6	14.7
8	14.6	14.8
9	14.7	14.8
10	14.8	14.9
Mean	14.54	14.56

Table I. Measurements of water temperature by Teams Alpha and Zed; for convenience results are sorted numerically.

Team Alpha				
	Degrees C	difference	squared	dif/sd
1	14.3	-0.24	0.0576	-1.59411
2	14.4	-0.14	0.0196	-0.9299
3	14.4	-0.14	0.0196	-0.9299
4	14.5	-0.04	0.0016	-0.26568
5	14.5	-0.04	0.0016	-0.26568
6	14.6	0.06	0.0036	0.398527
7	14.6	0.06	0.0036	0.398527
8	14.6	0.06	0.0036	0.398527
9	14.7	0.16	0.0256	1.062738
10	14.8	0.26	0.0676	1.726949
Mean	14.54			
Sum of the squares			0.204	
Standard Deviation			0.150555	

Team Zed				
	Degrees C	difference	squared	dif/sd
1	14.2	-0.36	0.1296	-1.52128
2	14.3	-0.26	0.0676	-1.0987
3	14.4	-0.16	0.0256	-0.67612
4	14.4	-0.16	0.0256	-0.67612
5	14.5	-0.06	0.0036	-0.25355
6	14.6	0.04	0.0016	0.169031
7	14.7	0.14	0.0196	0.591608
8	14.8	0.24	0.0576	1.014185
9	14.8	0.24	0.0576	1.014185
10	14.9	0.34	0.1156	1.436762
Mean	14.56			
Sum of the squares			0.504	
Standard Deviation			0.236643	

Table II. Initial measurements of Teams Alpha and Zed examined more critically.

perimenter must take into account all possible sources of error, bias, and uncertainty. Statistics only helps identify and quantify errors.

The usual technique for minimizing error is to make a large number of independent measurements. In general, this is the best approach for avoiding bias.

Now let's look at the last column in Table II, the difference divided by the standard deviation. This leads us into the normal error curve. The normal error curve, or normal distribution, is pervasive in nature. Such a curve describes a wide variety of natural phenomena. For our purposes, the normal error curve is almost always a good fit when describing the scatter of values apparent in any large set of measurements. In brief, the normal error curve shows us that the number of measurements close to the average will predominate over measurements farther from average. Quantitatively, about two-thirds of all measurements will reside within one standard deviation on either side of the average. This is an extraordinary generalization in that it holds for such diverse phenomena as the growth of mollusks, the dripping of rainwater from the gutter of my house, the sizes of digestive stones in a dinosaur's gut, and student scores on standardized tests. See Fig. 1 for a graph of a perfect normal error curve.

This consideration shows us graphically why we desire a small standard deviation in any set of measurements. The width of the curve is entirely dependent on the standard deviation. If we are to say with any confidence that a mean approximates some real value, then we need a narrow curve. Looking at the last column for the Team Alpha data confirms, even in a very small sample, our generalizations. Seven of the ten measurements lie within 1 to −1 standard deviations. For additional information on achieving a normal error distribution, see the sidebar on page 496, A Straight Line to a Normal Curve.

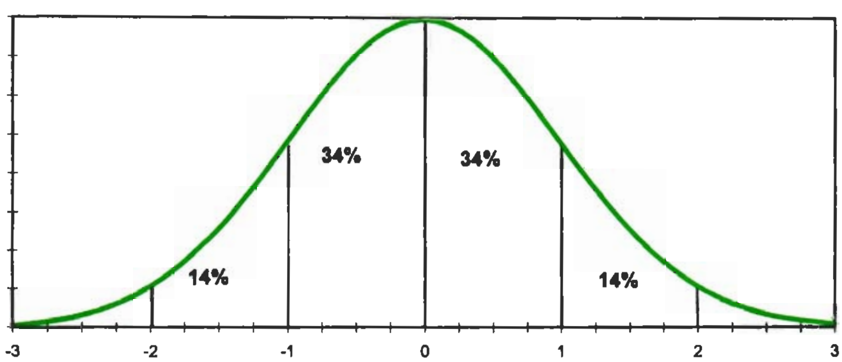


Fig. 1. Normal law of error. For sample populations that are distributed normally, the shape of the curve is defined by the mean and the standard deviation. As shown, 68% of measurements fall within one standard deviation of the mean; another 28% lie between one and two standard deviations.

After learning about standard deviation and the normal error curve, we will assign teams Alpha and Zed the task of improving their estimates of the temperature of the water. In the examples, we have cited rather small samples of data. In part, this was for economy of space. But it also matches the situation faced by the man known as Student.

William Sealy Gosset (1876–1937)

The firm of Arthur Guinness Son & Company was already 140 years old when young William Sealy Gosset began his employment there in 1899. Gosset was fresh from Oxford University, having studied mathematics and taken a degree in chemistry. His employment was, in itself, something of an experiment, for Guinness had only recently initiated the practice of hiring university men with training in science. This novelty was compounded by incorporating these graduates directly into the work of the brewery, not relegating them to strict scientific or technical assignments.

William Sealy Gosset, the "Student" of Student's *t*-test, spent his entire professional career with Arthur Guinness Son & Company as a brewer. A collaborating statistician, E.S. Pearson, remarked that "the name of 'Student' was associated in statistical circles with an atmosphere of romance." Photograph courtesy of Arthur Guinness Son & Company.

Fairly soon after his arrival in Dublin, Ireland, where he would make his home for the next 36 years, Gosset was provided with an abundance of raw data relating to the operations of the brewery. The Guinness product was the result of exacting procedures carried out with only four basic ingredients: barley, hops, water, and yeast (see photo at bottom of this page). However, the simplicity of the ingredients belies the complexity of the process. Guinness had earlier established an experimental unit to research the effects of variations in their procedures. Slight changes in initial conditions, ingredients, processing times, and temperatures could greatly alter the finished product. The managers of the company wished to better understand these variations. Thus, Gosset, being the most mathematically gifted of the recent university graduates, was directed to analyze the data.[2]

In 1904, Gosset reported on his progress. He advocated the use of statistical methods, commenting: "It may seem strange that reasoning of this nature has not been more widely made use of, but this is due: (1) To the popular dread of mathematical reasoning. (2) To the fact that the results [conventionally obtained] are well within the accuracy required."[3] A central problem facing the laboratory was that the complexity of their procedures prevented the numerous replication of experiments in attempts to minimize error. This constraint, which was common in industrial situations, had not been adequately addressed by the more theoretical academic statisticians of the time. Gosset concluded his initial report with the recommendation that more advanced statistical techniques be brought to bear on the work of the laboratory.

As a result of his findings, communication was established between Gosset, working in Dublin, and Karl Pearson (1857–1936), probably the leading statistician of the day in Britain. Pearson headed the Biometric Laboratory of University College London and had recently begun publication of the journal *Biometrika*. In advance of their first meeting, Gosset composed four questions that came to characterize his lifelong interest in the analysis of variance. Gosset sought assistance in:

- relating the economic impact of changes in procedure to the costs of experimentation.
- determining the probability that his measurements closely describe an unknown value. As Gosset wrote, if the number of measurements, "*n* were infinite, I could say 'it is 10:1 that the truth lies within 2.6 of the result of the analysis'. As however *n* is finite and in some cases not very large, it is clear that I must enlarge my limits, but I do not know by how much."
- identifying correlation between supposedly independent observations.
- and finally, identifying books that would be useful.[4]

From Pearson, Gosset obtained immediate assistance particularly in methods of dealing with correlation. A more important upshot of their introduction was that in 1906, Gosset was sent by his employer to study for a year at the Biometric Laboratory. While in London, Gosset tackled the problems outlined in his first letter to Pearson. The results were published as a series of papers in *Biometrika*. The second of these papers, "The Probable Error of a Mean," marks Gosset's place in history (see illustration on page 491).

Agricultural experiments with test fields of barley and hops were perfect candidates for Gosset's new techniques of evaluating small sample populations. The bags of hops in this photograph await inspection at the Guinness Dublin brewery around the turn of the century. Photograph courtesy of Arthur Guinness Son & Company.

For publication, Gosset assumed the pseudonym "Student." Recall that he was essentially an industrial chemist working in a very competitive field. His employer, conscious of the new techniques they were introducing into their processing, wished to protect whatever competitive advantage was offered by this statistical work. Thus, the writing was pure statistics, containing no link to the barley, hops, water, and yeast that inspired the problems; the pseudonym prevented ready identification of the author with the employer. Competitors would have little help in relating Gosset's theoretical discoveries with changes in brewing. The rather unusual pseudonym "Student" was chosen by Gosset in honor of Karl Pearson, his "professor."[5]

The t-Test

In the introduction to "The Probable Error of a Mean," Gosset succinctly states his problem: "it is sometimes necessary to judge of the certainty of the results from a very small sample." This is the second question originally posed to Pearson. Gosset continues by explaining, "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 'large' and 'small' samples is to be drawn. The aim of the present paper is to...furnish alternative tables for use when the number of experiments is too few."[6]

Gosset arrived at his new understanding by analyzing the relationship of the mean of a population, the estimate of that mean, and the standard deviations. For experimental data, he drew upon a table listing the height and left middle finger measurements of 3,000 criminals. He first calculated the parameters for the entire population under study and then drew four random samples, each of 750 measurements, for comparison against the whole. This approach was decidedly empirical; Gosset was perhaps lucky in quantifying the error in a way that was later demonstrated to be optimum.

In this 1908 paper, Gosset expressed the original formulation of what has since become known as Student's *t*-test of statistical hypothesis. In its modern form, as we shall see, Gosset related a table of constants, *t*, with the probability that a mean closely approximates the unknown value it attempts to describe. The formula,

$$\Delta = t\,\frac{s}{\sqrt{n-1}} \qquad (2)$$

produces a delta which, when subtracted from and added to the mean, provides an interval in which the unknown value may lie according to the probability specified with the constant *t*. For an example, refer to the data from Team Alpha. The standard deviation of the sample, 0.15 divided by the square root of the number of measurements less one, is equal to 0.05. Now we must select an appropriate value for *t*.

If we wish for the interval to have a 50% chance of encompassing the unknown value, then we select a value for *t* close to 0.7. The product is 0.035, providing an interval of 14.51 to 14.58 degrees. In other words, we have a fifty-fifty chance that the temperature of the water actually lies between 14.5° and 14.58° Celsius. We can increase our certainty by using another value for *t*. To achieve 99% probability, substitute for *t* the value 3.2. This provides a delta of 0.16 and an interval of 14.38 to 14.70 degrees.

Note that as the level of confidence increases, the delta grows larger, and the interval widens. Now carefully consider the equation. The number of measurements, in the denominator, is taken into account only through its square root. Thus, it is difficult to achieve greater accuracy merely by increasing the number of measurements. The most benefit accrues by lowering the standard deviation in the numerator. Also, refer to the more complete list of *t* values in Table III. The values of these constants reinforce the diminishing returns relationship. Scan the column for 99% probability. It should be obvious that increasing the number of measurements beyond ten or so does not greatly improve the chances that our interval includes the actual value.

	Probability					
n-1	0.99	0.9	0.8	0.7	0.6	0.5
1	63.655898	6.3137486	3.0776846	1.962612	1.376382	1.000001
2	9.92498826	2.9199873	1.885619	1.386206	1.06066	0.816497
3	5.84084773	2.353363	1.6377453	1.249778	0.978472	0.764892
4	4.60408046	2.1318465	1.5332057	1.189567	0.940964	0.740697
5	4.03211743	2.0150492	1.4758848	1.155768	0.919543	0.726687
6	3.70742782	1.9431809	1.4397551	1.134157	0.905703	0.717558
7	3.49948095	1.8945775	1.4149236	1.119159	0.89603	0.711142
8	3.35538061	1.8595483	1.3968156	1.108145	0.88889	0.706386
9	3.24984285	1.8331139	1.3830288	1.099716	0.883404	0.702722
10	3.16926162	1.8124615	1.3721842	1.093058	0.879057	0.699812
11	3.10581527	1.7958837	1.3634303	1.087667	0.87553	0.697445
12	3.05453796	1.7822867	1.356218	1.083212	0.872609	0.695483
13	3.01228283	1.7709317	1.3501722	1.079469	0.870151	0.69383
14	2.97684892	1.7613092	1.3450313	1.07628	0.868055	0.692417
15	2.94672645	1.753051	1.3406054	1.073531	0.866245	0.691197
16	2.92078767	1.7458842	1.3367571	1.071137	0.864667	0.690133
17	2.8982322	1.7396064	1.3333795	1.069034	0.863279	0.689195
18	2.87844159	1.7340631	1.3303907	1.067169	0.862049	0.688364
19	2.86094291	1.7291313	1.3277281	1.065507	0.86095	0.687621
20	2.84533598	1.724718	1.3253407	1.064016	0.859965	0.686954
21	2.83136615	1.7207435	1.3231875	1.06267	0.859075	0.686352
22	2.81876055	1.7171442	1.3212366	1.061449	0.858266	0.685805
23	2.8073373	1.71387	1.3194608	1.060337	0.85753	0.685307
24	2.79695087	1.7108823	1.3178351	1.059319	0.856855	0.68485
25	2.78743755	1.7081402	1.3163458	1.058385	0.856236	0.68443

Table III. Values of Student's *t* constant for selected probabilities between 99% and 50%, for two to 26 measurements. Table generated using Microsoft Excel with the TINV function.

Although Gosset immediately put his findings into use at the Guinness laboratory, his work was largely ignored for the next several years by academic statisticians. The reason for this relates to Gosset's original motivation: only in an industrial setting was there such concern over reaching supportable conclusions on the basis of so few measurements. Years later, Gosset articulated his predicament to a correspondent:

> You see one must experiment and frequently it is quite out of the question, from considerations of cost or of impossibility of duplicating conditions in the time scale, to do enough repetitions to define one's variability as accurately as one could wish. It's no good saying "Oh these small samples can't prove anything." Demonstrably small samples *have* proved all sorts of things and it is really a question of defining the amount of dependence that can be placed on their results as accurately as we can. Obviously we lose by having a poor definition of the variability but *how much do we lose?*[7]

Nevertheless, Gosset's preoccupation with small sample sizes stressed existing statistical technique and eventually led to improvements in handling the more customary larger populations. According to statistician Egon S. Pearson, son of Karl Pearson, the existing imperfections were of "small consequence" with large samples but completely evident in the "absurdly small numbers" of Gosset's.[8] In this way, Gosset's contributions extended greatly beyond small sample technique and truly set the foundation for the analysis of variance.

Although Gosset was first motivated by problems uncovered in the experimental brewery, he was later quite active in extending his techniques to agricultural experiments conducted by the Irish Department of Agriculture. The techniques pioneered by Gosset were eventually extended and popularized by Ronald A. Fisher (1890–1962). Writing to Gosset while still a student, Fisher proposed that the numerator in the standard deviation be $n - 1$ instead of simply n, as Gosset had assumed. Responding to this correction, Karl Pearson wrote that such a refinement mattered little "because only naughty brewers take n so small that the difference is not of the order of the probable error!"[9] From this beginning, Gosset and Fisher established a lifelong collaboration. Indeed, it was Fisher in 1924, who first introduced the term "t-test." Gosset immediately accepted the new formulation and published extended values for t.[10]

To see Student's t-test in action, let's return to teams Alpha and Zed and their attempts to determine more accurately the temperature of some water. Not knowing of Student's t-test, the two teams approached the problem differently. Their results are shown in Table IV. Team Alpha again made only ten measurements. However, they found a magnifying glass that they put to good use in reading the thermometer. Team Zed kept the usual technique but made 20 readings. Fortunately, both averages are very close. The real difference shows with the standard deviations. Team Alpha's standard deviation is three times less than that of Team Zed's. This advantage translates into a higher degree of confidence as can be seen through the t-test. Using values for 90% probability, Team Alpha has an interval of 14.5 to 14.6 degrees. By contrast, Team Zed, even with ten additional samples, has an interval from 14.44 to 14.68 degrees. For 90% confidence, Team Zed has an interval more than twice as wide as Team Alpha's!

Team Alpha			
	Degrees C	**difference**	**squared**
1	14.4	-0.15	0.0225
2	14.5	-0.05	0.0025
3	14.5	-0.05	0.0025
4	14.5	-0.05	0.0025
5	14.5	-0.05	0.0025
6	14.6	0.05	0.0025
7	14.6	0.05	0.0025
8	14.6	0.05	0.0025
9	14.6	0.05	0.0025
10	14.7	0.15	0.0225
Mean	**14.55**		
Sum of the squares			0.065
Standard Deviation			**0.08498366**

Team Zed			
	Degrees C	**difference**	**squared**
1	14	-0.56	0.3136
2	14	-0.56	0.3136
3	14.2	-0.36	0.1296
4	14.3	-0.26	0.0676
5	14.3	-0.26	0.0676
6	14.4	-0.16	0.0676
7	14.5	-0.06	0.0036
8	14.5	-0.06	0.0036
9	14.6	0.04	0.0016
10	14.6	0.04	0.0016
11	14.6	0.04	0.0016
12	14.7	0.14	0.0196
13	14.7	0.14	0.0196
14	14.7	0.14	0.0196
15	14.7	0.14	0.0196
16	14.7	0.14	0.0196
17	14.8	0.24	0.0576
18	14.9	0.34	0.1156
19	15	0.44	0.1936
20	15	0.44	0.1936
Mean	**14.56**		
Sum of the squares			1.63
Standard Deviation			**0.2891**

Table IV. Results of different strategies by Team Alpha and Team Zed to improve their estimates of the water temperature.

Student's t-test has, from time to time, been taken to provide a perhaps unwarranted legitimacy to small samples. Gosset himself never fell into this trap. In his 22 papers published between 1907 and his death in 1937, he referred frequently to the necessity of gathering significant numbers of observations, striving rigorously to eliminate error and bias, and routing out spurious correlation. Indeed, he repeatedly warned that "tables can only be an aid to, and not a substitute for, common sense."[11]

As Team Zed learned, Gosset not only left us a powerful tool for evaluating small sample populations, but he also showed that we cannot compensate for the lack of accuracy in individual measurements merely by increasing their number.

References

1. Although standard deviation is nowadays easily available at the touch of a button on most scientific calculators or through the use of a function in spreadsheet software, the detailed form is presented here since the method is germane to our goal of reducing error. Incidentally, it may prove interesting to compare a manual calculation with the result from your calculator.
2. Launce McMullen, Foreword to *"Student's" Collected Papers* (Cambridge University Press, Cambridge, 1958), p. xi. Hereafter cited as *Collected Papers*.
3. E.S. Pearson, "'Student' as Statistician," in *Studies in the History of Statistics and Probability*, edited by E.S. Pearson and M. Kendall (Charles Griffin & Co., London, 1970), Vol. 1, p. 364. Hereafter cited as *Studies*.
4. Ibid., pp. 365–366.
5. E.S. Pearson, "Some Reflexions on Continuity in the Development of Mathematical Statistics, 1885–1920," *Studies*, p. 350.
6. W.S. Gosset, "The Probable Error of a Mean," *Collected Papers*, p. 12.
7. W.S. Gosset to E.S. Pearson, May 11, 1926, *Studies*, p. 396.
8. E.S. Pearson, "'Student' as Statistician," *Studies*, p. 368.
9. K. Pearson to W.S. Gosset, Sept. 17, 1912, *Studies*, p. 368.
10. Gosset's original work used a z constant, which Fisher later related to t as $\sqrt{n-1}z$. The new tables appeared in publication together with a theoretical contribution by Fisher.
11. W.S. Gosset, "Mathematics and Agronomy," *Collected Papers*, p. 128.

Additional Reading

- W.J. Youdon, *Experimentation and Measurement*. This volume, the best little introduction to statistics and measurement, is available free of charge. Request an individual copy of NIST Special Publication 672 from Standard Reference Materials Program, Building 202, Room 204, National Institute of Standards and Technology, Gaithersburg, MD 20899.
- James W. Tankard, Jr., *The Statistical Pioneers* (Schenkman Publishing, Cambridge, MA, 1964).
- Clifford E. Swartz, *Used Math* (AAPT, College Park, MD, 1993).
- Philip J. Boland, "A biographical glimpse of William Sealy Gosset," *Am. Statist.* **38** (3), 179–183 (August 1984).
- George W. Snedecor and William G. Cochran, *Statistical Methods* (Iowa State University Press, Ames, IA, 1989). This is a definitive and supremely practical book of detailed statistical technique for the scientific worker.