Part III, Interpreting Results: An Introduction to Basic Statistics

Luck. An uneducated gambler in Las Vegas depends upon it, usually to their surprise and dismay, while the more experienced gambler carefully calculates the probability that they can win. Similarly, as a professional and as a superintendent, you prefer not to rely on lucky guesses when making management decisions on the golf course. Instead, you strive to make sure that your key decisions are based on factual information that allows you to accurately predict how new products and management practices will perform on the golf course.

When designed properly, a good testing program helps to support you in this effort. While you can never eliminate the possibility of unexpected results, you can surely reduce the possibility that you will be unpleasantly surprised by basing your decisions on data from a sound testing program.

Statistics: Managing the Game of Chance

When a "fair" coin is tossed into the air, the likelihood that the coin lands with the heads facing up is 1/2, or 50%. This probability represents the number of heads on the coin (1) divided by the total number of sides on the coin (2, heads and tails). Probability theory tells us that there is a 50% chance that you will win a bet every time the coin is tossed regardless of whether you select heads or tails. Even if five tosses of the coin come up tails, the chance that the next toss of the coin will be heads is still 50% — no more and no less than for any other toss of a fair coin.

When a field test is conducted, the odds are not so easily calculated as they are for a coin toss. Why is this? The answer is that the number of **variables**, or factors than can contribute to the outcome, are much higher for a field test than for a coin toss. In a field test, the turfgrass variety, turfgrass stress, soil type, traffic patterns, weather etc. can have a big effect on the performance of products and practices. In contrast, the number of variables contributing to the outcome of a coin toss are limited.

Because we cannot use guesswork, probability theory or any other system to predict how a product or practice will perform, field tests are conducted to give us information that can be used to make the best possible predictions and decisions.

Statistics is the tool that allows you, as objectively as possible, to analyze the information collected from field tests, and to predict, with as much confidence as possible, which products or practices will give you the best results. In the first two installments of this series (Stowell and Gelernter, 1997; Stowell and Gelernter, 1998), we described how to set up field tests, and how to collect the results. In this 3rd and final installment, we will describe the final steps — how to analyze the results statistically, by calculating the **mean**, the **standard deviation**, and the **confidence interval.** In addition, we'll review methods that will allow you to clearly represent the results in the form of **line graphs, bar charts** and **data tables,** for use in your own records and for presentations to greens committees, general managers and others.

A Real Life Example

To give our discussion of statistics some grounding in reality, we will use results from a field test conducted by the PACE Turfgrass Research Institute in 1997. This test was conducted with the assistance of Bill Gallegos, CGCS at Los Coyotes Country Club, and with financial support from Valent Corporation The objective of the test was to look at the performance of 3 different rates of an experimental fungicide (procymidone, Valent Corporation) and to compare it to a standard fungicide, iprodione (Chipco 26019, Rhone-Poulenc) for control of dollar spot, *Sclerotinia homeocarpa,* on a creeping bentgrass nursery. The five different treatments tested are listed in Table 1. Each treatment was replicated three times, and treatments were randomized. Results were collected two weeks after the fungicides were applied by making a visual estimate of percent turf damage due to dollar spot.

How do we use this data to make a decision on the best product and best rate to use for controlling dollar spot? By calculating the mean, the standard deviation and the confidence interval, as described below.

Calculating the Mean

Usually, the first statistic that is calculated is the **mean**, or average rating for each treatment. The mean is calculated by summing the values for each replicate of a given treatment, and then dividing by the number of replicates. For example, in our dollar spot experiment, the mean for percent dollar spot in plots treated with procymidone at 0.5 oz active ingredient/1000 sq ft is 4.667 (rounded to 4.7 in Table 1):

sum of values (5 + 7 + 2) ÷ number of replicates (3) = mean (4.667)

This process is repeated for each treatment, as illustrated in Table 1. The mean can be easily calculated with pencil and paper. If you are using a calculator, the mean may be represented by the symbol "**X**" with a horizontal line drawn above it.

Although the mean is a powerful statistic, when used by itself, it can be misleading and can push you towards poor decisions. This is because the calculation of the mean doesn't take into account the **variability** of the results.

Variability: a Complicating Factor

There are many factors beyond our immediate control at work on a golf course, such as microclimate, moisture, turf quality, pest pressure, etc. As discussed in Part 2 of this series (Stowell and Gelernter, 1998), these factors exert a powerful force on the way a product or practice performs, and how consistently it performs. The use of replication (repeating a treatment in two or more locations) in designing your field test helps to minimize the effects of variability, but it can't erase them. As a result, it is extremely rare for a given treatment to produce the same result each time it is applied. In the dollar spot experiment, for example, procymidone applied at 0.5 oz active ingredient/1000 square feet produced three different disease incidence levels in each of three identical plots — 5%, 7% and 2% (Table 1).

How does variability affect your interpretation of the results? Let's assume that in the dollar spot experiment above, variability was much lower. In that case, the percent dollar spot values for the procymidone 0.5 oz treatment would be much more similar, for example, 4.6%, 4.7% and 4.7%. The mean for these hypothetical values would be identical to the mean calculated above — 4.667, but the variability would be less.

Which set of data gives you a greater guarantee that the product will perform the same way the next time you apply it? Which data set gives you a greater sense of confidence? Statisticians tell us that the data set with the lowest variability gives us the best predictions for how products will perform. So, even when the means are the same for two data sets, we still want to know how variable the data was.

Measuring Variability: the Standard Deviation and the Confidence Interval

There are a variety of statistics used to measure variability, but the most commonly used measure is the **standard deviation,** frequently represented by the symbol "**S**" on a hand calculator. A small standard deviation indicates that there is less variability associated with the mean — the data is more consistent — than the same mean with a large standard deviation. In the dollar spot example presented in Table 1, the highest standard deviation (6.3) occurs in the non-treated check treatment, and the lowest standard deviation (0.6) occurs in two of the procymidone treatments — the 1.5 oz and 2.5 oz rates.

Calculating the standard deviation is more complicated than calculating the mean, and we encourage you to purchase a calculator (most simple scientific calculators include standard deviation), or use a spread sheet program, such as Microsoft Excel, that performs the standard deviation function.

Looking at the means and standard deviations in Table 1, which treatment or treatments do you think gave the best dollar spot control? We still have one more calculation to perform before we can answer that question — **the confidence interval.**

The confidence interval is related to the standard deviation, and is an easy way to represent the interval, or range of values, or degree of variability associated with a mean. The lower end of the interval is calculated by subtracting the standard deviation from the mean, and the higher end of the interval is calculated by adding the standard deviation to the mean. Staying with the example of procymidone at 0.5 oz, the confidence interval for this treatment would range from 2.2 (4.7 - 2.5) to 7.2 (4.7 + 2.5). In other words, we have a high level of confidence that the mean value for this treatment falls between 2.2% and 7.2%; and our best estimate for that mean is 4.7%.

To find out which treatments performed statistically differently from another, look for the treatments where the range of values of the confidence intervals do not overlap. For example, the non-treated check, with confidence limits of 13.0 - 25.6, is statistically different from all of the other treatments, whose confidence intervals never get as high as 13.0. In contrast, procymidone at 1.5 oz and 2.5 oz have overlapping confidence intervals. This means that, based upon the data from this trial, the treatments did not perform differently

Once all of your calculations have been completed, make a **summary table** similar to that in Table 2. This table shows a letter following each mean value, something you will frequently encounter when reading scientific papers. These letters are a way of illustrating which confidence intervals overlap, and which don't. For example, values (such as 1.3%, 1.7% and 2.3%) followed by the letter "a" have overlapping confidence intervals and are therefore not statistically different from one another. In contrast, values that are followed by different letters ("b" or "c" in the case of Table 2), are statistically different from those followed by "a"s.

In fact, all of the information required to determine which treatments are best, which are worst, and which are the same, is contained in Table 2, but it's difficult for most of us to read tables. That's where graphs come in.

One Picture is Worth aThousand Words

One of the best approaches towards interpreting results is to graph the information. There are two types of graphs that are used to illustrate data collected from field trials such as the fungicide trial described above, **the bar chart** and **the line chart**.

For either type of chart, there are two axes, or lines, that define the chart — the horizontal axis, also called the "X" axis, and the vertical axis, also known as the "Y" axis. Figure 1 illustrates the results of the dollar spot fungicide experiment presented in a bar chart. The X axis has no numerical units, just treatment names. The Y axis represents the mean percent dollar spot values presented in Table 1. Thus, the bar for the non-treated check is the tallest bar, registering at 19.3% dollar spot. The vertical lines extending above and below the tops of each bar are called **error bars** and represent the confidence intervals for each treatment mean.

We suggest that you always try to graph your data. You can plot the results by hand using graph paper, or you can let a spreadsheet program on the computer do it for you, automatically.

Finale

This series of articles has described simple methods for designing a field testing program — from developing an experiment plan with clear objectives, to executing an experiment, to analyzing the results. Although this is a cursory look at the scientific process, we hope it encourages you to begin, or if you have already started, to continue testing new ideas. Remember to take care to record your objectives, materials

and methods, results, and conclusions. The next time someone asks you why you selected a particular practice or product, you may be able to pull a notebook from the shelf and point to a graph illustrating the advantages of your approach. Aside from personal pride, there is no better way to answer an agronomic practices question than to run a carefully designed, simple experiment.

References:

Little, T.M., Hills, F.J., 1978. Agricultural experimentation. John Wiley and Sons, NY. 350 pp.

Stowell, L.J., and Gelernter, W. 1997. How to test products and practices: Part I, Getting started. Calif. Fairways, November/December, 1997. pp. 20-22.

Stowell, L.J., and Gelernter, W. 1998. How to test products and practices: Part II, Experiment design. Calif. Fairways, Jan/Feb 1998. pp. ???

Table 1. Results of a fungicide trial for control of dollar spot on creeping bentgrass. Rates of fungicides are represented as ounces of active ingredient per 1000 sq ft (oz ai/M). Check refers to the non-treated check plot.

	Procymidone oz ai/M		Iprodione Oz ai/M		Check
	0.5	1.5	2.5	2.0	
Percent dollar spot (replicate 1)	5	2	2	2	13
Percent dollar spot (replicate 2)	7	1	1	4	25
Percent dollar spot (replicate 3)	2	1	1	1	20
Total	14	5	4	7	58
Number of replicates	3	3	3	3	3
Mean (=total ÷replicates)	4.7	1.7	1.3	2.3	19.3
Standard Deviation (S)	2.5	0.6	0.6	1.5	6.3
Confidence Interval (mean - S) to (mean + S)	2.2 - 7.2	1.4 - 2.0	0.7 - 1.9	0.8 - 3.8	13.0 - 25.6

Table 2. Summary of dollar spot control results using tabular format. The numbers in the percent Dollar Spot column followed by the same letter are not significantly different using the standard deviation as the confidence interval. If the confidence intervals overlap, it is unlikely that the means are different.

Treatment (oz/M)	Mean % Dollar Spot	Confidence Interval
Procymidone 2.5	1.3 a	0.7 - 1.9
Procymidone 1.5	1.7 a	1.4 - 2.0
Iprodione 2.0	2.3 ab	0.8 - 3.8
Procymidone 0.5	4.7 b	2.2 - 7.2
Check 0.0	19.3 c	13.0 - 25.6