Statistics and The Science of Turfgrass Research

By Tom Schwab, Superintendent
O.J. Noer Turfgrass Research and Education Facility

What comes to mind when you think of statistics? A most prevalent use of statistics is relating it to sports activities. A batting average, free throw percentage, use of handicaps, and a win/ loss record are all forms of statistics. Many of them are used to predict the outcome of a given event. My favorite statistic is the last time the Packers started out 8 and 1 they went on to win the title. The science of statistics is used to make unbiased inferences about outcomes in many walks-of-life such as sports, but also business, medicine, engineering, law, education, and turfgrass management.

The basic concept behind statistical inference is to observe a small part of a large group and from that small group make a prediction of the whole. Two simple terms need to be defined here, population and sample. The large group or real life situation which you are really interested in is considered the *population*. The small group or study from which you will be drawing inference to make a prediction on the population is called the *sample*. At the Noer Facility a *sample* might be to observe a bentgrass putting green that has three different mowing heights. We may be investigating whether the different mowing heights relate to having more or less disease incidence. The *population* of interest in this study would be all bentgrass putting greens.

A criticism that a study like this sometimes receives is that this putting green is not like the real world where a green may also have 60,000 rounds of golf played on it. That is true, but this study was not designed to study the whole real life situation. It was designed to investigate one aspect. Then you can take this one aspect and consider how it applies within your whole management scheme.

With careful planning and design of the experiment, using statistics, the researcher can quantify a claim that a certain input will have a chance of giving a certain effect. For example, they can state with a certain level of confidence (say 95%) that XYZ product will increase rooting on native soil bentgrass putting greens. Statistics uses calculations to give a numerical value of whether this difference is meaningful. Then you can judge whether with numerical value is enough for you to be confident in the claim.

Oftentimes we hear from advertisers about how well a new product works. Sometimes we are given testimonials as proof such as, "Joe at Bigwig Golf Links used schmuckum juice to speed germination and his members were playing the new green two week after planting." Because most of us value the scientific method over hearsay we can't be convinced of these claims unless we know more about the testimonial. You rarely hear any statistical techniques mentioned when you hear a testimonial. A testimonial could be believable if the following occurred: the same outcomes were repeated at many different sites, at each site individual treatments were replicated manipulations that you are observing in a study. For example in a fungicide study, each different product would be considered a separate treatment. Then you may be observing which treatment lasts the longest.) When analyzing statistical results: find out what tools of measurement were used in analyzing the data, and question whether the conclusions are biased by the people making the claims.

Many of the ideas behind statistical procedures seem like common-sense yet their importance in designing an investigation and producing factual analysis can not be underestimated. I can also tell you that taking a graduate level statistics course is significantly harder than most undergraduate courses I took 15+ years ago. For this novice, although many of the ideas seem obvious, their actual examination and calculation are very complex. There are many terms and procedures used in statistical science when analyzing data such as Null Hypothesis, Significance Testing, Confidence Interval, Power, Least Significant Difference, Sampling, and many others. I'll try to explain what some of these concepts mean.

Most statistical analysis starts out with a *null hypothesis*. That is an expectation that there is no distinction between the different treatments that you are observing. If there is an observed difference, statistics can be used to determine the probability that

this difference is a chance occurrence. If the probability, due to chance, is calculated to be very small than we conclude that the difference is due to the treatment. In actuality, what you most likely want to find is that there is difference between treatments. You would love to learn that one fungicide, for example, will last significantly longer than all the others, unless it costs an arm and a leg. In other words, you'd like to reject the null hypothesis which states that "the treatments are the same."

The statistical technique used to determine whether to reject or accept the null hypothesis is *significance testing.* Significance is a measure of probability that a certain difference is not due to chance. If the probability that these results could have occurred by chance is calculated to be extremely low, much less than 5%, then the null hypothesis (treatments are the same) should be rejected. If the null hypothesis can be determined not to be true, then you can be pretty sure that this observed treatment will be different from other treatments. Significance is just a way of telling you to what degree the treatment contributed to the difference. Then you can judge for yourself if the calculated significance value will make you believe that the particular treatment will make a difference.

Another similar approach to derive information from statistical inference is to calculate a *confidence interval.* A confidence interval is a plausible range for the true mean of the population. Say you wanted a type of turfgrass that could survive underwater for periods of time because you have some lowlands on your golf course that get flooded. You pay a researcher to investigate five different types of grasses that may be able to withstand submersion. They may find ABC grass can be underwater and survive for a period of one to seven days. The confidence interval in this case is one to seven days. (This is in contrast to significance testing discussed above which would tell you that ABC grass is distinctly better in it's flooding tolerance than the other four turfgrasses, but it would not tell you the upper and lower limit.) If that one to seven day interval is too broad and you absolutely need to know that a grass will last from three to five days underwater, that is over a narrower interval, the researcher could direct a study towards achieving this. This investigation would be more costly because the sample sizes would have to be much larger to achieve the same amount of confidence in the prediction. Thus more lab time would be needed and the possibility that no turfgrass can confidently be recommended to withstand that interval may be the conclusion. The confidence interval produces a plausible upper and lower limit where most of the data in the population will fall between. This confidence interval approach is related to significance testing because significance will also tell you to what degree the particular range predicts the population.

Power is the name of another statistical calculation. Power considers that the sample results may not be an exact prediction about the population. The power calculation theoretically asks, "If the true population is different from the null hypothesis, to what degree of probability can the test results be expressed?" Say that Organosmellystuff (OSS) fertilizer is expected, from previous knowledge, to keep Kentucky bluegrass growing vigorously for an average period of 40

(Continued on page 19)

(Continued from page 17)
days. A study is planned to see if this expectation is realized. The study finds that OSS lasts for an average of 45 days. The researcher of the study may ask, "If the true population average longevity of OSS fertilizer would be 40 days, then with what probability can I make the statement that my expected longevity of 45 days is true?" Power is a measure of experimental sensitivity that determines how likely the sample results are correct based on previous knowledge about the true population. Power and some basic algebra can also be used to help you determine how large a sample size, or how many replications that you will need to achieve a certain level of probability in your results.

Least Significant Difference (LSD) is a very common statistics calculation used when analyzing data for a turfgrass research report. LSD is a calculated range which tells you that any treatments separated by this value or less are not to be considered different. Their difference is only a matter of experimental error that occurs by chance. For example, let's say you want to look at a National Turfgrass Evaluation Program to see which cultivar of fine fescue performed best. If you see a calculated LSD of 0.8 in the report, then a cultivar that is rated 6.6 is just as good as one rated 7.3. If two cultivars are rated with a difference of 0.9 or greater, then the higher rated one can be considered significantly better in this example.

Sampling is a process which explains that the larger the sample size the closer the study will approximate the true population distribution. This means that the more times a given treatment is replicated in a study the closer it will approximate the true population. If a study continues for a longer period it is also considered more accurate. This makes sense because if a study takes place over many years the climactic extremes that are experienced in the short term are balanced out. Longer studies are obviously more expensive so the cost/benefits have to be weighed with the degree of confidence that you are trying to achieve. Another example is if a researcher is trying to prove his conclusion to a very high degree to a very critical peer group then he would need to replicate the study many more times than if he was just setting up a demonstration for a group of non-professionals.

There is a difference between a demonstration and an experimental research design. Experimental research designs need consistent and controlled data collection and statistical analysis. In contrast, demonstrations often lack that sophistication, but they are still very important. An example of one of these demonstrations at the Noer Facility is our turfgrass variety plot. We planted 20 different types, mixtures, or blends of turfgrass that grow in our upper midwest climate for homeowners to observe. There is plenty of information already known about many of the characteristics of these turfs. To investigate a certain characteristic to yet a higher degree of confidence than is already available would take considerable amounts of time, effort, and money. Like most demonstration plots, this variety plot was inexpensive to install and yet it provides great information for non-professionals like homeowners. Non-professionals often just want more general information about what varieties to use. Demonstrations are also important in that they may, by chance, present a characteristic or input that should be investigated in much more detail in an experimental research design.

The Noer Facility has many demonstrations as well as experimental designs. An example of the latter is a study that has 15 different treatments with each treatment replicated 10 times. That study was designed by a department statistician. There is another study at the Noer Facility that will take place over ten years. Those are the kind of *samples* that will be able to make very accurate, statistically confident statements about the whole *population.*

There are many products that claim to positively effect turf health, such as decrease thatch, increase rooting, reduce pesticide applications, enhance winter hardiness, increase green speed, and on and on. To make those types of claims, a product needs to be evaluated using some statistical knowledge. If a salesperson or researcher can show you the statistical evaluation done in studying the product then you'll have more confidence in trying it. Every consumer needs some basic understanding of statistics in order to question and understand the claims that are made about a product or practice. Statistics, in spite of its complicated calculations, strengthens our understanding in many walks-of-life, from turfgrass management to sports. I just hope the original Packer outcome prediction holds true. The way that they're pounding the Bears today, as I write this, the prediction looks true with 100% confidence.

References:

Unpublished *Course Notes for Statistics/Forestry/Horticulture* Madison, Wisconsin Erik V. Nordheim & Murray K. Clayton

Statistical Methods for Food and Agriculture Filmore E Bender, Larry W. Douglass, Amihud Kramer

Statistical Methods in Agricultural Research Thomas M Little, F. Jackson Hill