CONDUCTING AND UNDERSTANDING EXPERIMENTS IN MAPLE OPERATIONS. *Reprinted from Farming, the Journal of Northeast Agriculture. February, 2011*

At the University of Vermont Proctor Maple Research Center, experiments take place using a variety of methods and in many locations—wires and probes stuck in trees, experimental tubing arrangements across the sugarbush, calibrations of boiling efficiency in parts of the evaporator, special taste testing panels, and so on. These types of research studies must follow certain rules in order for the findings to be valid. In this column I will discuss a few of these simple rules: comparison of treatment vs. control, replication, dealing with natural variation, and statistical validity. These rules are very useful for sugarmakers who might want to try their own experiments and use their findings to help improve the efficiency of their operations.

Suppose that we wanted to find out if white spouts yielded more sap than black spouts. Perhaps this idea was inspired by thoughts that black spouts heat up in the sun, and breed more microorganisms on warm days. We could compare a black plastic spout with one of the many non-black spouts currently on the market, such as a stainless, or clear polycarbonate spout—but these spouts, in addition to being non-black, are made of other materials and are different in other ways from a black plastic spout. In order to answer the question properly, we would need to compare two spouts that were the same in every way except color. The black spout would be the *control*—which is the part of the study that we collect data on, but haven't changed in any way. The *treatment* is to color a black spout white. This is fundamental—there always has to be a control which is exactly the same as the treatment, except for the one factor that is being examined.

In this study, sap yield is being measured in order to determine if black spouts and white spouts are different. If we install black spouts in some trees and a white one others, then everything else about the trees and the installation must be as similar as possible—for example, the same depth, aspect, timing, etc. of all tapholes. The trees should be roughly the same diameter, and as close to each other as possible to avoid differences in microclimate. Sap yield even from adjacent trees of the same size is likely to be different, because sap flow is naturally quite variable from tree to tree. Thus, we need to make this comparison of black and white spouts on many trees. This is the important concept of *replication*. Perhaps we could compare 20 trees with white spouts and 20 trees with black spouts. If we collected sap with buckets, we could measure the volume in the bucket every day, and at the end of the season we might have some worthy results.

Things are rarely this simple, and there are many pitfalls to this or any experiment. First, we must make sure that we didn't bias the study in some way by, for example, painting the black spout white and introducing the paint into the taphole where it might have had unintended effects. Second, we only compared the spouts in one season, and results might be quite different in another year, with another temperature regime in the spring. Third, we might have gotten results that seem inconclusive—in some trees the white spout yielded more sap, in some trees the black yielded more. We could sum all the sap from the white and black spouts and report which one totaled more, but this would not be valid. We need to determine if the difference between the amount of sap collected from the white and black spouts is greater than would be expected by random variation among trees in the amount of sap they produce. In this case we should perform a statistical test that takes all the sap yields from each separate tree into account. The more replicates—separate trees we used for each treatment—the more likely that we can

determine the difference between chance and a real distinction between spouts. Our test will tell us if the differences between the two sap yields are *statistically significant*, which would mean that we can be confident that sap yields from black and white spouts are in fact different.

In the experiment described above, we collected sap by gravity. Many sugarmakers use vacuum exclusively, and we cannot be sure that any differences between black and white spouts will be similar for gravity and vacuum. Replication becomes much more difficult and expensive with vacuum, which is why this kind of experiment is more likely to be done as a grant-funded study. At the Proctor Center, vacuum chambers, described in a previous column, are used to isolate the sap from a single taphole, but these chambers are costly. Some people might attempt this experiment under vacuum by collecting all of the sap from white spouts in one tank, and from black spouts in another tank. Even though each tank might be connected to a hundred or more spouts, this does not mean that the experiment is replicated a hundred times—instead it consists of one sample from each type of spout. While it might be interesting to note that one tank had more sap than the other, there is no way to be sure that the white-spout trees are not in a part of the stand with a different microclimate from the black-spout trees, or that some other factor other than spout color made one group of trees yield more sap than the other. If there were many tanks available, each collecting sap from many black or many white spouts, then the experiment would be similar to the one described above for gravity, only substituting tanks for trees. This would be expensive, but in this case we could again perform statistical tests on the results, and our findings would be statistically valid.

We encourage producers to do experiments in their own woods or sugarhouse. If you have an idea that you want to try out, and need some help designing your experiment, feel free to contact me at the Proctor Maple Research Center, 802-899-9926.