Embracing Variability: How to Hug a Cactus

Most agricultural engineers, including yours truly, enjoy working with statistics about as much as undergoing a root canal. This statistical aversion starts early in our careers. For example, at every conference or a student's dissertation defense, there's always that unobtrusive statistician seated in the middle of room, who keeps silent until the data slides appear. Meanwhile, the student feels comfortable, since the committee seems to appreciate the slides in which the instrumentation, algorithms, and software are discussed. But with the appearance of the data, it's time for the statistician to rant. He slams the experimental design (or lack thereof), he criticizes how the analysis was done, he gripes about the lack of skewness and kurtosis measures, and he condemns the improper testing, the lack of statistical power, the improper use of significance measures, and the conclusions based on small regression coefficients. The student is stunned, and in desperation looks to the rest of the committee for help, but the committee members are too busy trying to remember whatever statistics knowledge they still have on tap since they graduated themselves.

Engineers tend to make a mess of statistics. We do a poor job even in basic work like regression analysis, and often we don't go beyond simple means and standard deviations. We report numbers with too many significant digits (often by converting from English to metric units) and without error margins, mostly because we either have only one rep or we don't want to draw attention to imperfections in the data. This is inherent in our way of thinking: There is something irreconcilable between engineering and statistics. As engineers, we develop systems in which we try to control everything, and we measure phenomena as accurately as possible. Our mission in life is to beat errors into submission, not to embrace them! In college, we all took classes in linear differential equations, with their beautiful, smooth solutions. But now we have to make sense of a bunch of dots with a regression curve. Every time I look at a regression plot, I wonder: where did we go wrong and who is to blame?

That's a bad habit, I know. Variability is the essence of nature, and I have been doing this kind of work for 20 years, so I should be able to accept the reality of it. But, alas, I can't. I have many arguments with my colleagues in crop sciences, who are sometimes satisfied with a regression coefficient of 0.5. I tell them that they don't understand half of the problem, and I usually get yelled at for that. This is another trait of engineers: we want to understand the whole thing. In fact, our designs require complete understanding before we can build a reliable artifact. Science, as it is pursued in crop sciences, is different. The crop sciences crowd seems to be content with understanding half of the problem. But maybe that's a misconception on my part.

Before pursuing my doctorate, I used statistics in a "they tell me I have to so I will" kind of way. I vividly remember how hard it was to understand concepts in books whose titles always seemed to start with a misleading "Introduction to...". Student's t-test to compare means was about as far as I would go, and forget about anything multivariate. However, during my doctoral studies, I learned a little more about statistics. I worked on a sensor for granular mass flow measurement with Kate Crespi, a bio-statistician at UCLA. She took me to the cleaners many times when I tried, once again, to transform stochastic circles into deterministic squares. With Kate's help (a lot of it, I might add), I realized that, to measure a granular mass flow accurately, it is essential to have a fully developed, Poisson-driven, random arrival process. In other words, make the biggest mess possible, and then the theory (and the measurement) actually works! It seems to me that we only have a good grip on problems when they are either completely deterministic (Engr-land) or completely stochastic (Stat-land), but what if they are somewhere in between? Fortunately, with the exception of self-organizing systems, if we leave a system alone, it tends to become more disorderly with time: Just wait a while, and you'll be drawn closer to Stat-land. This realization is not that intuitive, and it even baffled the editor of a prestigious journal, who rejected a paper that Kate and I wrote on the subject, with the comment that our data was "too good to be true". This same paper received the Biosystems Engineering Outstanding Paper Award in 2010, which is awarded to one out of a hundred papers. In the end, this confirmed to me once more that data are much more valuable than any one person's opinion.

After all these years, I still have not become a statistics fan, but at least I can appreciate the work of statisticians, and I have developed a détente with them. I hope this editorial can convince some engineers that statisticians are like cactuses. They can be very helpful when approached with care. Have you hugged your statistician today?

ASABE member Tony Grift, associate professor, Department of Agricultural and Biological Engineering, University of Illinois, Urbana, USA; grift@illinois.edu.