

Statistics Spotlight



STATISTICAL ENGINEERING

The Foundation Of Statistical Engineering

How Ronald A. Fisher's work feeds into statistical engineering

by Lynne B. Hare

Physics has been around since Isaac Newton (1642-1727) and before, depending on whom and how you count. Chemistry has been around since Merlin, a fiction of 1136. Statistics, especially modern statistics, has been around for about 120 years. No wonder it was slow to catch on in some quarters and still hasn't been accepted in others.

One impediment to acceptance has been the reluctance to forgo determinism in favor of empiricism. When Newton broke down light into its color component parts and when he studied the music of the spheres, he thought in terms of absolutes—orbital paths can be measured and predicted to the nearest nano. Evidence is that Copernicus (1473-1547) and Galileo (1564-1642) thought the same: no room for variation.

Another impediment is the failure to recognize a multiplicity of factor activity. To this day, chemists teach their students that to determine the effect of a factor in a chemical experiment, you must hold all other factors constant and vary only that one factor. While this might technically be true, it is not the only truth: This technique ignores boundless opportunities to harvest knowledge using planned experiments

simultaneously involving multiple factors. Further, it is oblivious of three fundamentals:

1. It fails to recognize that responses may have more than one peak and one valley.
2. It assumes that factors are independent of one another—that there are no interactions or joint effects.
3. It assumes, contrary to experience, experimental variation is small relative to the response.¹

Vestiges of this old-world thinking persist to this day. They can be seen in the surprised looks on the faces of participants in industrial short courses about data-driven decisions and strategies for process improvement as they learn that it "ain't necessarily so."

It helps to know something about the genesis of modern statistics.

Fisher's influence

During the early 20th century, England found itself unable to feed its population. Insufficient agricultural yield was augmented by imports. While agricultural experiment stations existed, their research was helter-skelter by modern standards. Enter onto the scene a young mathematician, a Cambridge University honor graduate.

With degree in hand, Ronald A. Fisher left the university to work for an investment organization, entered farming and then school teaching. None of these occupations suited him, and his mentors and tormentors weren't impressed with his performance. Some say he wasn't tactful. Others say he didn't suffer fools lightly. He returned to academe.

Further in his research into modeling genetic behavior (for which he was knighted), Fisher didn't see eye-to-eye with the key statistical leadership of his day, Karl Pearson. He dropped his early work on eugenics because he saw it was leading nowhere, but that activity further ostracized him from the statistical mainstream even though its members were greatly impressed by his mathematical skill and insight.²

In 1919, Fisher was offered a position at the Rothamsted Agricultural Experiment Station where staff had accumulated a mountain of data but had little knowledge of how to extract its meaning. In essence, he had encountered a large, unstructured data set. While he spent years poring over Rothamsted's volumes, it is doubtful that he learned much of real value from them.

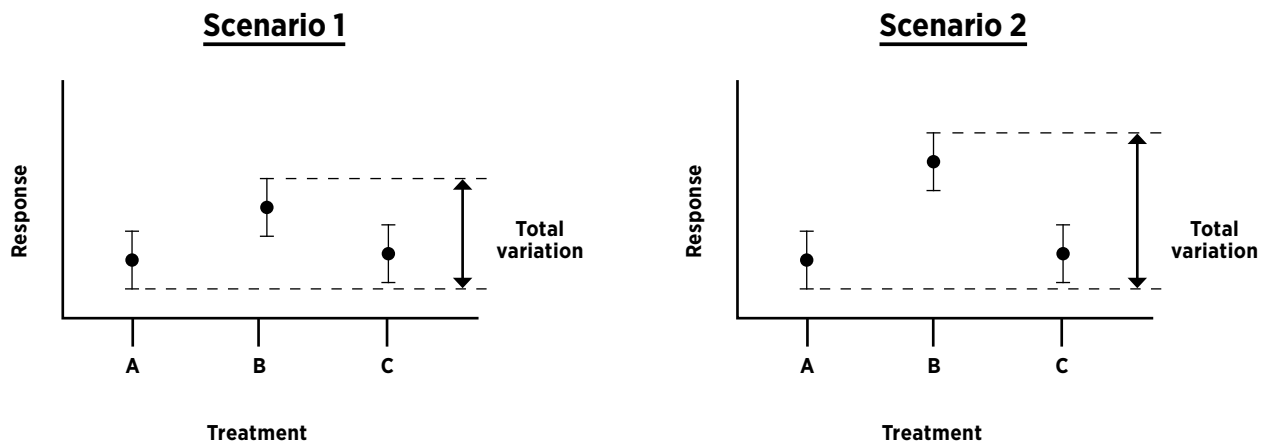
The problem, in brief, was one-factor-at-a-time experimentation. Information-less data often were the consequence of attempting a lone, sans control, experiment one year with a follow-up characterized by a slight variation the following year. Of course, the yearly



Statistical engineering can build on the foundation set by Ronald A. Fisher, a British statistician and geneticist.

FIGURE 1

How the analysis of variance identifies mean differences



A Formal ASA Statement Concerning P-Values

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

—American Statistical Association (ASA) press release, March 7, 2016

difference was influenced by many of Mother Nature's favorite attributes, including shifting temperatures, varying rainfall and differing soil conditions.

What Fisher engineered were sensible, planned, multifactor experimentation techniques that minimize the effects of natural and human-induced variation through randomization, balance and replication. He sat back while those clinging to their old ways raged. Experimental design had been used before, but randomization—especially outside the laboratory—was foreign to a deterministic world.

“Those who don't take randomization seriously are condemned to make conclusions at random.”

—Stephan Senn, “The Revenge of R.A. Fisher: Thoughts on Randomgate and Its Wider Implications,” presentation, July 17, 2017, www.slideshare.net/StephenSenn1/the-revenge-of-ra-fisher.

Then he explained, in mathematical terms, how this new way of experimentation works and what it brings, and he introduced what he called the analysis of variance. Explanations were later published³ with more underpinning mathematical detail to follow.

You might wonder how information about variances might impart knowledge of differences among means. Consider alternative scenarios as in Figure 1 (p. 49). Here you see three

treatment means depicted graphically and surrounded by statistical intervals.⁴ In scenario one, the intervals overlap, hinting that the variation among the means may be due to chance alone. The opposite is true in scenario two: the intervals surrounding the means fail to overlap.

Also notice the total variation—as depicted in each scenario as the distance between the lower bound of the lowest mean and the upper bound of the highest mean—differs between scenarios. It is much larger in scenario two because of the differences among means. This is because the total variation in all the data is a function of the variation associated with an individual treatment combined with the variation among the treatment means. Fisher's genius is in recognizing this phenomenon and generalizing it to multiple factors, their levels and their interactions. Hmm ... why didn't we think of that?

Largely due to Fisher's genius and advice, and due to the diligence of those who followed it, England now feeds itself. But it didn't end there, and it certainly didn't end with agricultural experimentation, while many detractors thought it should.

The propagation of the use of statistics

After World War II, George E.P. Box and J. Stuart Hunter, accompanied by various others and sponsored in part by ASQC's Chemical Division (now ASQ's Chemical and Process Industries Division), traveled the United States preaching the gospel of statistical methods, especially the statistical design of experiments, greatly to the influence of chemical, automotive and other technically based industries.^{5,6}

This sparked a revolution in the approach to all manners of experimentation beyond those initial applications and extending to pharmaceuticals, foods, electronics, healthcare, psychology and countless other activities. Their efforts changed the way the world engages in scientific inquiry.

Fisher's genius extended beyond those already mentioned. He brought us the concept of statistics itself. Earlier, Karl Pearson had argued that data, even sampled data, contained all the necessary information for decisions, whereas Fisher's

view—the one that stuck—was that sampled data are used to obtain statistics, which are estimates of parameters of a larger population.

Fisher also brought us the notion of degrees of freedom, discriminant analysis, maximum likelihood estimation and many other original ideas. He is correctly called the Father of Statistics (and the father-in-law of George E.P. Box).

For the record, Fisher was not a fan of the formalized Neyman-Pearson school of statistics. Jerzy Neyman, a Polish-born mathematical statistician, and Egon Pearson (yes, Karl's son) taught together at University College in London. Together, they espoused a rigorous, highly probabilistic form of hypothesis testing, which is taught today in beginning statistics courses and is highly embraced by government agencies globally and nearly worshiped in many industries, including pharmaceutical research.

While Fisher had done the math and calculated exact probabilities (for example, F-distribution tail areas), his approach to research downplayed the probabilities in favor of allowing seemingly rare or near-rare events to guide further research. He advocated using what he named “p-values” to categorize factors into “significant” (meaning potentially influential), and others which did not emerge above the noise inherent in an experiment. Those others were to be kept in mind for future experimentation.

It is likely that he did this because of the murky waters of probability itself. Philosophically, what is probability? How can you embrace the calculation of probabilities based on a “null hypothesis” with a related experiment that would never be run if those in control really believed it is true?

For more on the shaky ground of probability and hypothesis testing see Herbert I. Weisberg, *Willful Ignorance*, in which the author offers suggestions regarding why the results of some population studies are not reproduced in follow-up studies.⁷ And for more about the perils of reliance on p-values, see Ronald L. Wasserstein and Nicole A. Lazar's article on the American Statistical Association's (ASA) statement on p-values.⁸ You can also read ASA's actual summary of main points in the sidebar.

In Fisher's footsteps

We build on the foundation Fisher set. Research into new and more fitting experimental design techniques persists, and many of the tools embedded in such initiatives as total quality management and lean Six Sigma are of Fisher's derivation and are being distributed broadly.

This is also true of the strong and increasing leadership role taken on by many statisticians. Recently launched is a new

professional society, the International Statistical Engineering Association.⁹

Statistical engineering is to statistics as chemical engineering is to chemistry. Its focus is on integrating scientific tools and engineering to solve large, complex and unstructured problems like those that Fisher faced when he accepted his position at the Rothamsted station.

Formally, statistical engineering is the study of the systematic integration of statistical concepts, methods and tools—usually with other relevant disciplines—to solve important problems sustainably. It is a proactive leadership and technological force, following in Fisher's footsteps.

He would be proud. **QP**

REFERENCES AND NOTE

1. Roger W. Hoerl and Ronald D. Snee, *Statistical Thinking: Improving Business Performance*, second edition, John Wiley and Sons, 2012.
2. David Salsburg, *The Lady Tasting Tea*, W.H. Freeman and Co., 2001.
3. Ronald A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, 1925.
4. Gerald J. Hahn and William Q. Meeker, *Statistical Intervals*, John Wiley and Sons, 1991.
5. George E.P. Box, J. Stuart Hunter and William G. Hunter, *Statistics for Experimenters*, John Wiley and Sons, 2005.
6. George E.P. Box, *An Accidental Statistician*, John Wiley and Sons, 2013.
7. Herbert I. Weisberg, *Willful Ignorance*, John Wiley and Sons, 2014.
8. Ronald L. Wasserstein and Nicole A. Lazar, “The ASA's Statement on p-Values: Context, Process and Purpose,” *American Statistician*, Vol. 70, No. 2, 2016, pp 129-133.
9. For more information, see the International Statistical Engineering Association's website at <https://isea-change.org>.

ACKNOWLEDGEMENT

The author thanks Ronald D. Snee for insights and contributions to this column.

Copyright 2019 Lynne B. Hare.



Lynne B. Hare is a statistical consultant. He holds a doctorate in statistics from Rutgers University in New Brunswick, N.J. He is past chair of the ASQ Statistics Division and a fellow of ASQ and the American Statistical Association.