

Statistical Thinking

Lynn D. Torbeck

The past 25 years have seen many advances in the field of statistics. Some of the better known relate to the advances in computing — better and faster software, for example. Computationally intensive methods have become possible, including bootstrapping, resampling, and permutation tests. Meta-Analysis also is a recent addition to the set of statistical tools. A less well-known advance that promises to play a large role in the future is the formation of a set of principles referred to as *statistical thinking*. Although not an advance in statistical theory, it is nonetheless a substantial advance in the application of statistics in industry. Of the changes in the field of statistics in the past 25 years, statistical thinking has the potential to be the most useful and beneficial. Perhaps more important, it seems to be on a solid foundation and not just another flavor-of-the-month fad.



Lynn D. Torbeck

Some statistical techniques have been known for a long time: regression analysis, probability theory, and the normal distribution date from the 1700s. However, statistics as a field only began to take shape in the late 1800s in London as biologists and geneticists became increasingly quantitative in their research. Karl Pearson (1857–1936) (1), his son, Egon Pearson (1895–1980), and Francis Galton, (1822–1911), a cousin to Charles Darwin, were leaders. Sir Ronald A. Fisher, (1890–1962), Jerzy Neyman, (1894–1981), and many others made major contributions to the theoretical basis of statistics. In the United States, Walter A. Shewhart (1891–1967) developed the control chart (2) at Western Electric, while W. Edwards Deming (1900–1994) and many others helped develop the application of statistics to quality control. The Kepner-Tregoe approach to systematic problem solving in the 1950s helped those in industry to formulate problems and seek pragmatic solutions. The introduction of calculators and computers in the 1960s and 70s made statistical computations more palatable. Arguably, the first cookbook for statistics was published in 1963 by Mary Natrella titled *Experimental Statistics* (3). In Japan the very well-known book and pragmatic *Guide to Quality Control* was published in 1968 by Professor Kaoru Ishikawa (4).

About 1975, the US industry was astonished to discover that the Japanese industry had exceeded the United States in quality and productivity. As US companies tried to catch up, many fads in quality were created. The year 1977 found the publication of John Tukey's landmark text *Exploratory Data Analysis* (5), which has been very influential in changing applied statistics. On 24 June 1980, NBC aired their white paper, "If Japan Can ... Why Can't We?" This featured Ed Deming as the statistician to the Japanese industry and the source of many of their approaches to quality. His classic book *Out of the Crisis* was published in 1982 (6). The pursuit of Japanese fads in quality and statistics continued into the 1980s and early 90s with many of the philosophies, ideas, and techniques becoming ingrained into industry practice. Some of this was formalized in a number of books, with *The Team Handbook* by Peter Scholtes being one of the best and the most popular (7).

The term *statistical thinking* traditionally has been applied to the rise of the field of statistics, but more recently it has taken on a broader meaning. This newer meaning was discussed in 1991 by Ronald Snee in his paper "Statistical Thinking and Its Contribution to Total Quality" (8). This manner of thinking was further promoted by the Statistics Division of the American Society for Quality (ASQ) in 1994 when they set a goal to "enable broad application of statistical thinking" (9,10). Their efforts resulted in the book *Improving Performance Through Statistical Thinking* (11).

Looking beyond the formulas

Too often the field of statistics has been seen as abstract theory with little application or relevance to real industrial problems. Historically, there has been a love/hate relationship between the field of quality and the field of statistics. It's true that an overemphasis on math and formulas has placed many statistical tools and techniques out of the reach of practitioners, but the roots of the problem go to the American educational system. Statistical concepts are not taught in grade or high school to any degree. Students don't encounter the formal concept of randomness and variation until their first college course in statistics. By then, statistics simply becomes another hurdle to be surmounted for graduation. Further, college professors seldom include statistics in subject matter courses such as chemistry or biology in the same way that, say, algebra is used. This results in industry's poor appreciation for statistics at both the technical and the managerial level. A few highly motivated staff will use statistics, but in general, the benefits of statistics are not known, understood, or utilized.

A solution is to look beyond the formulas and focus on the underlying ideas, concepts, and the philosophy of applying statistics in industry. There needs to be a connection to the applications and to the big picture. The mathematics and formulas are presented as the need arises, not up front, and out of context. Computers and current software facilitate understanding by making graphics and analysis easy and interesting.

The working principles of statistical thinking

The three key principles in statistical thinking, as stated by the ASQ Statistical Division, are philosophical and fundamental:

- All work occurs in a system of interconnected processes.
- Variation exists in all processes.
- Understanding and reducing variation is key to success.

These three key principles can be expanded further into 10 concepts:

- Work occurs in systems of processes and subprocesses of interconnected and interrelated steps, known as the SIPOC model: processes have Suppliers that provide Inputs into the Process activity; the end result is the Output that then goes to the Customer. This model provides a global overview useful for broadly defining the process.
- Processes can be mapped (12), flowcharted, studied systematically, understood, and improved. However, optimizing each step separately can result in a suboptimum process.
- Work is conducted by teams of people with differing backgrounds, education, expertise, skills, needs, and expectations.
- Process outputs vary as a result of both special or systematic causes and common or random causes. These can be found, studied, and understood. Note that dysfunctional teams are a source of variation in the system.
- Cause-and-effect relationships are the foundation of science. These relationships can be found, studied, quantitated, and understood.
- Variability is the enemy of CGMPs, validation, quality, productivity, efficiency, and profits.
- Variability can be measured, studied, and understood.
- Statistics is the science of variation.

- Variability in processes can be reduced. It is not fixed. Adopting a philosophy of running to target and working for consistency is needed.
- Organizations succeed by continuous improvement using teams to reduce variation and bring processes into statistical stability.

By connecting statistics to real industry problems using these 10 concepts, statistical thinking can achieve the integration of management, funding, problem solving, and statistical techniques. Applications abound in manufacturing, production, and quality assurance and control. Management processes such as purchasing, customer complaints, and accounts payable can also be studied with the same approach.

Conclusion

Statistical thinking is the summation of the history of problem solving, statistics, and quality. It combines the best from many areas and fields into a coherent and internally consistent approach. Of the changes in the field of statistics in the past 25 years, it has the best chance of enduring. The benefits of implementing statistical thinking into industry processes include fewer out-of-specification results, fewer rejected and recalled batches, reduced cycle time, faster and better validations, improved productivity and efficiency, and, of course, better quality for the consumer.

References

1. N.L. Johnson and S. Kotz, *Leading Personalities in Statistical Sciences* (John Wiley & Sons, New York, 1997).
2. W.A. Shewhart, *Statistical Methods from the Viewpoint of Quality Control* (Graduate School of the Department of Agriculture, Washington, DC, 1939, Dover edition 1986).
3. M.G. Natrella, *Experimental Statistics* (National Bureau of Standards, Department of Commerce, Washington, DC, 1963, revised 1966).
4. K. Ishikawa, *Guide to Quality Control* (Asian Productivity Organization, Akasaka, Japan, 1968).
5. J.W. Tukey, *Exploratory Data Analysis* (Addison-Wesley Publishing Co., Reading MA, 1977).
6. W.E. Deming, *Out of the Crisis* (MIT, Cambridge, MA, 1982).
7. P.R. Scholtes, *The Team Handbook* (Joiner Associates, Madison, WI, 1988).
8. R.D. Snee, "Statistical Thinking and Its Contribution to Total Quality," *ASQ Statistics Division Newsletter* (Winter 1991).
9. G. Britz et al., "Statistical Thinking," in a special publication of the *ASQ Statistics Division Newsletter* (Spring 1996).
10. M.G. Leitnaker, "Using the Power of Statistical Thinking," in a special publication of the *ASQ Statistics Division Newsletter* (2000).
11. G.C. Britz et al., *Improving Performance Through Statistical Thinking* (ASQ Statistics Division, ASQ Quality Press, Milwaukee, WI, 2000).
12. D. Galloway, *Mapping Work Processes* (ASQ Quality Press, Milwaukee, WI, 1994). **PT**

Lynn D. Torbeck is an international consultant and trainer specializing in applied statistics and experimental design for pharmaceutical, biopharmaceutical, and medical device research and development, validation, quality control/assurance, and production, 2000 Dempster Plaza, Evanston, IL 60202, tel. 847.424.1314, www.Torbeck.org. He has been active in the pharmaceutical industry in statistics, quality, management, and validation since 1975. Mr. Torbeck is a member of the Editorial Advisory Board of *Pharmaceutical Technology*.