

Essays on *The Lady Tasting Tea: How Statistics Revolutionized
Science in the Twentieth Century* by David Salsburg

Quang Nguyen

Wittenberg University

Spring 2020

One of the themes you should notice from the book is the way in which the discipline of statistics has influenced and been influenced by other disciplines and by major events and forces in the world. Cite and describe some of these connections between statistics and other disciplines, events, and forces.

Today, statistics is widely used in numerous areas, from the politics and business worlds, to the sciences, or even in sports and other pastimes. Nevertheless, this relationship did not just become a common theme recently, as there is a long history of how the discipline concerned with the collection, summary, and analysis of data has influenced and been influenced by other fields of study, or major events around the world.

It is fair to say that the natural sciences had paved the path for modern statistics, as many familiar statistical concepts today were invented from scientific research. For instance, Francis Galton discovered the important concepts of “regression to the mean” and “correlation coefficient” while working with data collected at his own biometrical lab. In addition, R.A. Fisher, whom many believed to be the founder of modern statistics, created the theory of experimental design and introduced a now well-known method of analysis of variance when trying to solve problems in agricultural science. These discoveries all had immense impact on many subfields of science later on, as statistical models based on Galton’s regression method dominated many areas of study; or Fisher’s work on experimental design laid the foundation for different design rules being developed and applied in important fields like biology, chemistry, or medicine.

In relation to the social sciences, statistics had some considerable contributions to many subjects. In history, for example, statistical methods of textual analysis were used to determine the authorship of the infamous Federalist papers. The industry world also benefited from the implementation of statistical methods, as illustrated by W.E. Deming. For instance, the Japanese industry was transformed by statistics, as they, inspired by Deming, effectively executed methods of quality control as well as promoted the use of statistical methods to improve all activities. Regarding social issues, statistics has been used to illustrate healthcare questions, with Jerome Cornfield’s establishment of a cause-and-effect relationship between smoking and lung cancer, or the development of the martingale approach in medical studies to analyze congestive heart failure complications.

Statistics was impacted by major events during the twentieth century, specifically World War II, which is arguably the biggest global-scale event in the last 100 years. Statisticians were recruited to participate in research to solve strategic and tactical war problems everywhere in the world, from the Nazis in Europe to the United States. The term “operations research” was coined during this period, as scientists were trying to come up with statistical models of weapons usage or food supplies for soldiers. The American version of operations research, led by Sam Wilks, were able to apply statistical methods to describe to effectiveness of explosive weapons. Moreover, many statistical innovations were discovered, including the “sequential analysis” method, which is commonly used in sociology, medicine, and industrial quality control nowadays. After the war, the discipline of operations research took a further step and translated into the business world, as many operations research department were formed in most large companies.

Overall, statistics has had countless significant contributions to every subject in our lives such as science, business, industry, war, and social issues. This trend will certainly continue in the future, as statistics and the new, exciting, and emerging area of data science are expected to play a major part in every aspect of the world in this day and age.

The subtitle of our book is “How Statistics Revolutionized Science in the Twentieth Century.” So how did statistical thinking revolutionize science? Characterize the kind of scientific thinking before the statistical revolution, and describe the new way of scientific thinking that eventually superseded the old way.

Revolution happens everywhere, every day, and in every discipline we can think of. In addition to famous revolutions in political history such as the American Revolution or the Bolshevik Revolution, there are also notable revolutions in other aspects of our lives, from the technological revolution in the industry, to the recent three-point revolution in basketball. Among those significant events, the statistical revolution in science surely stands out, as it has completely changed the way we interpret, predict, and understand different phenomena based on empirical evidence from observation and experimentation.

Before the statistical revolution, there existed a strong philosophical vision in science known as the “clockwork universe,” which indicates the similarity between the universe and a mechanical clock ticking along as a perfect device. In this deterministic approach, precise mathematical formulas were a huge part of science, as many equations and laws were being discovered in popular fields such as physics, chemistry, and biology. These unambiguously defined laws were used to describe many phenomena in the universe, with a specific example of the development of classical mechanics in physics by Sir Isaac Newton, a popular figure in science back then. “Newtonian mechanics” (or “Newtonian physics”) has proven to effectively provide explanations for the behavior of objects in our solar system.

In addition, there was a strong belief that everything that occurs is strictly determined by the circumstances that happened in the past, or in other words, the initial conditions of the universe can be used to predict future events. Not to mention, according to scientific paradigm at the time, people were almost certain that the more refined the measurements, the better the physical reality being examined. However, given the vastly imprecise nature of measuring, Pierre-Simon Laplace came up with the “error function”, which stated that errors associated with scientific observations would follow a distribution. This seemingly simple idea essentially laid the foundation for the statistical revolution. While it makes sense to think that high measuring precision would shrink down the error component, in reality, more errors would arise in relation to more precise

measurements. Consequently, the deterministic view of the world was starting to be seriously in doubt and challenged.

As the nineteenth century progressed, the old scientific paradigm continued to lose its status in science. At the same time, a new statistical paradigm began to emerge and soon replaced the old and outdated approach. In this new statistical way of thinking, the parameters of a distribution do not always have a physical reality and can be estimated with error, regardless of how high the level of precision a measurement might have. A forerunner of the statistical revolution was Karl Pearson, who proposed a probabilistic approach that said scientific data, not the associated error, are probabilistic in nature and have inherent randomness. As a result, the bygone scientific discoveries of Newton and others had now become just rough approximations. For example, the gravitational constant were no longer a fixed empirical physical constant representing the force between two objects caused by gravity as it used to be. Instead, its value varied between measuring trials, and what scientists now were interested in looking at is the scatter of the values' distribution, which would then allow them to analyze and understand the data.

From then, statistical models went on to become a powerful tool and dominate science in the next century. Many well-known and important concepts in statistics nowadays were being discovered, including rules for experimental design, regression to the mean, correlation coefficient, significant testing, confidence interval, or exploratory data analysis. By the end the previous millennium, almost all of science had shifted to using statistical models. We can certainly expect this trend to continue to develop in the field of statistics, as advanced technology and modern methods such as machine learning and computational algorithms will undoubtedly revolutionize statistics even more and take it a whole new level in near future.