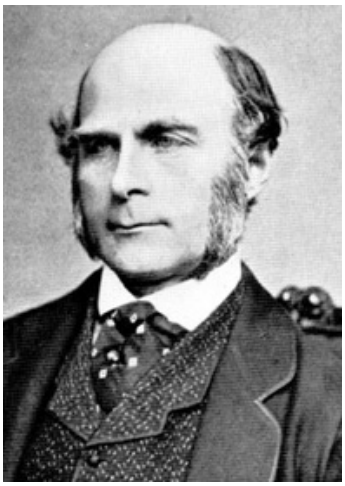
The incredible story of statistics

By Jean-Marie Gogue

In the 20th century, seven mathematicians unknown to the general public led to profound transformations : they invented modern statistics. Not the one we usually think of, that of economic studies and opinion polls. No. They invented methods of statistical analysis, algorithms, as they say today, which accelerated scientific research in all fields.

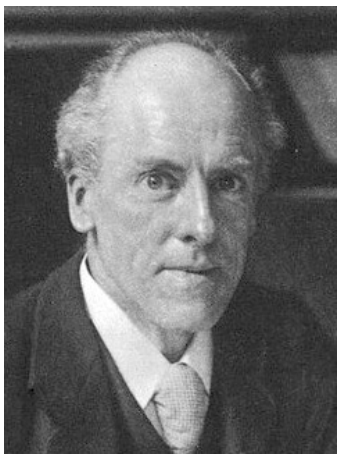
Francis Galton (1822-1911)



Francis Galton was born in England, in the Midlands. His father was a banker ; he was a cousin of Charles Darwin. He began studying medicine in Birmingham and then in London, but he was not interested in medicine. In 1840 he entered *Trinity College* in Cambridge to attend mathematics lessons, but he left the university because teaching did not suit him. He went back to medical school in London. At the death of his father in 1844, he inherited a great fortune that will allow him to satisfy his passion : travelling. After a first trip to Egypt, along the Nile, he led in 1852 a great expedition in southern Africa.

In 1859, the publication of Darwin's book on natural selection gave him a new subject of interest. He gave up his ideas of travelling and devoted most of his activities to studying the variations of the characters in the human populations, the color of the eyes for example. This requires finding methods to measure, collect and analyze data. This is how he invented regression and correlation. His work was pursued by other statisticians.

Karl Pearson (1857-1936)



Karl Pearson was born in Yorkshire. His parents enrolled him at *University College* London, but he left it at sixteen because of poor health. Helped by a preceptor, he passed the Cambridge entrance examination in 1875 and entered *King's College*. He studied mathematics. In 1879, when he graduated from university, he obtained a scholarship that allowed him to travel. He went to Germany to study German literature, then returned to England to study law. In 1884 he was appointed professor of mathematics at *University College* London. He has written numerous articles in mathematics, physics and history of science.

In 1891 he was appointed to the chair of geometry at *Gresham College*, London, where he met Galton. He has just published a book : "The Natural Heritage". Having become his friend, he creates statistical methods that he will apply to the study of natural selection. In 1900, he founded with Galton the journal *Biometrika* in which he published many articles. He invented the "chi-square test" he presents in an article the same year. This test makes it possible to know if chance is the only cause of the variations observed during an experiment.

Karl Pearson has a son, Egon, also a prominent statistician. He succeeded him in 1933 as head of the Department of Applied Statistics at *University College* London.

Jerzy Neyman (1894-1981)

Jerzy Neyman was born in 1894 in a Polish family in eastern Romania. His father died when he was twelve. As a widow, his mother moved to Kharkov, where he attended college and university. He spoke Polish, Russian, German and French fluently. In 1912 he decided to study mathematics. He obtained a postgraduate degree in Kharkov in 1917. The following year he was appointed professor at the Kharkov Institute of Technology.

His family moved to Poland in 1921 because of the Russo-Polish war. He obtained a doctorate in mathematics, before leaving for London in 1924 to study under the authority of Karl Pearson. He got in touch with his son, Egon.

In 1926, he obtained a Rockefeller scholarship to study in Paris. He attended lectures by Borel, Hadamard and Lebesgue at the *College de France*. In 1927, he returned to Poland and became a laboratory head at an institute in Warsaw.

In 1934, Karl Pearson retired. His department at *University College* was cut in half ; a part was led by his son, Egon, who invited Neyman to work with him on the theory of hypothesis testing. They were the first to recognize the need to make alternative hypotheses. They defined the two types of errors known today by most researchers : on the one hand, reject a true hypothesis ; on the other hand accept a false assumption. A considerable progress.

Also in 1934, Neyman developed the theory of sampling in surveys. This theory met the US demand for opinion polls, a demand that sometimes turned into an obsession. He was invited to give lectures in the United States.

In 1937, he joined the University of Berkeley to create a statistics department. He lectured at the *Graduate School* of the Department of Agriculture in Washington. This was the first time American statisticians had the opportunity to listen to him.

In 1948, he launched a symposium on mathematical statistics. This event, which was repeated every five years until 1972, strongly influenced American statisticians.

William Gosset (1876-1937)

William Gosset was born in Canterbury. He attended *Winchester College*, then *Oxford University* where he studied mathematics and natural sciences. A chemistry graduate in 1899, he was hired by the *Guinness brewery* in Dublin, Ireland, as a technician. He was immediately interested in the validity of the results obtained in the laboratory.

He took a break in 1906 to study statistics under Karl Pearson at *University College* London. This one helped him to write articles, but did not understand their importance. Gosset questioned the relevance of a statistical calculation for small samples.

To prevent employees from disclosing confidential information, Guinness had prohibited them from publishing documents, regardless of their content. However, after explaining to management that his mathematical and philosophical conclusions could not be used by

competing breweries, Gosset was authorized to publish them, but under a pseudonym. So Gosset's articles have been published in *Biometrika* under the pseudonym *Student*.

Guinness attached great importance to scientific research to improve its products. How to increase production while keeping the flavor of beer ? Gosset had to answer this question ; his team aimed to improve the selection process of hops.

Gosset sought to understand in what limits a small sample is less representative than a larger sample. This is how he developed the "Student distribution" which is the basis of modern statistics.

A young mathematician named Ronald Fisher was excited about Student's articles. He was particularly seduced by the idea that the Student's distribution could be used to know the probability of a result. We will see that Fisher's interest in Gosset's work has influenced all statistical science.

However Gosset remains little known to the public. It's probably because he worked in a company rather than in academia. If he had been a university professor, he could have used his own name in his articles and published a book to expose his methods and ideas. But he would never have had the opportunity to address the practical problems that stimulated his imagination.

Ronald Fisher (1890-1962)

Ronald Fisher was born in London where his father was head of an auctioneer firm. Very young he was attracted by physics. In 1912 he graduated from *Cambridge University* with a degree in astronomy and taught physics in private schools. In 1919 he was hired at the *Rothamsted Agricultural Research Center*. He stayed there until 1933.

Fisher came into contact with Gosset in 1912. Ten years later he invited him at Rothamsted to discuss mathematical statistics. It is during this period that he developed the concept of analysis of variance.

Analysis of variance consists of simultaneously comparing the measurement results of several samples subjected to different treatments. This is a big step forward from the two-to-two comparison method.

In 1925, Fisher published the book "Statistical Methods for Research Workers", which was reissued several times. His fame was growing ; he began to travel and give lectures. In 1931, he spent six weeks in the United States at the Statistics Center at *Iowa State College*, where he gave three lectures a week and met with many American statisticians, including George Snedecor. He returned there in 1936.

In 1933, Karl Pearson retired. Part of the department of applied statistics was entrusted to Egon Pearson, the other one to Ronald Fisher. But this part disappeared in 1940 and it was only three years later that he could find a job : a post of professor of genetics in Cambridge, where he remained until 1957.

Worldwide mathematicians recognize sir Ronald Fisher as a genius, as the main pillar of mathematical statistics. His revolutionary concepts have allowed research laboratories in all disciplines to make considerable progress.

George Snedecor (1881-1974)

George Snedecor was born in Memphis, Tennessee. He grew up in Florida and Alabama, where his father had moved with wife and children to fulfill a religious vocation.

Snedecor graduated in mathematics and physics at the University of Alabama in 1905, followed by a master's degree in physics at the *University of Michigan* in 1913. In the same year, he was appointed assistant professor of mathematics at *Iowa State College* in Ames. He was promoted to associate professor in 1914 and finally professor in 1931. He had a relationship with Fisher for many years.

As soon as he arrived in the state of Iowa, Snedecor became interested in statistics, especially methods for agricultural research. He collaborated with Henry Wallace, a former student at the university who had given a series of lectures on statistics in 1924. Henry Wallace will later be Vice President of the United States, between 1941 and 1945, under the presidency of Roosevelt. His lectures provoked great interest in statistics throughout the state of Iowa.

A growing demand for advice in statistical methods, especially from farmers, prompted the State of Iowa to set up a statistical section in 1927 in the department of mathematics of the university and to entrust it to Snedecor. In 1933, he created a statistical laboratory. Under his leadership, this laboratory became one of the largest statistical centers in the United States. He gave up his administrative duties in 1947, but continued to teach and do full-time research until 1952.

In 1958, Snedecor was hired as a consultant by the *Navy Electronics Lab* in San Diego, California.

During his career, Snedecor has written numerous articles and several statistical books that are less difficult to read than Fisher's books. He published in particular in 1937 the book "Statistical Methods" which was reissued six times until 1967 and translated into several languages, including French.

Walter Shewhart (1891-1967)

Walter Shewhart was born in New Canton, Illinois. After graduating from the University of Illinois, he graduated in 1917 with a PhD in physics at *Berkeley*, University of California. He was first appointed physics assistant at the University of Illinois and then head of the physics department at a Wisconsin college.

In 1918, he was hired by the *Western Electric company*, which gave him the mission to improve the production of a large telephone factory near Chicago. In 1925, he was transferred to *Bell Telephone Laboratories* in New York, the headquarters, where he remained until his retirement.

In 1932, Shewhart made a trip to England where he met Egon Pearson and William Gosset, which made him known by the scientific community. The previous year, he had published a big book about his experience at *Bell Telephone Laboratories*, but he had not caught the attention of the American public. The reception of Pearson in London was a consecration. In

1935, he founded the Institute of Mathematical Statistics, which he presided. From 1935 to 1944, the Ministry of Defense gave him a consulting mission. In 1943, he founded the collection of mathematical statistics at the publisher *John Wiley and Sons*. He directed it until 1965.

In 1938, Deming invited Shewhart to lecture at the *Graduate School of the Department of Agriculture*. The text of these conferences was published one year later in the book "Statistical Method from the Viewpoint of Quality Control" which was the base of the Deming's teaching in Japan after the Second World War. The Deming's teaching had a great part in the amazing success of Japanese industry.

In the Western Electric plant, Shewhart had significant resources that allowed him to experiment with real data. He understood that all quality issues were caused by variations in product characteristics. As a result, the company could not improve the quality without knowing the variations of the characteristics and without looking for the causes. It was a new concept.

At the base of this concept is the notion of statistical control of a process. Shewhart has shown that when studying a process, it must first be determined whether it is under control or out of control. Thus he elaborated the theory of variations and invented a statistical tool: the control chart.

To build the theory of variations, Shewhart used the concept of "system of random causes" that a French mathematician, Pierre-Simon de Laplace, had already mentioned in 1815 in his "Analytical Theory of Probabilities". The control chart is a tool that identifies among random causes the presence of significant causes.

Shewhart was a modest and selfless man. He conceived the role of the statistician in a firm as that of coordinator of several disciplines involved in a collective learning process.

Conclusion

Here are the seven mathematicians : two Americans, four Britons and a Pole, who invented the modern statistical methods, methods used today in many laboratories all over the world.

Several of them knew each other ; they worked together to create new theories and methods. Those who were employed in companies : Gosset and Shewhart, exchanged regularly with their university colleagues.

Thanks to them, science and industry have made tremendous progress in all areas, progress that no one could have imagined at the beginning of the twentieth century.

August 2017