

Karl PEARSON¹

b. 27 March 1857 - d 27 April 1936

Summary. Karl Pearson was Founder of the Biometric School. He made prolific contributions to statistics, eugenics and to the scientific method. Stimulated by the applications of W.F.R. Weldon and F. Galton he laid the foundations of much of modern mathematical statistics.

Founder of biometrics, Karl Pearson was one of the principal architects of the modern theory of mathematical statistics. He was a polymath whose interests ranged from astronomy, mechanics, meteorology and physics to the biological sciences in particular (including anthropology, eugenics, evolutionary biology, heredity and medicine). In addition to these scientific pursuits, he undertook the study of German folklore and literature, the history of the Reformation and German humanists (especially Martin Luther). Pearson's writings were prodigious: he published more than 650 papers in his lifetime, of which 400 are statistical. Over a period of 28 years, he founded and edited 6 journals and was a co-founder (along with Weldon and Galton) of the journal *Biometrika*. University College London houses the main set of Pearson's collected papers which consist of 235 boxes containing family papers, scientific manuscripts and 16,000 letters.

Largely owing to his interests in evolutionary biology, Pearson created, almost single-handedly, the modern theory of statistics in his Biometric School at University College London from 1892 to 1903 (which was practised in the Drapers' Biometric Laboratory from 1903-1933). These developments were underpinned by Charles Darwin's ideas of biological variation and 'statistical' populations of species - arising from the impetus of statistical and experimental work of his colleague and closest friend, the Darwinian zoologist, W.F.R. Weldon (q.v.). Additional developments emerged from Francis Galton's (q.v.) law of ancestral heredity. Pearson also devised a separate methodology for problems of eugenics in the Galton Eugenics Laboratory from 1907-1933.

In his creation of biometrics, out of which the discipline of mathematical statistics had developed by the end of the nineteenth century, Pearson introduced a new vernacular for statistics (including such terms as the stan-

¹Abridged version of an article in *Encyclopedia of Biostatistics*, Eds. P. Armitage and T. Colton, and published by kind permission of John Wiley & Sons Ltd.

dard deviation, mode, homoscedasticity, heteroscedasticity, kurtosis and the product-moment correlation coefficient).

Family and Education

Karl was the second of three children born to William Pearson and Fanny Smith. His father was a barrister and QC. The Pearsons were of Yorkshire descent. They were a family of dissenters and of Quaker stock. By the time he was in his 20s, Pearson had rejected Christianity and had become a Freethinker which involved the 'rejection of all myths as explanation and the frank acceptance of all ascertained truths to the relation of the finite to the infinite'. Politically, he was a socialist whose outlook was similar to the Fabians, but he never joined the Fabian Society. Socialism was a form of morality for Pearson; the moral was social and the immoral was anti-social in conduct.

Pearson's father William was a very hard-working and taciturn man. In a letter to Karl, his elder brother Arthur described the experience of being home with their father as 'simply purgatory...the governor never spoke a word'.

When they went up to Cambridge, at least one of the Pearson boys was expected to read mathematics. The Cambridge Mathematics Tripos was, at that time, the most prestigious degree in any British university. Although his father urged him to read mathematics, Arthur settled on Classics. Thus when Karl was 15 years old, his father was looking for a good Cambridge Wrangler to prepare him for the Mathematics Tripos.

By the Spring of 1875, Pearson was ready to take the entrance examinations at various colleges at Cambridge. His first choice was Trinity College, where he failed the entrance exam; his second choice was King's College from whom he received an Open Fellowship in April 1875. Pearson found that the highly competitive and demanding system leading up to the Mathematical Tripos was the tonic he needed. Though he had been a rather delicate and sickly child with a nervous disposition, he came to life in this environment and his health improved. Students of the Mathematics Tripos were also expected to take regular exercise as a means of preserving a robust constitution and regulating the working day.

Pearson took the Mathematics Tripos examination in January 1879. He graduated with honours being the Third Wrangler; subsequently, he received a fellowship from King's College which he held for seven years.

A couple of weeks after Pearson had taken his degree, he began to work in

Professor James Stuart's Engineering workshop and read philosophy during the Lent Term in preparation for a trip to Germany.

Germany and University College London

Pearson's time in Germany was a period of self-discovery, philosophically and professionally. While in Heidelberg Pearson read Berkeley, Fichte, Locke, Kant and Spinoza, but he subsequently abandoned philosophy. He studied physics under Quincke and metaphysics under Kuno Fischer. He considered becoming a mathematical physicist, but decided not to pursue this. He went to Berlin to hear Kirchoff and Helmholtz and began to study Roman Law. A year later, he took up rooms at the Inner Temple and read law at Lincoln's Inn. He was called to the Bar at the end of 1881 and practised the law for a very short time only. Still searching for some direction when he returned to London, Pearson lectured on socialism, Marx and Lassalle at the working men's clubs and on Martin Luther at Hampstead from 1880 to 1881.

By 1882 Pearson had decided that he did not want to pursue the law. From 1882 to 1884, he lectured on German society from the medieval period up to the sixteenth century. He became so competent in German that by the late spring of 1884, he was offered a post in German at Cambridge.

Nevertheless, Pearson found all these pursuits dissatisfying and he then began to write some papers on the theory of elastic solids and fluids as well as some mathematical physics papers on optics and ether squirts. Between 1879 and 1884 he applied for more than six mathematical posts and he received the Chair of 'Mechanism and Applied Mathematics' at University College London (UCL) in June of 1884.

During Pearson's first six years at UCL, he taught mathematical physics, hydrodynamics, magnetism, electricity and his speciality, elasticity, to engineering students.

The Gresham Lectures on Geometry and Curve-Fitting

Pearson was a founding member of the Men's and Women's Club established in 1885 'for the free and unreserved discussion of all matters in any way connected with the mutual position and relation of men and women'. Among the various members was Marie Sharpe whom he married in June 1890. They had three children, Sigrid, Helga and Egon. Six months after his marriage, he took up another teaching post in the Gresham Chair of Geometry which he held for three years concurrently with his post at UCL. As Gresham Professor, he was responsible for giving 12 lectures a year. These

were free to the public. Between February 1891 to November 1893, Pearson delivered 38 lectures. His first eight lectures formed the basis of his book, *The Grammar of Science* which was published in several languages.

Pearson's earliest teaching of statistics can, in fact, be found in his lecture of 18 November 1891 when he discussed graphical statistics and the mathematical theory of probability with a particular interest in actuarial methods. Two days later he introduced the histogram - a term he coined to designate a 'time-diagram' to be used for historical purposes. He introduced the standard deviation in his Gresham lecture of 31 January 1893. Pearson's early Gresham lectures on statistics were influenced by the work of Edgeworth (q.v.), Jevons (q.v.) and Venn (q.v.). Up until November 1893, these lectures covered fairly conventional statistical and probability methods. Whilst the material in these lectures was not original in content, Pearson's approach in teaching was highly innovative. In one of his lectures, he scattered 10,000 pennies over the lecture room floor and asked his students to count the number of heads or tails.

Pearson's last twelve Gresham Lectures signified a turning-point in his career owing, in particular, to his relationship with Weldon - who was the first biologist Pearson met who was interested in using a statistical approach for problems of Darwinian evolution. Their emphasis on Darwinian population of species not only implied the necessity of systematically measuring variation, but it prompted the re-conceptualisation of statistical populations. Moreover, it was this mathematisation of Darwin which led to a paradigmatic shift for Pearson from the Aristotelian essentialism underpinning the earlier use and development of social and vital statistics. Weldon's questions not only provided the impetus for Pearson's seminal statistical work, but this led eventually to the creation of the Biometric School at UCL.

In Pearson's first published statistical paper of 26 October 1893, he introduced the method of moments as a means of curve fitting asymmetrical distributions. One of his aims in developing the method of moments was to provide a general method for determining the values of the parameters of a frequency distribution. In 1895 Pearson developed a general formula to use for subsets of six types of frequency curves. In his first supplement in 1901, he defined two further types and a final two were added in his second supplement in 1916. Many of his curves were J-shaped, U-shaped and skewed. Pearson derived all of his curves from a differential equation whose parameters were found from the moments of the distribution. As Churchill Eisenhart remarked in 1974, 'Pearson's family of curves did much to dispel the almost

religious acceptance of the normal distribution as the mathematical model of variation of biological, physical and social phenomena'.

The Biometric School

Following the success of his Gresham lectures, Pearson began to teach statistics to students at UCL in October of 1894. By 1895 he worked out the mathematical properties of the product-moment correlation coefficient (which measures the relationship between two continuous variables) and simple regression (used for the linear prediction between two continuous variables). By then, Francis Galton had determined graphically the idea of correlation and regression for the normal distribution only.

In this seminal paper on 'Regression, Heredity and Panmixia' in 1896, Pearson introduced matrix algebra into statistical theory. In the same paper, Pearson also introduced the following statistical methods: eta (η) as a measure for a curvilinear relationship, the standard error of an estimate, multiple regression and multiple and partial correlation, and he also devised the coefficient of variation as a measure of the ratio of a standard deviation to the corresponding mean expressed as a percentage.

Pearson introduced various methods of correlation. By the end of the nineteenth century he began to consider the relationship between two discrete variables. In 1900, he devised the tetrachoric correlation and the phi-coefficient for dichotomous variables. The tetrachoric correlation requires that both X and Y represent continuous, normally distributed and linearly related variables whereas the phi-coefficient was designed for classes having qualitative attributes. Nine years later, he devised the biserial correlation when one variable is continuous and the other is discontinuous. With his son Egon, he devised the polychoric correlation in 1922 (which is very similar to canonical correlation today). Though not all of Pearson's correlational methods have survived him, a number of these methods are still the principal tools used by psychometrists for test construction. Following the publication of his first three statistical papers in Philosophical Transactions of the Royal Society, Pearson was elected a Fellow of the Royal Society in 1896. He was awarded the Darwin Medal from the Royal Society in 1898.

Pearson's chi-square tests

At the turn of the century, Pearson reached a fundamental breakthrough in his development of a modern theory of statistics when he found the exact chi-square distribution from the family of Gamma distributions and de-

vised the chi-square (χ^2, P) goodness of fit test. The test was constructed to compare observed frequencies in an empirical distribution with expected frequencies in a theoretical distribution to determine ‘whether a reasonable graduation had been achieved’ (i.e., one with an acceptable probability).

Four years later, he extended this to the analysis of multiple contingency tables and introduced the ‘mean square contingency coefficient’ which he also termed the chi-square test of independence (which R.A. Fisher termed the chi-square statistic in 1923). Pearson’s conception of contingency led at once to the generalisation of the notion of the association of two attributes developed by his former student, G. Udny Yule (q.v.). Individuals could now be classed into more than two alternate groups or into many groups with exclusive attributes. The contingency coefficient and the chi-square test of independence could then be used to determine the extent to which two such systems agreed.

Pearson’s four laboratories

Pearson set up the Drapers’ Biometric Laboratory in 1903 following a grant from the Worshipful Drapers’ Company (who funded Pearson annually for work in this laboratory until his retirement in 1933). The methodology incorporated in the Drapers’ Biometric Laboratory was twofold: the first was mathematical, and included the use of Pearson’s statistical methods, matrix algebra and analytical solid geometry. The second involved the use of such instruments as integrators, analysers, curve-plotters, the cranial co-ordinatograph, silhouettes and cameras. The problems investigated by the biometrists included natural selection, Mendelian genetics and Galton’s law of ancestral inheritance, craniometry, physical anthropology and theoretical aspects of mathematical statistics. By 1915, Pearson established the first degree course in mathematical statistics in Britain.

Though Pearson did not accept the generality of Mendelism, he did not reject it completely as is commonly believed. When William Bateson published his fiercely polemical attack on Weldon in 1902, Bateson saw Mendelism as a tool for discontinuous variation only. As a biometrician, most of the variables that Pearson and his co-workers analysed were continuous and only occasionally did they examine discontinuous variables. Whilst Pearson and Weldon used Galton’s law of ancestral inheritance for continuous variables, they used Mendelism for discontinuous variables. Indeed, Pearson argued that his chi-square test of independence was the most appropriate statistical tool for the analysis of Mendel’s discrete data for dominant and recessive

alleles (such as colour of eyes where brown is dominant and blue is recessive). Even today, Pearson's chi-square tests remain the most widely used technique for analysing Mendelian data.

A year after Pearson had established the Biometric Laboratory, the Drapers' Company gave him a grant so that he could establish an Astronomical Laboratory. Pearson was interested in determining the correlations of stellar rotations, and the variability in stellar parallax. He was also instrumental in setting up a degree course in astronomy in 1914 at UCL.

In 1907, Francis Galton (who was then 85 years old) wanted to step down as Director from the Eugenics Record Office which he had set up three years earlier, and he asked Pearson if he would take it on. Pearson agreed reluctantly. He renamed the office the Galton Eugenics Laboratory when he became its director. Pearson made very little use of his biometric methods in this Laboratory; instead he developed a completely different methodology for problems relating to eugenics. This methodology was underpinned by the use of actuarial death rates and by a very highly specialised use of family pedigrees assembled in an attempt to discover the inheritance of various diseases (which included, for example, alcoholism, cancer, diabetes, epilepsy, paralysis and pulmonary tuberculosis). These family pedigrees became the vehicle through which Pearson could communicate statistical ideas to the medical community by stressing the importance of using quantitative methods for medical research. This tool enabled doctors to move away from concentrating on individual pathological cases or 'types' and to see, instead, a wide range of pathological variation of the disease (or condition) of the doctors' speciality.

In the spring of 1909, Galton was discussing the future of the Eugenics Laboratory with Pearson. Whilst Galton thought that Pearson would have been 'the most suitable man for the first Galton Professor', Pearson let Galton know that he was 'wholly unwilling to give up superintendence of the Biometric Laboratory [he] had founded and confine [his] work to Eugenics Research'. A month later, Galton added a codicil to his will stating that he desired that the first Professor of the post should be offered to Pearson on such condition that Pearson could continue to run his Biometric Laboratory. After Galton's death in 1911 Pearson relinquished the Goldsmid Chair of Applied Mathematics after 27 years of tenure to take up the Galton Chair. The Drapers' Biometric and the Galton Eugenics laboratories, which continued to receive separate funding, then became incorporated into the Department of Applied Statistics.

Pearson then proceeded to raise funding for a new building for his Department of Applied Statistics. In the early summer of 1914, the new laboratory was complete and preparations were underway for the occupation and fitting up of the public museum and the Anthropometric Laboratory. It was hoped that the building would be occupied by October 1915. These developments and further biometric work were shattered by the onset of the First War. The new laboratory building was taken over by the government to be used as a military hospital. Pearson and his co-workers took on special war duties. They produced statistical charts for the Board of Trade's Labour Department as well as for its Census Production. Pearson was also involved with elaborate calculations of anti-air craft guns and bomb trajectories. It was not until December 1922 that Pearson's building was reoccupied.

His wife, Marie Sharpe, died in 1928 and in 1929 he married Margaret Victoria Child, a co-worker in the Biometric Laboratory. Pearson was made Emeritus Professor in 1933. From his retirement until his death in 1936, he published 34 articles and notes and continued to edit *Biometrika*. Pearson was offered an OBE in 1920 and a knighthood in 1933, but he refused both honours. He also declined the Royal Statistical Society Guy Medal in their centenary year in 1934. Pearson believed that 'all medals and honours should be given to young men, they encourage them when they begin to doubt whether their work was of value'.

Bibliography

- Eisenhart, Churchill. (1974). Karl Pearson. *Dictionary of Scientific Biography*, 10, Charles Scribner's Sons, New York, pp. 447-73.
- Hilts, Victor. (1981). Statist and Statistician. Arno Press, New York. Reprint of his PhD thesis, Harvard University, 1967.
- Mackenzie, Donald. (1981). *Statistics in Britain 1865-1930: The Social Construction of Scientific Knowledge*. Edinburgh University Press, Edinburgh.
- Magnello, M. Eileen. (1993). Karl Pearson: Evolutionary Biology and the Emergence of a Modern Theory of Statistics. DPhil thesis, University of Oxford.
- Magnello, M. Eileen. (1996). Karl Pearson's Gresham Lectures: W.F.R. Weldon, Speciation and the origins of Pearsonian Statistics. *British Journal for the History of Science*, **29**, 43-64.
- Magnello, M. Eileen. (forthcoming). Karl Pearson's mathematisation of inheritance. From Galton's ancestral heredity to Mendelian Genetics (1895-1909). *Annals of Science*.

- Norton, Bernard. (1978). Karl Pearson and Statistics: The Social Origin of Scientific Innovation. *Social Studies of Science*, **8**, 3-34.
- Pearson, Egon. (1936-1938). Karl Pearson: An Appreciation of Some Aspects of his Life and Work. Part 1, 1857-1905, *Biometrika*, (1936), 193-257; Part 2.1906-1936, (1938), 161-248. (Reprinted by Cambridge University Press: 1938).
- Pearson, Karl, (1914-1930). *The Life, Letters and Labours of Francis Galton*. 3 vol. in 4 parts, Cambridge University Press, Cambridge.
- Porter, Theodore M. (1986). *The Rise of Statistical Thinking. 1820-1900*. Princeton Univ. Press, Princeton.
- Semmel, Barnard. (1958). Karl Pearson: Socialist and Darwinist. *British Journal of Sociology*, **9**, 111-125.
- Stigler, Steven M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press of Harvard University Press, Cambridge, MA.

Eileen Magnello