Francis GALTON
b. 16 February 1822 – d. 17 January 1911

Summary. Francis Galton developed the basic statistical concepts of regression lines and correlation between variables. He helped in founding the journal *Biometrika* and left most of his fortune to establish the Chair of Eugenics subsequently occupied by K. Pearson and R.A. Fisher.

Francis Galton was born in Birmingham (England) and died in Haslemere (England). His grandfathers Erasmus Darwin and Samuel Galton were both members of an informal society of scientists, inventors, and industrialists that is known as the Lunar Society of Birmingham as its meetings were held on nights of the full moon. Other members included Matthew Boulton, Richard Lovell Edgeworth, Joseph Priestley, James Watt, and Josiah Wedgwood.

The children and grandchildren of members of this society tended to marry one another. In particular, Samuel Tertius Galton, the second son of Samuel Galton married Frances Anne Violetta Darwin, the eldest daughter of Erasmus Darwin by his second marriage. Thus Charles Darwin and Francis Galton were first cousins. Francis Galton was also related by marriage to Francis Ysidro Edgeworth (q.v.), a grandson of Richard Lovell Edgeworth, as Galton's wife's brother, Arthur Gray Butler, married Edgeworth's first cousin, Harriet Jessie Edgeworth. Another brother, Henry Montagu Butler was Master of Trinity College, Cambridge, when Galton was elected an Honorary Fellow of the College in 1902. For a detailed account of the ramifications of this type of social grouping (sometimes known as an 'intellectual aristocracy') in the subject areas of statistics and eugenics, see McKenzie (1981).

Galton's father and grandfather were both prosperous gun manufacturers and bankers (despite their also being members of the Society of Friends or Quakers). Nevertheless Galton intended to follow Erasmus Darwin and his Darwin uncles into a medical career. He began his medical training at the Birmingham General Hospital in 1838 and continued it at King's College, London, in 1839, but broke his training off to study mathematics at Trinity College, Cambridge, in 1840. However, poor health obliged him to abandon the honours course and take a 'poll' or pass degree in 1844.

Galton's father died in 1844 leaving his third son (and youngest child) a substantial income. He abandoned any thought of a medical career and gave himself up to amusements for a period of five years, but this style of life was

not able to satisfy his ambition.

In his student days Galton had travelled widely in the territory of the Ottoman Empire. On one trip he had visited Vienna, Constantinople (then the capital of the Empire), and Smyrna, and on a second he had travelled in Egypt and Syria.

In 1850 Galton mounted an expedition at his own expense, but under the aegis of the Royal Geographical Society, to the little-known Territory of Damaraland in South-West Africa (now Namibia). He Landed at Walfisch or Walvis Bay and penetrated into the interior. On his return to England in 1852, he wrote an account of his expedition entitled *Tropical South Africa* (1853) and a practical guide to *The Art of Travel* (1853) which is still in print.

Galton was elected a Fellow of the Royal Society of London in 1856 and frequently served on the council of the Society. He was active in the British Association for the Advancement of Science and Served as General Secretary of the Association 1863-67. In 1863 he published a book entitled *Meteorographica, or Methods of Mapping the Weather*, which led to his serving as a member of the Governing Council of the British Meterological Office 1868-1900. The term 'anticyclone' was introduced by him.

In his *Hereditary Genius* (1869) he developed the argument that the descendents of eminent persons are more likely to be 'eminent' than others, though to a lesser degree than their immediate ancestors. A difficulty in defining such abstract concepts as 'eminence' obliged him to turn his attention to physical measurements, and in 1884 Galton set up an Anthropometric Laboratory in connection with the International Health Exhibition of 1884-85 in which he took physical measurements of individuals who presented themselves. Some of the results of these experiments are presented in his *Natural Inheritance* (1889).

With his social background, his interests in heredity, and in the context of the ideas on natural selection discussed by Darwin and Huxley, it is not surprising to find that Galton believed in the superiority of certain human types. He recommended breeding from the 'best' social types and restricting the offspring of the 'worst'. He coined the term 'eugenics' (though not the name Eugene). He devoted much of his income in later life to the promotion of this subject and, after his death, he bequeathed most of his estate to University College, London, to found a Chair of Eugenics. Karl Pearson (q.v.) and Ronald Fisher (q.v.) were the first two holders of this chair. Pearson (1914-30) naturally included extensive genealogical tables of the Galton

family in his three volume biography of Francis Galton.

Galton's statistical work began well before 1885, but, because it contents are relatively familiar, it is convenient to begin our account of his work in this area with his 1885 paper published in the *Philosophical Transactions of the Royal Society.* In this paper he compiled a table of the heights of 928 adult individuals and the weighted average of their parents' heights. After a little smoothing, Galton found that the contours of equal frequency for this table clearly took the form of a system of concentric ellipses. Further, when a system of vertical lines were inserted in this diagram, he found that they each identified a single point of maximum frequency, and that these points of maximum frequency lay on a single straight line which has since come to be known as a 'regression line'.

These familiar results from a bivariate distribution with elliptical contours may be regarded as a natural development of Galton's earlier work on a system of univariate distributions. In 1877 he had established that the frequency distribution of the seeds produced by plants grown from sweet pea seeds of a given size had a symmetric distribution about a particular value. This value was smaller than the size of the parent seeds if the latter were larger than the average for the population as a whole, and *vice versa*. This is what is meant by the phrase 'regression towards the mean'.

Galton also developed a simple mechanical device for illustrating the generation of the univariate normal distribution as a limiting case of the binomial distribution. Lead shot is poured in a steady stream in such a way as to fall on a particular pin in an array of pins in the form of a *quincunx*, or a lattice of equilateral triangles, on an inclined plane. Once the shot has passed through the quincunx it is collected in a set of rectangular compartments of equal size and the empirical distribution examined. This device is still in existence, see Stigler (1986, 1989).

Galton soon realised that his system of elliptical contours gave rise to two regression lines and that these regression lines had the same slope when the data were expressed in standard units. In 1888 he therefore extended his analysis to define the common slope as a measure of the 'correlation' between the chosen pair of variables. However, Galton does not seem to have discussed the possibility of negative values for the correlation between variables as most of his practical examples were concerned with the relationship between the lengths of different limbs from the same individual.

Galton's ideas on regression and correlation were promptly taken up and given a formal mathematical development by K. Pearson and G.U. Yule

(q.v.). At almost eighty years of age, Galton's attention passed onto other interests. He wrote three books on the use of fingerprints in forensic science and investigated the operation of visual memory.

In his analyses of statistical data Galton employed the median value and the median absolute error (or probable error) where we would use the mean value and the root mean squared error (or standard error). He justified his preference for median values by analogy with a simple majority voting procedure: if there are an odd number of electors and they all have single-humped preference curves, then it is readily established that the value chosen by the committee as a whole is the most preferred point of the middlemost voter. In this context, it is interesting to observe that Galton and Edgeworth were both champions of the median against the mean and that they were both concerned with the points of tangency between two mutually convex systems of contours or indifference curves. Indeed, Edgeworth named the line of tangency a 'contract curve' in his theory of economic exchange of 1881 and he was concerned with the intersection of two conditional median curves in his implementation of the least absolute deviations fitting procedure in 1887, see Farebrother (1987).

Galton received numerous honours during his lifetime: he was awarded honorary doctorates by the Universities of Oxford and Cambridge in 1894 and 1895 respectively, and he was knighted by King Edward VII in June 1909.

References

[1] Farebrother, R.W. (1987). The theory of committee decisions and the double median method, *Computational Statistics and Data Analysis* **5**, 437-442.

[2] Forrest, D.W. (1974). *Francis Galton: The Life and Works of a Victorian Genius*, Paul Elek, London.

[3] McKenzie, D.A. (1981). *Statistics in Britain 1865-1930*, Edinburgh University Press, Edinburgh.

[4] Pearson, K. (1914-30). *The Life, Letters and Labours of Francis Galton*, Three Volumes, Cambridge University Press, Cambridge.

[5] Stigler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Harvard University Press, Cambridge, Massachusetts.

[6] Stigler, S.M. (1989). Francis Galton's account of the invention of correlation, *Statistical Science* **4**, 73-86.

R.W. Farebrother