

KARL PEARSON (March 27, 1857 – April 27, 1936)

by HEINZ KLAUS STRICK, Germany

CARL PEARSON was born the second child of a London lawyer. He was educated at home until the age of nine, after which he attended *University College School*. Due to illness, he had to stop attending school at the age of 16.

A private tutor prepared him for the entrance examination at Cambridge University, which he passed as second best in his year and he received a scholarship to *King's College*.

He was fortunate to be assigned EDWARD JOHN ROUTH (the university's most successful "coach") as tutor to prepare him for the Tripos examination.

In his studies, which were not limited to mathematics and physics, he also dealt with topics from the history of literature and philosophy. PEARSON read DANTE, GOETHE and ROUSSEAU in the original and he wrote reviews of books on MAIMONIDES and SPINOZA.

Although his protest against the obligation for all students to attend theology lectures, which had existed since 1441, led to the abolition of the regulation, he continued to attend them, voluntarily and out of personal interest. With the help of his father's legal support, he also secured the abolition of compulsory attendance at church services, but here too, he did not change his personal behaviour.

In 1879 PEARSON passed the Mathematics Tripos examination as third best in his year. He then travelled to Germany for a year. In Heidelberg and Berlin he attended lectures on German literature of the Middle Ages and the Renaissance as well as on the history of the Reformation. He dealt with the role of women in society and studied the teachings of KARL MARX and FERDINAND LASALLE, and from this time on, he spelt his first name with a "K".

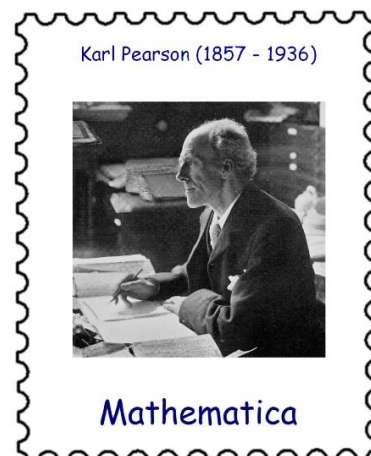
Thanks to a scholarship, he was financially independent during the next years. PEARSON attended law lectures and was admitted to the bar in 1882.

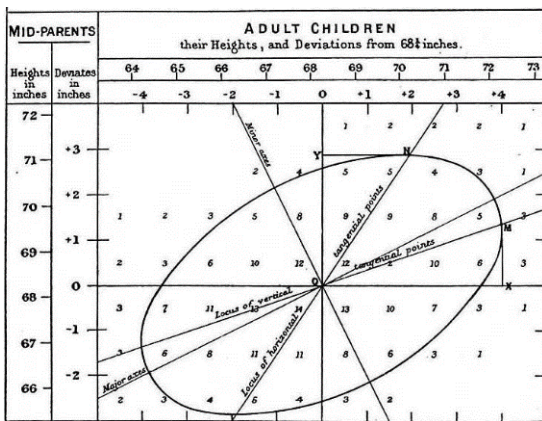
He wrote a book on science and philosophy (*The New Werther*), lectured on MARTIN LUTHER and offered KARL MARX the translation of *Das Kapital* into English.

From 1880 onwards PEARSON deputised for various professors of mathematics at *King's College*, and also at *University College* London (UCL), and there he was appointed *Professor of Applied Mathematics and Mechanics* in 1884. In the course of this activity, he developed a special interest in graphical and statistical methods.

When the historian of mathematics ISAAC TODHUNTER fell ill and died, he took over the final editing of his two-volume work *History of the Theory of Elasticity*. In an extremely productive phase, PEARSON, who had meanwhile become an agnostic, wrote *The Ethic of Free Thought*, as well as *The Grammar of Science*, in which he emphasised the evaluation of quantitative data in sciences such as biology, medicine and the social sciences.

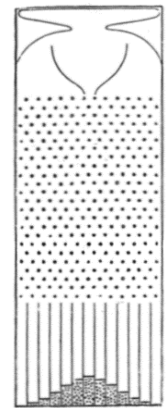
In 1889, the polymath FRANCIS GALTON, a cousin of CHARLES DARWIN, published *Natural Inheritance* – a summary of the research he had conducted on heredity up to that time.





Among other things, GALTON had discovered the phenomenon of *regression to the mean* (he called it *reversion to mediocrity*) when comparing the heights of parents and their adult children.

The work also contained a description of the *GALTON board* named after him, cf. GALTON's illustration on the right.

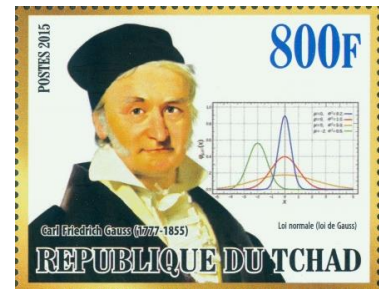


PEARSON was so taken with GALTON's correlation investigations that he spent the following decade and a half almost exclusively developing the mathematical foundations of statistics.

His hope was to arrive at findings *comparable to those in physics*, especially with regard to human behaviour. He was aware of the difference between correlation and causality – he saw the latter as the “limit” of correlation.

PEARSON was supported in his research work by the zoologist WALTER FRANK RAPHAEL WELDON, who asked him concrete, technical questions about statistical analysis – including, for example, how to proceed in the case of asymmetrical or two-tailed non-normally distributed distributions.

PEARSON wrote a wealth of papers which were published as *Mathematical Contributions to the Theory of Evolution*. In this context, he coined the term *standard deviation* – until then, the term *mean error*, originating from CARL FRIEDRICH GAUSS, was used.



Following preliminary work by GALTON and AUGUSTE BRAVAIS, he defined the so-called *PEARSON'S correlation coefficient* as a measure for the linear correlation of the random variables X, Y :

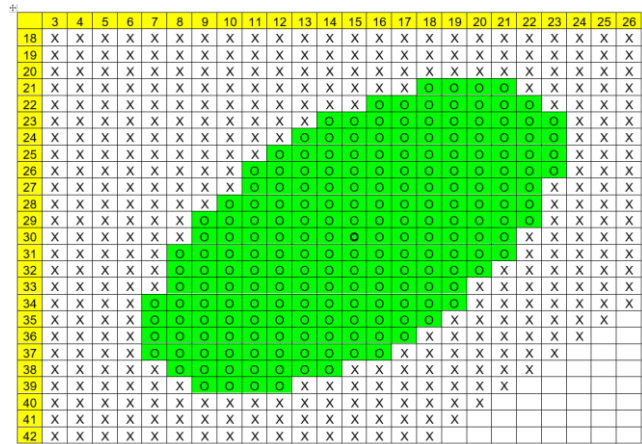
$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}} = \frac{E(X - \mu_X) \cdot E(Y - \mu_Y)}{\sigma_X \cdot \sigma_Y}$$

An estimate for this is

$$r_{X,Y} = \frac{s_{X,Y}}{s_X \cdot s_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}, \text{ the empirical correlation coefficient.}$$

In 1900, PEARSON developed the *chi-squared fit test*, which measures the deviation of observed data μ_i from expected data: $\chi^2 = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\mu_i}$. To this end, he published tables with the associated probabilities for different degrees of freedom. By doing this he laid the foundations for test and decision theory by determining *p*-values for the measured χ^2 values of empirical results.

In the case of a binomial distribution (degree of freedom 1), the random variable χ^2 coincides with the square of the standardised random variable $\frac{X - \mu}{\sigma}$.



The figure on the right shows a 95% ellipse for a trinomially distributed random variable with $n = 60$; $p_1 = 0.25$; $p_2 = 0.5$, i.e. $p_3 = 0.25$, such as is present in intermediate inheritance – the different phenotypes occur in a ratio of 1:2:1.

In his article on the χ^2 test, PEARSON dealt, among other things, with a random experiment by WELDON, who had thrown twelve dice 26306 times: For various groupings of results, he determined values of χ^2 for the deviations of the theoretical probabilities from the recorded frequencies.

PEARSON was the first to systematically use the representation form of a *histogram*. In one of his contributions, the term *random walk* appears for the first time.

In 1896, PEARSON was admitted as a member of the *Royal Society*, and in 1898 and 1900 he received the Society's DARWIN Medal.

Around 1900, the laws of heredity already described by GREGOR MENDEL in 1866 were rediscovered.

Two camps formed in the *Royal Society*, which fought each other fiercely. On the one side – led by WELDON and PEARSON – were the supporters of GALTON's *Law of Ancestral Heredity*, which states that the two parents contribute on average half of the total hereditary material of the offspring, the four grandparents a quarter, etc.

On the other side – led by WILLIAM BATESON, a former student of WELDON, creator of the term *genetics* – were biologists in particular who found it difficult to accept that one can derive statements about biological facts at all with the help of mathematical methods. For them, the rediscovered MENDEL's laws contained the sole truth and everything else was heresy.



The bitterly fought controversy led to the foundation of the journal *Biometrika* by GALTON, PEARSON and WELDON.

In this journal papers were published on variation, heredity and selection in the animal and plant kingdoms, based on the study of a large number of objects, and on the development of statistical methods for the analysis of biological problems and the associated mathematical theory.

The conflict in the *Royal Society* only lost its intensity when WELDON died unexpectedly in 1906. For the next thirty years PEARSON was the sole editor of the journal.

After GALTON's death in 1911, PEARSON took over the *GALTON Chair in National Eugenics* at *University College* (later renamed the *GALTON Chair of Human Genetics*).

The term *eugenics* had been coined by GALTON in 1883 as a science concerned with *all influences which improve the innate characteristics of a race, including those which develop them to the greatest possible advantage*.

He held that the inheritance of positive traits was to be selectively encouraged and the inheritance of negative traits was to be avoided as far as possible.

In contrast to ADOLPHE QUETELET, who in his investigations into measurable characteristics of people and their living conditions spoke – purely *descriptively* – of the *statistical average human* being and described this as *normal*, GALTON deliberately deviated from this by assessing the *average human* being as *mediocre*.



PEARSON also argued that *Nations should be a homogeneous whole, not a mixture of high-value and low-value breeds* – with corresponding legal regulations regarding immigration.

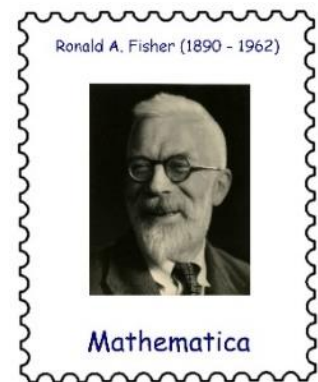
At the time, demands for selective reproduction were approved of by many – in principle, they are still applied today by countries such as Canada and Australia in their immigration policies.

During the Third Reich, the term *racial hygiene* was used in Germany instead of the previously common term *hereditary health theory*; it served the National Socialists as a justification for euthanasia programmes (so-called *destruction of life without value*) and for human experiments in the concentration camps.

PEARSON was respected as head of the *Department of Applied Statistics* at UCL, admired by his staff – but also feared. Disagreements could easily escalate to discriminatory, unforgiving comments.

Around the year 1914, a conflict developed – partly played out in public – with RONALD AYLMER FISHER, who submitted an article for publication in *Biometrika*, which PEARSON initially commented favourably on, but then rejected – perhaps because of a misunderstanding. A dispute developed from this, which led to FISHER turning down the prestigious post of chief statistician at the GALTON laboratory in 1919, as he would then have had to work under PEARSON.

Both were competent advocates of statistical methods, but their approaches were different: PEARSON tried to infer correlations using large samples. FISHER, on the other hand, used small samples to find causes.



PEARSON had been married since 1890 to MARIA SHARPE, the former secretary of the "Men's and Women's Club", which PEARSON had founded in the mid-1880s. The marriage produced three children – a son (EGON) and two daughters (SIGRID and HELGA).

After KARL PEARSON's retirement in 1933, the *GALTON Chair* was divided: PEARSON's son EGON took over the Department of Statistics and RONALD AYLMER FISHER the Department of Eugenics – new conflicts were inevitable ...

First published 2022 by Spektrum der Wissenschaft Verlagsgesellschaft Heidelberg

<https://www.spektrum.de/wissen/karl-pearson-der-vollblut-statistiker/2079426>

Translated 2022 by John O'Connor, University of St Andrews
