

## **Harold HOTELLING**

b. 29 September 1895 - d. 26 December 1973

**Summary.** A major developer of the foundations of statistics and an important contributor to mathematical economics, Hotelling introduced the multivariate  $T^2$ , principal components analysis, and canonical correlations.

Harold Hotelling was born in Fulda, Minnesota, USA. He graduated with a B.A. in Journalism in 1919, and subsequently obtained a Master of Science degree in mathematics in 1921, both from the University of Washington. In 1924, he received his doctorate in mathematics from Princeton University, writing a dissertation on topology. Upon leaving Princeton, he joined the Food Research Institute at Stanford University where he was a research associate from 1924 to 1927, and an associate professor in the Department of Mathematics from 1927 to 1931. It was during this time that Hotelling, already involved in economics and mathematics research, began his work in statistics. He became aware of the writings of R. A. Fisher (q.v.) and began a correspondence with him that led to their long friendship. In fact, Hotelling wrote in 1927 the book review of the first edition of Fisher's *Statistical Methods for Research Workers* for *The Journal of the American Statistical Association*, and also spent a number of months visiting Fisher at the Rothamsted Experimental Station in the second half of 1929. While at Stanford, Hotelling wrote several ground-breaking papers in both statistics and mathematical economics. Hotelling developed his generalization of Student's  $t$ -ratio to deal with multivariate correlated response data (Hotelling, 1931) and this test known as Hotelling's  $T^2$ , remains one of his best known results. Also he worked with Holbrook Working, and together in 1929 they developed results on the standard error of the estimated regression line leading to the well known Working-Hotelling curved simultaneous confidence regions for a regression line. In addition to his statistical research at Stanford, Hotelling worked in mathematical economics. Of the notable papers he wrote, one in 1929 dealt with a problem of optimizing price and location in a competition between two entities in a spatial setting. The other paper in 1931 dealt with exhaustible natural resources, in which he showed that in equilibrium, the prices of such natural resources will have a tendency to rise over time at a percentage rate that equals the national interest rate.

In 1931, Hotelling was recruited to Columbia University to be a professor of economics, and also to undertake the initiation of mathematical statis-

tics at Columbia. On July 1, 1942, Hotelling, W. Allen Wallis and Jacob Wolfowitz became the charter members of the renowned Statistical Research Group (SRG) which was based at Columbia University during World War II and remained in existence until September 30, 1945. (See Wallis (1980) for details concerning the SRG.) The SRG attracted an extraordinary group of research statisticians to Columbia, and its goal was to support the Armed Forces in improving the quality and efficiency of their war efforts. As part of the SRG effort, Hotelling developed methods for control charts for multivariate data (Hotelling, 1947), and Abraham Wald did his fundamental research establishing the field of sequential analysis.

During his career at Columbia University prior to the War, Hotelling continued with his innovative developments in statistics and mathematical economics. He introduced the idea of principal component analysis in his 1933 and 1936 papers as a way of understanding the structure of large numbers of correlated multivariate observations, and he generalized the notions of correlation and multiple correlation to introduce canonical correlation analysis which allows one to measure the strength of relationship between two dependent sets of multivariate observations, and also understand the structure of the relationships between them (Hotelling, 1936). In 1940, he wrote with great foresight a paper on how statistics should be best taught in universities. Hotelling (1940) anticipated the growth of statistics and, at that time, innovatively argued that the subject of statistics should be taught in universities by statisticians in departments of statistics. In mathematical economics, two notable papers of this period were one in 1935 uniting demand and utility and one in 1938 introducing the “welfare equilibrium principle.”

Hotelling was president of the Institute of Mathematical Statistics in 1941, having been one of the three original founding leaders of the Institute in 1935. He also served as president of the Econometrics Society in 1936-37, and was actively involved with the Cowles Commission essentially from its beginning.

One year after the end of the World War II, Hotelling was recruited in 1946 to the University of North Carolina at Chapel Hill to start upon his arrival a Department of Mathematical Statistics. This department was to be a complement to the Department of Experimental Statistics at North Carolina State University at Raleigh, and together the two departments plus groups in social sciences, sociology, and biostatistics were to constitute an Institute of Statistics. Hotelling remained chairman at Chapel Hill until 1952 and formally retired in 1966; however, afterwards he continued to remain active in the department. At North Carolina, he continued his research, publishing

papers in 1947 and 1951 on hypothesis tests related to multivariate analysis of variance, and introduced the criteria which continues to be referred to as the Lawley-Hotelling trace.

At both Columbia University and at the University of North Carolina, Hotelling excelled at attracting outstanding colleagues. Hotelling was intimately involved in bringing Abraham Wald to Columbia, as well as helping to attract Jacob Wolfowitz and W. Allen Wallis. At Chapel Hill such renowned statisticians as R. C. Bose, Wassily Hoeffding, P. L. Hsu, William Madow, Herbert Robbins, and S. N. Roy, were drawn there by Hotelling. He also was successful at developing young researchers who later would be renowned in their careers. He was an early mentor of both Milton Friedman and Kenneth Arrow, both of whom were later to win the Nobel Prize in Economics.

Hotelling's best known statistical contributions in multivariate analysis remain vital and current to present times. These specifically include Hotelling's  $T^2$ , principal components, and a number of correlational techniques.

Although introduced in 1908, Student's  $t$ -test remains as one of the most used statistical procedures for univariate data. Hotelling (1931) had recognized that experiments often have multiple measurements on each individual, and thus multiple univariate  $t$ -tests would be correlated. Hotelling's elegant solution was to propose a vector formulation of Student's test which yields a quadratic form whose distribution under the null hypothesis is that of an  $F$ -distribution. In particular, he showed that the general distribution of the  $T^2$  does not depend on nuisance parameters, but only on a quadratic form in the population mean vector, in which the matrix of the quadratic form is the inverse of the population covariance matrix. In showing this, Hotelling made use of invariance, thereby anticipating a theory developed much later. The multivariate version of Student's  $t$ -statistic remains known as Hotelling's  $T^2$  statistic.

To untangle the correlation and variability that exist among multiple measurements  $x_1, \dots, x_q$  on an individual, Hotelling (1933) introduced principal components. Principal components are uncorrelated linear combinations of the original measurements, each successfully decreasing in variation, and yet in total preserving the variation of the original measurements. Hotelling showed that the theoretical solution to the principal component problem involved finding the characteristic roots of a population covariance matrix. This then naturally led to the study of the distribution of the roots of the sample covariance matrix and thereby opened a new research area of statis-

tics involving roots of determinantal equations. Because characteristic roots of a covariance matrix were not readily computed numerically at that time, Hotelling suggested a power method which had the effect of accentuating the largest and smallest roots. This procedure was adopted for sometime by numerical analysts until more effective factorization methods were later developed.

Beginning with his early research, Hotelling was interested in relations between variables. His first publication in 1925 dealt with the distribution of correlation ratios. In 1936, he introduced the concept of canonical correlations. To handle relationships between two sets of multiple measurements on an individual, he extended the multiple correlation coefficient which had been previously introduced to study the correlation between a single variable  $y$  and multiple measures  $x_1, \dots, x_q$ . The first canonical correlation between measures  $y_1, \dots, y_p$  and  $x_1, \dots, x_q$  is the maximum correlation between separate normalized linear combinations of the  $x$ 's and the  $y$ 's. Subsequent canonical correlations are similarly defined, but constrained to be uncorrelated of earlier chosen canonical variates.

In data analysis, the first canonical variate provides what might be termed “the most predictable criterion,” which was studied by Hotelling in 1935 in the context of educational psychology. In 1936, Hotelling, jointly with Margaret Pabst, studied rank correlations, and in 1953 presented a Royal Statistical Society paper on correlations and their transforms. This paper remains as one of the definitive papers that discusses properties of correlation and Fisher's  $z$ -transformation on Fisher's  $z$ -transformation.

For other reviews and discussions about Hotelling, see Olkin, Ghurye, Hoeffding, Madow and Mann (1960), Rubin (1960), and Darnell (1990). A complete bibliography for Hotelling is given in Olkin *et al* (1960) and his articles on mathematical economics are reprinted in Darnell (1990).

Hotelling was married in 1920 to Floy Tracy and they had two children. She died in 1932, and Hotelling later married Susanna Edmundson in 1934 and together they had five sons, and a daughter (who died in her infancy).

Hotelling was elected to the National Academy of Sciences in 1970 and received a number of other prestigious honors during his life. After a stroke the preceding year, Harold Hotelling died on December 26, 1973.

## References

- [1] Darnell, Adrian (1990). *The Collected Economics Articles of Harold Hotelling*. Springer Verlag, New York.
- [2] Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics*, **2**, 360-378.
- [3] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417-441, and 498-520.
- [4] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **27**, 321-77.
- [5] Hotelling, H. (1940). The teaching of statistics. *Annals of Mathematical Statistics*, **11**, 457-70.
- [6] Hotelling, H. (1947). Multivariate quality control, illustrated by the air testing of sample bombsights. In *Selected Techniques of Statistical Analysis*. (Eds. C. Eisenhart, M. W. Hastay, and W. A. Wallis). Chapter 3. McGraw-Hill, NY.
- [7] Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., and Mann, H. B. (1960). *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press, Stanford, CA.
- [8] Rubin, H. (1960). Preface to "Three papers in honor of Harold Hotelling at 65," *The American Statistician*, **14**, 15.
- [9] Wallis, W. A. (1980). The Statistical Research Group, 1942-45. *Journal of the American Statistical Association*, **75**, 320-330.

Ingram Olkin and Allan R. Sampson