



Biometrics and Psychometrics: Origins, Commonalities and Differences

John Gower
The Open University

Abstract

Starting with the common origins of biometrics and psychometrics at the beginning of the twentieth century, the paper compares and contrasts subsequent developments, informed by the author's 35 years at Rothamsted Experimental Station followed by a period with the data theory group in Leiden and thereafter. Although the methods used by biometricians and psychometricians have much in common, there are important differences arising from the different fields of study. Similar differences arise wherever data are generated and may be regarded as a major driving force in the development of statistical ideas.

Keywords: biometrics, psychometrics, history of statistics.

1. Introduction

Biometrics and psychometrics have common roots at the beginning of the 20th century. Karl Pearson was engaged with the multi-normal distribution and its regression properties, while Spearman in the UK and Thurstone in the USA were immersed in studying correlation, also in the context of multi-normal distributions. [Pearson \(1901\)](#) also introduced principal components analysis (PCA) in his paper "On Lines and Planes of Closest Fit" and [De Leeuw \(1983\)](#) showed how [Pearson \(1906\)](#) had come close to correspondence analysis. Galton was behind both approaches and behind him was Albert Gifi!

About twenty years later Fisher was appointed to Rothamsted, initially for six months but he stayed for 14 years. Like many of his generation Fisher was interested in Eugenics and thence population genetics, but this was not his job at Rothamsted, which was to see if additional information could be gleaned from the previous 75 years of agricultural field experiments. In 1992, the centenary of Fisher's birth, Springer in their journal *Chance* published comments from statisticians giving what they thought were Fisher's greatest contributions to statistics.

Most mentioned likelihood and other aspects of scientific inference but a few, including me, said experimental design. At the time I surprised myself by selecting experimental design but I still think I was right. Design includes all aspects of data collection and subsequent analysis and interpretation. Only recently, through the work of [Parolini \(2015\)](#) in Bologna, did I come to realize how closely Fisher was involved in making sure that a daily detailed diary of the Rothamsted experiments was recorded in the famous White Books (which detailed experimental work at Rothamsted). Without good data, collected in an unbiased way, analysis and interpretation are of questionable value.

My own career began at Rothamsted in 1956, in the biometric field, where Frank Yates was Head of the Statistics Department. I was concerned with early work on statistical computing which involved writing standard programs and, crucially, with what Yates termed the Research Statistical Service, which was concerned with miscellaneous statistical problems that might benefit from the newly available computing facilities. These were quite primitive and everything, including a division subroutine and floating-point arithmetic, had to be programmed in an esoteric machine code. It is astonishing what could be done and rapid progress was made in providing matrix inversion and eigenvalue code as well as the first programs for analysing experiments (see [Cox 2015](#); [Gower 2015](#), for further details). Of relevance here is the canonical variate analysis of teeth measurements pertaining to populations of modern and fossil hominids ([Ashton, Healy, and Lipton 1957](#)) which was in progress at the time of my arrival. A year or two later I was asked to program the hierarchical classification scheme devised by Peter Sneath. This initiated my interest in (dis)similarity coefficients and in numerical taxonomy. Another influence was W. T. Williams who at the time was Professor of Botany at Southampton University. It is good to recall that much of the impetus to these new developments came from biologists. Williams was interested in botanical quadrat data and had devised an “association index” based on 2×2 chi-squared tables which could be assembled into a symmetric matrix. He then formally did a principal component analysis which seemed to give results consistent with ecological expectations. Sorting out the tangle gave rise to principal coordinate analysis. I became convinced that the fundamental idea behind many of these problems was some concept of “distance” and although the data were often presented in what looked like a cases \times variables data-matrix, they had only a very tenuous connection with the concept of correlation and, indeed, may not necessarily be open to probabilistic development. Thus I became involved with “classification” but I was surrounded by colleagues who influenced me with their work on design, agricultural surveys, generalized linear models and much else. Only in 1976 did I become aware of parallel work in psychometrics, where principal coordinate analysis was known as classical scaling. I had met Jan occasionally in the 1980’s but apart from getting a prepublication copy of [Gifi \(1990\)](#) I had had little direct contact. So it was a pleasant surprise when after retirement from Rothamsted in 1990, I was invited by Willem Heiser to spend a year with the data theory group in Leiden; the initial year extended to over two years. By the time that I arrived in Leiden, Jan had moved to California but nevertheless his influence on me, despite his geographical remoteness, has been immense.

2. Some examples of biometric and psychometric contrasts

The following is essentially an account of how the change from a biometric to a psychometric milieu impinged on my own work.

2.1. Correlation versus regression

A glance at *Statistical Methods for Research Workers* (Fisher 1938) shows that Fisher did not hold correlation in the same esteem as did psychometricians. There were two contributing factors: one is that correlation (at least product moment correlation) is essentially concerned with parameters in the bi(multi)normal distribution and is of marginal importance when analysing experiments and the other is that linear models (as expressed in Analysis of Variance) are a better and more useful tool (see what Fisher had to say about inter and intra class correlation). Another reason is that Fisher's analysis of experiments is based on the Gauss linear model and its extensions and not on the multivariate normal distribution. I return to this major difference in my concluding remarks.

2.2. Data matrix, correlation matrix

With PCA it is not the correlations that are the primary interest but the underlying data matrix from which they derive. Thus we are more in tune with the singular value decomposition (SVD). Stewart (1993) showed how the least-squares properties of the SVD were first given by Schmidt (1907), in the context of approximating the kernel functions of integral equations, but for those working in data-analysis the classical reference continues to be Eckart and Young (1936). Pearson (1901) came close to discovering the SVD, as did Fisher and Mackenzie (1923) but both were concerned with fitting only one dimension. It was only with Hotelling (1933) that a link with factor analysis emerged, where the objective is to find a best-fitting correlation matrix, perhaps excluding its diagonal elements. The confusion persists between fitting a data matrix (more prevalent in biometrics) with fitting some form of correlation matrix (more prevalent in psychometrics). I shall return to data-structure in Section 2.6 below or, for a more extended discussion, see Gower (2006).

2.3. Experiments versus observational data

Frank Yates was expert in the design and analysis of agricultural surveys (Yates 1949) and he wrote one of the first computer programs for survey analysis. It was slow and certainly not user-friendly but it had surprisingly good facilities for coping with a variety of survey designs (e.g., stratified, multi-stage, multi-phase). It produced multiway tabulations, with or without margins, and had a facility to print quantitative tables of totals/means with "associated counts". The associated counts were useful for deriving error estimates, so it was something of a surprise to find that in the Social Sciences the associated counts were often the main objective and themselves analyzed, especially by correspondence analysis or loglinear models and so on. I recall Frank being very dismissive of a survey program that only delivered tables of counts. Observational data are a poor substitute for experimental data; they are fine for finding out how things are but not so good for understanding causal relationships. I recall an example where a survey had shown that potato yields were associated with depth of ploughing, so agronomists wished to advise farmers to plough deeper. But the statisticians said that first it would be wise to do some experiments; these were done and showed that depth of ploughing was not important. Subsequent investigations showed that depth of ploughing was a sign of general overall efficiency, so the more conscientious farmers tended to have the better yields.

2.4. Diagonalization

In my own work I have often found a requirement for the diagonalization of two matrices but I am not sure who did it first. Perhaps it can be found in the first book on matrix algebra by [Weierstrasse \(1868\)](#); soon after, [Rayleigh \(1877\)](#) was using simultaneous diagonalization in his work on vibrating strings, where the ratio of two quadratic forms is termed the Rayleigh coefficient, to this day generating a significant literature in applied physics journals. In statistics, [Hotelling's work \(Hotelling 1933\)](#) on canonical correlation was among the first applications of simultaneous diagonalization in statistics, though a year earlier [Hirschfeld \(1935\)](#), in his development of correspondence analysis as the optimal scaling of the correlation between two categorical variables, had allowed one quadratic form to be diagonal, entailing that so was the other. Astonishingly, it was not until [Newcomb \(1961\)](#) that two semidefinite quadratic forms could be handled, which, with improvements from [De Leeuw \(1982\)](#), brought the advances into psychometrics; we had to wait until [Albers, Critchley, and Gower \(2011\)](#) to handle indefinite cases (see [Gower 2014b](#), for a brief history of algebraic forms used in data analysis). The latest manifestation of simultaneous diagonalization seems to be in the trust region optimization problems of numerical analysis.

2.5. Quantitative versus categorical versus counting

It has long been recognized that categorical variables and counts are much more prevalent in the social sciences and psychometrics than in the biological sciences. This is true and is a natural consequence that biologists find it much simpler to measure quantitative variables than do social scientists. The distinction is far from absolute. For example, biometrics has Fisher's optimal scores ([Fisher 1938](#)) while in psychology, quantifying categorical data has been most profoundly influenced by [Guttman \(1941\)](#). Nevertheless, I was surprised to find that [Gifi \(1990\)](#) actually advocated that when the data are quantitative it may be good to first transform them into categorical form. This seemed perverse but I have come to accept it might be a reasonable thing to do. It is confusing that the reverse process of transforming categorical variables to give optimal numerical scores, perhaps constrained to be ordinal, is also a reasonable thing to do. Both processes could be applied to the same basic data, perhaps iteratively. Behind these shenanigans seems to be the desire to transform the data to a simpler manageable form - perhaps additive, or linear or low-dimensional. A similar desire occurs with generalized linear models, though there the transformation is specified in a link function, whereas the Gifi approach more flexibly derives transformations from the data. This example, is far from the only major difference between biometric and psychometric practice that I was confronted with in the psychometric world. In biometrics it is perfectly respectable to study genotype/environment interaction when breeding new varieties of cereal crops. For obvious reasons this is not acceptable in the social sciences.

2.6. Computing, ALS and constraints

Statistical computing was a big thing at Rothamsted and Frank Yates and John Nelder were both deeply involved. So was I in the beginning but after ten years I retired to the periphery of computing activities and concentrated on data-analysis. Gifi is well-known for his use of alternating least-squares (ALS) with its bewildering plethora of ALS acronyms. I do not know when ALS was first used but it was well-known in the 1930s for the iterative solution of fitting constants to two-way tables and for three-way-models too ([Stevens 1948](#)). In the 1960s

Yates had used an ALS algorithm for fitting constants to multiway tables, allowing for main effects and multi-factor interactions and had extended this to handle quantal data using the probit transformation. All this involved ALS in trumps. Using ALS algorithms to fit least-squares criteria is to be expected but it is probably worth noting that the algorithm used to fit the likelihood criterion of generalized linear models (GLIM) is alternating reweighted least squares (John Nelder insisted on reweighted rather than weighted). Linear or generalized linear models were important in both Yates' and Nelder's work. One of Gifi's substantial contributions was to identify a wide range of nonlinear data analytical problems that could be recast into two components each compatible separately with the ALS approach and thus allow ALS to fit nonlinear models. The numerical scoring of categorical or ordinal variables is readily included within the ALS framework. The separable criteria developed often suggest generalizations of more elementary criteria but care has to be taken with how constraints are handled. With criteria in elementary form, mild constraints are often unexceptional but when generalized the difference between identification constraints and substantive constraints can be crucial. The Gifi generalization of canonical correlation compares very favorably with other methods for handling K sets of variables that have been proposed in the literature. It offers a very good example because the normalized weak identification constraints convenient when $K = 2$ need care when extended to $K > 3$, with weak constraints becoming substantive. Another example is provided by Healy and Goldstein (1976) who showed that the scaling of canonical vectors when calculating Guttman's optimal scores can create serious anomalies. Gower (1998) showed that the anomalies arise from fitting a constrained criterion rather than the original ratio criterion. Basically, in $\max(\frac{x^\top Ax}{x^\top Bx})$ then x may be scaled arbitrarily, provided $x \neq 0$: in particular $x^\top Bx = 1$ is convenient. Equivalently the ratio may be written as $\max(x^\top Ax)$ subject to $x^\top Bx = 1$ but if we change the constraint to, say, $x^\top 1 = 1$ as was considered by Healy and Goldstein (1976), then we are solving a completely different problem which may, or may not, have any practical value. The basic ratio form of the criterion is consistent with the constrained form but only when the quadratic constraint is imposed.

In general, the growth of computing has narrowed the distinction between closed form solutions and iterative solutions and in my opinion to a large measure it has vanished. Actually, it may never have existed. A closed form solution seems to have been one that can be expressed in terms of known functions. A known function is one that can be expressed in terms of more elementary known functions, has known properties and has been tabulated – which implies that its values can be calculated. We know that there is no general closed form solution for the roots of a polynomial of degree greater than four, so iterative solutions have to be admitted as part of the definition of closed form even for simple eigenvalue problems. With this perspective, anything that can be calculated algorithmically, has a potential closed form solution expressed in terms of other algorithms; the relationships may be recursive. It does not matter how complicated is the algorithm or how many arguments it has or how many new arguments are generated. So what attributes of closed form remain in the computer age? Well, tabulation is usually out of the question but, equivalently, the algorithm itself can generate any solutions required so tabulation is unnecessary. That leaves the study of the properties of any proposed algorithm included in any computer package or in stand-alone implementations. The study of algorithms includes comparisons of different implementations of essentially the same basic algorithm as well as different criteria proposed for fitting the same model. This is a vast area which includes the study of the propensity for non-unique solutions or suboptimal solutions or inaccurate solutions and much more. Incidentally, non-unique so-

lutions may be acceptable but the nature of any non-uniqueness has to be understood. Thus the requirement that closed-form functions have known properties is the one with most need of attention when applied to functions defined by algorithmically based software.

Data structure is very important in both biometrics and psychometrics but it is more frequently discussed in the psychometric literature. Psychometrics often deals with human subjects, which prevents psychologists from applying certain operations for ethical reasons and so they use surrogate variables and factor analysis methods to identify latent variables. Although these methods have impinged on biology they are far less used than in psychology. In my own work on biplots ([Gower, Lubbe, and Roux 2010a](#)), I have preferred to exhibit the measured variables themselves, as generalized calibrated axes and dispense altogether with any latent variables, although, of course, these can always easily be displayed if they are really needed. Because of the difficulties of measurement in psychometrics, many “scales” are in use (e.g., quantitative, categorical, ordinal, a 10-point scale, rankings, dichotomies, multiple choice, paired comparisons and many more). [Nishisato \(1993\)](#) give a good discussion on why scales chosen for one form of analysis cannot necessarily be transferred when used for data on another scale. In other words, data type and data structure should be carefully distinguished. I had a footnote in history here, as [Knuth \(1968\)](#) refers to a paper of mine ([Gower 1962](#)) as being one of the first to separate the structure of data (in this case multiway tables) from what operations may be required. Similar choices of scale as those used in psychometrics also occur in biometrics but not with anything like the same frequency. For computing purposes all scales including, those of qualitative variables, have to be presented in a coded digital form. Sometimes the coding tends to be conflated with data values themselves and hence gets mixed up with the data structure. Even a simple dichotomy A, B has its problems. Do A and B represent two different forms, or does B mean not A , or does B mean not known. I prefer to reserve the word structure for geometrical properties and leave the values contained in the structure to be handled independently. Thus acceptable structures may be multiway tables (crossed and/or nested), matrices (diagonal, symmetric, skew-symmetric) or sets of previously defined structures.

2.7. Hedra: Three-way, skew-symmetry, orthogonality, indscal

I shall finish with one of the major differences between biometrics and psychometrics. As said above, in biometrics the norm is to base analyzes on linear or generalized linear models, which have no restriction on the number of “ways” a table is classified and where crossing and nesting (and other) classifications are routinely allowed. Psychometrics, has followed the work of [Tucker \(1966\)](#) who proposed three way PCA, probably motivated by the factor analytical interpretation of PCA. Jan (see [Kroonenberg and De Leeuw 1980](#)) has contributed to the algorithmic basis of this extension and [Kroonenberg \(2008\)](#) has written an excellent account of the vast amount of work which has ensued. I here intend to comment only on the one parallel between the SVD and three-way algebraic decompositions where attempts have been sought to unify the algebraic trilinear decompositions in a similar way to what has been achieved for bilinear models. It is not difficult devise an ALS algorithm to fit the individual scaling model (effectively, the Tucker2 model). This gives what looks like an SVD but with product terms replaced by triple-product terms. With SVD the terms of the decomposition are ordered, orthogonal and have a minimal rank property and, as rightly pointed out by [Stewart \(1993\)](#), it is this property rather than the SVD itself which makes the SVD so useful. Thus, an obvious problem is to find a unique trilinear decomposition with optimal least-

squares properties. The SVD depends on inner products and their associated distances and I conjecture that three-dimensional content might fulfil a similar role for three-way arrays (see [Albers and Gower 2014](#)). This involves the concept of hedra, that is two-dimensional visualizations of orthogonal rank-two entities which can replace the usual one-dimensional coordinate axes. The concept of hedra has already been used for underpinning skew-symmetry and orthogonal matrices ([Gower 2014a](#)) and has some links to representations of complex numbers. It is unfortunate that the fundamental inner-product concept is expressed in terms of cosines $\text{abcos}(\phi)$ rather than the area $\text{absin}(\phi)$ which would be appropriate for content but $\text{abcos}(\phi + \pi/2) = \text{absin}(\phi)$ shows that it is easy to transform inner products to areas merely by rotating one axis through a right-angle ([Gower, Groenen, and Van De Velden 2010b](#)). The hedron concept suggests that perhaps we should be looking at conventional r -dimensional visualizations to approximate lower dimensional hedral representations, as was done by [Albers and Gower \(2014\)](#) for underpinning the Indscal model. There, a rank three model is represented by three orthogonal planes to which all constituent points are allocated. This is an interesting geometrical structure – if only it could be endowed with similar least-squares properties that the SVD has for bilinear models.

3. Conclusion

I have discussed the common origins of biometrics and psychometrics. Although there were many differences in how they developed, much common ground remained. It was lucky that Fisher went to work in agriculture, where experiments took a year to reach harvest and where the simultaneous comparison of treatment effects was necessary for efficiency. It was not a disaster if a plot of wheat performed poorly, or even died. Things were different in fruit and animal experiments and when things moved into the area of clinical and pharmaceutical experiments, a whole new range of problems were encountered, some ethical and some due to variations in human life span, perhaps associated not only with age but also with the given experimental treatments. In industrial experimentation the limitations of having to work within an annual harvest cycle were not important, and many small experiments can be conducted within a single year. Each field of application, not least the social sciences, has its own boundaries and each has to respond to its own constraints on what measurements are acceptable and desirable. New measures are constantly being created. It was only gradually that I became aware of the importance of substantive fields of research on the growth of statistical ideas. My early involvement in taxonomic problems began a process that was crystallized when I made the transition from biometrics to psychometrics. To a major extent, statistical methodology follows the science, not the other way around.

We have been encouraged to suggest things for Jan to do in retirement. I don't think that Jan will have any problems in finding things to do for himself but it would be splendid if he were to write an account of the development of data analysis in the first half of the 19th century. [Hald \(1998\)](#) in his marvellous history of statistics, has given a detailed account of the beginnings of the linear model under Gauss and Legendre and he gives an account of how Galton made the bivariate (soon to become multivariate) normal distribution a cornerstone of his work in genetics but [Hald \(1998\)](#) leaves off just at the point when so many new ideas were being developed. We have already touched on the beginnings of factor analysis and Karl Pearson's development of multivariate correlation and regression is well-known. Fisher used the Gauss linear model as the basis of models to analyze experiments, with an emphasis on a

difference between the dependent variable and associated independent variables. Especially the independent variables might be treated as “dummy variables”, to denote categorical variables, sometimes ordered categories. Who first used dummy variables in a regression context? Fisher also discussed variance components, showing that he was well aware that independent variables could be random variables. Random variation could be a manifestation of natural biological behavior but could also contain a component of measurement error. Mixed models have had something of a resurgence in the 21st century. Even in the 19th century split plots were used in field experiments and Fisher incorporated the practice into his 20th century experiments, so initiating the first multi-level statistical models. And, of course, there are always the linear structural relationships and problems of separating cause from effect.

It would be a service to all if Jan could disentangle all this.

References

- Albers CJ, Critchley F, Gower JC (2011). “Applications of Quadratic Minimisation Problems in Statistics.” *Journal of Multivariate Analysis*, **102**(3), 714–722. doi:10.1016/j.jmva.2010.11.009.
- Albers CJ, Gower JC (2014). “A Contribution to the Visualisation of Three-Way Arrays.” *Journal of Multivariate Analysis*, **132**, 1–8. doi:10.1016/j.jmva.2014.07.013.
- Ashton EH, Healy MJR, Lipton S (1957). “The Descriptive Use of Discriminant Functions in Physical Anthropology.” *Proceedings of the Royal Society of London Series B*, **146**, 552–572. doi:10.1098/rspb.1957.0030.
- Cox DR (2015). “A Conversation with John C. Gower.” *International Statistical Review*, **83**(3), 339–356. doi:10.1111/insr.12094.
- De Leeuw J (1982). “Generalized Eigenvalue Problems with Positive Semi-Definite Matrices.” *Psychometrika*, **47**(1), 87–93. doi:10.1007/bf02293853.
- De Leeuw J (1983). “On the Prehistory of Correspondence Analysis.” *Statistica Neerlandica*, **37**(4), 161–164. doi:10.1111/j.1467-9574.1983.tb00810.x.
- Eckart C, Young G (1936). “The Approximation of One Matrix by Another of Lower Rank.” *Psychometrika*, **1**(3), 211–218. doi:10.1007/bf02288367.
- Fisher RA (1938). *Statistical Methods for Research Workers*. 7th edition. Oliver and Boyd, Edinburgh.
- Fisher RA, Mackenzie WA (1923). “Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties.” *The Journal of Agricultural Science*, **13**(3), 311–320. ISSN 1469-5146. doi:10.1017/s0021859600003592.
- Gifi A (1990). *Nonlinear Multivariate Analysis*. John Wiley & Sons, New York.
- Gower J, Lubbe S, Roux N (2010a). *Understanding Biplots*. John Wiley & Sons. doi:10.1002/9780470973196.refs.

- Gower JC (1962). “The Handling of Multiway Tables on Computers.” *The Computer Journal*, **4**(4), 280–286. doi:10.1093/comjnl/4.4.280.
- Gower JC (1998). “The Role of Constraints in Determining Optimal Scores.” *Statistics in Medicine*, **17**(23), 2709–2721. doi:10.1002/(sici)1097-0258(19981215)17:23<2709::aid-sim37>3.0.co;2-8.
- Gower JC (2006). “Analyzing and Visualising Two-Way Arrays.” In M Greenacre, J Blasius (eds.), *Multiple Correspondence Analysis and Related Methods*. Chapman and Hall, Boca Raton.
- Gower JC (2014a). “Skew Symmetry in Retrospect.” *Advances in Data Analysis and Classification*, pp. 1–9. doi:10.1007/s11634-014-0181-7.
- Gower JC (2014b). “Some History of Algebraic Forms and Data Analysis.” In J Blasius, M Greenacre (eds.), *Visualisation and Verbalization of Data*, pp. 17–29. Chapman and Hall, Boca Raton, Florida.
- Gower JC (2015). “The Development of Statistical Computing at Rothamsted.” *International Statistical Review*, **83**(3), 357–370. doi:10.1111/insr.12093.
- Gower JC, Groenen PJJ, Van De Velden M (2010b). “Area Biplots.” *Journal of Computational and Graphical Statistics*, **19**(1), 46–61. doi:10.1198/jcgs.2010.07134.
- Guttman L (1941). “The Quantification of a Class of Attributes: A Theory and Method of Scale Construction.” In P Horst (ed.), *The Prediction of Personal Adjustment*, pp. 319–348. Social Science Research Council, New York.
- Hald A (1998). *A History of Mathematical Statistics from 1750–1930*. John Wiley & Sons, New York.
- Healy MJR, Goldstein H (1976). “An Approach to the Scaling of Categorized Attributes.” *Biometrika*, **63**(2), 219–229.
- Hirschfeld HO (1935). “A Connection between Correlation and Contingency.” *Mathematical Proceedings of the Cambridge Philosophical Society*, **31**(4), 520–524. doi:10.1017/s0305004100013517.
- Hotelling H (1933). “Analysis of a Complex of Statistical Variables into Principal Components.” *Journal of Educational Psychology*, **24**, 417–441. doi:10.1037/h0071325.
- Knuth DE (1968). *The Art of Computer Programming: Volume 1, Fundamental Algorithms*. Addison-Wesley, Reading.
- Kroonenberg PM (2008). *Applied Multiway Data Analysis*. John Wiley & Sons, Hoboken.
- Kroonenberg PM, De Leeuw J (1980). “Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms.” *Psychometrika*, **45**(1), 69–97. doi:10.1007/bf02293599.
- Newcomb RW (1961). “On the Simultaneous Diagonalization of Two Semi-Definite Matrices.” *Quarterly Journal of Applied Mathematics*, **19**, 144–146.

- Nishisato S (1993). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Psychology Press. doi:10.4324/9781315806907.
- Parolini G (2015). “The Emergence of Modern Statistics in Agricultural Science: Analysis of Variance, Experimental Design and the Reshaping of Research at Rothamsted Experimental Station, 1919–1933.” *Journal of the History of Biology*, **48**(2), 301–335. doi:10.1007/s10739-014-9394-z.
- Pearson K (1901). “On Lines and Planes of Closest Fit to Systems of Points in Space.” *Philosophical Magazine*, **2**(6), 559–572. doi:10.1080/14786440109462720.
- Pearson K (1906). “III. On Certain Points Connected with Scale Order in the Case of the Correlation of Two Characters Which for Some Arrangement Give a Linear Regression Line.” *Biometrika*, **5**(1-2), 176–178.
- Rayleigh L (1877). *Theory of Sound*. Dover Publications, New York.
- Schmidt E (1907). “Zur Theorie der linearen und nichtlinearen Integralgleichungen.” *Mathematische Annalen*, **63**(4), 433–476. doi:10.1007/bf01449770.
- Stevens WL (1948). “Statistical Analysis of a Non-Orthogonal Tri-Factorial Experiment.” *Biometrika*, **35**(3-4), 346–367.
- Stewart GW (1993). “On the Early History of the Singular Value Decomposition.” *SIAM Review*, **35**(4), 551–566. doi:10.1137/1035134.
- Tucker LR (1966). “Some Mathematical Notes on Three-Mode Factor Analysis.” *Psychometrika*, **31**(3), 279–311. doi:10.1007/bf02289464.
- Weierstrasse K (1868). “Zur Theorie der bilinearen und quadratischen Formen.” *Monatsberichte Königlich-Preussische Akademie der Wissenschaften Berlin*, pp. 310–338. doi:10.1017/cbo9781139567817.002.
- Yates F (1949). *Sampling Methods for Censuses and Surveys*. C. Griffin, London.

Affiliation:

John Gower
 Department of Mathematics and Statistics
 The Open University
 Walton Hall
 Milton Keynes MK7 6AA, United Kingdom
 E-mail: John.Gower@open.ac.uk

Journal of Statistical Software

published by the Foundation for Open Access Statistics

September 2016, Volume 73, Issue 5

doi:10.18637/jss.v073.i05

<http://www.jstatsoft.org/>

<http://www.foastat.org/>

Submitted: 2016-03-24

Accepted: 2016-06-01
