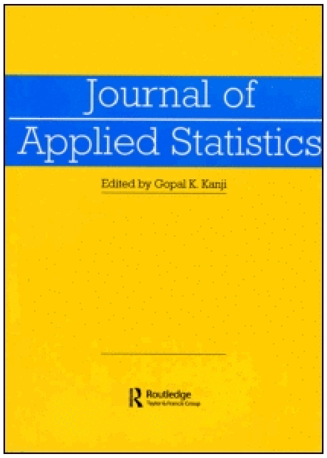

Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713428038

A note on the history of regression

David J. Finney

Online Publication Date: 01 October 1996
To cite this Article: Finney, David J. (1996) 'A note on the history of regression', Journal of Applied Statistics, 23:5, 555 - 558
To link to this article: DOI: 10.1080/02664769624099
URL: http://dx.doi.org/10.1080/02664769624099

Journal of Applied Statistics, Vol. 23, No. 5, 1996, 555–557

A note on the history of regression

DAVID J. FINNEY, *Edinburgh, UK*

SUMMARY *The first notion of statistical regression is usually attributed to Francis Galton. Recent work in a special field of applied statistics has brought to light the work of the neglected pioneer R. J. Adcock. Despite errors of execution, he deserves remembering in the history of our science. He came close to inventing linear regression; he also saw the need for specifying criteria for optimal estimation of parameters, and, with a little clearer understanding, he might have become the first recorded maximizer of likelihood.*

1 Introduction

The World Health Organization (WHO) has specified (WHO, 1983) the statistical analysis appropriate to the calibration of thromboplastins. This calibration is an essential step before a new thromboplastin preparation is approved for use in the clinical management of blood coagulation. The laboratory procedure provides clotting times for paired blood samples, measured in the presence of the standard thromboplastin preparation and of the new. If x and y denote the logarithms of those times, the N pairs of values (x, y) commonly accord with a hypothesis that both are subject to the same error distribution about an underlying linear functional relation (Kirkwood, 1983) From a set of N pairs, the WHO method determines the line that minimizes the sum of squares of the N perpendicular distances of points from the line. The result has often appeared in the literature of statistics, sometimes as a proposal for avoiding unidentifiability of parameter estimation for a supposedly optimal fitted line:

$$y = \alpha + \beta x$$

when both variables are subject to random error.

Simple derivations of the estimators of α and β that result from the minimization have been published in many places, notably by Pearson (1901), Kostitzin (1939), Kermack and Haldane (1950), Madansky (1959) and Ricker (1973) to name only

Correspondence: D. J. Finney, 13 Oswald Court, South Oswald Road, Edinburgh EH9 2HY, UK.

0266-4763/96/050555-03 $6.00

a few. Especially when the strict conditions for familiar regression methods seem inappropriate, the line so determined has appealed to many non-statisticians as an alternative to the simple linear regression, commonly attributed to Francis Galton (of course also firmly rooted in classical least squares). Although sometimes termed the 'orthogonal regression line', the procedure appears to have found few practical uses and is not widely known. Logical confusion may follow if in statistical contexts the word 'regression' ceases to be reserved for a locus of conditional expectations.

2 The history

In an attempt to study the suitability and practical aspects of the WHO method, I was led to a valuable source of details of calibration practice (Van den Besselaar *et al.*, 1984). There, in a chapter by E. A. Van der Velde, I first found a reference to R. J. Adcock—a name previously unknown to me. Various writers (notably Sprent, 1989) have referred respectfully to Adcock's work, but none has given any exact account of what he did!

Adcock contributed frequently to a little-known American mathematical journal, *The Analyst*, that originated in Des Moines, Iowa, in the 1870s. Writing from Monmouth, a small town in southern Illinois where he was an attorney (Stigler, 1978), he began a short general paper (Adcock, 1877) on least squares by stating:

> When a greater number of points are given or observed than are sufficient to determine any point, line or surface, that point, line or surface which makes the sum of the squares of the errors of situation a minimum, has the greatest probability, and is therefore the one determined by these points.

He was evidently thinking of the general problem of finding the 'best' equation of linear or other specified form to fit an assembly of N points in two or three dimensions. After his oddly worded introduction, he commented on maximizing the probability density for the observations as a function of the parameters of the equation, though he was a trifle confused on how to form the compounded element of the probability density for N observations. Under an implicit assumption of the normality of the probability distribution of coordinates relative to the unknown equation, he saw that this process—which today we would term 'maximization of the likelihood'—led to the 'determination' of the best equation, by minimizing the sum of squares of the distances of the N points perpendicular to the line or surface.

Adcock appeared to restrict himself to linear functions, although he may have envisaged wider applicability. Later (Adcock, 1878a), in broad outline but reading strangely today, he discussed the assignment of a probable error to his estimates. In a third paper (Adcock, 1878b), he purported to give the full solution for the special case of N points in a plane and a straight line to be estimated. It may surprise a reader today that any editor was prepared to publish an article on mathematics that, despite conceptual originality, required no more than elementary differential calculus. Adcock obtained a correct expression for the sum of squares of perpendicular distances, differentiated correctly with respect to the two parameters of the line, and showed that his 'best' line must pass through the centre of gravity of the N points.

3 The errors

Unfortunately, Adcock ended by publishing an expression for estimating β that was so dimensionally absurd as to occasion surprise that no referee noticed the error. Possibly, Adcock's clumsy symbolism caused a mistake in handling the minimizing equations. He then stated a simple invented example for $N = 3$, with all coordinates being small integers. After making a gross arithmetical error in substitution into his own incorrect formula, he concluded with a numerical result that agreed neither with his own formula nor with today's well-known correct formula! After drawing attention to Adcock's algebraic error, Kummell (1879) presented the correct result for the estimator of β. Pearson (1901) formulated a problem in terms very similar to those of Adcock, but surprisingly seems to have been unaware of the earlier author or his mistakes.

I am tempted to speculate that Adcock was at heart a pure mathematician unaccustomed to expressing results in simple algebraic terms, or to performing scientific arithmetic. He made no mention of any practical problem that had first stimulated his interest, although there are hints of an origin in astronomy. Despite twentieth century mentions of Adcock's name, I have found in them no tribute to his innovatory importance. Nevertheless, I think he should be recognized as an original thinker in the long story of a search for an algebraic function that can best represent a set of observations, and in specifying criteria to give explicit interpretation to the word 'best'.

Acknowledgements

I am grateful to Dr Antonius Van den Besselaar, whose book first put me on the track of Adcock; to Professor M. M. Goodman (from whom I first obtained a copy of Adcock's paper); and also to Professors Anders Hald, Albert Madansky, Norman W. Simmonds and Peter Sprent for invaluable help in following the trail.

REFERENCES

ADCOCK, R. J. (1877) On the method of least squares, *The Analyst*, 4, pp. 183–184.

ADCOCK, R. J. (1878a) Least squares, *The Analyst*, 5, pp. 21–22.

ADCOCK, R. J. (1878b) A problem in least squares, *The Analyst*, 5, pp. 53–55.

KERMACK, K. A. & HALDANE, J. B. S. (1950) Organic correlation and allometry, *Biometrika*, 37, pp. 30–41.

KIRKWOOD, T. B. L. (1983) Calibration of reference thromboplastins and standardisation of the prothrombin time ratio, *Thrombosis and Haemostasis*, 49, pp. 238–244.

KOSTITZIN, V. A. (1939) *Mathematical Biology* (London, Harrap).

KUMMELL, C. H. (1879) Reduction of observation equations which contain more than one observed quantity, *The Analyst*, 6, pp. 98–105.

MADANSKY, A. (1959) The fitting of straight lines when both variables are subject to error, *Journal of the American Statistical Association*, 54, pp. 173–205.

PEARSON, K. (1901) On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2, Series 6, pp. 559–572.

RICKER, W. E. (1973) Linear regressions in fishery research, *Journal of the Fisheries Research Board of Canada*, 30, pp. 409–434.

SPRENT, P. (1989) Some history of functional and structural relationships. *In:* D. J. Brown & W. A. Fuller (Eds), *Statistical Analysis of Measurement Error Models and Applications*, pp. 3–16 (American Mathematical Society).

STIGLER, S. M. (1978) Mathematical statistics in the early states, *Annals of Statistics*, 6, pp. 239–265.

VAN DEN BESSELAAR, A. M. H. P., GRALNICK, H. R. & LEWIS, S. M. (1984) *Thromboplastin Calibration and Oral Anticoagulant Control* (Boston, Martinus Nijhoff).

WHO (1983) Expert Committee on Biological Standardization, 33rd Report, Appendix 1, *WHO Technical Report* 687 (World Health Organization).