Jour. Ind. Soc. Ag. Statistics
57 (Special Volume), 2004 : 3-10

First (?) Occurrence of Common Terms in Agricultural Statistics, with Help from JSTOR

H.A. David
Iowa State University, Ames, Iowa, USA

SUMMARY

A list of presumed first occurrences of terms important in Agricultural Statistics is constructed. Terms obtained earlier without a computer are combined with terms generated by the electronic journal archive JSTOR. The advantages and limitations of JSTOR for this purpose are described.

Key words : History of agricultural statistics, Terminology.

1. Introduction

In this article we attempt to list the first occurrence in print of terms that are of particular importance in agricultural statistics. The coining of such terms is often an important step in fixing and propagating the underlying concept. Of course, the date of a term does not necessarily signify the beginning of the concept, but together with the identification of the originator adds to our understanding of the development of the field of agricultural statistics.

Except for just a few additions (diallel cross, partial diallel cross, significance test) the list of terms is a selection from a much larger list in David [14], where the criteria for inclusion and the method of construction are described in detail. However, in quite a few instances, indicated by asterisks in the present list, the date of the term has been improved, possibly with change of the originator. This updating has been made possible by use of the electronic journal archive JSTOR, as discussed in the next section.

The selection of terms is necessarily somewhat personal and it will be easy for the reader to spot absences. Each term is given in only one form, usually as a noun, even if the first occurrence is in a different form. A few major more general statistical terms have also been included.

Many individuals have contributed to the list in David [14], where their help is acknowledged.

Table 1. First (?) Occurrence of Terms in Agricultural Statistics

Additivity (in ANOVA)	McCarthy, M. D. [42]
Alias	Finney, D. J. [18]
Analysis of variance	Fisher, R. A. [20]
_____ of covariance	Fisher, R. A. [28]
Association scheme	Bose, R. C. and Shimamoto, T. [4]
Balanced incomplete blocks	Fisher, R. A. and Yates, F. [30]
_____ _____ _____, partially	Bose, R. C. and Nair, K. R. [3]
Bioassay	Wood, H. C. [64]
Biometry	Whewell, W. [39]
Biostatistics	Webster's Dictionary (1890)
Change-over design	Cochran, W. G. *et al.* [11]
Column effect	Wilks, S. S. [62]
Components of variance	Daniels, H. E. [13]
Confounding	Fisher, R. A. [25]
*_____, partial	Yates, F. [65]
Correlation	Galton, F. [34]
_____ coefficient	Pearson, K. [52]
Covariance	Fisher, R. A. [27]
*Cumulative distribution function (cdf)	Wilks, S. S. [61]
Degrees of freedom	Fisher, R. A. [22]
Diallel cross	Hayman, B. I. [37]
_____ _____, partial	Kempthorne, O. and Curnow, R. N. [38]
Efficiency	Fisher, R. A. [21]
Errors of first and second kind	Neyman, J. and Pearson, E. S. [45]
F	Snedecor, G. W. [57]
*Factorial design	Yates, F. [66]
Fixed effects	Eisenhart, C. [17]
*_____ _____model	Thompson, W. A., Jr. [58]
Fractional replication	Finney, D. J. [18]
*Goodness of fit	Pearson, K. [51]
Hypothesis, alternative	Neyman, J. and Pearson, E. S. [45]
_____, null	Fisher, R. A. [29]
*Incidence matrix	Schützenberger, M. P. [56]
Interaction	Fisher, R. A. [25]
*Interval estimation	Pearson, E. S. [48]

*Latin square	Cayley, A. [8]
Lattice (design)	Yates, F. [67]
*Least squares	Rumker, C. [54]
Level of significance	Fisher, R. A. [24]
*Linear model	Crump, S. L. [12]
Maximum likelihood	Fisher, R. A. [21]
*Mean square (of errors)	Ordnance Survey [47]
*Method of least squares	Brinkley, J. [7]
Mixed model	Mood, A. M. [44]
Model I, II (in ANOVA)	Eisenhart, C. [17]
Monte Carlo method	Metropolis, N. and Ulam, S. [43]
Multiple comparisons	Duncan, D. B. [16]
Nested	Ganguli, M. [35]
Nonadditivity	Cochran, W. G. [10]
Noncentral	Fisher, R. A. [26]
Nonparametric	Wolfowitz, J. [63]
Orthogonality (in design)	Yates, F. [65]
P value	Deming, W. E. [15]
Percentage point (essentially)	Fisher, R. A. [24]
Percentile	Galton, F. [32]
Power function	Neyman, J. and Pearson, E. S. [46]
*Probability density function	Wilks, S. S. [60]
Random effects model	Scheffé, H. [55]
_____ sampling	Pearson, K. [53]
*_____ variable	Frisch, R. [31]
Randomization	Fisher, R. A. [25]
Range	Lloyd, H. [41]
Ranked-set sampling	Halls, L. K. and Dell, T. R. [36]
Recovery of interblock information	Yates, F. [68]
Regression	Galton, F. [33]
Resolvable design	Bose, R. C. [2]
Response surface	Box, G. E. P. and Wilson, K. B. [6]
Rotatable design	Box, G. E. P. and Hunter, J. S. [5]
Row effect	Wilks, S. S. [62]

σ	Pearson, K. [50]
Scatterplot	Kurtz, A. K. and Edgerton, H. A. [39]
Significance test	Latter, O. H. [40]
Standard deviation (σ)	Pearson, K. [50]
*_____ error	Yule, G. U. [69]
Student's t (essentially)	Fisher, R. A. [23]
*Studentized range	Pearson, E. S. [49]
t	Fisher, R. A. [23]
Test of significance	Fisher, R. A. [24]
*Treatment effect	Cochran, W. G. [9]
Variance	Fisher, R. A. [19]

*Improved by JSTOR

2. JSTOR

The primary purpose of the electronic journal archive JSTOR is to make all but the recent 3-5 years of the journals featured available to users in subscribing institutions. Major journals from a variety of disciplines are covered, including 13 journals in Statistics and 17 in Mathematics, all 30 apparently published in Britain or the USA. There is no Agriculture category, but Botany and Ecology are represented.

For our search the specific feature of interest is the program's ability to pick out quickly, from all the journals in the category or categories specified by the user, the page(s) containing a designated term. The process can be speeded up further by narrowing the years to be covered by the search. For example, in looking for "analysis of variance" we might specify 1900 to 1930 if starting from scratch. Searching the Statistics and Mathematics journals JSTOR then leads us to a 1923 *Biometrika* paper by Student. But David [14] gives a 1918 Fisher article that appeared in a journal not covered by JSTOR. This is therefore also the entry in the present table.

It will be seen from our listing that JSTOR improved the earlier entries in quite a few instances. In addition, JSTOR arrived at the same entry for 9 terms. The remaining terms carried over from the older list first appeared either in books or in journals not included in JSTOR.

Some Cautions in the Use of JSTOR: It is essential to check a term generated by JSTOR for its relevance, as the term may well be used in a different sense from that intended. This may take time, especially as the term searched is not highlighted on the page. For obvious reasons JSTOR is not helpful with common words, such as "efficiency" or with symbols such as "F". For terms consisting of two or more words JSTOR may cite a page containing all the individual words, but these may be scattered over the page.

On the other hand, JSTOR is almost totally reliable in finding the earliest occurrence for journals in its coverage; other methods simply cannot compete in this respect. Also JSTOR can quickly build up a chronological sequence of references within its domain, making it a valuable aid to understanding the development of a topic and to historical research.

Finally, a recent paper (Aldrich [1]), itself of interest in dealing with an important period in the development of statistical terminology, draws attention to a remarkable web site, "Earliest Known Uses of Some of the Words of Mathematics" (http://members.aol.com/jeff570/c.html) by J. Miller. This large compilation includes entries on mainline statistical terms, usually with a commentary. It has led us to add "Significance test" to the previous entry on "Test of significance" and to correct the entry for "Analysis of covariance". The entry for the latter in David [14] applies to a different use of the term, as pointed out by Aldrich, a major contributor to the statistical terms included in the web site.

REFERENCES

[1] Aldrich, J. (2003). The language of the English biometric school. *Intern. Statist. Rev.*, **71**, 109-129.

[2] Bose, R. C. (1942). A note on the resolvability of balanced incomplete block designs. *Sankhyā*, **6**, 105-110.

[3] Bose, R. C. and Nair, K. R. (1939). Partially balanced incomplete block designs. *Sankhyā*, **4**, 337-372.

[4] Bose, R. C. and Shimamoto, T. (1952). Classification and analysis of partially balanced incomplete block designs with two associate classes. *J. Amer. Statist. Assoc.*, **47**, 151-184.

[5] Box, G. E. P. and Hunter, J. S. (1957). Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Statist.*, **28**, 195-241.

[6] Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc.*, **B(13)**, 1-45.

[7] Brinkley, J. (1821). ... observations ... for investigating the effects of parallax and aberration on the places of certain fixed stars .... *Phil. Trans. Roy. Soc.*, London, **111**, 327-360,

[8] Cayley, A. (1889). On the theory of groups. *Amer. J. Math.*, **11**, 139-157.

[9] Cochran, W. G. (1939). Long-term agricultural experiments. *Suppl. J. Roy. Statist. Soc.*, **6**, 104-148.

[10] Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, **3**, 22-38.

[11] Cochran, W. G, Autrey, K. M., and Cannon, C. Y. (1941). A double change-over design for dairy cattle feeding experiments. *J. Dairy Sci.*, **24**, 937-951.

[12] Crump, S. L. (1951). The present status of variance component analysis. *Biometrics*, **7**, 1-16.

[13] Daniels, H. E. (1939). The estimation of components of variance. *Suppl. J. Roy. Statist. Soc.*, **6**, 186-197.

[14] David, H. A. (2001). First (?) occurrence of common terms in statistics and probability. In : David, H. A. and Edwards, A. W. F. *Annotated Readings in the History of Statistics*. Springer, New York.

[15] Deming, W. E. (1943). *Statistical Adjustment of Data.* Wiley, New York.

[16] Duncan, D. B. (1951). A significance test for differences between ranked treatments in an analysis of variance. *Virgina J. Sci.*, **2**, 171-189.

[17] Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, **3**, 1-21.

[18] Finney, D. J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen.*, **12**, 291-301.

[19] Fisher, R. A. (1918a). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.*, **52**, 399-433.

[20] Fisher, R. A. (1918b). The causes of human variability. *Eugen. Rev.*, **10**, 213-220.

[21] Fisher, R. A. (1922a). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, London, **A222**, 309-368.

[22] Fisher, R. A. (1922b). On the interpretation of χ^2 from contingency tables, and the calculation of P. *J. Roy. Statist. Soc.*, **85**, 87-94.

[23] Fisher, R. A. (1924). On a distribution yielding the error functions of several well known statistics. In : *Proc. Intern. Congress Math.*, Toronto, **2**, 805-813.

[24] Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.

[25] Fisher, R. A. (1926). The arrangement of field experiments. *J. Ministry Agric., Great Britain*, **33**, 503-513.

[26] Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.*, London, **A 121,** 654-673.

[27] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford University Press.

[28] Fisher, R. A. (1933). *Statistical Methods for Research Workers*, 4th edn. Oliver and Boyd, Edinburgh.

[29] Fisher, R. A. (1935). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

[30] Fisher, R. A. and Yates, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, London.

[31] Frisch, R. (1931). A method of decomposing an empirical series into its cyclical and progressive components., *J. Amer. Statist. Assoc.*, **26**, 73-78.

[32] Galton, F. (1885a). Some results of the Anthropometric Laboratory. *J. Anthrop. Inst.*, **14**, 275-287.

[33] Galton, F. (1885b). Section H. Anthropology. Opening address by Francis Galton. *Nature*, **32,** 507-510.

[34] Galton, F. (1888). Co-relations and their measurement. *Proc. Roy. Soc.*, London, **45**, 135-145.

[35] Ganguli, M. (1941). A note on nested sampling. *Sankhyā*, **5**, 449-452.

[36] Halls, L. K. and Dell, T. R. (1966). Trial of ranked-set sampling for forage yields. *Forest Sci.*, **12**, 22-26.

[37] Hayman, B. I. (1954). The analysis of variance of diallel tables. *Biometrics*, **10**, 235-244.

[38] Kempthorne, O. and Curnow, R. N. (1961). The partial diallel cross (abstract). *Biometrics*, **17**, 164-189.

[39] Kurtz, A. K. and Edgerton, H. A. (1939). *Statistical Dictionary of Terms and Symbols*. Wiley, New York.

[40] Latter, O. H. (1902). The egg of Cuculus Canorus. *Biometrika*, **1**, 164-176.

[41] Lloyd, H. (1848). On certain questions connected with the reduction of magnetical and meteorological observations. *Proc. Roy. Irish Acad.*, **4**, 180-183.

[42] McCarthy, M. D. (1939). On the application of the z-test to randomized blocks, *Ann. Math. Statist.*, **10**, 337-359.

[43] Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *J. Amer. Statist. Assoc.*, **44**, 335-341.

[44] Mood, A. M. (1950). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

[45] Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc.*, London, **A231**, 289-337.

[46] Neyman, J. and Pearson, E. S. (1936). Contributions to the theory of testing statistical hypotheses. *Statist. Res. Memoirs*, **1**, 1-37.

[47] Ordnance Survey (1856). On the deflection of the plumb-line at Arthur's seat, and the mean specific gravity of the earth., *Phil. Trans. Roy. Soc.*, London, **146**, 591-606.

[48] Pearson, E. S. (1937). Some aspects of the problem of randomization. *Biometrika*, **29**, 53-64.

[49] Pearson, E. S. (1939). William Sealy Gosset 1876-1937. "Student" as statistician. *Biometrika*, **30**, 210-250.

[50] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Roy. Soc.*, London, **A185**, 71-110.

[51] Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Phil. Trans. Roy. Soc.*, London, **A186**, 343-414.

[52] Pearson, K. (1896). Regression, heredity and panmixia. *Phil. Trans. Roy. Soc.*, London, **A187**, 253-318.

[53] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.*, 5th Ser., **50**, 157-175.

[54] Rumker, C. (1821). Calculations of some observations of the solar eclipse on the 7th of September, 1820. *Phil. Trans. Roy. Soc.*, London, **111**, 311-315.

[55] Scheffé, H. (1956). Alternative models for the analysis of variance. *Ann. Math. Statist.*, **27**, 251-271.

[56] Schützenberger, M. P. (1951). An extension problem in the theory of incomplete block designs. *J. Roy. Statist. Soc.*, **B, 13**, 120-125.

[57] Snedecor, G. W. (1934). *Calculation and Interpretation of Analysis of Variance and Covariance*. Collegiate Press, Ames, IA.

[58] Thompson, W. A. Jr. (1955). On the ratio of variances in the mixed incomplete block model. *Ann. Math. Statist.*, **26**, 721-733.

[59] Whewell, W. (1831). In : *Letters II (1876).* Macmillan, London, p. 135 (Nov.12).

[60] Wilks, S. S. (1938a). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60-62.

[61] Wilks, S. S. (1938b). Shortest average confidence intervals from large samples, *Ann. Math. Statist.*, **9**, 166-175.

[62] Wilks, S. S. (1943). *Mathematical Statistics*. Princeton University Press.

[63] Wolfowitz, J. (1942). Additive partition functions and a class of statistical hypotheses. *Ann. Math. Statist.*, **13**, 247-279.

[64] Wood, H. C. (1912). The purpose and limitations of bio-assay. *J. Amer. Medical Assn.*, **59**, 1433-1434.

[65] Yates, F. (1933). The principles of orthogonality and confounding in replicated experiments. *J. Agric. Sci.*, **23**, 108-145.

[66] Yates, F. (1935). Complex experiments. *Supp. J. Roy. Statist. Soc.*, **2**, 181-247.

[67] Yates, F. (1937). *The Design and Analysis of Factorial Experiments*. Imperial Bureau of Soil Science, Techn. Comm. No. 35, Harpenden.

[68] Yates, F. (1939). The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugen.*, **9**, 136-156.

[69] Yule, G. U. (1896). On the significance of Bravais' formulae for regression, & c., in the case of skew correlation. *Proc. Roy. Soc.*, London, **60**, 477-489.