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

The most-cited statistical papers

Thomas P. Ryan [a]; William H. Woodall [b]
[a] National Institute of Standards and Technology, Gaithersburg, Maryland, USA [b] Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA

Online Publication Date: 01 July 2005

To cite this Article Ryan, Thomas P. and Woodall, William H.(2005)'The most-cited statistical papers',Journal of Applied Statistics,32:5,461 — 474

To link to this Article: DOI: 10.1080/02664760500079373

URL: http://dx.doi.org/10.1080/02664760500079373

Journal of Applied Statistics
Vol. 32, No. 5, 461–474, July 2005

Routledge
Taylor & Francis Group

The Most-Cited Statistical Papers

THOMAS P. RYAN* & WILLIAM H. WOODALL**

**National Institute of Standards and Technology, Gaithersburg, Maryland, USA, **Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA*

ABSTRACT *We attempt to identify the 25 most-cited statistical papers, providing some brief commentary on each paper on our list. This list consists, to a great extent, of papers that are on non-parametric methods, have applications in the life sciences, or deal with the multiple comparisons problem. We also list the most-cited papers published in 1993 or later. In contrast to the overall most-cited papers, these are predominately papers on Bayesian methods and wavelets. We briefly discuss some of the issues involved in the use of citation counts.*

KEY WORDS: Citations, history of statistics

Citations in General

There has been much discussion of the uses of citation counts in the literature, although not with respect to the statistics literature with the exceptions of Stigler (1994), Altman & Goodman (1994), and Theoharakis & Skordia (2003). Austin (1993) assessed the reliability of citation counts in making tenure and promotion decisions in academia, while Gilbert (1977) and Edge (1979) have considered citation counts as measures of the influence of research. See also Cronin (1984).

Edge (1979) criticized citation counts as being overused to measure intellectual linkages. Others have made similar criticisms. Despite such criticisms, however, the use of citation counts seems to be increasing. The National Research Council, for example, uses citation rates as one measure to rank PhD programmes in statistics and other fields. In addition, citation counts appear to be increasingly used in promotion decisions in academia, in addition to ranking scientific journals. Using ISI *Journal Citation Reports*, for example, one can determine that among the 71 journals in the Statistics and Probability category, *Statistical Science* ranked fifth in citation impact factor and 16th in the total number of citations received in 2002 with 1,051.

In attempting to determine the causal factors for highly cited papers, Donoho (2002) gave a list of suggestions for writing papers that would receive a large number of citations. At the top of his list was ‘Develop a method which can be applied on statistical data of a kind whose prevalence is growing rapidly’. For example, if someone could develop ‘the’ approach to data mining, the paper would undoubtedly garner a huge number of citations.

Correspondence Address: W. H. Woodall, Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0439, USA. Email: bwoodall@vt.edu

0266-4763 Print/1360-0532 Online/05/050461–14
DOI: 10.1080/02664760500079373

There is generally a time lag of several years before new methodology is implemented in software, so it is not surprising that number 2 on Donoho's list was 'Implement the method in software, place examples of the software's use in the paper, make the software of broad functionality, and give the software away for free.'

Garfield (1998) reported that for the period 1945–1988 the majority of cited papers in science were cited only once. In a controversial citation analysis, it was shown that 55% of papers published during 1981–1985 received no citations within five years of their publication (Hamilton, 1990). From the same data, Hamilton (1991) broke down the 55% of uncited papers and indicated that there was a huge variation across various disciplines, ranging from 9.2% of papers uncited in atomic, molecular, and chemical physics, to 86.9% in engineering. Pendlebury (1991), however, disagreed with Hamilton's analysis and reported that only 22.4% of science articles published in 1984 remained uncited by the end of 1988.

Papers are cited at different rates in different fields. ScienceWatch (1999) reported that for the years 1981–1997 a paper in mathematics needed at least 291 citations to rank in the top 0.01%, while it took 1,823 citations for the corresponding ranking in molecular biology and genetics.

What should we make of these numbers? A sceptic might contend that these studies show that much research has little or no value. Indeed, it seems apparent that most published papers do not influence the work of other researchers, although there have of course been innumerable instances in which researchers have failed to acknowledge related work. Overall, however, it seems clear that the distribution of papers in regard to their impact has a huge amount of right skewness.

The 25 Most-Cited Papers

In this section we provide our list of the 25 most-cited statistical papers. We did not limit ourselves to the primary statistical journals. We considered for inclusion only papers in which the author(s) proposed a new statistical method, modified an existing statistical method, or used an existing statistical method in a novel way to address an important scientific problem. The application of this set of criteria is necessarily subjective to a large extent. As discussed by Straf (2003), there is no generally accepted definition of 'statistics'.

The citation counts are those given on the Institute for Scientific Information (ISI) Web of Science (as of 1 December 2003). Since the Web of Science does not include all scientific journals, the counts are all undercounts. In addition, we did not attempt to make adjustments for incorrect citation information, e.g., citations that had incorrect page or volume numbers. Taking into account these factors could lead to some reordering of the top 25 or even to some papers dropping off the list. A more significant issue, however, is the fact that the ISI Web of Science citation counts does not include citations before 1945. This is thus a problem for papers published well before then, and the problem is compounded by the fact that it was more difficult for the early papers to accumulate citations since the number of scientific journals was much smaller at the time they were published than is now the case. In addition, as a method becomes a generally accepted part of statistics, e.g., the one-sample *t*-test, the citation rate of the paper in which the method was initially proposed decreases. We also calculated a *current* annual citation rate for some papers (reported in parentheses when available). This is a conservative value since it was obtained by doubling the number of citations received during a period of less than six months in the last part of 2003.

Some very highly cited papers on fuzzy logic, such as the one by Zadeh (1965) with 5,022 citations (338 per year), were not considered to be statistical papers even though there is a connection between fuzzy logic and statistics, as discussed by Laviolette *et al.* (1995). Similarly, Hopfield (1982), on the topic of neural networks with 3,574 citations (156 per year), was not included. Reed & Muench (1938), with 10,974 citations (242 per year), was not included due to the simplicity of the proposed method. Wright (1931) with 2,218 citations (144 per year) was not included since the method and results were judged to be primarily probabilistic, not statistical. In addition, there are some highly cited papers by well-known statisticians that we did not consider to be statistical enough to warrant inclusion in our list, e.g., Cooley & Tukey (1965) with 2,872 citations (78 per year) and Nelder & Mead (1965) with 5,635 citations (426 per year). Some of these decisions are debatable since Cooley & Tukey (1965) was included by Kotz & Johnson (1997) with an introduction written by I. J. Good.

It would not be surprising if some papers with significant statistical content have been overlooked in our study. In addition, it might be argued that some of the papers on our list should have been excluded for one reason or another. We welcome input on these issues from the readers of our paper.

The following is our list with some brief commentary.

(1) With 25,869 citations (currently cited 1,984 times per year),

Kaplan, E. L. & Meier, P. (1958) Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, 53, pp. 457–481.

Kaplan & Meier (1958) proposed a non-parametric method for estimating the proportion of items in a population whose lifetime exceeded some specified time t from censored survival data. This type of data is very common in medical studies. This paper not only has by far the highest number of citations of all statistics papers, but it has also been ranked among the top five most cited papers for the entire field of science. Based on data from *Journal Citation Reports*, the total number of citations received by this paper exceeds twice the number of citations received by *all Journal of the American Statistical Association* papers in 2002. This paper appeared in Kotz & Johnson (1992b, pp. 311–338) as a breakthrough paper in statistics with an introduction written by N. E. Breslow.

Kaplan (1983) reported that he and Meier had, in fact, each submitted separate manuscripts to the *Journal of the American Statistical Association.* Due to their similarity, the editor recommended that their papers be combined into one manuscript. It took them four years to resolve the differences between their approaches, during which time they were concerned that someone else might publish the idea.

Interestingly, Garfield (1989) gave this paper as an example of one that was slow to receive recognition. Indeed, Figure 3 in Garfield (1989) shows that the paper received very few citations per year through the early 1970s (i.e., for the first 15 years after it was published). It was cited only 25 times from 1958–1968. But, starting in 1975, the number of citations per year began to increase sharply and continued to increase monotonically through 1989, the last year covered by the graph. Meier is quoted in personal communication that year as stating that the needs of applied researchers were 'quite well met' by the existing methodology, and it was not until the advent of computers and the increasing mathematical sophistication of clinical researchers that the Kaplan–Meier method grew in importance and eventually was recognized as the standard.

Despite its popularity, the Kaplan–Meier method has not been without controversy. Miller (1983) wrote a paper entitled 'What price Kaplan–Meier?' in which he claimed that the Kaplan–Meier estimator was inefficient and suggested that analysts should use some parametric assumptions whenever possible. Eighteen years later, Meier (2001) responded to the paper with a talk entitled 'The price of Kaplan–Meier.' Meier believed Miller's (1983) conclusions were incorrect and initially believed that references to it would taper off for that reason. His presentation was motivated in part by the number of citations of Miller's paper. Also, see Meier *et al.* (2004).

(2) With 18,193 citations (1,342 per year),

Cox, D. R. (1972) Regression models and life tables, *Journal of the Royal Statistical Society, Series B*, 34, pp. 187–220.

The topic of this paper is the regression analysis of censored failure time data, which has far-reaching applications in the biomedical sciences. Cox (1972) used a semiparametric model for the hazard function, which has significant advantages over using parametric models for the failure time.

This paper appeared in Kotz & Johnson (1992b, pp. 519–542) as a breakthrough paper in statistics with an introduction written by R. L. Prentice. See Reid (1994) for some interesting background on this paper from D. R. Cox. Interestingly, it is reported that a key insight into the statistical analysis method first came to Professor Cox when he was quite ill with the flu and was recalled later only with some difficulty. Cox (1986) also provided some background on the paper.

(3) With 13,108 citations (256 per year),

Duncan, D. B. (1955) Multiple range and multiple *F*-tests, *Biometrics*, 11, pp. 1–42.

David Duncan presented his now-famous multiple range test for comparing the means of several populations at the Joint Meetings of the Institute of Mathematical Statistics and the Eastern North American Region of the Biometric Society in March of 1954. Although Duncan also proposed multiple *F*-tests, and in fact this was his original emphasis, these tests have not enjoyed the popularity of his multiple range test because they were more cumbersome to use.

Duncan (1977) gave some historical background on this paper. He also recommended that the methods in Duncan (1975) be used in place of his multiple range test.

(4) With 9,504 citations (488 per year),

Marquardt, D. W. (1963) An algorithm for least squares estimation of non-linear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 2, pp. 431–441.

The Marquardt algorithm proposed in this paper is used to estimate the parameters in a nonlinear model. See Hahn (1995) and Marquardt (1979) for some interesting background information on this paper.

(5) With 8,720 citations (114 per year),

Litchfield, J. T. & Wilcoxon, F. A. (1949) A simplified method of evaluating dose-effect experiments, *Journal of Pharmacological and Experimental Therapeutics*, 96, pp. 99–113.

The authors proposed a rapid graphical method for approximating the median effective dose and the slope of dose-percent effect curves. Litchfield (1977) credited Wilcoxon's intense interest in collaboration for the development of the proposed method. When Litchfield joined the laboratories where Wilcoxon was working, the two were discussing the method at Wilcoxon's request even before Litchfield had seen his employer or checked in with the personnel department.

(6) With 8,151 citations (1,590 per year),

Bland, J. M. & Altman, D. G. (1986) Statistical methods for assessing agreement between two clinical measurements, *Lancet*, 1 (8476), pp. 307–310.

The authors described simple statistical methods and graphs originally proposed by Altman & Bland (1983) for using paired data to assess the differences between measurements obtained by two different measurement systems. (The paper is available online at http://www.users.york.ac.uk/∼mb55/meas/ba.htm) See Bland & Altman (1992) and Bland & Altman (1995) for descriptions of the genesis and impact of this paper.

(7) With 6,788 citations (914 per year),

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, 39, pp. 783–791.

The context of evolutionary biology is phylogeny, the connections between all groups of organisms as understood by ancestor/descendant relationships. According to I. Hoeschele (personal communication), the human genome project and sequencing projects for other organisms provide an unprecedented amount of data to which the methods in this paper and those in Nei (1972), our Number 13 paper, can be applied. The resulting information is immensely valuable in understanding questions in evolution and in inferring the functions of genes. Phylogenetics is a very active area of research, in particular in the context of comparative genome analysis and genome-scale adaptation of methods. For more information on this topic, the reader is referred to Holmes (2003).

Felsenstein (1985) considered an application of the bootstrap method, whereas more fundamental statistical issues were addressed in the bootstrap paper in Efron (1979), which narrowly missed being in our list with 1,889 citations (156 per year). Efron (1979) appeared in Kotz & Johnson (1992b, pp. 519–542) as a breakthrough paper in statistics, with an introduction written by R. J. Beran.

(8) With 6,579 citations (126 per year),

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. & Smith, K. G. (1977) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part II. Analysis and examples, *British Journal of Cancer*, 35, pp. 1–39.

Sir Richard Peto, the first author, and Sir David Cox and Nathan Mantel, who appear in other places on this list, are among the distinguished group of co-authors of this paper. The paper is the second of a two-part report to the UK Medical Research Council's Leukemia Steering Committee. This report was focused on efficient methods of analysis of data from randomized clinical trials for which the duration of survival among different groups of patients is to be compared.

(9) With 6,006 citations (422 per year),

Mantel, N. & Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute*, 22, pp. 719–748.

These authors proposed a chi-square test with one degree of freedom for testing the association of disease incidence using 2 × 2 contingency tables.

(10) With 5,260 citations (300 per year),

Mantel, N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports*, 50, pp. 163–170.

Mantel (1966) was also cited by Garfield (1989) as a paper that was slow to receive recognition. Mantel was apparently philosophical about this, stating in personal communication to Garfield in 1989, 'Actually, slow initial rise characterizes nearly everything', and also reasoned that his method was slow to gain recognition by statisticians and epidemiologists because it was published in a cancer journal.

(11) With 4,306 citations (492 per year),

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (C/R: pp. 22–37), *Journal of the Royal Statistical Society, Series B*, 39, pp. 1–22.

The Expectation Maximization (EM) algorithm is used for maximum likelihood estimation with data for which some variables are unobserved. Much has been written about the algorithm, which coupled with its various applications, including those involving censored data and truncated data, helps to explain the large number of citations. A well-regarded book by McLachlan & Krishnan (1997) has been written about the algorithm. The name 'EM' was coined by Dempster, Laird & Rubin in this paper, but the method was apparently used in some form much earlier by a few researchers, including McKendrick (1926) and Hartley (1958), who introduced the procedure for calculating maximum likelihood estimates for the general case of count data.

(12) With 3,819 citations (32 per year),

Wilkinson, G. N. (1961) Statistical estimations in enzyme kinetics, *Biochemical Journal*, 80, pp. 324–336.

The author gave an account of the weighted linear and nonlinear regression methods applicable to general problems in enzyme kinetics. The Michaelis–Menten model, which is used frequently in enzyme kinetics, was used to illustrate aspects of nonlinear regression.

(13) With 3,672 citations (142 per year),

Nei, M. (1972) Genetic distance between populations, *The American Naturalist*, 106, pp. 283–292.

Nei (1972) proposed a measure of genetic distance based on the identity of genes between populations. The measure can be applied to any pair of organisms.

(14) With 3,511 citations (118 per year),

Dunnett, C. W. (1955) A multiple comparison procedure for comparing several treatments with a control, *Journal of the American Statistical Association*, 50, pp. 1096–1121.

A very interesting article about Professor Dunnett and how his work on multiple comparisons against a control evolved can be found at www.ssc.ca/main/about/history/dunnett_e.html. Additionally, Professor Dunnett has been kind enough to provide us with information relating to this paper and subsequent developments. His work with Bob Bechhofer and Milton Sobel on ranking and selection led to development of the multivariate-*t* distribution (Dunnett & Sobel, 1954, 1955), which fortuitously turned out to be the appropriate distribution for making multiple comparisons involving a control. Dunnett (1955) formulated the problem in terms of simultaneous confidence intervals, which was the same approach that John Tukey and Henry Scheffé had taken in their work.

Because of the great extent to which multiple comparison procedures are used by researchers outside the field of statistics, it is relevant to question the extent to which more recent papers that have defined the current state-of-the-art may have been overlooked. Indeed, Dunnett & Tamhane (1991, 1992) presented step-up and step-down methods (somewhat analogous to forward selection and backward elimination in linear regression) with these methods being superior to one-stage procedures in terms of maximizing power. Despite the superiority of these procedures, Dunnett & Tamhane (1991), for example, has only 31 citations.

(15) With 3,444 citations (280 per year),

Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19, pp. 716–723.

This is a paper in which Akaike proposed a criterion for estimating the dimensionality of a model using the criterion now known as Akaike's Information Criterion (AIC). This paper has over three times as many citations as Akaike (1973), which was included in Kotz & Johnson (1992a, pp. 599–624) as a breakthrough paper in statistics, with a discussion written by J. de Leeuw.

(16) With 2,837 citations (376 per year),

Liang, K.-Y. & Zeger, S. (1986) Longitudinal data analysis using generalized linear models, *Biometrika*, 73, pp. 13–22.

This paper was reprinted by Kotz & Johnson (1997, pp. 463–482) as a breakthrough paper in statistics with a discussion by P. J. Diggle. Liang & Zeger (1986) dealt with longitudinal studies in which the response measurement was a count. They derived a generalized estimating equations (GEE) methodology, which is now widely used.

(17) With 2,810 citations (22 per year),

Cutler, S. J. & Ederer, F. (1958) Maximum utilization of the life table method in analyzing survival, *Journal of Chronic Diseases*, 8, pp. 699–712.

The authors presented the rationale and computational details of the actuarial or life-table method for analysing data on patient survival. The method makes use of all survival information accumulated up to the closing date of a study. Cutler (1979) reported that he and Ederer were sharing a hotel room at a scientific meeting when the question leading to the paper came to him at 5 a.m. He promptly woke Ederer to discuss his idea. Cutler (1979) also stated that the paper did not represent a methodological breakthrough. The authors demonstrated that the life-table method could be used to extract the maximum amount of information from the data being collected in the newly organized cancer reporting system.

(18) With 2,764 citations (240 per year),

> Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 721–741.

This paper was included by Kotz & Johnson (1997, pp. 123–126) as a breakthrough paper with a discussion by P. J. Huber. Geman & Geman (1984) modified Markov chain Monte Carlo methods and applied them to Bayesian models for the computation of posterior probabilities.

(19) With 2,529 citations (120 per year),

> Box, G. E. P. & Cox, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, pp. 211–243 (discussion pp. 244–252).

DeGroot (1987) provided some interesting background on this paper from an interview with Professor Box. Box recounted, for example, that he and Cox were on a committee of the Royal Statistical Society and several people suggested that they collaborate. Their motivation and the idea of the paper sprung, to some extent, from the similarities of their family names.

Box & Cox (1964) presented a very useful family of power transformations that have typically been used to transform the dependent variable in a regression model so as to try to meet the assumptions of homoscedasticity and normality of the error terms. The right side of the model can then be transformed in the same manner so as to retrieve the quality of the fit before the dependent variable was transformed.

(20) With 2,512 citations (76 per year),

> Mantel, N. (1963) Chi-square tests with one degree of freedom: extensions of the Mantel–Haenszel procedure, *Journal of the American Statistical Association*, 58, pp. 690–700.

The author extended the methods in Mantel & Haenszel (1959), Number 9 on our list, in two ways, as it was recognized that the methods are not limited to retrospective studies and the number of levels of the study factor of interest was allowed to be greater than two.

(21) With 2,456 citations (46 per year),

> Dunnett, C. W. (1964) New tables for multiple comparisons with a control, *Biometrics*, 20, pp. 482–491.

In this paper, exact critical values are given for the method of Dunnett (1955), Number 14 on our list, when two-sided comparisons are made with a control.

(22) With 2,302 citations (42 per year),

> Kramer, C. Y. (1956) Extension of multiple range tests to group means with unequal numbers of replications, *Biometrics*, 12, pp. 307–310.

Kramer (1956) proposed an approximate method for extending multiple range tests to cases for which the sample sizes are unequal. Kramer's work was strongly related to the methodology proposed by John Tukey in 1953, whose work was not published. Nevertheless, because of the close connection, Kramer's method for the unbalanced case is known as the Tukey–Kramer procedure. (See Benjamini & Braun, 2002, for a discussion of this issue.)

(23) With 2,248 citations (72 per year),

Fisher, R. A. (1953) Dispersion on a sphere, *Proceedings of the Royal Society of London, Series A*, 217, pp. 295–305.

Fisher (1953) presented a theory of errors that is believed to be appropriate for measurements on a sphere and derived a test of significance that was stated as being 'the analogue of "Student's test" in the Gaussian theory of errors'. The paper can be viewed online at http://www.library.adelaide.edu.au/digitised/fisher/249.pdf). According to Garfield (1977), this paper had only 277 citations between 1961 and 1975, but was Fisher's most frequently cited paper during that time period.

(24) With 2,219 citations (240 per year),

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, 6, pp. 461–464.

Schwartz's Bayesian Information Criterion (BIC), introduced in this paper, is a criterion for model selection that is often mentioned with Akaike's AIC criterion.

(25) With 2,014 citations (382 per year),

Weir, B. S. & Cockerham, C. C. (1984) Estimating *F*-statistics for the analysis of population structure, *Evolution*, 38(6), pp. 1358–1370.

As Professor Weir informed us, the number of citations of this paper has risen every year since its publication as different groups of researchers have become interested in genetic population structure. These groups include ecologists, conservationists and, interestingly enough, forensic scientists.

Comments on the Top 25 List

The most-cited statistical papers fare well when compared to the most-cited papers in science. Garfield (1990) ranked the 100 most-cited papers in the 1945–1988 *Science Citation Index.* Duncan (1972), Litchfield & Wilcoxon (1949), Kaplan & Meier (1958), Marquardt (1963), and Cox (1972) ranked Numbers 24, 29, 55, 92 and 94, respectively. Kaplan & Meier (1958) and Cox (1972) had 'only' 4,756 and 3,392 citations, respectively, in Garfield's study.

All papers on our list were published prior to 1987. A dotplot of the publication years of the 25 papers on our list is shown in Figure 1.

There is no question that the field of a paper is related to the number of citations. This is evident from the number of papers in biostatistics on our list. Similarly, of the 27 'highly cited authors in mathematics and statistics' listed by Kruse (2002) in *AmStat News*, the

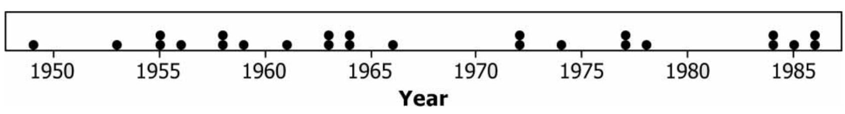

Figure 1. Dotplot of publication year for the 25 most cited papers

four who would rank the highest in terms of the number of citations per paper (for papers published during 1991–2001 and also cited during that period) are all biostatisticians.

If we traced the development of statistical methodology and theory (as Efron, 2001, did, concentrating on 1950–1980), we would certainly expect that there would be a strong correlation between the influence of a paper and its number of citations. The most influential papers in statistics have large citation counts, but only a few have enough citations to make our list. Efron (2001) listed non-parametric and robust methods first in impact, followed by the Kaplan–Meier method and Cox's method, with logistic regression and generalized linear models (GLM) mentioned third, while stating that logistic regression has had a huge effect on biostatistics. It has been known for some time that Kaplan & Meier (1958) and Cox (1972) were the two most-cited papers in statistics. See, for example, Stigler (1994).

Only a few of the most influential papers on the field of statistics are included on our list. Only five are included in Kotz & Johnson (1992a, 1992b, 1997) as representing 'breakthrough papers in statistics'. Four of our most cited papers, Duncan (1955), Kramer (1956), and Dunnett (1955, 1964) are on the topic of multiple comparisons. Multiple comparison methods are widely used in statistical practice, but without a major influence on the field of statistics itself. Tukey (1991), for example, downplayed the importance of the method of Duncan (1955) calling it a 'distraction'. To more effectively measure impact with respect to the field of statistics, it would be better to count only citations that appeared in statistical journals.

It is interesting to note that Nathan Mantel was author or co-author on four of our 25 papers. For more information on his contributions to statistics, the reader is referred to his obituary in *AmStat News* (July, 2002, pp. 35–36) or to http://members.aol.com/savilon/nmantel.html. He did much of his work at the National Cancer Institute, retiring from there in 1974. He was a very active consultant and undoubtedly many of his major research contributions had their origins in his consulting work. Sir D. R. Cox was author or co-author of three of the papers on our list. See Reid (1994) for information on his background.

Most-cited Papers Published Since 1993

For a perspective on the changing emphases of statistics over time, we also studied the most-cited statistical papers published in 1993 or later. Our list of the fifteen most-cited papers was obtained by first obtaining citation counts for all papers written during this time by the ten most-cited statisticians (for citations received for papers written and citations received between 1 January 1993 and 30 June 2003) listed in the November 2003 *AmStat News* (also see http://www.in-cites.com/top/2003/third03-math.html). These were the following: David L. Donoho (1,354 citations), Iain M. Johnstone (1,203 citations), Adrian E. Raftery (1,117 citations), Adrian F. M. Smith (866 citations), Peter Hall (827 citations), Donald B. Rubin (792 citations), Jianqing Fan (768 citations), Gareth O. Roberts (725 citations), Robert E. Kass (723 citations), and Siddhartha Chib (708 citations). We then checked the citation counts for all papers published in 1993 or later in the following statistical journals given by ISI *Journal Citation Reports* in the Statistics and Probability category as having the most citations in 2002 (number of citations is in parentheses): *Journal of the American Statistical Association* (11,318), *Econometrica* (9,458), *Biometrics* (7,469), *Biometrika* (6,742), *Annals of Statistics* (5,566), *Statistics in Medicine* (4,755), *Journal of the Royal Statistical Society, Series B* (4,755), and *Technometrics* (2,514). (Note that *Fuzzy Sets and Systems* with 3,626 citations was not included in our search.) It is possible that some papers were overlooked.

The following is our list:

1. Breslow, N. E. & Clayton, D. G. (1993) Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, 88, pp. 9–25. (558 citations)
2. Tierney, L. (1994) Markov-chains for exploring posterior distributions, *Annals of Statistics*, 22, pp. 1701–1728. (541 citations)
3. Kass, R. E. & Raftery, A. E. (1995) Bayes Factors, *Journal of the American Statistical Association*, 90, pp. 773–795. (533 citations)
4. Donoho, D. L. & Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81, pp. 425–455. (480 citations)
5. Smith, A. F. M. & Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov-chain Monte-Carlo methods, *Journal of the Royal Statistical Society, Series B*, 55, pp. 3–23. (444 citations)
6. Green, P. J. (1995) Reversible jump Markov-chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82, pp. 711–732. (479 citations)
7. Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B*, 57, pp. 289–300. (294 citations)
8. Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995) Wavelet shrinkage – asymptopia, *Journal of the Royal Statistical Society, Series B*, 57, pp. 301–337. (293 citations)
9. Donoho, D. L. (1995) De-noising by soft thresholding, *IEEE Transactions on Information Theory*, 41, pp. 613–627. (292 citations)
10. Grambsch, P. M. & Therneau, T. M. (1994) Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika*, 81, pp. 515–526. (261 citations)
11. Donoho, D. L. & Johnstone, I. M. (1995) Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, 90, pp. 1200–1224. (257 citations)
12. Bound, J., Jaeger, D. A. & Baker, R. M. (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association*, 90, pp. 443–450. (252 citations)
13. Albert, J. H. & Chib, S. (1993) Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, 88, pp. 669–679. (246 citations)
14. Stock, J. H. & Watson, M. W. (1993) A simple estimator of cointegrating vectors in higher-order integrated systems, *Econometrica*, 61, pp. 783–820. (244 citations)
15. Chib, S. & Greenberg, E. (1995) Understanding the Metropolis–Hastings algorithm, *The American Statistician*, 49, pp. 327–335. (240 citations)

The most cited papers presented here tend to be on topics related to Bayesian methods and wavelets, although the topics of multiple testing and proportional hazards modelling are represented. It is interesting to note that it often takes quite a few years for the number of citations of a paper to reach its maximum rate. A number of the 25 overall most-cited papers are cited now at much higher rates than the most-cited papers of the last decade.

Conclusions

We find the study of citation counts to be very interesting. It is surprising that relatively little research has been done on citation counts, rates and patterns in the field of statistics.

Garfield (1979: 16) described early work in this area, including the Citation Index for Statistics and Probability, 'a cumulative one-time effort that covers the journal literature of the field from its inception, early in the twentieth century, through 1966'. This was compiled by John Tukey and published in 1973 as part of the 'Information Access Series' of R&D Press. It provided comprehensive coverage of 40 statistics journals and selective coverage of an additional 100 journals.

In our view it would be very interesting to examine a list of the most-cited papers in each of the top statistics journals (see Campbell & Julious, 1994) or in different application areas of statistics. Also, it would be useful to identify papers projected to enter the top 25 most-cited statistical papers and, more generally, 'hot papers' that have attained unusually high citation rates shortly after publication. For more on this latter topic, the reader is referred to Garfield (2000).

Acknowledgements

The authors gratefully acknowledge input from I. Hoeschele, C. W. Dunnett, J. Felsenstein, C. King, B. S. Weir, J. M. Bland, E. Garfield, I. J. Good and D. G. Altman.

References

Akaike, H. (1973) Information theory and the maximum likelihood principle, in: B. N. Petrov & F. Csàki (Eds) *Second International Symposium on Information Theory* (Budapest: Akademiai Kiàdo).

Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19, pp. 716–723.

Altman, D. G. & Bland, J.M. (1983) Measurement in medicine: the analysis of method comparison studies, *The Statistician*, 32, pp. 307–317.

Altman, D. G. & Goodman, S. N. (1994) Transfer of technology from statistical journals to the biomedical literature: past trends and future predictions, *Journal of the American Medical Association*, 272, pp. 129–132. Available at http://www.ama-assn.org/public/peer/7_13_94/pv3108x.htm.

Austin, A. (1993) The reliability of citation counts in judgments on promotion, tenure, and status, *Arizona Law Review*, pp. 829–829.

Benjamini, Y. & Braun, H. (2002) John Tukey's contributions to multiple comparisons, *Annals of Statistics*, 30, pp. 1576–1594.

Bland, J. M. & Altman, D. G. (1986) Statistical methods for assessing agreement between two clinical measurements, *Lancet*, 1 (8476), pp. 307–310.

Bland, J. M. & Altman, D. G. (1992) Comparing methods of clinical measurement. [Citation Classic], *Current Contents*, 20, p. 8.

Bland, J. M. & Altman, D. G. (1995) Comparing two methods of clinical measurement: a personal history, *International Journal of Epidemiology*, 24 (Suppl. 1), pp. S7–S14.

Box, G. E. P. & Cox, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, pp. 211–243 (discussion pp. 244–252).

Campbell, M. J. & Julious, S. A. (1994) *Statistics in Medicine*: citations of papers in the first ten years, *Statistics in Medicine*, 13, pp. 3–10.

Cooley, J. W. & Tukey, J. W. (1965) An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation*, 19, pp. 297–301.

Cox, D. R. (1972) Regression models and life tables, *Journal of the Royal Statistical Society, Series B*, 34, pp. 187–220.

Cox, D. R. (1986) This week's citation classic, *Current Contents*, CC/Number 42, p. 16.

Cronin, B. (1984) *The Citation Process, the Role and Significance of Citations in Scientific Communication* (Taylor Graham).

Cutler, S. J. (1979) This week's citation classic, *Current Contents*, Number 16, p. 356.

Cutler, S. J. & Ederer, F. (1958) Maximum utilization of the life table method in analyzing survival, *Journal of Chronic Diseases*, 8, pp. 699–712.

DeGroot, M. H. (1987) A conversation with George Box, *Statistical Science*, 2, pp. 239–258.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (C/R: pp. 22–37), *Journal of the Royal Statistical Society, Series B*, 39, pp. 1–22.

Donoho, D. (2002) How to be a widely cited author in the mathematical sciences, *in-cites*. Available at http://www.in-cites.com/scientists/DrDavidDonoho.html.

Duncan, D. B. (1955) Multiple range and multiple F–tests, *Biometrics*, 11, pp. 1–42.

Duncan, D. B. (1975) t tests and intervals for comparisons suggested by the data, *Biometrics*, 31, pp. 339–359.

Duncan. D. B. (1977) This week's citation classic, *Current Contents*, Number 4, p. 10.

Dunnett, C. W. (1955) A multiple comparison procedure for comparing several treatments with a control, *Journal of the American Statistical Association*, 50, 1096–1121.

Dunnett, C. W. (1964) New tables for multiple comparisons with a control, *Biometrics*, 20, pp. 482–491.

Dunnett, C.W. & Sobel, M. (1954) A bivariate generalization of Student's t–distribution with tables for certain special cases, *Biometrika*, 41, pp. 153–169.

Dunnett, C.W. & Sobel, M. (1955) Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t–distribution, *Biometrika*, 42, pp. 258–260.

Dunnett, C. W. & Tamhane, A. C. (1991) Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts, *Statistics in Medicine*, 10, pp. 939–947.

Dunnett, C. W. & Tamhane, A. C. (1992) A step-up multiple test procedure, *Journal of the American Statistical Association*, 87, pp. 162–170.

Edge, D. (1979) Quantitative measures of communication in science: a critical review, *History of Science*, 17, p. 102.

Efron, B. (1979) Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, 7, pp. 1–26.

Efron, B. (2001) Statistics in the 20th century, and the 21st, Invited Lecture at the Statistiche Woche 2001 in Vienna, Austria, 16–19 October 2001. Published in Fetschrift 50 Jahre Österreichische Statistische Gesellschaft, Seiten 7–19. Available at http://www.statistik.tuwien.ac.at/oezstat/festschr02/papers/efron.pdf.

Engle, R. F. & Granger, C. W. J. (1987) Co-integration and error correction: representation, estimation and testing, *Econometrica*, 55, pp. 251–276.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, 39, pp. 783–791.

Fisher, R. A. (1953) Dispersion on a sphere, *Proceedings of the Royal Society of London, Series A*, 217, pp. 295–305.

Garfield, E. (1977) The 250 most-cited primary authors, 1961–1975. Part III. Each author's most-cited publication, *Current Comments: Essays of an Information Scientist*, 3, pp. 348–363. Available at http://www.garfield.library.upenn.edu/essays/v3p348y1977–78.pdf.

Garfield, E. (1979) *Citation Indexing – its Theory and Application in Science, Technology, and Humanities* (New York: Wiley).

Garfield, E. (1989) Delayed recognition in scientific discovery: citation frequency analysis aids the search for case histories, in: *Essays of an Information Scientist: Creativity, Delayed Recognition, and other Essays*, 12, pp. 154–160, and in *Current Contents*, 23, pp. 3–9, June 5, 1989.

Garfield, E. (1990) The most-cited papers of all time, SCI 1945–1988. Part 1A. The SCI top 100 – will the Lowry method ever be obliterated?, *Current Comments*, 7, pp. 3–14. Available at http://www.garfield.library.upenn.edu/essays/v13p045y1990.pdf.

Garfield, E. (1998) The use of journal impact factors and citation analysis for evaluation of science, Presented at Cell Separation, Hematology, and Journal Citation Analysis, Mini–Symposium in tribute to Arne Bøyum, Rikshospitalet, Oslo.

Garfield, E. (2000) The evolution of 'Hot Papers', *The Scientist*, 14. Available at www.garfield.library.upenn.edu/commentaries/tsv14(14)p04y20000710.pdf.

Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 721–741.

Gilbert, G. N. (1977) Referencing as persuasion, *Social Studies of Science* 7, pp. 113–122.

Hahn, G. J. (1995) A conversation with Donald Marquardt, *Statistical Science*, 10, pp. 377–393.

Hamilton, D. P. (1990) Publishing by and for ? the numbers, *Science*, 250, pp. 1331–1332.

Hamilton, D. P. (1991) Research papers: who's uncited now?, *Science*, 251, p. 25.

Hartley, H. O. (1958) Maximum likelihood estimation from incomplete data. *Biometrics*, 14, pp. 174–194.

Holmes, S. (2003) Bootstrapping phylogenetic trees: theory and methods, *Statistical Science*, 18, pp. 241–255.

Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences*, 79, pp. 2554–2558.

Kaplan, E. L. (1983) This week's citation classic, *Current Contents*, CC/Number 24, p. 14.

Kaplan, E. L. & Meier, P. (1958) Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, 53, pp. 457–481.

Kotz, S. & Johnson, N. L. (1992a) *Breakthroughs in Statistics Volume I* (New York: Springer-Verlag).

Kotz, S. & Johnson, N. L. (1992b) *Breakthroughs in Statistics Volume II* (New York: Springer-Verlag).

Kotz, S. & Johnson, N. L. (1997) *Breakthroughs in Statistics Volume III* (New York: Springer-Verlag).

Kramer, C. Y. (1956) Extension of multiple range tests to group means with unequal numbers of replications, *Biometrics*, 12, pp. 307–310.

Kruse, M. (2002) Members, *JASA* top math citations, *AmStat News*, Issue 305, November, pp. 6–7.

Laviolette, M., Seaman Jr., J. W., Barrett, J. D. & Woodall, W. H. (1995) A probabilistic and statistical view of fuzzy methods, (with discussion) *Technometrics*, 37, pp. 249–292.

Liang, K.-Y. & Zeger, S. (1986) Longitudinal data analysis using generalized linear models, *Biometrika*, 73, pp. 13–22.

Litchfield, J. T. (1977) This week's citation classic, *Current Contents*, Number 7, p. 8.

Litchfield, J. T. & Wilcoxon, F. A. (1949) A simplified method of evaluating dose-effect experiments, *Journal of Pharmacological and Experimental Therapeutics*, 96, pp. 99–113.

Marquardt, D. W. (1979) This week's citation classic, *Current Contents*, CC/Number 27, p. 20.

Mantel, N. (1963) Chi-square tests with one degree of freedom: extensions of the Mantel–Haenszel procedure, *Journal of the American Statistical Association*, 58, pp. 690–700.

Mantel, N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports*, 50, pp. 163–170.

Mantel, N. & Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute*, 22, pp. 719–748.

McKendrick, A. G. (1926) Applications of mathematics to medical problems, *Proceedings of the Edinburgh Mathematics Society*, 44, pp. 98–130.

McLachlan, G. J. & Krishnan, T. (1997) *The EM Algorithm and Extensions* (New York: Wiley).

Meier, P. (2001) The price of Kaplan–Meier, paper presented at the 9th Merck–Temple Conference on Research Topics in Pharmaceutical Statistics. Available at http://www.sbm.temple.edu/~biostat/conf9.html.

Meier, P., Karrison, T., Chappell, R. & Xie, H. (2004) The price of Kaplan–Meier, *Journal of the American Statistical Association*, 99, pp. 890–896.

Miller Jr., R. G. (1983) What price Kaplan–Meier?, *Biometrics*, 39, pp. 1077–1081.

Nei, M. (1972) Genetic distance between populations, *The American Naturalist*, 106, pp. 283–292.

Nelder, J. A. & Mead, R. (1965) A simplex method for function minimization, *Computer Journal*, 8, pp. 308–313.

Pendlebury, D. A. (1991) Science, citation, and funding, *Science*, 251, pp. 1410–1411.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. & Smith, K. G. (1977) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part II. Analysis and examples, *British Journal of Cancer*, 35, pp. 1–39.

Reed, L. J. & Muench, H. (1938) A simple method of estimating 50 percent endpoints, *American Journal of Hygiene*, 27, pp. 493–497.

Reid, N. (1994) A conversation with Sir David Cox, *Statistical Science*, 9, pp. 439–455.

ScienceWatch (1999) Citations reveal concentrated influence: some fields have it, but what does it mean?, *ScienceWatch*, 10(1). Available at http://www.sciencewatch.com/jan-feb99/sw_jan–feb99_page2.htm.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, 6, pp. 461–464.

Stigler, S. M. (1994) Citation patterns in the journals of statistics and probability, *Statistical Science*, 9, pp. 94–108.

Straf, M. L. (2003) Statistics: the next generation, *Journal of the American Statistical Association*, 98, pp. 1–6.

Theoharakis, V. & Skordia, M. (2003) How do statisticians perceive statistics journals?, *The American Statistician*, 57, pp. 115–123.

Tukey, J. W. (1991) The philosophy of multiple comparisons, *Statistical Science*, 6, pp. 100–116.

Weir, B. S. & Cockerham, C. C. (1984) Estimating *F*–Statistics for the analysis of population structure, *Evolution*, 38(6), pp. 1358–1370.

Wilkinson, G. N. (1961) Statistical estimations in enzyme kinetics, *Biochemical Journal*, 80, pp. 324–336.

Wright, S. (1931) Evolution in Mendelian populations, *Genetics*, 16, pp. 97–159.

Zadeh, L. A. (1965) Fuzzy sets, *Information and Control*, 8, pp. 338–353.