Statistical Research for Science and Industry - CSIRO

Cameron, Murray
CSIRO Mathematical & Information Sciences
Locked Bag 17
North Ryde NSW 1670 Australia
Murray.Cameron@csiro.au

Summary

I outline the history and evolving role of statistics in CSIRO, leaning heavily on earlier papers by Speed (1988) and Field, Speed, Speed & Williams (1988). In addition, I consider the research output from statisticians in CSIRO which has been most influential and hypothesise a strategy for maximising the scientific impact of statistical work.

1. Statistics and the beginnings of CSIRO

The Council for Scientific and Industrial Research (CSIR) was created in 1926. CSIR evolved into the Commonwealth Scientific and Industrial Research Organisation, known universally as CSIRO, in 1949.

CSIR's early focus was agricultural and, at least in part influenced by the experience of Rothamsted and the contributions of R.A. Fisher. The first statistician (Betty Allan) was appointed to CSIR in September 1930. Field, Speed, Speed & Williams (1988) give more details of Allan and two other early statisticians (Mildred Barnard and Helen Newton Turner).

Until March 1954, statistical work occurred within the "Biometrics Section". Research in CSIRO has generally been organised through (semi)autonomous business units known as Divisions. Since 1954, most statistical work has been centralised within a Division known variously as Mathematical Statistics (1954-1974), Mathematics & Statistics (1974-1996), or Mathematical & Information Sciences (1996-). Between 1987 and 1996 there were also 3 "Biometrics Units" which had an internal focus, in CSIRO while the Division had a largely external focus.

Speed (1988) gives a good overview of statistics in CSIRO from 1930 to 1987. His paper ends with a note added in proof: "A sequence of decisions beginning in June 1987 has led to a complete rejection by the Board of CSIRO of the role for the Division whose nature and evolution has been described in this paper." In fact, some of those decisions were later modified so that, although 1987 was a turning point for statistics in CSIRO, leading to a split of the Division and the creation of the Biometrics Units, it was not the end of statistics as a significant force and contributor to CSIRO's science.

CSIRO's research emphasis has broadened over time. Statistics work has broadened to match this although, of course, statistics does not contribute equally to all research areas. The number of statisticians in CSIRO grew, being 12 in 1950, 25 in 1960, 49 in 1970 and about 60 in 1975. Today the Division of Mathematical & Information Sciences has over 100 research staff, though the breadth of the Division has grown to include applied mathematics, operations research and image analysis so that the number of statisticians is about 65.

2. The Role of Statistics in CSIRO

Generally the focus of the Division has been applications from other Divisions and industry, and statistical research driven by those applications. However, from 1944 until at least the early

1970's there was also a small 'core' research group in Adelaide. The applications focus in many ways follows the Rothamsted model which was also followed in other similar organisations.

Since 1987 the Division has not been constrained by any expectation that it will provide (free) support to other parts of CSIRO, though the Biometrics Units were. It has been constrained by an expectation that it will generate at least 30% of its total expenditure from external sources. This has meant that (a) research has been either aligned with the current needs of applications or directed towards perceived future applications (b) it has often been financially unattractive to work with other parts of CSIRO (c) publication of research has been much less important.

3. Science Output

Any research group will assess its performance through measures such as the numbers of published papers and patents, numbers of citations, numbers of research students and post-doctoral fellows trained, and the adoption and impact of research results. For statistics in a broader research organisation another relevant measure might be the number of cross-disciplinary research papers and their impact.

I shall focus here on research publications. The published research of a statistics research group can be subdivided into 3 parts – (1) *core* research (general statistical methodology appearing in a statistical research journal), (2) *subject specific methodology* (statistical methods developed or adapted for a specific discipline) and (3) *applications* (research in another discipline where expert statistical analysis is an important component).

3.1 Monographs

CSIRO statisticians have authored or co-authored a number of research monographs which have generally been inspired by their CSIRO experiences in work with those from other disciplines. For example, Williams (1959) was motivated by his experiences in working with the CSIRO Forest Products Division. There have been 13 monographs, including 4 co-authored by Venables on aspects of the S (and R) statistical programming language. These books average in excess of 200 citations each.

3.2 Journal Publications

Data on journal publications are influenced by the growth in the number of journals, changes in publication and authorship practices, changes in the numbers of research staff and their focus and also changes in the quality of the data on publication and citation. Publication and citation are themselves imperfect surrogates for productivity, influence and impact.

In terms of productivity, Genest (1997) provides a comparison of publication rates by institutions in 16 statistics journals which focus on methodology ('Core' journals) and shows that among institutions, CSIRO is 22nd (Genest's Table 6), though among non-academic institutions CSIRO is 2nd to Bell Laboratories (Genest's Table 7).

We have undertaken an analysis of the citations of publications by CSIRO statisticians. We do not claim that the analysis is complete, but are confident that the main features of the analysis would not change. The analysis arbitrarily considers all publications that have been cited more than 50 times. The role and organisation of statistics in CSIRO changed significantly in 1974 and again in 1987 and so I have separated the data to cover, roughly, those 3 periods.

Table: Distribution of most cited CSIRO statistical publications

	All published Monographs	No of papers with more than 50 citations		
		Core Journals	Subject Journals	Applications
Before 1974	1	13	4	2
1974-1987	2	7	8	5
1988-2000	10	3	5	4
Total	**13**	**23**	**17**	**11**

3.2.1 Core statistics journals

Before 1973, the highly cited publications in core journals were in the area of distribution theory (Cornish-Fisher polynomials) and multivariate distribution theory. This was a time when there was a research group focused on 'methodology' and the highly cited papers were in the Annals of Mathematical Statistics.

In the period 1974-1987, reflecting the broader research agenda that was followed, the highly cited research in core journals had no clear focus and covered growth curves, causal models, robustness, subset selection in regression and classification of markov chains. In the period from 1988 to 2000 the amount of high impact core research was reduced - reflecting changes in direction for statistics in CSIRO and an increased focus on industry impact (and possibly the time it takes for a statistical paper to achieve a large number of citations).

3.2.2 Subject specific methodology

Before 1974 there were a small number of highly cited papers in subject-specific journals, most notably the paper by GN Wilkinson (1961) - cited more than 3800 times. The paper gave statistical methods for estimating the parameters in the Michaelis-Menten equation used routinely in enzyme kinetic studies.

Two other papers in subject specific journals are notable in that period. In both, G.A. McIntyre was the statistical author. The first paper (McIntyre, 1952) introduced the method known as ranked set sampling (the subject of another session at this meeting) and the second (McIntyre, Brooks, Compston and Turek, 1966) provided the basic methodology for analysing Rb-Sr isotope ratios for the aging of rocks. In essence, this was a weighted least squares method for estimating linear relationships between two variables when there are errors in both variables.

In the period from 1974-1987, in subject specific methodology, the major contribution was from a sequence of papers on the use of measurements of fission track properties in apatite for estimating the thermal history of rocks for petroleum exploration. A sequence of 8 papers, all involving Geoff Laslett, in the period from 1982 until 1993 averaged nearly 150 citations each. The work on fission tracks supported the formation of the company Geotrack International (see www.geotrack.com.au).

3.2.3 Applications publications

It is perhaps surprising that relatively few of the applications papers in which CSIRO statisticians have made substantial contributions have been highly cited. Since CSIRO researchers are highly cited in Agriculture and Environmental Science we hypothesise that (a) the work of statisticians working with researchers has not been effectively prioritised to contribute to the most important research and that (b) the role for statisticians which makes the biggest contribution to

research is that of developing and adapting modern statistical methodology for new measurement technologies.

4. Conclusion

To maximise the scientific impact of their work, statisticians should aim, where possible, to collaborate on projects that are the most important, particularly in areas where new measurements technologies are being used. This points to fields such as environmental science (where new, cheap data collection methods are being developed), and biotechnology.

5. Acknowledgements

I thank Linda Whitford and Frank de Hoog who initiated the collection of the citation data used here.

REFERENCES

Brillinger, D.R. (1977) Division of Mathematics & Statistics Newsletter No 38, December.

Field, J.B.F., Speed, F.E., Speed, T.P. and Williams, J.M. (1988) Biometrics in the CSIR:1930-1940. Australian J of Statistics **30B**, 54-76.

Genest, C. (1997) Statistics on statistics: measuring research productivity by journal publications between 1985 and 1995. The Canadian J of Statistics **25**, No 4, 427-443.

McIntyre, G. (1952) A method for unbiased selective sampling using ranked sets. Australian J Agric Res **3**, No 4, 385-90.

McIntyre, G., Brooks, G.C., Compston, W. And Turek, A. (1966) The statistical assessment of Rb-Sr isochrons. J Geophys Res **71**, No 22, 5429-68.

Speed, T.P. (1988) The role of statisticians in CSIRO: Past, Present and Future. Australian J of Statistics **30**, 15-34.

Wilkinson, G.N. (1961) Statistical estimation in enzyme kinetics. *Biochem. J* **80**, 324-32.

Williams, E.J. (1959) Regression Analysis. New York: Wiley

RESUMÉ

Je donne un aperçu de l'histoire et de l'évolution de la statistique au CSIRO, en m'inspirant fortement des écrits de Speed (1988) et de Field, Speed, Speed & Williams (1988). Un examen des réalisations les plus marquantes des statisticiens du CSIRO m'amène à proposer une stratégie visant à maximiser l'impact scientifique des travaux statistiques.