The Open University

Statistics Department

1997 Report

1. Membership

Professor D.J. Hand (Head of Department)
Professor J.C. Gower
Dr N. Adams
Dr F. Daly (to November)
Dr S.E. Hitchcock (to May)
Dr G.E. Iossif
Dr M.C. Jones
Dr H.G. Li (to August)
Dr A.D. Lunn
Dr K.J. McConway
Dr R. Ngouenet
Dr P.C. Taylor
Dr S.K. Vines
Mrs J.E. Williams
Dr K. Yu (to December)

Visitors

Professor B. Johnston (University of Manitoba)
Professor C.E. Lunneborg (University of Washington)

Full-time research students

Mr M.G. Kelly
Miss M. Kanaan
Mrs J. Warwick (from October)

Part-time research students

Mr G. Blunt
Mr S. Cresswell
Mr P.M. Fayers
Mr A. Garrett
Ms E.K. Kemsley
Dr M. King
Mr M. Lunt
Mr D. F. Signorini
Mr K. Yu (to March)

Support staff

Ms E. Ostrowski, Departmental Secretary

2. Introduction This year was another very productive and successful one for the Department. We hold a grade 4 rating from the 1996 Research Assessment Exercise, and research activity continued strongly in 1997, particularly in the areas of statistics in finance, classification methods, smoothing and biplots. During 1997, Department members published 1 book and 27 papers, book chapters etc. In addition, our teaching continued to flourish, with the completion of production of our new third level course in linear modelling. This brings back to four the number of core statistics courses presented by the Department.Unfortunately, two longstanding members of the Department left us this year. In May, Shirley Hitchcock retired and in November, Fergus Daly moved to Scotland (taking a post at the University of Dundee). Shirley and Fergus are both thanked for their sterling service to the Department over the years, and we wish them both well in their future endeavours.We are pleased to be able to signal the promotion of John Gower to Professor during 1997.

Information on the Department is available through our Web pages at http://Statistics.open.ac.uk/index.html

3. Teaching

Presentation The Department presented three courses (30 CATS points each), MDST242 *Statistics in Society*, M246 *Elements of Statistics* and M343 *Applications of Probability.* This was the first presentation year of the revised version of M343. Altogether the Departmentís courses attracted a total of over 1,500 students. Each of the Faculty of Mathematics and Computingís first level courses, MU120 *Open Mathematics* and MST121 *Using Mathematics*, contains substantial amounts of statistical material (roughly one quarter of each) with which Gillian Iossif is heavily involved. MST121 was presented for the first time in 1997.Both MST121 and M246 are also presented to students of the Singapore Institute of Management. Jane Williams, Staff Tutor in the Department, has been to Singapore to train and advise local tutors of these courses.

Production Production of the new course M346 *Linear Statistical Modelling*, which replaces M345 *Statistical Methods*, last presented in 1996, was completed and the course will be presented for the first time in 1998. The production of M346 met the highest standards of sticking to schedule, a fact of which we can be proud. M346 takes an even more applied line than its predecessor, following on from M246 *Elements of Statistics*, and will involve the students in heavy use of *Genstat for Windows* on a home computer. As well as linear models *per se*, the new course covers generalized linear models and some aspects of experimental design.

Other Kevin McConway also contributed to the presentation of U205 *Health and Disease* and D300 *Professional Judgement and Decision Making.*

4. PhDs awarded Keming Yu was awarded a PhD for his thesis entitled *Smooth Conditional Quantile Estimation* (supervisor: M.C. Jones).

5. Research interests The Department is the home of two research groups, in *Statistics in Finance* (Hand, Adams, Kelly, Li, McConway) and *Classification Methods* (Hand, Adams, Gower, Ngouenet, Taylor, Yu). In addition, Dr McConway is a member of the Universitys *Ecology and Conservation* research group.Members of the Department are involved in a variety of research projects, including:

Statistical computing (Daly)

Applied multivariate analysis (Gower)

Multidimensional scaling, Procrustes methods (Gower)

Biplots (Gower, Hand, Ngouenet)

Measurement theory and metadata (Hand)

Assessing classification rules (Hand, Adams)

Credit risk scoring (Hand, Adams, Kelly, Li, McConway)

Discrimination and classification (Hand, Adams, Taylor, Yu, Webb (DERA))

Deconstructing statistical questions (Hand, Lunneborg)

Queues with priorities (Hitchcock)

Stochastic birth-death processes for individuals in clusters (Hitchcock)

Distribution theory (Jones)

The local dependence function (Jones)

L-moments (Jones)

Minimum distance estimation (Jones, Basu (Indian Statistical Institute), Harris (University of Northern Arizona), Hjort (University of Oslo))

Semiparametric density and probability estimation (Jones, Faddy (University of Queensland), Hjort (University of Oslo))

Density estimation with biased data (Jones, Lloyd (University of Hong Kong))

Conditional quantile estimation (Jones, Yu)

Bayesian methods in meta-analysis (Lunn)

Medical statistics, clinical trials (categorical data, repeated measures, generalised estimating equations, Bayesian methods) (Lunn)

Reliability of software (Lunn)

Performance-based handicapping systems in sport (Lunn, Cresswell)

Robust estimation with circular and spherical data (Lunn, Fisher (CSIRO))

Analysis of spectroscopic data (discrimination, multivariate methods of density reduction) (Lunn, Kemsley)

Bayesian statistics (Lunn, McConway, Vines)

Statistical methods in plant ecology (McConway)

Comparative methods using phylogenies (McConway, Hitchcock, Silvertown (Biology, OU))

The determinants of species richness in flowering plants (McConway, Silvertown (Biology, OU), Chase (Royal Botanical Gardens, Kew))

Methods for comparing cladogenesis rates in phylogenies (McConway, Silvertown (Biology, OU), Dodd (Biology, OU))

Ecological and genetic correlates of long-term population trends in plants (McConway, Silvertown (Biology, OU), Miller (Rothamsted Experimental Station), Macnair (University of Exeter))

Nonparametric binary regression (Signorini, Jones)

Image analysis (Taylor, Denham (University of Reading), Pursey (University of Reading))

Spatial statistics (Taylor, Kanaan, Mugglestone (Rothamsted Experimental Station), Alexander (Rothamsted Experimental Station))

Bayesian analysis of case-control studies (Vines)

Parameterisation of random effects models (Vines, Gilks (MRC Biostatistics Unit))

6. Publications

6.1 Books

Hand, D.J. (1997) *Construction and Assessment of Classification Rules*, Wiley.

6.2 Publications in refereed academic journals

Al Jowder, O., Kemsley, E.K. and Wilson, R.H. (1997) Mid-infrared spectroscopy and authenticity problems in selected meats: a feasibility study. *Food Chemistry*, 59, 195201.

Defernez, M., and Kemsley, E.K. (1997) The use and misuse of chemometrics for treating classification problems. *Trends in Analytical Chemistry*, 16, 21621.

Defernez, M., Kemsley, E.K. and Wilson, R.H. (1997) Mid-infrared as a new tool for detecting adulteration in fruit products. *Mikrochimica Acta*, S14, 1812.

Delaney, G., Gebski, V., Lunn, A.D., Lunn, M., Rus, M., Manderson, C. and Langlands, A.O. (1997) Basic treatment equivalent (BTE): a new measure of linear accelerator workload. *Clinical Oncology*, 9, 2349.

Delaney, G., Gebski, V., Lunn, A.D., Lunn, M., Rus, M., Manderson, C. and Langlands, A.O. (1997) An assessment of the basic treatment equivalent (BTE) model as a measure of radiotherapy workload. *Clinical Oncology*, 9, 2404.

Fayers, P.M. and Hand, D.J. (1997) Generalising from phase III clinical trials: survival, quality of life, and health economics. *The Lancet*, 350, 10257.

Fayers, P.M. and Hand, D.J. (1997) Factor analysis, causal indicators, and quality of life. *Quality of Life Research*, 6, 13950.

Fayers, P.M., Hand, D.J., Bjordal, K., and Gr¯nvold, M. (1997) Causal indicators in quality of life research. *Quality of Life Research*, 6, 393406.

Hand, D.J. and Henley, W.E. (1997) Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society, Series A*, 160, 52341.

Hand, D.J., McConway, K.J., and Stanghellini, E. (1997) Graphical models of applicants for credit. *IMA Journal of Mathematics Applied in Business and Industry*, 8, 14355.

Henley, W.E. and Hand, D.J. (1997) Construction of a k-nearest neighbour credit scoring system. *IMA Journal of Mathematics Applied in Business and Industry*, 8, 30521.

Holland, J.K., Kemsley, E.K. and Wilson R.H. (1997) Transfer of spectral data between Fourier-transform infrared spectrometers for use in discriminant analysis of fruit purees. *Journal of the Science of Food and Agriculture*, 75, 391400.

Jones, M.C. (1997) A variation on local linear regression. *Statistica Sinica*, 7, 117180.

Jones, M.C. and Karunamuni, R.J. (1997) Fourier series estimation for length biased data. *Australian Journal of Statistics*, 39, 5768.

Jones, M.C. and Signorini,, D.F. (1997) A comparison of higher order bias kernel density estimators. *Journal of the American Statistical Association*, 25, 15966.

Krzanowski, W.J. and Hand, D.J. (1997) Assessing error rate estimators: the leave-one-out method reconsidered. *Australian Journal of Statistics*, 39, 3546.

Lunn, R.J., Lunn, A.D. and Mackay, R. (1997) Development of a hydroecological model of the Borrowdale volcanics at Sellafield. *Hydrology and Earth System Sciences*, 1, 3546.

Mohan, D.J., Murray, K., Taylor, P.C. and Steed, P. (1997) Developments in the use of regional secure unit beds over a 12-year period. *Journal of Forensic Psychiatry*, 8, 32135.

Park, B.U., Kim, W.C. and Jones, M.C. (1997) On identity reproducing nonparametric regression estimators. *Statistics and Probability Letters*, 32, 27990.

Park, B.U., Kim, W.C., Ruppert, D., Jones, M.C., Signorini, D.F. and Kohn, R. (1997) Simple transformation techniques for improved nonparametric regression. *Scandinavian Journal of Statistics*, 24, 14563.

Silvertown, J. and McConway, K.J. (1997) Does ìpublication biasî lead to biased science? *Oikos*, 79, 1678.

Yu, K. and Jones, M.C. (1997) A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis*, 25, 15966.

6.3 Refereed conference contributions and book chapters

Gower, J. C., Hand, D.J. and Ngouenet, R.F. (1997) The development of interactive graphical software for a unified biplot methodology. In *Computing Science and Statistics*, vol. 29, eds E.J. Wegman and S.P. Azen, International Association for Statistical Computing, 227.

Hand, D.J. (1997) Intelligent data analysis: issues and opportunities. In *Advances in Intelligent Data Analysis: Reasoning about Data*, eds X. Liu, P. Cohen and M. Berthold, Springer, 114.

Hand, D.J. (1997) Scientific and statistical hypotheses: bridging the gap. In *Understanding Social Research: Perspectives on Methodology and Practice*, eds G. McKenzie, J. Powell and R. Usher, Falmer Press, 12436.

Hand, D.J. and Henley, W.E. (1997) Some developments in statistical credit scoring. In *Machine Learning and Statistics: the Interface*, eds G. Nakhaeizadeh and C. Taylor, Wiley, 22137.

Williams-Thorpe, O., Tindle, A.G. and Jones, M.C. (1997) Characterisation studies: magnetic susceptibility. In *Survey and Excavations: Mons Claudianus 19871993, Vol 1, Topography and Quarries*, eds D.P.S. Peacock and V.A. Maxfield, Institut Francais díArcheologie Orientale, 287-316.

7. Seminars and conferences at the Open University The Department runs an active seminar programme with invited external speakers. During the year we heard the following:

Dr N.R.J. Fieller (University of Sheffield) *Archaeostatistics: old statistics in ancient contexts*

Professor V.S. Isham (University College London) *Epidemic models for macroparasite infections*

Dr T. Sparks (Institute of Terrestial Ecology) *Phenology: 25 years of dodgy data*

Dr N. Thacker (University of Sheffield) *Statistical training of neural networks and the Bhattacharrya metric*

Dr S.G. Walker (Imperial College, London) *Mixtures of beta distributions*Professor R.L. Chambers (University of Southampton) *Maximum likelihood under informative sampling*

Professor B. Johnston (University of Manitoba) *Assessing inter-observer measures of reliability*

In December, the Department again ran the Annual Open University Statistics Conference, the eighth in the series. This year's topic was

Spatial Statistics. The conference was opened by the Vice-Chancellor and the speakers at the meeting were:

Dr M.-C. van Lieshout (CWI, Amsterdam) *Measures of interaction in spatial point patterns*

Dr I.L. Dryden (University of Leeds) *Procrustes shape analysis of spatial point patterns*

Dr M.A. Mugglestone (Rothamsted Experimental Station) *Analysing and modelling binary spatial data using oddsratiograms*

Dr P.C. Taylor (The Open University) *Investigating the relationship between a point process and a lattice process using spectral analysis*

Professor P.J. Diggle (Lancaster University) *Space-time calibration of radar-rainfall data*

Dr J. Perry (Rothamsted Experimental Station) *Analysing spatially-referenced counts using spatial analysis by distance indices (SADIE)*

In February, the Department ran an internal Research Afternoon at which the speakers were Paul Taylor, David Hand, Chris Jones and Niall Adams.

8. Conference papers presented

1. **N. Adams**

Sixth International Workshop on AI and Statistics, Fort Lauderdale, FL, USA: *Assessing and improving classification rules*

J.C. Gower

Tenth European Meeting of the Psychometric Society, Santiago de Compostela, Spain: *The relationships between distance, projection and biplots in multidimensional scaling*

Second International Association for Statistical Computing World Conference, Pasadena, CA, USA: *The development of interactive graphical software for a unified biplot methodology*

D.J. Hand

Conference on Applied Statistics in Ireland, Bunratty, Ireland: *Where do the classes come from?* (Keynote address)

Intelligent Data Analysis97, Birkbeck College, London: *Intelligent data analysis: issues and opportunities* (Keynote address)

Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA, USA: *Modelling data and discovering knowledge* (Invited tutorial)

Microsoft/University of Washington Data Mining Workshop, Seattle, WA, USA: *Statistics and data mining* (Invited talk)

Neural Computing Applications Forum, Bath: *Assessment and generalizability of classification rules* (Invited talk)

Equifax New Technology Forum Conference, Coventry: *New uses of statistics in credit scoring* (Invited talk)

Zeneca Biostatisticiansí Conference: *Path analysis, Lisrel, and other such models*

M.C. Jones

Gordon Research Conference on Statistics in Chemistry and Chemical Engineering, Newport, RI, USA: *Kernel smoothing* (Invited session)

Art of Nonparametric Statistics: Methodologies and Applications, Louvain-la-Neuve, Belgium: *Two separate topics: semiparametric discrete smoothing and local dependence* (Invited talk)

M.N. Kanaan

U.K. Young Statisticians Conference, Lancaster: *Spectral analysis and spatial data* (Poster)

M.G. Kelly

Credit Scoring and Credit Control V, Edinburgh: *Credit scoring with uncertain class definitions*

H.G. Li

Credit Scoring and Credit Control V, Edinburgh: *Direct versus indirect credit scoring classifications*

P.C. Taylor

Open University Statistics Conference: *Investigating the relationship between a point process and a lattice process using spectral analysis*

9. Other seminars and talks given

J. C. Gower

University of Rochester, NY, USA: *Algebra, geometry and statistical graphics* (Ninth Odoroff Memorial Lecture)

RSS Local Group, South Wales: *Nonmetric linear biplots*

Silsoe Research Institute: *Biplots: the graphical representation of multivariate data*

J.C. Gower and R.F. Ngouenet

Unilever Research, Cosworth: *New biplot methods for food technology*

D. J. Hand

RSS, London: *Breaking misconceptions: statistics and its relationship to mathematics* (RSS discussion paper)

Abbey National Plc, Milton Keynes: *Some new approaches to credit scoring*

University of Birmingham: *Statistical aspects of credit scoring*

Zeneca Pharmaceuticals, Macclesfield: *Path analysis*

University of Bath: *Statistical aspects of credit scoring*

Isaac Newton Institute, Cambridge: *Comparing classification rules*

M.C. Jones

University of Newcastle upon Tyne: *Two separate topics: semiparametric discrete smoothing and local dependence*

M.G. Kelly

University of Sheffield: *Granting loans: who and how?*

University of Bath: *Supervised classification when the class definitions are initially unknown*

A.D. Lunn

University of Sydney, Australia: *Estimating treatment effects in an unbalanced two period cross-over trial with Poisson outcomes*

University of East Anglia: *A question of sport*

K. Yu

RSS TIMSAC Study Group, London: *Kernel smoothed conditional quantiles*

10. Editorships

J. C. Gower

Associate editor: *Psychometrika*

Associate editor: *Journal of Classification*

Associate editor: *The Mathematical Scientist*

Statistical editor: *Journal of Agricultural Science*

D. J. Hand

Editor-in-Chief: *Statistics and Computing*

Joint Editor: *Royal Statistical Society Book Series*

Editorial board member: *Journal of Applied Statistics*

Editorial board member: *Statistical Methods in Medical Research*

Editorial board member: *Data Mining and Knowledge Discovery*

Editorial board member: *Studies in Classification, Data Analysis, and Knowledge Acquisition* (book series, Springer-Verlag)

Editorial advisory committee member: *Research in Official Statistics*

M. C. Jones

Joint Editor: *Journal of the Royal Statistical Society*, Series B

Associate editor: *Statistics and Computing*

C. E. Lunneborg

Software reviews editor: *Statistics and Computing*

11. Grants awarded

D. J. Hand

Criteria for assessing classification rules. Grant extension. £7,121 from DERA.

D.J. Hand and J.C. Gower

Biplot methods applied to various data sets. £14,584 from Unilever Research.

M.C. Jones

Minimum distance estimation. Research studentship from Open University Research Committee.

K.J. McConway

(with J. Silvertown (Biology, OU) A.C.E. Miller (Rothamsted Experimental Station), M.R. McNair (University of Exeter)) Ecological and genetic correlates of long-term population trends in plants. £209,430 from Natural Environment Research Council.

12. Other activities

F. Daly

spent most of 1997 on study leave, visiting the University of Strathclyde.

J. C. Gower

is Honorary Secretary of the Sir Ronald Aylmer Fisher Memorial Committee of Great Britain.

is a Foreign Member of the Advisory Board of IOPS (Interuniversitaire Onderzoekschool voor Psychometrie en Sociometrie).

D. J. Hand

is external examiner for the BSc and MSc in Statistics at University College London, and for the MSc in Statistics at the University of Kent.

examined PhDs at the Universities of Surrey, Southampton, Lancaster, and Imperial College London.

spent a month on study leave at the Isaac Newton Institute, Cambridge.

was a member of the programme committees and/or advisory boards for the following conferences: Second International Association for Statistical Computing World Conference, Pasadena, CA, USA; Sixth International Workshop on AI and Statistics, Fort Lauderdale, FL, USA; Softstat 97, Heidelberg, Germany; First European Symposium on Principles of Data Mining and Knowledge Discovery, Trondheim, Norway; Intelligent Data

Analysis97, London; Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA, USA.

was appointed to the Statistics Expert Advisory Panel of Zeneca Pharmaceuticals and to the Advisory Board of the Edinburgh University Credit Research Centre.

M. C. Jones

completed his term as external examiner for the MSc in Computational Statistics at the University of Bath.

examined a PhD at the University of Manchester.

is a member of the Editorial Policy Board, the Programme Committee and the Research Section Committee of the Royal Statistical Society.

M.N. Kanaan

although only in her first year as a PhD student, won the prize for the best poster at the U.K. Young Statisticians Conference.

A.D. Lunn

spent six weeks on study leave at the NHMRC Clinical Trials Centre, Sydney, Australia

organised and ran a series of two-day research workshops on decision making and cost-effectiveness in medicine at the NHMRC Clinical Trials Centre, Sydney, Australia.

was appointed as assessor of Final Honour School of both Mathematics and Mathematics and Computation at the University of Oxford.

K.J. McConway

examined PhDs at the Universities of York and Warwick.

P.C. Taylor

is a member of Statistical Computing Section Committee and Programme Committee of the Royal Statistical Society.

S.K. Vines

continues to act as advisor on statistical matters to all research students associated with the Open University, and to give appropriate courses and talks.