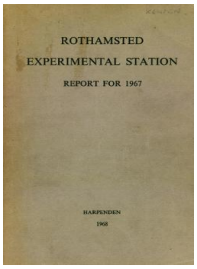
ROTHAMSTED RESEARCH

Rothamsted Report for 1967

Full Table of Content



Statistics Department

F. Yates

F. Yates (1968) *Statistics Department ;* Rothamsted Report For 1967, pp 220 - 229 **- DOI: https://doi.org/10.23637/ERADOC-1-120**

STATISTICS DEPARTMENT

F. YATES

The Orion computer continued to give good service. Productive work increased by 65%, to three times what it was in 1965. The additional equipment obtained from the Rutherford Laboratory in 1966 and an earlier start by the maintenance engineers enabled us to cope with the increased demands without altering the present system of running the machine unmaintained till 10.30 in the evening. Teleprinter links with the National Vegetable Research Station, East Malling, and the Grassland Research Institute were set up as a first step in establishing closer computer links with other institutes.

At the request of the Agricultural Research Council we investigated possible replacements for the Orion. An English Electric 4-70 seems most suitable for our needs, and will have the advantage that we shall be compatible with the Edinburgh Regional Computer Centre, which serves the agricultural research institutes in that area. A new building was also planned, and the plans are with the quantity surveyors.

Because of the increasing need in agricultural and biological research to apply computers to problems not involving statistics, and the organisational problems of running a large computer installation, it was decided to set up a separate Computer Department, thus leaving the Statistics Department free to concentrate on statistical work. The split will be made in January 1968, with D. H. Rees as head of the new department. The Statistics Department will still be responsible for statistical work on the computer and for programming and development of statistical applications.

In spite of loss of staff, we had a productive year. The analysis of the large 1966 Survey of Fertiliser Practice went smoothly and showed interesting changes in the manuring of grassland and cereals, on which nitrogen applications increased by almost 50% and 20% respectively between 1962 and 1966. Several other survey analyses, including the survey of calf wastage, were completed and reported. Several long-term experiments were also summarised.

Our work for the Economics Division of the Ministry of Agriculture gave us the opportunity of preparing a new type-of-farming map on the computer.

The demands for overseas work increased considerably; in 1967 it absorbed some 17% of the time of the scientific staff other than that devoted to theoretical research and general programming work, compared with 7% in 1963. Non-agricultural work, by contrast, has greatly decreased, and now absorbs only 4% of our time, compared with 11% in 1963. Work for Rothamsted increased from 20 to 24% between 1963 and 1967, and for the National Agricultural Advisory Service and other institutes in the British Isles decreased from 62 to 55%; much of this last category, for example summaries of fertiliser trials and the Survey of Fertiliser Practice, is of direct interest to Rothamsted. Theoretical re-

https://doi.org/10.23637/ERADOC-1-120

search and general programming work occupied some 30% of the time in both years; of this about two-thirds was devoted to programming in 1963 and four-fifths in 1967.

The Orion computer

Performance. By rearranging the times of routine maintenance (by arrangement with International Computers and Tabulators Ltd., the engineers now start at 8 a.m. and the machine is made available at 10 a.m.), and by increased evening work, for which the staff deserve credit, 14% more useful time was obtained from the machine (Table 1). This, with the greater output per hour made possible by the enlarged installation, just enabled us to cope with the increased demands for work.

TABLE 1

Orion performance, 1965–7 (hours worked)

	Maintained			Unmaintained			Total			%
	1965	1966	1967	1965	1966	1967	1965	1966	1967	(1967)
Useful time	1543	1426	1686	374	804	856	1917	2230	2542	91
Faults	159	146	152	9	23	18	168	169	170	6
Restarts	35	52	66	4	15	18	39	67	84	3
Total	1737	1624	1904	387	842	892	2124	2466	2796	100
Maintenance							*	752	549	
Total running time							*	3218	3345	

* Not available.

The percentage of time lost by computer faults was marginally less than in 1966; maintenance time was 16% of the time the machine was switched on, compared with 23% in 1966. As previously, most of the faults were transient, with no major breakdown, but there were many irritating minor faults necessitating restarts. Faults occurred on 80% of the working days, and the longest fault-free period was three days.

Volume of work. Table 2 shows the volume of work, measured in terms of the costing program used by the I.C.T. for their commercial costing.

TABLE 2

Work done, 1965–7 (Nominal value, £,000)

	1965	1966			1967
		Jan.–	Oct.–	Total	
Program development	85	88	33	121	135
Production	69	86	38	124	205
Systems	9	8	3	11	12
Total	163	182	74	256	362
Work per hour of useful time	£85	£104	£155	£115	£143
Production per working day	£270	£460	£670	£510	£812

The actual cost at the 1967 rate of working, allowing for depreciation, maintenance and operating staff, is about one-third of the accounting cost. The figures for 1966 are separated into two periods, marking the increase in size of the installation. Production work increased by 65% over 1966,

https://doi.org/10.23637/ERADOC-1-120

and is about three times what it was in 1965. For the first time program development was much less than production.

In interpreting this apparent increase in volume of work, it should be borne in mind that the costing program is likely to give somewhat greater cost for the same job if done on a larger installation. This is because the amounts of store and peripheral equipment reserved contribute to the nominal cost. With more core and drum store available it pays us to reserve more store in our general programs, to be able to cope with large jobs without modifying the programs. Similarly, in the new Genfac program a magnetic tape deck is always reserved, even if none of the experiments in the batch being processed requires storage on magnetic tape.

Table 3 gives an indication of the direct use made of the computer by other institutes, either for jobs programmed by the institutes themselves or for jobs run on Rothamsted programs for which the institutes wrote

TABLE 3

Direct use of the computer by other institutes
(Nominal value, £,000)

	1965	1966	1967
Development	24	40	41
Production	26	33	52
Experiments*	7	16	27
Total	57	88	120
% of total nominal value	35%	33%	33%

* Estimated—see text.

their own control instructions and prepared their own tapes. For experiments for which Rothamsted and outside work is batched together, no direct separation in terms of nominal value is possible—the total nominal value of experimental runs was therefore split in proportion to numbers of variates analysed. Although the percentage of the total nominal value of all work was slightly less in 1966 and 1967 than in 1965, the amount of production work done directly by other institutes increased substantially, particularly in 1967, when the development work also bore a much better relation to production work.

The distribution between institutes is very similar to that of previous years. Figures for 1965 and 1966 are given in the 1965 and 1966 reports; these figures, however, may not fully reflect the relative use made by different institutes, as experiment analysis could not be included.

Responsibilities of the Computer Department. The Computer Department will be responsible for the organisation and running of the computer, acquiring new equipment and advising other institutes on equipment, providing good routine program-punching and routine card-punching services, organising and maintaining program libraries, editing and circulating program descriptions and establishing and maintaining reference data files. It will also be responsible for the whole of the systems side of the programming, including adequate programming aids for information retrieval and input and output, and for providing general programming assistance and advice. One of its primary tasks will be to develop the use

https://doi.org/10.23637/ERADOC-1-120

of the computer by other departments in the many applications, other than statistics, for which computers are proving of value in biological research. This will involve developing new methods and writing programs where none exist, and seeking and making available on our computer programs for non-statistical work, such as crystallographic analysis, and systems synthesis, written for other machines.

This division will enable the Statistics Department to concentrate on statistical problems. Computer work from other institutes needing the intervention of statisticians will first go to the Statistics Department and will be returned to them for examination of the results. They will also provide a data-preparation service for statistical work. Statistical programming, and the development of common statistical systems jointly with other statistics departments, will remain their responsibility, but they will be able to make use of the general programming services of the Computer Department for the systems side of their work; at present statisticians have to deal with many programming problems, such as card input, magnetic-tape operations, file management and table printing, which should be the concern of a systems group. They will also no longer be expected to be experts on the use of the computer for operations such as crystallographic analysis and the solution of differential equations.

Programming developments

Programs for the analysis of experiments. The elimination of residual errors from the revised version of our General Factorial Program proved unexpectedly troublesome, but the program has been in use throughout the year, and now seems to be substantially free of errors. (Preece and Alvey) The revised version of the General Input for Experimental Designs (GIED) also ran into trouble, and had to be completely rewritten. (Yates and Preece) This revised version, which is incorporated in the General Factorial Program, provides several useful additional facilities, of which the most important are: reading data from cards, direct specification of various types of linear functions of variates, derivation of conversion factors from plot dimensions, reordering of data recorded in serpentine fashion (i.e. in opposite directions along alternate lines of plots). Provision for transferring data to and from magnetic tape has now been added. (Anderson and Lowe) Experience with these revised programs has already demonstrated that the improvements were well worth the effort required.

Lauckner extended Ross's General Lattice Program. This will now analyse lattice squares, square lattices and rectangular lattices; a card input is being tested. Patterson and Lowe further improved the Long-Term Experiments Program, described in the 1966 Report.

Programs for summarising sets of experiments. With the new version of the General Factorial Program, the results of analyses can be stored on magnetic tape, so that sets of experiments can be subsequently summarised. A program for making these summaries is now complete and working. (Yates, Alvey and Preece) Hierarchical summaries with up to three levels (sub-groups, groups and totality) can be made, and summaries can, if

https://doi.org/10.23637/ERADOC-1-120

required, be made by stages, incorporating the results of previous partial summaries. Linear functions of the values in the factorial tables (linear effects) can be specified in the same manner as in the General Factorial Program.

The results of the summary operations, and the information relating to each experiment, are assembled on magnetic tape, and the results are printed from it by a separate part of the program. This enables different features of the results to be set out in separate tables, as the magnetic tape can be run through repeatedly to pick out different items of information. In particular, tables can be printed in which each line contains a selected set of values appertaining to an experiment, such as the means of chosen treatment combinations, or linear effects, and their associated standard errors. The experiments are printed in order, line by line, with interspersed lines for sub-group, group and general means; if desired, printing can be restricted to the sub-group, etc., means. Printing can also be done by stages, with each stage examined to decide what further is required. Hierarchical analyses of variance of the factorial table and of the linear effects can be printed. Flexible provisions are made for the specification of the required standard errors.

The construction of the General Summary Program led to the construction of an *Ad hoc* Summary Program, which performs similar functions for data stored on magnetic tape (or read directly from paper tape) but not processed by the General Factorial Program. In this program certain routines controlling the input of data and calculation of standard errors are written by the user in EMA, thus enabling very varied arrangements of data, etc., to be dealt with. The summary results are stored in the same form as for the General Summary Program and are printed by the print section of that program.

The *Ad hoc* Program proved very useful in making a rapid summary of the set of long-term residual phosphate experiments and a similar set of long-term ley fertility experiments done by the National Agricultural Advisory Service on the Experimental Husbandry Farms. It does not take account of components of error arising from differences between plots that are constant over the years, but presents the results better than the Long-Term Experiments Program.

General Survey Program. Good progress was made in revising this early in the year, but it then had to be postponed because of other programming. It was taken up again later, and a usable version should be ready shortly. In this program, also, we are endeavouring to make analysis by stages easy, so that the user can determine what further is required by examining a selected set of complete or part tables, thus avoiding the need to guess at the outset all he is likely to require, and then being confronted with piles of folded stationery covered with figures from which he has to extract whatever (often regrettably little) is of value. (Anderson and Yates)

Other statistical programs. Anderson wrote a program that does multiple-regression analysis much more elegantly than our previous program. The output is well annotated, and various useful additional facilities are

https://doi.org/10.23637/ERADOC-1-120

provided. The results of analyses can be stored on magnetic tape so that the analysis can be resumed later if required. He also wrote a program for discriminant analysis, and together these will form the nucleus of a set of compatible multivariate analysis programs.

Ross added some further special routines to his maximum likelihood program and is translating it into EMA (the program is written in Orion machine code) with a view to its transfer to other machines.

Simpson produced a routine for constructing scatter diagrams on the line-printer. Various other small general statistical programs were written.

Non-statistical and systems programs. Programming for non-statistical and systems work was handicapped by shortage of staff, but some useful work was done. We completed an analysis of experimental records of temperature gradients in electrically stimulated nerves. Interpretation is very difficult, as the observations are near the limits of instrumental accuracy; we are investigating further how far computers can help in the objective assessment of such records. We are also developing the use of the computer for interpreting traces from automatic chromatographic equipment. Further work was done on the routines required to work the incremental plotter, and the plotter can now be used to produce dendograms from the output of one of our classification programs. Various small jobs were programmed for other Rothamsted departments. (Rees, Bicknell and Martin)

Experiments

The results of the Woburn Ley–Arable Experiment from 1956 to 1967 were analysed and are reported on p. 316.

The number of experiments analysed on the computer was almost the same as in 1966, but there was again an increase (19%) in the number of variates analysed (Table 4). Rothamsted programs were used for 94% of the experiments and 90% of the variate analyses. (Dunwoody, Hill, Alvey and Ryan)

TABLE 4

Numbers of replicated experiments analysed in the department

	Number of experiments			Number of variates on computer	Variates per experiment
	By hand	On computer	Total		
1951	437	—	437	—	—
1957	98	1253	1351	5041	4·0
1963	72	2770	2842	14357	5·2
1964	88	3383	3471	18054	5·3
1965	69	4751	4820	28663	6·0
1966	109	6162	6271	40826	6·6
1967	17	6124	6141	50373	8·2

Data for experiment analyses (often including specification of the design and the analysis required) are increasingly being punched at other institutes. In 1965 the data for only 26% of the experiments analysed were prepared at other institutes, compared with 36% in 1966 and 42% in 1967. The corresponding figures for the associated variate analyses are 39, 50 and 55%. The Rothamsted punching load (measured by variates analysed)

https://doi.org/10.23637/ERADOC-1-120

therefore only increased by 11% in 1967. This is a gratifying trend; there are many advantages in decentralising data preparation.

In interpreting the above figures it should be borne in mind that the number of variates analysed gives only a rough indication of the amount of computer work and data preparation. Experiments differ greatly in size and complexity. Moreover, there is no fixed relationship between the number of variates that have to be punched and read in and the number analysed: the variates analysed are frequently functions, which require preliminary calculation, of the variates read in. However, more elaborate statistics are not thought to be worth the effort required to compile them.

National Agricultural Advisory Service

In addition to the routine analysis of experiments on Experimental Husbandry Farms and commercial farms (Hill and Lessells), the set of six residual phosphate experiments at the Experimental Husbandry Farms, which ran from 1951 to 1964, and a similar set of ley fertility experiments, which ran to 1967, were summarised. (Alvey) These are being written up by members of the National Agricultural Advisory Service, in collaboration with Boyd.

A summary of 77 experiments testing P and K fertilisers done in 1964–66 was made. The responses to both P and K were small and varied little from site to site (12.15). Various other sets of manurial trials have been summarised, and reports are being agreed for publication.

Lessells collaborated with members of the National Institute for Research in Dairying and the National Agricultural Advisory Service in an investigation of ways of improving the efficiency of experiments on farm animals. He also continued to collaborate in work on the Agricultural Research Council metabolisable energy system for ruminants. Various other analyses of continuing animal experiments were made. (Lessells, Hills)

Surveys

The main analysis of the 1966 Survey of Fertiliser Practice was completed and the results reported (12.2). Supplementary analyses were made to determine the types of fertiliser used on different crops, fertiliser practice on horticultural crops and variation in rates of fertiliser application between farms. Cereal manuring is also being investigated in more detail. The summary tables for the individual districts were produced by the computer in their final form on paper tape, which was then used to cut stencils on a flexowriter. This saved much typing and checking.

In collaboration with the Potato Marketing Board and the National Institute for Agricultural Engineering, plans were prepared for a survey in 1968 of the practices used in growing early and maincrop potatoes. (Church and Hills) This is similar to the surveys of maincrop potatoes done in 1958 and 1963. Analysis of data of the Potato Marketing Board survey on quality of potatoes sampled from store in 1965–66 and 1966–67 continues as required.

We collaborate with the Ministry of Agriculture Plant Pathology Lab-

https://doi.org/10.23637/ERADOC-1-120

oratory on their surveys on the use of pesticides. The analysis of the data of the soft fruit survey was completed, and work on other surveys is being done as required. (Church) We also organised direct recording of data, using a paper tape punch, of the Plant Pathology Laboratory's surveys of *Rhynchosporium* and other foliar diseases of barley. This enabled the information to be rapidly summarised as required, and is likely to become standard practice for such surveys. (Rees, Martin and Hills)

The analysis and reporting of the Calf Wastage Survey was completed (12.10). Data for the second period of the Guernsey progeny testing scheme and for the surveys on the disposal of sheep-dip residues were also analysed and the results reported. (Leech)

Commonwealth and overseas

The service we provide for the routine analysis of experiments is of value to many of the tropical countries in the Commonwealth. Sets of experiments were analysed for Pakistan (tobacco), Tanzania (cotton), Sarawak (rice, maize, sweet potatoes), Nigeria (cocoa, kenaf, cotton and millet), Trinidad (cocoyam), Barbados and Antigua (yams, maize, etc.). Other similar sets are being analysed. In addition to the work dealt with directly by Rothamsted, East Malling Research Station sends us data from experiments on tree crops for analysis on the computer. Up to now we have for the most part only provided analyses of the individual experiments, leaving it to the workers concerned to make their own summaries. The recently completed Summary Program should enable us to make useful computer summaries as each set is analysed. (Walker and Dunwoody)

Various other statistical problems of varying degrees of complexity were dealt with by Ross and Walker, in particular the calculation of genetic components of variance (Cotton Research Station, Namulonge), classification of marine bacteria (New Zealand), ecological classification (Nigeria), comparison and calibration of the Gunn-Belloni and Kipps solarimeters (Uganda), and screening of cocoa varieties for resistance to swollen-shoot disease (Ghana).

Theory

Gower derived some important properties of a general coefficient of similarity (12.7). He also examined various methods of cluster analysis that have been proposed and showed that they can all be interpreted in terms of distance in multi-dimensional space between the centroids of the clusters (12.6). He surveyed statistical methods of classification used in taxonomy and drew attention to some of their deficiencies (12.8).

Preece continued his investigations on balanced incomplete block designs (12.12, 13, 14). Patterson completed a paper on his investigations on serial factorial designs (12.11), and Yates made a small contribution to the problem of drawing valid conclusions from cloud seeding experiments (12.16). Simpson resolved some further theoretical points that arose from anomalous results and failures in our programs for fitting constants to quantitative and quantal data.

https://doi.org/10.23637/ERADOC-1-120

Other work

Ross's work on the population dynamics of the potato cyst-nematode *Heterodera rostochiensis* was completed (12.9). Many statistical problems arising in other Rothamsted departments, and some for other institutes, were dealt with, including regression, discriminant and classification problems, and routine processing of results of herbicide screening trials for the Weed Research Organisation.

Gower and Simpson analysed a set of tests on perception of different odours for the Food Research Institute, Norwich. Thirty-five individuals tested 45 materials, scoring these for the presence and intensity of 44 different types of odour. This provided a useful opportunity for developing techniques of analysis for this type of test. The conclusions will be reported in the Annual Report of the Food Research Institute.

Dr. J. W. B. Douglas, Miss Jean Ross and Simpson completed their book on the results of the National Survey of Health and Development of Children relating to secondary education; a paper was also read to the Royal Statistical Society on the ability and attainment of short-sighted children (12.5). The appointment of a statistician by the National Foundation for Educational Research enabled us to discontinue any direct work for them, but they still use our computer from time to time in the evenings with their own operating staff. Gower assisted in the analysis of a survey of science students at five universities by the Oxford University Department of Education; a preliminary report was published by them.

Investigational work on agricultural census, grants and subsidy data for England and Wales was continued for the Economic Division of the Ministry of Agriculture, using our General Survey Program. Some of the census information was printed by the computer in map form, giving various characteristics for 10-km grid squares. This led to the construction of a new type-of-farming map. The map is being published in colour, together with the definitions used to delimit the various farming types, and certain census characteristics of the delimited areas (12.3). The only previous map of this sort was one published in 1941 by the Land Utilization Survey of Britain. This was in part based on subjective assessments, whereas the new map uses objective criteria. Using computers, maps of this kind, constructed on a consistent objective basis, can be prepared periodically to illustrate geographical changes in the pattern of farming.

Staff and visiting workers

Mr. J. A. Nelder, Head of the Statistics Department of the National Vegetable Research Station, was appointed to succeed F. Yates, who retires from the headship of the department in March 1968.

H. D. Patterson left to join the A.R.C. Statistics Unit, University of Edinburgh, under Professor D. J. Finney. J. A. Lewis was appointed to a lectureship in the Statistics Department, Leeds University, and Alison Macfarlane also left.

P. Walker, Margaret A. Watson and K. C. Ryan were appointed, the

https://doi.org/10.23637/ERADOC-1-120

first-named to a post financed by the Ministry of Overseas Development. A. J. Vernon completed his secondment to the Cocoa Research Institute, Ghana, and is now with the Fiji Department of Agriculture.

J. C. Gower spent six months at the Bell Telephone Laboratories, New Jersey, and G. J. S. Ross spent two months at the Sugar Industry Research Institute of Mauritius. B. M. Church attended two CENTO Agricultural Statistics discussions in Ankara as U.K. representative. P. Walker visited the Ahmadu Bello University, Nigeria, and the University of the West Indies, Trinidad. Eight temporary workers, two from the United Kingdom, two from Australia, and one each from Denmark, India, Uganda and the Philippines, spent various periods in the department.

https://doi.org/10.23637/ERADOC-1-120