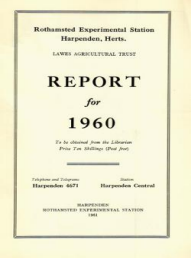
ROTHAMSTED RESEARCH

Report for 1960

Full Table of Content



Statistics Department

F. Yates

F. Yates (1961) *Statistics Department ;* Report For 1960, pp 176 - 185 **- DOI: https://doi.org/10.23637/ERADOC-1-93**

STATISTICS DEPARTMENT

F. YATES

Muriel E. Davis resigned during the year to take up a post with the Department of Agriculture, New Zealand. S. F. Buck, Agricultural Research Council Scholar, joined the Pig Industry Development Authority on completing two years' research for a Ph.D. New appointments were: J. R. Davis from the National Coal Board, A. Frater from the R.A.F. Bomber Command Development Unit and M. P. Vessey from University College London Medical School.

Healy returned in July from a year's visit to the U.S.A., most of which was spent at the Bell Telephone Laboratories, Murray Hill, N.J. While there he gained valuable experience working on a large electronic computer, the I.B.M. 704. Church spent two months in Kenya at the invitation of the East African Extract Corporation Ltd., investigating methods of sampling pyrethrum. Leech spent two months in Southern Rhodesia planning a survey of the fertility of beef cattle. Yates attended an International Biometric Meeting organised by the Société Adolphe Quetelet as part of the Centenary Celebrations of the Institut Agronomique, Gembloux, and gave a paper on the use of electronic computers in the analysis of experimental results (12.4). Leech attended a Symposium on Quantitative Methods in Pharmacology held by the Biometric Society in Leyden.

Yates was awarded the Guy Medal in Gold of the Royal Statistical Society; he was appointed President of the British Computer Society for 1960/61 and Chairman of the Committee on Agriculture of the Ministry of Science (formerly the Sub-Committee on Agriculture of the Natural Resources Committee, Lord President's Office).

Work started early in 1961 on an additional building for the Department, which will house the new electronic computer and provide much-needed accommodation for staff.

THE ELECTRONIC COMPUTER

The Elliott 401 continued to give satisfactory service until July, when a bearing of the magnetic disk store failed. The manufacturers did a good job of work in replacing this store by a drum store of the type that is fitted to the 402. This was repaired in a month, but unfortunately there has since been recurrent trouble from interference between the motor driving the drum store and the electronic circuits. A further two weeks were wasted in investigation of the trouble by the manufacturers, but it has not yet been cured.

The breakdown of the machine and the subsequent troubles have seriously interfered with the work of coupling up the magnetic tape unit to the computer, and although considerable progress has been made, we still have a fair amount to do.

In spite of these troubles, we have been fairly successful in coping with current demands, but at the cost of considerable strain to

https://doi.org/10.23637/ERADOC-1-93

members of the Department, and particularly to the engineering staff. We have now introduced for a trial period a regular system of overtime, working from 5.30 p.m. to 9.30 p.m. for three nights a week, which it is later hoped to extend to five nights a week. Previously overtime running in the evenings was almost wholly done by scientific staff. Systematic overtime working will, of course, only be possible if the computer gives reasonably reliable service.

TABLE 1

Record of machine operation for 1957–60

Percentage of time for:	1957	1958	1959	1960 *
System modification	0·6	2·4	0·0	0·2
Scheduled maintenance ...	12·4	13·9	9·4	9·8
Unscheduled maintenance:				
Computer	6·2	2·3	3·1	3·1
Peripheral	1·8	1·5	1·1	0·4
Programme checking	17·8	13·3	8·9	14·8
Programme tape preparation ...	4·8	3·5	1·4	2·6
Production runs	49·8	58·4	73·9	67·2
Abortive efforts:				
Computer	1·3	1·4	0·7	1·6
Peripheral	1·0	1·7	0·8	
Idle time	4·4	1·7	0·7	0·4
Total	100·1	100·1	100·0	100·1
Total hours worked	2,622	2,869	2,993	2,757
Percentage overtime	32·6%	45·1%	50·8%	55·9%

* 46 weeks.

Table 1 shows the record of machine operations over the last 4 years, excluding the 6 weeks for which the machine was out of commission in 1960. Although there is little difference from previous years in the percentage of total time during which the machine was recorded as working incorrectly or out of commission (apart

TABLE 2

Numbers of replicated experiments analysed in the department

	Number of experiments: By hand	On computer	Total	Number of variates on computer
1934	115	—	115	—
1951	437	—	437	—
1955	384	419	803	834
1956	181	683	864	1,701
1957	98	1,253	1,351	5,041
1958	182	1,664	1,846	6,260
1959	67	2,649	2,716	11,102
1960	85	3,687	3,772	11,147

from the 6 weeks referred to above), there was a considerable difference in efficiency before and after the breakdown. Of the 85 hours spent on unscheduled maintenance, for example, 74 occurred in the last 5 months of the year. Moreover, some time that has been booked to production for survey analysis should have been included in abortive effort.

The main increase in work this year has been in survey analysis,

https://doi.org/10.23637/ERADOC-1-93 *pp 3*

largely as a result of completing the General Survey Programme referred to below. 25% of the total production time was used for survey work. If machine time is available a further substantial increase in survey work may be expected in 1961.

The number of experiments analysed on the machine (Table 2) again increased, because of a large series of very simple fertiliser trials (894 trials) analysed for Pakistan, and of the fact that part of this series (523 trials) had to be re-analysed because we were not informed that the treatments had been changed. The number of variate analyses is almost the same as last year.

Replacement of the Elliott 401

The replacement of the Elliott 401 has been under consideration for some time, and it has now been decided to install a Ferranti Orion. An order has been placed, and delivery has been promised for the autumn of 1962. The Orion is a considerably larger and faster machine than the 401. It will have a computing speed of the order of 25–50 times that of the 401, and input and output of information with the ancillary equipment we envisage will be of the order of 10–20 times that of the 401.

Certain items of equipment for the input and output of information are still under discussion, but the installation will probably be as follows:

Central computer

4,096 words of immediate access storage
32,768 words of drum storage
Built in floating-point facilities

Magnetic tape

Two Ampex units

Input

Paper tape and punched cards

Output

Paper tape, punched cards and line printer

One important advantage of a larger and more modern machine is that less effort is required for programming. During the last 5 years much progress has been made in the simplification of programming, and in the art of constructing programmes which are simpler to use and cover a wider variety of problems of the same general type.

(a) *Improvement in order codes*

The order codes of the more modern machines are much more straightforward and simple to use than those of the early machines. In part this is a consequence of improvements in electronic techniques, in part from recognising the importance of lightening the programmer's task if a computer is to be used effectively. These simplifications are greatest in large machines.

https://doi.org/10.23637/ERADOC-1-93

(b) *Autocode facilities*

The use of autocodes is rapidly spreading. An autocode enables the operations required to be specified in simple form; thus mathematical operations can be specified by ordinary mathematical formulae. This specification is translated into the machine code by means of a programme held in the machine. Effective autocodes can only be developed on large machines. (Ferranti have provided an autocode for the Mercury, and propose to provide a similar autocode for the Orion.)

(c) *Generality of programmes*

Many statistical computations, such as the analysis of experiments, the analysis of surveys and much multivariate work, are largely standardised. Within these standard types, however, there are many variants, and it is important that programmes should be such that all variants normally encountered can be called for without any special programming. Such general programmes are necessarily complex, and can only be effectively written for a large machine. Our 401 programmes for the analysis of experiments are too inflexible to deal conveniently with exceptional designs, and some designs have not been programmed at all. These defects were not of great importance initially, when the total number of experiments analysed was small, but are serious now the load has increased tenfold. Similarly in the analysis of surveys, machine limitations prevent the construction of a programme which is sufficiently powerful and speedy to deal efficiently with large, complex surveys.

(d) *Ease of specification*

A programme must not only be sufficiently general. It is also important that the specification of exactly what is required in any particular analysis is itself simple, for this makes the specification easier and substantially reduces errors. Equally, in the presentation of results it is important to provide a good lay-out and introduce printed headings where required, particularly when the results have to be sent to research workers in other stations. Here again adequate storage and speed are required; our 401 programmes leave much to be desired in these respects.

(e) *Floating-point arithmetic*

Although in much statistical work fixed-point arithmetic can be used, this inevitably causes programming difficulties which are entirely avoided by the use of floating-point facilities. Fixed-point arithmetic also complicates the specifications of the work required, because instructions regarding scaling must often be included in these specifications. To evaluate complicated mathematical functions, it is almost essential to work in floating point. This can be done on a fixed-point machine, but only with considerable loss of speed.

Programming has been a major limiting factor in using the 401. Although theoretically much more work could be handled by round-the-clock working, this would require a considerable increase in programming staff, and from the national point of view it would be

https://doi.org/10.23637/ERADOC-1-93

very uneconomic to devote scarce programming ability to the programming of such a slow machine, even if the necessary staff could be obtained.

The greater ease of programming and the availability of autocode facilities and easily used general programmes for such work as the analysis of experiments and surveys should enable statisticians at other research stations to participate much more directly in programming the Orion for their own problems than has been possible with the 401. We shall make every effort to encourage this co-operative approach.

Programming Developments

We reported last year that a general system for the specification of survey analyses had been worked out. The view was then expressed that it would not be practicable to write a programme for the 401 which would interpret these specifications directly, so avoiding special programming for survey analysis. Further investigation showed, however, that a useful general programme could be constructed, and this has now been done. (Yates and H. R. Simpson 12.23, 12.26, 12.27.) This programme, though slow, has revolutionised the analysis of surveys on the computer. It will also serve as a prototype for a faster and more sophisticated programme for the Orion.

The construction of the general survey programme has suggested a similar approach to the analysis of experiments. This may provide a solution to the analysis of long-term experiments and one-year experiments with exceptional and rarely occurring designs. A preliminary investigation on these lines has been started. (Yates and Gower.)

No further work has been done on the development of a communications language suitable for statistical programmes; it was decided that the change of scale between the 401 and the Orion was such that it would be more profitable to construct much more general programmes for the Orion than to attempt to transpose the programmes we at present use on the 401. We are, however, investigating the possibility of writing a programme for the Orion that will enable it to accept 401 programmes, so that as an interim measure little-used programmes can be used on the Orion without any re-programming. This task is more complicated than appears at first sight, mainly because in B-modification on the 401 the whole of the order is modified, and thus under certain circumstances the address of the next order or the operation performed may be changed. The simplest procedure is to use a straight simulation routine in which each order of the 401 is re-interpreted every time it is obeyed, but this would be very slow; a combination of translation and simulation appears to be what is required, simulation being used for modified orders when the contents of the modifying register are such that the address of the next order or the operation is changed. (Bonsall and Rees.)

We may also write a routine for simulating Orion operations on the 401. Such a routine will enable Orion programmes to be tested before our machine is installed. It will be worth while to do this

https://doi.org/10.23637/ERADOC-1-93 *pp 6*

only if adequate facilities for testing Orion programmes cannot be obtained elsewhere.

A set of programmes has been written for classifying different types of organism according to the similarities revealed by a set of tests. (Gower.) This problem has attracted attention in various quarters, and a method was recently proposed by Dr. P. H. A. Sneath, and tested on some data on strains of bacteria, using a computer. (The application of computers in taxonomy, *J. gen. Microbiol.* 1957, **17**, 201–226.) Essentially the method consists of calculating coefficients of similarity between every pair of strains and then grouping the strains in such a manner that, for any given level of similarity, each strain in a group has a similarity greater than the given level with at least one other member of the group; this grouping is performed successively at decreasing levels of similarity until all the strains are grouped together. Although simple in principle, the procedure involves heavy computing and a considerable amount of classification work when the number of strains and tests is at all large. The method appears to be of wide interest as we have already had requests for:

(1) Taxonomic grouping of a number of types of bacteria. (Dr. M. Pleasance, Low Temperature Research Station.)

(2) Taxonomic grouping of a number of types of fungi. (Miss G. W. Waterhouse, Commonwealth Mycological Institute.)

(3) Grouping of Pacific islands by the species of birds occurring in them. (Mr. I. C. J. Galbraith, Natural History Museum.)

(4) Grouping of psychiatric patients by their various symptoms. (Dr. A. E. Maxwell, Maudsley Hospital.)

Of the other general programmes developed during the year, the following may be mentioned:

(1) Programmes for centroid factor analysis and for rotating and plotting factor loadings (Healy).

(2) An auto-covariance programme and a Fourier transform programme for the estimation of power spectra in time series. The construction of recurring digital filters for time-series work has also been investigated (Healy).

(3) The fitting of constants to large two-way tables which are beyond the capacity of our present programme (Bonsall).

(4) A general programme for the fitting of Chebyshev polynomials to observed data (Bonsall).

(5) A programme for determining the distribution of distances between pairs of diseased plants in a field. This forms part of an investigation in the spatial spread of disease in crops (Gower, with A. J. Gibbs, Plant Pathology Department).

Experimental Design and Analysis

The results of recent fertiliser trials on potatoes were studied and compared with those of earlier series. Although there appeared to be no large changes in response to the standard nutrients, the study revealed several new facts and in particular showed that there were

https://doi.org/10.23637/ERADOC-1-93

not only large interactions between nutrients but also that the shape of the response curves varied considerably at different levels of the other nutrients. Evidence was obtained that the exponential form of the response curve is unsatisfactory and that yield may be decreased by excessive dressings (Boyd (6)). The responses of potatoes to manuring in the peaty Fenlands and their relation to soil types were also examined and reported, in collaboration with the National Agricultural Advisory Service. (Dr. N. H. Pizer *et al.*, and Boyd (12.21).)

Recent experiments on nitrogen responses of cereals were studied for the National Agricultural Advisory Service (Lessells). Nitrogen response is very variable and sometimes negative, the following factors being important: yield without nitrogen, time of sowing, time of application, previous crop, nutrient status of soil, and disease incidence.

Current fertiliser practice was examined in the light of the results of manurial experiments on potatoes, sugar beet and cereals. The large variations that occur in fertiliser applications from farm to farm appear to be largely unrelated to crop requirements. However, differences in average applications from districts known to have substantially different requirements are small (Boyd (12.8)).

At the request of the Soil Fertility Committee of the Soil Chemists' Conference, the correlation of soil analysis with crop responses to fertilisers was studied and a joint paper is to be published (Boyd and E. P. Simpson).

A survey of long-term agricultural field experiments has been started. So far this includes studies of the structure of errors, the value of an initial period of uniform treatment in improving the accuracy of subsequent results and an assessment of the value of time-trends as indicators of the long-term effects of treatments (Patterson).

The results to date of six long-term experiments on Experimental Husbandry Farms on the use of various forms of phosphatic fertiliser were summarised (Patterson), and reports prepared for the Animal Experiments Sub-Committee on the preliminary results of the Beef Progeny Tests and Beef Carcase Assessment from the experiments done at Experimental Husbandry Farms (Boyd, Frater). We have analysed many other experiments, both on crops and animals, for the National Agricultural Advisory Service, and advised on experimental design (Dunwoody, E. P. Simpson, Frater and Turner). An article was written on the planning of series of experiments (Boyd (12.7)).

The Bulletin on changeover designs by Patterson and Professor H. L. Lucas (Institute of Statistics, North Carolina), referred to in the 1958 Report, was completed (12.20).

An analysis of some experiments on the effects on the growth of *Vicia* with light and temperature treatments on both roots and shoots has been completed for Professor W. T. Williams, Botany Department, Southampton University (Gower). This involved fitting constants and a covariance analysis between two variates, both with incomplete data.

https://doi.org/10.23637/ERADOC-1-93

Surveys

Surveys of fertiliser practice were made by the N.A.A.S. Soil Chemists in 10 districts of England and Wales in 1960. Field work was delayed by the wet summer and by foot-and-mouth disease, but four districts have been analysed. Parallel surveys in Scotland were made in six districts, of which four have been analysed. Two summary papers on the 1957–58 Survey of Fertiliser Practice were completed (Boyd, Church and Hills (12.9), (12.12)), as was a report on the survey of fertiliser practice on market garden crops of the Evesham–Pershore district.

The report on the 1958 Maincrop Potato Survey, done jointly with the National Institute of Agricultural Engineering and the Potato Marketing Board, was published by the Potato Marketing Board (Mr. J. K. W. Slater, Church and Hills (12.22)).

Analysis of the Grassland Survey of England was completed for the Grassland Research Institute (H. R. Simpson and Mr. C. D. Kemp).

The herbicide survey, started in 1959 in collaboration with the A.R.C. Unit of Experimental Agronomy, Oxford, was extended to a further five districts in 1960 (Church).

The analysis of the first National Survey of Disease in Dairy Cattle, covering the year 1957–58, was completed and a report published (Leech, Davis, Vessey (12.18)), and the 1958–59 data were analysed (Vessey). Field work started on a survey of *Brucellosis* in dairy herds in conjunction with the Ministry of Agriculture. A preliminary national survey of losses in breeding ewes was analysed (Davis), and the field work of a second more detailed survey is now completed.

Commonwealth Work

We continued our advisory work for Commonwealth countries and analysed many experiments on the computer (Vernon and Turner). The 1959–60 results of the co-operative fertiliser trials on rice in the State of Bihar, India, were analysed, and a combined analysis of the 4 years' results was also made; this provided an interesting exercise in the analysis of the results of a large series of trials extending over several years, and it is intended to publish a paper on the methods adopted (Yates and Gower). A further year's results from a similar and simpler series of experiments in Pakistan were analysed (Gower). The analysis of a long-term experiment on coconuts for the Coconut Research Institute, Ceylon, was completed by Dr. T. Eden (working in this Department) and Gower.

A report was prepared on the results of a capsid control experiment run jointly by the Shell Chemical Company and the Western Nigerian Ministry of Agriculture. Similar data from work in Ghana is now being examined (Vernon).

We are collaborating with Mr. J. L. Gregory, of the Federal Agricultural Research Organisation of Nigeria, in the examination of a voluminous set of data from light traps and insecticidal studies of yam beetles (Vernon and Morris).

At the request of the East African Extract Corporation, supported

https://doi.org/10.23637/ERADOC-1-93

by the Department of Agriculture and the Pyrethrum Board of Tanganyika, Church visited Kenya to investigate large-scale commercial methods of sampling dried pyrethrum flowers. An unbiased method of reference sampling was developed suitable both for direct use in experimental work and as a standard for testing the adequacy of routine commercial sampling. The mechanical sampling method used by the East African Extract Corporation was tested against this standard (12.11).

Four Commonwealth workers worked in the Department for varying periods. Leech spent 2 months in Rhodesia, planning a Federal survey of the fertility of beef cattle.

Other Work

Of the many miscellaneous problems from Rothamsted and other research institutes and organisations dealt with by members of the Department, most of which have involved the use of the computer, the following may be mentioned:

1. A comparison of theoretical and empirical results for some stochastic population models (Professor M. S. Bartlett, Gower and Dr. P. H. Leslie (12.5)).
2. The derivation of marginal percentages in quantal data with disproportionate frequencies, analysed by use of a transformation (Yates (12.25)).
3. A method of dealing with incomplete multivariate data, suitable for use on an electronic computer (Buck (12.10)).
4. Continuation of the investigation on the multiple transfer method of sampling (Gower with A. J. Gibbs, Plant Pathology Department (12.15)).
5. A programme for Darroch's capture–recapture method (described in *Biometrika*, **46**, 349–351) for F. Raw, Entomology Department, and Dr. P. H. Leslie, Oxford Bureau of Animal Population (H. R. Simpson).
6. An investigation of the joint action of four poisons, following the methods of Hewlett and Plackett. Two models for the joint action of poisons based on the known responses of each poison separately both gave satisfactory agreement with experimental results (Gower with R. M. Sawicki, Insecticides Department).
7. Biological studies of the wheat-bulb fly (*Leptohylemia coarctata* (Fall)) were continued in collaboration with the Entomology Department (Morris, R. M. Dobson and D. B. Long (12.13)).
8. Least-squares estimation of components of variance in genetical data. This analysis was undertaken for the Department of Genetics, Birmingham University, using a corrected form of the methods applied by Mr. J. A. Nelder, National Vegetable Research Station (Bonsall).
9. Fitting of constants and analysis of variance on some non-orthogonal data on herrings for the National Institute for Research in Dairying (Bonsall).
10. A study of foetal growth in the mouse, and an explanation of the inverse relation between number in the litter and

https://doi.org/10.23637/ERADOC-1-93 *pp 10*

size of young at birth (Healy, with Dr. Anne McLaren and Dr. D. Michie (Royal Veterinary College) (12.17)).

11. The relation of body build to the excretion of 17-keto-steroids and 17-ketogenic steroids in healthy young men (Dr. J. M. Tanner, Mr. R. H. Whitehouse, Mr. A. C. Edgson, Institute of Child Health, University of London, and Healy (12.24)).

12. Research on the effect of temperature on the yield of crops, and on Penman's potential transpiration theory. A paper is being prepared (Buck).

13. Examination of data relating to sulphur dioxide pollution of the atmosphere at Great House Experimental Husbandry Farm. Contrary to what has previously been reported in the literature, values recorded by the conventional SO_2 candle method seem to depend on wind speed as well as SO_2 concentration (Vernon and Mr. C. H. Mudd).

14. The effects of heating and lighting variation on the ripening of tomatoes (Gower with Dr. A. J. Cooper, Glasshouse Crops Research Station).

In addition, work on medical and other problems has been undertaken for the London University Institute of Child Health, the Imperial Cancer Research Fund, the Medical Research Council, the National Survey of the Health and Development of Children, the National Foundation for Educational Research, the Institute of Criminology, the Home Office and the Prison Commission.

Mr. N. E. G. Gilbert of the John Innes Institute continued his work on animal breeding problems in conjunction with the Animal Breeding Research Organisation; in particular, the analysis of data from pig progeny testing stations of the Pig Industry Development Authority was completed with the estimation of genetic components of variance and heritabilities. Latent vectors were used successfully to summarise the numerous genetic correlations.

Papers 12.14, 12.16 and 12.19 relate to previously reported investigations.

https://doi.org/10.23637/ERADOC-1-93