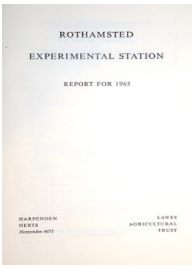
This work is licensed under a Creative Commons Attribution 4.0 International License.

Thank you for using eradoc, a platform to publish electronic copies of the Rothamsted Documents. Your requested document has been scanned from original documents. If you find this document is not readible, or you suspect there are some problems, please let us know and we will correct that.

ROTHAMSTED RESEARCH

Rothamsted Experimental Station Report for 1965

Full Table of Content



Statistics

F. Yates

F. Yates (1966) *Statistics ;* Rothamsted Experimental Station Report For 1965, pp 203 - 210 **- DOI: https://doi.org/10.23637/ERADOC-1-60**

This work is licensed under a Creative Commons Attribution 4.0 International License.

STATISTICS DEPARTMENT

F. YATES

M. J. R. Healy left to develop a Statistics and Computer Department for the new Clinical Research Centre of the Medical Research Council. Judith M. Stone and M. P. Vessey also left and K. E. Bicknell, J. A. Lewis and Alison Macfarlane were appointed. Marjory G. Morris transferred to the Entomology Department. Church returned from secondment to the United Nations Technical Assistance Board in Ethiopia.

Yates gave a paper at the Fifth Berkeley Symposium on Mathematical Statistics and Probability (12·15). Rees attended the Congress of the International Federation for Information Processing in New York, and Gower and Ross attended the International Symposium for Methods of Field Experimentation held in Halle (E. Germany), where Gower gave a paper (12·10). Leech took part in a training course on the evaluation of foot-and-mouth disease campaigns at the Pan American Foot-and-Mouth Disease Center (World Health Organisation) in Rio de Janeiro. Boyd spent two months in Rome as F.A.O. Consultant on the fertiliser requirements of Near East Countries, and attended a meeting in Paris of agricultural economists and scientists called by the Organisation for Economic Co-operation and Development to discuss interdisciplinary research in beef and milk production, crops and poultry. Anderson completed his year with the department as an Agricultural Research Council Scholar. Five other workers spent various periods in the department.

General

We are making good progress with the Orion computer. We now have a set of programmes providing excellent analyses of most types of replicated experiment; the improved service that we can give for such analyses is reflected in the increase this year of 40% in experiments analysed and 60% in variate analyses. We also have two powerful and flexible programmes for the analysis of surveys and similar material. One of these was successfully used to analyse the Agricultural Research Council and Ministry of Agriculture survey of calf wastage and husbandry, a very large and complex survey requiring 150,000 punched cards for recording the data. Once the initial problems of transferring the data to magnetic tape and doing preliminary editing had been solved, a series of analyses was made very speedily. The results show that many widely held beliefs on the causes of calf losses have little substance. Losses among calves moved from one farm to another were only one-fifth more than among calves not moved; 80% of all losses occurred on the farms where the calves were born. Total losses were estimated at 5·3% for the year; losses in the Jersey breed were about four times this figure. As 99% of calves were reported to have received colostrum as their first feed, lack of colostral antibodies could be a major cause of loss only if the quantity of colostrum was grossly deficient

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

for many of the calves or the first feed was witheld until some hours after the calves were born.

Our efforts to make available a large computer system for direct use by research workers at other institutes are also proving successful. The computer is being increasingly used in this manner, both by means of the general programmes we have developed and by special programmes written in Extended Mercury Autocode by the workers concerned. This decentralisation considerably lightens the burden on our own research staff.

Through the computer we are establishing closer relations with agricultural economists. Both the Advisory Economists and the Ministry of Agriculture's Economics Division are finding our survey programmes useful. Much work is being done on animal records, both from experiments and from surveys and progeny tests, as well as on experiments and surveys on agricultural and horticultural crops. A further extensive survey of fertiliser practice, particularly oriented to grassland, has been organised for 1966.

The Computers

The load on the Orion increased steadily, and reliability also improved (Table 1). Whereas in 1964 21% of the time when the machine was nominally available was wasted by faults in the central processor and by

TABLE 1

Orion performance, 1964–65

	1964		1965			
	Apr. –Dec.	%	Maintained	Unmaintained	Total	%
Useful time	1071	79	1543	374	1917	90
Computer faults	235	17	159	9	168	8
Restarts after faults	55†	4	35	4	39	2
Total	1361	100	1737	387	2124	100

† Estimated.

restarts, in the first six months of 1965 this dropped to 12%, and in the last six months to 8%. The engineers achieved this improvement through their increased familiarity with the machine, by more thorough investigation of all incidents under marginal conditions during normal maintenance, and through improvements in the fault-monitoring programmes. Fortunately most faults are transient, and we can consequently work late without engineers in attendance; all operators are trained to reload the Organisation and Monitoring Programme. The performance of the peripheral equipment also improved. The card-reader gave most trouble, and was once out of action for two days; this, of course, did not preclude running jobs not requiring the reader.

The increase in load (useful time) in 1965 was 34% over 1964 (April–December). The increase during the last half of 1965 over the first half was 40%. We expect this increase to continue. Late working (to 10.30 p.m.) is now frequently required, and averaged 2·1 hours per day over the last six months of 1965; in this we have had most excellent co-operation

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

by the operating and punching staff. In the near future it may well be necessary to organise regular two-shift working.

An additional 4,096 words of core store were fitted in December 1965. This will enable us to make better use of time-sharing. A line-printer is ordered, and should be installed by April 1966. This is badly needed, as printing results on flexowriters from paper tape is a slow process and often delays the return of work. The line-printer will also increase the capacity of the computer by shortening output time. We hope that additional magnetic tape decks may also be obtained. Permission has been given to install a graph-plotter, and for some form of Telex equipment; in the first instance a data-link will be established with the National Vegetable Research Station, which makes most outside use of the computer. A third paper-tape output channel and punch was added in December; this will be used for five-track output for the data-link, and for output for the graph-plotter.

TABLE 2

Use of computer by institutes, 1965 (% total accounting costs)

Institute	Development	Production	Total
Rothamsted Experimental Station†	37·1	14·5	51·6
National Vegetable Research Station	4·5	7·5	12·0
East Malling Research Station	3·8	1·8	5·6
Animal Breeding Research Organisation	1·3	0·9	2·2
National Institute of Agricultural Engineering	1·2	0·7	1·9
Glasshouse Crops Research Institute	0·7	0·8	1·5
Grassland Research Institute	1·1	0·3	1·4
Others (15)	2·4	3·7	6·1
TOTAL	52·1	30·2	82·3
Experiments (Production, Rothamsted programmes)			12·3
Systems			5·4
			100·0

† Survey and Experiments Programme accounted for 17·0% development and 3·2% production.

Table 2 shows the use of the computer by other institutes. This is based on accounting costs which are calculated (by the special programme used by International Computers and Tabulators for their commercial costing) from the information on such items as mill-time, elapsed time, store allocations, peripherals used, etc., provided by the Organisation and Monitoring Programme and recorded on the consol flexowriter tape. The total costings on a commercial basis were £180,000, of which £68,000 was production. It is not easy to relate these commercial costings to our actual costs, but from internal costing studies made by another Orion user we judge that they should be about halved. This gives an average cost per variate analysis in an experiment of about 7*s*., excluding programme development; with allowance for punching and verifying, the average cost of analysing an experiment is about £4. On a similar basis the Calf Survey (described below), which comprised about 150,000 cards, has so far cost about £3,500, including loading the data on tape, editing and correcting, testing of the job programmes and all tabulations so far made.

Research workers at various institutes are showing increasing interest in

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

installing self-recording instruments to provide information in a form that can be read and processed directly by a computer. We are concerned with the critical assessment of the value of such instrumentation in different circumstances, and with the most suitable form of recording. If many such instruments are installed, requirements for computer time will increase considerably. Ultimately it may be more convenient and efficient to process the records locally, but during development we may expect many demands to be made on the Orion by other institutes, in addition to Rothamsted.

The 402 computer was transferred to the Watford Technical College in March. The 401 computer was finally switched off on 9 July at a farewell party arranged by Elliott-Automation, which was attended by many of those concerned with its development, including Lord Halsbury (Managing Director of the National Research Development Corporation during the conception of the 401) and Mr. George Duckworth (present Managing Director). The 401 has been accepted by the Science Museum, where it will be permanently exhibited in 1966. To the end it continued to give reliable service.

Programming Developments

Programmes for the analysis of experiments. A programme for the analysis of general factorial designs deals with experiments with up to seven factors, which may be on successive splits, in randomised blocks or one or more Latin squares or quasi-Latin squares. There may be confounding, partial or complete, of the whole plot treatments, and some forms of split-plot confounding, and even such tiresome designs as balanced or unbalanced confounded $3^p \times 2^q$ designs and criss-cross designs are conveniently handled. For most types of design the correct analysis of variance is provided automatically, including estimates of error from high-order interactions in single replicate designs. Though this was not planned, the programme can analyse a 3^5 one-third fractional replicate. (Yates and Anderson)

Patterson is working on the problem of storing the results of long-term experiments on magnetic tape as they accrue, so as to produce summaries or more elaborate analyses when required. A plot-by-plot form of the general input for experimental designs (G.I.E.D.) from punched cards was organised by the National Vegetable Research Station Statistics Department. (Mr. G. Berry) Various more specialised programmes for experimental analysis were written by members of the department and other institutes.

Survey programmes. The parts of the Survey and Experiments Programme (S.E.P.) required for simple survey analysis are now working, and the programme has been used on various jobs. Experience in its use indicates that it is a very powerful and convenient language for survey analysis and other table operations. It will therefore be developed as opportunity offers, because it may prove to be a language of general interest for statistical work on the large machines with very fast central processors that are now being installed (12·10). (Gower, Simpson, Martin, Ross) In its present

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

form, however, it is considerably slower than the E.M.A. General Survey Programme (G.S.P.) described in the 1964 report (12·2), which is therefore used for large jobs. G.S.P. seems likely to be of permanent value for the Orion, so we intend to modify it and add extra features to make it more convenient for general use.

Fitting constants to multiway tables. This follows the general lines of the 401 programme for quantitative variates, but the constants are estimated by inverting the matrix instead of by the iterative method of Stevens. This enables standard errors and an informative analysis of variance to be provided. Constants can be fitted for all levels of each factor and for interactions between pairs of factors; constants representing linear functions of single factors, e.g. linear regression, or interactions between such functions for different factors can also be fitted. A linear function of one factor can also be fitted separately for each level of another. Up to seven-way tables can be handled. The basic data are at present read from paper tape, but provision will be made for reading tables produced by G.S.P. or S.E.P. and recorded on magnetic tape. The programme will also be extended to provide for the fitting of constants to quantal data, using the logit transformation in the same manner as on the 401. (Yates and Lewis)

Multiple regression. This is a flexible regression programme which permits any one or more of a set of variates to be nominated as dependent variates and calculates the regression of these on any one or more of the remaining variates. Extra independent variates can be added to the regression equation and alternative independent variates can be added in turn, or the best of a set of independent variates can be selected. Deletions from the regression equation can be made similarly. (Healy)

Maximum likelihood programme. This provides for the estimation of parameters by the method of maximum likelihood. The stages of the programme are: (1) setting up of likelihood functions appropriate to the data and the model; (2) maximisation of the likelihood by steepest ascent methods; (3) presentation of the results (or re-entry to stage (2)). The programme incorporates standard chapters for the more frequent problems and a flexible system for adding chapters to deal with special problems. Up to four parameters can be estimated. Standard chapters are included for probit analysis, negative binomial, asymptotic regression, Gompertz curve and logistic curve, and special chapters have been written for blood group gene frequencies and the double normal distribution. Further special chapters for concurrent regression, capture–recapture and the solution of second-order differential equations are being prepared. (Ross)

Probit analysis. A programme exists for the fitting of simple probit lines. (Healy)

Classification programme. This is similar to the classification programme already written for the 401. It calculates similarities between pairs of

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

species from taxonomic data and classifies the species in groups in a hierarchical (tree) classification. (Ross)

Experiments

The demand for routine analysis of experiments increased greatly (Table 3); 41% more experiments, involving 59% more variate analyses, than in 1964

TABLE 3.

Numbers of replicated experiments analysed in the department

	Number of experiments			Number of variates on computer	Variates per experiment
	By hand	On computer	Total		
1934	115	—	115	—	—
1951	437	—	437	—	—
1955	384	419	803	834	2·0
1957	98	1,253	1,351	5,041	4·0
1959	67	2,649	2,716	11,102	4·2
1961	89	2,862	2,951	15,184	5·3
1963	72	2,770	2,842	14,357	5·2
1964	88	3,383	3,471	18,054	5·3
1965	69	4,751	4,820	28,663	6·0

were analysed on the computers. The data and instructions for 74% of these experiments, comprising 61% of the variate analyses, were punched at Rothamsted. (Dunwoody and Rossiter) The tapes or cards for the remainder were prepared at other research institutes (12 in all) and sent to us to run; 18% of these, from five institutes, used programmes prepared by the institutes concerned. The average number of variate analyses per experiment for experiments punched at other institutes was 9·0, compared with 5·0 for those punched at Rothamsted; this difference is mainly attributable to these institutes having more experiments with horticultural crops.

Boyd and Mr. W. Dermott summarised the results of 125 fertiliser experiments on potatoes and investigated the relation of responses to soil types and other soil characteristics (12·7). Boyd also summarised the two Saxmundham rotation experiments which have been running since 1899 (12·14, 12·8). Boyd and Preece prepared a set of explanatory notes for the National Agricultural Advisory Service on the interpretation of the routine experimental analyses produced by the computer.

Because of the computer we get many requests both from within the station and elsewhere to assist in the analysis of large amounts of experimental data, often representing many years of work. Such analyses are almost always more troublesome than they appear at first sight, and we have still much to learn about the best ways of handling them.

Surveys

A large survey of fertiliser practice is planned for 1966; the preliminary organisation for this is already done. It is hoped that this survey will provide more detailed information about fertilisers used on grassland and their relation to sward use. Additional questions relating to grassland were

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

tested in a survey in 1965 by the National Agricultural Advisory Service of part of Cheshire. (Church and Hills) The relative use of fertiliser compounds of differing composition on different crops and types of farming was examined, using the data of the 1962 survey (12·6).

The analysis of the calf wastage survey made good progress. Some of the results that have already emerged are commented on above (General section). Further questions arising from these results are being investigated and a report will be issued. A second survey of bovine brucellosis by the Ministry of Agriculture was analysed on the 402 and 401. The results, from a small sample of the herds in Britain, showed that 10·1% of cattle give positive reactions and a further 11·4% give inconclusive reactions to the serum agglutination test, indicating that there are about half a million reactors of each type. The survey showed that most reactions could be attributed to vaccination, and only a small proportion of all reactions seem to be caused by actual infection. The results of an investigation by the National Agricultural Advisory Service into the relation between udder health and solids-not-fat were also analysed. (Leech and Vessey)

We are co-operating in the analysis of much other survey and census material, including the survey of diseases in seed potatoes with J. M. Hirst, Plant Pathology Department, various investigations for the Economics Division of the Ministry of Agriculture, and a national economic survey of the wheat crop organised by the Farm Economics Branch of the School of Agriculture, Cambridge. (Church and Hills)

A small survey planned for the Egg Marketing Board on the causes of farm-to-farm variation in the percentage of cracked eggs showed clearly that the cracking occurred before the eggs were removed from the nest and was not correlated with shell thickness; poor design of battery cages seems a likely cause, but this requires further investigation. (Leech and Vessey)

Commonwealth and Overseas

We arranged with the Ministry of Overseas Development to transfer the work on tree crops to East Malling Research Station, where it will be in the charge of Dr. S. C. Pearce, who has more experience of the statistical problems of experiments on such crops than we have. The Ministry have provided an additional post for this work. East Malling, of course, has access to our computer.

Ross again analysed the cotton progeny trials for Namulonge, Uganda, and some discriminant function work on the results of these trials is planned. Preece analysed an extensive set of South Korean fertilizer trials on rice and barley for Dr. Wang Keun Oh, who was working in the Chemistry Department, and advised him on the design of future trials. Various other experiment analyses and design problems were dealt with.

Other Work

Preece investigated designs for the superposition of a new set of treatments on an experiment with previous treatments and two-way elimination of soil or other heterogeneity (12·11). He also classified various sizes of

https://doi.org/10.23637/ERADOC-1-60

This work is licensed under a Creative Commons Attribution 4.0 International License.

Youden rectangles; this led to a method of generating certain designs of the above type (12·12). Yates examined various logical points arising in the design and analysis of experiments, and in particular condemned the distinction between fixed and random effects models, which have a certain vogue in America, and the use of non-factorial response surface designs in place of ordinary factorial designs in agricultural field trials (12·15). Patterson is doing more detailed research on the last problem, and on other problems in experimental design. Gower made progress in various problems connected with classification techniques and multivariate analysis.

Discriminant function analysis was successfully applied to measurements on the shoulder girdles of primates (12·3), on barnacles (12·4) and on white-toothed shrews (12·9). Vessey assisted in the analysis of the effects of social and racial factors on human pregnancy (12·5). Genetic parameters of British Landrace pigs were estimated from data from five progeny testing stations (12·13).

https://doi.org/10.23637/ERADOC-1-60