Norwegian Journal of Agricultural Sciences

Supplement No. 22 1996

A pioneer in the use of mathematical statistics in Norwegian agricultural research

A tribute to Øivind Nissen

Lars Strand (ed.)

Ås Science Park Ltd., Norway

...VAL OF AGRICULTURAL SCIENCES

The Norwegian Journal of Agricultural Sciences is a forum for publishing Norwegian research with international interest within the following areas: Aquaculture, Animal Science, Soil and Water Science, Agricultural Engineering and Technology, Natural Resources and Environment, Food Technology, Crop Science, Forestry, Economics and Society Planning.

PUBLISHER

The journal is published by the Ås Science Park Ltd., Sagabygget, N-1432 Ås, Norway. Norwegian Journal of Agricultural Sciences (ISSN 0801-5341) is published four times a year, each issue containing approximately 100 pages. Four issues comprise one volume. Annual subscription NOK 500,-. Supplement issues can be ordered separately through the publisher.

REFEREE

All papers submitted for publication will be sent to competent international referees.

CORRESPONDENCE

All correspondence, editorial or otherwise, should be adressed to Ås Science Park Ltd., Sagabygget, N-1432 Ås, Norway

The drawing on the cover: Kjell Aukrust

ISSN 0802-1600

Norwegian Journal of Agricultural Sciences

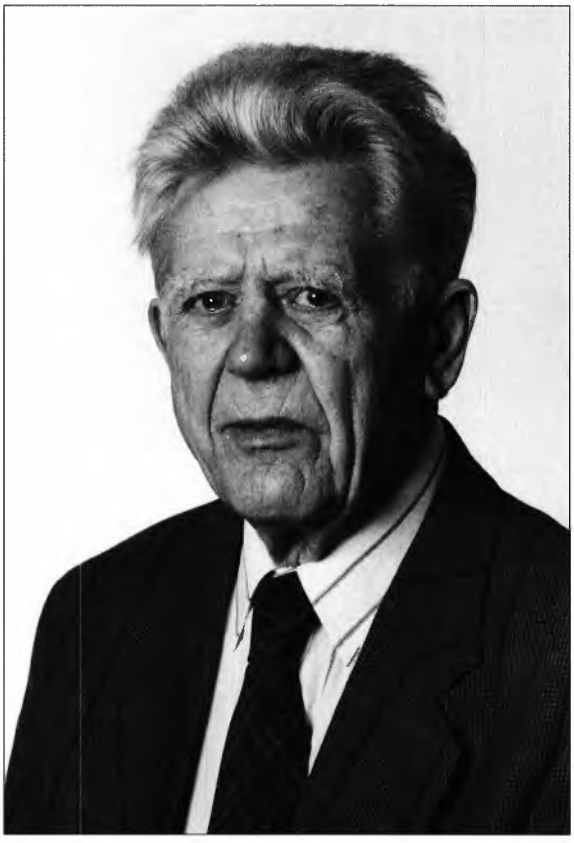
Supplement No. 22 1996

A pioneer in the use of mathematical statistics in Norwegian agricultural research

A tribute to Øivind Nissen

Lars Strand (ed.)

Ås Science Park Ltd., Norway



Preface

Many of the employees at the Agricultural University of Norway have through the years made the name of our university known over great parts of the world. One of them is Professor Øivind Nissen who has been a pioneer in Norway in the field of statistics and research methodology. As a tribute and recognition to Professor Nissen a seminar was arranged in May 1995. Norwegian and foreign scientists were giving lectures showing the importance of Professor Nissen's work. Internationally Professor Nissen is above all known for his MSTAT-programme which is used in more than 110 countries.

You will find all the lectures given at the seminar in this publication. Reading it you will have an impression of a very active, enthusiastic and competent researcher. He is still at work at an age of nearly 86.

On behaf of the Agricultural University of Norway I would like to take this opportunity to thank Professor Nissen for his achievement in improving the data analysis in the agricultural research process. In this way he has been an outstanding representative of our university.

I would also like to thank Norwegian College of Veterinary Medicin, Planteforsk and Statkorn for their financial support to the seminar.

Nils Kolstad
Rector

Content

Øivind Nissen – Contributions to Science

KÅRE RINGLUND
Department of Horticulture and Crop Sciences,
Agricultural University of Norway, Ås, Norway

The human population as well as populations of other animals, fish, plants or simply sets of observations, can be described by two parameters, the mean and the standard deviation. Most individuals in a population have characteristics around the population mean, but some individuals are extremes on either side of the mean. Øivind Nissen is one of the extreme positive deviations from the mean of the human population. He has a unique ability to understand and solve problems. Almost 20 years after he reached retirement age, he continuous to contribute to problem solving in many areas of science.

Øivind was born in Karasjok 3 February 1910. He completed his high-school exam at the very young age of seventeen, two years ahead of the population mean, and he continued his education at the agricultural school Gibostad and at the Agricultural University of Norway where he graduated with excellence in 1933. In 1934 he supplemented his agricultural education with an exam in botany at the University of Oslo.

In his early professional career Øivind Nissen worked as Research Fellow on inheritance and breeding of forage grasses. Already at this stage he was genuinely interested in experimental design and data analyses, and in his publications he always emphasized experimental methodology. He was appointed professor of Crop Science in 1951, and served as Chairman of the Department of Farm Crops through 1964. During this time Øivind Nissen continued his research and teaching in forage grasses, root crops and crop science in general. He also lectured in experimental design, and thus continued a strong tradition established by his predecessors, Bastian Larsen and Knut Vik. His interest in experimental design and data analyses brought him into what was to become his second, and probably most important career, as adviser on experimental design and data analyses and as developer of computer programs for statistical analyses of data in general, and of data from field experiments in particular.

Øivind Nissen was the head of the Center for Experimental Design and Data Analyses (the FDB Center) funded by the Norwegian Council for Agricultural Research, from its establishment in 1965 until he retired in 1977. Since then he has been working with program development at the Department of Crop Science, and since 1993 at the Department of Horticulture and Crop Sciences.

In 1948-49 Nissen was visiting professor at Pennsylvania State College where he lectured in experimental design and acted as consultant in analyses of experimental data. During the 1950's he visited British and American research institutions that were pioneers in the use of computers for data analyses. Øivind was probably the first Norwegian Agricultural Scientist who was familiar with and analyzed data using the two-digit system. In 1961 he developed the first programs for analyses of data from field

experiments. These programs were developed on a FACIT mainframe computer at the Meteorological Institute in Oslo. Inputs and outputs were via paper tapes, and the programs were written in hexadecimal machine codes. When the FDB Center was established, more sophisticated programs were developed on different IBM machines with data storage on discs and magnetic tapes. The program package FDBPRO was used extensively by Norwegian agricultural research institutions and others during the 1970's and the early 1980's. This was a unique program for its time combining a data manipulation system with statistical analyses and printing of results in different forms.

Most of the routines in FDBPRO were built into Øivind Nissen's program package for microcomputers, MSTAT, which was written in interpretative BASIC. In addition this program package contained a set of programs for development of plans, maps, labels and field books for field experiments. A system for keeping track of plant breeding materials from generation to generation, and for printing of field books and labels for such materials was also included. This program package has been internationalized in cooperation with Michigan State University and has obtained very wide use.

The latest program package that Øivind Nissen has developed is NM, or the english version ENM, which is written in Turbo Pascal. This is a menu driven data system containing all the elements from MSTAT and a number of new programs.

It is not the purpose of this short intervention to describe in detail the development and use of Nissen's computer programs. It is rather an effort to pay tribute to the man behind this very important contribution to Norwegian and to international agricultural science. As one of those who has had the privilege of having Øivind Nissen as adviser through a graduate study, I would like to emphasize his ability to clarify problems. Øivind is not only an expert in analyzing data, he is also an expert in discussing the relevance of the questions asked. In fact, when he is asked how to solve a problem, he always spend some time clarifying the questions asked before suggesting solutions.

It is a great asset for a University to have individuals like Øivind Nissen on its staff and as contributors without pay after retirement. For the Department of Crop Science he has been a valuable contributor in many ways. In addition to his statistical and data manipulation programs, he has also developed programs for accounting, literature databases, and before the time of Word Star, Word Perfect and Word, he even wrote a simple word processing program using the line editor of the microcomputer ABC80. I would be remiss if I did not mention his contributions to a positive social atmosphere at the Department through his short and well composed speeches and through his very pleasant personality.

Your colleagues thank you, Øivind Nissen, for your continued interest and help in solving our data problems.

Make the numbers tell their story

DAVID J. FINNEY
Emeritus Professor of Statistics, University of Edinburgh, Edinburgh, England

"So do not write with your hand except that which will delight you if you see it on the day of judgment." (Translation of inscription, inlaid in silver and gold, on a pen box in The National Museum of Kuwait.)

Every research scientist has a duty to make his findings add to the store of human knowledge. Effective publication on observational or experimental science requires conclusions to be so presented that a qualified reader can compare them with those known to him from other sources. This implies unambiguous statement of all methods employed: every reader should expect to be told what an author has done, precisely and in language so comprehensible as to facilitate adoption of similar procedures in his own research Responsibility for maintaining these standards rests with journal editors and their chosen referees, who should endeavour always to minimize the frequency with which a reader needs an inspired guess or an ingenious personal inference in order to determine meaning! I stress the need for unambiguous terminology. I comment on benefits and dangers from statistical software packages, and on avoidable flaws of presentation. Excessive emphasis on significance testing is still too common. After suggestions on the planning of tables and diagrams, I emphasize the practical difficulty of choosing the statistical method appropriate to a specific problem. Finally, I mention ethical considerations that relate to how a scientist tells stories about numbers.

We meet today to honour Øivind Nissen for his long career and his distinctive contributions to Norwegian agricultural research. I am not competent to review that career, or to give a comprehensive account of his own research. Having known him as professional colleague and friend for 40 years, I think of him especially as a communicator, concerned to be himself understood, and to help others to understand, all topics on which he writes. I think he has always seen the place of statistical science in biological research as aiding logical appreciation, not as an extra torture for the biologist. I doubt whether Øivind has ever shared my presumption in lecturing others on how to communicate statistical ideas effectively. I believe that any seminar such as today's should be an occasion for open exchange of ideas: I propose to be deliberately provocative, in the hope that any who hold contrary views will express them freely.

Research in biology is more difficult and more complex than in the purely physical sciences. Any biological research that is related to agriculture or to forestry or to medicine is necessarily interdisciplinary: the entities and concepts that are studied refuse to be divided tidily into distinct compartments of botany and zoology, biochemistry and genetics, bacteriology and pathology, and the like. A scientist who regards himself as primarily an agronomist or an animal nutritionist may require great care in his writings if he is to avoid solecism when he refers to methods and ideas taken from other specialist fields.

In the analysis and interpretation of quantitative measurements, counts, or other records, modern biology increasingly uses sophisticated statistical techniques; these may involve a scientist in use of procedures for which he lacks either training or inclination. His papers may be submitted to journals where neither editor nor referee has the ability to detect inappropriate statistical techniques, or to seek amendment of forms of presentation that are defective in identification of methods used, or in logical justification for and clear exposition of quantitative conclusions.

Terminology

Every science requires special names for its concepts, procedures, and methods, often adapted from colloquial speech or manufactured from a classical language. Failure to distinguish between specialized and common meanings is a source of confusion in any interdisciplinary writing, especially where mathematical and statistical ideas are involved (Finney, 1994b). I envy chemists. The systematic and readily expandable nomenclature created by Lavoisier is now so universally accepted that every scientist, whatever his special field, is expected to use chemical terms precisely, without intrusion of meanings that might be natural in his daily home life. Remember such words as 'salt', 'base', 'radical', that have technical meaning in chemistry yet with colloquial meanings that are usually clarified by context. I have difficulty in developing my theme without conveying an impression that I assume English to be the only language for scientific communication. That is not so. I illustrate from the one language that I know well: I urge you who listen to look for comparable examples in Norwegian.

Outside strict technical writing, ambiguities can arise even with chemical nomenclature. I once attended a committee discussion on soil deficiencies, at which a leading sugar beet farmer displayed his confusion between magnesium and manganese! In any account of the planning of an experiment or of a sampling study, the reader should be able to trust the word 'random' and its derivatives. This is not merely a point of literary style. The supporting logic of conclusions may be unreliable if the reader cannot be certain that every mention of randomization or of random selection relates to explicit use of a process designed to be strictly random.

Although 'probability' is a familiar word, acceptance of its measurement on a scale from 0 to 1 is perhaps less general. Consequently, journalists and others often use an alternative expression of quantitative uncertainty in terms of "odds". Are odds more widely understood than probability? This may be true among gamblers, but the word has dangers when used in relation to exact quantitative reasoning. In colloquial speech, the words "chance", "odds", "probability", "likelihood", and even "risk" are often used as though synonymous, but this habit should be avoided. Too often, the vital words "for" ("in favour of") and "against" are omitted, or are left to the reader's imagination. In strict statistical usage, the statement: "The odds that event E will occur are m to n", should mean: "In a very long series of independent trials, event E will occur more frequently than not in the ratio m:n". The statement can be translated exactly as:

"The probability of occurrence of E is m/(m+n)", and this is the safer form of words for scientific writing.

In an article about a proposed

experiment in space, I read some months ago: "NASA calculates that there is less than a million-to-one chance of a mishap"; did the writer really mean that a mishap was almost certain, or that the odds against a mishap were at least a million to one?

I do not assert the right of statisticians to give special meaning to a common word and thereafter to demand that it be reserved for sole use in that sense. For example, I have read criticism of statisticians who confuse engineers and others by pedantically insisting that 'a common average' be called 'an arithmetic mean'. Is this any grosser pedantry than a chemist's renaming of caustic soda as sodium hydroxide? I believe there to be good reason for this kind of insistence, as long as there is recognition of the right of any author to use a word differently provided that he declares explicitly what he is doing.

Software

Fifty years ago, biologists often refused to use what a professional statistician then advised as the optimal method of analysis because of the difficult and time-consuming computations. Today, personal computers and sophisticated software packages have done much to remove such inhibitions. A computer is now normal equipment for many biologists. Although neither reliable hardware nor clever software can convert poor data into good science, at least they destroy all argument that the labour of statistical arithmetic excuses adoption of an unsuitable analysis of data.

The weighty manuals that accompany a software package are rarely adequate substitutes for familiarity with underlying principles. "Even a good and thoroughly tested statistical package can be grossly misused by a scientist whose experience does not fit him to choose a method appropriate to his present problem" (Finney, 1995b). He may be led into undue complexity or even absurdity by misunderstanding assumptions and conditions implicit in the package. His subsequent publications may include symbols and phrases copied from computer output, in the belief that they are standard statistical idiom but that are in fact peculiar to an idiosyncratic software writer. On these matters, the documentation of statistical software needs much improvement.

An editor or referee who fails to detect irregularities of notation and terminology may contribute to future confusion. A year ago, I first tried to use that popular package MINITAB. Analysis of a simple set of data caused my screen to state: "Test is significant at $p=0.0000$"! This extraordinary output may account for statements I have seen in print that "$p \leq 0.0000$", or even "$p < 0.0000$". The explanation may lie in a rounding of p to four digits, but I see no excuse for software writers providing absurd phrasing that a user may believe to be authentic statistical terminology. My quoted MINITAB output illustrates a further flaw, the failure to distinguish between "test" and "test statistic".

Searle (1989) and Dallal (1990) have expressed similar concern about the quality of statistical software. An author whose statistical analyses have depended substantially upon a particular package should identify that package with the detail that is customary in respect of special materials or instruments used in a research project. A group of editors of medical journals recommended: "Describe statistical methods with enough detail to enable a knowledgeable reader with access to the

original data to verify the reported results" (ICMJE,1988). This is a desirable standard for all biological publication.

Inform and clarify

I have elsewhere (Finney, 1993) urged that the editor of a journal in which many authors report conclusions dependent upon interpretation of statistical analyses should inform authors that, without explicit definition in the text, certain symbols will be accepted as having their usual statistical meanings. Such symbols might be t, r, b, F, χ^2, p, and s.d., s.e., and d.f. A suitable code might include:

"b: Simple linear regression coefficient of y on x, where x, y have already been clearly identified; b may be modified to $b_{y.x}$ if the variables need to be made explicit. The word 'slope' is not an acceptable synonym for 'regression coefficient'" (Finney & Harper, 1993)

To persuade editors not merely to formulate such a code but to adhere to it will not be easy; authors can help their readers by voluntarily adopting the conventions, and therefore explaining clearly any different uses of symbols. The code should insist that, if numerical values of the common statistics t, r, F, χ^2, need to be stated, their degrees of freedom must accompany them, perhaps as $t_{[17]}$, $\chi^2_{(28)}$, $F_{3,27}$ and the like; a reader should not be expected to infer these numbers from the text.

Too many published papers use symbols inexactly or inconsistently. The symbol 'n' or 'N' is often found in contexts that require the reader to guess whether number of observations or number of degrees of freedom is being stated. Unless totally ignorant of chemistry, no scientist would, without stating his convention, use 'Ca' as identifier of data collected in California, yet too often the user of a statistical technique does the equivalent!

The ease with which a computer can operate to great arithmetical accuracy may encourage biological nonsense. Arithmetically correct computer output from data on 70 organisms might show the mean weight as 42.571429 g, but to publish this figure is foolish. Software readily available today can produce apparently very exact probabilities in significance tests. If computer output for a test gives "p=0.0000327016", before publishing this an author should consider whether implicit assumptions about the nature of his data justify anything more seemingly exact than "p=0. 0003" or possibly only "p < 0. 005". Some editors, especially of medical journals, appear to demand these strings of untrustworthy digits.

To most non-mathematical scientists, natural logarithms were once a mystery. The revolution in computing facilities has made them easily usable by all: pocket calculator and computer give them more directly than logarithms to base 10. Many scientists no longer use base 10, except perhaps with conventional reference to "orders of magnitude" or to "pH". I would like to see the clumsy "ln" or "$\log_e$" abandoned, and "log" accepted as always referring to a natural logarithm.

Remember also that, in telling the story of inferences and conclusions from a piece of research, detail of statistical operations must not obscure the view of biological meaning. Except in exposition of new methods, a definitive publication seldom needs to display features of the arithmetic such as tables of analyses of variance or separate terms that sum to a χ^2. Close attention should be given to the source and nature of data, to all conceivable sources of bias, and to objectives

consistent with the design of a project. When due account of these has been taken, in my experience multiple comparison procedures rarely answer any realistic question; these should not be used unless the relevance of their inherent assumptions is evident.

I have never seen a set of data to which I would apply a multiple comparison test. I have lacked interest in them because I have never been satisfied that the hypotheses proposed for testing were of real importance or have useful meaning. Some early methods seemed to be based upon the assumption that, if an experiment studied 6 different rates of fertilizer application, then the true treatment means would fall into k (number unknown, but <6) groups in such a way that, within any group, the parametric values for the means are identical. If this were so, to estimate k and to assess which treatments belong to each group might be an interesting theoretical problem, but is such an assumption ever plausible?

Even percentages may produce confusion. Some writers forget the convention that a percentage change is always to be assessed relative to the basal value. Consider a quality (perhaps a disease) present in 65% of a population. A year later, this rate of occurrence has become 40%. Is this to be stated as a 25% reduction, or as a 38% (=100x25/65)? A careful scientist should so write as to leave no doubt of his meaning. Very recently, in a semi-popular article on world food needs, I read: "...since 1974 the rate of population growth in Indonesia has fallen by 17% to 1.6% per annum..."! A related clumsy idiom, not unknown in scientific writing, is a statement that a specified character was "three times less common in population A than in B"; I guess this to mean that the frequency in A was one-third of that in B, but why not say so? I similarly deplore: "During this period, the death rate from XXXX has been more than halved".

Statistical significance

Few biologists find the logic of statistical significance tests easy. When first introduced, these tests were a needed corrective to: "*Post hoc, ergo propter hoc*", the uncritical assumption that any numerical difference between mean measurements for two differently treated groups of animals could properly be termed an EFFECT of treatment. Unfortunately, this introduction has led some to teach, and many to write, as though the chief objective of statistical analysis is the testing of significance. In colloquial English, the word "significant" can mean no more than 'interesting'. Therefore, a biologist will be wise to insert "statistically" if he refers to a formal statistical test; he will keep in mind that significance at probability 0.05 does NOT mean: "The probability of the effect being real is 0.95"!

Agricultural journals are perhaps less afflicted than medical by editorial pressures to accompany every quantitative statement by a test of statistical significance. This practice encourages publication of silly statements such as: "Significance was only 0.33", or: "The difference between the two means is slightly statistically significant ($p = 0.046029$)", or: "Passed the χ^2 test for Normality". These statistical solecisms may be transcriptions of output from software the writers of which could program statistical arithmetic but lacked understanding of its meaning. Every teacher of statistics must have emphasized to students that a non-significant departure from a null hypothesis is logically far short of demonstrating truth

for the null hypothesis. One common form of words that always amuses me is:

"This test failed to achieve statistical significance"; I immediately visualize a vast numbers of tests competing in the great *SIGNIFICANCE OLYMPICS*, in which only 5% will receive a Medal for success! Neither reader nor science benefits from referees who fail to require amendment of statistical illiteracy as great as I have illustrated.

Significance tests can be important to the interpretation of data. Often, a close look at the logical content of the implied null hypothesis will suffice to convince that its trivial nature removes all real interest from the exact form of test! Undue emphasis on significance testing can distort the message obtained from a research project. Much purposeless testing is encouraged by poor textbooks and by some software packages.

Tables and diagrams

Effective communication of statistically oriented research findings demands thoughtful planning for each table and diagram. It may have any one of several distinct functions: is it included as support for statements in the text, as visual evidence helping to clarify a complex argument, as in itself part of a definitive summary of findings, or perhaps as archival material for future reference? Even the simplest table may, if ill-designed, fail in its duty of having effective impact upon a reader. Modern constraints on journals have made publication of the full data from a piece of research a rarity. In the past, the availability in libraries of journals containing complete data from many fields of research has been an important aid to those who develop new methods for statistical biometry.

Biological publication must deal with many matters that are visually observable: the common practice rightly is to illustrate these by photographs, and by manually drawn diagrams that emphasize anatomical features. For presenting exact quantitative information, diagrams are usually less suited than tables, yet they may intensify dramatic effect. Sophisticated modern software has made graphical representation of numerical data much easier than when it required tedious operations of measuring and drawing on graph paper. This facility may not only aid preliminary scanning of data and steps towards choice of a definitive method of analysis, but may also be used in preparing effective diagrams to illustrate a lecture or a paper for publication. Every numerate scientist can enjoy and benefit from the superb book by Tufte(1983). Dangers exist (Finney, 1986a, b).

Whether a table is to be an integral part of the text, for support and clarification of the scientific argument or as information for archiving and future reference rather than consecutive reading, its title and legend must clearly state exactly what it contains, leaving no uncertainty about units of measurement and requiring no guesswork by the reader. Row and column labels must be concise but exact, and a reader should have no difficulty in relating the contents to the author's description of his experiments and data acquisition.

Similar considerations apply to diagrams that illustrate statistical features of research. A deplorable practice is now common in many journals, especially in medicine and in ecology. Numerical tables are totally excluded. Instead, all quantitative features of results, whether mean yields, total or mean numbers, or other interesting summarizers of information are represented by block diagrams, in

which each value is shown as a vertical rectangle of appropriate height, sometimes elegantly shaded or drawn to suggest a three-dimensional perspective. A vertical scale at the extreme left may be included. Because of their visual impact on me, I refer to these as Manhattan Skyline diagrams (Finney,1995a). Editorial ignorance often fails to prevent an author from calling them "histograms", a name that is recognized in statistics only for diagrams that show relative frequencies. A popular software package produces these well drawn diagrams and encourages their use as supposedly more easily understood by readers.

Manhattan diagrams may be useful in popular science writing or in newspaper articles, but I question the ethics of employing them as the sole quantitative summaries in definitive papers for major journals. On the skyscraper roofs, vertical radio masts are often shown and named as "error bars"; the legend may explain these as standard deviations, or standard errors, or confidence limits, but sometimes leaves them undefined. This is not a generally recognized convention for representing variability.

Author and editor may regard a Manhattan diagram as dramatically effective in communicating the broad character of results. My sympathy lies with a reader who wants a numerical value for comparison with his own work or beliefs; in order to extract something as simple as a statement that the mean for treatment D is 20% greater than that for treatment A, he must take a ruler to the published diagram, measure the height of each rectangle, and apply the factor that converts his millimetres into units of the vertical scale. To this tedious and inaccurate task, he may need to add guesses about the precise meaning of the radio masts, the degrees of freedom, and whether a pooled estimate of standard deviation or separate estimates from replicates of each treatment have been used. Unless there is near certainty that no responsible reader will ever need such information, I think it improper that research, possibly financed by public funds or by a major Foundation, is reported only in a diagram that conceals the facts: no reader should be expected to spend time reconstructing values which the original author would have had in front of him when drafting his paper.

Oral presentation

Special considerations may arise when tables and diagrams are to be used in oral presentation, perhaps as a lecture to a scientific meeting or as part of a course of instruction. If the aim is didactic, it is obviously important that the best possible examples be shown. If there is to be any display upon a screen, the speaker must remember the limited time of exposure of even his most informative tables or his most aesthetically pleasing diagrams. I have great personal dislike of overhead projectors, dating from my only use of one in 1949. They have merits for displaying diagrams, but they are ill-suited to showing equations or tables: have we not all heard a lecturer say: "...and now equation (6) leads to the conclusion...", where (6) is something that was on the screen five minutes earlier but has now disappeared. Either as speaker or as hearer, I prefer the old-fashioned chalk board, if possible supported by a typed or printed handout containing all equations or tables to which repeated reference may be needed.

Choice of method

For the biologist who has had no formal training in biometric method, the computer revolution has in one respect made life more difficult. Faced with measurements or other numerical records from his latest research, he may have easy access to one or more of several popular packages, such as BMDP, GENSTAT, GLIM, MINITAB, SAS, SPSS and others; each offers him choice among a number of statistical procedures but may give him no indication how to choose one appropriate to his present data. A statistician may have advised him to select his method of analysis before beginning his research, but that is not always practicable.

Statisticians have been more successful in the invention of new methods than in giving guidelines on their use. Neither mathematical theory nor rigid rules based upon theory will guarantee avoidance of disastrous faults. This is not the place for attempting any thorough discussion, beyond emphasizing once more that choice and exact specification of method, with due regard to any assumptions implicit in its use, may vitally influence conclusions. Here is the most difficult problem that confronts a biologist whose training and experience have not given him the necessary skill to decide. As one who is frequently a referee for papers submitted to biological journals, I am very conscious of occasions when I would like to have advised an author to start analysis again with a different method. Too often, I suspect that a biologist has been led astray by a notion gleaned from some inferior textbook, or from publicity associated with a piece of software, that a list of measurements from an experiment, classified according to treatments or other factors, suffices to determine the appropriate statistical analysis. The result may be nonsense.

Riley (1994) has recently presented some general guidance in combination with recommendations on presentation of results. A few years ago, I published a series of papers (1988a-1989c) on related topics that seemed to me to have escaped the attention of writers of textbooks. Several writers on medical statistics have published papers relevant to this matter, notably Altman et al. (1983), Bailar & Mosteller (1988), Gore et al. (1977), and the awful warnings of Mills (1993); all of these contain lessons that can benefit agricultural science.

Scientific ethics

As indicated in Section 6, presentation of conclusions from biological data, whether analyzed by professional statisticians or by biologists who undertake their own statistical work,can raise issues of scientific ethics. In all publication of quantitative research, an integral part of the story to be told is a truthful and accurate report on the source and handling of relevant numerical data. A biologist may not always easily recognize what information may prove important to some readers. Rarely if ever can legitimate confidentiality justify secrecy in respect of method of data collection or method of statistical analysis. Exact identification of a software package used, unless self-evident, is very desirable. He who analyzes his own data accepts for himself all related responsibilities. I have addressed these matters more thoroughly elsewhere (Finney, 1990, 1991, 1994a).

With increasing sophistication of laboratory equipment, a manufacturer may incorporate into an instrument a chip programmed to undertake routine statistical analysis of the data it will produce. If

the implementation of the program is regarded as commercially confidential, details on which he is not informed constrain the user of the instrument and may lead him into an analysis not conforming to the best advice from statistical science. This last point can be especially relevant to the handling of outliers, always a difficulty for statistical practice. A software writer can easily incorporate into a preliminary scan of data as they are input an automatic rule for rejecting outliers based upon chosen limits of plausibility. However good the software in other respects, this does nothing to remove an ethical problem; it amounts to arbitrary interference with interpretation of the data actually recorded. Any resulting publication must report what has happened and exactly what action was taken. In a medical context such considerations may raise important ethical issues

References

Altman, D.G., S.M. Gore, M.J. Gardner & S.J. Pocock 1983. Statistical guidelines for contributors to medical journals, British Medical Journal, 286, 1489-1493.

Bailar, J.C. & F. Mosteller 1988. Guidelines for statistical reporting in articles for medical journals: Amplifications and explanations, Annals of Internal Medicine, 108, 266-273.

Dallal, G.E. 1990. Statistical computing packages: Dare we abandon their teaching to others?, American Statistician, 44, 265-266.

Finney, D.J. 1982. Statistics and statisticians for agricultural research, Inaugural Address, International Conference on Frontiers of Research in Agriculture (Calcutta), 5-9.

Finney, D.J. 1986a. On presenting tables and diagrams, Scholarly Publishing (Toronto University Press), 17, 327-342.

Finney, D.J. 1986b. Tables and Summaries: will they help or hinder? Journal of the Indian Society of Agricultural Statistics, 38, 60-70.

Finney, D.J. 1988a. Was this in your statistics textbook? I. Agricultural scientist and statistician, Experimental Agriculture, 24, 153-161.

Finney, D.J. 1988b. Was this in your statistics textbook? II. Data handling, Experimental Agriculture, 24, 343-353.

Finney, D.J. 1988c. Was this in your statistics textbook? III. Design and analysis, Experimental Agriculture, 24, 421-432.

Finney, D.J. 1989a. Was this in your statistics textbook? IV. Frequency data, Experimental Agriculture, 25, 11-25.

Finney, D.J. 1989b. Was this in your statistics textbook? V. Transformation of data, Experimental Agriculture, 25, 165-175.

Finney, D.J. 1989c. Was this in your statistics textbook? VI. Regression and covariance, Experimental Agriculture, 25, 291-311.

Finney, D.J. 1990. Statistical data - Their care and maintenance, Indian Society of Agricultural Statistics, Bull. no. 1.

Finney, D.J. 1991. Ethical aspects of statistical practice, Biometrics, 47, 331-339.

Finney, D.J. 1993. On presenting the results of statistical analyses in scientific journals, Biometrical Journal, 35, 499-509.

Finney, D.J. & J.L. Harper 1993. Editorial code for presentation of statistical analyses, Proe. R. Soc. Lond., β254, 287-288.

Finney, D.J. 1994a. Professional ethics and the statistician: A basis for discussion, Amstat News, no. 210, 14-15.

Finney, D.J. 1994b. On biometric language and its abuses, Biometric Bulletin, 11 (4), 2-4.

Finney, D.J. 1995a. Statistical principles of met-analysis, Journal of Clinical Epidemiology, 48, 87-103.

Finney, D.J. 1995b. Statistical science and effective scientific communication, Journal of Applied Statisties, 22, (22, 291-306).

Gore, S.M., I.G. Jones & E.C. Rytter 1977. Misuse of statistical methods: critical assessment of articles in the B.M.J. from January to March 1976, British Medical Journal, i, 85-87.

ICMJE (International Committee of Medical Journal Editors) 1988. Uniform requirements for papers submitted to biomedical journals, Annals of Internal Medicine 108, 258-265.

Mills, J.L. 1993. Data torturing, New England Journal of Medicine 329, 1196-1199.

Riley, J. 1994. Aspects of statistical technique and presentation, Experimental Agriculture, 30, 361-394.

Searle, S.R. 1989. Statistical computing packages: Some words of caution. American Statistician, 43, 189-190.

Tufte, E.R. 1983. The Visual Display of Quantitative Information, Connecticut USA, The Graphics Press.

The Role of Computers in International Development

RUSSELL D. FREED
200 International Centre, Michigan State University, East Lansing, Mi. 48824 USA

"There is today a vast technological fault line. On the one side are the fast growing, adolescent, undercapitalized, undereducated societies; on the other the rich, technologically inventive yet demographically aging populations. The greatest challenge that global society faces is preventing this fault line from erupting into a world shaking crises." (Paul Kennedy Herald Tribune July 23-24,1994.)

Today the world is struggling with three challenges which appear to show no signs of slackening: 1) population growth, 2) poverty and 3) environmental degradation. The population today is 5.6 billion people. One hundred million people are added each year – 93% of these are in the LDC countries. Tomorrow morning there will be 270,000 more mouths to feed. Food demand will double by 2020 and triple by 2050.

In 1995, 12-14 million children will die because of undernourishment. That translates to 35,000 children dying each day. In addition to those dying, undernourishment will have tremendous impact on societies because of (1) neurological damage due to low protein, (2) poor muscular coordination; (3) lack of energy; and (4) poor immunulogical system development.

Poverty haunts much of the world. One fifth (1.1 billion) of the world has a household income of less than 1US dollar/day. One billion people do not have access to clean water. Two billion people do not have electricity or adequate sanitation.

Population, agriculture and industrial pressures have stressed the Global environment. Peasants fighting their constant cycle of despair are often using fragile lands for food production. Food production comes before saving the environment.

When we look at the three challenges – population growth, poverty and environmental degradation, we see that agriculture plays a pivotal role in each of these challenges. Agriculture provides food and fiber for the people. In many LDCs, 50-90% of their economies are based on agriculture. Agriculture is the foundation for the employment as well as income generation. Agriculture is the engine of growth for the economies of the LDCs.

When devising a program for international development, three issues must be addressed; (1) growth, (2) equity and (3) justice. Agriculture provides the vehicle to address all three. Everyone must eat. Agriculture can stimulate significant economic growth.

The most effective way to lower the birth rate is to improve the economic status of the people. As income increases, population growth rate decreases. Increasing agricultural production will help to alleviate poverty. Low food prices will also help to alleviate poverty in the cities. Agricultural research that has increased yields has prevented the use of millions

of hectars of fragile lands. Yield increases on the good lands will pave the way for food production needs for the future.

If we are going to cope with population, poverty and environment, agriculture must play a pivotal role. If the world of 8 billion people is to have any hope of providing a minimal standard of living, a way must be found to provide food and fiber in an 1) economically, 2) socially, 3) technologically feasible and 4) environmentally sustainable manner.

The use of computers will play an important role in developing the agricultural programs of developing countries. Development means to enhance (1) institutions, (2) human resources, and (3) technologies. Computers will play a role in all three.

Computers are inexpensive, accurate, speedy and can handle unlimited quantities of data. Computers can help with financial and personnel management. Computers can help with report writing, information exchange and improved communication. Computers can facilitate the research process, including experimental design, data management and the quick and efficient processing of experimental data.

The weakest link in the agricultural research process is data analysis. In my travels I have seen many cases where people just collect data, and never analyse it. Many times I have seen researchers plan their experiments before they analysed the data from the previous season. Some research systems have central data processing units. Many times I have heard stories of data never returning from the central data processing units. Computers get researchers closer to their data. The closer researchers are to their data, the better their programs will be.

MSTAT is a software package which captures the low-cost power and speed of computers to help generate solutions to developing country's food crises. Efficiency, timeliness and confidence of conclusion are vital concerns, and those engaged in research must have access to every technique that can contribute to the achievement of the research goals.

MSTAT, developed by Professor Øivind Nissen, is used by agricultural researchers all over the world. It is used in over 110 countries. MSTAT is used by many of the International Agricultural Research Centers – CIMMYT, WARDA, CIAT, CIP, IRRI, AVRCD, IFDC and ILCA. MSTAT is used by companies including Pioneer Hi-Bred, Northrup King, Sandoz, Merck, Coors Beer, Gallo Wines, Stauffer Chemical, Dekalb-Pfizer, Rohm and Haas, Funk Seeds, Agrigenetics and Cargill Seeds. MSTAT is used by many university researchers in the US, Canada and other parts of the world. I recently received a letter from an Ecuadorian who wrote "Because of its versatility and facility the students of INIAP and the Central University are required to know MSTAT as a requirement for their graduation."

As I said earlier MSTAT is used in over 110 countries. One person wrote "For my part I have used MSTAT in Thailand, Salomon Islands, Mexico, Tanzania, Gabon, Nigeria, China as well as Europe." Another person wrote "Our work covers 8 Pacific countries; Papua New Guinea, Salomon Islands, Vanuatu, Fiji, Tonga, Western Samoa, Tuvalu and Kiribati."

MSTAT has several immediate and significant effects on agricultural research systems. It helps researchers design and manage both simple and complex experiments. It enables researchers to analyse experiments over years and locations. The ease and speed of data analysis provides more timely results and thereby facilitates the generation of new and appropriate

technologies. MSTAT allows agricultural researchers to quickly determine the biological and economical significance of research results. MSTAT also enables agricultural researchers to exhaustively analyse past experiments in ways not possible before, and to glean information from past investments in research.

MSTAT maximizes the most scarce resource in the LDCs – trained people. The number of agricultural researchers per one million hectars in the developed world is 47, while in the LDCs there are 32. There are only 7 researchers/1 million hectars in sub-Saharan Africa.

Computer use in LDCs does have some problems. These include power supply, dust, heat, humidity, maintenance, repair, level of training and software availability. However, all these problems can be solved and should not present a long-term problem.

"MSTAT represents an appropriate software technology with wide applications to the needs of agricultural scientists around the world. Clearly it is a product which strongly complements and may justify expanded use of microcomputer technolgy in 3rd world agriculture." (Director General, M.S. Swaminathan, IRRI.)

I have given MSTAT workshops in Asia, Africa, Latin America, US and Canada. Some of my students had never even used a typewriter, let alone a computer. Some were very scared the first day or so, because they were afraid they would break the computer. However after the first day they were all smiles because they saw what MSTAT would do to help their research programs. Many write to me about how they continue to use MSTAT in their programs.

Crop yield growth rates declined in Asia from 3% in the 70's, to less than 2% in the late 80's. In some parts of the world, rice yields have stagnated. Wheat yields show the same trend. Last year when I reviewed the agricultural program in Sri Lanka, I saw that rice and other crop yields had stagnated. Last year China became a net importer of foodgrains.

If we look at the task ahead we see that we have a big job to do. The challenges of population, poverty and environmental degradation are great. The technological fault line is "real". My suggestion is to use computers to bridge the technological gap. Let's use the "power of the chip" to bridge the gap.

Because of the vision and hard work of Professor Nissen, thousands of agricultural researchers have been given the opportunity to bolt into the 21st century with computers.

Professor Nissen has given scientists around the world a software package which will help bridge the gap between the rich and poor. The recognition given today to Professor Øivind Nissen is well deserved.

Analysis of variance testing in the presence of random effects

INGE S. HELLAND
Department of Mathematical Sciences, Agricultural University of Norway, Ås, Norway

Helland, Inge S. 1995. Analysis of variance testing in the presence of random effects. Norwegian Journal of Agricultural Sciences. Supplement No. 22: xx-xx. ISSN 0802-1600.

The history of the random effect concept in the analysis of variance is briefly reviewed. Next basic terms like 'random', 'fixed' and 'nested' are defined and related to graphs connected to experimental designs. Using these graphs, simple and general rules for finding the correct error term in balanced designs are formulated. These rules hold in general when all random effects are modelled as independent random variables, which implies a different conventions for defining mixed interaction than the one usually found in textbooks. Some arguments for the chosen convention are given. The two conventions usually lead to the same tests for fixed effects. The rules are illustrated on some concrete situations, in particular split plot designs and series of experiments.

Key words: Analysis of variance, choice of error effect, expected mean squares, fixed effects, random effects, series of experiments, split plot design.

Inge S. Helland, Agricultural University of Norway, Department of Mathematical Sciences, P.O.Box 5035, N-1432 Ås, Norway.

Professor Nissen has numerous times complained about misuse of statistical packages. I think that he very often has been right in these complaints, and I also think he has been right in his next point: One of the most common mistakes has been that a wrong error has been used to test against in the analysis of variance tests, which in many cases have led to seemingly very sensitive tests, far more sensitive than one has any basis for in the data and in the situation that generated the data. Hence false conclusions may easily result.

The background for these mistakes is easy to spot: Most statistical packages, like SAS, Systat and Minitab do their computations as if all effects in any analysis of variance are fixed, unless told otherwise, and many of the packages have limited capability to handle random effects in analysis of variance models. Therefore the packages make their default calculations assuming that all effects should be tested against the residual mean square, which may or may not be correct. If it is not correct, the typical case is that the correct error mean square is both larger and has fewer degrees of freedom than the residual mean square. Both these factors contribute to making the tests formally made by the packages too sensitive.

At the same time packages of this kind are invaluable in carrying out the necessary calculations on real data. Most of

them tend to get more and more user-friendly, which also implies that they get more and more users. It is of course no solution to try to work against the statistics packages as such; it is the incorrect use that must be fought. One purpose of this paper is to give simple and general graphical tools that can be used to find the correct error to test against in many cases without the need of any complicated mathematical analysis. I will also discuss briefly two different conventions that have been used to define certain random effects, and some of the consequenses of these differences. As illustrations I will use several more or less standard situations, among others some from series of experiments.

Some background

There may be many reasons why random effect models to some extent have been neglecled in statistical textbooks, one is historical. Sir Ronald Fisher, the father of the analysis of variance, concentrated mainly on fixed effect analysis of variance, although he did mention random effect models in relation to concepts like intra class correlation. A systematic treatment of these models was not taken up until the paper by Eisenhart (1947), which inspired several important articles on mixed models and random models for the analysis of variance, most of this is summarized in the early review paper by Plackett (1960). The papers by Nelder (1965a,b) laid the foundation of the socalled Genstat-school, treating block-experiments with orthogonal and often unbalanced block structure. In the present paper we will concentrate on the balanced case, so this development will not be pursued further here. A modern discussion of experiments with balanced blocks is given by Tjur (1991), which forms the basis for much of the discussion below. But first one should mention on the unbalanced side the very general restricted maximum likelihood (REML) method for estimating variance components, first proposed by Patterson and Thompson (1971) in a block experiment setting and whose more recent use has been reviewed by Robinson (1987). This method and relatively recent prediction methods developed for mixed models have been much used in animal breeding, and there is now available computer software that seems to function reasonably well in a large class of unbalanced mixed models.

To use computer software for general situations without understanding the simplest case, is dangerous, however. Therefore we will concentrate here on the simple balanced case, where we in principle are able to calculate everything by hand, and (more importantly) where it is possible to see in a simple way from the situation which effect that should be tested against which. On the other hand, very general combinations of fixed and random factors will be allowed. Well-known designs like balanced factorial block designs, split plot designs and series over one and several years are only special cases.

Basic concepts

An experiment is always done on a collection of *experimental units*, which may be plots, animals, trees, plants or more complicated constructs like single measurements on animals. We assume that the experiment gives a number y for each experimental unit.

We will use the word *factor* in a wide sense which includes both treatment

factors and blocking factors. From the point of view of the analysis of variance, a factor is just a way to group the experimental units into disjoint groups. Thus plots may be grouped one way into plots with the same varieties, in another way into plots belonging to the same block. Then both varieties and blocks are regarded as factors. The different groups of experimental units corresponding to a factor are associated with different levels of the factor.

From this point of view the *interaction* between two factors is looked upon as the grouping implied by using the grouping criteria corresponding to both factors. This may be an unusual way to handle interactions, but it will turn out to be useful in the graphs below, where it also will be illustrated by several examples. We will let the term *'effect'* denote either a single factor or an interaction between factors insofar as the implied grouping is assumed to have some net influence on the response y. Then, from another point of view, the 'effects' correspond to the different terms on the righthand side of the model equation the way it is formulated either in textbook discussions of analysis of variance or in several computer systems used to analyze data.

A factor is *balanced* if the number of experimental units in each of the corresponding groups is the same. Similarly, an interaction is balanced when the implied division gives groups of equal sizes. We will only consider designs where all factors and interactions are balanced.

An effect B is *nested* in another effect A if the division implied by B is finer than the division implied by A. A simple example is if A is litters of animals B. Then animals are nested in litters whether the animals themselves are the experimental units, or the experimental units imply a still finer division. The interaction of two factors C and D is always nested both in C and in D. In the graphs below we will always draw B below A with a line from A to B when B is nested in A.

Finally we have the very important distinction between *fixed* and *random* effects. A factor is called fixed if the mean response for each single level of the factor is of interest to the researcher being responsible for the experiment. A related definition was given in Nissen (1989). On the other hand, the levels of a random factor may be regarded as sampled from some population, and it is only the variation in the population that is of interest. This last idea is more systematically exploited in Helland (1995).

It is of some significance that the interest of the researcher enters as an important part of this definition. In most cases it is fairly obvious what is of interest. For instance, in a variety trial the mean response of each variety is clearly of interest, hence variety is a fixed factor. On the other hand, the function of a block in such an experiment is mainly to reduce the error variation. The mean for each block is of no intrinsic interest, just the variation between blocks. Hence block is a random factor.

There do exist situations, however, where a factor can be random from one point of view and fixed from one point of view, and where both types of analyses can be of interest on the same data set. The following example is a case in point:

Assume that each of two therapies requires special training and equipment, so that a physician can be trained and equipped for only one of them. 10 physicians are randomly divided into two groups of 5, to be trained and equipped for the two therapies. Finally, 60 patients

(the experimental units) are randomly divided into 10 groups, 6 patients for each physician. Some response is measured on each of the patients after the treatment. (After White, 1975).

What is mainly of interest here, is to compare the two treatments, and then we want our conclusions to be valid for a large populations of patients and also for a large population of physicians. Hence physicians and patients are regarded as random effects. After this is done, however, a second analysis may be done to compare the physicians that actually took part in the investigation, and in this stage the physicians will be fixed effects.

In Figure 1a the factor graph corresponding to the main analysis is shown: It is a completely nested design: Patients are nested in physicians, which are nested in treatments. In this graph as in all graphs in the rest of this paper, random effects are singled out by drawing a ring around them. In Figure 1b the graph for the second analysis is shown.

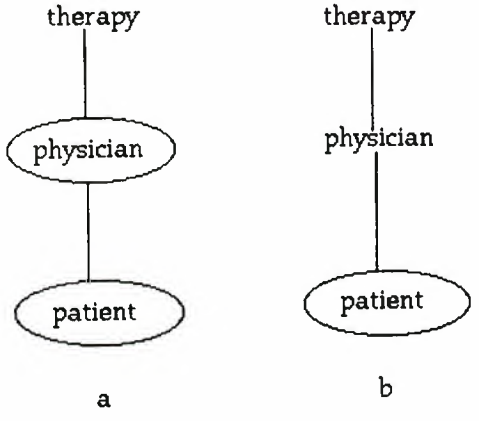


Figure 1. White's example with therapies, physicians and patients; a: physician as random effect; b: phycisian as fixed effect.

In general, when A is random and B is nested in A, then necessarily B must be random. Hence in all graphs, all effects below a random effect must be random. In particular, an interaction is random if at least one of the factors behind it is random. The residual error effect E, connected to the experimental unit and always at the bottom of the graph, must always be random.

The important theme of this paper is which effects should be tested against what error. To illustrate the solutions arrived at, we will in general give tests for all effects in the models used as examples, also sometimes tests that are of less interest in practice.

An extreme case and the general rule for finding the correct error

Assume a very simple experiment where two varieties are to be tested against each other in a block experiment with 5 blocks, each consisting of 2 plots. The character that one is interested in, is such that each plant has to be investigated. So 100 plants are randomly selected from each plot, and each plant measured to give a number y. The graphical structure of this experiment is shown in Figure 2.

Usually, a researcher carrying out such an experiment will only use the means over the 100 plant from each plot as input for an analysis of variance, and he will experience no difficulties. Nevertheless, suppose for the sake of argument that all the 1000 numbers are punched into a computer. A 'straightforward' analysis of variance using the effects varieties, block and plots will then give an 'error mean square' which presumably is fairly small, and which has 990 degrees of freedom.

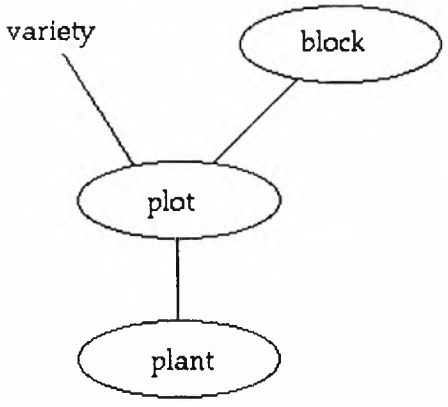


Figure 2. A block experiment with measurements on single plants included.

No serious researcher will use this as an error, but similar failures have been done many times.

What has gone wrong, and what is the correct procedure? The answer is simply that the last line in the analysis of variance output simply does not give the correct error to test the variety against. Looking at Figure 2, the error should of course be the plot-effect, with just 4 degrees of freedom. This is an example of the following simple rule, which is always correct when it works:

Rule 1: Test any effect A against the first random effect that is found by following lines downwards from A.

Thus the error for testing varieties (and also for testing blocks) is the plot-effect in this case. Note that when we have found the error to use in the F-test for an effect, the same error – and the corresponding degrees of freedom – should be used when computing LSD, in testing contrasts and so on. In the example above all these tests are equivalent to the tests obtained by just using the mean-per-plot numbers as inputs for the analysis.

Going back to the situation in Figure 1a, it is clear that the random effect 'physician' (with 8 degrees of freedom) is the correct error for testing treatments. In principle, in the situation of Figure 1b, one should test treatments against 'patients', but this is a test that hardly anybody can be interested in: It applies only to the situation where just the 10 specific physician involved in the experiment perform the treatment, 5 of them one treatment, and the other five the other treatment.

In the same way, it is theoretically possible to look upon the plots of Figure 2 as fixed, and then the test against plants could be accepted in some strange sense: The test is then done as if the only comparison of interest is for just those specific plots that are involved in this experiment. Experiments with such aims are never done. We always look upon plots as representatives of some population; hence the plots are random.

Further examples and further rules

The first observation we make is that in an ordinary randomized block experiment all factors should be tested against the residual error. This can be seen from a simple graph (like Figure 2 with the effect 'plant' deleted), and the conclusion is independent of whether we look upon blocks as random or fixed. In a similar way, Rule 1 works in the sense that it gives a unique error to test against, and then this error is the correct one.

A more complicated design is given by the split plot design, as exemplified in Figure 3, where irrigation is a factor on

main plots and variety is a factor on subplots, say in the following concrete sense: 4 blocks each contain 2 mainplots, one irrigated and one not. Each main plot is divided into 10 subplots, and 10 varieties are randomized over these subplots, independently for each main plot. Note that the form of the graph follows immediately from the nesting structure of the design, and it constitutes no problem to say which effects are fixed and which are random.

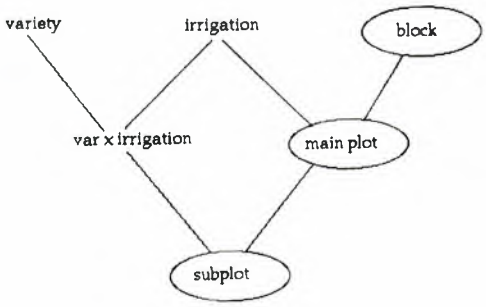


Figure 3. A split plot example.

Looking at Figure 3, it is obvious from Rule 1 that varieties and the interaction between varieties and irrigation should be tested against the subplot effect. It also seems reasonable from the figure that irrigation should be tested against the main plot effect, but strictly speaking, Rule 1 gives two possibilities: By starting with the line to the right from 'irrigation', we end up with 'main plot'; by staring with the line to the left, we end up with 'subplot'. To be able to choose the correct solution, we need a further rule.

Rule 2: When Rule 1 gives several possibilities, try out each of them using the following criterium: The set of random effects below and including the chosen error should be the same as the set of random effects below the effect to be tested.

In the split-plot example: The set of random effects below 'irrigation' is 'main plot' and 'subplot', and rule 2 gives then without questions that 'main plot' must be the correct error to use for irrigation. The same rule also gives the correct error for each factor in the split-split-plot design and several related designs.

The mathematical proof of these two rules is built upon calculation of expected mean squares, using the following way of thinking, which is explained in more detail in many statistical textbooks, for instance Montgomery (1991): The expected mean square for the effect to be tested is always equal to a term connected to the effect itself plus additional terms. The correct error must then have an expected mean square equal to these additional terms in order that the corresponding F-test, which consists in comparing the two mean squares, shall be unbiased.

In fact this way of thinking gives a precise derivation of Rule 1 and Rule 2 by using the following mathematical results: Each effect in the graph gives a unique contribution to the mean squares. (For random effects, this c_{effect} is equal to $n_{effect}\sigma^2_{effect}$, the product of the number of experimental units at each level of the effect and the corresponding variance component, and for fixed effects a slightly more complicated expression exists. However, we will not need these details here.) Furthermore, these contributions combine by

Rule 0: The expected mean square for any effect A is equal to c_A *plus the sum of the contributions* c_{effect} *for all random effects below A in the graph.*

Using this rule, it is easy to derive the criterion in Rule 2, and by a little further thought it follows that Rule 1 leads to the correct solution when it leads to a unique solution.

When one does not find a unique error effect by combining these two rules, it is impossible to find an exact F-test for the effect in question, and one must use an approximation. Such cases exist (and are not uncommon), as illustrated by the following example from Lea et al (1991):

10 judges are in a sensory panel to judge 5 brands of coffee. 4 cans of each brand are made, and every judge is given a sample from each of the 20 cans. The response for some specified attribute is given as a mark on some continuous scale. We look upon judges and cans as random effects. The corresponding graph is given in Figure 4.

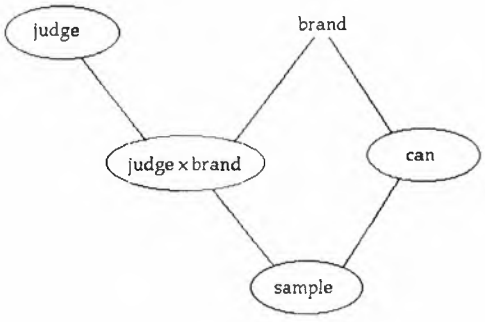


Figure 4. An example with judges evaluating samples of coffee from cans of coffee of various brands.

It is clear that we have a problem here when we want to test the brands of coffee, which is the test of main interest here. By Rule 1 the possible error effects are either 'can' or the interaction between judge and brand, but neither satisfy the criterion given in Rule 2. The solution is an approximate test using linear combinations of different mean squares.

Rule 3: When no unique error exists, do as follows: Using Rule 0, seek a linear combination

$$MS = MS_{effect1} \pm MS_{effect2} \pm \ldots \quad (1)$$

such that the expectation of the means square of A is equal to the contribution from A itself plus the expectation of MS. Then test A against MS with approximate error degrees of freedom

$$\frac{MS^2}{\frac{MS_{effect1}^{\ 2}}{df_{effect1}} + \frac{MS_{effect2}^{\ 2}}{df_{effect2}} + \ldots} \quad \textit{(rounded to the nearest integer).} \quad (2)$$

This is called Satterthwaite's approximation; see Montomery (1991) for more details and further references.

In the example in Figure 4 we find from the figure

$$\begin{aligned} E(MS_{brand}) &= c_{brand} + c_{judge \times brand} + c_{can} + c_{sample} \\ E(MS_{judge \times brand}) &= c_{judge \times brand} + c_{sample} \\ E(MS_{can}) &= c_{can} + c_{sample} \\ E(MS_{sample}) &= c_{sample} \end{aligned} \quad (3)$$

By inspecting these expressions, we find that the correct error for 'brand' is

$$MS = MS_{judge \times brand} + MS_{can} - MS_{sample} \quad (4)$$

with the approximate degrees of freedom calculated as in Rule 3.

Often, an alternative approximation is

recommended in this case, using sums of mean squares both in the numerator and in the denominator of the test variable F . This alternative usually gives a better approximation (one problem with the approximation sketched above is that MS can be negative). The advantage with the approximate procedure used above is that we have a single error to relate to in the usual way, and this can directly be used further for instance in testing of contrasts.

The mixed model muddle – alternative formulas for expected mean squares

Unfortunately, relatively simple and general rules for finding the correct error to test against, are very rarely found in statistical textbooks, at least not in the exact form as given above. In general, all textbooks agree on how to test fixed effects, and in most cases these are the tests of main interest. But for random effects one can find many variants, and these are usually derived from formulas for expected mean squares that are more complicated than the one formulated in Rule 0 above.

To understand the difference, it is necessary to look upon statistical models behind the analysis. To be concrete, look at a block experiment for varieties where the blocks are so large that the same variety can be used on several plots in the same block. This gives us the possibility to include the interaction between varieties and block as an effect in the model. The standard model has the form

$$y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk}, \qquad (5)$$

where α_i is the fixed effect of varieties, b_j is the block effect, c_{ij} is the interaction between varieties and blocks and e_{ijk} is the plot effect. The rules 0-3 above are derived under the assumption that in all such situations the terms of the model equation are normalized in such a way that all random effects like b_j, c_{ij} and e_{ijk} are independent normal variables. This is always possible if the data themselves are normally distributed variables, and in my view it is the simplest and most natural convention.

Unfortunately, the convention chosen in most textbooks is different. The interaction effect c_{ij}, is in most cases defined in such a way that the sum over the 'fixed' index i is zero. This is also possible to do. The terms b_j and c_{ij} will be different under the two normalizations, but from a purely mathematical point of view the two model formulations are completely equivalent. (See for instance Hocking, 1985.) When it comes to interpretation of the different terms, however, one can give many argument for the conventionwith all random effects independent:

- The fact that independent interactions gives the simple formula for the expected mean square formulated in Rule 0 above, with the result that the simple Rule 1 and Rule 2 are valid in all cases, is in itself a strong argument. The standard textbook solution requires complicated rules involving hidden and open indices to calculate expected mean squares. (See Montgomery, 1991, where the convention recommended here is only mentioned briefly as an alternative.)

- Looking at Figure 2 and Figure 5, the two figures have the same structure, the two situations are described by similar linear models of the form (5), and it would seem reasonable to use the same convention for the terms of this model in the two cases. Now by taking the

means over k in (5) in the situation described by Figure 2, we see that the independent formulation is completely consistent with the intuitive notion that the model for the means over the plants for each plot should give an equivalent analysis of variance model for testing varieties and blocks.

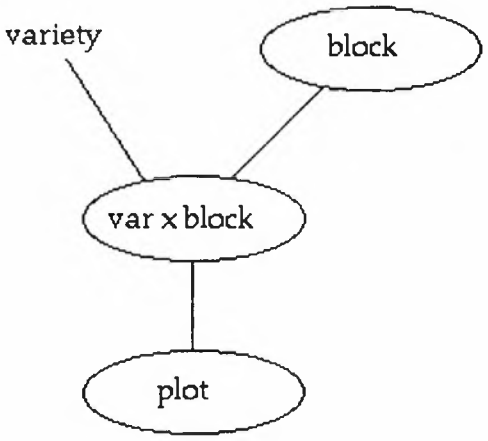


Figure 5. A block experiment with blocks large enough to contain several plots of each variety.

- By using the standard textbook model and taking the means over both the indices i and k in (5), the interaction/plot-term c_{ij} disappears, which in the situation of Figure 2 implies that the variation among the blocks in this model should be tested against the variation among the single plants within plots, which seems to be a very strange recommendation. In the independent normalization, the blocks should be tested against plots in Figure 2 and against the variety x block interaction in Figure 5.

- When it comes to unbalanced data, at least in animal breeding, independent interaction terms are nearly always used. This is also the convention used by the computer package SAS when a user asks for an expected mean square to be printed out. (An argument that may add to the confusion, though, is that BMDP, another major package, uses the standard convention.)

Further arguments against the standard textbook convention for defining interactions in mixed models can be found in Helland (1995) and further arguments in both directions are given in Samuels et al. (1991) and the discussion there.

If one wants to go into the discussion concerning the two alternative normalization in mixed models in full detail, many technical issues will have to be taken into account. Some people will claim that the situation described in Figure 2 is so different from the situation described in Figure 5 that there is no reason to use the same convention in the two cases. This may be a valid point, of course, but on the other hand, things will simplify considerably if we always can normalize random effects in the same way.

One (relatively weak) argument for the standard convention is: In certain special cases (an example will be given in the next section) it seems like one can reduce the need for the complicated Rule 3 above by using the standard convention instead of the independent one.

In my view, however, the arguments in favour of the independent normalization of interactions strongly outweight the argument against it. Main issues for me are the simplicity of the models, the simplicity of the accompanying rules to find expected mean squares, and the fact that the way to test fixed effects – usually the effects of main interest – is the same in both normalizations.

Series of experiments

Once one has general tools for formulating models and for testing effects in models, these tools can be used in fairly general situations. In particular, there is very little difference in these respects between single experiments and series of experiments as long as these series are balanced.

The graph for series over one year with several fields is shown in Figure 6. It is obvious that the variety effect should be tested against the interaction between varieties and fields.

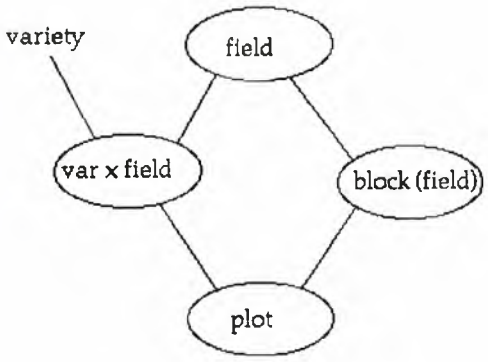


Figure 6. A one-year series of experiments.

In Figure 7 we give the graph for the ideal situation of a series over several years where new fields are selected each year. This means that the effect 'field' is nested in the effect 'year'. We look upon both years and fields as random. It is clear from the graph (and Rule 1) that the variety effect should be tested against the variety times year interaction, that this interaction should be tested against the interaction between variety and field, and that this latter interaction should be tested against the plot-effect. To test the main effects of year and of field, we would need to apply Rule 3, but in many cases these main effects are so clearly significant that a formal test is of less interest.

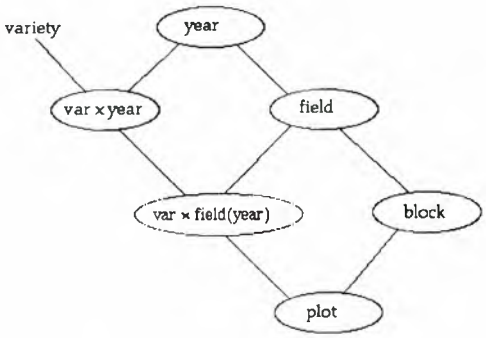


Figure 7. A series of experiments over several years with new fields each year. Fields are taken as random.

Finally, in Figure 8 we show the graph for a series carried out on the same field each year. We see that Rule 3 must be applied for all main effects (except blocks), while all two-factor-interactions are tested against the three-factor-interaction. The test for varieties is against the approximate mean square error $MS_{var\ x\ year} + MS_{var\ x\ field} - MS_{var\ x\ field}$ with approximate degrees of freedom found from Rule 3, and this solution is the same for both normalizations, confirming what has been said above in general for tests of fixed factors.

Going through all tests for the situation described in Figure 8, we find 3 cases were the normalizations lead to different solutions:

1) The test of years in the standard normalization is simply against $MS_{year\ x\ field}$, while the independent normalization (from the figure) leads to a Rule 3-case. As in the previous situation, one should guess that the year-effect is relatively clearly significant, so a formal test may not be too important.

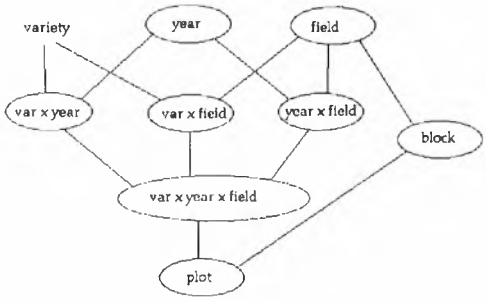


Figure 8. A series of experiments over several years with the same fields each year. Fields are taken as random.

2) The test of fields in the standard formulation is (approximate) against $MS_{year\ x\ field} + MS_{block} - MS_{plot}$, while the independent formulation demands a test against $MS_{var\ x\ field} + MS_{year\ x\ field} + MS_{block} - MS_{var\ x\ field\ x\ block} - MS_{plot}$.

3) The test of the interaction between year and field is against MS_{plot} in the standard formulation, but against the three-factor-interaction in the independent formulation.

To the user it may be confusing that two different conventions for defining mixed interactions give different tests in this way. It should be repeated, though, that the differences usually occur in the less important test. Also, once one get used to applying figures of the kind given in this paper, the tests read off from these figures more or less always seem to be the natural ones. For instance, looking at Figure 8, it seems very unnatural to test the year times field interaction against anything else than the three-factor-interaction.

Concluding remarks

The main purpose of this paper has been to describe qualitatively some simple rules for finding the correct error mean square in balanced analysis of variance situations. As approximations these rules can also be used for moderately unbalanced designs, for instance when a few observations are missing. In these cases several standard computer packages can do the computations of the mean squares. The important issue is to use these mean squares in a correct way.

The rules are based upon graphs connected to the various designs. It may be of some interest that the same type of graphs can be used to give simple general formulas for degrees of freedoms and sums of squares for balanced designs; see Tjur (1991). Note that the graphs described in the present paper always have one node for each effect in the model equation. When used for calculating degrees of freedom and sums of squares, it it also necessary to include a node for the constant term, placed at the top of each graph.

For the purpose developed here, to find the correct error mean square, the rules formulated are quite general. It is natural, however, to look upon the split plot experiment as the typical case where this sort of thinking applies. It may be appropriate to end by the following citation from perhaps the most famous european statistician living today, Sir David Cox. He recently gave a long interview in a statistical journal (Reid, 1994), where he among other things said:

"There is a tendency if you see random variation on different individuals, to model it by independent identically distributed random variables. The split plot is a warning against this: there may well be correlation structure in the error which mean that some comparisons have quite different precision than others. That's a point of importance far beyond just the classical design of experiments.

Although one could have a working life and never come across a split plot design, there are areas where it's a wholly natural design and there are other areas where it's being used implicitly, without people quite understanding that this is what is being done, and getting incorrect estimates of precision. So I think the moral is that these very important, absolutely central issues, like split plot, have to be understood, and it's more important to understand them at a qualitative level than to plough through a lot of algebra."

The graphical approach may be a way to avoid too much algebra.

References

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. Biometrics 3: 1-21.

Helland, I. (1995). A population approach to analysis of variance models. Submitted.

Hocking, R.R. (1985). The Analysis of Linear Models. Brooks/Cole Publishing Company. Monterey, California.

Lea, P., T. Næs & M. Rødbotten (1991). Variansanalyse for sensoriske data. Matforsk, Ås.

Montgomery, D.C. (1991). Design and Analysis of Experiments. Wiley, New York.

Nelder, J.A. (1965a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. Proc. Royal Soc. London A 283: 147-162.

Nelder, J.A. (1965b). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. Proc. Royal Soc. London A 283: 163-178.

Nissen, Ø. (1989). Bruk og misbruk av statistikk. Notat, Institutt for plantekultur, NLH.

Patterson, H.D. & R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58(3): 545-554.

Plackett, R.L. (1960). Models in the analysis of variance. J.R. Statist. Soc. B 22: 195-217.

Reid, N. (1994). A conversation with Sir David Cox. Statistical Science 9(3): 439-455.

Robinson, D.L. (1987). Estimation and use of variance components. The Statistician 36: 3-14.

Samuels, M.L., G. Casella & G.P. McCabe (1991). Interpreting blocks and random factors. J. American Statist. Ass. 86(415): 798-821.

Tjur, T. (1991). Analysis of variance and design of experiments. Scand. J. Statist. 18(4): 273-322.

White, R.F. (1975). Randomization and the analysis of variance. Biometrics 31: 552-572.

Nearest neighbour analysis

ØIVIND NISSEN
Department of horticulture and crop sciences,
Agrigultural University of Norway

Field experiments with a large number of entries will usually have a large experimental error. The reason is that the variation in soil fertility is not random. Plots far apart tend to be more different than plots close to each other.

A number of methods to increase the precession of such experiments have been proposed.

I I will mention 4 of these:
One of the oldest is the check plot method. The check entry is planted on a number of plots, usually at even distances across the experimental field. The yields of the other entries are then corrected by means of interpolated check plot yields. This method was used in Norway 100 years ago by Bastian Larsen, founder of the Norwegian agricultural experimentation.

II Approximately 60 years ago Yates introduced the Lattice designs for comparing a large number of entries. Since 1949 we have used these designs quite extensively in Norway, and we have found them valuable in reducing the error compared to complete block designs.

Lattice designs have some limitations:

1. The number of entries must be a square (4, 9, 16 etc.) or a rectangular (6, 12, 20, 30 etc.) number. This limitation is seldom serious: If we want to compare 18 new oat lines with a standard variety, we may either use two standards or duplicate the one we are using, in order to get 20 entries.
2. The number of replicates must be the same for all entries. This may sometimes be difficult to obtain if we are comparing new varieties and do not have the same amount of seed for all.
3. The block size must be constant. In field experiments this very seldom cause any difficulty.

III A general incomplete block or nonorthogonal anova
In each block we may have from zero to a large number of plots for each entry. The data are analyzed with an iteration program that finds the block and entry parameters which minimize the block times entry interaction. If the number of replications of an entry in each block is zero or one, the result will be the same as with a two-way anova with estimation of missing values.
In Norway this Nonorthogonal Anova has been used mainly for analyzing series of variety trials where the same varieties are not in all experiments or years.

IV A Nearest Neighbor program
Such programs usually correct the yield on each plot by a comparison with the average of the two neighbor

plots, and the average yields of the entries on these plots. This process must be repeated until the results are stable.

The program I have written compares each plot with the average of two plots on each side, I have therefore called it NNNA, for Next Nearest Neighbour Analysis.

This program can be used for any number of entries and any number of replications for each entry. Each plot of a certain entry should have as neighbours as many of the other entries as possible.

The two endplots on each end of a tier do not have two plots on each side, and must be compared with the four other endplots.

The tiers should therefore be as long as possible, a length of 5 plots is the absolute minimum.

My NNNA program calculates only the corrected means for each entry, and does not give any computed error for these means.

Comparing lattice analysis and NNNA

Two analyzing methods, say Lattice and NNNA can be compared in two very different ways:

1) By analyzing real experiments with the two methods, and
2) By means of statistical simulation.

If we, as in my NNNA-program, do not have a computed error, the first method requires a series of similar experiments. All experiments are analyzed by the two methods, and the corrected means are analyzed "over experiments", separate for the two methods. The error "over experiments" is the sum of the interaction experiments x entries and the error within experiments and can, therefore, be used to compare the two methods. A weakness is that the error over experiments usually is much greater and more variable than the error within experiments.

I have been using the other method, statistical simulation.

First a design is constructed, for instance a randomized lattice with 3 replications and 25 treatments. For each plot a "yield" is computed as a sum of a "mean", say 1000, a linear trend within each tier and a random error.

The trend, with average zero, is measured in "% per plot".

The random error is a normal variable with mean zero and standard deviation in percent.

The data are then analyzed by the lattice and the NNNA method. Without error the corrected means for all treatments should be 1000, and the standard deviations of these means are direct measures of the standard error for the two analyzing methods.

A few preliminary studies with differed tier length and different number of replications with the NNNA method gave these conclusions:

1 A tier length of 7 plots is too short.

2 Experiments with 2 or 4 replications for half of the treatments gave nearly as good results as experiments with 3 replications for all treatments.

3 If up to 60% of the treatments were on one plot only and the rest on 2 plots, the NNNA method reduced the error considerably compared to using the direct means.

4 The comparison Lattice versus NNNA

depends on the linear trend compared to the random error.

To study this relationship in more detail three replicated "experiments" were simulated.

Two of these will be discussed here.

Each had 10 Lattices with r=2, t=25 and a tier length of 25. Data were the 16 combinations of: Random error 2, 4, 6 or 8%, linear trend of 0.3, 0.6, 0.9 or 1.2 %/ plot.

In one of the series the trend was upwards from one end of the tier to the other end, in the other series the trend went up from both end of the tier to the middle plot. the average trend was therefore zero.

The results can be summarized as follows: For both methods the error increased with increasing random error.

The error for Lattice increased with increasing trend.

Increasing "one way trend" had no effect on NNNA, and "two way trend" only a slight effect.

The conclusion from these simulation studies is that Lattice and NNNA are on pair when the linear trend, measured in percent per plot is about 5% of the standard deviation of the random error. If the trend is larger the NNNA will give the best results.

One important question now remains: How are the linear trends compared to the random errors in actual experiments?

To answer this question I have used two series of experiments from the State Experiment Station Apelsvoll and one series from the Farm Crop Institute.

1. Early barley varieties, a total of 53 experiments with 2 replications and 10 or 12 entries in each.
2. Potato varieties, a total of 26 experiments with 3 replications and 7 entries.
3. Spring wheat varieties. 13 experiments with 2 replications and 16 or 25 entries.

After correcting for entry effects I computed the linear trend, the random

Table 1. Test of the difference between Lattice and NINA.

Trend %/plot	Random error	Trend in % of error	Trend one way	Trend up/down
0.3	8	3.75	-	-
0.3	6	5	-	-
0.6	8	7.5	+	-
0.3	4	7.5	+	-
0.6	6	10	+	-
0.9	8	11.25	+	-
0.3	2	15	+	-
0.6	4	15	+	+
0.9	6	15	+	+
1.2	8	15	+	+
1.2	6	20	++	+
0.9	4	22.5	++	+
0.6	2	30	+++	+
1.2	4	30	+++	++
0.9	3	45	+++	++
1.2	2	60	+++	+++

Legend:			
-	Lattice	best	P>5%
+	NNNA	best	P>5%
++	"	"	P<5%
+++	"	"	P<1%

error and the linear trend in percent of the random error for each replication separately.

The interesting results can be seen in table 2.

The 2 series from Apelsvoll gave very similar results. For both crops and for both characters studied 75 to 85% of the blocks shoved a linear trend larger than 6% of the random error.

The wheat experiments differed. For moisture only 25% and for grain yield 39% of the replications had a linear trend of more than 6% of the random error. This difference can only partly be explained by the different tier length in the three series. In the barley experiments the length was 10 or 12 plots, in the potato experiments 7 and for wheat 16 or 25. I have studied this "length effect" by dividing the wheat replications in blocks of 8 plots and computing the trend and the random error for each such blocks separately. The result is given in the last line in table 2.

My final conclusions from these studies are:

Even when experiments are designed as Lattice the NNNA analysis seems to give the "best" results in round 50% of the cases. An additional advantage of NNNA is that this method is not restricted to a certain number of entries or a fixed number of replications, and does not use blocks.

Table 2.

	No.of reps.	Percent reps. with linear trend in % of random error								
		0	6	12	18	24	30	36	42	Sum
Barley moisture	104	28	20	26	9	7	5	5		100
" grain yield	106	26	21	19	11	9	5	9		100
Potato dry matter %	78	22	20	8	12	15	10	13		100
" tuber yield	78	14	17	15	13	17	9	15		100
Wheat moisture	24	75	17	4	4					100
" grain yield	26	61	31	8						100
Wheat grain yield blocks of 8 plots	70	54	30	10	3	2	0	1		100

Presentasjon av forsøksresultat

ROLF ENGE
Institutt for plantefag, NLH, Ås

Jeg vil i dette innlegget se presentasjon av forsøksresultatene fra rådgiverens side. Forsøksresultatatene kan framstilles ved tabeller, figurer enkle funksjoner og modeller. Tidligere ble forsøksresultatene her i landet stort sett framstilt i tabeller. Av og til ble brukt stolpediagram og der det var nødvendig ble det brukt regresjonsfunksjoner. Med den tekniske utviklingen vi har hatt i de siste årtiene, har vi fått mange måter å framstille forsøksresultatene på. Mangfoldet er blitt stort.

Tabellene var stort sett greie å ha med å gjøre for rådgiveren, men ikke så sjelden falt rådgiveren for fristelsen til å gjengi alt for kompliserte tabeller i sin helhet og det gjør det vanskelig for tilhørere eller lesere å få tak i poengene. Når en bruker tall, kan en få fram mye bare ved å bruke noen få tall. Ved skriftlig framstilling, kan en bruke en noe mer komplisert framstilling. I dette innlegget er det imidlertid forskerens presentasjon som skal under lupen.

Tabellene

Fra forskernes side bør tabellene være så omfattende at de gir dekning for konklusjonene i teksten og det må være samsvar mellom tekst og tabeller. Det er irriterende når en med en rask gjennomlesing finner feil. Sjøl noe så banalt som at summen av prosentene på grunn av avrundinger ikke blir 100, bør ikke forekomme.

Eksempler til skrekk og advarsel

Verre er det når forskeren presenterer tall som er feil eller som forskeren burde vite var feil. I en rapport "Jordsmonnovervåking i Norge 1992-1996" Rapport fra programmet 1992, er det ved en bekk ikke langt fra Mjøsa målt en nedbørmengde på 1065 mm i 1991. Året 1991 var meget tørt. På den meteorologiske stasjonen på Kise (7-8 km fra bekken) ble det i 1991 målt en nedbørmengde på 467 mm. I det høgdelaget (under 300 m oh) som denne bekken ligger i, er det i dette århundre ikke målt nedbør på over 1000 mm i noe år på Hedemarken. Målingen er feil og det ble det gjort oppmerksom på, men forskeren valgte likevel å publisere denne feilmålingen. Det er her klart at forskeren skulle ha brukt observasjonene fra Kise og forklart hvorfor disse observasjonene ble brukt. Etter min mening er det bak mål å publisere en måling som en vet er feil. I dette tilfelle vil en som ikke skjønte at det som stod i tabellen var feil, komme til å dra den feilslutningen at store nedbørmengder i dette området ikke fører til økt avrenning av næringsstoffer og økt erosjon. Dersom en ellers får resultater som opplagt er feil, skal de ikke publiseres som en sannhet i noe fall. I dette tilfelle kan resultatet også komme til å bli brukt i samfunnsplanlegging og dermed påføre samfunnet unødvendige utgifter.

I en utredning om:"Miljøavgifter gitt på kunstgjødsel-N og -P og på plantevernmidler" fra 1989, er det i tabells form presentert et tallmateriale som slår alle rekorder.

	Kg N per dekar	
	Fra utredningen	Statistisk sentralbyrå (1979)
Korn, Romerike	12,5	10,6
" Jæren	14,5	6,1 x)
Gras, Romerike	25,3	14,3
" Jæren	27,5	24,2 x)

x) I 1989 var talla for Time på Jæren fra Byrået ca 18,6 kg N per dekar for gras og ca 6,5 kg N per dekar for korn. Gjødselstatistikken viser at innkjøpet av gjødsel er sterkt redusert de siste åra i Rogaland. Gjennomsnittlig gjødsling med nitrogen var i 1989 ca 11 kg per dekar i Norge. Tabellen viser ellers at det på Jæren til korn blir gjødslet med mer enn dobbelt så mye nitrogen som det som faktisk brukes. I tillegg til det som i tabellen er oppgitt som kunstgjødsel, kommer så husdyrgjødsla. Utredningen er laget etter oppdrag fra Landbruksdepartementet. Dette er derfor en meget "farlig" utredning fordi den brukes i den offentlige forvaltningen. Det er utført mye arbeid for å legge denne utredningen død. Om vi har lykkes med det, er jeg ikke sikker på.

Andre måter å framstille resultater på

I seinere år har det blitt mer og mer vanlig å bruke funksjoner og med grunnlag i dem lage kurver. Det har også blitt vanlig å lage mer komplekse modeller som gis ut for å være allmengyldige. Allmenngyldige er de som regel ikke. I det følgende skal jeg se på noen uheldige eksempel på bruk av funksjoner.

Først skal jeg vise et uheldig eksempel som jeg sjøl har publisert. I "Fagnytt, Plantefag nr.1 1994", har jeg på grunnlag av materiale fra Statistisk sentralbyrå, blant annet framstilt avlingsutviklingen for kveite fra 1960 til 1992. Det er gjort ved bruk av en andregrads regresjonsfunksjon og berekningene er utført ved hjelp av Nissens programmer:

$Y = -1558+44,52x-0,251x^2$, Y=avling og x=år med $R^2 = 0,73$.

Funksjonen viser at avlingene fikk maksimumsverdi i 1988. Dersom denne regresjonen forlenges ut over 1992, vil det bli helt ubrukelige resultater slik som kurveframstillingen viser. Avlingene vil ikke gå ned med mindre det blir restriksjoner på bruk av nitrogengjødsel. Årsaka til at kurven har fått denne formen er flere. Den kraftige stigningen mot slutten av 1960-tallet skyldes flere faktorer og noen av de viktigste er: klima, nye sorter, bedre tilpasset gjødsling og det at gardbrukerne lærte seg å dyrke korn i ensidig åkeromløp. Årsaka til nedgangen mot slutten av perioden er de klimatiske faktorene.

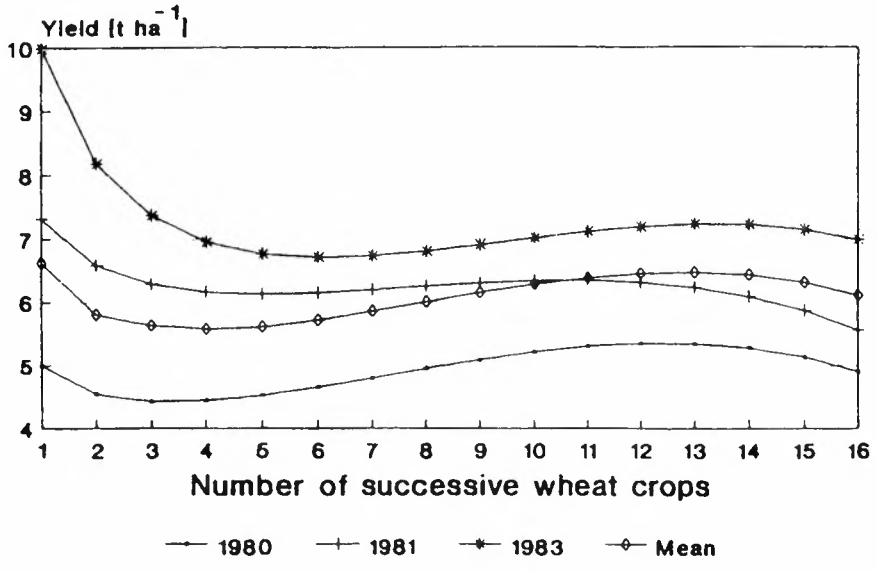
Det er dessverre ikke alle som er klar over at en ikke bør forlenge en regresjonsfunksjon ut over det området en har observasjoner for. I teksten skrev jeg derfor at funksjonen bare gjelder for perioden 1960-1992 og at den ikke kan brukes til å estimere avlinger etter 1992.
Øivind Nissen bruker å si at det ikke er nok med god tilpasning til materialet, men funksjonen må også gi et fornuftig resultat. En funksjon slik som den refererte, er farlig fordi den i ukyndige hender kan bli misbrukt. Basert på det samme materiale laget Øivind Nissen funksjonen: $Y=595,7-9195/(x-36)$, hvor x er år og Y er avling som før. Denne hyperbelen gir en litt dårligere tilpasning med $R^2 = 0,71$. Dersom en forlenger denne funksjonen ut over 1992, får en, som en ser, i alle fall fornuftige resultater.

I ei bok som kom ut i 1992 "Bønder og gjødsling" har forfatterne forsøkt å lage funksjoner for bruk ved gjødslingsplanlegging. I publikasjonen er brukt tredjegradsfunksjoner. For bygg dyrket på dårlig jord kom forfatterne fram til funksjonen:

$Y=147{,}514+26{,}979x-0{,}1337x^3$.

Her betyr x kg nitrogen per dekar og Y som før avling. Denne funksjonen gir ei maksimal avling ved ei nitrogengjødsling på 8,2 kg N per dekar. Årsaka til at de har valgt en tredjegradsfunksjon uten andregradsledd er nok at de vil ha en brå utflating av avlingsøkingen ved legdegrensa. Jeg tviler på at legdegrensa går ved lågere gjødslingsnivå på dårlig jord enn på god jord. Tvert om tror jeg at legdegrensa kommer ved sterkere gjødsling, men responsen for hvert gjødseltrinn vil være svakere og likne mere på avlingskurver på god jord under tørkestress. Det er ikke gitt klare grenser for funksjonens område, men det er antydet at funksjonen passer innenfor det aktuelle området for gjødsling. Egentlig passer ikke funksjonen til å estimere effekten av gjødsling på dårlig jord i det hele tatt fordi den virkelige vekstkurva vil være langt flatere hele vegen og neppe gi avlingsnedgang før ved meget sterk gjødsling på slik jord. Øivind Nissen har tidligere laget et gjødslingsprogram som etter hvert fungerte. Det faglige grunnlaget var det andre som stod for. Programmet ble flittig brukt av rådgivingstjenesten.

Til slutt skal jeg vise et eksempel fra en tysk publikasjon: "Yield Reaction of Winter Wheat in Monoculture in Dependence upon Weather and Soil" Publisert i: "Journal of Agronomy and Crop Sciense 1990 Vol 165 (hefte 2-3) p 151-168." Det er mye som kan sis om denne publikasjonen også, særlig når det gjelder forsøksmetodikken som er brukt. Jeg skal la det ligge nå, men bare presentere noen kurver. Kurvene er laget ved hjelp av en multippel regresjonsanalyse. Formelen er ikke oppgitt. R^2 er oppgitt til å være 0,52 for gjennomsnittlig avling. Forsøka omfatter i alt 16 år med monokultur av vinterkveite. Resultatet av



undersøkelsen er, som en ser, at avlingene gikk betydelig ned de første åra med monokultur for så å stige og stabilisere seg og deretter gå ned etter 13-14 år. Det siste fallet i avling er vanskelig å forstå og skyldes med temmelig stor grad av sikkerhet den slette metodikken som er brukt. Klimafaktorene kan også ha hatt effekt her. Kurvene kan ikke under noen omstendighet forlenges. Det ville ikke gå mange åra før avlingene kommer ned på null. Igjen kurver som er pedagogisk dårlige og i tillegg er de farlige fordi de kan føre til feilslutninger.

Konklusjoner

Annen og særlig tredje og høgere grads polynomer er dårlig egnet til å framstille avlingskurver. De er enkle å integrere og derivere. Slik sett kan de være morsomme å leke seg med. For oss som arbeider med rådgiving er det imidlertid viktig at det som kommer fram i forskningen presenteres på en slik måte at det ikke blir misforståelser. Her har Nissen gjort et veldig verdifullt arbeid gjennom et langt liv. Konklusjonene skal forøvrig være dratt av forskerne og vi skal ikke være redde for å bruke de resultatene som kommer fram. Jeg har vist noen eksempler på til dels grove feil og feil som i de fleste tilfelle hadde vært meget enkle unngå. Funksjoner, tabeller og kurver må presenteres på en slik måte at vi ikke går oss "vill". Jeg vil igjen ,minne om det jeg refererte fra Øivind Nissen, at en funksjon ikke bare skal gi god tilpasning til forsøksmatertialet, men den skal også gi et fornuftig resultat.

Etter 1970 har jeg bare brukt Nissens EDB-programmer sporadisk, og når jeg skal berekne noe søker jeg hjelp hos andre på Instituttet, fordi jeg arbeider som rådgiver og ikke som forsker. Jeg har imidlertid hatt ham som sjef i flere år og har også dratt nytte av hans spesielle kompetanse i andre stillinger jeg har hatt. Det som har vært kjennetegnet er at alt legges til side når Nissen får besøk og at han kaster seg over problemene som blir presentert.

For oss i rådgivingstjenesten er det umåtelig viktig at vi får resultater som vi kan stole på og ikke slike ting som jeg har referert her.

Summary

Presentation of Scientific Reports

This presentation is a review of scientific reports as seen from the advisors point of view. Examples are given of reports where results contradicting common knowledge are given without acceptable justification.

1. A survey on the amount of nitrogen fertilizer used in Norwegian agriculture, shows important deviation from Official Statistics of Norway, presented as "Statistics Norway".

Table 1. Aplication of nitrogen (N) as kg N per ha

	Survey data	"Statistics Norway"
Cereals, Romerike	125	106
", Jæren	145	61
Grasses, Romerike	253	143
", Jæren	275	242

A comparison of the survey data from "Statistics Norway" with the statistics from companies selling fertilizers would have revealed the lack of agreements between the datasets and should have prompted the scientist to careful revaluation of the survey data before publishing.

2. In "Fagnytt Plantefag nr 1, 1994" I presented a regression function for the yield of spring wheat (Y) as affected by years (x) for the time period 1960–92, $Y=-1558+44,52x-0,251x^2$, $R^2=0,73$. The function shows how the yield has improved from 1960–1992. The data used are taken from the "Statistics Norway". Extrapolation of this function gives very low yield in the year 2 000. In the paper I stressed that the function cannot be extrapolated. Later on I contacted proff. Nissen and he formed a hyperbolic function for the same material, $Y=595,7-9195/(x-36)$, $R^2=0,71$. This function will give reasonable results also if extrapolated. Proff. Nissen use to say: "It is not enough that a function gives a high correlation it must also give reasonable results".

3. A scientific report gives a function estimating the relationship between yield of barley and nitrogen fertilizer as $Y = 147,514+26,979x-0,1337x^3$, where Y=yield in kg per decare, x= kg nitrogen per decare. This function predicts maximum yield at 82 kg N per ha, wheras most research indicates maximum yield at a much higher N-level.

4. In "Journal of Agronomy and Crop Science 1990, vol. 165, p 151-168" in a paper entitled "Yield Reaction of Winter Wheat in Monoculture in Dependence upon Weather and soil" multiple regression functions are presented to describe the relationship between yield and climatic parameters and a trend over years. $R^2=0,52$, indicating that 52% of the variation in yield can be explained by the data, however if extrapolated over years the dependent variable very soon approaches zero.

The extension service depends on scientific reports to form the practical advice to farmers. It should be unnecessary to stress that unreliable or misinterpreted data causes confusion and creates difficulties for the implementation of new technologies based on serious scientific reports.