ISSN 0962-8452

The Royal Society

Proceedings: Biological Sciences

Volume **249** Pages 1–117 Number **1324** 22 July 1992

Proceedings of the Royal Society of London SERIES B

PROCEEDINGS: BIOLOGICAL SCIENCES

NOTICE TO AUTHORS

Scope. *Proceedings* series B welcomes papers on any aspects of biological science. It publishes, rapidly, announcements of important results, normally not exceeding 4000 words (five printed pages). The Editor will also consider short reviews, but only if they contain original and interesting new ideas. Short additions to, or criticisms of, papers that have already been published (subject to their originality and interest, and to the rules of good manners) will also be acceptable. Preliminary reports ('letters') are not encouraged. The target publication time is three months from receipt (excluding any period that the manuscript is in the hands of the authors).

Submission. There is no page charge. Authors should send their papers directly to the Editor. Up to six keywords should be supplied with the typescript. Four copies of the typescript and figures are required, and one set of originals of any figures should also be submitted. Papers should be prepared in accordance with the Instructions to Authors printed at the end of every issue of *Proceedings* B and also available from the Editorial Office.

Offprints. Authors will receive 50 free offprints.

SUBSCRIPTIONS

Proceedings of the Royal Society series B (ISSN 0962-8452) is published in monthly parts. The subscription price in 1992, including postage, is £188 in the U.K. and £200 (U.S. $380) overseas. Air-speeded post is available for overseas subscribers at a small extra charge. Full details of subscription rates may be obtained on request from the Publications Sales and Marketing Department, The Royal Society, 6 Carlton House Terrace, London SW1Y 5AG.

Editorial code for presentation of statistical analyses

INTRODUCTORY GENERAL COMMENTS

The proliferation of statistical methods and software increases the need for authors to state, unambiguously and informatively, the methods they have used. The reader of any paper should expect to learn precisely what the author has done, and to be able to adopt similar procedures in his or her own work.

An author who uses an algebraic symbol or technical term that is not common enough to be generally understood should state exactly (by definition or by bibliographic reference) the meaning given to it. Here we set out a code for the statistical terms and symbols that we regard as part of the general language of science. The Editor will accept them without detailed explanations or definitions. There can be no absolute standards of *correctness*: new developments in statistical methods continually appear, and statisticians themselves do not employ a unique system of symbols and nomenclature. Confusion is inevitable, however, if authors fail to make clear their own usages.

THE CODE

The code that follows is not intended to restrict an author's freedom to choose symbols or statistical methods, but only to simplify the task of writing. The Editor will accept the following without definition, but will expect any different usages to be made explicit:

t: Student's t-statistic (when used for a difference between two arithmetic means, the text must clearly identify the means that are being compared).

r: Product moment or Pearson correlation coefficient (*not any other*), modifiable to $r_{x,y}$ if the variables involved need to be specified.

b: Simple linear regression coefficient of y on x, where x, y have already been clearly identified, modifiable to $b_{x,y}$, if the variables need to be specified.

F: Ratio of two independent error mean squares, usually, but not invariably, in an analysis of variance (degrees of freedom must be stated).

χ^2: Statistic used in many well-known tests of significance relating to tables of frequencies, where, if correctly calculated, it has a probability distribution approximately that of a sum of squares of several independent N(0,1) variables, 'N(0,1)' referring to a Normal distribution with mean 0 and variance 1. The degrees of freedom must be stated. If it is necessary to do so, this statistic can be described in the text as 'chi-squared', but it is preferable to use the symbol.

p: Probability obtained in the course of applying a test of statistical significance (it should be printed in the lower case).

$\pm$: Introduces the standard error of the mean whose numerical value has preceded it. Special care is needed, because this sign has been used by some writers as introducing the standard deviation, or even the maximum range of a measurement error.

s.d.: Standard deviation, the square root of an estimate of variance.

s.e.: Standard error, always of a mean or of an estimate of some other parameter.

d.f.: Degrees of freedom.

$\log_e$: Natural logarithm (base e).

$\log_{10}$: Logarithm to the base 10; other bases can be specified by subscripts.

Using any of these abbreviations or symbols with different meanings (for example, using t for a time, or r for a ratio) requires that they be defined. However, the use of F, N or P in chemical contexts without definition is unlikely to cause misunderstandings. Referees will be asked to note undefined usages, and to regard them as requiring that a paper be returned to its author for clarification. In particular, any use of N or n should include a statement of whether this is the number of observations, the number of degrees of freedom, or something else. Authors need not be inhibited from using variants of these symbols, such as 'r_s' for a Spearman rank correlation coefficient, provided that the symbolism is defined.

The description of χ^2 as 'well-known' does not imply that it can be used uncritically. Many of the commonest errors in the elementary application of statistical methods are associated with the misuse or miscalculation of this statistic, but for comparing observed and expected frequencies it remains one of the most widely used techniques. Mention of F does not imply that its numerical value must always be given, for often no great interest attaches to it.

Every mention of t or F or χ^2 should explicitly state the degrees of freedom, either as subscripts or in brackets, for example, as t_{17}, $F_{3,27}$ or $F[3, 27]$, $\chi^2_{[28]}$, and so on, unless the accompanying text makes the numbers of d.f. apparent. For a correlation coefficient, the degrees of freedom should always be stated unless the number is evident as being two less than a previously stated total number of pairs of values.

Authors should remember that the physical dimensions of any s.d. or s.e. are those of the measurement to which it relates. A regression coefficient of y upon x has dimensions 'units of y per unit of x', and an explicit statement of these units may be needed in any discussion of a particular coefficient. The statistics t, r, F and χ^2 are dimensionless, as are some statistics that appear in less-familiar methods. For all others, the physical dimensions should be stated wherever numerical values are shown unless the text has made the units very obvious. Any unstated assumption that time is measured in days or in seconds, or that mass is measured in kilograms, may confuse a reader!

Proc. R. Soc. Lond. B (1992), **249**, 1–2
Printed in Great Britain

Within the written text of a paper, statistical symbols should be printed in italics rather than roman letters, except in circumstances where this practice might itself cause confusion. Special care is needed with the names of methods. Some, such as 'multivariate analysis' and 'cluster analysis', relate to broad classes of statistical technique, and are likely to be ambiguous unless they are specified more exactly. Research workers often give newly fashioned names to new concepts, and whether or not these come into widespread use is largely determined by natural selection. Some methods are commonly named eponymously: convenient though this practice is, such a name may be meaningless even to a professional statistician who is not familiar with a highly specialized field of application. In all cases of doubt, the Editor will expect an author to make clear the meaning, and to provide appropriate bibliographic references for any method whose name has not become generally known in the particular field of application.

Many biologists today analyse their data by using statistical packages on mainframe or desktop computers. Among the ever-growing number of general-purpose packages available today, a few have become so widely used that their names can now be regarded as assimilated to the general language of science. Hence, in the text of a paper, a reference to BMDP, GENSTAT, GLIM, SAS or SPSS is admissible without further identification. However, an author should describe, in the section headed 'Materials and methods' or elsewhere, any use of a different package, exactly as the standard conventions of scientific writing would lead him or her to do in respect of special apparatus or special reagents, with appropriate identification of its source and the address of its producer or supplier. When using any software package, an author cannot safely assume that the computer output will be presented solely in language and with terminology that is totally acceptable without explanation. The single aim of this recommendation is to ensure that what an author has done is made clear to the readers.

Authors should seek to avoid giving a spurious appearance of numerical exactness. Few sets of biological data can sensibly be summarized by means, special-purpose coefficients, or other statistics purporting to be accurate to 1 part in 10000. For example, it would be inappropriate to record all the numerals if the mean mass of an organism, given by a computer program, came out as 42.57146 mg. The same applies to probabilities obtained by the application of any significance test. Arithmetically correct computation might seem to lead to '$p = 0.0002583$', but it is most unlikely that the underlying assumptions about the nature and source of the data would justify anything more exact than '$p = 0.0003$', or possibly only '$p < 0.005$'.

David J. Finney
Bryan C. Clarke
May 1992