Presenting statistical analyses

Authors must be explicit about the methods they have used and should ensure that they have applied the most appropriate statistical test(s). Equally, results must be discussed unambiguously (e.g. instead of saying 'datasets satisfied the assumptions of the ANOVA', be specific, i.e. 'datasets did not deviate significantly from the assumptions required by the ANOVA'). The following should be described or stated with absolute clarity:

- the exact design of manipulated experiments, with details of replication and nesting;
- means by which data were produced for observational studies (e.g. sources, selection criteria, exclusion criteria);
- power calculations undertaken to ensure that sample sizes were adequate to test the hypothesis considered;
- absence of prior information on sample variances, e.g. non-availability of data;
- the type of analysis undertaken, along with an appropriate bibliographic reference.
- any transformations applied to the response variable or of explanatory variables;
- any model simplification (e.g. by deletion from a maximal model);
- how temporal or spatial autocorrelations (if any) were dealt with (pseudo-replication);
- how, exactly, randomization was conducted, or why it was not.
- how post-fit model checking was conducted, including tests of normality and constancy of variance.
- for bootstrap tests, why, if so, less than 10 000 realizations were done.

Authors should use the following code as a guide. If the same abbreviations are used for a different purpose, then such a purpose must be defined. That said, the use of undefined F, N or P in chemical contexts is unlikely to cause confusion. Statistical symbols are always printed *italic* except if this might cause confusion.

b: simple linear regression coefficient of *y* on *x*, where *x, y* have already been clearly identified, modifiable to b_{yx}, if the variables need to be specified. 'Slope' is not an acceptable synonym for 'regression coefficient'.

F: ratio of two independent error mean squares, usually, but not invariably, in an analysis of variance. Degrees of freedom (d.f.), plus the probability of the null hypothesis being true, must be stated with each mention (e.g. $F_{2,26} = 5.57$, $p = 0.01$). *F* ratios always have two values for d.f., one each for the numerator and denominator.

n: sample size or number of observations (not *N*).

p: probability (not *P*). In significance tests, reduce $p = 0.00\ 025\ 83$ to $p = 0.0003$ or $p < 0.005$, unless there is a specific need to be more exact.

r: product moment or Pearson correlation coefficient (*no other*), modifiable to r_{xy}, if the variables involved need to be specified.

r_S: Spearman's rank correlation coefficient.

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

χ^2: Statistic used in many well-known tests of significance relating to tables of frequencies where, if properly calculated, it has a probability distribution of several independent $N(0,1)$ variables, where '$N(0,1)$' denotes a normal distribution with mean 0 and variance 1. The degrees of freedom must be stated.

d: differential symbol (not *d*).

d.f.: degrees of freedom. These should be presented as subscripts (e.g. $F_{3,27}$) unless the accompanying text includes the d.f. For a correlation coefficient, the d.f. should be stated unless the number is evident as being two less than a previously stated number of pairs of values.

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

s.e.: standard error, always of the mean (s.e.m.) or of an estimate (s.e.e.) of some other parameter. Make clear that if an s.e. rather than a s.d. or confidence interval is being used, e.g. write 'the mean was 4.67 ± 0.32 (1 s.e., $n = 23$)'.

ln: natural logarithm (i.e. $\log_e$).

log *x*: common logarithm ($\log_{10}$ of *x*).

±: this sign may introduce the s.e.m. whose numerical value has preceded it.