The presentation of statistical results in journal articles

John W. Hall

Research Centre, Agriculture and Agri-Food Canada, Summerland, British Columbia V0H 1Z0. Contribution no. 985

Hall, J. W. 1997. **The presentation of statistical results in journal articles.** Can. J. Plant Sci. **77**: 11–14. Of the instructions about statistics given to authors, compliance in the Canadian Journal of Soil Science is lowest in providing a measure of variability when reporting quantitative data. The standard error is the most useful of these measures. The coefficient of variation, standard deviation or range are sometimes appropriate alternatives. The least significant difference (LSD) is sometimes used in this role, but authors are then tempted, incorrectly, to omit it when differences are not statistically significant. The LSD can be replaced by the standard error and the rule of thumb that the LSD is approximately three standard errors. Tables of means are easier to read when rows and columns are ordered so the most important variable is in the leftmost column, trends are evident and data relating to comparisons of greatest interest are close together. Tables of correlation coefficients can also be sorted to highlight relationships among the variables. In future, some readers may demand more statistical detail so that statistical methods can be used in literature reviews. Electronic publishing may allow this demand to be met without compromising good communication.

Key words: Data presentation, measures of variation, table layout

Hall, J. W. 1997. **La présentation des résultats d'analyse statistique dans les articles scientifiques.** Can. J. Plant Sci. **77**: 11–14. De toutes les recommandations faites aux auteurs sur la présentation des données statistiques, c'est celle concernant les mesures de la variabilité des données quantitatives qui, dans les articles publiés dans le Canadian Journal of Soil Science, est la moins bien respectée. L'erreur-type est la plus utile de ces mesures, bien que les coefficients de variation, l'écart-type et l'amplitude soient également des solutions valables. La plus petite différence significative (PPDS) est parfois utilisée pour cela, mais les auteurs ont alors tendance, à tort, de l'omettre lorsque les différences ne sont pas significatives au plan statistique. On peut remplacer la PPDS par l'erreur-type qui, grosso modo, correspond à environ le tiers de la PPDS. Les tableaux des moyennes sont plus faciles à lire lorsque les rangées et les colonnes sont classées de telle sorte que la variable la plus importante se trouve dans la colonne de gauche, que les tendances soient évidentes et que les données portant sur les comparaisons les plus intéressantes soient étroitement regroupées. On peut aussi disposer les tableaux des coefficients de corrélation de façon à faire ressortir les rapports existants entre les variables. Il se peut qu'à l'avenir les lecteurs exigeront plus de détails d'ordre statistique de manière à pouvoir les utiliser dans les revues de la bibliographie. L'édition électronique devrait aider à satisfaire cette demande sans pour autant nuire à la bonne compréhension des résultats.

Mots clés: Présentation des données, mesures de la variabilité, disposition des tableaux

In the Agricultural Institute of Canada (AIC) journals, the instructions given to authors concerning statistics are:

1. The statistical design should be described briefly and clearly.
2. Data should be analysed and summarized by appropriate statistical methods.
3. Authors should examine closely their use of multiple comparison procedures.
4. A measure of variability, e.g. standard deviation or standard error must be provided when reporting quantitative data.

Overall, the Canadian Journal of Soil Science does well in maintaining these requirements. But scrutiny of the journal since 1991 suggests there is room for some improvement in about half the papers published. The other half, which pass muster, include studies in which statistical methods were not appropriate tools and were not used. Those papers needing improvement include isolated examples of problems in the use of statistical terminology, the choice of statistical methods and overuse of multiple comparisons. But by far the most common failing is in omitting a measure of variability. The primary objective of this paper is to discuss the reporting of variability. As well, some features of good tables and statistical terminology will be discussed and some future trends in statistical reporting will be considered.

REPORTING VARIABILITY

The reader should see some measure of variability whenever mean responses to the experimental treatments are reported. The same requirement applies to regression coefficients and other statistical estimates. Measures of variability are statistics such as the standard error, coefficient of variation, standard deviation or range. Such a statistic shows the reader the precision of the results and allows him to assess them further. Proportions based on counts (i.e. binomially distributed data) and totals of Poisson distributed counts are their own measure of variability so no additional statistic is needed.

In addition to reporting quantitative responses to treatments, papers often contain initial tables summarizing quantitative information about the conditions of the study. They may describe the soils involved, the weather, other agronomic information or the composition of the diets used in an animal study. Measure of variability are not required in such tables although one, such as the range of soil properties, may sometimes be useful.

Why is Variability Neglected?

One can speculate why scientists continue to have such difficulty in reporting variability. The possibility that they are unaware of it can be ruled out immediately. Anyone who works with animals, crops or soil encounters it continually. The reasons are more likely to reflect bad habits of the past, idiosyncrasies of human cognition or lack of an intuitive feel for the meaning of the appropriate statistics.

Bad habits of the past are still with us. A lack of significant differences among reported means does not eliminate the need to report a measure of variability. Neither does the presence of results from a multiple comparison procedure. Statistical tests and statistical estimates such as the standard error are conceptually distinct and are not interchangeable. So a measure of variability should always be presented.

In spite of training and experience, scientists may still be falling prey to the “law of small numbers” (Tversky and Kahneman 1971). This “law” is the observation that everyone has an exaggerated belief in the likelihood of successfully replicating an obtained finding. It is seen in untrained subjects and trained scientists alike. This false belief naturally leads to ignoring quantitative information about the imprecision of results and consequently to the failure to provide it for readers.

It is also likely that scientific training provides insufficient practice to become adept with these statistics and to counter faulty intuition. This is regrettable when 1996 is the 100th anniversary of the introduction of the standard error and coefficient of variation. One can become skilled in their use by learning some rules of thumb. This is analogous to the way people raised on Fahrenheit have learned to think in Celsius.

The Standard Error

The standard error is the most useful of the statistics that measure variability. In addition to indicating precision, it can be used for carrying out further statistical tests should the reader desire this or for comparison with other studies. A useful rule of thumb for normally distributed variables is that the observed sample mean will be within one standard error of the true mean 66% of the time, two standard errors 95% of the time and three standard errors 99% of the time. The standard error of the difference between two sample means is about 1.5 standard errors and the least significant difference is about three standard errors.

Wherever possible, the pooled standard error, which applies to a group of sample means, should be reported rather than presenting a separate standard error for each mean. This reflects the requirement of constant variance in analysis of variance and can be readily calculated from the residual mean square. Using pooled standard errors also reduces the size and increases the readability of tables.

Other Measures of Variability

Although other measures of variation are sometimes appropriate, they are less often of use. They should be used only if the context warrants. The standard deviation measures the variation among individual observations rather than the precision of the mean. The coefficient of variation incorporates the standard deviation and so also refers to individual observations. Further, the coefficient of variation is only meaningful for variables that cannot take on negative values and for which the standard deviation is proportional to the mean. These are variables for which the logarithm would be the appropriate variance stabilizing transformation in analysis of variance. Ranges have limited use as summary statistics because they are dependent on the sample size.

The least significant difference (LSD) is sometimes reported as a measure of variation. This has an advantage that the reader can readily identify treatment differences. However, an author using a “protected” LSD for means separation (that is, doing a means separation only if the analysis of variance shows a significant difference) is tempted to omit the value when there is no significant difference. This violates the rule that a measure of variation should be presented regardless of the outcome of any statistical tests. The LSD has the further disadvantage that any reader wishing to use a different means separation test or do further statistical analyses must work backward from it to the standard error before proceeding. Reporting the standard error and remembering the rule of thumb that the LSD is about three standard errors is more effective.

TABLES

Tables of results in scientific manuscripts generally have two roles. They serve in part to provide quantitative support for the statements in the text or to clarify a complex argument and in part for archival and reference purposes (Finney 1986). In this latter role, they may be put to uses that will not be immediately evident to the author such as being included in a review of the field of research. This leads to a tension in designing a table. For the first role the table should be as simple as possible but for the second it should be more extensive. What is done in actual practice is probably a reasonable resolution of this tension.

Means Tables

Many tables present the means for several variables from an experiment with a factorial design. The data have often been analysed using analysis of variance. The outcome of the analysis should be reflected in the layout of the table. In interpreting an analysis of variance, the higher order interactions, if statistically significant, must be understood before the lower order ones or the main effects may be examined. To ignore this procedure leads to misinterpretation of the results as the effects of higher order interactions are masked in lower ones and in main effects. A table should display the means corresponding to the highest order interaction that is statistically significant and shows a biologically important effect. For example, in a factorial experiment with two rates of phosphorous and three of nitrogen, a significant interaction between these two nutrients would be reflected in a table showing the means for all six treatment combinations. In the absence of an interaction the means for the three rates of nitrogen and two rates of phosphorous would provide all the information.

Reflecting the results of the statistical analyses in the tables of means will lead to different layouts for different variables. It may sometimes be possible to divide the variables into groups depending on format and to have a different table for each group. When multiple tables are not practical, the best strategy may be to remember that including more detail than absolutely necessary is preferable to omitting essential information.

Improving the Layout

The layout of many tables could be improved with little effort. Ehrenberg (1977, 1981) of the London Business School, gave the following guidelines:

1. Giving marginal averages to provide visual focus;
2. Ordering the rows or columns of the table by the marginal averages or some other measure of size (keeping to the same order if there are many similar tables);
3. Putting figures to be compared into columns rather than rows (with larger numbers on top if possible);
4. Rounding to two effective digits;
5. Using layout to guide the eye and facilitate comparisons;
6. Giving brief verbal summaries to lead the reader to the main patterns and exceptions.

Ehrenberg had in mind business and official statistics like those produced by Statistics Canada. For scientific work some modification of these guidelines may be necessary. Marginal totals are rarely appropriate. The rows in tables are often treatment rates or concentrations and these rates should be in increasing order downward even though this leaves the smaller responses on top. The columns in tables are often different variables and can be ordered by putting the variable of most interest or importance on the left. If some choice is left in ordering the rows, as it would be if the factor levels were qualitative, the values of this leftmost variable can be sorted in descending order. Variables in tables of correlation coefficients can also be sorted to make relationships among them clear by putting the larger correlations next to the diagonal. A good order for the variables may be given by the tree diagram from a cluster analysis.

Analysis of Variance Tables

Analysis of variance tables do not need to be published unless their numerical values are essential to the presentation. This might be the case in studies of quantitative genetics or if the complexities involved in the analysis can be more readily explained with reference to the table. When an analysis of variance table is published, it should include the sources of variation, degrees of freedom and mean squares. The residual error should always appear in the analysis of variance table. This does not preclude the need for standard errors in the tables of means.

Analysis of variance tables usually include some indication of statistical significance. Statisticians disagree on whether tail probabilities ("*P* values") should be reported. They allow readers to choose their own critical values (Bailar and Mosteller 1988) but may give a false sense of accuracy because the distribution may not be exactly normal (Finney 1993). On balance, tail probabilities seem preferable to the usual stars.

Restraint should be exercised in presenting extensive tables of tail probabilities or stars to indicate statistical significance. Some authors argue that these tables allow them to summarize large amounts of information about the statistical tests concisely so their use cannot be entirely ruled out. However it is far more important to show how big the effect is and how precisely it has been measured. Extensive tables of stars have been referred to in the statistical literature as "statistical star wars".

STATISTICAL TERMINOLOGY

The correct use of statistical terminology is, and will probably remain, a challenge for authors because so few fully trained statisticians are available to provide guidance. A statement like "... differences are statistically different at the 95% confidence interval" (found in a meteorology journal) shows complete conceptual confusion. The authors are apparently reporting the results of statistical tests. If so, they probably meant "... differences are statistically different at the 5% significance level." The significance level in a statistical test is the probability of rejecting the null hypothesis (usually of no difference) when it is in fact true. It is a small value like 5% or 1% and never a large value like 95%. Confidence intervals are not statistical tests at all but, like means, are estimators of the statistical parameter of interest. The mean estimates the parameter by a single value while the confidence interval indicates a range in which the parameter probably lies. The term "confidence coefficient" should be used only in reference to a confidence interval and never to a statistical test.

Statements in the text like the "the mean was 14.2 ± 1.9" are ambiguous because it is unclear whether the 1.9 is a standard deviation, standard error, or indicates confidence limits. It is better to write "the mean was 14.9 (SE 1.9)" so there cannot be any confusion. The addition of plus and minus signs as in "the mean was 14.9 (SE ±1.9)" would never be seen in a statistical journal. The notation 14.2 ± 1.9 is sometimes used in a single column in a table. This is acceptable provided the statistic represented by the 1.9 is clearly indicated in the table. When a standard error, or standard deviation, occurs alone either in the text or in a separate column in a table its sign is always positive and need not be indicated.

LOOKING TOWARD THE FUTURE

How will statistical results be reported in the future? More graphs will likely be used to enhance communication. This positive trend has already started (Vessey 1996). Two other developments bear watching. These are an increased interest in and manuscripts about computerized management systems and an increased use of "meta-analysis".

Statistical methods for the development and assessment of computerized management systems are somewhat different from those that have been commonly used in agriculture. Good sampling designs are required to provide sound data sets for calibration and validation of the models. Regression methods are used to develop prediction equations. Finally multivariate tests are often incorporated in the software to automatically identify conditions outside the range of calibration when the systems are in actual use. This is the medley of statistical techniques now being used, for example, in the analysis of plant materials by near infrared spectroscopy. The appropriate graphics and statistics, such as the prediction error, are somewhat different from those used for factorial experiments. Scientists need to be aware of this and to become familiar with the appropriate statistical techniques.

In writing a review paper, the author must often determine whether differing conclusions in the published results of similar experiments reflect genuine disagreement or are only the consequence of random variation. Using statistical techniques for resolving these issues and combining the results to get an overall assessment of treatment efficacy is called “meta-analysis”. While much meta-analysis can be done with only means and standard errors, analysis of variance tables are useful to the analyst while the graphical displays are a hindrance. Thus the meta-analyst emphasizes the archival or reference role of a paper. This conflicts with the need for economy of space and clarity of presentation. If meta-analysis is to prove useful in agricultural research as it has in medicine, this conflict between the archival and communicative role of papers must be resolved. Perhaps the transition from print to electronic publishing will allow for publications with two parts: one whose role is to communicate and a second whose role is to provide the more detailed material needed for meta-analyses.

CONCLUSION

The aim of statistical presentation is to provide a clear and informative account of the results of a piece of research. It assumes that the underlying experimental design and statistical analyses were appropriately done. The material presented here cannot be exhaustive and has deliberately focused on a few topics where the need for improvement appears greatest or might easily be achieved. It is hoped that this may contribute to the continued improvement of reporting in the AIC journals.

Bailar, J. C. and Mosteller, F. 1988. Guidlines for statistical reporting in articles for medical journals. Amplifications and explanations. Ann. Intern. Med. **108**: 266–273.

Ehrenberg, A. S. C. 1977. Rudiments of numeracy. J. R. Statist. Soc. A, **140**: 277–297.

Ehrenberg, A. S. C. 1981. The problem of numeracy. Am. Statist. **35**: 67–71

Finney, D. J. 1986. Tables and diagrams: Will they help or hinder? J. Ind. Soc. Agric. Statist. **38**: 60–70

Finney, D. J. 1993. On presenting results of statistical analyses in scientific journals. Biom. J. **35**: 499–509

Tversky, A. and Kahneman, D. 1971. The belief in the “law of small numbers”. Psychol. Bull. **76**: 105–110.

Vessey, J. Kevin 1996. The conversion of cellulose to electrons: The future of scientific journals from a scientist’s perspective. Can. J. Plant Sci. **77**: 1-9.

This article has been cited by:

1. Michèle Bernier-Cardou, Francine J. Bigras. The Analysis of Cold Hardiness Experiments 403-435. [Crossref]
2. S.D Siciliano, J.J Germida. 1999. Taxonomic diversity of bacteria associated with the roots of field-grown transgenic Brassica napus cv. Quest, compared to the non-transgenic B. napus cv. Excel and B. rapa cv. Parkland. *FEMS Microbiology Ecology* **29**:3, 263-272. [Crossref]
3. C. G. Kowalenko. 1997. Letter to the Editor. *Canadian Journal of Soil Science* **77**:2, 331-331. [Abstract] [PDF] [PDF Plus]