

Statistical Guidelines

Bal K Joshi
Biotechnology Unit-NARC, Khumaltar

For interpretation of the results, there are a number of statistical tools those can help for getting logical inference. Following points might be suggested to be considered during running in the road from experimentation to publications.

- Appropriate statistical procedures depending on the objectives, treatments structure and experimental design should be selected for analyzing the data.
- To correctly analyze an experiment all terms must be accounted for because most computer software packages automatically pool any unaccounted source of variation with the error term.
- For a set of data with summary statistics, useful statistics to present are the mean, the number of observations and a measure of the variation or scatter of the observations as well as the units of measurement. The range or the SE is useful measures of the variation in the data. SE is not relevant in this context, since it measure the precision
- Significance testing is not the be-all and end-all of a designed experiment. The tests, which we do, are designed to give confidence that any treatment differences, which we may have observed, are real differences and not chance ones.
- To report that treatments are significant is meaningless. It is essential to state clearly which pairs or groups of treatments are being tested against what.
- Multiple comparison tests (MCTs) may be used for unstructured qualitative treatments. Indiscriminant use of MCTs can result in loss of information and reduce efficiency when more appropriate procedures are available. It is more appropriate to form planned sets of comparisons in experiment involving qualitative treatment.
- For experiment involving factorial sets of treatments or graded levels of quantitative factors there is almost always statistical procedures which can be specified in advance and which is more appropriate than MCTs.
- Whatever the methods of presentation some appropriate measure of variance eg SE of means should be included.
- Present treatment means from results of single factor experiments and SE or SED or LSD and the other possibly significance probabilities (report actual probabilities)
- Report additional statistics (CV, SE, etc) to assess data quality
- Present relevant treatments means, a measure of their precision and significance probabilities from the result of ANOVA
- The measure of precision should be either a SE or a SED or LSD. The SED and LSD are usually suitable for balanced design. Only one of these three statistics is necessary and it is important to make it clear which is being used
- Make clear whether the number after the \pm sign is a SD or a SE or a confidence interval
- For unbalanced experiments each treatment mean has a different precision. Presenting results is to include column of SEs next to each column of means and also a column containing the number of observations
- If there is no interaction then the main effect means should be presented along with SEs and significance probabilities for each variable
- If there are interactions then main effect means alone are of limited use. The individual treatment means should be presented.
- For a balanced design, there is only one SE per variable (except for split plot designs) and three rows giving F test probabilities for the two main effects and the interactions.
- For unbalanced factorial experiments additional columns for numbers of observations per group and SEs for the individual means should be added
- The key results of a linear regression analysis are usually the b, its standard error, the intercept or constant term, r and the residual standard deviation. For multiple regressions there will be a numbers of coefficients and SEs, and the coefficient of determination (R^2) will replace r.
- Line graphs or histograms portraying quantitative data usually are much more meaningful and easier to understand than are massive amounts of tabular data. It is hoped that researchers will recognize a more powerful statistical tool to be used in their research.
- If error bars are displayed on graph and charts, it must be clear whether the bars refer to SD, SE, LSD, confidence interval or range.
