Notes on the writing of reports **Statistics 347** **March 2012**

1. Reports should be no longer than necessary. A short report that makes the salient points is preferable to a long rambling philosophical essay, even if the longer essay makes the same points. Above all, have compassion for the reader (and grader).
2. A plot either of the raw data or of the residuals is almost always essential at some point in the analysis. Not all plots will be helpful or especially interesting. Although you should indicate what plots were made, it is generally not necessary to include in the report a copy of all plots made and all analyses performed. If necessary for examination purposes, extra plots and lengthy analyses can be included in an appendix.
3. All major conclusions should be stated at the beginning in a summary intended for a scientifically literate reader who is not a statistician. Technical terms associated with the context of the problem are unavoidable, but technical statistical terms should so far as possible be avoided. One page is the upper limit. Remember, few readers progress beyond the summary. It is up to the author to state the conclusions early in as persuasive a manner as possible if the reader is to be convinced.
4. Following the summary, the report should describe the models fitted, the tests performed, and how these support the conclusions. The relevance of the models to the context under study is important. Technical statistical terms are acceptable here only if they are essential to support the conclusions.
5. Adequate precision is important, but ordinarily two significant digits are sufficient for standard errors. Parameter estimates should always be given with standard errors, or standard errors of differences in the case of factor levels. It is often sufficient to say that the standard error is 9–15%, or the standard error of pairwise differences is 0.35–0.45, if the range is not excessive. Parameter estimates should be accurate to 10% of a standard error. By convention, p-values are given as a percentage: rarely is there a need for more than two significant digits. The listing of excessively many uninformative digits in estimates and standard errors betrays a lack of statistical sense, and will be penalized.
6. While it is necessary to demonstrate that you have mastered the computer system or statistical package, it is more important to demonstrate that you are not a slave to the computer by tailoring the computer output to the problem at hand. (i) From the computer-generated analysis-of variance table, list only the parts that are relevant to your analysis. (ii) If the model matrix is non-standard and cannot be generated by a model formula, as for example the additive skew-symmetric formula $E(Y_{ij}) = \alpha_i - \alpha_j$, you need to explain what the structure of the matrix is. (iii) Do not quote a p-value without stating the hypothesis under test and how the value supports your conclusion. (iv) Parameters have a physical interpretation: do not pass up the opportunity to remind the reader what the physical interpretation of $\hat{\beta} = 0.684$ is in the context of the problem.
7. Physical variables, unlike mathematical variables, always have units such as 'length in mm.,' 'temperature in °K,' 'mm. Hg.,' 'age in months,' or 'depth in fathoms.' If you lose sight of the units your conclusions are liable to be ridiculous. For a published example, see p. 105 in Andrews and Herzberg (1985) where, despite the fact that Adelaide borders on the Australian desert, its annual rainfall is given as 1530 mm., or an astonishing 60 in.
8. Reports should be logically organized and written in grammatical English. In particular, each sentence should have one, and only one, main verb. Poor logical organization betrays a confused mind, and poor sentence structure indicates a lack of attention to detail.
9. Good grammar is important insofar as poor grammar betrays confusion or faulty logic. For example, the basketball coach may *substitute* a bench player B for a starter A, or he may *replace* the starter with a substitute from the bench. An active player may be replaced, in which case an inactive player is substituted. Likewise, the upstream region of a gene may be rich in certain motifs, meaning that these motifs are abundant in the upstream region. The upstream region is enriched with motifs, but the motifs themselves are neither rich nor enriched. Coal is abundant in Wyoming and fruit is plentiful in Florida, but coal is not enriched in Wyoming nor is fruit rich in Florida.
10. It is sometimes helpful to indicate what further analyses might have been helpful had there been more time or different software available.