What constitutes responsible statistical criticism?

'Much that nowadays passes as statistical criticism is superficial and
sophomoric in character and serves to obscure a scientific discussion
rather than to clarify it' (Bross, 1960).

Any piece of statistical analysis involving human beings, and quite
possibly any involving real data, will suffer from the violation of the
standard mathematical statistics assumptions, assumptions which we were
all taught in our statistics courses were necessary if the operations were
at all legal.   However this does not mean that the entire corpus of
research using statistical techniques should be dismissed immediately.
There is a tendency to attack the statistical or research techniques used
if one disagrees with the result of a piece of research, and this is
perhaps inevitable given the strongly legitimating function of the use
of statistical techniques.

Not all such criticism is sensible. While, except possibly in one's
own research, one tends to examine conclusions one likes less closely
than those one dislikes, we have all (I think) come across examples of
attempted hatchet jobs which have taken relatively minor departures and
attempted to blow them up into fundamental weaknesses, and such off-the
mark attacks have frequently strengthened rather than weakened the piece
of research they were assaulting - it would be invidious to give examples.

Part of such attacks is of course simply a deliberate strategy and not
motivated by statistical respectability ('I may not know much about
statistics, but I know what I like') but it would be helpful to those
involved in carrying on or spectating at a debate to know what constitutes
'responsible' criticism.

If we attempt to draw up a set of guidelines, should they be relevant
particularly to radical statisticians, or to all statisticians? If we
do decide on the latter, are we to some extent prejudging the issue of
professionalisation and putting ourselves forward as 'experts' by setting
up a 'code of conduct'? I feel that we should try to draw up such a set
of guidelines as, whether we like it or not, statistics are used in this
legitimising way, and such an agreement amongst statisticians should help to
get rid of this mystificatory function. It should also be accompanied
by a determined attempt to demystify the question under discussion.

Irving Bross in an article entitled 'Statistical Criticism' (1960) sought
to present guidelines. He claimed that almost all criticism could be
understood as the presentation of a counter-hypothesis. For example,
to criticise a report for not using significance tests is, he argues, to
claim that the observed differences could be explained by sampling varia-
tion, rather than the proponent's preferred explanation; if the differences
are actually large enough to have been statistically significant had the
tests been carried out, then, argued Bross, the objection was 'trivial'.
I think that this is a mistaken emphasis, since unless the sample sizes
and the differences are so large as to make such tests redundant, it is
normal to use them, and the fact that the differences are subsequently
found to be statistically significant seems more to be an example of being
right for the wrong reason. It has the same logical status as saying
it's all right to fire a machine gun in a public street provided you don't
hit anyone.

For Bross, the counter hypothesis must be not merely 'conceivable' but
'tenable'. 'Tenable' means not just 'plausible' (in Campbell and Stanley's
sense of 'consistent with existing theory in the field') but also 'corrobo-
rated' by the data in the proponent's report, or other existing data, and
also that accepting the counter-hypothesis would make a difference to the
conclusions reported.

This raises the question of where the onus of proof lies, with the critic
or with the researcher. Bross's position seems to me to put the onus
far to much on the critic, and lean over backwards to support the writer.
For example under Bross's suggestions, one could escape criticism by
doing particularly bad research, not giving proper descriptions and
not collecting data relating to any alternative explanations. I am sorry
to have to say that given the standard of much of present educational
research, one cannot be certain that all published work is even competent
without much closer inspection than is possible from published descriptions.

Also he seems to adopt a scientist's matter-of-fact 'Either it (probably)
is there or it (probably) isn't' approach. Yet it seems to me that research
may be misleading in much more subtle ways. In order of increasing subtley
and difficulty in countering examples (a) The interpretation is wrong,
i.e. the data contradict it. This, naturally enough, is almost unheard
of in anything of any consequence in published research. Rutter et al
(1979) seem to give a description of one table on p. 92 which is the opposite

17

of what it actually shows. However, this does not in any way affect the conclusions of the study.

(b) The data do not either confirm or contradict the interpretation. Pratt et al. (1972) in a study of local education authorities compared rateable resources and expenditure, and came to the conclusion that Scottish authorities were more 'generous' than English ones for given resources. It was however quite unwarranted to reach such a conclusion since the Scottish data were for an earlier year than those for England.

(c) Something like the interpretation given may be true but it is not justifiable to come to such black and white conclusions. I discuss this in relation to the work of Rutter et al below.

Statistically or Politically Responsible?

Should one's notion of responsibility in criticism address itself to the statistical respectablity of the research or it s political acceptability? Personally, I feel the former is more important; firstly because an agreed procedure among statisticians should help in the cause of demystification, so that non-statisticians are better able to make up their own minds; and secondly because 'radical' (or other) positions may change and we should not shut our minds to what could be valuable evidence.

I list below, some possible criteria for guidelines to responsible criticism. People will probably want to argue.

A possible outline of a guide

1. The aim of a statistical criticism should be to inform general discussion, making the issues clear as far as possible and interested non-statistician. Of course, this is closely related to the demystifying process.

2. Criticism should be confined to indicating departures from assumptions only when there is a real danger of invalidating results. Such 'real danger' need not be confined to evidence from the research under consideration, but could quote results from allied fields or statistical theory.

3. This last point applies particularly to 'third variable' explanations. It has been said that any social scientist who has difficulty in thinking of at least three explanations for any correlation of interest should probably look for another job. However, a third variable' explanation should not be put forward as an explanation unless one has good reason to believe that this differs between

subgroupsand that differences are related to differences in the two correlated variables.

4. This presupposes a certain quality of research competence and presentation. If the research appears to ignore real difficulties, then this should render it suspect, and one should say so.

5. Statistical criticisms should be made in such a way that the important arguments are clear to a non-statistician.

6. The possibility of secondary analysis is important. While this is something that is generally difficult for an individual to do anything about, refusing to deposit data with the SSRC Archive ought to be viewed in the same light as a second hand car salesman refusing to permit an expert inspection.

7. The critic should state his own political/ideological stance.

8. Discrepancies between careful interpretations in the research chapters and wide claims in the conclusions are important.

9. It is important to explain any 'hidden' implications of techniques.

10. One should not make a criticism which the researcher in principle could not have met.

11. It is important to clarify the difference between '.statistical' and 'substantiative' significance and between significance testing and parameter estimation.

Other developments

Points (4) and (6) relate to the statistician or researcher working with the project rather than the critic. In particular, however all concerned should emphasise the importance of secondary analysis.

There should be more statistical examination of robustness of techniques

I shall attempt to illustrate the question of responsible statistical criticism by referring to 15000 Hours by Rutter et al. I am indebted to Ludi Simpson for the following summary of the book and the round up of statistical criticisms.