

ONE-DAY CRICKET MATCHES AND CUSUMS

by Peter Thakurdas

CUSUMS are fast becoming the simplest way of watching performance and checking whether targets are being met.

One interesting CUSUM application was in the 2nd innings of the NZ vs India 1-Day Cricket match. The particular match in the example below had a breath-taking finish and was held at the MCG in Melbourne, Australia, on 23rd January, 1986.*

India had completed the 1st innings with an impressive 238 runs for 8 wickets. NZ batsmen were chasing a total of 239 runs from 50 overs, which translates to a target run rate of 4.78 runs per over.

The CUSUM, in simple recursive equation form, is:

$$C(j) = C(j-1) + (R(j) - T)$$

where: $C(j)$ is the CUSUM in the j 'th over.

$C(0) = 0$, ie. the initial value of the CUSUM.

$R(j)$ is the runs scored in the j 'th over.

T is the CUSUM Target or reference value. ($T=4$)

For ease of mental arithmetic and plotting, I chose an integer reference or target rate of 4 runs per over (ie. set $T = 4$). I found I could plot the CUSUM quite easily while watching the match on TV.

Figures 1 and 2 are plots of the runs, $R(j)$, and CUSUM, $C(j)$, respectively. (The original graphs were of course drawn by hand).

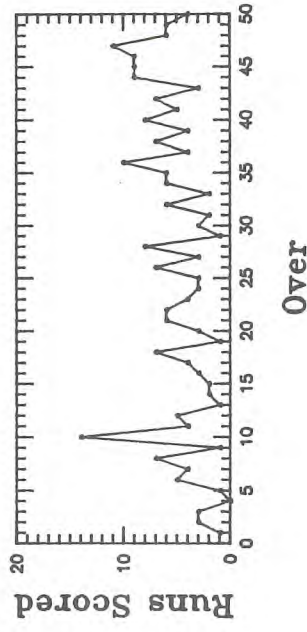
To interpret the CUSUM plot, bear in mind that the SLOPE represents the run rate. Dramatic changes in slope correspond to dramatic changes in run rate. Use the protractor in Figure 2 to convert the CUSUM slope to a local average run rate. I could get an idea of the required run rate at any time by looking at the slope from the latest CUSUM point to NZ's goal, marked by an "X".

Note how clearly the 34th over shows up as a change point on the CUSUM plot. Some may recall, this was when India's bowlers Ravi Shastri and Yadav were in the tail end of their overs and NZ batsmen Martin Crowe and John Reid were taking advantage of this. From then on, NZ maintained a high run rate (6.7). NZ batsmen, Jeff Crowe and Jeremy Coney, helped maintain the high run rate in the final 7 overs to clinch victory from India with only one ball of the match to spare.

Another interesting observation on the CUSUM plot was a temporary drop in the run rate whenever a new batsman entered the game or a new set of bowlers began their overs.

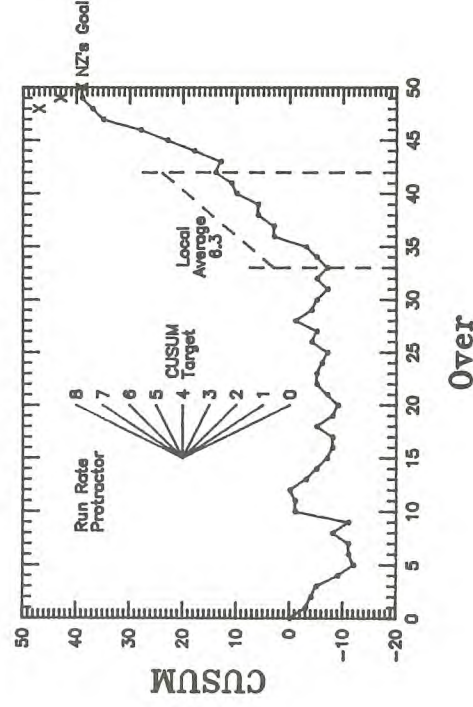
Both of these observations are not nearly as clear in Figure 1, (ie. the plot of runs scored in each over). Figure 2 exemplifies the effectiveness of the CUSUM plot for presentation of time series data.

Figure 1. NZ vs India (NZ Run Rate) 23 January 1986



© Copyright 1986, Applied Maths Div. DSIR.

Figure 2. NZ vs India (NZ Run Rate) 23 January 1986



© Copyright 1986, Applied Maths Div. DSIR.

BOOK REVIEW

by John H. Maindonald

"Directory of Statistical Microcomputer Software", 1985 edition, by W. A. Woodward, A. C. Elliott and H. L. Gray. Marcel Dekker, Inc., New York. Price \$US54 (outside USA).

Microcomputer statistical packages are in much the state of the mainframe packages of ten years ago. There is great variety, and the quality is uneven. Most are untested, and there have been few reviews that are based on the experience of competent statistical users. This volume catalogues much of what is available, but will

be of limited help in choosing between rival packages. Most of the space is taken up with lists of features that summarise responses to a questionnaire sent out in 1983 and 1984 to statistical software developers. There has been no attempt at the daunting task of providing critical evaluations of the 140 packages. The modest information which it does provide is welcome.

In a few cases references are given to "reviews". Alas, a "review" is often an uncritical description that has appeared in a popular computing magazine, such as "Interface Age" or "Popular Computing". By contrast BMDPC and MINITAB claim no reviews, even though their mainframe versions have featured in several comparative reviews undertaken by recognised statistical professionals. It is likely that there are similar anomalies in the features listed.

It is interesting to look at the computer systems used by the 140 listed packages. The older "8-bit" style of microcomputer, such as the Apple II (with 49 packages offered), the Tandy Model III etc. (21 packages), and various CP/M machines (31 packages), have largely given way to the IBM PC and its lookalikes (109 packages). These are usually claimed (not quite accurately) as "16-bit" machines, and run either the PC-DOS or (in the case of lookalikes) the MS-DOS operating system. Most major manufacturers are now offering or planning to offer this style of machine. Software developers that have a continuing interest in their product and wish to get it widely used must soon provide a PC-DOS or MS-DOS version. Even if your use will be on an Apple II or Tandy Model III you may be wise to restrict your attention to packages for which a PC-DOS or MS-DOS version is available.

What of machines that belong to the "avante-guard" of microcomputing? The Apple Macintosh is mentioned once only. Much of what is available has appeared too recently for inclusion in this volume. Package developers require time to learn how to make good use of any innovative machine. The Commodore Amiga now offers an attractive alternative to the Macintosh, and is certain to seduce some current Macintosh devotees. It is too soon to look for scientific and statistical software for the Amiga, or to assess its place in the market.

The UNIX operating system caters for the high end of the microcomputer range. Three packages that run under UNIX get a mention.

Who will use this directory? I am glad to have it on my shelf, mainly as a contribution to the taxonomy of statistical packages. It will be useful in finding the very small number of packages that cater for some specific requirement, for example "Life Table and Survival Analysis." But I do not find it very helpful in discriminating between the large number of packages that cater for a variety of elementary analyses. Readers must search with some care to discover which packages are good prospects for processing survey data.

Given the limited information available from such volumes as this, how should users choose a statistical package? My first preference is for reputable packages, widely used by statistical professionals, that have been carried across from conventional computers. For these packages it is usually possible to find critical comparative reviews, which may not however be entirely relevant to the most recent version or to the microcomputer version. The speed of the microcomputer version should be checked—it may run slowly. The user interface is unlikely to be up to the standard of Lotus 1-2-3, but then most of the statistical packages written specifically for micros are no better in this respect. So far there's little that does for statistics

what electronic spreadsheets do for row and column calculations. What little there is merits, if it's alright in other respects, our support.

In all instances I'd like to be assured that packages have been written by or under the supervision of a statistical professional. Unfortunately the Directory does not list names of the people primarily involved in developing the package, or give references to papers or books they have written that might give some clue of their knowledge and competence. No slot was provided where the vendor might list a small number of published technical references explaining the statistical methodology, the guiding motives of the package developers, algorithms used, etc. Where a newsletter is published that allows users to exchange comments and ideas, details would have been helpful. Examination of the newsletter will provide clues to the quality of the package.

The best statistical packages bring a wide range of statistical analyses, some part of current statistical practice and some yet to be devised, within the scope of one of a small number of conceptual frameworks. Such conceptual frameworks are, in their turn, embodied in a statistical language. ANOVA in Genstat, and GLM's in GENSTAT and GLIM, are perhaps the best examples. EDA, as incorporated e.g. in MINITAB or in 'S', is another. The result is a marvellous power and economy of thought and language, revealing connections between what were formerly seen as disparate analyses, and suggesting extensions to or variations of intended analyses. Presumably it was hoped that the request to describe: "the unique and most important features of your product, i.e. what makes it distinctive..." would elicit comments in these terms. I'd like to add the question: "What concepts and major themes unify your program and make it a powerful tool for statistical analysis?" Indicate how these are reflected in the command language."

The only diagnostics that feature in the questionnaire are regression residuals, and some analyses (EDA routines and skewness and kurtosis) that are often used for diagnostic purposes. Information on any diagnostic routines ought to be requested under each category of analysis. Other questions I think should have been asked relate to:

- (1) Testing procedures; what basis is there for confidence that answers will be correct?
- (2) What sources were used for numerical and other algorithms?

The authors list as sources of reviews *The American Statistician*, *The Statistical Software Newsletter* (published by the International Association for Statistical Computing), such popular computing magazines as *Byte*, and *The Capital PC Users' Group*. An extensive bibliographic source for reviews is promised for the 1986 Directory. I conclude with my own list of other sources of information (which owes something to a list that Ivor Francis presented at the 1984 NZSA conference).

Sources of information on microcomputer statistical software:
 Cable, D. and Rowe, B. (1985), *Software for Statistical and Survey Analysis*. Study Group on Computers in Survey Analysis, 32pp. (c/o D. Cable, Central Statistical office, Great George Street, London SW1P 3AQ.)

Carpenter, J., Deloria, D. and Morganstein, D. (April 1984): Statistical Software for Microcomputers. (A Comparative Analysis of 24 Packages.) *Byte* 9:4 234-264.

MSU Development Working Papers (Dept. Agric.