

From life to general statistics

The latest release of Genstat represents evolution rather than revolution, Felix Grant discovers

I last wrote about Genstat exactly a year ago, based on field use of release 6.1 for Windows – the first under VSN, who now manage its development. It represented a great leap forward for this seminal statistics package – not least because its GUI finally came of age, putting it on a level with any competitor. VSN had, at that time, outlined a firm development timetable, with new releases in May (7.0 being due in May 2003) followed by an update in September. Something slipped, however, because that release was only announced for market right at the end of December. Now that it's here, though, there are solid if low-key developments right across the board.

The move to expand out of a perceived life-sciences persona, clearly visible in 6.1, is underlined by the six-sigma group of tools on the statistics menu. Likewise, the interface improvements have been built upon further, to make the program more familiar to new user groups – particular in the area of interactive exploratory graphics. On the downside of this, a couple of gremlins that occasionally interfere with Windows context switching have appeared – but they are benign, causing nothing worse than occasional user puzzlement. Survival analysis has been given a makeover, too, which points in the same direction, as does the improved provision for design and analysis of surveys.

From a more general viewpoint, the spreadsheet has been refined considerably. The most visibly dramatic of these changes is an increase in the maximum allowed number of columns from 250 to 16,000, but some of the subtler ones may, in the long run, have more impact. Improved routine management such as sheet expansion, column freezing and filling, subsetting and splitting are very welcome, for instance. So is direct read access to an expanded range of data file types. Genstat



'Genstat users are loyal, know the product to be a good one, and feel an intense relationship with development'

was already good at this in 6.1, but is better still now. Another strong area, data management facilities, is likewise built upon. On top of these procedural developments, which mimic generic spreadsheet products, come some original ones: a first row can be 'promoted', for instance, to provide not just column names but extended column descriptions. Particularly useful is calculation from a column of a randomly sampled subset. Of particular interest to me, from a selfish viewpoint, is set-up of GIS columns from the spreadsheet – one of several changes that strengthen Spatial Statistics (as geostatistics are now known, once again recognising general market-wide habits of usage).

There are numerous advances in detail, spread right across the range of facilities, with modifications or upgrades to many directives and procedures as well as some new additions. Summary statistics, for example, gain new menu options for weighting and tallies; at the other end of the scale, REML (an inboard version of VSN's residual maximum

likelihood tool) acquires menu access to mixed model prediction, as well as both regular and irregular spatial statistics grids. The multivariate analysis menu gains a scree plot (there are new graphics types elsewhere, too) as well as principal co-ordinate graphic options and loadings rotation. Useful cosmetic changes at the user level include colour-coded syntax-highlighting in text windows, sort-order display within data boxes, improved mouse control, intelligent data inputs, new progress indicators for processes likely to last noticeable lengths of time, and such like.

As with the last upgrade, and inevitably, given the shift in which Genstat is currently engaged, all this comes at a cost of some incompatibilities. For anyone upgrading from 6.1 to 7.0, problems should be trivial and infrequent, but need to be borne in mind; if you are upgrading in a leap across more than two generations, some debugging time is pretty much inevitable.

A lot has been happening at VSN over the past year, and the flow of information has become less predictable between releases six and seven. I have received queries from a number of users, contacted during previous reviews, who hoped (in vain) that I might know more than they about what was happening. This would not be worth commenting on, in most areas of the market, but Genstat is not an average case: it is an exceptionally good and unique product with a distinct history. Genstat users tend to be fiercely loyal, know the product to be a good one, and feel an intense relationship with development. An apparent shift of corporate culture from 'exceptionally open' to a more commercially usual 'somewhat closed' level is unfamiliar to them.

The release 7 that has emerged does not in any way disappoint, either as an inheritor of the heavyweight Genstat name or as a product in its own right. No moulds are broken, nor worlds shaken, this time around. It's an evolutionary release rather than a revolutionary one; but existing users will certainly want to upgrade, and the appeal has extended even more decisively into contention within the general applied statistics market.