

THE Cruncher

The year of upgrades

This is the year that many of our major suppliers have chosen to upgrade their products.

MATLAB 7 and its associated release 14 products has been out for several months now, and has become the regular version for us. This Cruncher is our first chance to tell you of some highlights, though.

STATISTICA 7 was released in August, just before we went to press. We have summarised some of the major points

from the StatSoft web site on page 2, and although we have yet to try it, the additions look as though they are really worth while.

In many cases they are improvements that were suggested by HRS (and probably by others as well).

Mathcad 12 has been eagerly awaited for some time. Most people will not find a large number of new features that they can see, but a lot of re-writing has gone on “under

the hood”, which results in increased speed and reliability, as well as better integration between Mathcad and other software used in today’s organisations. This version will be released at the same time as this Cruncher is published, so we are as keen as you are to see the final product.

To top this list off, Hydra Software has just released Version 3 of AULOS, as announced below.

SurveySolutions/EFM

Perseus has released an extension to its SurveySolutions Web surveying software that allows organisations to monitor and optimise relationships with customers, employees, Web-site visitors and others.

Enterprise information managers can centrally manage all of the survey practices of their organisation, with full control of the design, implementation, distribution and analyses of the feedback process. Contact Darrel for more information.

HRS Software Guide

We have just released our 2004 Software Guide. If you are reading this Cruncher in the tea room, or on the Web, and did not get the copy of the Guide that was mailed out with the Cruncher, then contact our office (admin@hrs.co.nz) to be sent your own free copy.

Many of the products are the same as in the previous Guide, of course, but we have a number of new ones, such as Strater, Surfer, NetZoom, Library Master, GrafICalc and SurveySolutions, among others.

We have re-organised and expanded sections on STATISTICA and MATLAB, to give a better overview of these products.

The HRS Software Guide is the ideal book to have on your shelf when you wonder “Is there a better way to do this?”

AULOS version 3

AULOS is a hydraulic modelling program that is made in New Zealand. HRS is involved in all aspects of preparing this program for the end user, although we do not write any of the programming.

Version 3 has just been launched, and it contains a lot of new features. The most notable one is the ability to show a map or photograph of a drainage catchment, as a background to help with drawing the channels in which the flow is to be modelled.

We believe this program is the most accurate and fastest program of its type. We face an obstacle selling it into a market that has for years used other software, but know that those who already use it are happy customers. Contact Marc to find out more.

An example of a Mathcad document on the Web

It can be hard for people who have not seen it to appreciate the strengths of Mathcad. Ray (our M.D.) was impressed by the following example – you might want to have a look it, even if you are not interested in the subject of riser-less ocean floor drilling.

Go to www.mathcad.com/support/item/04aug_user.asp.

| HRS Inside ... | |
|---|----------|
| STATISTICA version 7 | 2 |
| Metadata for cases and variables | 2 |
| Text searching | 2 |
| Selection conditions | 2 |
| Graphs use metadata | 2 |
| More data import options | 2 |
| Licensing changes | 2 |
| Analysis “By Group” | 2 |
| Major Users of STATISTICA | 2 |
| MATLAB 7 | 3 |
| Compiling..... | 3 |
| Graphing | 3 |
| User interface..... | 3 |
| GUI development environment | 3 |
| Publishing | 3 |
| Reports | 3 |
| Integer variables | 3 |
| Mathcad 12 new features | 3 |
| Graph enhancements | 3 |
| Units..... | 3 |
| Data import..... | 3 |
| Annotations | 3 |
| Provenance..... | 3 |
| Web controls..... | 3 |
| Correlation functions | 3 |
| XML..... | 3 |
| New STATISTICA Modules | 4 |
| Random Forests | 4 |
| STATISTICA NIPALS Algorithm | 4 |
| Link Analysis..... | 4 |
| Other | |
| MATLAB and Simulink Student versions..... | 4 |
| @RISK bundles..... | 4 |
| Seminars..... | 4 |
| Training Courses..... | 4 |

STATISTICA version 7

This great product has been improved! StatSoft have added some of the features that Ray really wanted, and a lot more of their own ideas that make analysis even more efficient. Take note of the “metadata” concept - it really is powerful.

We have described some of the more easily explained ideas on this page. There are lots more that are described at www.statsoft.com/v7list.htm

Metadata for cases and variables

Metadata is “Data about data”. For instance, in a spreadsheet with a column of numbers, the metadata might be that it represents a speed of a car. Knowing that makes the numbers much more useful. In *STATISTICA 7*, you can record metadata about both variables and cases.

One kind of metadata for variables allows you to specify that a variable is categorical or continuous in nature. This then allows the option of only showing variables with the appropriate type in variable selection list boxes. Another type of metadata property is used to define a variable as one that is used for naming points on graphs. A third type allows you to exclude variables from variable selection lists. There are others as well.

Case metadata contains additional properties of individual cases (records) in a data set. For instance, a particular subset of cases can be assigned a “star” marker, and will appear as such in all graphs. You can use case metadata to set the colour to be used in plotting points on a graph. Other properties are “Hidden”, “Labelled” and “Excluded”.

Text searching

STATISTICA now provides a broader selection of functions that can be used in searching for text strings in spreadsheet and case selection formulas (including so-called fuzzy text searching). These include searching for text, comparing text, text replacement, and comparing text using an operator similar to SQL’s LIKE keyword.

Selection conditions

STATISTICA 6 had great tools for selecting subsets of the total set of cases by entering conditional expressions. In version 7 you can automatically display the text of these conditions in the title areas of all subsequent output spreadsheets and graphs. This is a feature Ray (our M.D.) has been clamouring for!

Graphs use metadata

Because plots are now linked to the main data file (in version 6 they linked to a copy of the file) if you use scatter plot brushing techniques on one graph to set the case metadata properties of a set of points, then all other graphs based on that same data file will also show the same change in properties. This is extremely useful for graphical data exploration.

Statistical analyses take note of the metadata state, so that you can select subsets of cases graphically, perhaps to exclude outliers, and then perform any statistical analysis on this subset.

If you edit the data or recalculate it, graphs based on the data will automatically update.



Multiple graphs and the data file are linked in a way that allows you to see patterns in data

More data import options

You can now conveniently import or export SAS binary and portable formats, and binary JMP, Minitab, and SPSS files, providing easy migration from these legacy systems.

Licensing changes

STATISTICA Concurrent Licensing has been enhanced to allow for offline usage while a *STATISTICA* user is disconnected from the network, for supporting “trial period” usage of individual modules, and to allow more flexible combining of numbers of users for different modules.

See page 4 for new *STATISTICA* data analysis modules in version 7

Analysis “By Group”

“By Group” analysis means that complete analyses, whether Statistics or Graphs, are executed for each combination of levels of a set of “By Group” variables.

For example, selecting variables CITY and GENDER as “By Group” variables for a particular analysis will instruct *STATISTICA* to repeat this analysis for all subsets of cases that can be created by combinations of levels of CITY and GENDER (e.g., Wellington/Men, Wellington/Women, Auckland/Men, Auckland/Women,...). All levels within the CITY and GENDER variables are automatically identified and used for the categorisation.

This has been the single most requested feature from those people who have used SYSTAT in the past!

Major Users of *STATISTICA*

We are often asked “Who uses *STATISTICA*?” Since “Lots of people” is not a very useful answer, here is a partial list of the kinds of users we have, as at Aug 2004. Many of these licenses are multi-user licenses.

- 7 hospitals and DHB’s
- 7 Regional and district councils
- 5 Universities
- 4 major engineering organisations
- 4 market research organisations
- 3 Government departments
- 2 power companies
- 2 major environmental research institutions

There are many other customers as well, mostly with only one or two users.

Why don’t you contact Darrel to find out why this product is so successful?

MATLAB 7

You really needed to come to Ray's seminar to see all the nice new features! The 300 developers at The MathWorks have come up with a lot of very useful goodies in the 18 months or so since the last release.

Feedback at the seminars suggested the following are important points to note among the wide range of improvements and additions.

Compiling

All of MATLAB except proprietary GUIs is now able to be compiled, with no charge for deployment.

The compiling process uses only small amounts of C code, and instead the deployed application uses functionality from a large file called the "MATLAB component Runtime" (MCR).

This means that compilation is only used for deployment, not for speeding up computation. (MATLAB itself now runs faster, in order to bring you speed.)

Function calling

There are new and improved ways to define and use functions. Some commands will now give warnings when you use the old syntax, but they will still work till the next release.

Graphing

You can now use drag and drop methods to make quite complex graphs. You can create a MATLAB m-file from the interactively-created graph, and incorporate it in a program.

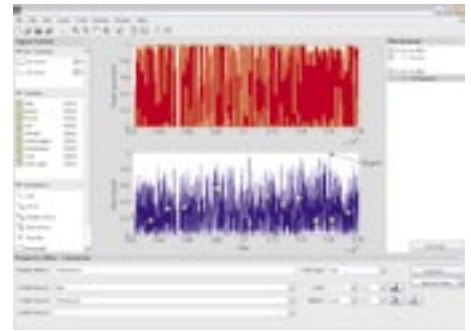
Write TeX and LaTeX annotations on your graphs, using the full LaTeX markup language.

User interface

Lots of nice ways to make life easy, including dockable windows and saved custom desktop layouts. Put a set of code in a button on a shortcut toolbar.

GUI development environment

New kinds of controls, and improvements to the existing controls.



MATLAB now has interactive plot tools to make it easy to create complex graphs

Publishing

You can publish your program files to HTML or other formats – headings will be created from your comments, and any graphs your program makes will be incorporated in the document.

Reports

7 different kinds of reports can be created from a set of files in a folder, including one that recommends changes in a program to allow you to follow best practise.

Integer variables

MATLAB now has full support for integer and single-precision maths.

Mathcad 12 new features

XML

You can now save Mathcad documents in XML format. This text-based format, related to the HTML used in Web browsers, is important to people who want to save Mathcad documents in a content-searchable form, so is particularly important in large organisations.

Because it is a format used by many different applications, it greatly facilitates collaboration between applications.

Correlation functions

New tools are useful for matching signals, telling the degree of similarity or difference, and providing information about signal lag, in one or two dimensions. (You can use two-dimensional correlation to determine similarities between a template image and a larger test image.)

Graph enhancements

Graphs with multiple traces can use both the left and right Y axes, with different scales on both. This allows you to use the data as calculated, rather than having to scale the data before plotting it.

The legends have more options than before.

Units

18 new units are included. You can define your own default units, so you no longer have to manually change units.

Data import

A new wizard allows you to read a wide variety of file formats. You can also access much of this functionality from a command, for use in programs.

Annotations

Although one of the strengths of Mathcad is that all the notation used is visible to any reader, unlike a spreadsheet, there are times when it is useful to annotate definitions of variables with a note that provides further explanation, to be read only when really needed. Mathcad now allows you to write notes that are readable in a pop-up window.

Provenance

Often you copy expressions from one document to another. The provenance property of an expression allows you to track this historical record.

Web controls

These are like the existing list boxes and buttons, but are much easier to set up for users who are not programmers.

New STATISTICA Data Analysis Modules

Random Forests

This is a new, powerful technique, now available in STATISTICA Data Miner, STATISTICA Text Miner and STATISTICA QC Miner. It is used for building flexible models for classification and regression and is particularly well-suited for extremely large numbers of predictor variables. See <http://stat-www.berkeley.edu/users/breiman/RandomForests/> for general literature on this method.

STATISTICA NIPALS Algorithm

This is an implementation of a number of techniques known as Principal Component Analysis (PCA) and Partial Least Squares (PLS). In STATISTICA, PCA and PLS are now implemented using the state of the art NIPALS algorithm (Non Linear Iterative Partial Least Squares) a mathematical procedure designed to extract systematic variations, relationships, and information in datasets. STATISTICA NIPALS simplifies the analysis at hand while effectively combating the curse of high dimensionality (typically present when the number of variables is large). STATISTICA NIPALS is also particularly suited for use in data diagnostics, making it an ideal tool for use in Quality Control in many areas of science and technology.

Link Analysis

Link analysis addresses the needs of clients in retailing, banking, insurance, etc., industries by implementing the fastest known, highly scalable sequence analysis algorithm with the ability to drive Association and Sequence rules in one single analysis.

MATLAB and Simulink Student versions.

The Student Version package for Release 14 is now available. It includes the newest releases of MATLAB and Simulink, the books *Learning MATLAB* and *Learning Simulink*, and a complete set of product documentation on a CD. Also available are toolboxes with specific capabilities in areas such as signal processing and communication system design, control design and analysis, physical modelling, and computational biology, among others.

If you are currently enrolled at a degree-granting institution you can buy these products directly from HRS. Staff members can request their student book store to stock them for sale to students.

@RISK bundles

Buy the tools and the knowledge of how to best use them, and save.

If you purchase the software and a one or two-day training course at the same time you can save up to \$300!

If you purchase training for more than one person at a time, you can get a group discount.

Contact Darrel for details.

Conferences supported by HRS

- NZWWA 46th Annual Conference and Expo 6-8 October
- NZWWA Modelling conference 10-11 November

Ethnicity

It occurred to one of our staff recently that HRS is ethnically diverse, probably reflecting the nature of New Zealand today. Although most of our staff were born in NZ, their family backgrounds reflect origins in Malawi, France, Switzerland, and Candian Inuit. Also Pakeha and Maori, of course!

Comment from a seminar attendee

"I would like to express my gratitude for Ray's talk. The room was overflowing with people and everyone took away something useful. I hope that these sorts of talks will be repeated in the future."

Seminars

There will be one brand-new seminar in the line-up in October - we think there is enough interest in Web-based surveys to run a seminar on SuveySolutions, from Perseus.

The seminars will be held in Auckland, Wellington, Christchurch, and Dunedin, in that order. In each city we will spend two days, on each day giving three separate seminars of 90 minutes duration.

Titles of the seminars are:

- Engineering and scientific computing with MATLAB
- Documenting design calculations with Mathcad
- Quantitative Risk Analysis for better decision making
- Web surveys with SurveySolutions
- SurveyPro – an integrated data collection environment
- Extracting information from data

To enrol, and to see details, please go to www.hrs.co.nz/seminars.

Training Courses

The following courses are scheduled in the next few months. Get detailed information on the courses and register for them at www.hrs.co.nz/training.

- Quantitative Risk Analysis
Wellington – 28-29 October
- STATISTICA Training
Wellington – 29 October
- MATLAB Training
Auckland – 22-23 November
- Mathcad Training
Auckland – 26 November

Contacting us ...

We are based in Hamilton, but supply, support and demonstrate our products throughout New Zealand.

Product information:
www.hrs.co.nz

Email Addresses

| | | |
|--------|--|-----------------------------|
| Darrel | darrel@hrs.co.nz | Business / Statistics sales |
| Marc | marc@hrs.co.nz | Engineering / Maths sales |
| Leon | leon@hrs.co.nz | Marketing |
| Ros | admin@hrs.co.nz | Accounts and shipping |
| David | support@hrs.co.nz | Support |
| Ray | ray@hrs.co.nz | Policy |
| | webmaster@hrs.co.nz | Web site |
| | info@hrs.co.nz | Everything else |

Phone 07 839 9102

Fax 07 839 9103

P O Box 4153

510 Grey St

Hamilton East

Free phone 0800 477 776