

Statistical Analysis on NW-GRID



Release of Sabre 5.0 Statistical Software and SabreR 5.0

Sabre 5.0 Workshop held on 17th - 19th July 2007

Longitudinal analysis is typically applied to studies of educational attainment or employment history, but it is also applicable to a large class of problem which encompasses such areas as: voting behaviour, accident proneness, demographic surveys, infertility in humans, animal husbandry, absenteeism studies, economic activity and migration.

Longitudinal or panel data on recurrent events are substantively important in social science research for two reasons. Firstly, they provide some scope for extending control for variables that have been omitted from the analysis. For example, differencing provides a simple way of removing time constant effects (omitted and observed) from the analysis. Secondly, a distinctive feature of social science theory is that it postulates that behaviour and outcomes are typically influenced by previous behaviour and outcomes, that is, there is positive 'feedback'.

Sabre (Software for the Analysis of Binary and other Recurrent Events), first appeared in 1989; for an introduction and potted history of Sabre see the Web site <http://sabre.lancs.ac.uk>. Sabre is a random effects statistical modelling package that has recently been extended and re-written to go parallel on the Grid. In this case, the Grid provides the computational environment that enables us to get multiple processors working concurrently on the same computational problem.

A new release of the Sabre statistical software was introduced during a recent workshop entitled *An Introduction to Multivariate Multilevel GLMs using SabreR 5.0 via R on the Grid*. Sabre 5.0 is now running on both the NW-GRID and UK National Grid Service, and outperforms alternative software used for similar statistical analyses; see the comparisons at <http://sabre.lancs.ac.uk/comparison3.html>.

The workshop demonstrated some of the new features of Sabre 5.0 (multilevel, multivariate generalised linear models). The Sabre site contains over 60 worked examples and exercises from published work.

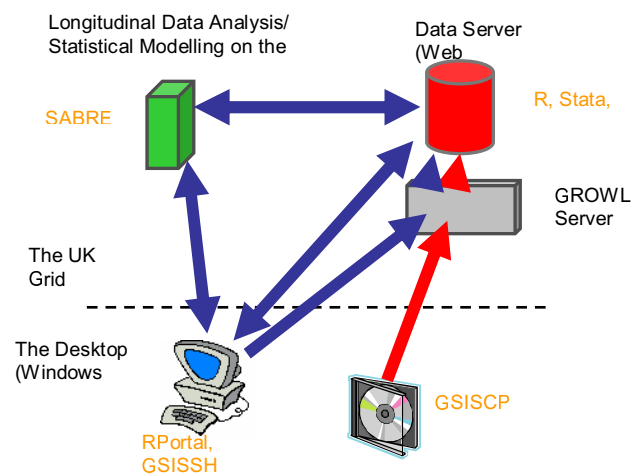


Figure 2: client, server and remote compute host accessed using GROWL toolkit


A new development for Sabre is the SabreR 5.0 plug-in for  - a powerful

open-source language and environment for statistical computing and graphics. The development of the plug-in makes Sabre 5.0 functionality available through the familiar R interface, which brings accessibility and ease of use benefits to large communities of scientists requiring specialised high performance statistical software. The workshop included sessions on running parallel Sabre 5.0 on the Grid using the GROWL middleware toolkit also from the R interface (for those very large and complex models you always wanted to estimate!).

1. how large data sets are to be moved from the source onto the server for manipulation prior to running the computational model;
2. how more flexible client access can be provided from desktop PCs running Windows.

The development of Sabre and associated software is a major part of the work of the NCESS node *Collaboratory for Quantitative e-Social Science* which is funded by the Economic and Social Research Council, ESRC. The associated ReDReSS Web site (ESRC and JISC funded) hosts the presentations and supporting material used during the workshop. Sabre, SabreR and GROWL will also be available as part of the ESRC e-Infrastructure.

For further information, material and downloads of Sabre and SabreR, see <http://sabre.lancs.ac.uk>

For further information about , see <http://www.r-project.org>

For GROWL, see <http://www.growl.org.uk>

To view the workshop presentations, see <http://redress.lancs.ac.uk>

```

Library(GROWL)

mysession=grid.session("User20","caman8",
  "lancs1.nw-grid.ac.uk","lancs1.nw-grid.ac.uk/
  jobmanager-sge","","")

#read the data
Hsb<-
read.table(file="/home/Shared/data/hsb.tab")
attach(hsb)

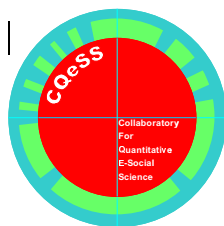
# create the models
Sabre.model.1<-sabre.grid(mathach~1,
  case=school, first.masss=64,
  first.family="Gaussian", script.file="hsb1.sab",
  log.file="hsb1.log", data.file="hsb1.dat"
  sabre.binary="/panfsl/la/staging/users/dlabl/
SABRE5/sabre_P.out",session=mysession,
  processors=1)

Print.sabre.arid(sabre.model.1)

```

Figure 2: Parallel (Grid) SabreR script

Two issues require longer-term investigation:



For more information about this work please contact Prof. R. Crouchley at University of Lancaster.

CQeSS is a collaborative project to develop and support Quantitative e-Social Science

