The Stata Journal publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those teaching, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the Stata Journal, including information for authors, see the web page

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates:

Subscriptions mailed to US and Canadian addresses:

- 3-year subscription (includes printed and electronic copy) $153
- 2-year subscription (includes printed and electronic copy) $110
- 1-year subscription (includes printed and electronic copy) $ 59
- 1-year student subscription (includes printed and electronic copy) $ 35

Subscriptions mailed to other countries:

- 3-year subscription (includes printed and electronic copy) $225
- 2-year subscription (includes printed and electronic copy) $158
- 1-year subscription (includes printed and electronic copy) $ 83
- 1-year student subscription (includes printed and electronic copy) $ 59
- 3-year subscription (electronic only) $153

Back issues of the Stata Journal may be ordered online at

http://www.stata.com/bookstore/sj.html

The Stata Journal is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the Stata Journal, StataCorp, 4905 Lakeway Drive, College Station TX 77845, USA, or email sj@stata.com.
The Stata Journal (2005) 
5, Number 1, pp. 43–45

Stata at 20: a personal view

Patrick Royston

I bought my first copy of Stata in May 1989, and it was version 2.05. I still use Stata most days, but now it is version 8.2. Why has it survived and prospered so long in my statistical armory?

I still have my 1989 shipment of Biturbo Stata version 2.05 and Stage 1.0 Graphics Editor, on five 3.5-inch, 720 Kb floppies. On opening at random my copy of the version 2.05 reference manual (copyright March 1989 © Computing Resource Center, Los Angeles, California), the page fell open at page 539, a summary of something called Stat.Kit. I recalled then the excitement of the arrival of “kits”; they were the forerunners of the now familiar and quintessential ado-files. In those days, there were 4 such kits (Stat.Kit, Graph.Kit, Data.Kit, Survive.Kit), and they provided a variety of programs in each relevant area. You typed, for example, run Stat.Kit, and the associated 17 programs were loaded into memory. After that you used them exactly as one now does an ado-file.

On reflection, Stata has a few key features that I value above all. First, I will illustrate one of these features in a little detail, and later I will briefly mention the others. This feature is backwards compatibility, which in Stata lingo is known as “version control”. As a simple experiment designed to test version control, I wrote and ran in Stata 8.2 the following do-file:

/*
   Experiment to show version control.
   Patrick Royston.
*/
* Set up version version 2.05
set logtype text
log using stata.log, replace
* clear out all ado-file support quietly forvalues j = 1/7 {
adopath -1 }
* Load Stat.Kit run stat.kit
* Load old auto data use auto
* Create mpg + uniform() * 5 gen mpg2 = mpg + uniform() * 5 
* Perform equal-variance t-test ttest mpg = mpg2
log close

What I am doing here is to set Stata 8.2 to version 2.05, remove all ado-file support, load Stat.Kit and the classic auto.dta dataset, generate a new variable mpg2 from mpg
with on average about 2.5 mpg greater values, and run an unpaired t-test comparing mpg with mpg2. Here is the output as stored in \texttt{stata.log}:

\begin{verbatim}
.log: e:\tsj\stata20\stata.log
log type: text
opened on: 28 Dec 2004, 10:03:35
.* clear out all ado-file support
.quietly forvalues j=1/7 {
.* Load Stat.Kit
.run stat.kit
Loading Stat.Kit Release 2.05 Copyright (c) 1986-1989 by ==C=R=C==
All rights reserved.
The following new commands are now available:
blogit genstd kwallis signrank ttest
bprobit glogit means signtest
dbeta gprobit ranksum spearman
genrank ksmirnov regdw teststd
.* Load old auto data
.use auto
(1978 Automobile Data)
.* Create mpg + uniform() * 5
.gen mpg2 = mpg + uniform() * 5
.* Perform equal-variance t-test
.ttest mpg = mpg2
Variable | Obs Mean Std. Dev. Min Max
-------------+--------------------------------------------------------
  mpg | 74 21.2973 5.785503 12 41
  mpg2 | 74 23.68922 5.945047 14.43152 44.18587
Test: means of mpg and mpg2 are equal (assuming equal variances)
  Difference = -2.3919258
  t-statistic = -2.48 with 146 d.f.
  Prob > |t| = 0.0143
.* log close
.log: e:\tsj\stata20\stata.log
log type: text
closed on: 28 Dec 2004, 10:03:35
\end{verbatim}

Lo and behold, Stata 8.2 runs correctly under version 2.05 control using 2.05 program code and data. This to my mind is impressive. How many other statistical packages would stand up to such a severe test? For me, version control is an apparently unglamorous property of Stata that shows StataCorp's strong commitment to the practical needs of its users, and is one reason why Stata still has a rosy future. I dread to think how many lines of C program code sit behind the scenes supporting the innocent-looking \texttt{version} \# command. Surely this is virtuoso computer programming of a high order?

Other essential features of Stata I particularly value are its graphics, its simple but elegant command syntax (which in essence has not changed since version 2.05), its seamless programmability via ado-files, and last, but by no means least, its excellent documentation. I do recall that having fast, high-resolution graphics was in 1989 an
important selling point and a key determinant of why I bought Stata. The fact that formulas are given in the manual even for quite complex mathematical calculations such as the partial likelihood in a Cox regression model has been immensely useful, and StataCorp even now continues to pay attention to improving this aspect of the product. I could add other far-sighted innovations that were even rather shocking at the time of their introduction, such as increasing levels of integration with the Internet for updating Stata and for installing add-on packages. Certainly the list-server Statalist has played a useful role in the development of user awareness, although I myself don’t find much time to access it nowadays. One must also mention the *Stata Journal* and its predecessor, the *Stata Technical Bulletin*. No serious package can be without such a publication.

Is Stata now perfect? Of course not. Areas in which I would like to see improvements include documentation of graphics and GUI dialog programming, greater searchability and flexibility of the help system (I have lost count of the number of times I have trawled through the help system in search of some obscure graphics option), acceleration of drawing a graph, and a facility to make it easier to write help files in Stata’s markup language, SMCL. A graphics editor would be nice too; my 1989 shipment included Stage 1.0, the first and only edition of the Stata graphics editor, a fine program that has not yet been replaced.

Above all, though, StataCorp’s encouragement of users to get involved and Stata-Corp’s habit of valuing and responding to their subsequent efforts have been some of the most delightful and productive aspects of Stata. If that interaction is ever lost, Stata will probably die.

**About the Author**

Patrick Royston is a medical statistician of 25 years of experience, with a strong interest in biostatistical methodology and in statistical computing and algorithms. At present, he works in clinical trials and related research issues in cancer. Currently he is focusing on problems of model building and validation with survival data, including prognostic factors studies, on parametric modeling of survival data, and on novel trial designs.