



Stat Studio: A New Programming Environment for Statistical Computing

**THE
POWER
TO KNOW®**

Robert N. Rodriguez
SAS Institute

Alaska SAS Users Group
May 21, 2008



What Is Stat Studio?

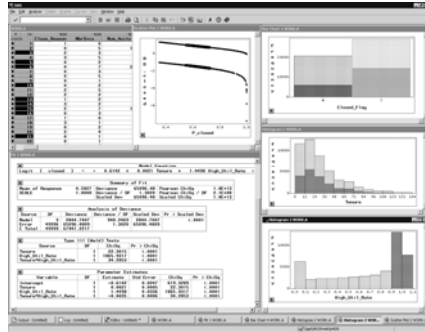
New statistical software in SAS 9.2 that blends

- Interactivity of SAS/Insight
 - Dynamically linked statistical graphics
 - Standard statistical modeling
- Extensibility of SAS/IML
 - Matrix programming language
- Power of SAS/STAT procedures

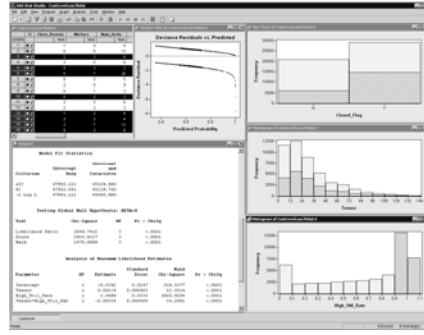
Copyright © 2006, SAS Institute Inc. All rights reserved.



Stat Studio as an Insight Successor



SAS/Insight



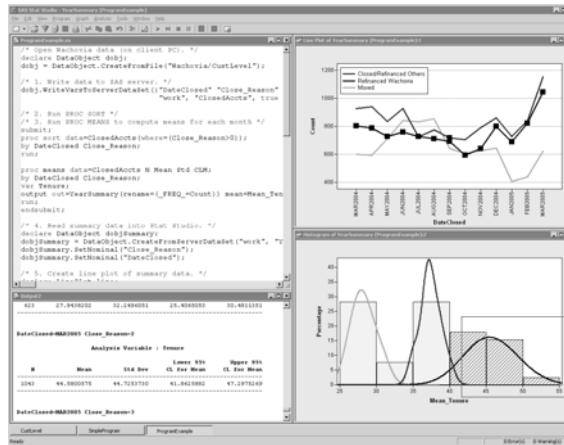
Stat Studio



Stat Studio as a Programming Environment

Program Editor
for IMLPlus
Language

STAT and IML
Output



Customized
dynamically
linked graphs

Audience: *High-End Data Analysts*

- Innovative problem-solvers
- Statisticians and researchers with
 - complex data used for critical business or scientific decisions
 - problems too difficult for standard models
- Need rich programming environment
 - "I want to write a program that visualizes the results with linked graphics ..."
 - "I wish I could call STAT procedures as functions in IML."
 - "We'd like to customize and extend Insight ..."

Copyright © 2006, SAS Institute Inc. All rights reserved.

Stat Studio Architecture

- Client runs on desktop Windows PC
- Client connects with one or more SAS servers
- Data can reside on client or server
- Designed around IMLPlus programming language

Copyright © 2006, SAS Institute Inc. All rights reserved.

IMLPlus Language

- IMLPlus is IML with *additional* statements for
 - calling SAS procedures and the DATA step
 - creating dynamic graphics
 - calling user-written programs in C/C++, Fortran, Java
- Runs on both client and server
 - Execution switches as needed
- You use IMLPlus to write programs in Stat Studio
 - Leverages your familiarity with either SAS or IML
- Stat Studio uses IMLPlus for the programs “behind” its analysis menus

What Does Stat Studio Provide?

- Point-and-click capabilities for graphical exploration and modeling ... no programming!
- Interface for writing your own IMLPlus programs
- Capability to
 - identify structure, features in multivariate data
 - experiment with standard modeling techniques
 - implement new methods
 - transform and pass output across methods
 - display results with interactive graphics
 - validate models and link outliers to original data

Demo Data from Wachovia

- Prime Equity Line (PEL) of credit
 - Secured loan using home equity as collateral
- 50,000 records (customers) and 30 variables
- Each account
 - open in March 2005 **or**
 - closed in the previous 13 months
- Why do customers close PEL accounts?
- How can we predict which customers will close their accounts?

Key Variables

Name	Description
Close_Reason	0 = account is open 1 = moved or sold home 2 = refinanced with Wachovia 3 = refinanced with competitor or no reason
Closed_Flag	Y = closed, N = open
NbrSvcs	number of services with Wachovia
Num_Accts	number of accounts with Wachovia
Tenure	number of months account has been open
Age_Open	customer's age when account opened
High_Util_Rate	ratio of highest balance to maximum the customer can borrow

Demo Outline

1. Exploratory analysis using GUI
2. Standard logistic model using GUI
3. Program to create customized graphs
4. Program for customized analysis
 - SAS procedures
 - IML
 - Linked graphics

Programming and Extensibility

You can

- draw arbitrary curves/shapes on graphs
- add legends and insets
- control graph attributes
- add menu items to call IML modules
- write and add user analyses

How Much IML Do I Need to Know?

You can do a lot with a little!

- Basic IML statements
 - DO, IF, THEN, ...
- IMLPlus statements
 - move data to/from data objects
 - create dynamic graphics
 - submit SAS statements

Knowing IML for matrix computations lets you do more ...

Benefits of Stat Studio

1. As a programming environment
 - access to data on client or remote servers
 - computational support on SAS server
 - ability to call STAT procedures or programs as functions
 - extensibility through IML, Fortran, C, C++, ...
 - programmable dynamic graphics through IMLPlus
2. As a point-and-click interface
 - Insight menus for graphical exploration and modeling
 - newer methods, such as robust regression
 - familiar SAS/STAT output