SAS In-Memory Statistics and SAS Studio

Steve Ludlow
What is SAS Studio?

What is IMSTAT?

Demonstration
New SAS Programming Environment

Available
Consistent
Assistive
Available through all modern web browsers

Available

Consistent

Assistive
One interface for SAS Programming

Available
Consistent
Assistive
Increased Developer productivity

Available
Consistent
Assistive
Programming Interface
Programming interface

- Color-coded editor with auto-complete & pop-up syntax help
- Table viewer to sort or filter a table … and see/copy the generated SQL
- Log navigator
- Customise to create a “classic” DMS look
Snippets & Tasks
Snippets

```sas
/* Stream a CSV representation of SASHELP.CARS directly to the user's browser. */

proc export data=SASHELP.CARS
dataout=_dataout
   dbms=csv replace;
run;

%let _DATAOUT_MIME_TYPE=text/csv;
%let _DATAOUT_NAME=cars.csv;
```
Tasks

Distribution Analysis

Exploring Data
- Histogram
- Add normal curve
- Add kernel density estimate
- Add inset statistics
- Inset Statistics

Checking for Normality
- Goodness-of-fit tests
- Histogram with normal curve
- Add inset statistics
- Normal probability plot
- Add inset statistics
- Normal quantile-quantile plot
- Add inset statistics
- Inset Statistics

Fitting Distributions
Tasks Authoring

• SAS Studio allows for a flexible task framework to create tasks.
Delivery - How you get SAS Studio

• Single-User. With Base SAS running on Microsoft Windows or with the SAS University Edition, you access SAS Studio through your local web browser on the same machine as SAS is installed.

• Multi-User. If you have SAS installed on a server, SAS Studio provides basic multi-user access to submit SAS code to that server. You can access SAS Studio from your browser (on any machine) by pointing to the server where SAS is running. You must have credentials to log into the SAS server machine. The content (files, data) you are allowed to access depends on the authority granted to you on the SAS server being accessed.

• Enterprise. If you have licensed Integration Technologies, you have full enterprise access to SAS Studio, which includes security & access control provided via SAS Metadata Server.
Delivery - Supported browsers

• SAS Studio is a pure HTML5 application that requires no browser plug-ins and supports all the major web browsers.

Microsoft Windows
• Microsoft Internet Explorer 9 or later
• Mozilla Firefox 21 or later
• Google Chrome 27 or later

Linux
• Mozilla Firefox 21 or later
• Google Chrome 27 or later

Apple Macintosh OS X
• Apple Safari 6.0 or later
• Mozilla Firefox 21 or later
• Google Chrome 27 or later

Mobile (iOS) = Beta
config.properties

browserpath=C:\Program Files (x86)\Google\Chrome\Application\chrome.exe
Delivery - How you get SAS Studio

Visit the SAS Studio product page (http://support.sas.com/software/products/sasstudio/index.html) for:

• Documentation
• Videos
• Papers
• Already on the Support Site
• SAS Studio 3.3 Users Guide
• SAS Studio 3.3 Administrators Guide
• SAS Studio 3.3 Developers Guide
SAS® In-Memory Statistics
SAS In-Memory Statistics

Provides a single *interactive* programming environment to perform

- analytical data preparation
- variable transformations
- exploratory analysis
- statistical modeling and machine learning
- integrated modeling comparison and scoring

- Takes advantage of distributed in-memory computing optimized for analytical workloads
SAS In-Memory Statistics

- All required components for successful data science (analytics lifecycle) at your fingertips in one integrated programming environment
- Supports the experimental, iterative nature of analytical model development
- Faster time to insight - nimble and agile
- Analyze data sets of any size and any structure
- Explore and compare different modeling approaches to build the optimal champion model faster
- Better collaboration by empowering multiple users to work on the same data concurrently
- Maintain even distribution of workload among the nodes in a distributed computing environment
SAS In-Memory Statistics

• Utilize compute resources more efficiently
  • Load data once for entire exploration and modeling process
  • Multiple User Concurrency
• Improve data governance by eliminating data duplication
• Utilize distributed computing environment for high-performance data science
• An in-memory architecture eliminates costly data movement
• Easier maintenance with web-based clients
Capabilities of SAS® In-memory Statistics

Data Manipulation
- Aggregate
- Compute
- Update
- Append
- Set
- Schema
- DeleteRows
- DropTables
- PurgeTempTables

Data Exploration
- Boxplot
- Corr
- Crosstab
- Distinct
- Fetch
- Frequency
- Histogram
- KDE
- MDSummary
- Percentile
- Summary
- TopK

Predictive Modeling
- Decision Tree
- Forecast
- Gen Linear Model
- Linear Regression
- Logistic Regression
- Random Forests
- Neural Networks

Descriptive Modeling
- Association
- Path Analysis
- Clustering (k-means)
- Clustering (DBSCAN)

Evaluation, Deployment
- Assess
  - Misclassification matrix
  - Lift, ROC, Concordance
- Score
- Training / Validation

Text Analytics
- Parsing
- SVD
- Topic generation
- Document projection

Recommendation Systems
- Association
- Clustering
- kNN
- SVD
- Ensemble

Utilities
- Where
- GroupBy
- TableInfo, ColumnInfo, ServerInfo
- Partition, Balance
- Store, Replay, Free
- Table, Promote

HDFS I/O
- Sasiola
- Sashdat
- Anyfile Reader
proclinstat:
   */-MERGE-*/
   table lastr.carinfo;
   schema carinfo (modelkey=modelkey / prefix=/info / mode=table);
   run;

41table lastr.carinfo;
42fetch / fcom=1 to=5 format;
43run;
44/**** EXPLORE ****/
45table lastr.cardatap;
46distinct _all_ / save=tab;
47run;
48/* the name list in LIST1 is automatically created;
49no need to type in variable names one by one;
50store tabl(where='Distinct is 5', 1)=list1;
51run;
52frequency list1;
53crossstab wheelTypeID*wheeltype;
54sort mu;
55ods output frequency=work.freq1;
56run;
57table lastr.cardatap (samenames=avgOdo) ;
58summary avgOdo / tempnames=avgOdo tempexpr=avgOdo = vehodo / var=Odo ;
59groupby=badbuy partition;
60run;
61quit;
62ods graphics / height=4in width=12in;
63proc sgpanel data=freq1;
64panelby column / novname columns=0 rows=1 uniscale=row;
65vbar formattedvalue=databar;
66run;
67proc imstat:
68/**** CLEAN-UP ****/
69table lastr.cardatap;
70schema carinfo (modelkey=modelkey / prefix=/info / mode=table);
71run;
72/**** CLEAN-UP ****/
73table lastr_i_templast ;
74truncate caridata2 ;
75where info_nationality ne 'NULL';
76run;
77table lastr.caridata2;
78where:
79compute delta1 = vehcost - mmreacquisitionretaileraverageprice;
80compute delta2 = vehcost - mmreacquisitionretaileraverageprice2;
81compute delta3 = vehcost - vehtotalaverageprice ;
Demonstration
SAS In-Memory Statistics

- **Faster** time to insight
- **Interactive** analytics
- **Seamless integration** of
  - Data Access and Preparation
  - Data Exploration
  - Model Development
  - Model Scoring
- Proven and **state-of-the-art** analytical and machine-learning algorithms
- **Scalable**
- **Concurrent** access to data loaded in-memory in a multi-user environment
Thank you