What's New in SAS 9.4

A High level Overview

Foundation SAS

What's new in 9.4

Agenda

- Base SAS
- SAS/GRAPH
- SAS/STAT

What’s New In BASE SAS for SAS 9.4

Two new programming languages

Base SAS

New Programming language

DS2 is a new SAS proprietary programming language.

- Enables DS2 language statements from Base SAS
- Includes additional data types, ANSI SQL types, programming structure elements, and user-defined methods and packages.
- Allows embedded FedSQL in some statements
- Runs anywhere – Base, In-Database (via In-Database Code Accelerator), High Performance Analytics (via HPDS2)

DS2

Why use it?

DS2 is beneficial in applications that

- require the precision that results from using the new supported data types
- benefit from using new expressions or write methods or packages
- need to execute SAS FedSQL from within DS2
- execute outside a SAS session, for example on High-Performance Analytics Server or the SAS Federation Server
- take advantage of threaded processing in products such as the SAS In-Database Code Accelerator, SAS High-Performance Analytics Server, and SAS Enterprise Miner

```
PROC DS2;
data _null_;
   method init();
      dcl varchar(20) foo;
      foo = '**> Starting';
      put foo;
   end;
   method run();
      set ds2_sas.banks;
      put _all_;
   end;
   method term();
      dcl char(11) bar;
      bar = '**> I quit!';
      put bar;
   end;
run; quit;
```

Initial processing

Execution loop

Final processing

FEDSQL

New Procedure

The FEDSQL Procedure

- The FEDSQL procedure enables you to submit FedSQL language statements from a Base SAS session.
- The FedSQL language is the SAS implementation of ANSI SQL:1999 core standard.

FEDSQL and DS2

Example

You can use an embedded FedSQL query to generate data within a DS2 method block.

```
method run();
  set {select * from work.titles natural join work.price
    order by publisher};
  by publisher;
    ... DS2 statements ...
end;
```

Enhancements to ODS

Output

Enhancements to the Output Delivery System enable you to

- Create EPUB, HTML5, Microsoft PowerPoint and Excel files.
- Use the ODS Report Writing Interface (RWI) to create and manipulate predefined ODS objects in a DATA step to create highly customized output.
- Arrange ODS output objects exactly where you want them on a page, or use dynamic placement of objects by using a grid structure. (ODS Layout)
- Animate multi-page GIF images and SVG files by setting system options.

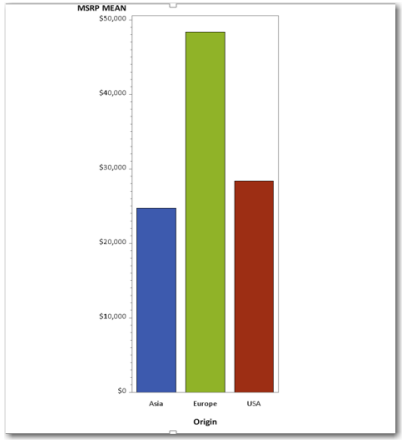
SAS ODS
PowerPoint

Sample Code

The ODS PowerPoint destination allows you to send SAS output directly to PowerPoint.

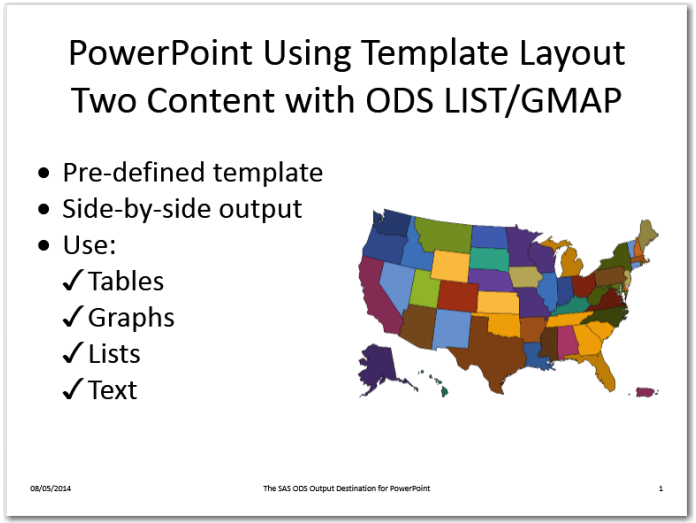
Like other ODS destinations, simply specify the POWERPOINT keyword on the ODS statement and use FILE= to name the resulting file.

```
ods powerpointfile='c:\test.pptx';

 proc gchart data=sashelp.cars;
       vbar origin / sumvar=msrp
 type=meanpatternid=midpoint;
 run; quit;

ods powerpoint close;
```



ODS Powerpoint

Two content layout

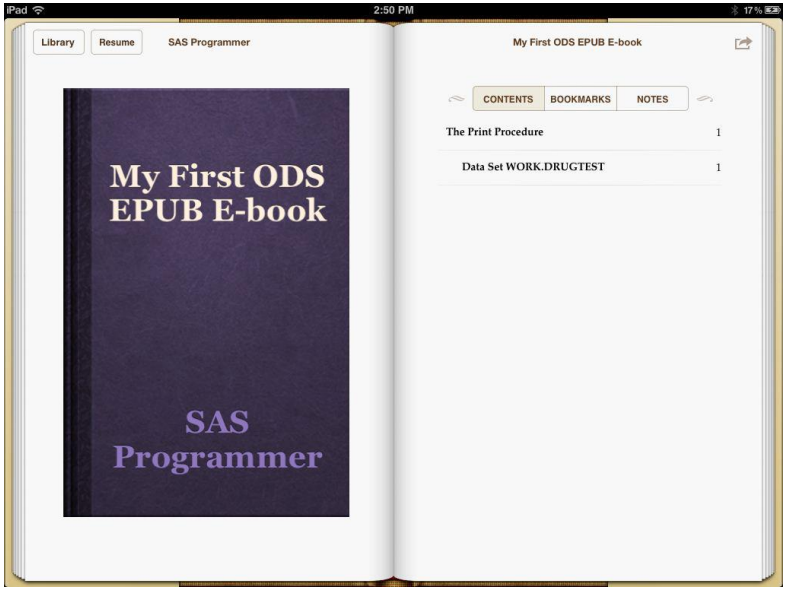



```
ods powerpoint file="Layout2List.ppt" layout=twocontent nogtitle nogfootnote
style=powerpointlight;

proc odslist;
  item 'Pre-defined template';
  item 'Side-by-side output';
  item;
     p 'Use:';
     list / style=[bullet=check];
      item 'Tables';
      item 'Graphs';
      item 'Lists';
      item 'Text';
     end;
run;
goptions hsize=4.5in vsize=4.5in;
proc gmap map=maps.us data=maps.us all;
  id state;
  choro statecode/statistic=frequency discrete nolegend;
run; quit;

ods _all_ close;
```

SAS ODS EPUB

Output SAS Reports to eBook format to be viewed on iPad & iPhone

```
ods epub file="glm.epub" title="My First ODS EPUB E-book"    options(creator="SAS Programmer" description="My First ODS EPUB Book" subject="PROC GLM" type="ODS EPUB book");

ods graphics on;
proc glm data=DrugTest;
   class Drug;
   model PostTreatment = Drug|PreTreatment;
run;
quit;

ods epub close;
```

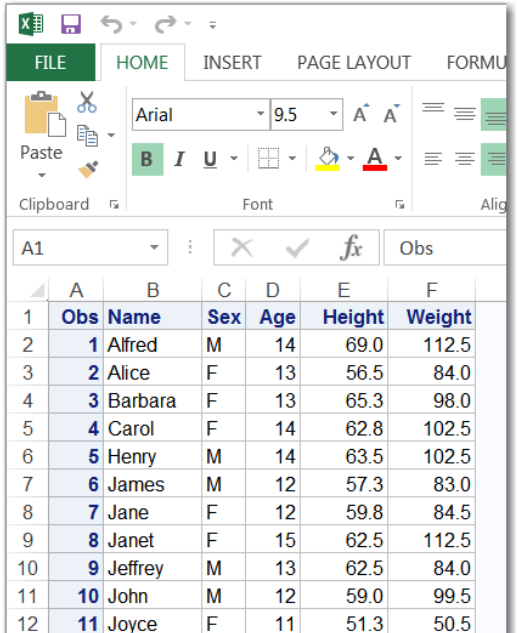


ODS Excel

Example

The ODS Excel destination opens, manages, or closes the ODS destination for Excel, which produces Excel spreadsheet files compatible with Microsoft Office 2010 and later versions.

```
ods excel file="c:\temp\test.xlsx"
          options(sheet_name='OFAC'
          zoom='100' row_heights='12.75');

   proc print data=sashelp.class;
   run;

ods excel close;
```



	A	B	C	D	E	F
1	Obs	Name	Sex	Age	Height	Weight
2	1	Alfred	M	14	69.0	112.5
3	2	Alice	F	13	56.5	84.0
4	3	Barbara	F	13	65.3	98.0
5	4	Carol	F	14	62.8	102.5
6	5	Henry	M	14	63.5	102.5
7	6	James	M	12	57.3	83.0
8	7	Jane	F	12	59.8	84.5
9	8	Janet	F	15	62.5	112.5
10	9	Jeffrey	M	13	62.5	84.0
11	10	John	M	12	59.0	99.5
12	11	Joyce	F	11	51.3	50.5

Output

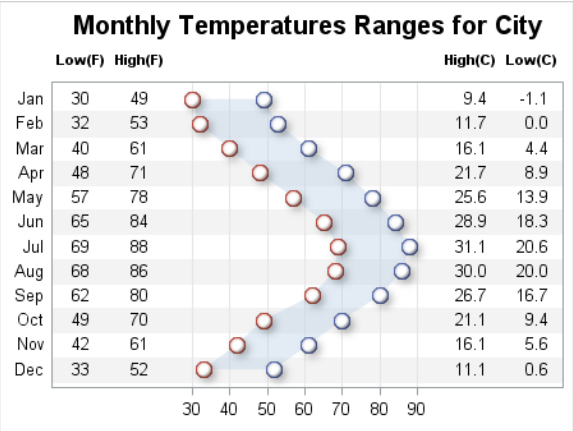
Enhancements to ODS Graphics provide

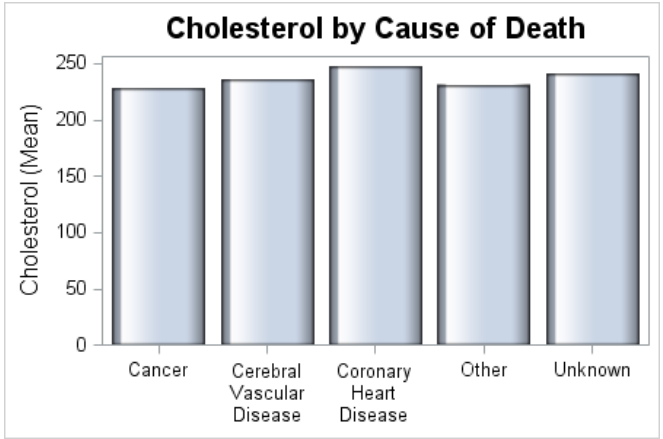
- Several new plot types, including axis tables that create an axis-aligned row or column of textual data.
- The addition of numerous plot layout, panel, and axis options to control and enhance the output of your graphs.
- A new sub-pixel rendering feature provides smoother curves for line charts and more consistent spacing in bar charts.

Output

Additional Information:

- The ODS Graphics products provide more options for fitting or splitting data labels, curve labels, and axis tick values when there is not enough room to display the text normally.
- The ODS Graphics Designer introduces an Auto Charts feature that generates a variety of graphs automatically, based on your data.





Other Important Enhancements to Base SAS

Security

SAS/SECURE is now delivered free of charge with Base SAS.

- Uses the industry standard Advanced Encryption Standard (AES) with 64-bit salt.

Metadata-bound libraries universally enforce metadata-layer permission requirements for physical tables—regardless of how a user requests access from SAS – closing the "LIBNAME hole"

- The **AUTHLIB procedure** is a new utility procedure that enables you to manage metadata-bound libraries.

Extended attributes

Customized attributes for Data Sets & Variables

Extended attributes are customized metadata for your SAS files. They

- Can be defined on a data set or on an individual variable
- Are organized into (name, value) pairs
- Can be numeric or character – no pre-defined limit on the number of bytes allowed for a character value
- Are managed by PROC DATASETS
- Base engine support with more to follow (MLE, ACCESS)

Other New Features

The Work library data sets and catalogs, and the values of global statements, macro variables, and system options can be **preserved between SAS sessions**.

The PRESENV Procedure

- The PRESENV procedure preserves all global statements and macro variables in your SAS code from one SAS session to another.

Other New Features

json

SAS data sets can be written to an external file in JSON representation.

The JSON Procedure

- The JSON procedure reads data from a SAS data set and writes it to an external file in JSON representation.

Other new features

The SAS language now supports time zones based on Universal Coordinate Time (UTC).

- Data sets and catalog time stamps can specify the time based on a specific time zone.
- SAS can also determine the time for an area, taking into account Daylight Savings Time.

Examples of new functions to support UTC include:

- TZONEID-returns the current time zone ID.
- TZONENAME-returns the current standard or daylight savings time and the time zone name.
- TZONES2U-converts a SAS datetime value to a UTC datetime value.

SAS Support for Hadoop

Support for Hadoop

Foundation SAS offers support for Hadoop through

- Base SAS
- SAS/Access Interface to Hadoop (Hive)
- SAS/Access Interface to HAWQ
- SAS/Access Interface to Impala

Hadoop Support

Base SAS

FILENAME statement – DATA step can read and write HDFS files.

PROC HADOOP - Copy or move files, execute MapReduce and Pig code, execute file system commands

Hadoop Support

SAS/ACCESS Interface to Hadoop

SAS/Access Interface to Hadoop supports

- SQL Pass-through
  - HiveQL queries passed to Hive for processing
- LIBNAME statement for Hadoop
  - Hive tables appear as SAS data sets
  - Access engine can translate to HiveQL to optimize data processing
  - Capable of significant in-database processing

What’s New in SAS/Graph

SAS/Graph

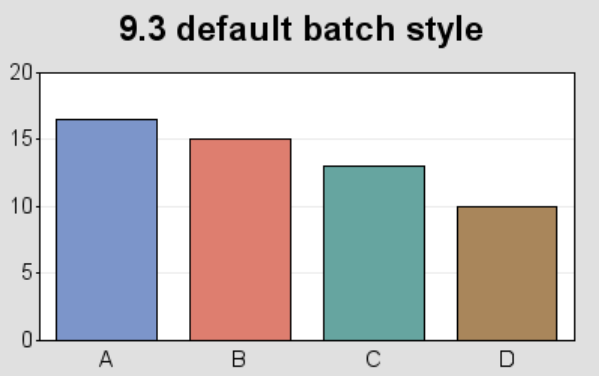
Selected Enhancements

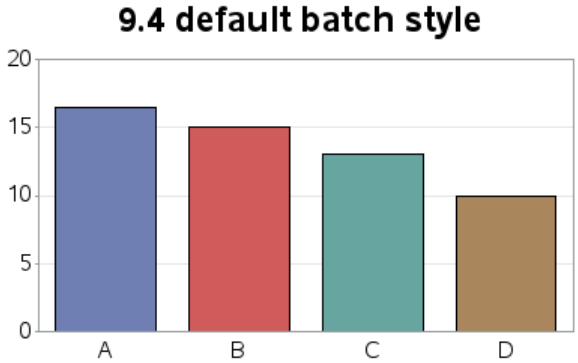
- New Default Style for batch mode
- New Graph styles (DOVE and RAVEN)
- The GIF device now supports RGBA color mode (transparency) and anti-aliasing.
- The SVG and GIF devices now support animation.
- Maps data sets have been updated
- The GEOCODE procedure now supports non-U.S. street geocoding

Sas/Graph

Default Style

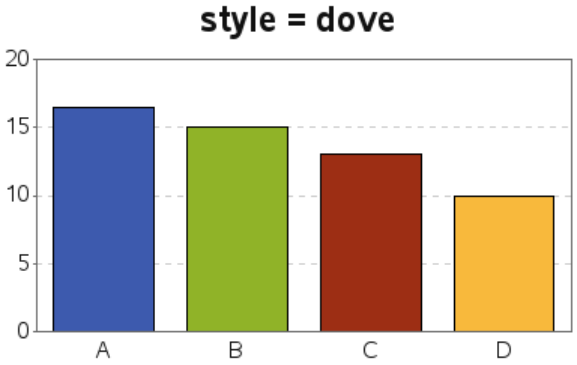
New Default Style for batch mode

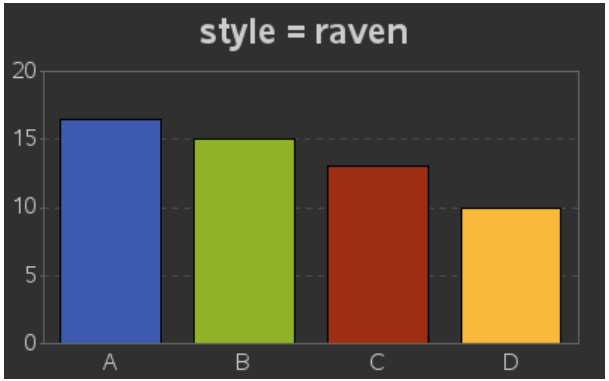




Sas/Graph

New Graph Styles





Sas/Graph

Anti-aliasing

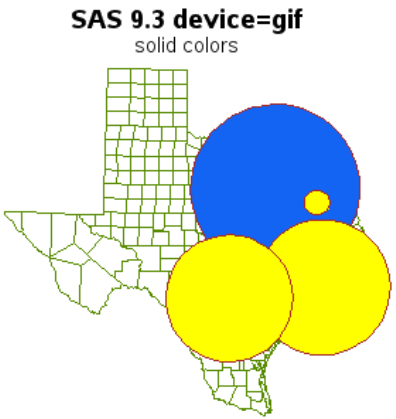
The GIF device now supports RGBA color mode (transparency) and anti-aliasing.

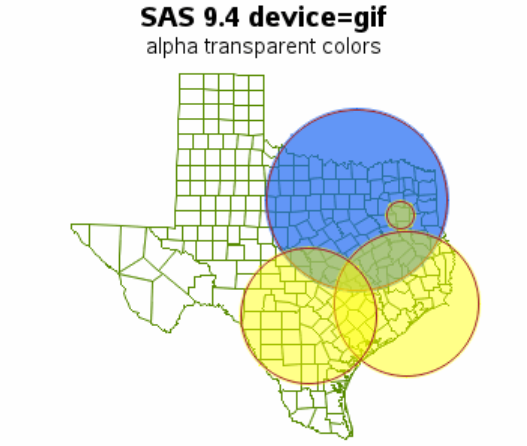
SAS 9.3 device=gif
jagged lines

SAS 9.4 device=gif
anti-aliased lines

Sas/Graph

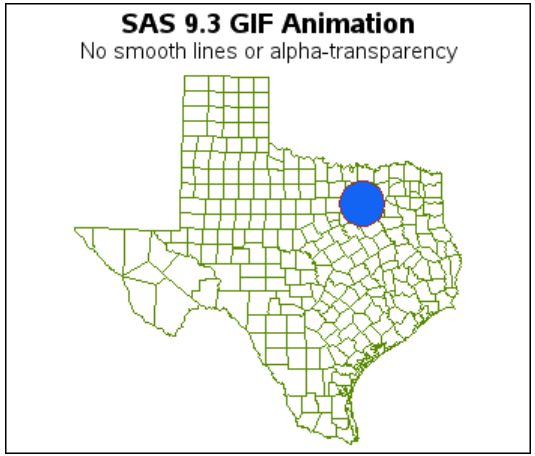
Gif Alpha-transparency

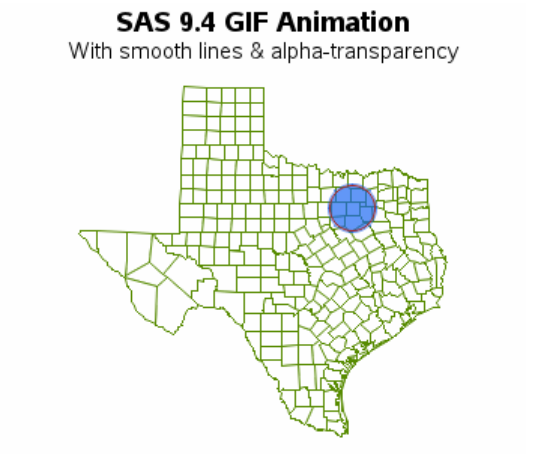




Sas/Graph

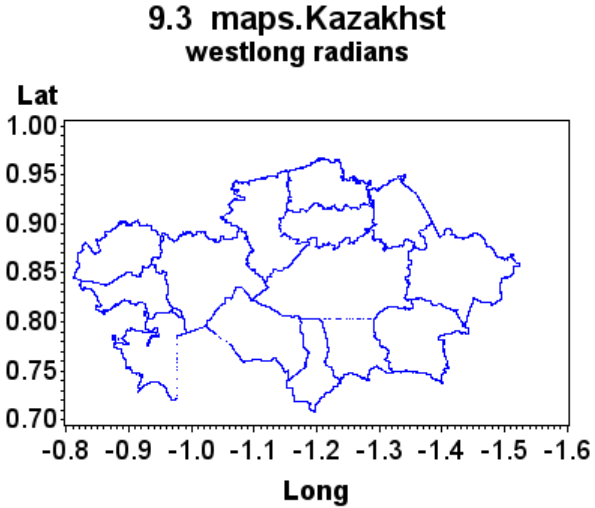
Better Gif Animations

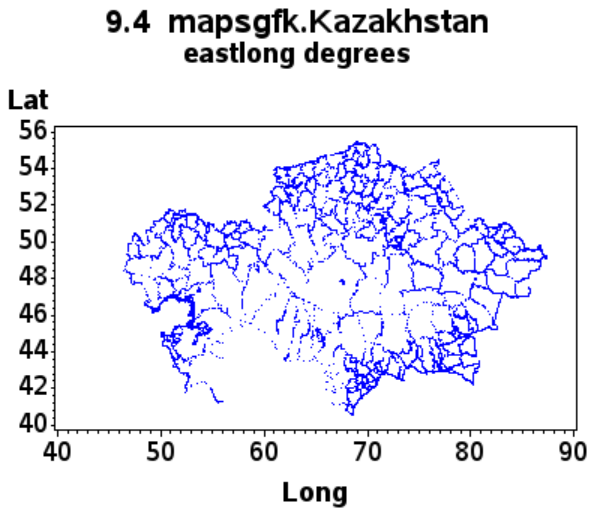




Sas/Graph

New Maps





Sas/Graph

- The GEOCODE procedure now supports non-U.S. street geocoding

9.3 Street-Level Geocoding
in United States

100 SAS Campus Drive, Cary, NC

9.4 Street-Level Geocoding
in Canada

55 Almond Place, Whitehorse, YT

SAS Studio

SAS Studio

SAS® Studio is a SAS developer environment that runs in a Web browser, enabling developers to program and interact with SAS wherever and whenever they want.

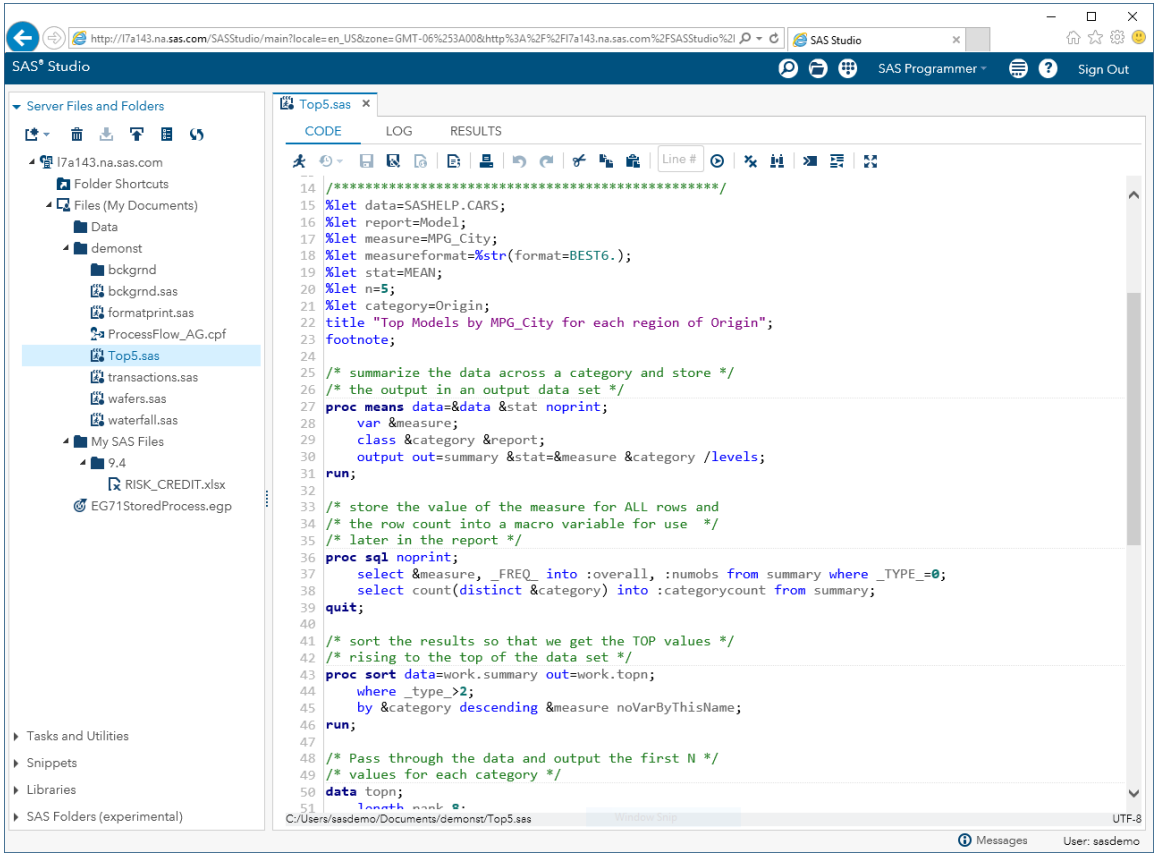
What's so great about SAS® Studio?

- **Availability.** SAS Studio allows SAS programmers to submit SAS code from a wide range of devices, from wherever they happen to be.
- **Consistency.** Become familiar with the SAS Studio user interface once and use it throughout your career.
- **Assistance.** In SAS Studio, SAS code is front and center. To speed development and promote consistent and efficient coding practices, functions similar to SAS® Enterprise Guide® (such as code-generating tasks and auto-complete) are available.

SAS® Studio

Programming Interface

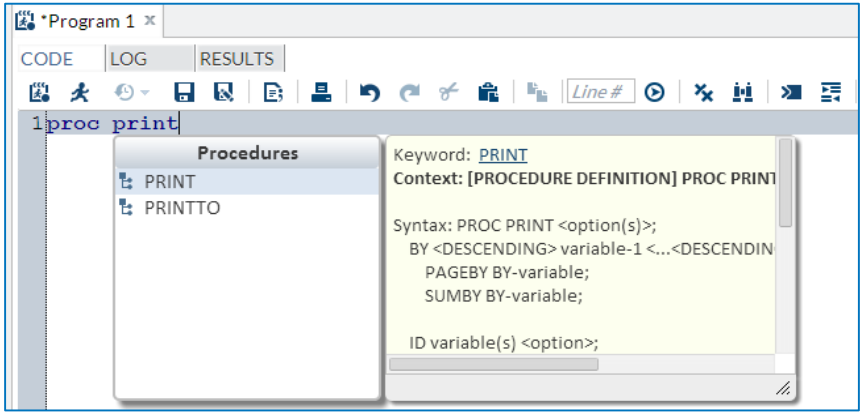
- Color-coded editor
- Submission history
- Code Formatting
- Search
- Tabs for easy navigation

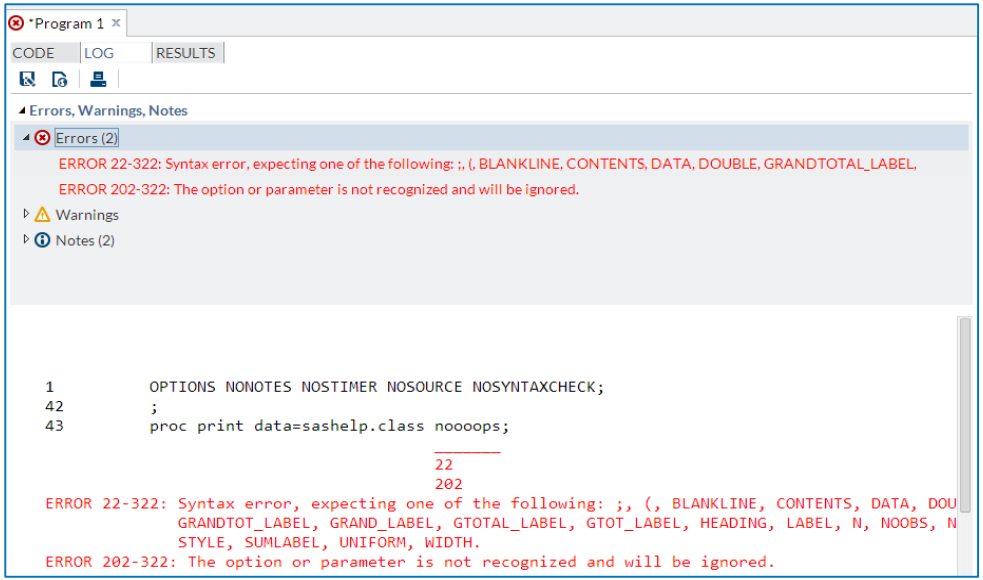


SAS® Studio

Programming Interface

- Auto-complete
- Pop-up syntax help



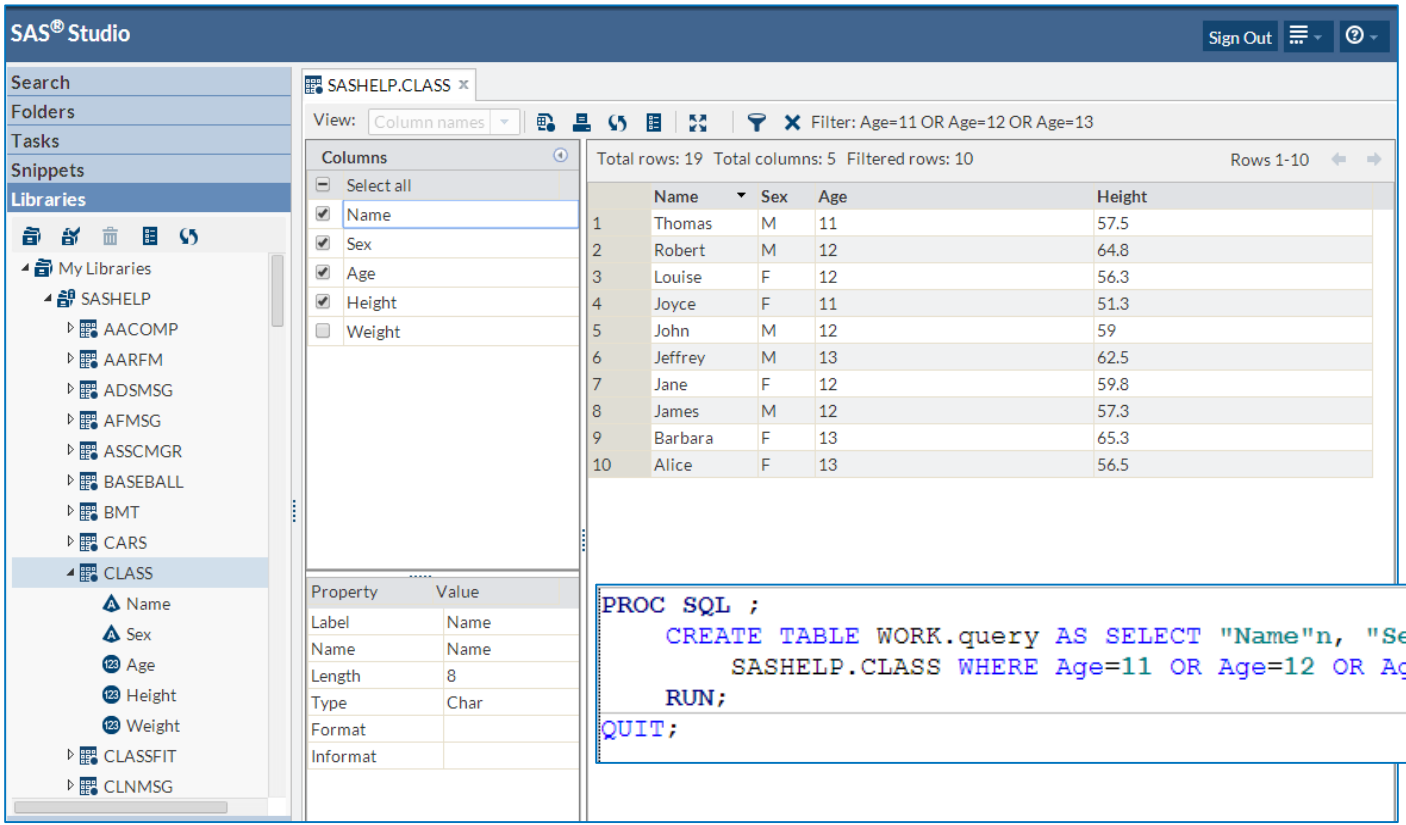


- Color-coded log navigator
- Categorized alerts:
  - Errors
  - Warnings
  - Notes
- Easily save or share logs

SAS® Studio

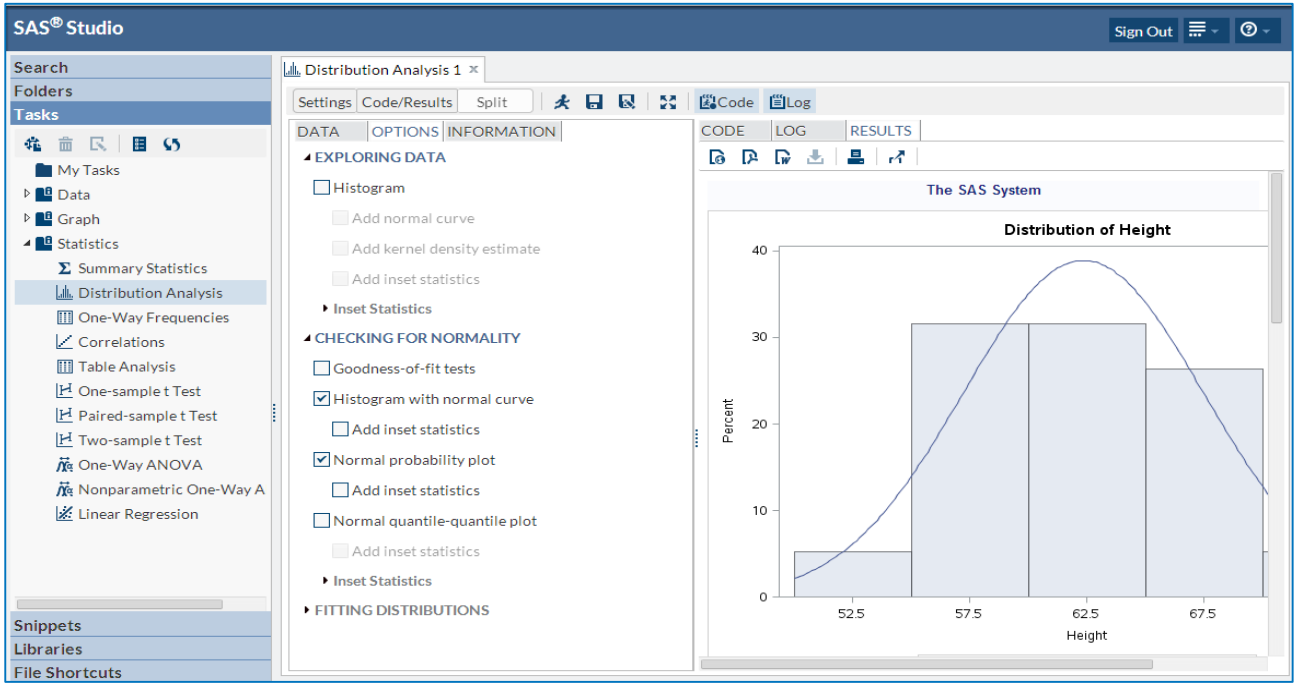
Table Viewer

- Sort or filter a table
- View or copy the generated SQL



Tasks

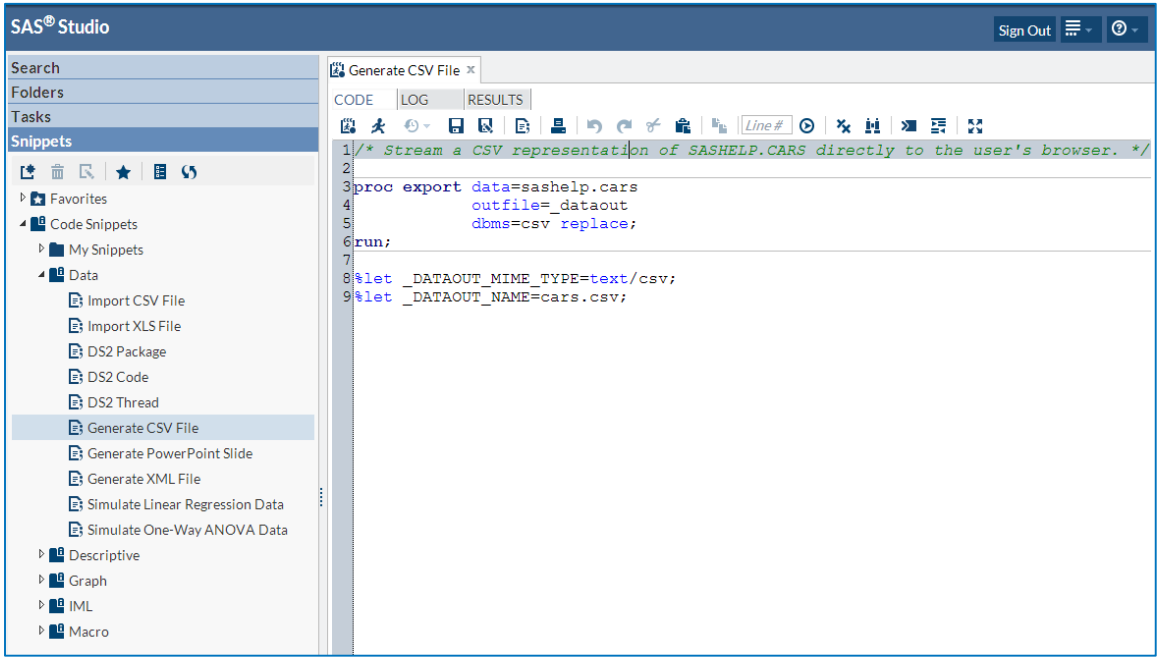
Tasks are point-and-click user interfaces which guide users through an analytical or other processes. Behind the scenes, SAS® code is generated.



Code Snippets

Overview

- Frequently used *code snippets* are provided in SAS® Studio
- Quickly insert SAS® Code
- Once inserted, you can modify the snippet code to meet your needs
- Easily create your own snippets
- Specify My snippets for easy access



What's New In SAS®/STAT for SAS 9.4

Lots of Change…

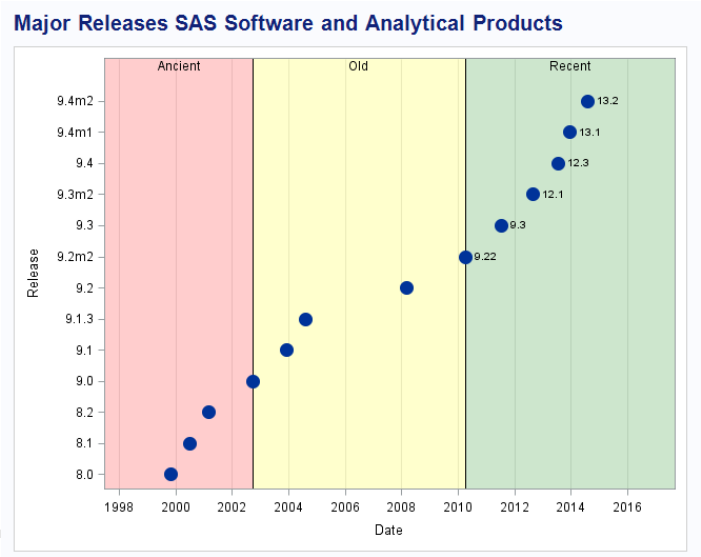
2 versions for SAS 9.3

5 versions for SAS 9.4

(so far)

New Numbering for Analytical Products

- SAS/STAT released independently of Base SAS®
- All analytical products follow new numbering scheme
- SAS/STAT released every 12–18 months



SAS/STAT

Focus Areas

Survival Analysis

- Interval Censored Survival Analysis
- Competing Risk Modeling

Methods for Handling Missing Data

- Multiple Imputation & Sensitivity Analysis with Proc MI
- Control Based Pattern Imputation

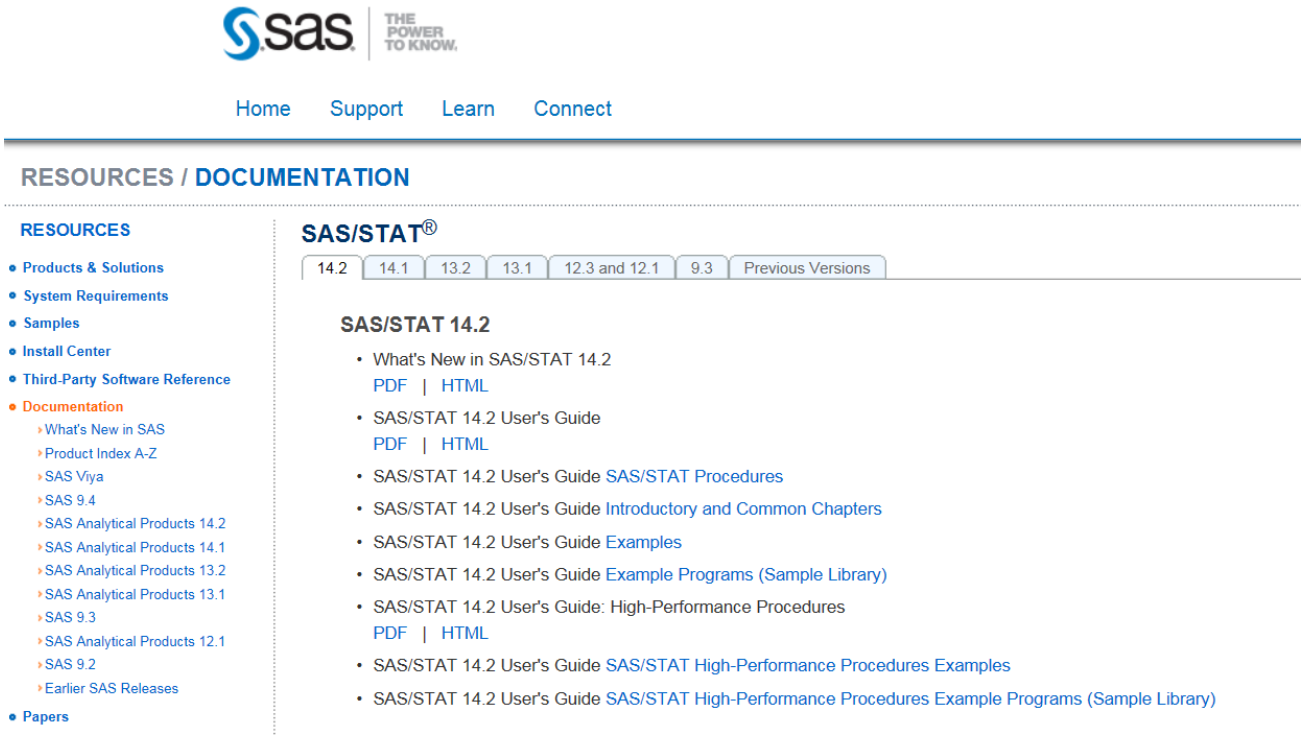
Bayesian Updates

- Bayesian Capabilities in Other Procedures
- PROC MCMC
- Procedures for Bayesian Applications

High Performance Statistics

What's New

Best Place to Go…



http://support.sas.com/documentation/onlinedoc/stat/index.html

Survival Analysis

Survival analysis

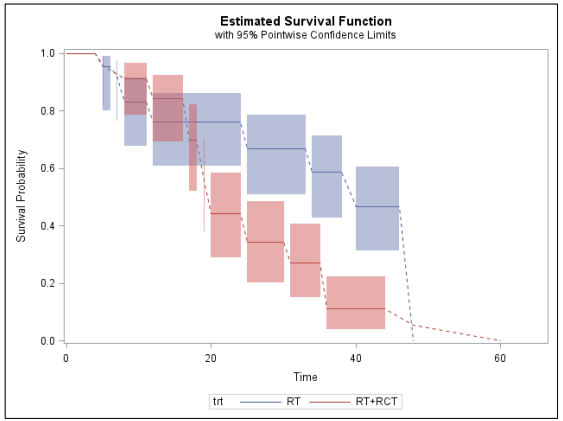
- Time-to-event data subject to censoring
- Goals are estimating survival function, testing equality of survival functions, and assessing covariate effects on lifetime

Procedure	Focus	Approach	Censoring	Covariates
LIFEREG	Lifetime	Parametric	Right or left	Yes
PHREG	Hazard function	Semiparametric	Right	Yes
QUANTLIFE	Lifetime	Semiparametric	Right	Yes
LIFETEST	Survival function	Nonparametric	Right	No
ICLIFETEST	Survival function	Nonparametric	**Interval**	No
ICPHREG	Hazard function	Multiple	**Interval**	Yes

Survival analysis

Analysis of Interval Censored Data

- Specialized methods are needed when events are known to have occurred within intervals of time.
- The ICLIFETEST procedure provides nonparametric methods
  - Supports multiple comparison
  - Estimates survival function with EMICM, Turnbull, ICM algorithms
  - Computes standard errors with imputation or bootstrap
  - Provides weighted generalized log-rank test for equality of survival functions



Survival analysis

Breast Cancer Study - Example

- Comparison of cosmetic breast deterioration after tumorectomy
- Two treatments: radiation therapy (RT) , radiation/chemotherapy (RCT)
- Deterioration times occurred between visits (lTime, rTime).

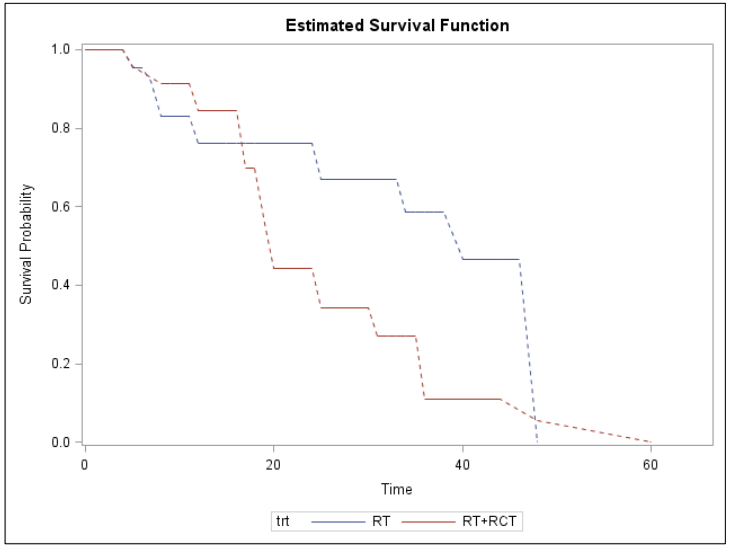
```
proc iclifetest data=BCS;
   time (lTime, rTime);
   test Trt;
run;
```

Source: Finkelstein and Wolfe (1985)

Survival analysis

Results

- Survival probabilities are estimated for disjoint intervals
- Group that only received radiation tended to survive longer



Test of Equality over Group			
Weight	Chi-Square	DF	Pr > Chi-Square
SUN	7.1907	1	0.0073

Survival analysis

Competing Risk Analysis of Time-to-Event Data

- Competing risks cause the event of interest to be impeded by a different type of event.

  Example: leukemia relapse is unobserved because patient dies.
- The Kaplan-Meier estimate of the survivor function is biased.
- An alternative is the cumulative incidence function, which is the marginal failure subdistribution of a given cause.
- The PHREG procedure fits the proportional subdistribution hazards model of Fine and Gray (1999).
  - Estimates for regression coefficients
  - Hazard ratio estimates for pairs of groups
  - Prediction for cumulative incidence

Survival analysis

Leukemia Study - Example

- Some patients experienced a relapse (Status=1) while others died while in remission (Status=0).
- Patients are classified into three disease groups (Disease).
- Do the hazard ratios differ for one group compared with another?

```
proc phreg data=Bmt plots(overlay=stratum)=cif;
   class Disease;
   model T*Status(0)=Disease / eventcode=1;
   hazardratio 'Pairwise' Disease / diff=pairwise;
   baseline covariates=Risk;
run;
```

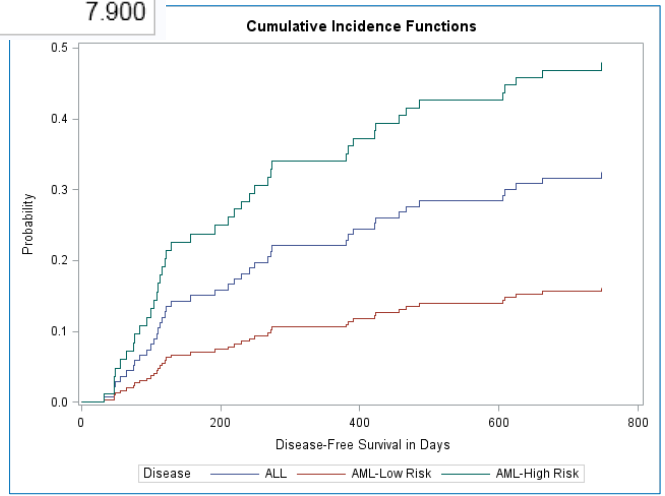
Source: Klein and Moeschberger (1997)

Survival Analysis

Results

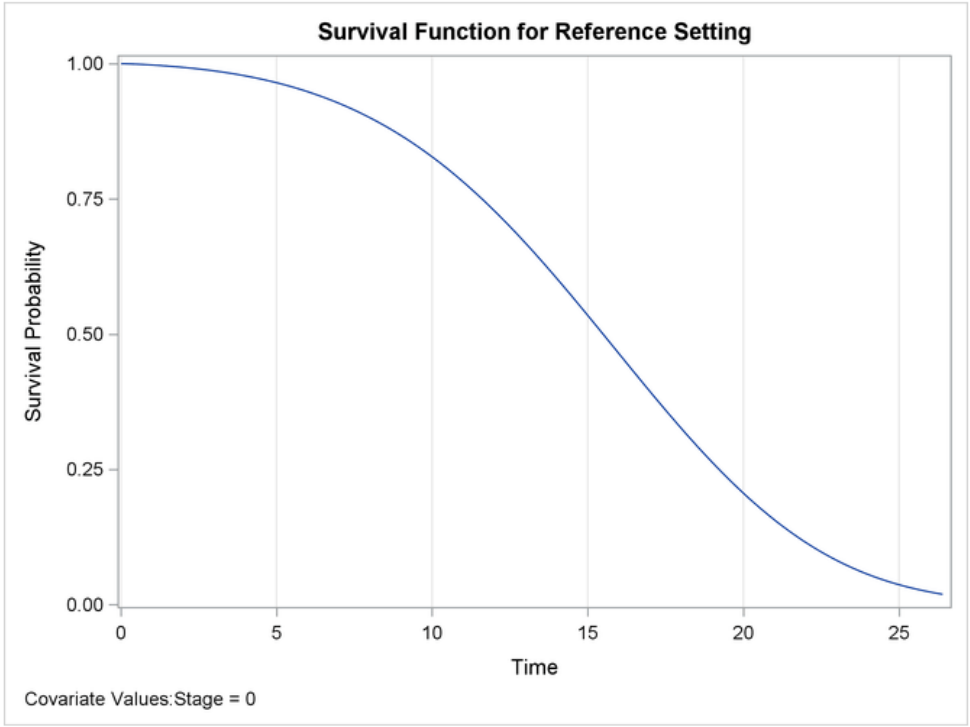
Pairwise: Hazard Ratios for Disease			
Description	**Point Estimate**	**95% Wald Confidence Limits**	
Disease ALL vs AML-Low Risk	2.233	0.964	5.171
Disease AML-Low Risk vs ALL	0.448	0.193	1.037
Disease ALL vs AML-High Risk	0.601	0.293	1.233
Disease AML-High Risk vs ALL	1.663	0.811	3.408
Disease AML-Low Risk vs AML-High Risk	0.269	0.127	0.573
Disease AML-High Risk vs AML-Low Risk	3.713	1.745	7.900

- Relapse in the AML high-risk group is more likely than in the ALL group.
- Relapse in the ALL group is more likely than in the AML low-risk group.



ICPHREG Procedure

The ICPHREG procedure is designed to fit proportional hazards regression models to interval-censored data.




```
ods graphics on;
proc icphreg data=hiv plot=surv;
   class Stage / desc;
   model (Left, Right) = Stage / basehaz=splines;
run;
```

QUANTLIFE Procedure

```
ods graphics on;
proc quantlife data=hiv log plots=quantplot seed=1268;
   class Drug;
   model Time*Status(0) = Drug Age / quantile=(0.25 0.5 0.75);
   Drug_Effect: test Drug;
run;
```

Figure 94.1: Model Fitting Information

The QUANTLIFE Procedure

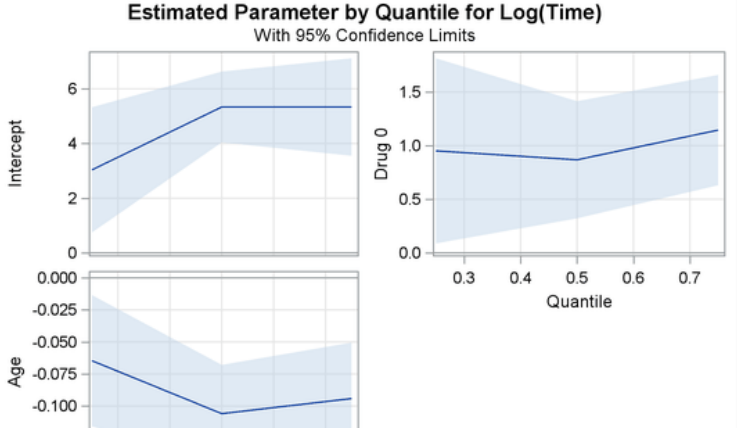
Model Information	
Data Set	WORK.HIV
Dependent Variable	Log(Time)
ring Variable	Status
ring Value(s)	0
er of Observations	100
d	Kaplan-Meier
ations	200
or Random Number Generator	1268

Test Drug_Effect Results

Quantile	DF	Chi-Square	Pr > ChiSq
0.2500	1	4.67	0.0307
0.5000	1	9.70	0.0018
0.7500	1	19.03	<.0001

Parameter Estimates

Quantile	Parameter		DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > |t
0.2500	Intercept		1	3.0373	1.1680	0.7482	5.3265	2.60	0.0108
	Drug	0	1	0.9516	0.4403	0.0887	1.8146	2.16	0.0331
	Drug	1	0	0	0	0	0	.	.
	Age		1	-0.0646	0.0261	-0.1158	-0.0135	-2.48	0.0150
0.5000	Intercept		1	5.3351	0.6605	4.0406	6.6296	8.08	<.0001
	Drug	0	1	0.8681	0.2786	0.3219	1.4142	3.12	0.0024
	Drug	1	0	0	0	0	0	.	.
	Age		1	-0.1059	0.0194	-0.1439	-0.0679	-5.46	<.0001
0.7500	Intercept		1	5.3351	0.9091	3.5532	7.1170	5.87	<.0001
	Drug	0	1	1.1451	0.2625	0.6307	1.6596	4.36	<.0001
	Drug	1	0	0	0	0	0	.	.
	Age		1	-0.0941	0.0223	-0.1378	-0.0505	-4.23	<.0001



METHODS FOR HANDLING MISSING DATA

MISSING DATA

Missing values are an issue. Most SAS statistical procedures exclude observations with missing values. Three options to handle:

- Some SAS procedures use all the available cases in an analysis.
- Another strategy is single imputation, in which you substitute a value for each missing value.
- Multiple imputation replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute.

The MI procedure is a multiple imputation procedure that creates multiply imputed data sets for incomplete p-dimensional multivariate data.

MISSING DATA

The FCS statement specifies a multivariate imputation by fully conditional specification (FCS) methods.

- For data with an arbitrary missing data pattern, the FCS methods enable you to impute missing values for all variables, assuming that a joint distribution for these variables exists.
- The FCS method requires fewer iterations than the MCMC method.

MISSING DATA

Sensitivity Analysis For Missing Data

- Multiple imputation usually assumes data are missing at random (MAR)
  - Under MAR, the probability that Y is missing for an observation depends only on the observed values of other variables, not on unobserved values of Y.
  - You *cannot* verify the MAR assumption from the data.
  - You *should* assess sensitivity of inference to departures from MAR.
- The new MNAR statement imputes missing values for plausible scenarios where data are missing not at random (MNAR)
  - Uses pattern-mixture model approach
  - Applies with fully conditional specification or monotone methods of imputation
- If these scenarios lead you to a conclusion that differs from inference under MAR, then you should question the MAR assumption.

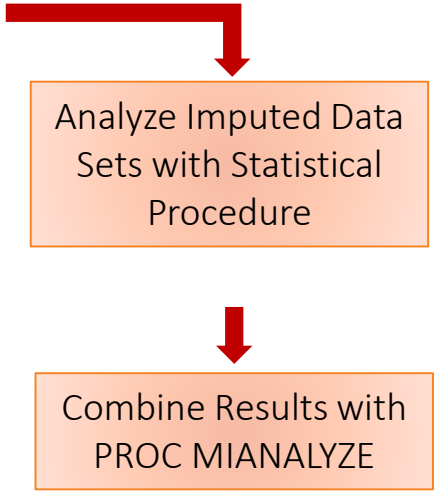
MISSING DATA

EXAMPLE: CONTROL-BASED PATTERN IMPUTATION

- Clinical trial for efficacy of a new drug with treatment and placebo groups
- Y0 is a baseline efficacy score, and Y1 is a follow-up efficacy score.
- The MI procedure assumes MAR by default.

```
proc mi data=Efficacy
out=OutMAR;
   class Trt;
   monotone reg(Y1);
   var Trt Y0 Y1;
run;
```

Source: Ratitch and O'Kelly (2011)



MISSING DATA

EXAMPLE: CONTROL-BASED PATTERN IMPUTATION

- An imputation model for missing values of Y1 in the **treatment** group is constructed from observed data in the **control** group.

```
proc mi data=Efficacy out=OutMNAR;
   class Trt;
   monotone reg(Y1);
   mnar model( Y1 / modelobs=(Trt='0')
);
   var Trt Y0 Y1;
run;
```

- Analyze OutMAR and OutMNAR with statistical procedure.
- For each scenario, combine p-values with PROC MIANALYZE.
- Question MAR if the p-values lead to different conclusions.

Bayesian Updates

Bayesian Models

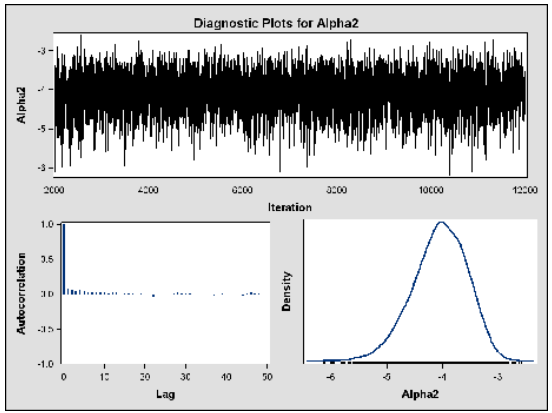
Advantages	Practical Issues
Flexibility for model building	Computational set-up
Incorporation of prior information	Specification of priors
Avoidance of asymptotics	Sampling algorithms
Intuitive interpretation of results	Convergence diagnostics

Bayesian Models

Three Avenues to Bayesian Modeling

1. **Built-in Bayesian analysis in widely-used procedures**
   - GENMOD, LIFEREG, PHREG, FMM
   - Broad range of models
   - Convenient syntax for users learning about Bayesian methods
2. **General-purpose modeling in MCMC procedure**
   - Likelihoods and priors are programmed by the user
   - Versatility for users trained in Bayesian methods
3. **Procedures for Bayesian applications**

Consistent diagnostics and posterior summaries



Bayesian Avenue 1

PROC GENMOD

- provides Bayesian analysis for generalized linear models.

PROC LIFEREG

- provides Bayesian analysis for parametric location-scale survival models.
- Supported prior distributions are normal and uniform.

Bayesian Avenue 1

PROC PHREG

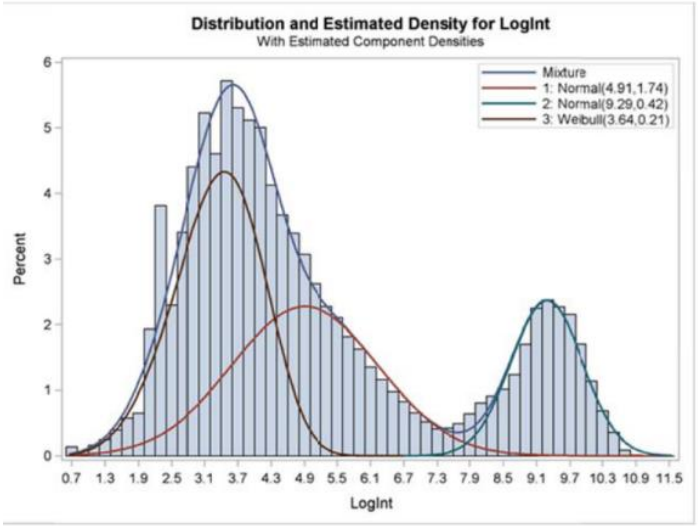
- provides Bayesian analysis for Cox regression models with time-independent and time-dependent predictor variables and accommodates all the methods handling ties.
- provides Bayesian analysis for piecewise exponential models where you can divide the time axis into sections having its own hazard rate.
- In SAS 9.4, Bayesian frailty models are supported and you can specify the gamma or lognormal distributions for the shared frailty.

Bayesian Avenue 1

PROC FMM fits statistical models to data where the distribution of the response is a finite mixture of univariate distributions.

- Performs maximum likelihood estimation for all models
- Provides Bayesian analysis for several models

- Useful for applications such as
  - estimating multimodal or heavy-tailed densities
  - modeling over-dispersed data.



Bayesian SAS/STAT®

GENMOD, LIFEREG, PHREG and FMM

- The **BAYES** statement requests Bayesian analysis.
- A set of standard prior distributions, posterior summary statistics, and convergence diagnostics are provided.
- You can specify **Adaptive rejection, Gamerman** or **Metropolis** sampling algorithms.

Bayesian Avenue 2

General purpose modeling in Proc MCMC

- *PROC MCMC* is a general purpose simulation procedure that uses Markov chain Monte Carlo (MCMC) techniques to fit a wide range of Bayesian models.
- It requires the specification of a likelihood function for the data and a prior distribution for the parameters.
- It enables you to analyze data that have any likelihood or prior distribution as long as they are programmable using SAS DATA step functions.

Bayesian Avenue 2

General purpose modeling in Proc MCMC

- You declare the parameters in the model and assign the starting values for the Markov chain with PARMS statements.
- Specify prior distributions for the parameters with PRIOR statements.
- Specify the likelihood function for the data with the MODEL statement.
- The model specification is similar to PROC NLIN and shares much of the same syntax as PROC NLMIXED.

```
PROC MCMC options;
    PARMS parameters and starting values;
    BEGINCNST;
      Programming Statements;
    ENDCNST;
    BEGINNODATA;
      Programming Statements;
    ENDNODATA;
    PRIOR  parameter ~ distribution;
    MODEL variable ~ distribution;
    RANDOM random effects specification;
```

Bayesian Avenue 2

The RANDOM statement

- Simulates the impact of random effects
- Allows for specification of hierarchical random effects
- Updated to support arbitrary hierarchy

The MODEL statement - updated so it fits more models with more convenient specifications.

Performance - Multithreaded

Bayesian Avenue 3

Procedures for Bayesian applications

Bayesian Choice Models

- Used in marketing research to model how people decide among alternative products and services.
- Relate P[individual *i* chooses alternative *j*] to personal characteristics and alternatives
- Account for individual differences with limited data
- Provide efficient, consistent estimation

Proc Bchoice

- Standard models: multinomial logit, nested logit, multinomial probit
- Random effects models that address individual preference
- Priors that can be specified or default to noninformative
- Metropolis sampling approach of Gamerman (1997)

High Performance STATISTICS

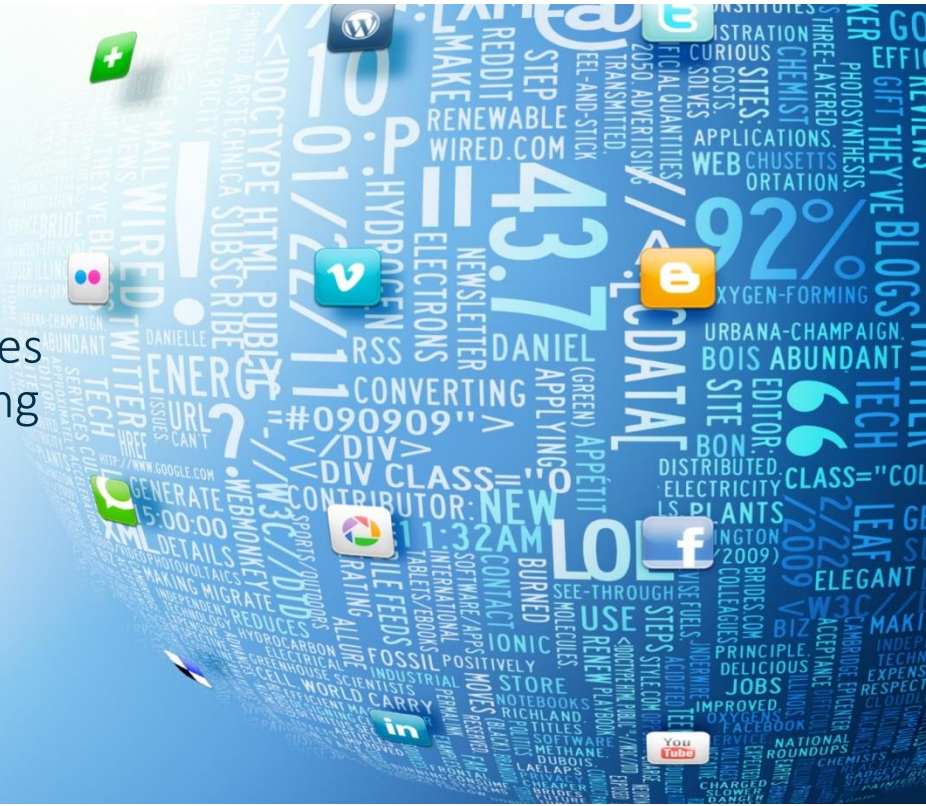
What's New

Big Data and High Performance

The Challenges

The data required to be analyzed continues to grow and is coming from ever increasing numbers of sources.

Simultaneously, the time required to get results is ever decreasing.

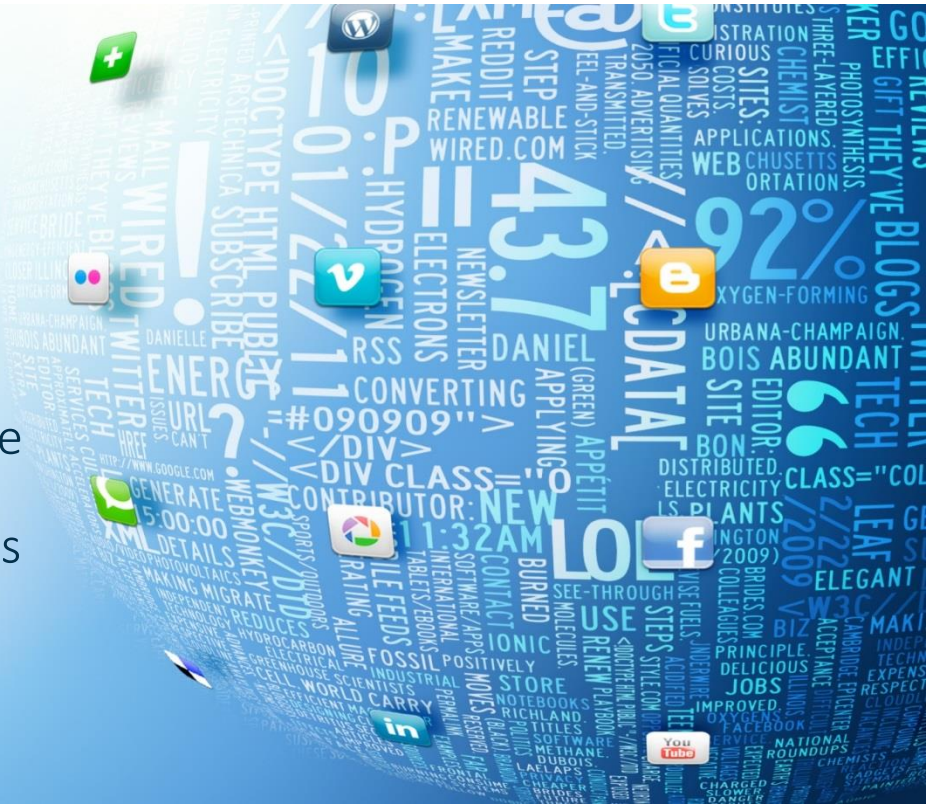


What's New

Big Data and High Performance

What if ….

Hardware could be exploited to perform computations in parallel, taking advantage of in-memory and making it faster to get results and possible to use larger amounts of data?



High Performance

One of the most significant changes in SAS 9.4 is the release of high performance procedures for use in a single-machine mode (sometimes referred to as SMP).

SAS® 9.4

SAS High Performance Procedures

High Performance Statistics	High Performance Data Mining	High Performance Econometrics	High Performance Forecasting	High Performance Optimization	High Performance Text Mining
HPLOGISTIC HPGENSELECT HPREG HPLMIXED HPNLMOD HPSPLIT HPFMM HPCANDISC HPPRINCOMP HPPLS HPQUANTSELECT	HPREDUCE HPNEURAL HPFOREST HP4SCORE HPDECIDE HPCLUS HPSVM HPBNET HPTSDR	HPCOUNTREG HPSEVERITY HPQLIM HPPANEL HPCOPULA HPCDM	HPFORECAST	OPTLSO Select features in OPTMILP OPTLP OPTMODEL OPTGRAPH HPCDM	HPTMINE HPTMSCORE
Common Set: HPDS2, HPDMDB, HPSAMPLE, HPSUMMARY, HPIMPUTE, HPBIN, HPCORR					

SAS/STAT 13.2 Documentation
SAS/STAT High-Performance Procedure Documentation

SAS® High-Performance Procedures

- Primary purpose is predictive modeling
- Sometimes features from multiple MVA procedures are combined into a single HP procedure (e.g. HPREG)
- Sometimes there is no MVA counterpart procedure
- Syntax has been standardized across all HP procedures:
  - Some MVA *options* have been standardized into *statements* for HP procedures (e.g. SELECTION)
- Calculations for statistical inference are not included

SAS® High-Performance Procedures

HPGENSELECT Procedure

EXAMPLE – the HPGENSELECT Procedure:

- Model selection for generalized linear models
- Model fitting with maximum likelihood
- Model building with forward, backward, stepwise methods with selection based on AIC, AICC, or SBC
- Supports variety of distributions: Normal, binomial, multinomial, Poisson, negative binomial, and Tweedie
- Supports variety of link functions
- Models for zero-inflated count data
- Models for ordinal and unordered multinomial data
- Multithreaded for fast performance

HPGENSELECT PROCEDURE

"Which procedure do I use?"

GENMOD Procedure

- Fits models
- Moderate-to-large data
- Designed for inferential analysis

HPGENSELECT Procedure

- Fits and builds models
- Large-to-massive data
- Designed for predictive modeling

HPGENSELECT Procedure

Three Benefits

1. The HPGENSELECT procedure provides variable selection for generalized linear models – new functionality in SAS/STAT®
2. You can run the procedure in single-machine mode and exploit all the cores.
3. As your problem size grows, you can take full advantage of all the cores and memory available in distributed computing environments with the additional license of SAS® High Performance Statistics.

HPGENSELECT Procedure

Proc hpgenselect

```
PROC HPGENSELECT < options> ;
    CLASS variable < (options)>. . . < variable < (options)> > </ global-options> ;
    CODE < options> ;
    MODEL response< (response-options) > = < effects> < / model-options> ;
    MODEL events/trials< (response-options) > = < effects> < / model-options> ;
    OUTPUT < OUT=SAS-data-set>
                < keyword < =name>>. . .
                < keyword < =name>> < / options> ;
    PERFORMANCE performance-options ;
    SELECTION selection-options ;
    FREQ variable ;
    ID variables ;
    WEIGHT variable ;
    ZEROMODEL < effects>< / zeromodel-options> ;
```

HP Procedures

SAS High-Performance Common Procedures

SAS High-Performance Procedure	Closest Traditional Procedure	High-Performance Analytics Procedure Key Functionality
HPBIN	None	Performs binning of numeric variables. Calculate Weight of Evidence (WOE) transformation for a binary target on binned numeric inputs.
HPCORR	CORR	Computes Pearson correlation coefficients and significance probabilities.
HPDMDB	DMDB	Calculates summary statistics for interval and categorical variables. Statistics can be saved in a SAS catalog for input into some SAS High-Performance Analytics procedures.
HPDS2	None	Provides the ability to execute DS2 (DATA step 2) in parallel in the grid. DS2 is designed to provide DATA step functionality in a database.

HP Procedures

SAS High-Performance Common Procedures

SAS High-Performance Procedure	Closest Traditional Procedure	High-Performance Analytics Procedure Key Functionality
HPIMPUTE	None/STDIZE	Impute missing values for interval variables and creates missing data indicator variables.
HPSAMPLE	SURVEY-SELECT	Selects simple random samples without replacement or creates a random partition. When a partition is created, an indicator variable is created identifying the partition to which a case belongs. The sample or partition can be stratified by a categorical variable.
HPSUMMARY	SUMMARY	Calculates descriptive statistics and estimates quantiles for numeric variables.

EXAMPLE


```
proc hpimpute data=sampsio.hmeq out=out1;
      input mortdue value clage debtinc;
      impute mortdue / value = 70000;
      impute value / method = mean;
      impute clage / method = random;
      impute debtinc / method = pmedian;
run;
```

Imputation Results

Variable	Imputation Indicator	Imputed Variable	N Missing	Type of Imputation	Imputation Value (Seed)
MORTDUE	M_MORTDUE	IM_MORTDUE	518	Given value	70000
VALUE	M_VALUE	IM_VALUE	112	Mean	101776
CLAGE	M_CLAGE	IM_CLAGE	308	Random	5.00000
DEBTINC	M_DEBTINC	IM_DEBTINC	1267	Pseudo Median	34.81696

HPIMPUTE Procedure Documentation

HP Procedures

SAS High-Performance Statistics Procedures

SAS High-Performance Procedure	Closest Traditional Procedure	High-Performance Analytics Procedure Key Functionality
HPGENSELECT	GENMOD and GLMSELECT	Provides model fitting and model selection for generalized linear models including the Tweedie. Also fits zero-inflated Poisson and negative binomial models
HPLMIXED	MIXED and HPMIXED	Fits linear mixed models. Correlation structures, including variance components, compound symmetry, unstructured, and AR(1), can be defined in a RANDOM statement.
HPLOGISTIC	LOGISTIC /DMREG	Fits models to model binary, ordinal, or nominal targets. Perform stepwise selection and choose a best model based on a several fit statistics including SBC. A CODE statement saves SAS scoring code to file.
HPNLMOD	NLIN and NLMIXED	Fits nonlinear fixed-effects models using OLS or ML. Parameter estimates can be constrained using a BOUNDS or RESTRICT statement.

HP Procedures

SAS High-Performance Statistics Procedures

SAS High-Performance Procedure	Closest Traditional Procedure	High-Performance Analytics Procedure Key Functionality
HPREG	GLMSELECT	Fits linear regression models to interval targets. Performs model selection using stepwise methods including LAR and LASSO. Best model can be chosen using validation data or cross validation.
HPSPLIT	/ ARBORETUM	Grows classification trees using entropy, Gini, or FastCHAID methods. Provides C4.5-style pruning using validation data.

Example


```
proc hpsplit data=sashelp.hmeq maxdepth=7 maxbranch=2;
    target BAD;
    input DELINQ DEROG JOB NINQ REASON / level=nom;
    input CLAGE CLNO DEBTINC LOAN MORTDUE VALUE YOJ / level=int;
    criterion entropy;
    prune misc / N <= 6;
    partition fraction(validate=0.2);
    rules file='hpsplhme2-rules.txt';
    score out=scored2;
run;
```

```
*-------------------------------------------------------------*
NODE = 2
*-------------------------------------------------------------*
DELINQ IS ONE OF 5, 6, 7, 8, 10, 11, 12, 13, 15
AND DELINQ IS ONE OF 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 1
    PREDICTED VALUE IS 1
    PREDICTED 1 = 0.9342( 71/76)
    PREDICTED 0 = 0.06579( 5/76)
*-------------------------------------------------------------*
NODE = 4
*-------------------------------------------------------------*
NINQ IS ONE OF 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 17
AND DELINQ IS ONE OF MISSING, 1, 2, 3, 4
AND DELINQ IS ONE OF 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 1
    PREDICTED VALUE IS 1
    PREDICTED 1 = 0.8714( 61/70)
    PREDICTED 0 = 0.1286( 9/70)
```

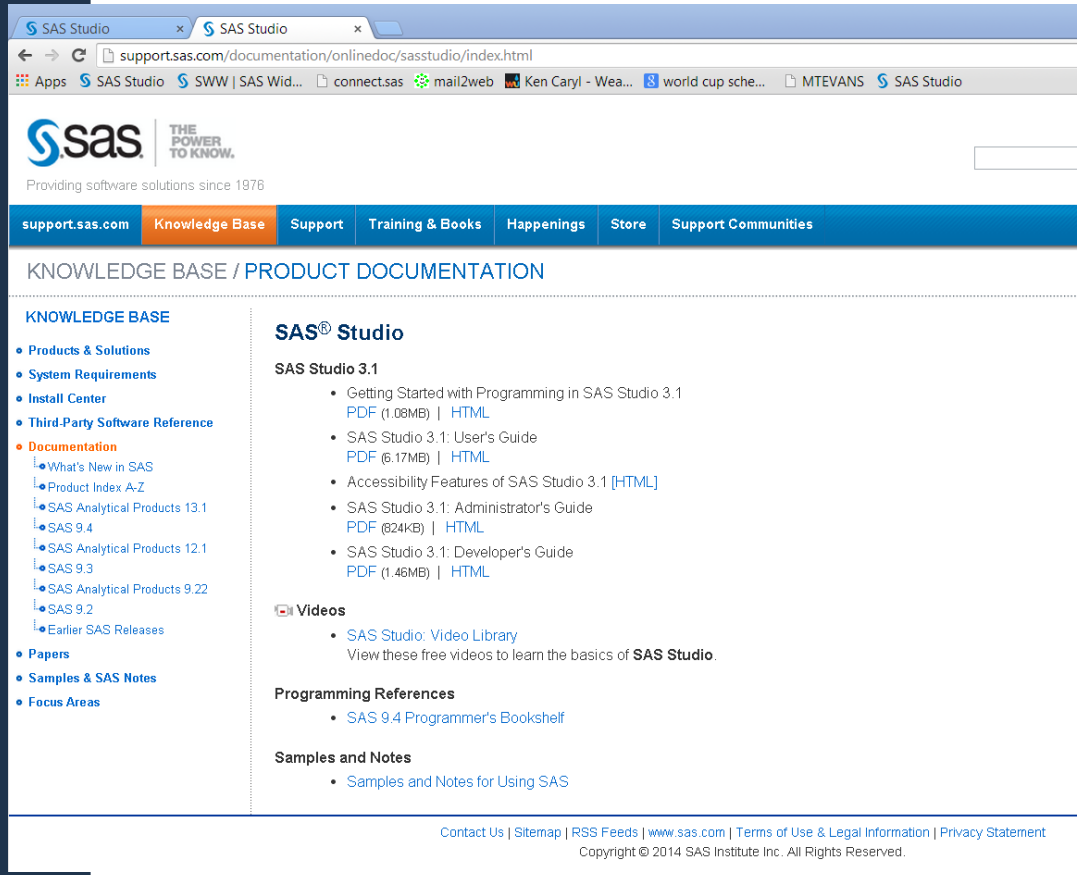
HPSPLIT Procedure Documentation

Resources

SAS® Studio

resources

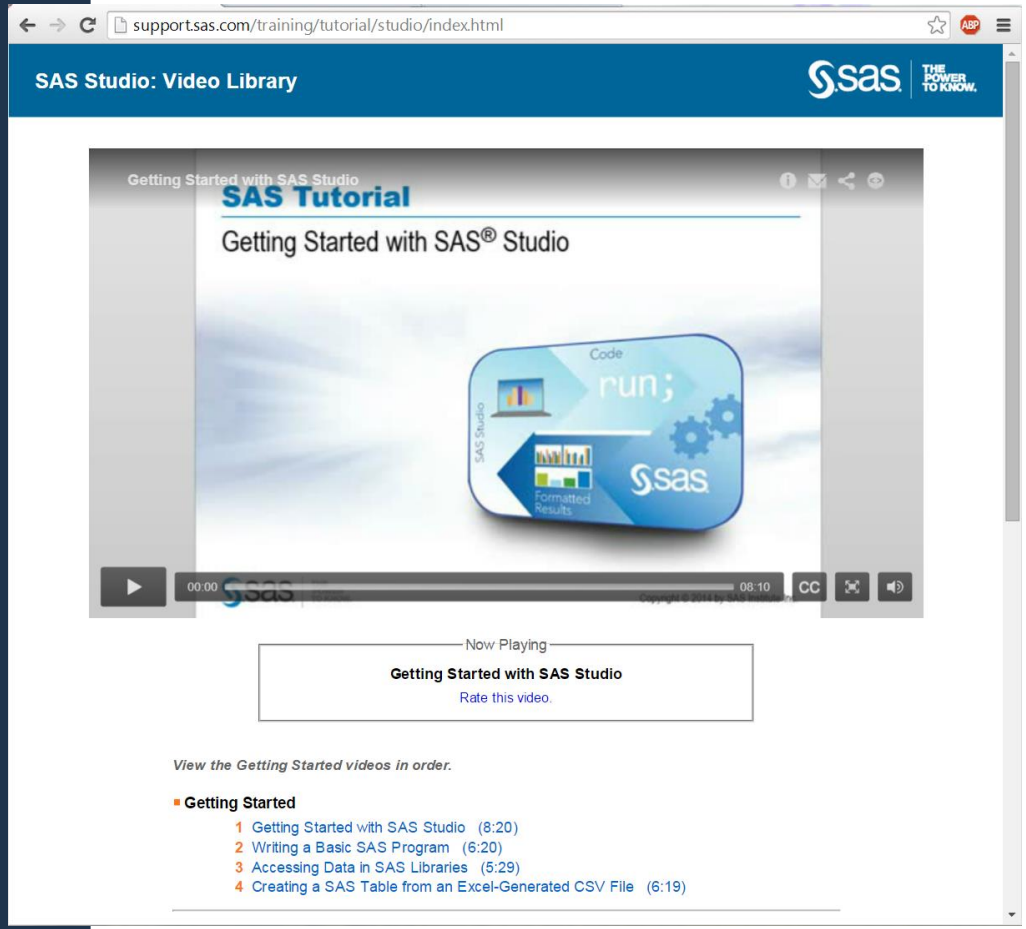
Main SAS® Studio Documentation page



SAS® Studio

resources

SAS® Studio
Video Library page



Resources

What’s New

What’s New in SAS

- http://support.sas.com/documentation/whatsnew/

What's New in SAS Enterprise Guide
http://support.sas.com/documentation/onlinedoc/guide/

Resources

SAS Analytics

- STAT, IML, OR, ETS Papers
- Communities
- STAT Videos

Resources

SAS 9.4

- Online documentation:
  http://support.sas.com/documentation/onlinedoc/base/index.html

- SAS/STAT High Performance Documentation
  http://support.sas.com/documentation/cdl/en/stathpug/66410/HTML/default/viewer.htm#stathpug_intro_sect001.htm

- Leveraging Big Data Using SAS High-Performance Analytics Server
  http://support.sas.com/resources/papers/proceedings13/399-2013.pdf

- High Performance Statistical Modeling
  http://support.sas.com/resources/papers/proceedings13/401-2013.pdf

- SAS Meets Big Iron: High Performance Computing in SAS Analytic Procedures
  http://www2.sas.com/proceedings/sugi27/p246-27.pdf

- Blog Post: Documentation for high-performance analytics *(definitions too)*
  http://blogs.sas.com/content/publishing/2012/04/17/documentation-for-high-performance-analytics/

SAS Academic Programs

https://www.sas.com/en_us/learn/academic-programs.html

Thank you for attending!

sas.com