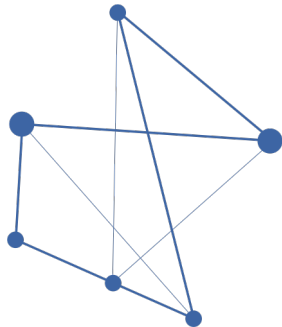




SAS/STAT High-performance procedures contrasted with their legacy counterparts



**Your 1st choice
of global FSP partner**

Author:
Oscar Bortolami (on behalf of Roman Ganzha)

GCE Position:
Principal Biostatistician

What?	Exactly what?
Intro	Disclaimer/An overview of what these procedures are
Body	An overview side by side of the differences in SAS between the two procedures
Is this this useful to me?	Many procedures are there but to think about if applicable to my work and whether it is worth investing money on that
Conclusion/What's next?	Simulations on Performance. Where these procedures can possibly reduce the time and/or cost of a project
References	SAS/STAT® 14.3 - User's Guide - High-Performance Procedures https://documentation.sas.com/?docsetId=stathpug&docsetTarget=titlepage.htm&docsetVersion=14.3&locale=en

Disclaimer:

I am presenting on behalf of Roman Ganzha

What is known on the topic?	the SAS/STAT high-performance procedures (HPPs) take advantage of computational capability (e.g. parallel processing or multithread processes on a single processor). This may enable a shorter execution time or different functionality compared to traditional procedures
What does this presentation add?	To increase awareness on such procedures per se and on the difference between HPPs and their SAS/STAT legacy counterparts
What is the focus on?	There will be more emphasis on Clinical trials related procedures (e.g. linear regression, GLM, mixed models)

2.1 Body – Overview

What	Use	HPP	Legacy Procedure
Generalized Additive models	Smoothing/non parametric regressions	GAMPL	GAM
Canonical discriminant analysis	Dimensional Reduction	HPCANDISC	CANDISC
Finite Mixtures Models	Modelling data that could be described by two or more underlying distributions	HPFMM	FMM
Generalized linear models		HPGENSELECT	GENMOD
Mixed Models		HPLMIXED	MIXED
Non Linear mixed models		HPNLMOD	NLIN/NLMIXED

2.2 Body – Overview

What	Use	HPP	Legacy Procedure
Partial Least Square regression models		HPPLS	PLS
Principal Component Analysis	Dimensional Reduction	HPPRINCOMP	PRINCOMP
Quantile regression models		HPQUANTSELECTION	QUANTSELECT
Regression models		HPREG	GLMSELECT, GLM AND REG

2.3 Body – Generalized Additive Models

What	GAMPL	GAM
Constructing Spline Basis Expansions	Different methodology compared to legacy procedure	
Fitting Generalized Additive Models	Different methodology compared to legacy procedure	
Distribution Families and Link Functions	More options available compared to legacy procedure. For instance PROC GAMPL fits models in the negative binomial family.	

2.4 Body – Canonical Discriminant analysis

What	HPCANDISC	CANDISC
Input	Only the ordinary SAS data set	Ordinary SAS data set and other types of special SAS data sets
ID statement	Available	Not available

2.5 Body – Finite Mixtures Models

What	HPFMM	FMM
CLASS statement	More options available compared to legacy procedure	
PERFORMANCE statement	More options available compared to legacy procedure	
A priori knowledge and uncorrelated observations	The number of components k in the finite mixture is known a priori and is not a parameter to be estimated. The parameters of the components are distinct a priori. The observations are uncorrelated	

2.6 Body – Generalized Linear Models

What	HPGENSELECT	GENMOD
CLASS STATEMENT	2 parameterizations: GLM and reference	13 parameterizations
Optimization Techniques	8 options	Newton-Raphson optimization with ridging
Post fitting analyses	wide variety	limited
BAYES statement	Not available	Available
EXACT statement	Not available	Available
PARTITION statement	Available (creation of training and testing datasets)	Not available

2.7 Body –Mixed Models

What	HPLMIXED	MIXED
Optimization Techniques	8 options	Newton-Raphson optimization with ridging
Method Techniques (estimation)	ML or REML	ML, REML or MIVQUE0 (plus type 1,2,3)
CODE Statement*	Not available	Available
CONTRAST statement	Not available	Available
PRIOR statement	Not available	Available (Bayesian analysis)
RANDOM statement		More options compared to HPMIXED
REPEATED Statement		More options compared to HPMIXED

*writes SAS DATA step code for computing predicted values of the fitted model either to a file or to a catalog entry. This code can then be included in a DATA step to score new data.

2.8 Body – Non Linear Mixed Models

What	HPNLMOD	NLIN/NLMIXED
Optimization techniques	Different methods compared to legacy procedure	
BOOTSTRAP Statement	Not available	Available (NLIN)
CMPTMODEL Statement*	Not available	Available (NLMIXED)
MODEL Statement	8 models (RESIDUAL or LS as new model)	7 models (NLMIXED)
RANDOM Statement	Not available	Available (NLMIXED)

* The CMPTMODEL statement computes predicted concentrations from a specified one-, two-, or three-compartment model

2.9 Body – Partial Least Square models

What	HPPLS	PLS
Algorithm techniques	3 techniques (RLGW not available)	4 techniques
EFFECT Statement*	Not available	Available

* The EFFECT statement enables you to construct special collections of columns for design matrices (e.g. splines/polynomials).

2.10 Body – Principal Component Analysis

What	HPPRINCOMP	PRINCOMP
PLOTS option	Not available	Available
Method Techniques	EIG ITERGS<(iter-options)> NIPALS<(iter-options)>	eigenvalue decomposition
CODE Statement*	Available	Not available
OUTPUT Statement**	Available	Not available

* The CODE statement enables you to write SAS DATA step code for computing the principal component scores either to a file or to a catalog entry.

** The OUTPUT statement creates a data set that contains observationwise statistics

2.11 Body – Quantile Regression Models

What	HPQUANTSELECT	QUANTSELECT
Algorithm option	Predictor-corrector interior point algorithm, which was originally designed to solve support vector machine classifiers for large data sets	SIMPLEX SMOOTH
EFFECT Statement*	Not available	Available
Confidence Intervals	Produced	Not directly produced

* The EFFECT statement enables you to construct special collections of columns for design matrices (e.g. splines/polynomials).

2.12 Body – Regression

What	HPREG	REG/GLM
Ridge option	Not available	Available (REG)
MANOVA Statement	Not available	Available (GLM)
MTEST Statement	Not available	Available (REG)
RANDOM Statement	Not available	Available (GLM)
REWEIGHT/REFIT statement	Not available	Available (REG)

3 Is this useful to me?

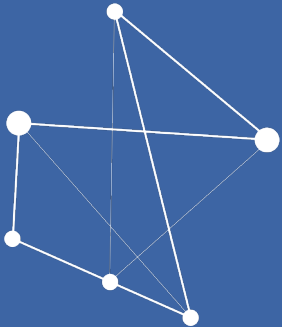
What?	Considerations
License	If the company does not have a licence for HPPs, evaluate whether it is worth to have it
Size/complexity of datasets	Think about the maximum size of datasets you are dealing with and which type of procedures. Think about the transfer speed of data (the transfer of data may be the bottle neck rather than the execution time)
How frequently I use computational intensive procedures	This could be one additional driver
Some computational intensive statements may not be available in HPPS	This could be one additional driver

4 Conclusion/ What's next?

- Where computational power is required, HPPs may provide an alternative solution to SAS/STAT base procedure.
- Unfortunately some computational intensive statements are not implemented in the HPP procedures
- Future work may include the evaluation of the performance (e.g. time to complete the procedure) in datasets with
 - increasing complexity (variables, observations)
 - when increasing the number of repetitions of a task (e.g. simulations)
 - when using different estimators
 - Other



Thank You for attending



**Your 1st choice
of global FSP partner**

Author:
Oscar Bortolami (on behalf of Roman Ganzha)

GCE Position:
Principal Biostatistician