Step Up Your Statistical Practice with Today's SAS/STAT® Software

Phil Gibbs

SAS Institute Technical Support

Are you over-relying on familiar procedures, and unaware of newer procedures that could benefit your work?

Should you *always* use

- PROC REG for building predictive models?
- PROC GENMOD for handling dropouts in longitudinal studies?
- PROC LIFETEST for analyzing interval-censored data?
- PROC MIXED for fitting linear mixed models?

This presentation explains the advantages of newer tools in four of the many areas where SAS/STAT is expanding

1. Regression model building
2. Inferential analysis of generalized linear models
3. Survival analysis
4. Analysis of mixed models

This is a high-level overview, which gives you the big picture without descending into details

SAS® users on balloon safari at Magaliesburg, South Africa, November 2015

Regression Model Building

Tech Support is often asked,
“Can you add a CLASS statement to PROC REG?”

Kathleen Kiernan
Analytical Technical Support

PROC GLMSELECT is now the flagship procedure for building standard regression models

Designed for

- Selecting the “best” model when you are choosing from hundreds of variables—or even thousands
- Continuous or categorical predictors
- Explanatory models or predictive models

PROC GLMSELECT provides many advantages for building regression models with large data

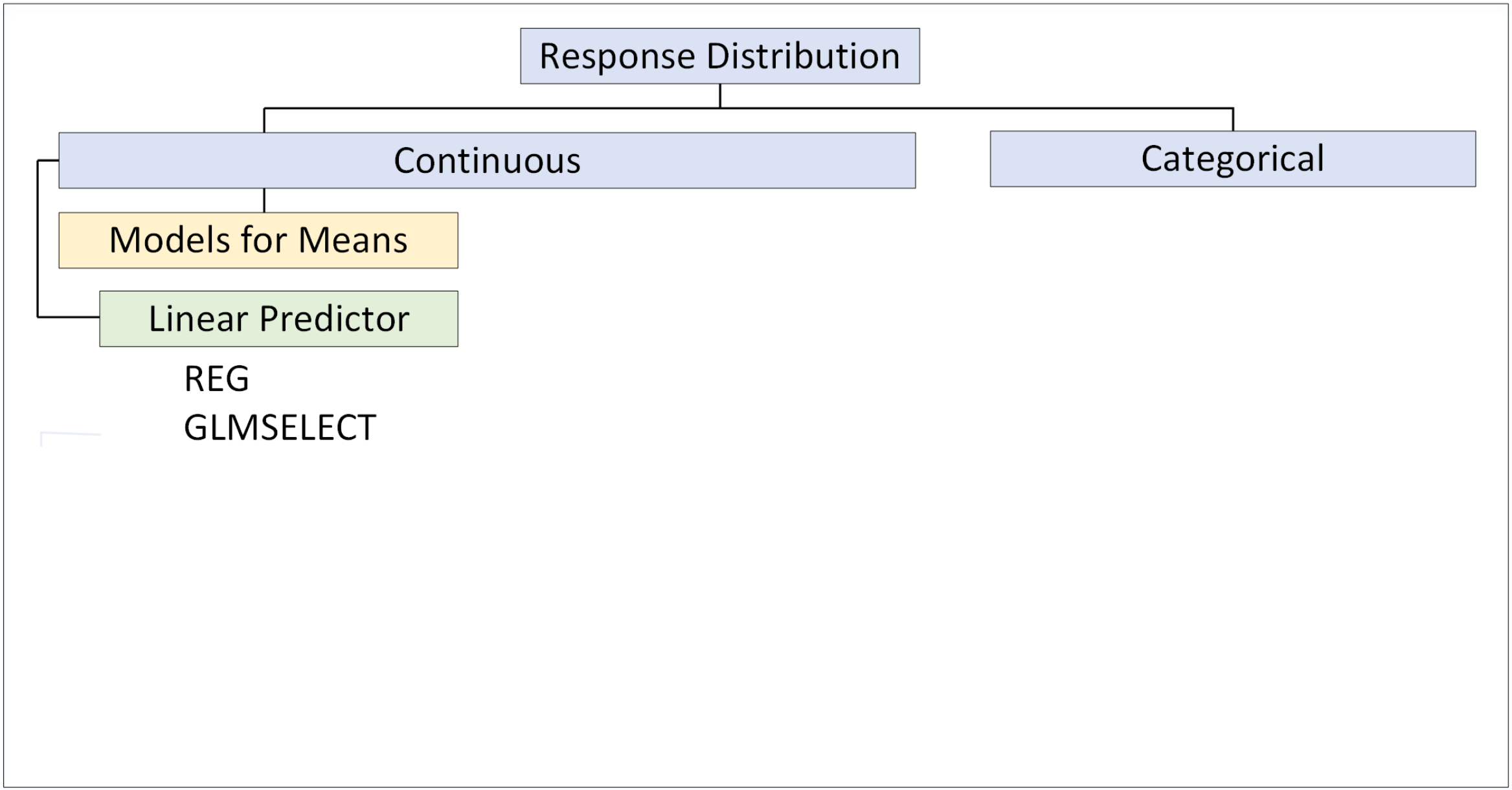
- Effect selection methods for general linear models

  Predictors can be main effects of continuous or classification variables, and interaction effects

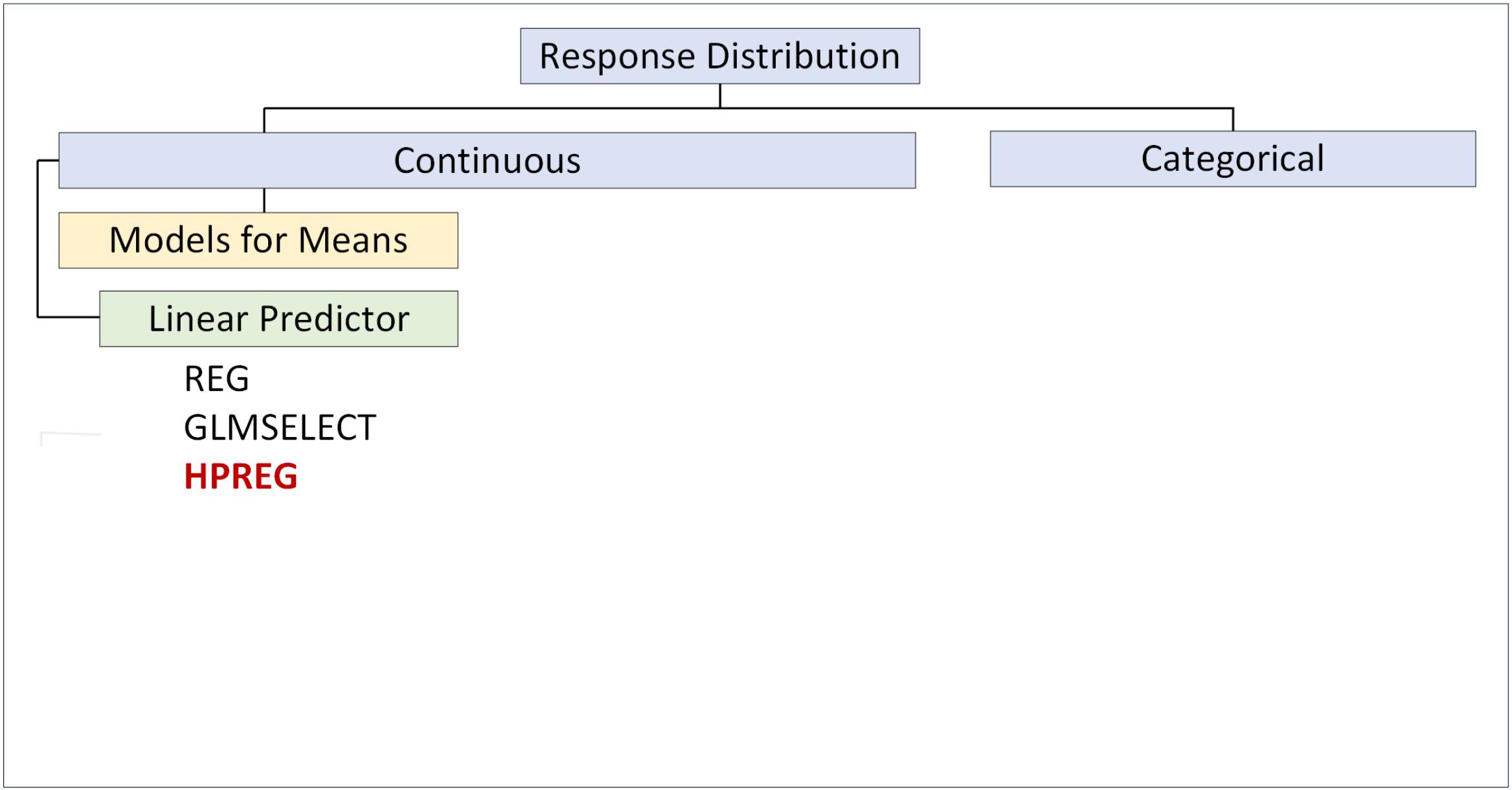
- Lasso methods for sparse, more interpretable models
- Data partitioning to avoid overfitting

Use PROC REG for fitting regression models when you need inferential methods, influence statistics, and diagnostics

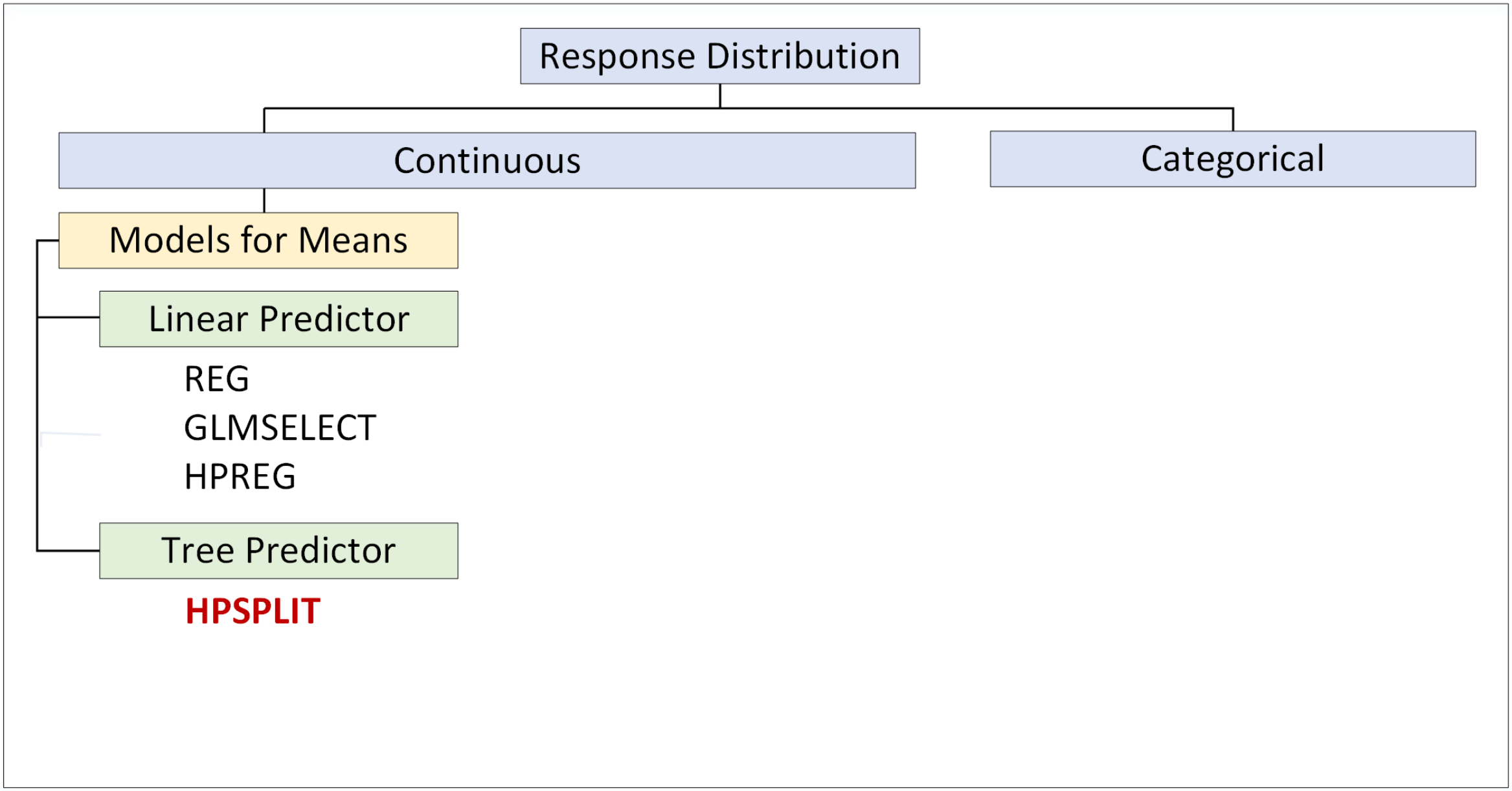
Model building procedures are available for a variety of goals and methods



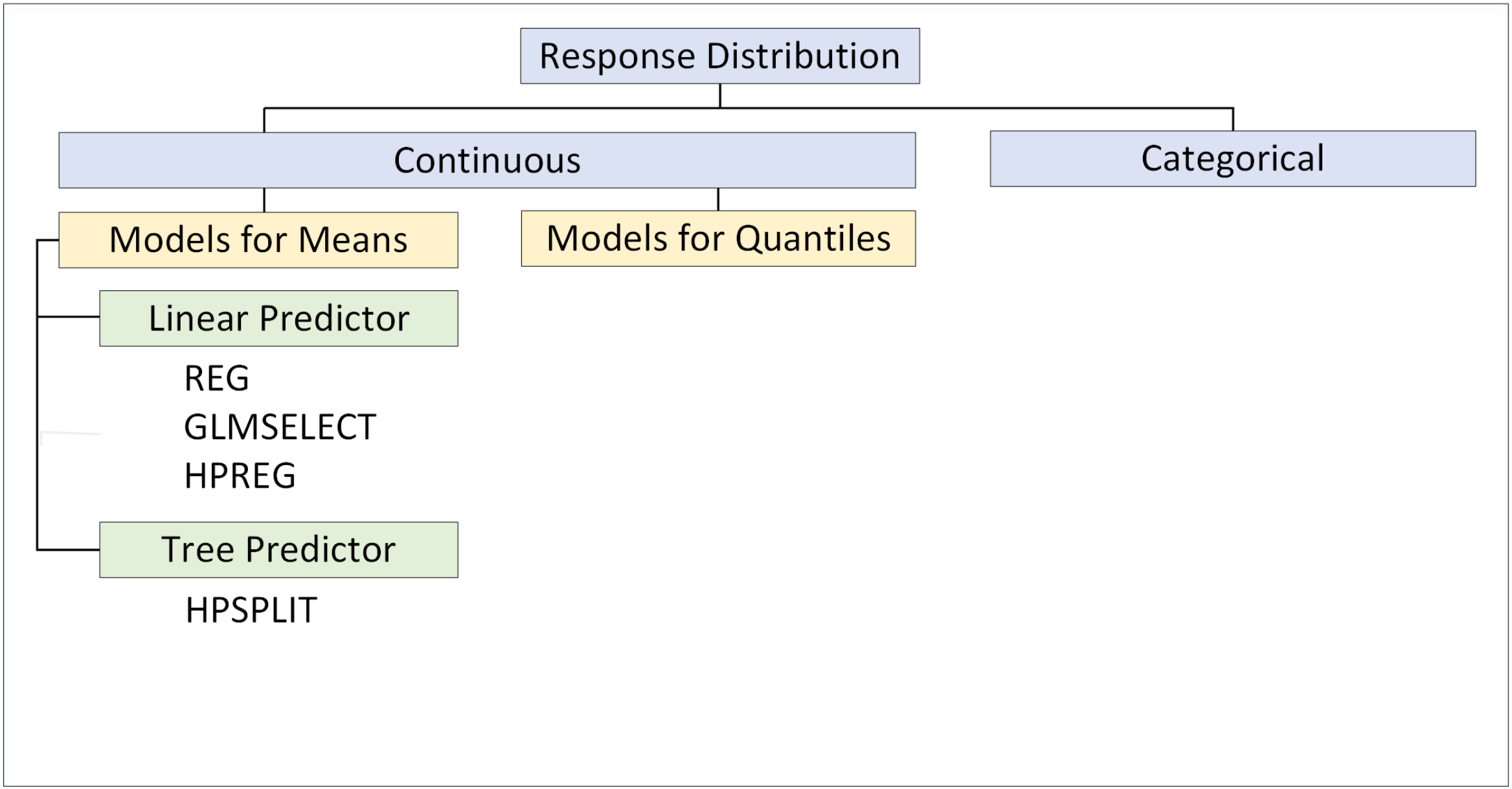
PROC HPREG is a high-performance regression modeling procedure



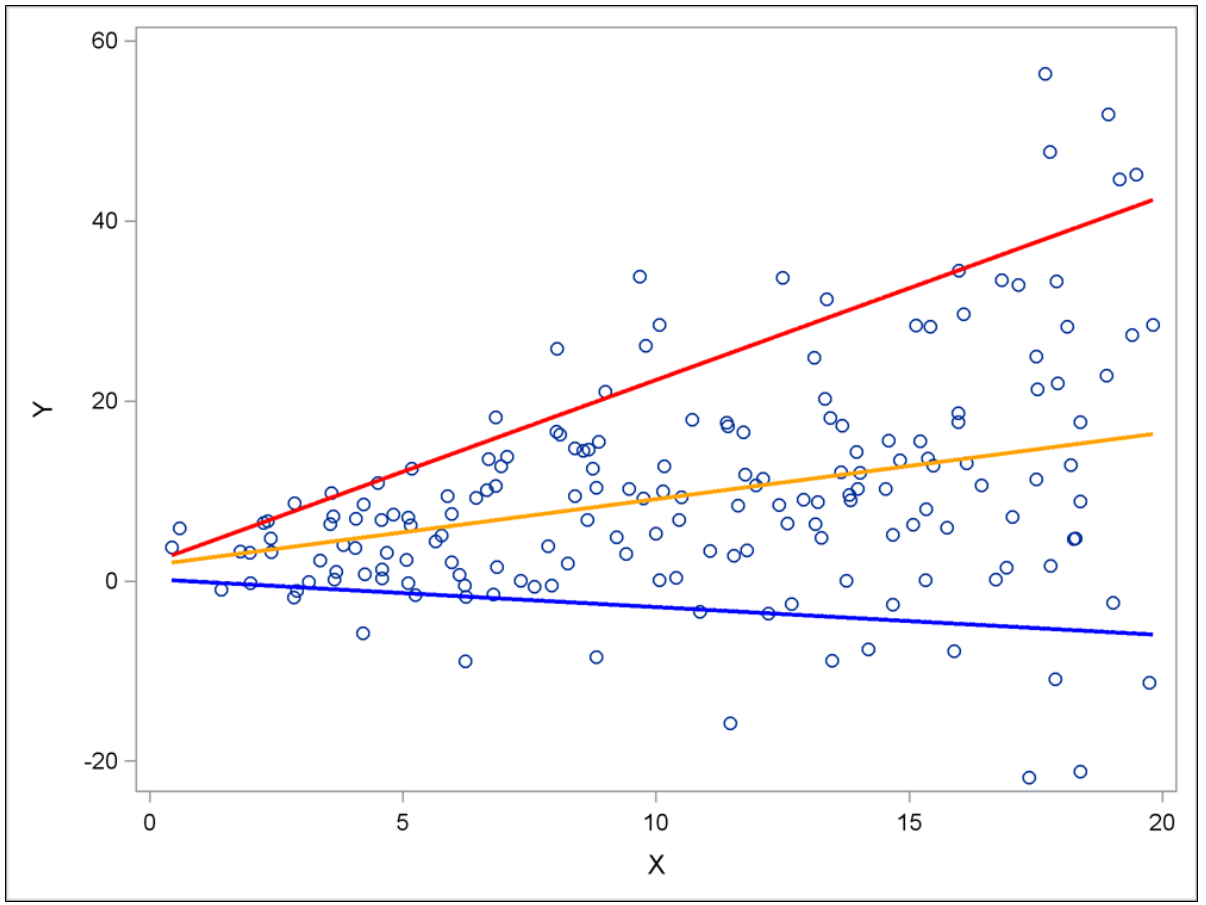
PROC HPSPLIT builds classification and regression trees



Models for means are not always adequate …



Regression models for quantiles (percentiles) are useful when the conditional distribution of the response varies with covariates

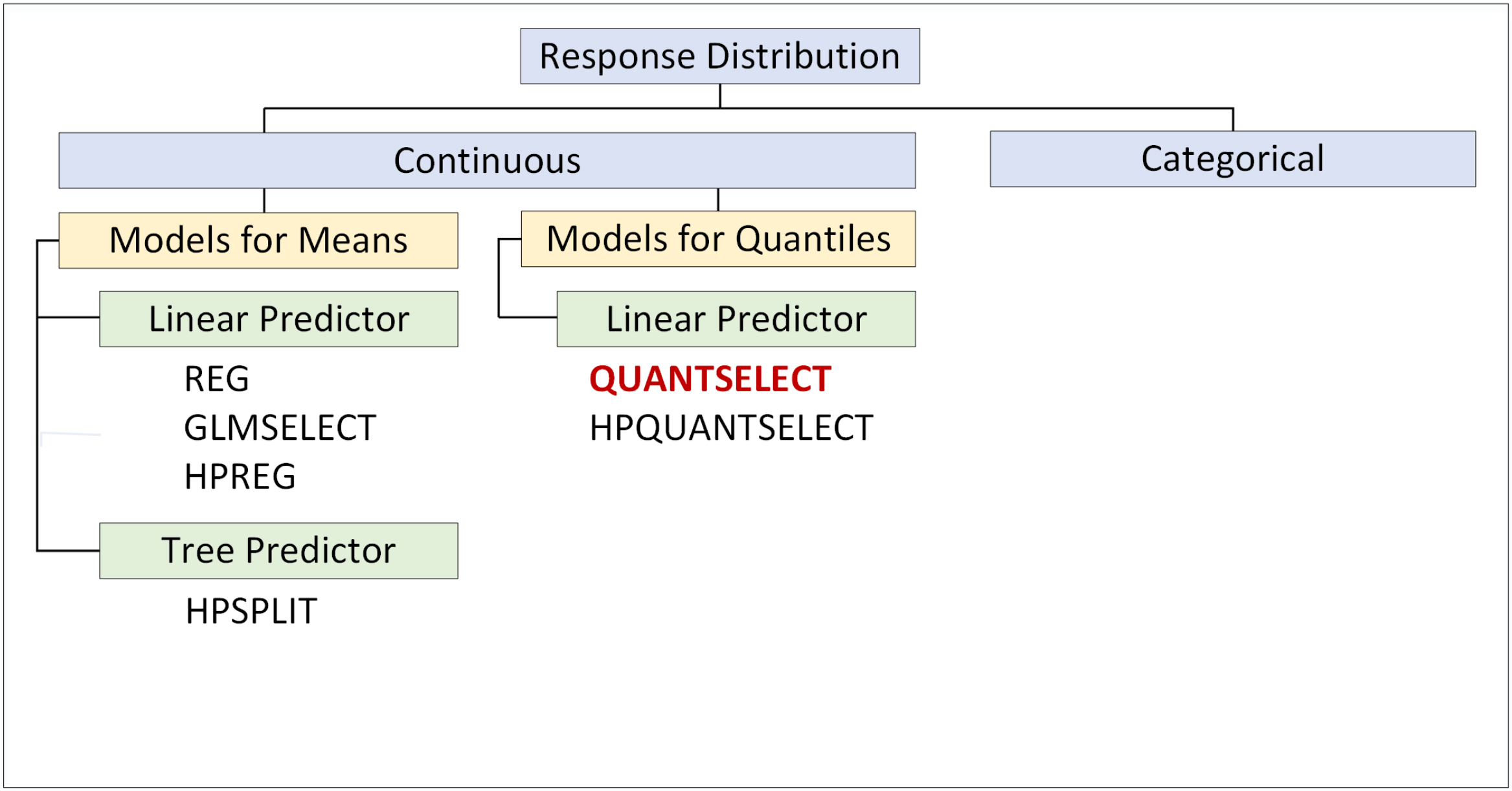


90th percentile

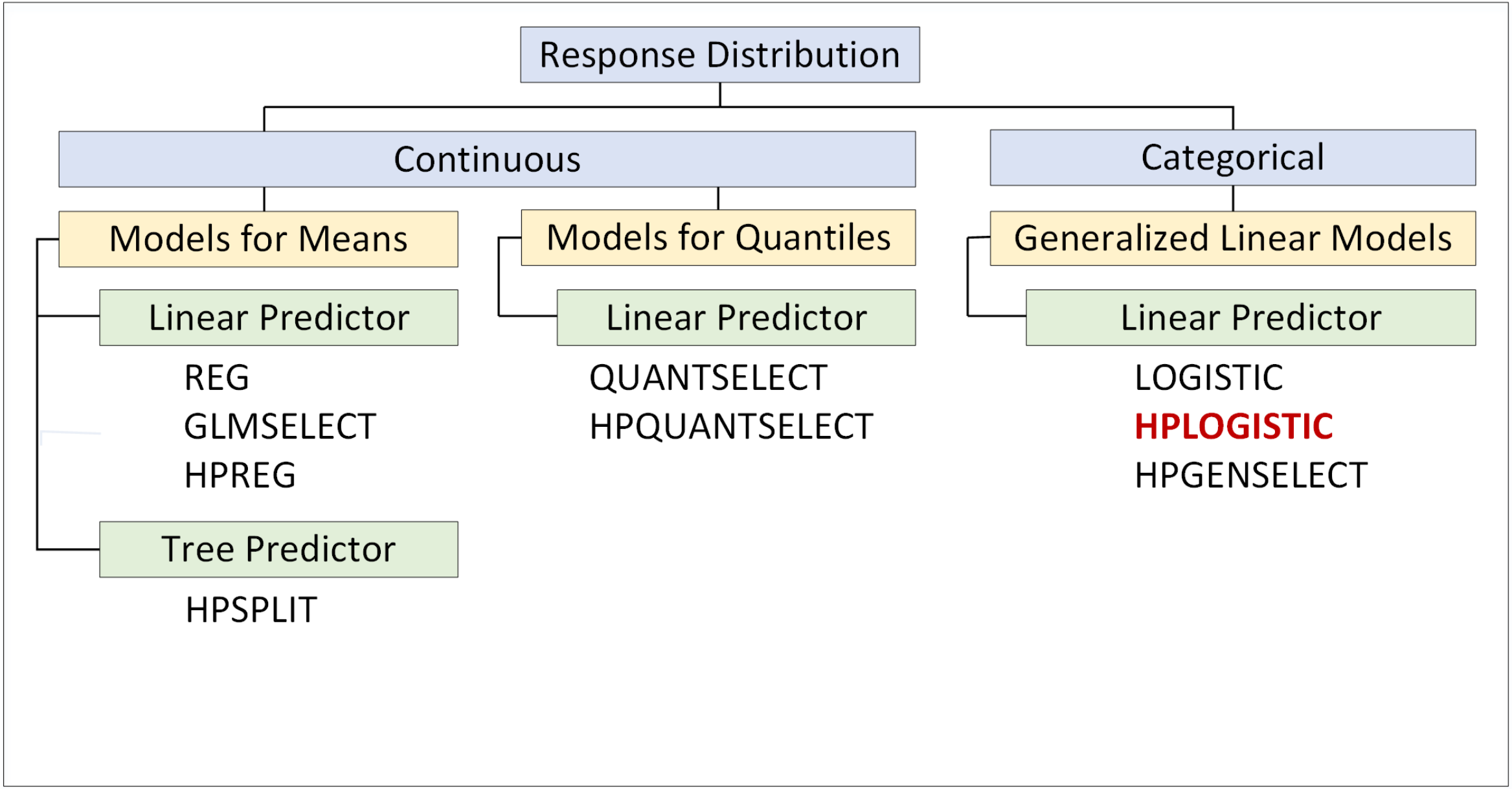
50th percentile

10th percentile

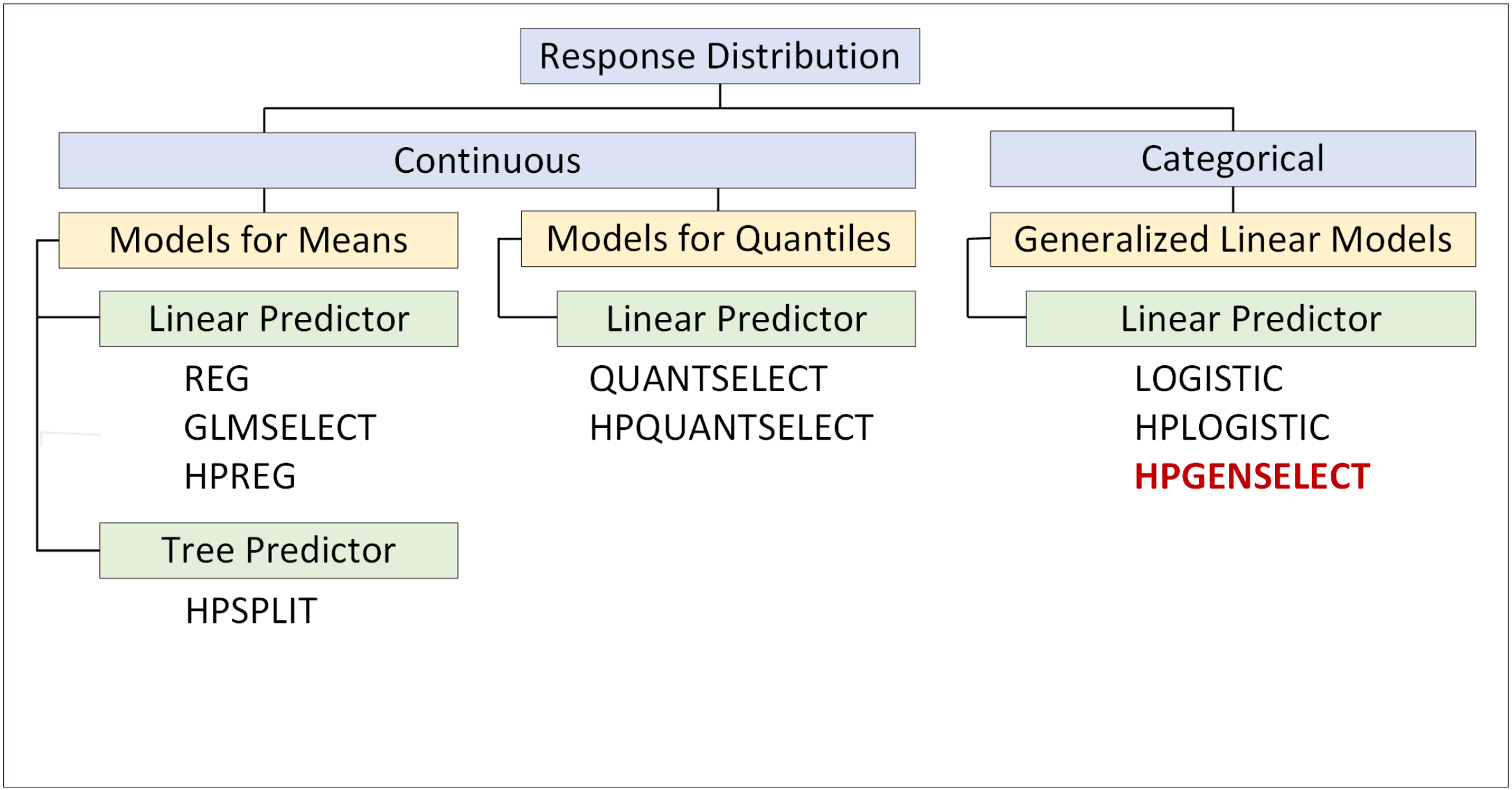
PROC QUANTSELECT builds quantile regression models



PROC HPLOGISTIC builds logistic regression models



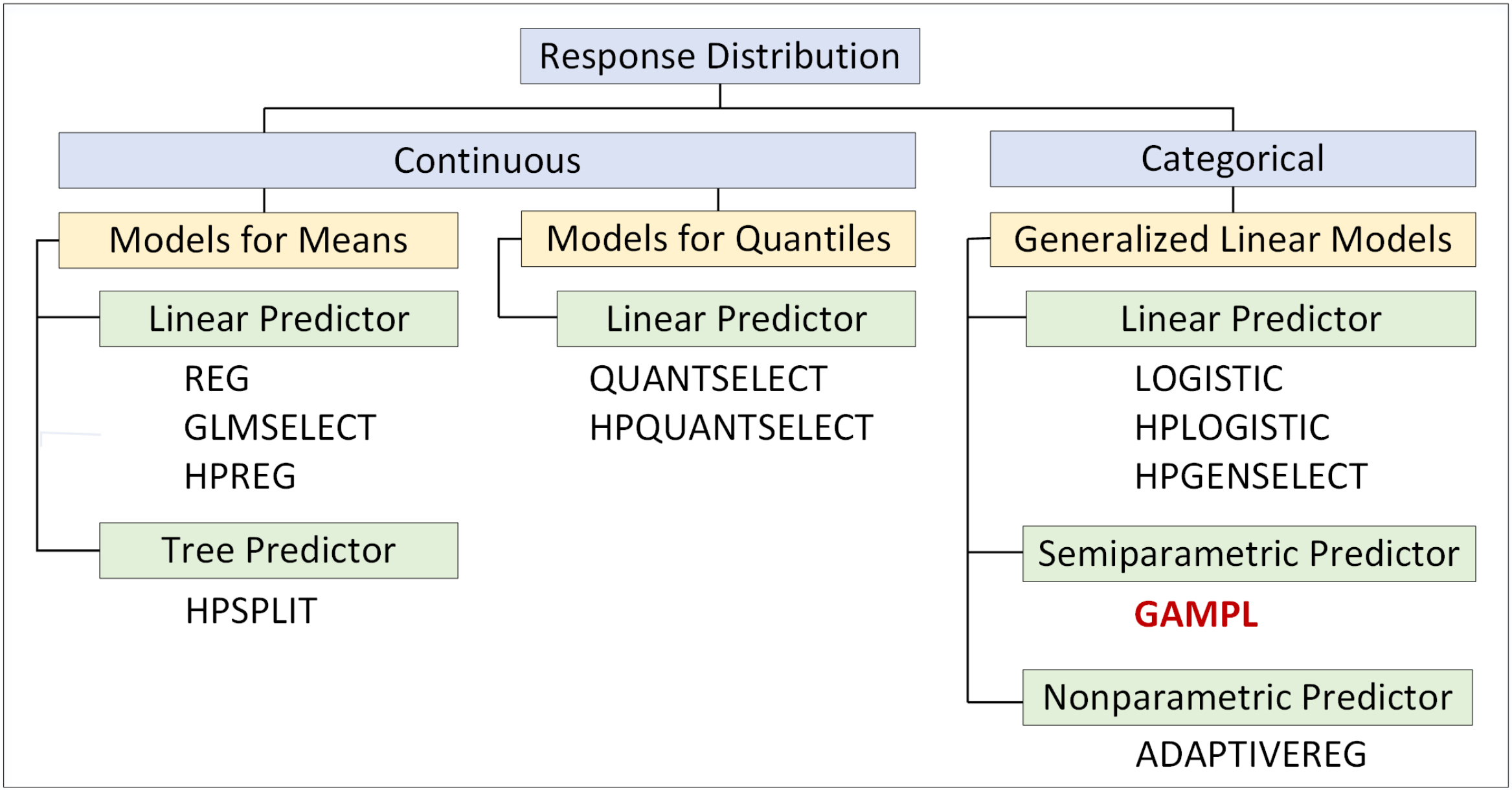
PROC HPGENSELECT builds generalized linear models



How does the HPGENSELECT procedure compare with the GENMOD procedure?

PROC HPGENSELECT	PROC GENMOD
Fits and builds models	Fits models
Large to massive data	Moderate to large data
Designed for predictive modeling	Designed for inferential analysis

The GAMPL procedure fits generalized additive models

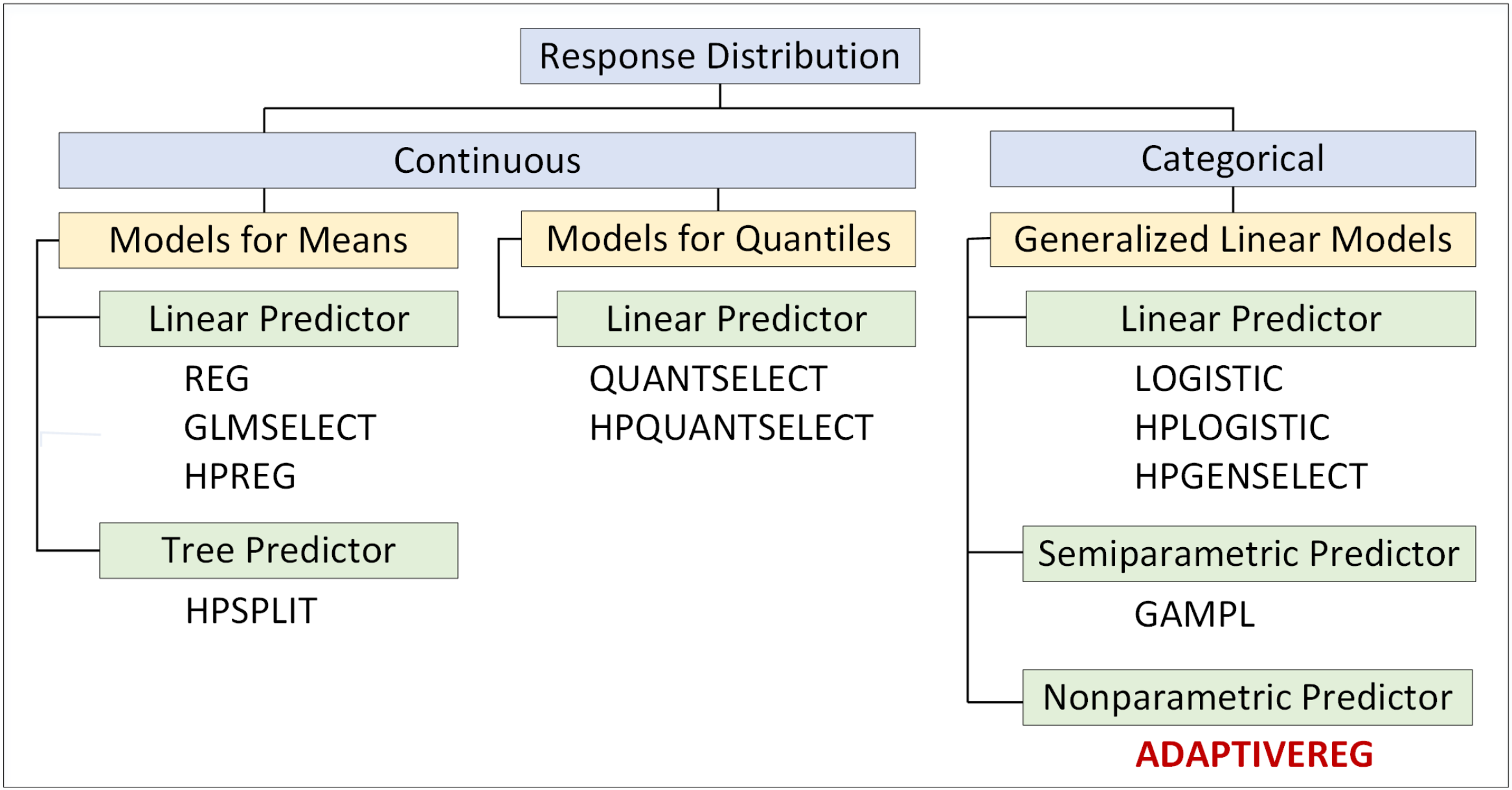


Generalized additive models provide greater flexibility for describing complex, unknown dependency relationships

Applications

- Analyzing claim rates for insured mortgages
- Environmental models with spatial effects
- Insurance ratemaking for geographic areas

The ADAPTIVEREG procedure fits multivariate adaptive regression splines



Inferential Analysis of Generalized Linear Models

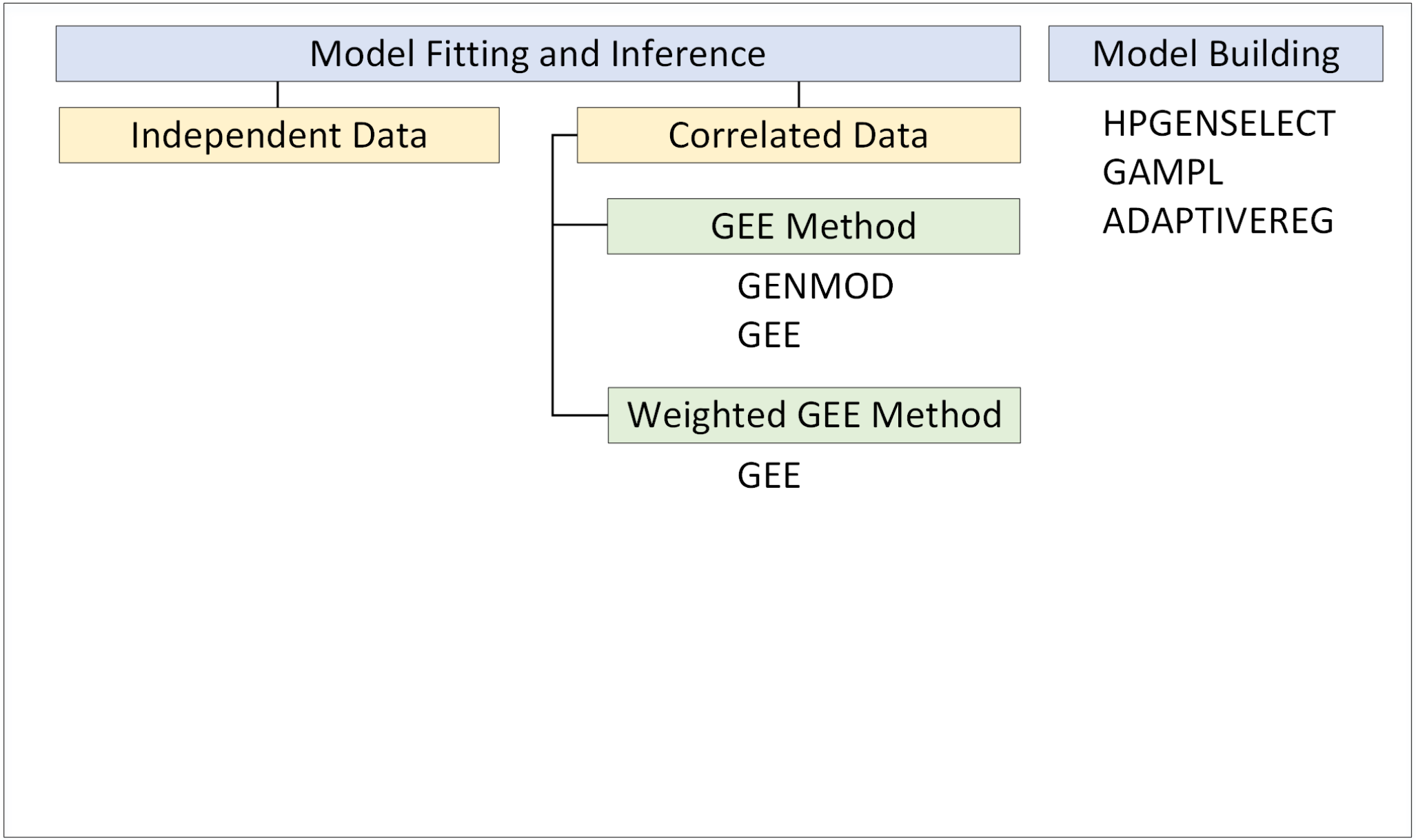
Tech Support is often asked, “I have longitudinal data with dropouts. Can PROC GENMOD do the right GEE analysis?”

Rob Agnelli and David Schlotzhauer, Analytical Technical Support

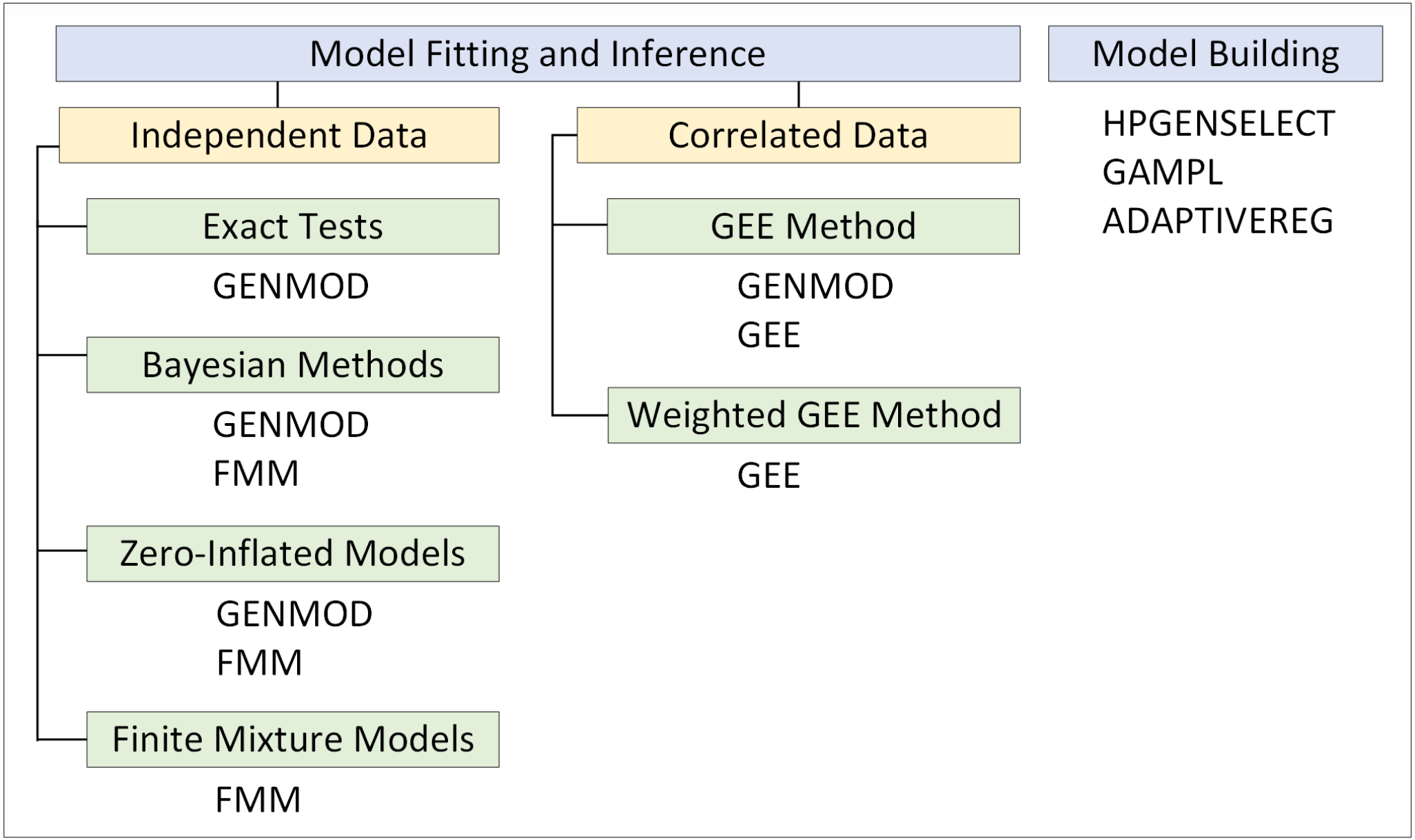
The new GEE procedure implements a weighted GEE method that accounts for dropouts that are missing at random (MAR)

	Standard GEE	Weighted GEE
Procedures	GENMOD and GEE	GEE
Specifications	Response model Correlation	Response model Correlation Missingness model
Inference assuming MCAR	Valid even if correlation is misspecified	Valid even if correlation is misspecified
Inference assuming MAR	Not generally valid	Valid even if correlation is misspecified

PROC GEE is just one new feature for analysis of generalized linear models



PROC GENMOD has been enhanced, and PROC FMM has been added



Survival Analysis

Tech Support is often asked, “Can I use PROC LIFETEST with time-to-event data that are interval-censored?”

Paul Savarese
Analytical Technical Support

Specialized methods of handling interval-censored data are available in the new ICLIFETEST and ICPHREG procedures

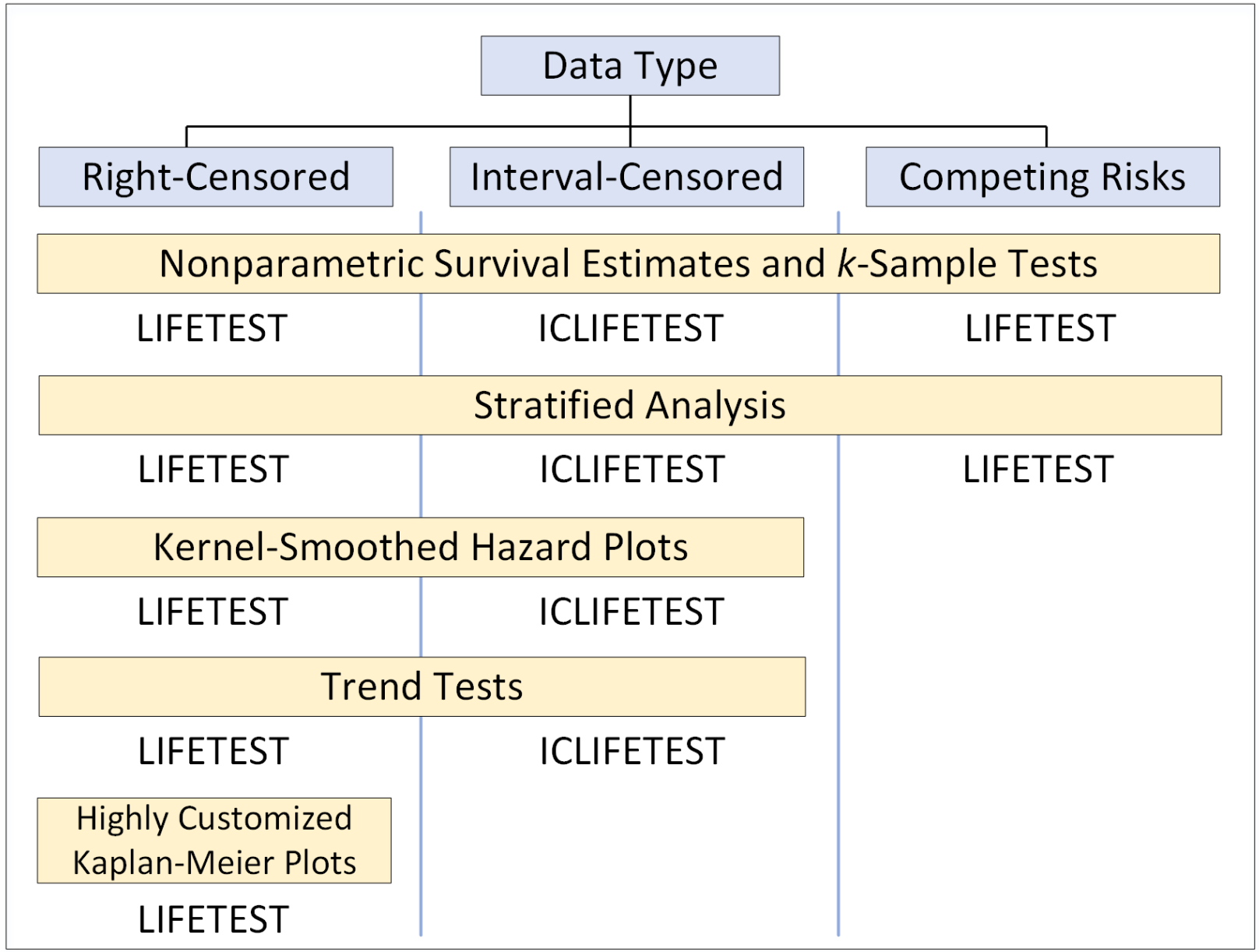
- PROC ICLIFETEST provides nonparametric methods of estimating survival functions and statistical testing
- PROC ICPHREG fits proportional hazards regression models

Imputing midpoints and using the LIFETEST and PHREG procedures is less efficient than applying specialized methods

There are now six procedures for analyzing time-to-event data, each with a different objective

Procedure	Focus	Approach	Modeling	Censoring
LIFETEST	Survival function	Nonparametric	No	Right
ICLIFETEST	Survival function	Nonparametric	No	Interval
LIFEREG	Lifetime	Parametric	Yes	Right, left, interval
PHREG	Hazard function	Semiparametric	Yes	Right
ICPHREG	Hazard function	Parametric	Yes	Interval
QUANTLIFE	Lifetime	Semiparametric	Yes	Right

Survival analysis capability for estimation and testing is growing



Specialized methods of analyzing competing risks are available in the LIFETEST and PHREG procedures

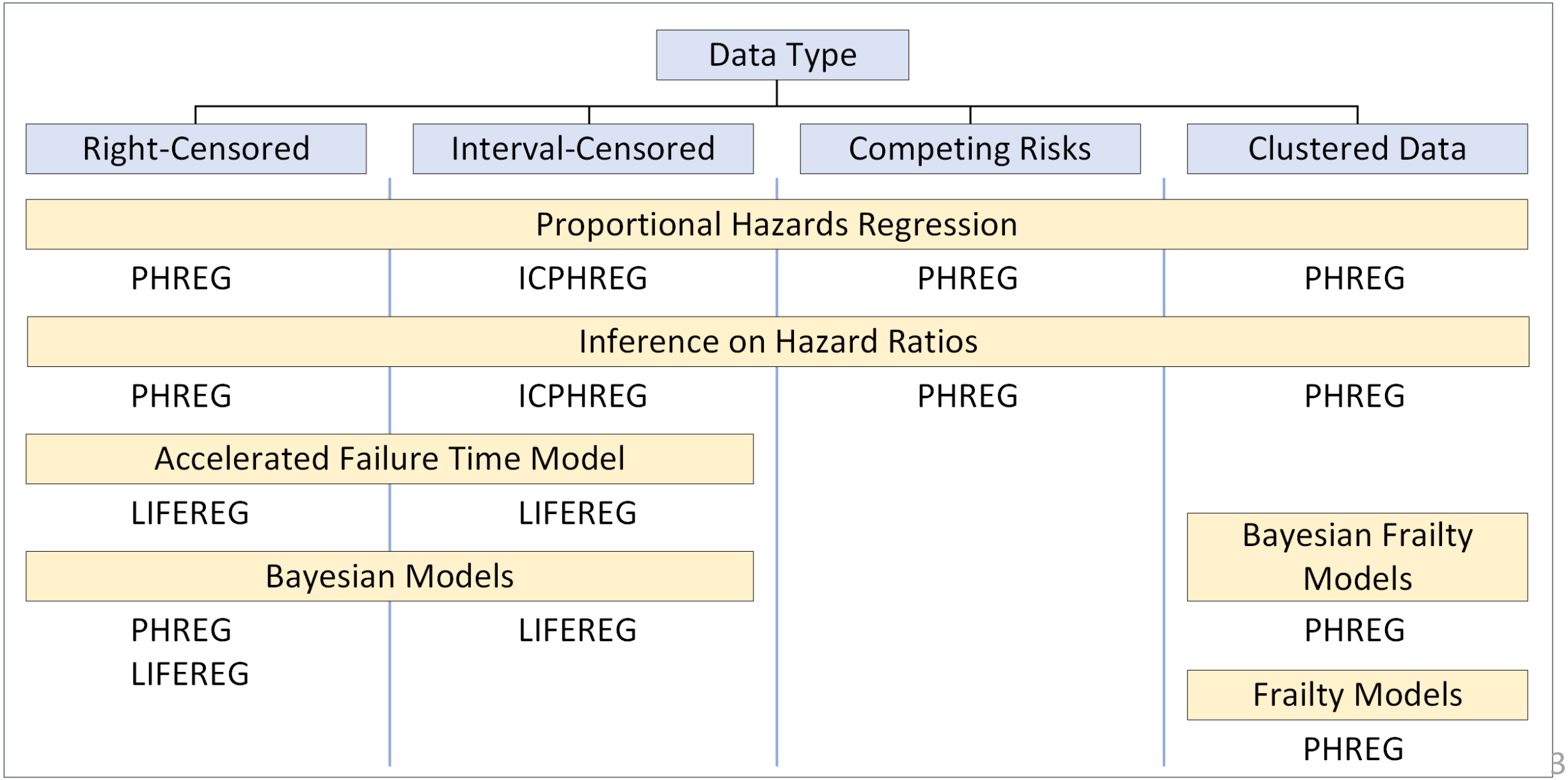
- The cumulative incidence function (CIF) replaces the survival function

  PROC LIFETEST estimates the CIF and provides Gray's test

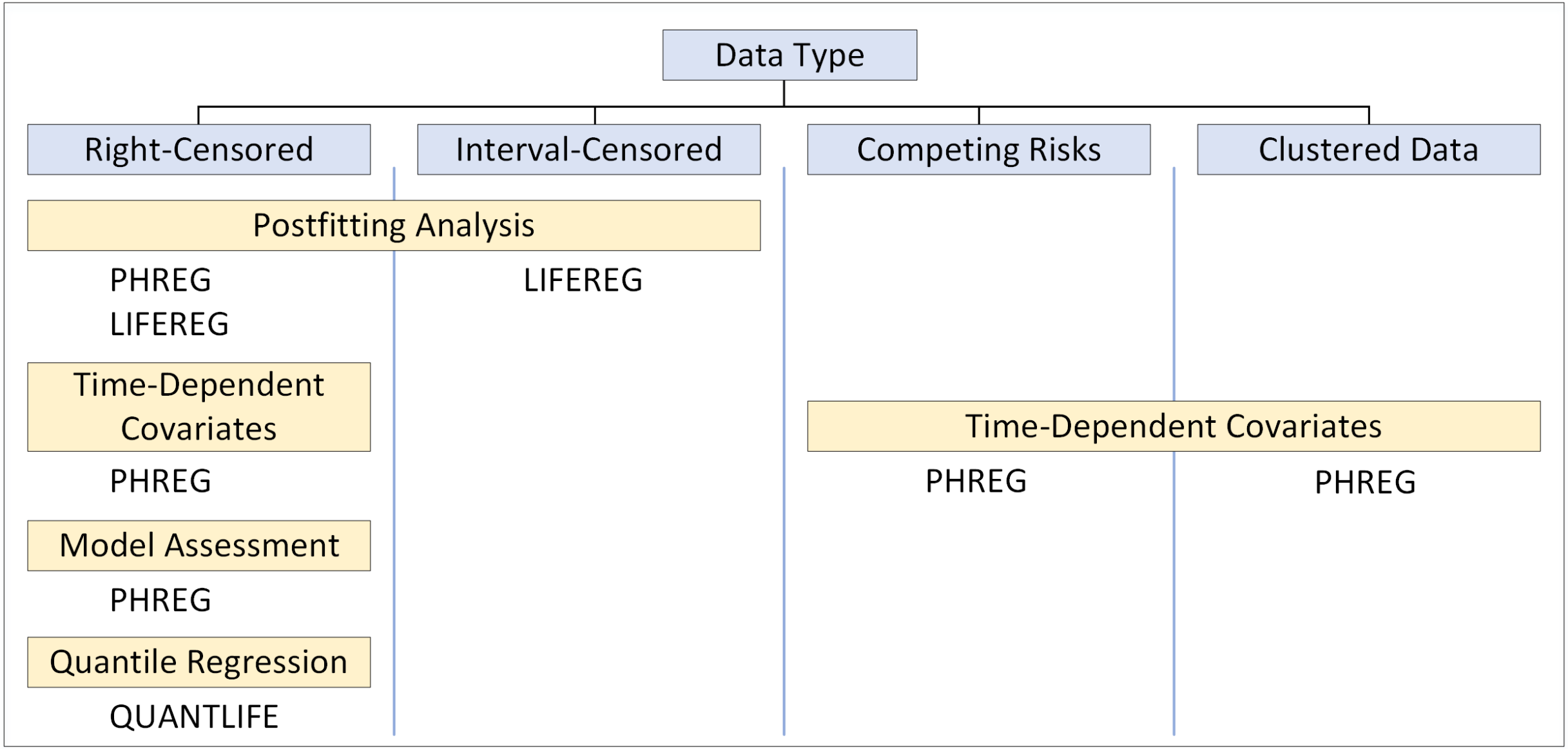
- The cause-specific hazard function (CSH) replaces the hazard function

  PROC PHREG implements the Fine and Gray model,
  which extends the Cox model to the CSH setting

Survival analysis capability for modeling is also growing



Survival analysis capability for modeling is also growing

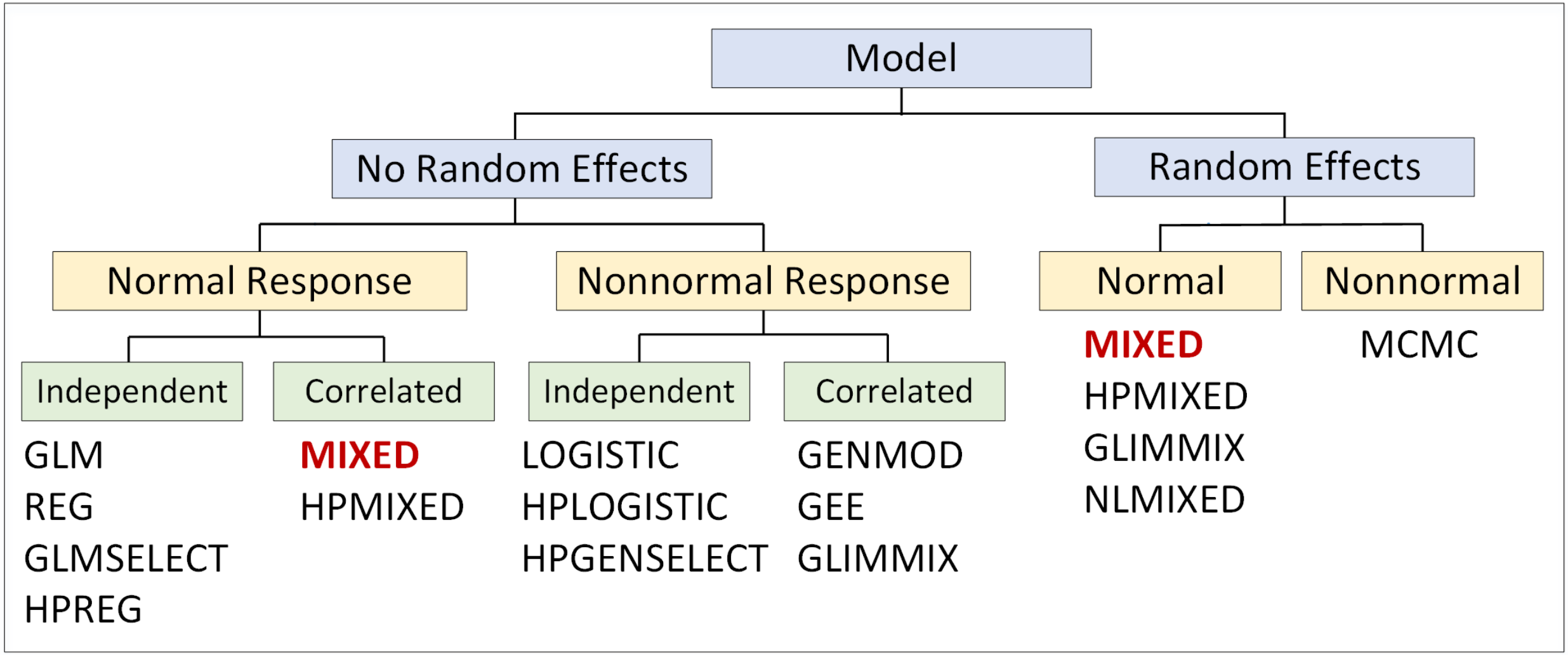


Mixed Models

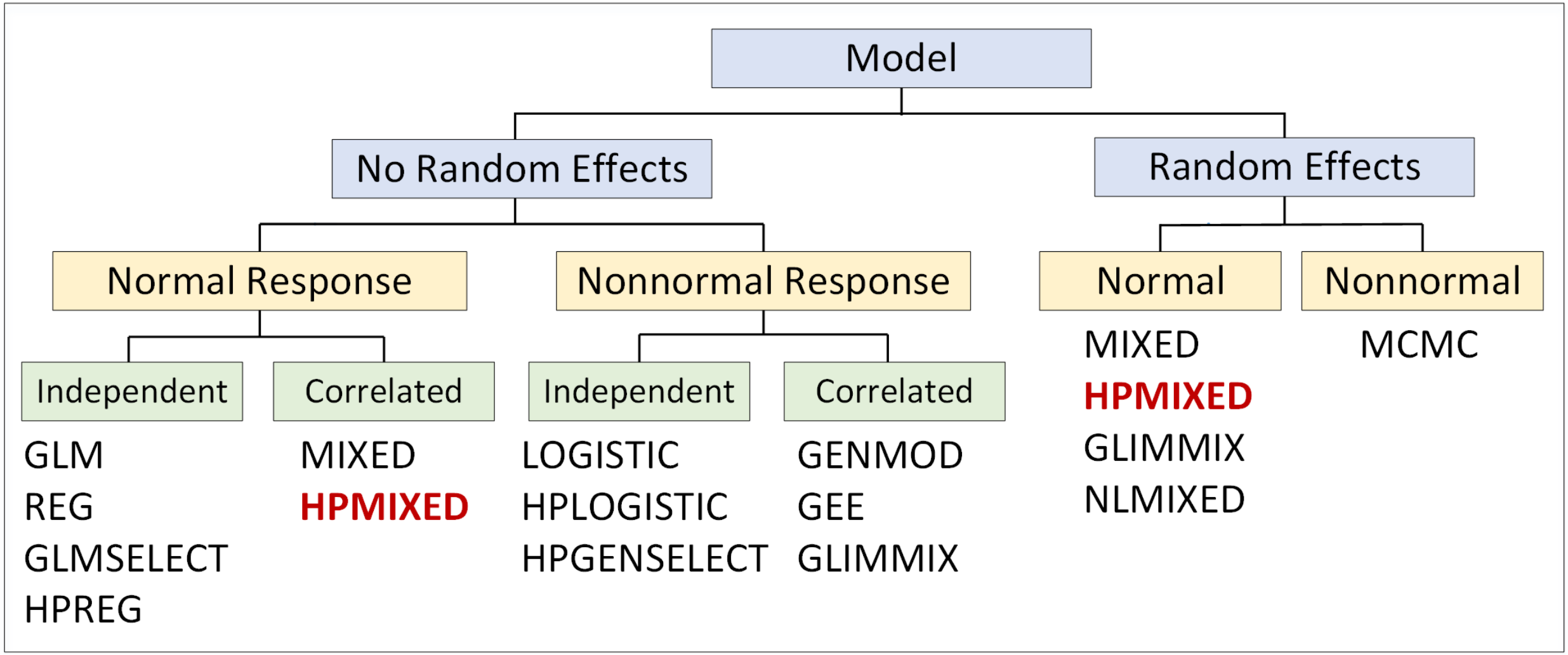
Tech Support is often asked,
“How do I decide which mixed model procedure to use?”

Jill Tao
Analytical Technical Support

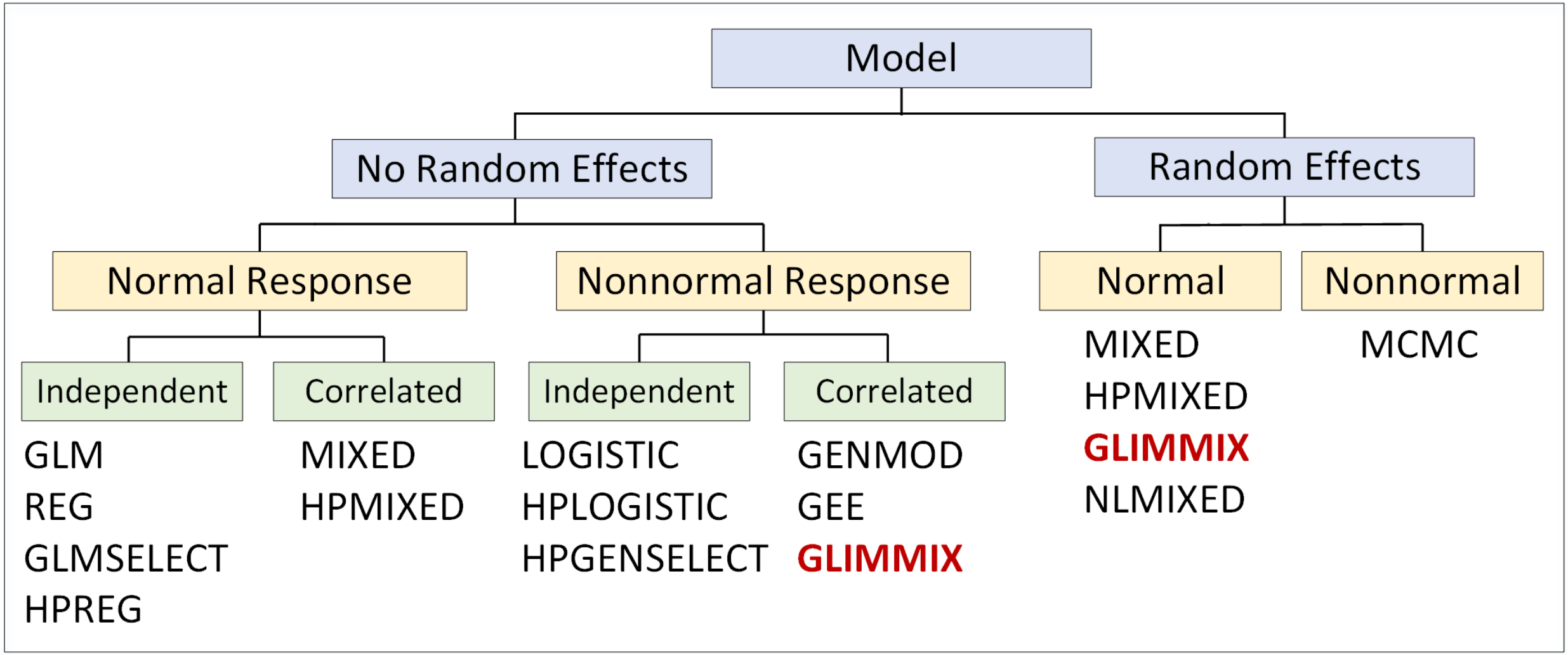
PROC MIXED is the flagship procedure for linear mixed models, providing generality for model estimation and postfit inference



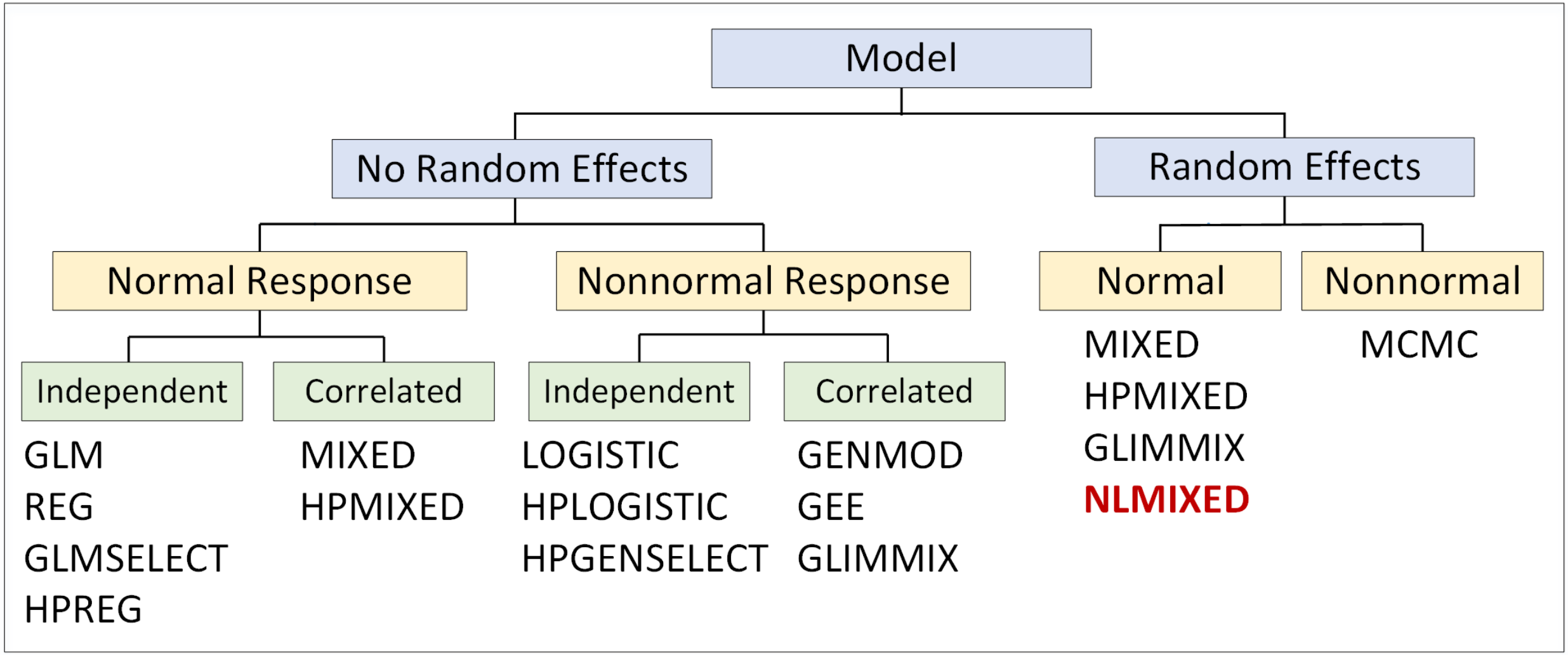
Use PROC HPMIXED when you need specialized computational methods for large, sparse mixed models



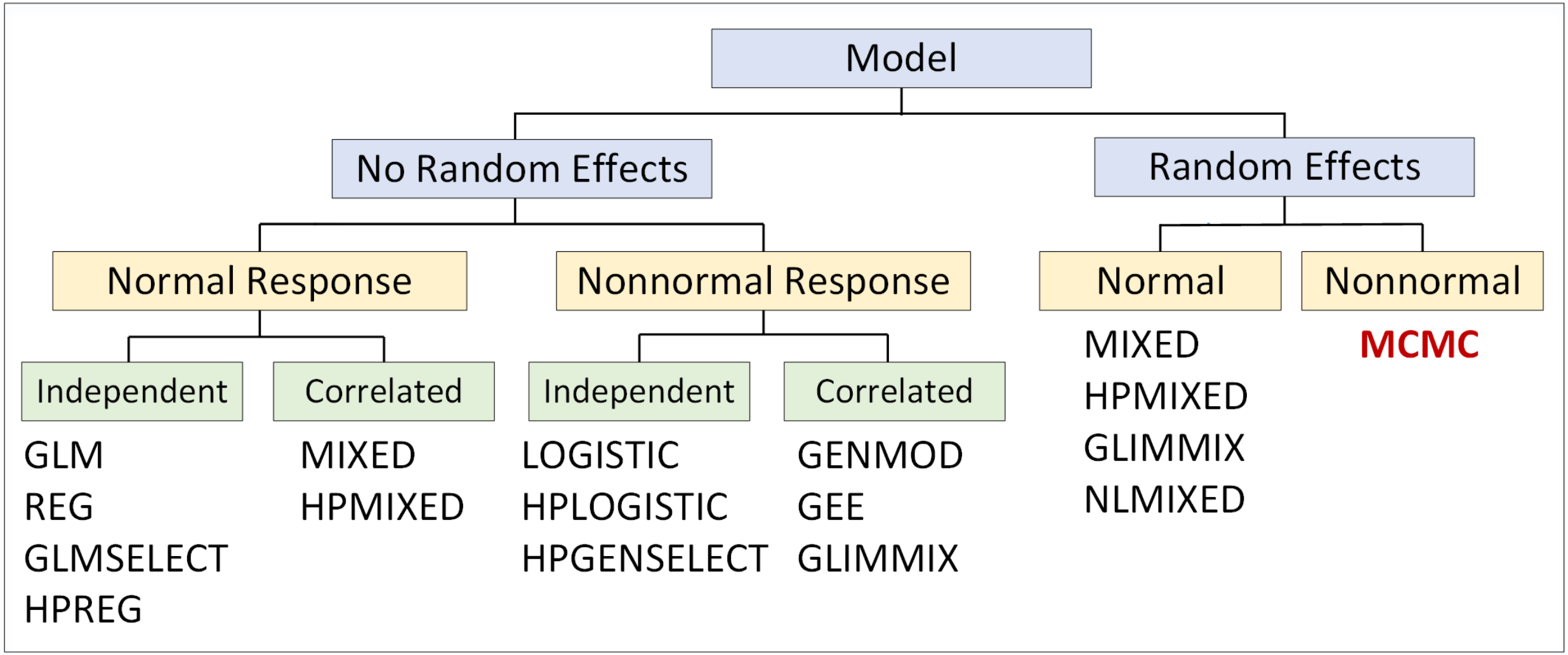
Use PROC GLIMMIX if your response has a nonnormal distribution that belongs to the exponential family



Use PROC NLMIXED to fit a random coefficients model in which the coefficients enter nonlinearly, or to fit PK/PD models … the list goes on



Use PROC MCMC for a wide range of Bayesian models and for models that the other procedures cannot handle



Summary

Our flyover has pointed out many new features—now, it's time to land and wrap up

Newer tools give you greater flexibility for regression modeling …

Benefit	Method	Procedures
Improved predictive ability and interpretability of regression models	Data partitioning	GLMSELECT, HPREG, HPSPLIT, QUANTSELECT, ADAPTIVEREG, HPLOGISTIC, HPGENSELECT
	Lasso methods and information criteria	GLMSELECT, QUANTSELECT, HPGENSELECT
Regression model building for a variety of response types and for complex dependence structures	Categorical responses	HPLOGISTIC, HPGENSELECT, GAMPL, ADAPTIVEREG
	Quantile regression	QUANTSELECT
	Regression trees	HPSPLIT
	Spline effects	GLMSELECT, GAMPL, ADAPTIVEREG

… specialized inference for complex data …

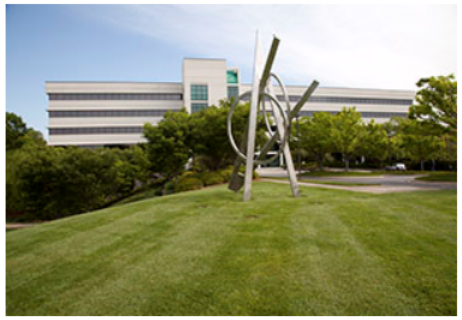
Benefit	Method	Procedures
Inference for special generalized linear models	Models for overdispersion	GENMOD, FMM
	Exact methods for small samples	GENMOD
	Weighted GEE methods for dropouts in longitudinal studies	GEE
Inference for special types of time-to-event data	Methods for interval-censored data	ICLIFETEST, ICPHREG
	Analysis of competing risks	LIFETEST, PHREG
	Analysis of heterogeneous data	QUANTLIFE

… versatile Bayesian methods, and high-performance computing

Benefit	Method	Procedures
Advantages of Bayesian methods, including model versatility and highly interpretable results	Generalized linear models	GENMOD
	Survival analysis models	LIFEREG, PHREG, MCMC
	Finite mixture models	FMM
	Mixed models	MCMC
	General Bayesian models	MCMC
High-performance computing for large data	Regression model building	HPREG, HPLOGISTIC, HPQUANTSELECT, HPGENSELECT, HPSPLIT
	Generalized additive models	GAMPL
	Regression trees	HPSPLIT
	Large, sparse mixed models	HPMIXED

Learn more at http://support.sas.com/statistics

Welcome to Statistics and Operations Research



SAS has long developed software for data analysis, econometrics, operations research, and quality improvement. The purpose of these pages is to provide our users with technical information about using this software, including details about software capabilities, examples, papers, e-newsletter, and communities.

SAS/ETS | SAS/IML | SAS/OR | SAS/QC | SAS/STAT

SAS/ETS software offers a broad array of time series, forecasting, and econometric techniques that enable modeling, forecasting, and simulation of business processes for improved strategic and tactical planning.

SAS/ETS 14.2 introduces the new SASENOAA and SASERAIN interface engines as well as the new SPATIALREG procedure along with new features for the HPCDM, HPSEVERITY, SEVERITY, QLIM, SSM, TIMESERIES, and VARMAX procedures and the SASEFAME, SASEFRED, SASEQUAN, and SASEXFSD interface engines.

Read about What's New in SAS/ETS 14.2.

Featured News

SAS Global Forum 2017

If you are still debating whether to attend this year, consider what you might miss:

- Statistical Tutorials on Sunday
- Over a dozen presentations by SAS staff on new capabilities
- Numerous emerging technology presentations

Upcoing Conferences

2017 American Statistical Association Conference on Statistical Practice
Feb. 23–25, 2017
Jacksonville, FL

ENAR 2017
Mar. 12–15, 2017
Washington, DC

SAS Global Forum 2017
Apr. 2–5, 2017
Orlando, FL

HIGHLIGHTS

- Latest e-Newsletter
- 14.2 Highlights and Information
- SAS Analytical Handouts
- SAS Macros for Experimental Design and Choice Modeling
- ODS Statistical Graphics

VIDEOS

- Computing an Optimal Blackjack Strategy with SAS/OR New
- SAS Simulation Studio: Power Up by Calling SAS Programs New
- What's New in Econometric Modeling in SAS/ETS 14.1
- Spatial Dependence, Nonlinear Panel Models, and More New Features in SAS/ETS 14.1
- Exploring Econometric Tasks in SAS Studio
- Statistical Quality Improvement with SAS/QC
- See all videos

PAPERS

- SAS/STAT 14.1: Methods for Massive, Missing, or Multifaceted Data
- Practical Applications of SAS Simulation Studio

Sign up for e-newsletter

Watch short videos

Download overview papers

Step Up Your Statistical Practice with Today's SAS/STAT® Software

Phil Gibbs

SAS Institute