

Popular SAS/STAT Procedures

Marc Huber, Danny Modlin

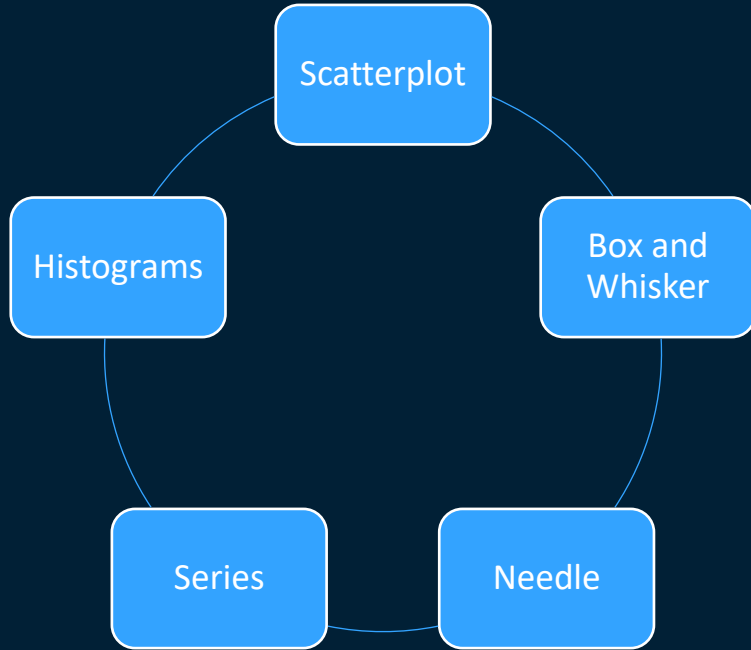
Sr. Analytical Training Consultants

Presentation Overview

- Data Exploration
 - PROC SGPLOT
 - PROC MEANS
 - PROC UNIVARIATE
 - PROC CORR
- Binary Response
 - PROC LOGISTIC
 - PROC GENMOD
- Continuous Response Analysis
 - PROC REG
 - PROC GLM
 - PROC GLMSELECT
- Post-Processing
 - PROC PLM

Data Exploration

PROC SGPLOT



- Used to make your own graphics
- Some graphs are automatically made in ODS

Data Exploration

PROC MEANS

- Commonly used to explore summary statistics.
- You control what statistics you want to see.
- By-group processing is allowed using a CLASS statement

N

Mean

Min

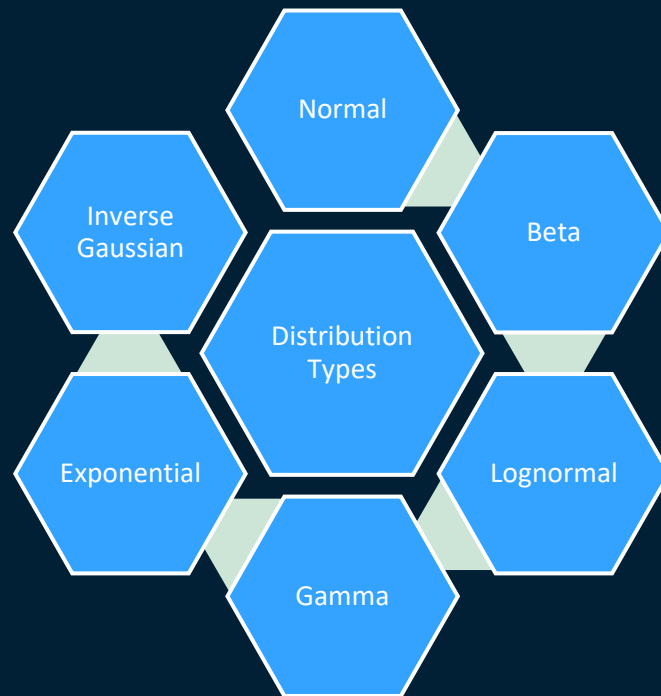
Max

Std Dev

Data Exploration

PROC UNIVARIATE

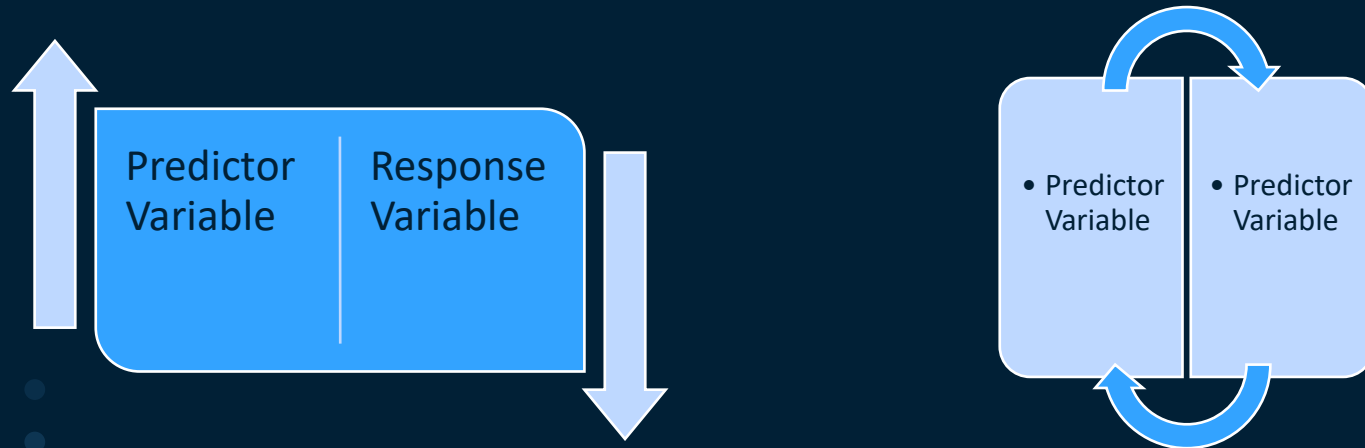
- Gives summary statistics and more
- Tests for Distribution type



Data Exploration

PROC CORR

- Determines strength and significance of linear relationships
- Helpful for variable selection and early collinearity detection






Demonstration Time

When do I use what?

SGPLOT – creating graphics on your own for exploration



MEANS – summary statistics in table format



UNIVARIATE – summary statistics along with distribution questions



CORR – checks relationships among variables for different purposes

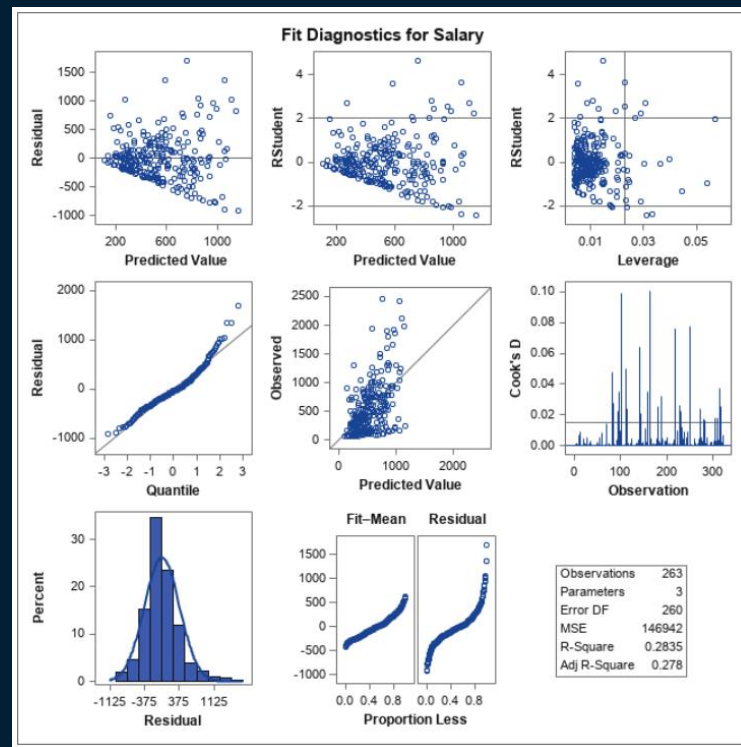
Continuous Response Analysis

PROC REG

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	15114214	7557107	51.43	<.0001
Error	260	38204899	146942		
Corrected Total	262	53319113			

Root MSE	383.33004	R-Square	0.2835
Dependent Mean	535.92588	Adj R-Sq	0.2780
Coeff Var	71.52669		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	3.70076	58.77334	0.06	0.9498
nRBI	RBI in 1986	1	6.39650	1.44511	4.43	<.0001
nRuns	Runs in 1986	1	3.56383	1.48196	2.40	0.0169



Continuous Response Analysis

PROC GLM

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	15442468.35	5147489.45	35.20	<.0001
Error	259	37876644.44	146241.87		
Corrected Total	262	53319112.79			

R-Square	Coeff Var	Root MSE	Salary Mean
0.289624	71.35610	382.4158	535.9259

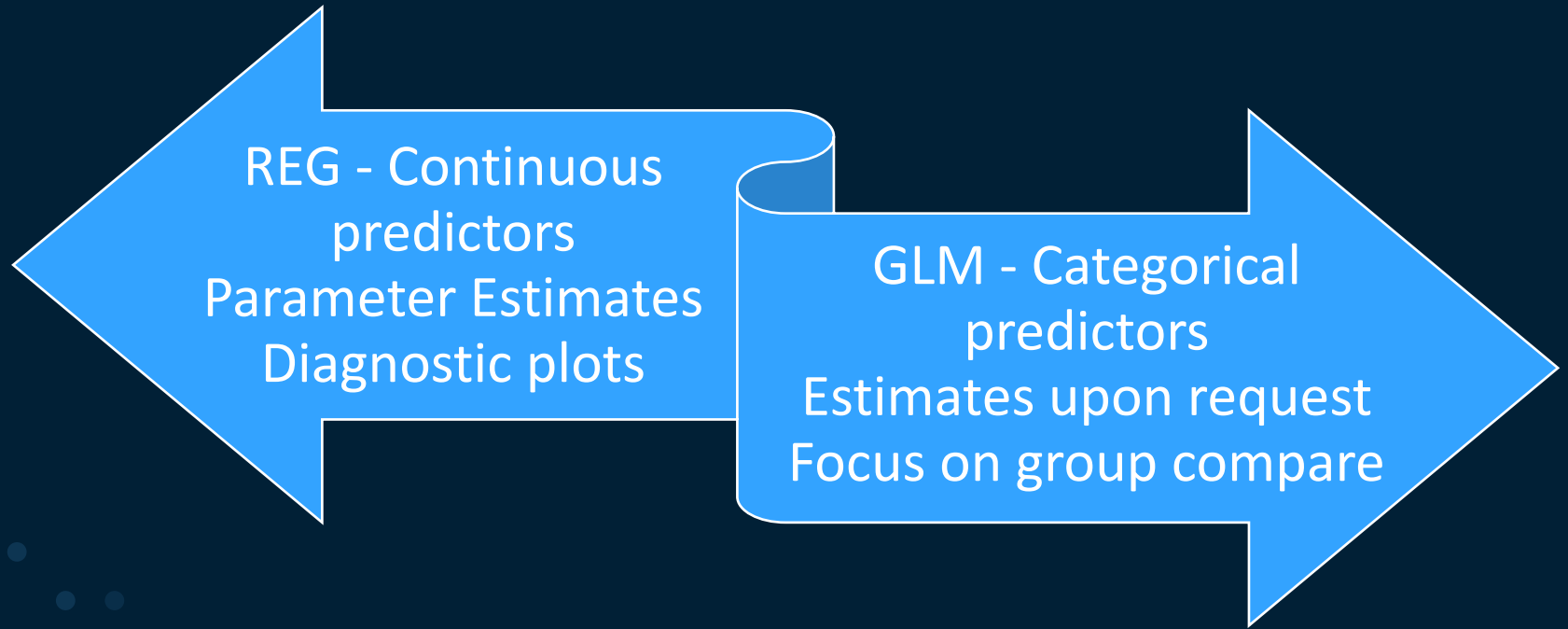
Source	DF	Type I SS	Mean Square	F Value	Pr > F
nRBI	1	14264437.26	14264437.26	97.54	<.0001
nRuns	1	849776.47	849776.47	5.81	0.0166
League	1	328254.62	328254.62	2.24	0.1353

Source	DF	Type III SS	Mean Square	F Value	Pr > F
nRBI	1	2928542.644	2928542.644	20.03	<.0001
nRuns	1	944690.724	944690.724	6.46	0.0116
League	1	328254.620	328254.620	2.24	0.1353

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	27.04586129	B	60.66837096	0.45	0.6561
nRBI	6.45365552		1.44216805	4.47	<.0001
nRuns	3.77455569		1.48510403	2.54	0.0116
League American	-71.95029400	B	48.02451808	-1.50	0.1353
League National	0.00000000	B	.	.	.

Continuous Response Analysis

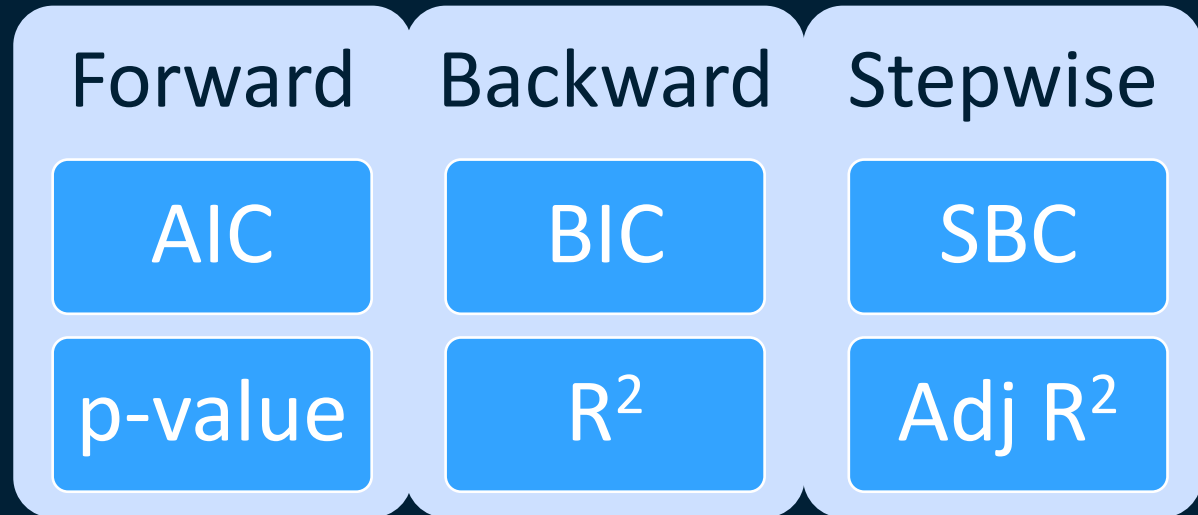
Difference between REG and GLM



Continuous Response Analysis

PROC GLMSELECT

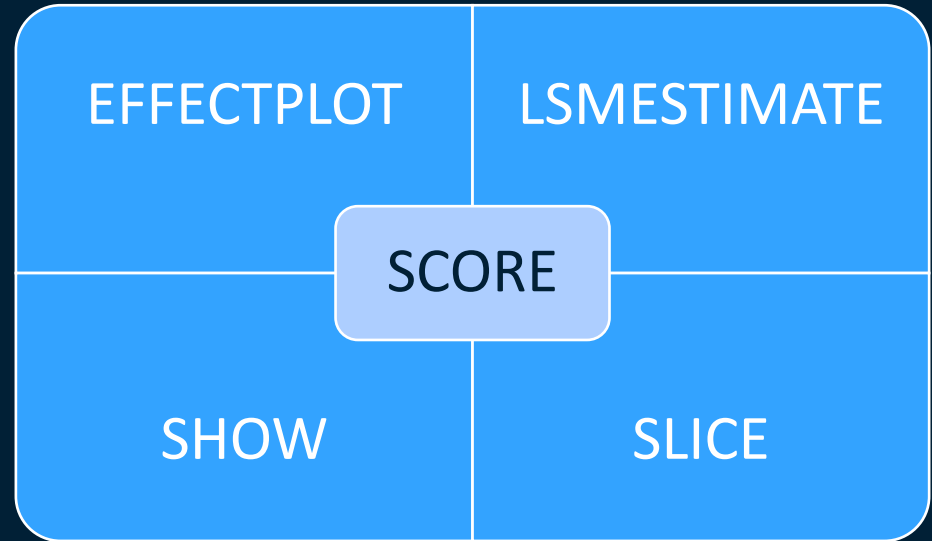
- Adds variable selection to GLM
- Goes beyond selection methods in REG



Continuous Response Analysis

PROC PLM

- Post-fitting for General Linear Models
- Requires use of a STORE statement during analysis
- Can be used without the need of original dataset





Demonstration Time

When do I use what?


REG – continuous predictor, diagnostic plots



GLM – categorical predictor, ANOVA, ANCOVA

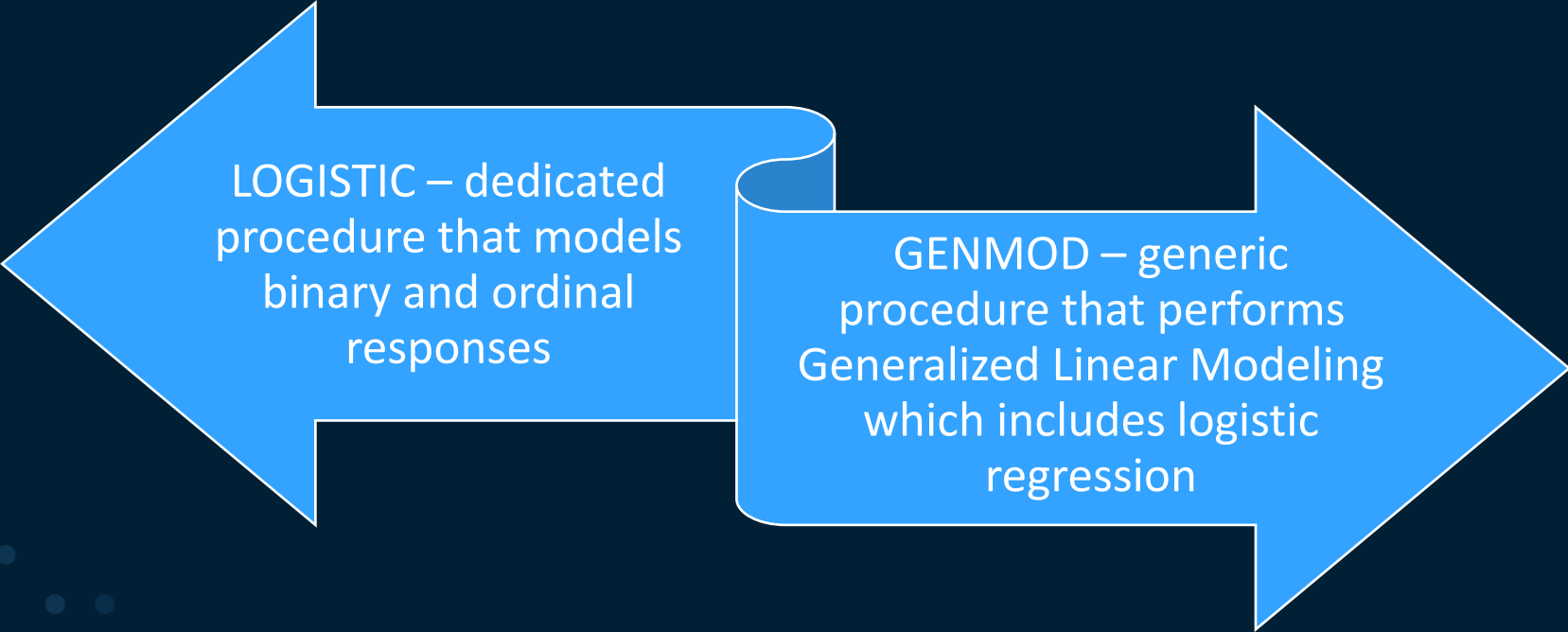


GLMSELECT – GLM plus additional variable selection



PLM – post analysis after model created

Binary Response Analysis



LOGISTIC – dedicated procedure that models binary and ordinal responses


GENMOD – generic procedure that performs Generalized Linear Modeling which includes logistic regression




Demonstration Time

When do I use what?


LOGISTIC – binary or ordinal response, predictors can be categorical or continuous



-- procedure is most prominently used for binary outcomes with output automated for typical questions



GENMOD – a more general procedure for generalized linear modeling which does include logistic regression



-- the output is not as full as that of logistic but with additional options and statements you can request

Common (Popular) PROCs for Other Analyses

Bayesian Analyses

PROC MCMC	PROC PHREG
PROC BGLIMM	PROC LIFEREG
PROC GENMOD	PROC COUNTREG

Survey Analyses

PROC SURVEYREG	PROC SURVEYFREQ	PROC SURVEYSELECT
PROC SURVEYLOGISTIC	PROC SURVEYIMPUTE	
PROC SURVEYMEANS	PROC SURVEYPHREG	

Survival Analyses

PROC ICLIFETEST	PROC LIFETEST
PROC ICPHREG	PROC PHREG
PROC LIFEREG	PROC SURVEYPHREG

Common (Popular) PROCs for Other Analyses

Cluster Analyses

PROC ACECLUS	PROC FASTCLUS	PROC VARCLUS
PROC CLUSTER	PROC MODECLUS	
PROC DISTANCE	PROC TREE	

Multivariate Analyses

PROC CANCELL	PROC FACTOR	PROC PRINQUAL
PROC CORR	PROC MDS	
PROC CORRESP	PROC PRINCOMP	

Mixed Model Analyses

PROC GLIMMIX	PROC PHREG
PROC NLMIXED	PROC VARCOMP
PROC MIXED	



<http://support.sas.com>

[SAS/STAT SAS/STAT Procedures A-Z](#)



Questions?