WBS Herbst Seminar
Statistische Software für Biometrische Auswertungen

Analytisches Datenmanagement und biometrische Auswertungen mit SAS 9.3

Dr. Gerhard Svolba
Wien, 27. November 2012

sas THE POWER TO KNOW.

Überblick über diesen Vortrag

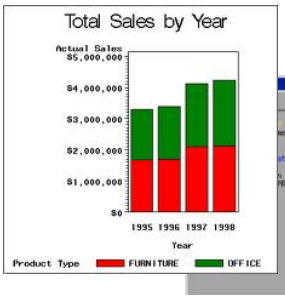
- „Historische Vorurteile“ als Einstieg
- Datenmanagement für biometrische Auswertungen
- Graphische Darstellungen mit SAS ODS Graphics
- Ausgewählte statistische Auswertungen
- Abschliessende Bemerkungen

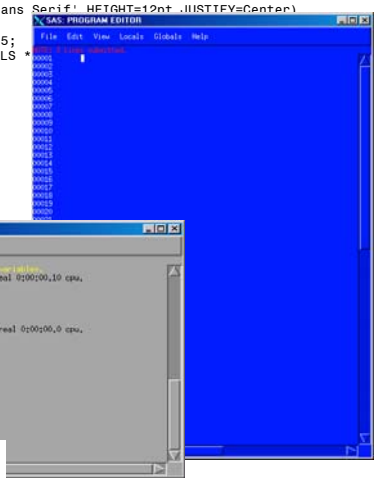
sas THE POWER TO KNOW.

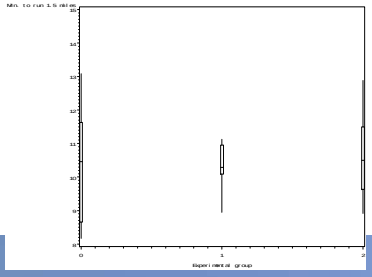
Historische Vorurteile

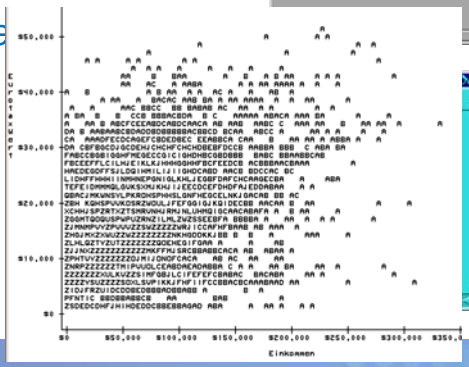
- *SAS Auswertungen können nur über den komplizierten SAS-Syntax erstellt werden.*
- *SAS Graphiken sind mühevoll zu erstellen – deren Aussehen ist bescheiden*
- *SAS Ergebnise können nur mühevoll in andere Programme übernommen werden*

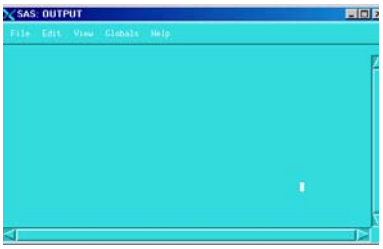
```
GOPTIONS xpixels=&_EGCHARTWIDTH ypixels=&_EGCHARTHEIGHT;
                    SYMBOL1 INTERPOL=BOXTJF          CV=CX3366FF
    CO=BLACK        VALUE=NONE       HEIGHT=1 ;
Axis1               STYLE=1          WIDTH=1         MINOR=NONE
                    LABEL=(FONT='Microsoft Sans Serif' HEIGHT=12pt
JUSTIFY=Right)
;
Axis2               STYLE=1          WIDTH=1         MINOR=NONE
    LABEL=(FONT='Microsoft Sans Serif' HEIGHT=12pt JUSTIFY=Center)
;
PROC GPLOT DATA=WORK.SORT2805;
                    PLOT LAUFPULS *
                    VAXIS=AXIS1
                    HAXIS=AXIS2
                    FRAME
;
RUN; QUIT;
```











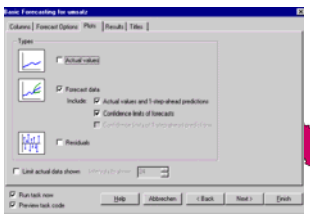
Evolutions-Schritte der Benutzer-Interfaces in der 35-jährigen SAS-Geschichte

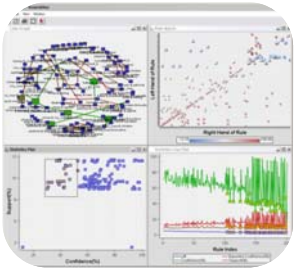
1999/2000

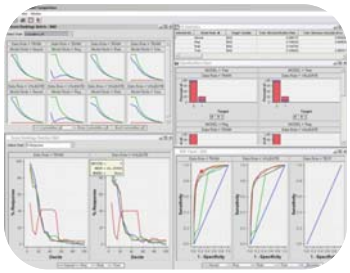
SAS 8, Enterprise Guide, Output Delivery System, HTML, PDF, Java, Active X

2004

SAS 9.1 Java Front Ends, Interactive Graphiken



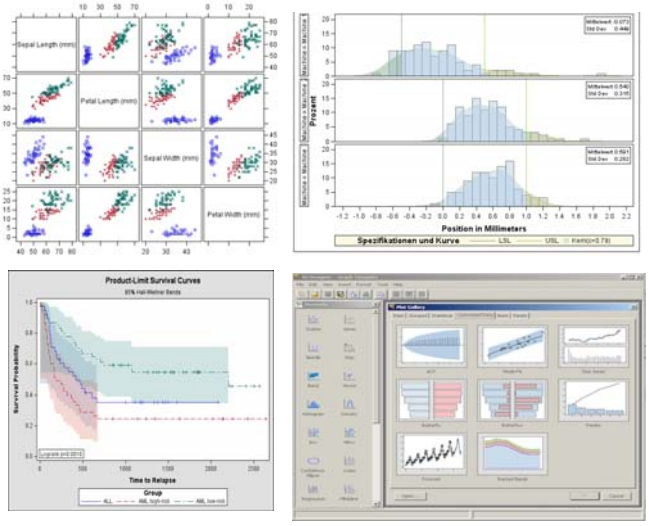




Evolutions-Schritte der Benutzer-Interfaces in der 35-jährigen SAS-Geschichte (Forts.)

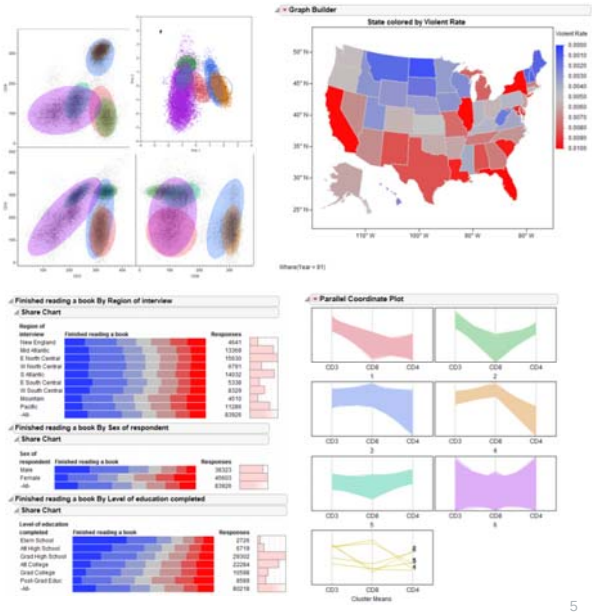
2008

SAS 9.2 Statistical Graphics



2011

SAS 9.3 und JMP-Integration



sas THE POWER TO KNOW.

Datenmanagement für biometrische Auswertungen

sas THE POWER TO KNOW.

Sehen Ihre Analyse Daten manchmal so aus?

PNR	Age	Sex	VisitDate	Visit_Code	TRIGL	CHOL
15	66	Female	14.02.1998	B0	131	276
			21.03.1998	T1	262	230
			25.04.1998	T2	339	231
			23.05.1998	T3	305	230
			22.08.1998	T6	200	190
			28.11.1998	T9	209	191
			06.03.1999	T12	192	220
17	69	Male	05.09.1997	B0	262	251
			05.10.1997	T1	153	230
			02.11.1997	T2	228	234
			30.11.1997	T3	222	231
			22.02.1998	T6	212	233
			17.05.1998	T9	224	232
			09.08.1998	T12	124	240
			01.11.1998	T15	227	220
			24.01.1999	T18	264	248
			18.04.1999	T21	169	216
			19.09.1999	T24	251	239
19	47	Female	31.07.1998	B0	159	236
			31.08.1998	T1	153	223
			28.09.1998	T2	220	222
			26.10.1998	T3	202	239

Kopieren

- Differenz zur letzten Messung
- Differenz zur ersten Messung
- Nummerierung der Messung

Wieviele Missing Values sehen Sie?

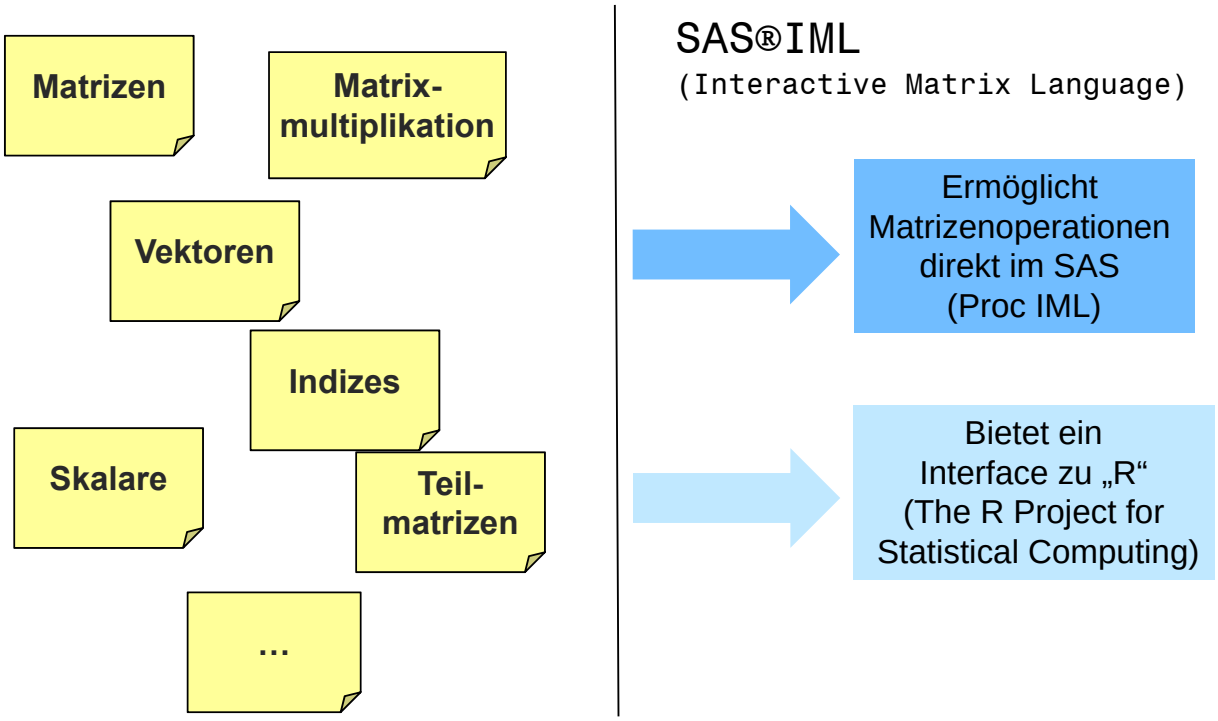
PNR	date	amount
56	2004-02-01	48
56	2004-03-01	51
56	2004-04-01	42
56	2004-05-01	36
56	2004-06-01	6
56	2004-07-01	.
56	2004-08-01	48
56	2004-09-01	36
56	2004-10-01	66
56	2004-11-01	15
56	2004-12-01	33
58	2005-06-01	39
58	2005-07-01	63
58	2005-08-01	84
58	2005-09-01	18
58	2005-12-01	69
58	2006-03-01	0
58	2006-07-01	90
58	2006-10-01	57
58	2007-01-01	48

Record vorhanden, Wert fehlt

?

Fehlende Records
Kontinuität nicht gegeben

Behandlung von Datenobjekten als Matrizen



Aufruf von R-Funktionalität
Austausch von Datenobjekten mit R

```
run ExportMatrixToR (amount,"amount_R");
*** Aufruf von R-Funktionalität;
submit / R;
 Smoothed_Amount <- smooth( amount_R )
 Smoothed_Amount <- as.matrix(Smoothed_Amount)

endsubmit;
** Import von R;
run ImportMatrixFromR( RSmoothed_Amount,
"Smoothed_Amount" );
```

Beispiele für IML Operationen und Funktionen

Operator	Description
` (accent grave)	Transpose (postfix)
- *(prefix)*	Negative prefix
[ ]	Subscript
**	Matrix exponentiation
##	Element-wise exponentiation
*	Matrix multiplication
#	Element-wise multiplication
/	Element-wise division
@	Direct (Kronecker) product
+	Addition
-	Subtraction
||	Horizontal concatenation
//	Vertical concatenation

Beispiele

- A+B : matrix addition
- A*B : matrix multiplication,
- A#B : element-wise multiplication
- A[5,2] : Element aus der 5. Zeile, 2. Spalte
- A[1:3,2:10] : die ersten drei Spalten für die 2. bis 10. Zeile
- W = INV(T(x)*x);

Matrizenoperationen direkt im SAS mit Proc IML (Beispiel)

$$\hat{\beta} = (X'X)^{-1}X'y$$

$\hat{Y}$=X*β

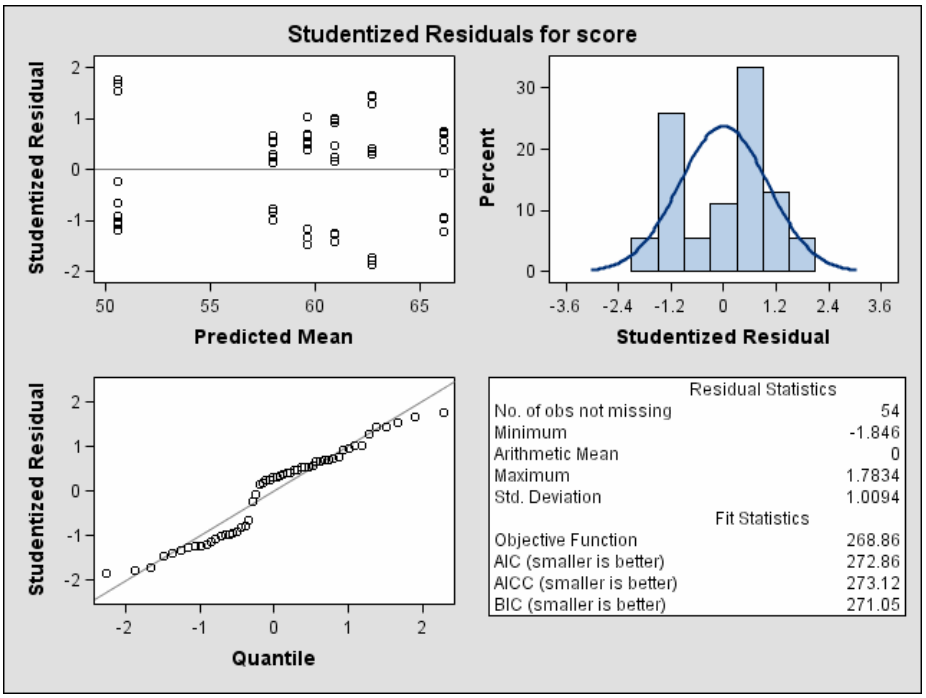
```
PROC IML;
 xpxi=inv(t(x)*x);          /* inverse of X'X          */
 beta=xpxi*(t(x)*y);        /* parameter estimate      */
 yhat=x*beta;               /* predicted values        */

 resid=y-yhat;                 /* residuals            */
 sse=ssq(resid);               /* SSE                  */
 n=nrow(x);                    /* sample size          */
 dfe=nrow(x)-ncol(x);          /* error DF             */
 mse=sse/dfe;                  /* MSE                  */
 cssy=ssq(y-sum(y)/n);         /* corrected total SS   */
 rsquare=(cssy-sse)/cssy;      /* RSQUARE              */
 stdb=sqrt(vecdiag(xpxi)*mse); /* std of estimates     */
 t=beta/stdb;                  /* parameter t tests    */
 prob=1-probf(t#t,1,dfe);      /* p-values             */
QUIT;
```

Graphische Darstellungen mit ODS Graphics

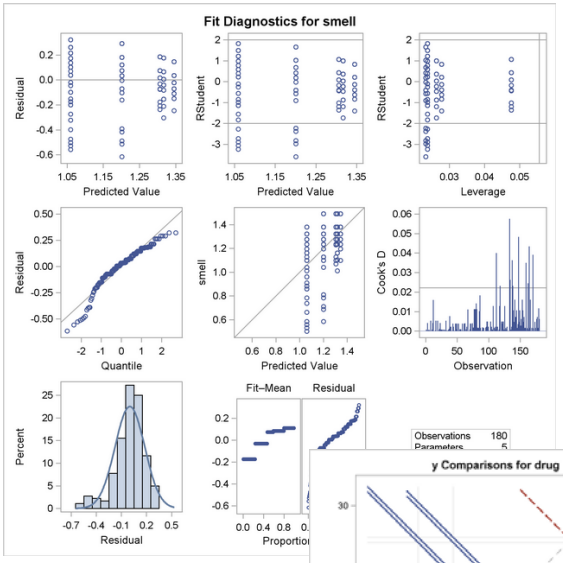
4 Neuerungen

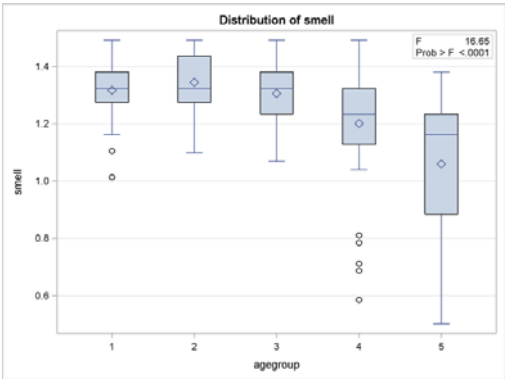
ODS Graphics – Graphiktemplates in vielen SAS/STAT Procedures

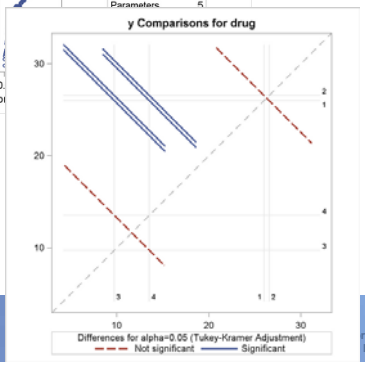


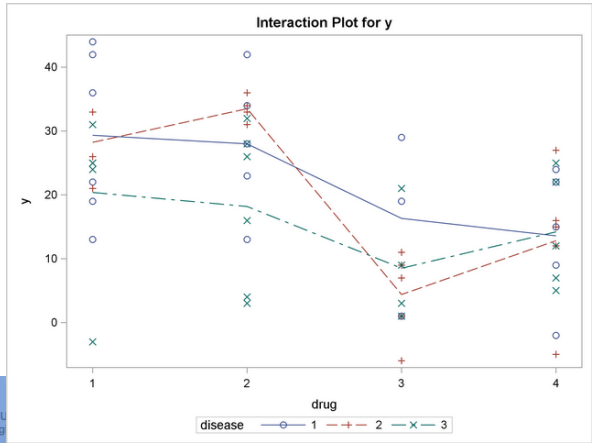
Residuals Plots with MIXED

ODS Graphics – Graphiktemplates in vielen SAS/STAT Procedures



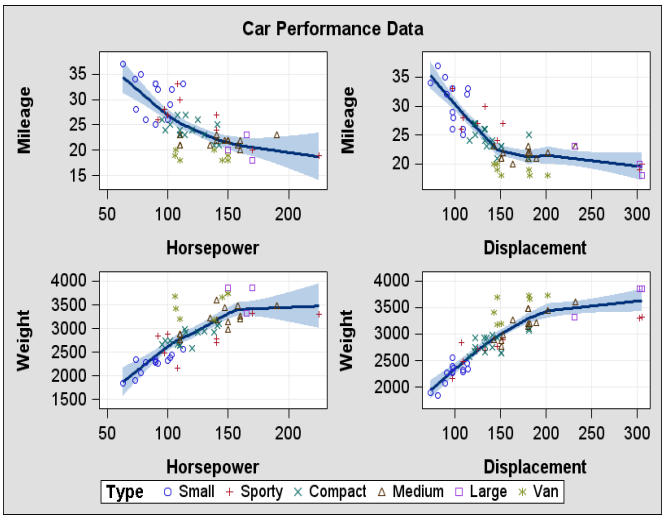


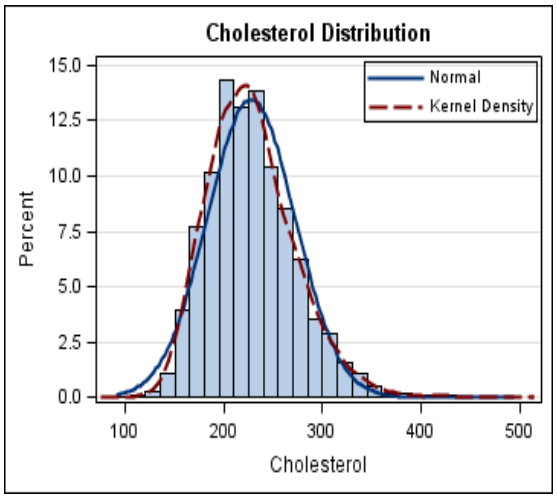




Neue SAS Procedures für statistische Graphiken (SGSCATTER, SGPLOT, SGPANEL)

```
title “Car Performance Data";
proc sgscatter data=cars;
plot (mpg_city weight) *
     (horsepower displacement)/
      group=type loess grid;
run;
```

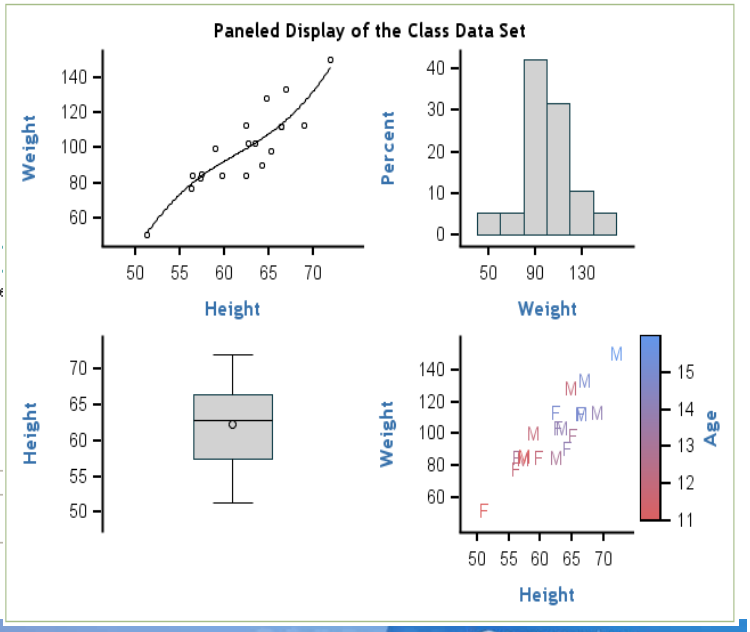




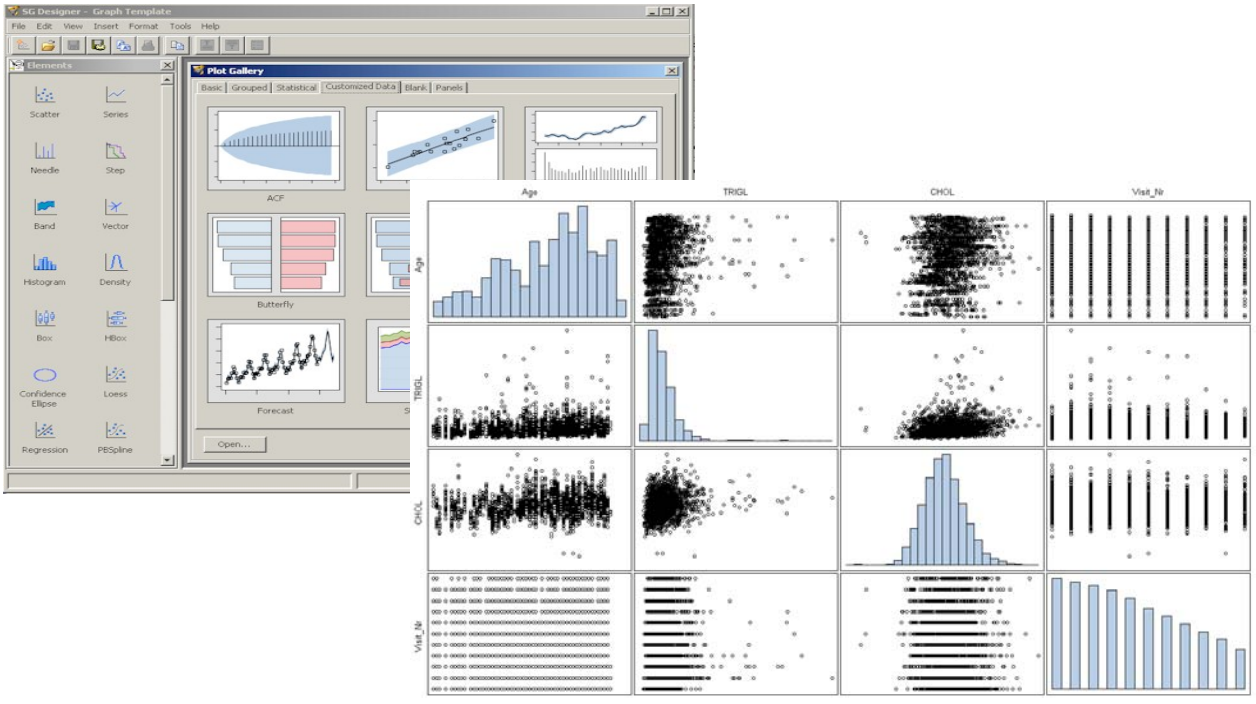

```
title "Cholesterol Distribution";
proc sgplot data=heart;
  histogram cholesterol;
  density cholesterol;
  density cholesterol/type=kernel;
  keylegend/location = inside
     position=TopRight across=1;
  yaxis grid;
run;
```

Erzeugen und Verwenden von Graphik-Templates mit PROC TEMPLATE und PROC SGRENDER

```
proc template;
    define statgraph Panel;
       begingraph;
          entrytitle "Paneled Display of the Class Data Set";
          layout lattice / rows=2 columns=2 rowgutter=10 columngutter=10;
             layout overlay;
                scatterplot y=weight x=height;
                pbsplineplot y=weight x=height;
             endlayout;
             layout overlay / xaxisopts=(label='Weight');
                histogram weight;
             endlayout;
             layout overlay / yaxisopts=(label='Height');
                boxplot y=height;
             endlayout;
             layout overlay / xaxisopts=(offsetmin=0.1 offsetmax=0.
                              yaxisopts=(offsetmin=0.1 offsetmax=0.
                scatterplot  y=weight x=height / markercharacter=se
                   name='color' markercolorgradient=age;
                continuouslegend 'color'/ title='Age';
             endlayout;
          endlayout;
       endgraph;
    end;
 run;
ods html style=meadow;
 proc sgrender data=sashelp.class template=panel;
 run;
ods html close;
```

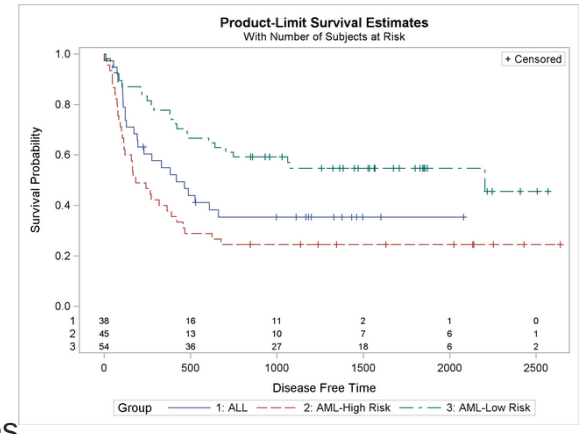


Verwenden des ODS Graphics Designer zur Erstellung der Graphiken



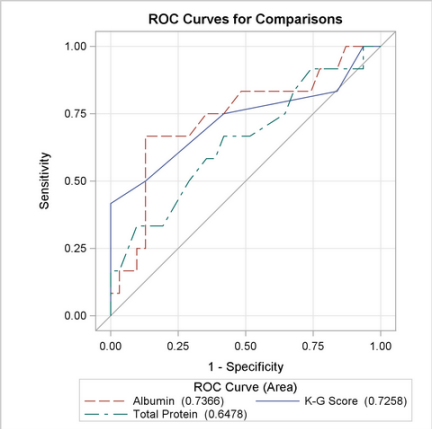
Ausgewählte statistische Auswertungen

New options in the survival analysis procedures



- Proc LIFETEST
  - SURVIVAL statement enables the creation of confidence bands for the survivor function S(t)
  - Number of subjects at risk can be displayed for Kaplan Meier survival curves
  - Smoother hazard function using the kernel method can be specified
- Proc PHREG
  - CLASS statement is available
  - HAZARDRATIO statement provides facitility to calculate hazard ratio in the presence of interactions
  - Firth‘s penalized likelihood method is provided

Proc LOGISTIC



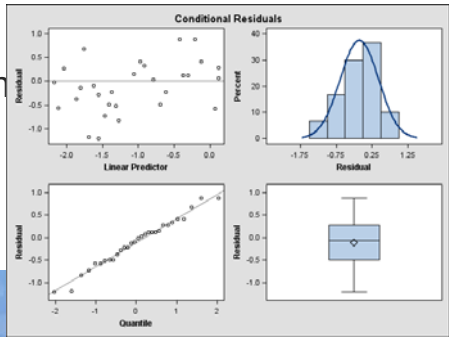
- Model can be output and input with the OUTMODEL and INMODEL option
- SCORE statement allows to score new observations
  - ROC values are calculated for the new observations
- Odds ratios in the presence of interactions are computed
- ROCCONTRAST compares different ROC models
- Performs Firth‘s penalized maximum likelihood

Bayesian analysis in SAS - Overview

- Bayesian analysis added to existing procedures
  - BAYES statement in Proc GENMOD, Proc LIFEREG, Proc PHREG
  - Gibbs sampling
- Proc MCMC
  - Markov Chain Monte Carlo simulations
  - Flexible simulation-based procedure that is suitable for fitting a wide range of Bayesian models

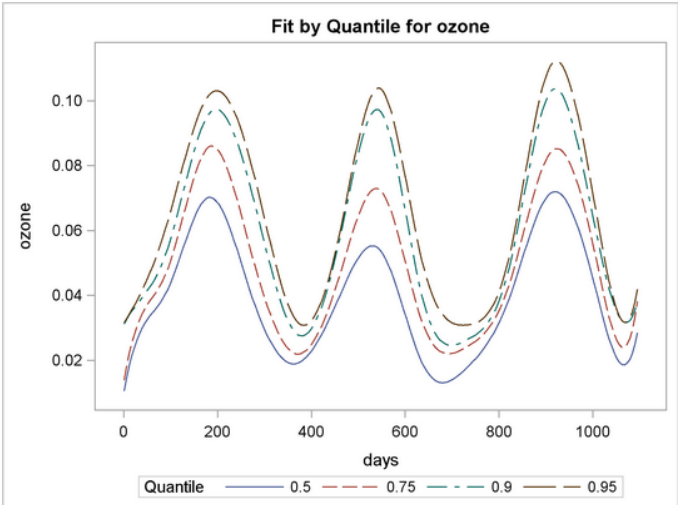
Generalized Linear Mixed Models – Proc GLIMMIX

- Conditional distribution of the data given the random effects is a member of the exponential family (binary, binomial, Poisson, gamma, beta, chi-square)
- Fixed-effects design matrix **X** is specified in the MODEL statement
- Random-effects design matrix **Z** is specified in the RANDOM statement
- Fits models based on likelihood based techniques
- Fits cumulative link models for ordinal data and generalized logit models for nominal data
- Empirical co-variance can be estimated through the EMPIRICAL= option
- Only GLM-type singular parameterization of CLASS variables is supported



Quantile Regression with Proc QUANTREG

- Models the relationship between X and the conditional quantiles of Y given X=x
- Usefull where extrems are important and shall be modeled.
- More complete picture of the conditional distribution of Y given X=x, when lower and upper quantiles are of interest.
  E.g. study of body mass index



Neuerungen in SAS STAT 9.3

- The experimental FMM procedure fits statistical models to data where the distribution of the response is a finite mixture of univariate distributions
- The EFFECT statement is now production. This statement is available in the HPMIXED, GLIMMIX, GLMSELECT, LOGISTIC, ORTHOREG, PHREG, PLS, QUANTREG, ROBUSTREG, SURVEYLOGISTIC, and SURVEYREG procedures.
- The MCMC procedure now supports the RANDOM statement.
- The METHOD=FIML option in the CALIS procedure is now production. This option specifies the full information maximum likelihood method. Instead of deleting observations with missing values, the full information maximum likelihood method uses all available information from all observations.
- The SURVEYPHREG procedure is now production.
- The HPMIXED procedure now provides a REPEATED statement and additional covariance structures.
- The MI procedure offers fully conditional specification methods for multiple imputation.

Neuerungen in SAS STAT 12.1

- The STDRATE procedure computes direct and indirect standardized rates and risks for study populations. With direct standardization, you compute the weighted average of stratum-specific estimates in the study population, using weights such as population-time from a standard or reference population. With indirect standardization, you compute the weighted average of stratum-specific estimates in the reference population by using weights from the study population. The procedure provides summary statistics such as rate and risk estimates (and their confidence limits) for each stratum, as well as graphs.
- The QUANTSELECT procedure performs model selection for linear quantile regression.
- The QUANTLIFE procedure implements two quantile regression approaches that have been developed to account for right-censoring and provide valid estimates

Neuerungen in SAS STAT 12.1

- The MCMC procedure now models missing values by default. The RANDOM statement supports multilevel hierarchy to an arbitrary depth. The procedure also implements faster and more efficient sampling algorithms.
- The PHREG procedure supports Bayesian frailty models.
- The FMM procedure for finite mixture models is now production and adds several truncated distributions.
- The LIFEREG and PROBIT procedures include additional postprocessing statements. They now support the TEST, LSMEANS, LSMESTIMATE, ESTIMATE, SLICE, and EFFECTPLOT statements.
- The FREQ procedure produces mosaic plots.
- The SURVEYSELECT procedure provides Poisson sampling.
- The SURVEYMEANS procedure now performs poststratification estimation.
- The GLM, MIXED, GLIMMIX, and ORTHOREG procedures support the REF= option in the CLASS statement.

Neuerungen in SAS STAT 9.22

- The experimental SURVEYPHREG procedure performs regression analysis based on the Cox proportional hazards model for sample survey data. The procedure provides design-based variance estimates, confidence intervals, and hypothesis tests concerning the parameters and model effects.
- The PLM procedure takes model results that are stored from SAS/STAT linear modeling procedures and performs additional postfitting inferences without your having to repeat your original analysis. The PLM procedure can perform tasks such as testing hypotheses, computing confidence intervals, producing prediction plots, and scoring new data sets by using familiar statements such as the ESTIMATE, LSMEANS, LSMESTIMATE, and SLICE statements.
- The EFFECT statement is now available in the GLIMMIX, GLMSELECT, HPMIXED, ORTHOREG, PHREG, PLS, QUANTREG, ROBUSTREG, SURVEYLOGISTIC, and SURVEYREG procedures. This statement enables you to construct a much richer family of linear models than you can traditionally define with the CLASS statement. Effect types include splines for semiparametric modeling, multimember effects for situations in which measurements can belong to more than one class, lag effects, and polynomials.
- Exact Poisson regression is now available with the GENMOD procedure.
- The MCMC procedure can create samples from the posterior predictive distribution.
- The zero-inflated negative binomial model is now available with the GENMOD procedure.
- The HPMIXED procedure is now production.
- The CALIS procedure has been completely revised and includes enhancements that were formerly available in the experimental TCALIS procedure.

Statistical Graphics in SAS 9.3

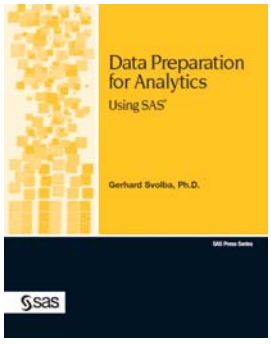
- Producing graphs with ODS Graphics no longer requires a SAS/GRAPH® license. In addition, the family of statistical graphics procedures (SGPANEL, SGPLOT, SGRENDER, and SGSCATTER) has moved from SAS/GRAPH to Base SAS® license.
- The MAXPOINTS= option has been added to the ANOVA, CLUSTER, GLM, LOGISTIC, MIXED, QUANTREG, and VARCLUS procedures. This option specifies a limit for the number of points that can be displayed on certain plots, and these plots are not created when this limit is exceeded. Note that the REG procedure already provided this option.
- The frequency plots and cumulative frequency plots of PROC FREQ and the weighted frequency plot of PROC SURVEYFREQ are no longer produced automatically when ODS Graphics is enabled. You can request these graphs with the PLOTS= option.
- In SAS 9.3, the default destination in the SAS windowing environment is HTML; in addition, ODS Graphics is enabled by default in the SAS windowing environment. These new defaults have several advantages. Graphs are integrated with tables, and all output is displayed in the same HTML file using a new style. This new style, HTMLBLUE, is an all-color style, which is designed to integrate tables and modern statistical graphics. You can view and modify the default settings by selecting **Tools Options Preference** from the menu at the top of the main SAS window. Then click the **Results** Tab.

Abschliessende Bemerkungen

Links and other ressources

- What's new in SAS®STAT 9.2
  http://support.sas.com/documentation/cdl/en/whatsnew/61982/HTML/default/statugwhatsnew.htm
- What's new in SAS®STAT 9.0, 9.1
  http://support.sas.com/documentation/whatsnew/91x/statugwhatsnew900.htm
- An Introduction to Quantile Regression and the QUANTREG Procedure
  http://www2.sas.com/proceedings/sugi30/213-30.pdf
- Introducing the GLMSELECT PROCEDURE for Model Selection
  http://www2.sas.com/proceedings/sugi31/207-31.pdf
- Robust Regression and Outlier Detection with the ROBUSTREG Procedure
  http://www2.sas.com/proceedings/sugi27/p265-27.pdf
- An introduction to partial least squares regression
  http://support.sas.com/techsup/technote/ts509.pdf
- Sample-Size Analysis in Study Planning: Concepts and Issues, with Examples Using PROC POWER and PROC GLMPOWER
  http://www2.sas.com/proceedings/sugi29/211-29.pdf
- Updates to SAS® Power and Sample Size Software in SAS/STAT® 9.2
  http://www2.sas.com/proceedings/forum2008/368-2008.pdf
- A Comparison of the Mixed Procedure and the Glimmix Procedure
  http://www2.sas.com/proceedings/sugi31/189-31.pdf
- Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models
  http://www2.sas.com/proceedings/sugi30/196-30.pdf
- Growing Up Fast: SAS 9.2 Enhancements to the GLIMMIX Procedure
  http://www2.sas.com/proceedings/forum2007/177-2007.pdf
- Old versus New: A Comparison of PROC LOGISTIC and PROC GLIMMIX
  http://www2.sas.com/proceedings/forum2008/226-2008.pdf
- Advanced Statistical and Graphical features of SAS® PHREG
  http://www2.sas.com/proceedings/forum2008/375-2008.pdf
- SAS Online Help for SAS®STAT, Chapter 7: Introduction to Bayesian Analysis Procedures
- Data Preparation for Analytics
  http://www.sascommunity.org/wiki/Data_Preparation_for_Analytics
- Makewide and Makelong Macro
  http://www.sascommunity.org/wiki/Gerhard%27s_Samples

Ausgewählte Bücher in SAS Press Wichtige SAS-Links



Gerhard Svolba
http://www.sascommunity.org/wiki/Data_Preparation_for_Analytics
http://www.sascommunity.org/wiki/Data_Quality_for_Analytics
http://www.sascommunity.org/wiki/New_Features_in_SAS_STAT_9.2

Paul Allison
Fixed Effects Regression Methods for Longitudinal Data Using SAS
Logistic Regression Using SAS: Theory and Application, Second Edition
Survival Analysis Using SAS: A Practical Guide, Second Edition

SAS Global Forum Proceedings
http://support.sas.com/events/sasglobalforum/previous/online.html
z.B: *Handling Missing Data by Maximum Likelihood* *(P. Allison)*

Veranstaltungen
- *SAS Club Austria*
- *KSFE (Konferenz für SAS in Forschung und Entwicklung) 28.2-1.3.2013 in Ulm*