DIE BESTEN TIPPS UND TRICKS AUS MEINEN LIEBSTEN SAS PRESS BÜCHERN – UND WARUM SIE ÜBERLEGEN SOLLTEN, SELBST EIN BUCH FÜR SAS ZU SCHREIBEN

DR. GERHARD SVOLBA
COMPETENCE CENTER ANALYTICS
GREIFSWALD, 3. MÄRZ 2016

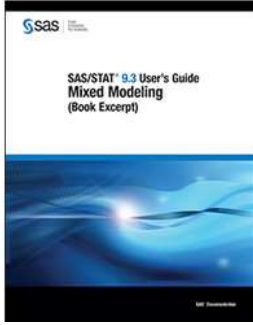
EINLEITUNG | DAS ERWARTET SIE IN ~~MEINEM~~ UNSEREM VORTRAG

- Tipps und Tricks aus ~~meinen~~ **UNSEREN** 6 liebsten SAS Press Büchern, u.a.
  - Longitudinal-Daten
  - SAS Graphiken
  - Survival Analyse
  - SAS Enterprise Guide

- Warum Sie selbst ein SAS Press Buch schreiben sollten
- Welche Schritte dafür notwendig sind

BUCH-TIPP #1 SAS®STAT USERS GUIDE

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



- 9429 Seiten Dokumentation ☺
- Ein Fundus an statistischem Wissen.
- Introductions zu den Procedures für unterschiedliche Themen

SAS/STAT 14.1 User's Guide - Introductory and Common Chapters

For the complete *SAS/STAT 14.1 User's Guide*, go to the SAS/STAT product documentation page.

- Introduction
  PDF | HTML
- Introduction to Statistical Modeling with SAS/STAT Software
  PDF | HTML
- Introduction to Regression Procedures
  PDF | HTML
- Introduction to Analysis of Variance Procedures
  PDF | HTML
- Introduction to Mixed Modeling Procedures
  PDF | HTML
- Introduction to Bayesian Analysis Procedures
  PDF | HTML
- Introduction to Categorical Data Analysis Procedures
  PDF | HTML
- Introduction to Multivariate Procedures
  PDF | HTML
- Introduction to Discriminant Procedures
  PDF | HTML
- Introduction to Clustering Procedures
  PDF | HTML
- Introduction to Scoring, Standardization, and Ranking Procedu
  PDF | HTML
- Introduction to Survival Analysis Procedures
  PDF | HTML
- Introduction to Survey Sampling and Analysis Procedures
  PDF | HTML
- The Four Types of Estimable Functions
  PDF | HTML
- Introduction to Nonparametric Analysis
  PDF | HTML
- Introduction to Structural Equation Modeling with Latent Variables
  PDF | HTML

Estimable Functions

Type I SS and Estimable Functions

In PROC GLM, the Type I SS and the associated hy
operator used to compute a generalized g_2-inverse
model $E[Y] = x_1\beta_1 + x_2\beta_2 + x_3\beta_3$, the Type I S

Effect
x_1
x_2
x_3

Note that some other SAS/STAT procedures comp
PROC MIXED and PROC GLIMMIX), but their tes
using those procedures to fit models that contain su

The Type I SS are model-order dependent; each
model.

There are numerous ways to obtain a Type I hypothe
matrix and then reduce $\mathbf{X}'\mathbf{X}$ to an upper triangular
zero diagonal. The nonzero rows of the resulting m

$$SS(H_0\colon \mathbf{L}\boldsymbol{\beta} = \mathbf{0}) = R(\beta_1)$$

together leads to the posterior distribution of the parameter. You use the posterior distribution to carry out all inferences. You cannot carry out any Bayesian inference or perform any modeling without using a prior distribution.

Objective Priors versus Subjective Priors

Bayesian probability measures the degree of belief that you have in a random event. By this definition, probability is highly subjective. It follows that all priors are *subjective priors*. Not everyone agrees with this notion of subjectivity when it comes to specifying prior distributions. There has long been a desire to obtain results that are objectively valid. Within the Bayesian paradigm, this can be somewhat achieved by using prior distributions that are "objective" (that is, that have a minimal impact on the posterior distribution). Such distributions are called *objective* or *noninformative* priors (see the next section). However, while noninformative priors are very popular in some applications, they are not always easy to construct. See DeGroot and Schervish (2002, Section 1.2) and Press (2003, Section 2.2) for more information about interpretations of probability. See Berger (2006) and Goldstein (2006) for discussions about objective Bayesian versus subjective Bayesian analysis.

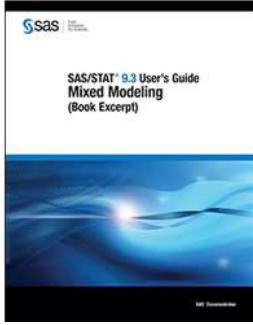
Noninformative Priors

Roughly speaking, a prior distribution is noninformative if the prior is "flat" relative to the likelihood function. Thus, a prior $\pi(\theta)$ is noninformative if it has minimal impact on the posterior distribution of θ. Other names for the noninformative prior are *vague*, *diffuse*, and *flat* prior. Many statisticians favor noninformative priors because they appear to be more objective. However, it is unrealistic to expect that noninformative priors represent total ignorance about the parameter of interest. In some cases, noninformative priors can lead to *improper posteriors* (nonintegrable posterior density). You cannot make inferences with improper posterior distributions. In addition, noninformative priors are often not invariant under transformation; that is, a prior might be noninformative in one parameterization but not necessarily noninformative if a transformation is applied.

See Box and Tiao (1973) for a more formal development of noninformative priors. See Kass and Wasserman (1996) for techniques for deriving noninformative priors.

BUCH-TIPP #1 SAS®STAT USERS GUIDE

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



- „Getting Started“ Sektion und Beispiele zu jeder SAS Procedure
  (und somit zu den wichtigsten Statistik-Themen)

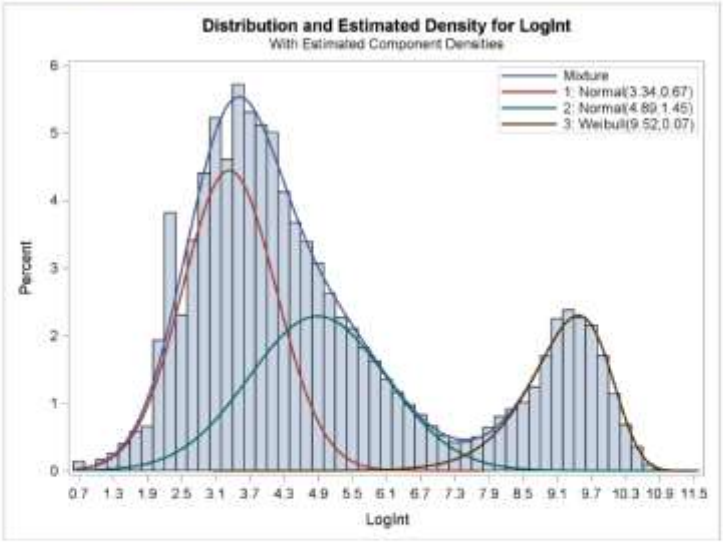
Getting Started: FMM Procedure

Mixture Modeling for Binomial Overdispersion: “Student,’ and Yeast

The following example demonstrates how you can model a complicated, two- distribution, either with maximum likelihood or with Bayesian methods, wit statements.

William Sealy Gosset, a chemist at the Arthur Guinness Son and Company b statistical laboratory of Karl Pearson in 1906–1907 to study statistics. At firs but one paper under the pseudonym “Student” because his employer forbad after a co-worker had disclosed trade secrets—worked on the Poisson limit using haemacytometer yeast cell counts. Gosset’s interest in studying small-san motivated by the small sample sizes he typically saw in his work at the brewery.

Output 39.2.5 Observed and Estimated Densities in the Three-Component Model



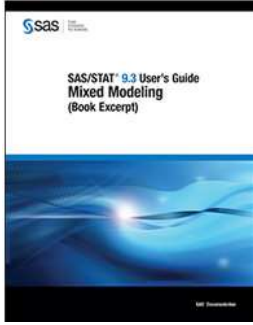
The estimated mixture density matches the histogram of the observed data closely (Output 39.2.5). The component densities are displayed in such a way that, at each point in the support of the LogInt variable, their sum combines to the overall mixture density. The three components in the mixtures are well separated.

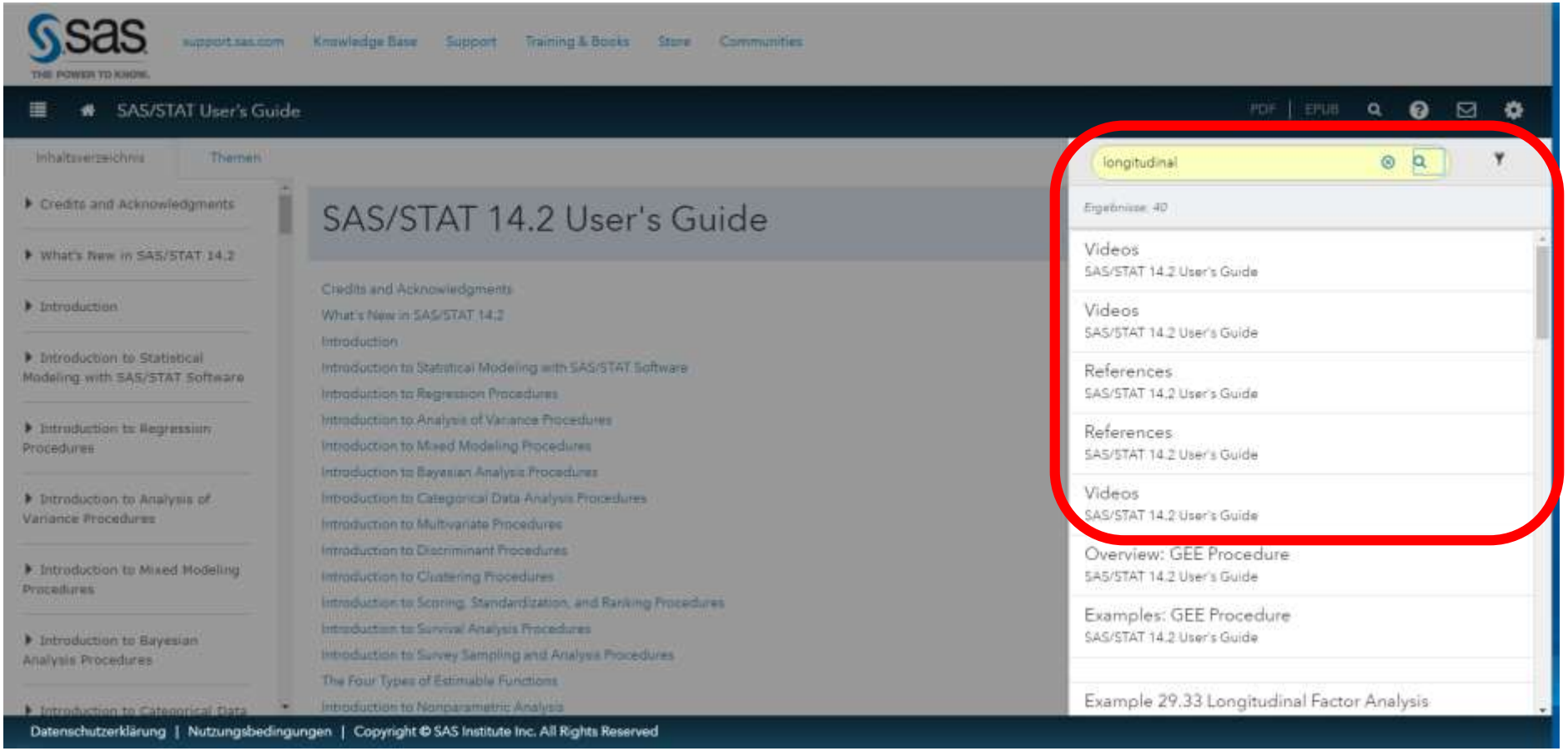
The excellent quality of the fit is even more evident when the distributions are displayed cumulatively by adding the CUMULATIVE option in the DENSITY option (Output 39.2.6):

```
ods select DensityPlot;
proc fmm data=cattle plot=density(cumulative) gconv=0;
   model LogInt = / dist=normal k=2 parms(3 1, 5 1);
   model        + / dist=weibull;
   freq count;
run;
```

BUCH-TIPP #1 SAS®STAT USERS GUIDE

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



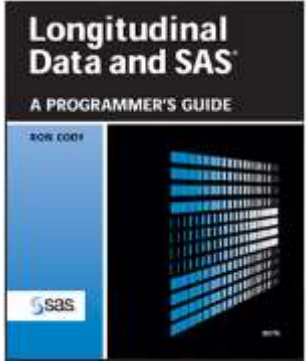


BUCH-TIPP #2

LONGITUDINAL DATA AND SAS®: A PROGRAMMER'S GUIDE

RON CODY

http://support.sas.com/publishing/authors/cody.html



- Sehr handliches Buch über die Aufbereitung von Transaktionsdaten und Daten im multiple-row-per subject Format
- (kein Buch über Zeitreihenanalyse)

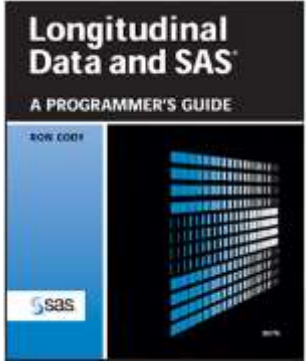
- Produktfokus: SAS Base: Datastep, Proc Means, Proc Freq
  - RETAIN, DIF, LAG, BY, FIRST, LAST
  - PROC MEANS: NWAY, CLASS, …
  - PROC TRANSPOSE

BUCH-TIPP #2

LONGITUDINAL DATA AND SAS®: A PROGRAMMER'S GUIDE

RON CODY

http://support.sas.com/publishing/authors/cody.html




```
proc means data=adp_ct.patients noprint chartype;
 class CentNr sex stage Treatment;
 var weight breslow;
 output out=recl_data mean=
run;
```

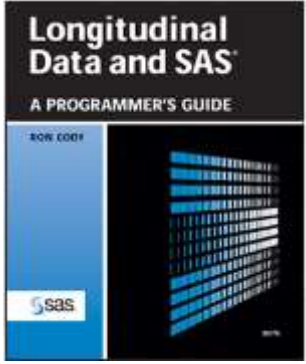
	CentNr	SEX	STAGE	Treatment	_TYPE_	_FREQ_	WEIGHT	BRESLOW
1	.	.	.		0000	407	77.3	3.19
2	.	.	.	A	0001	207	77.5	3.15
3	.	.	.	B	0001	200	77.0	3.23
4	.	.	1		0010	301	77.0	2.44
5	.	.	2		0010	106	78.1	5.40
6	.	.	1	A	0011	153	77.3	2.45
7	.	.	1	B	0011	148	76.6	2.44
8	.	.	2	A	0011	54	78.3	5.31
9	.	.	2	B	0011	52	78.0	5.50
10	.	0	.		0100	215	83.5	3.44
11	.	1	.		0100	192	70.3	2.90
12	.	0	.	A	0101	109	83.3	3.46
13	.	0	.	B	0101	106	83.7	3.42
14	.	1	.	A	0101	98	71.1	2.81
15	.	1	.	B	0101	94	69.4	3.01
16	.	0	1		0110	144	84.3	2.40

BUCH-TIPP #2

LONGITUDINAL DATA AND SAS®: A PROGRAMMER'S GUIDE

RON CODY

http://support.sas.com/publishing/authors/cody.html




```
*** Liste alle Aggregationen, wo nach CENTNR nicht detailliert wurde;
data CentNr_Aggr;
 set recr_data;
 where substr(_type_,1,1)='0';
run;
```

	CentNr	SEX	STAGE	Treatment	_TYPE_	_FREQ_	WEIGHT	BRESLOW
1					0000	407	77.3	3.19
2				A	0001	207	77.5	3.15
3				B	0001	200	77.0	3.23
4			1		0010	301	77.0	2.44
5			2		0010	106	78.1	5.40
6			1	A	0011	153	77.3	2.45
7			1	B	0011	148	76.6	2.44
8			2	A	0011	54	78.3	5.31
9			2	B	0011	52	78.0	5.50
10		0			0100	215	83.5	3.44
11		1			0100	192	70.3	2.90
12		0		A	0101	109	83.3	3.46
13		0		B	0101	106	83.7	3.42
14		1		A	0101	98	71.1	2.81
15		1		B	0101	94	69.4	3.01
16		0	1		0110	144	84.3	2.40

```
*** Liste alle Aggregationen, wo nach Treatment aufgeschlüsselt wurde;
data Treatment_Detail;
 set recr_data;
 where substr(_type_,4,1)='1';
run;
```

	CentNr	SEX	STAGE	Treatment	_TYPE_	_FREQ_	WEIGHT	BRESLOW
1				A	0001	207	77.5	3.15
2				B	0001	200	77.0	3.23
3			1	A	0011	153	77.3	2.45
4			1	B	0011	148	76.6	2.44
5			2	A	0011	54	78.3	5.31
6			2	B	0011	52	78.0	5.50
7		0		A	0101	109	83.3	3.46
8		0		B	0101	106	83.7	3.42
9		1		A	0101	98	71.1	2.81
10		1		B	0101	94	69.4	3.01
11		0	1	A	0111	71	84.7	2.45
12		0	1	B	0111	73	83.8	2.34
13		0	2	A	0111	38	80.7	5.65

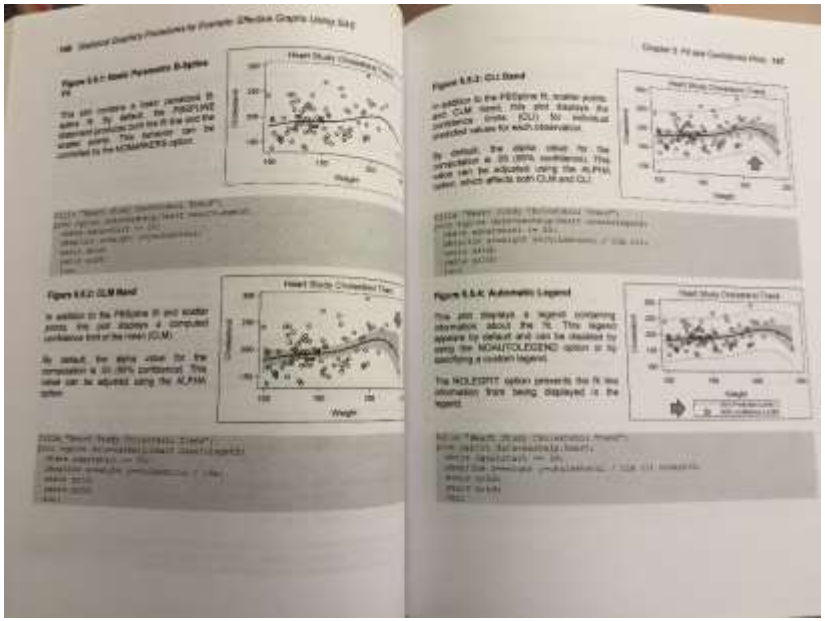
BUCH-TIPP #3

STATISTICAL GRAPHICS PROCEDURES BY EXAMPLE-EFFECTIVE GRAPHS USING SAS®

SANJAY MATANGE, DAN HEATH

http://support.sas.com/publishing/authors/matange.html

- Sehr hilfreiches Buch für die Nutzung der SGPLOT, SGSCATTER und SGPANEL Procedures
- Beispielorientierter Ansatz: Graphik + Code



BUCH-TIPP #3

STATISTICAL GRAPHICS PROCEDURES BY EXAMPLE EFFECTIVE GRAPHS USING SAS®

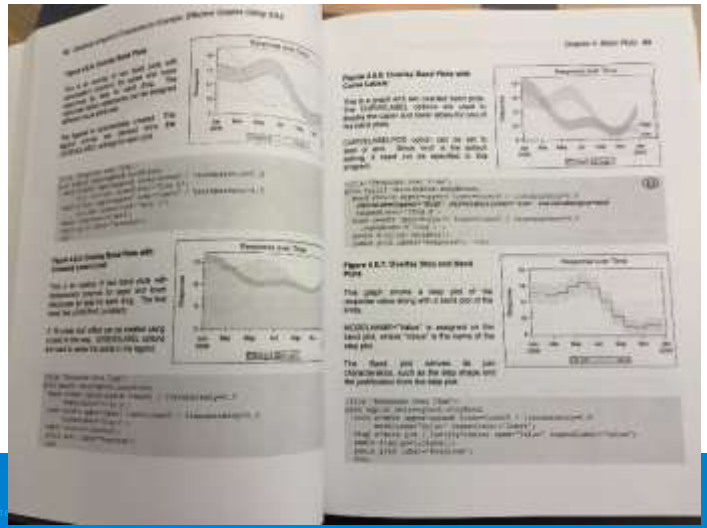
SANJAY MATANGE, DAN HEATH

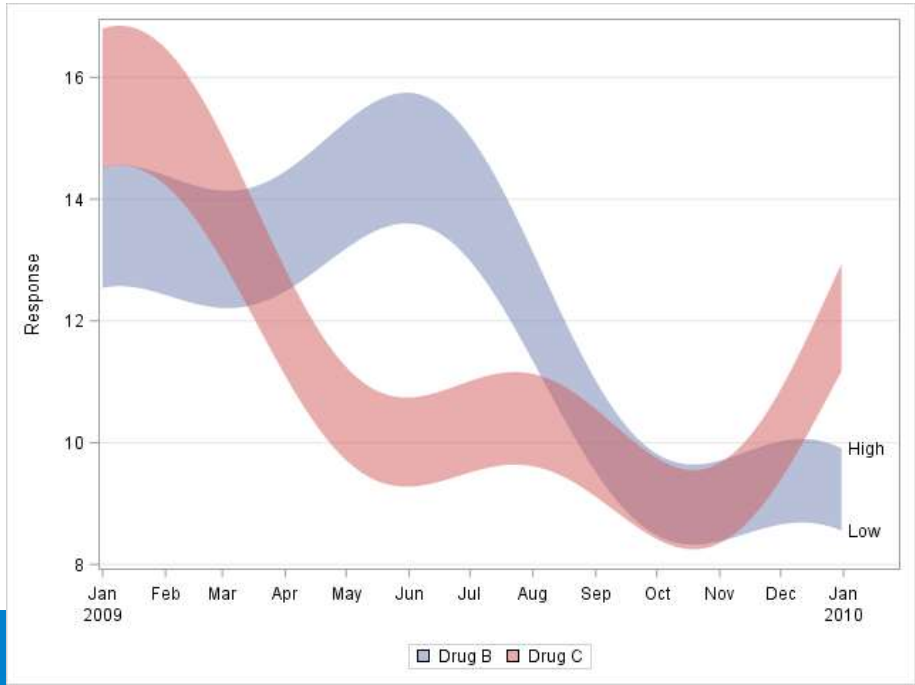
http://support.sas.com/publishing/authors/matange.html


```
title 'Response over Time';
proc sgplot data=BandBreak;
 band x=date upper = upperB lower=lowerB / transparency = 0.5
   curvellabelupper='High' curvelabellower='Low' curvelabelpos = end
   legendlabel = 'Drug B';

 band x=date upper = upperC lower=lowerC / transparency = 0.5
   legendlabel = 'Drug C';

 xaxis display=(nolabel);
 yaxis grid label = 'Response';
run;
```





BUCH-TIPP #3

STATISTICAL GRAPHICS PROCEDURES BY EXAMPLE EFFECTIVE GRAPHS USING SAS®

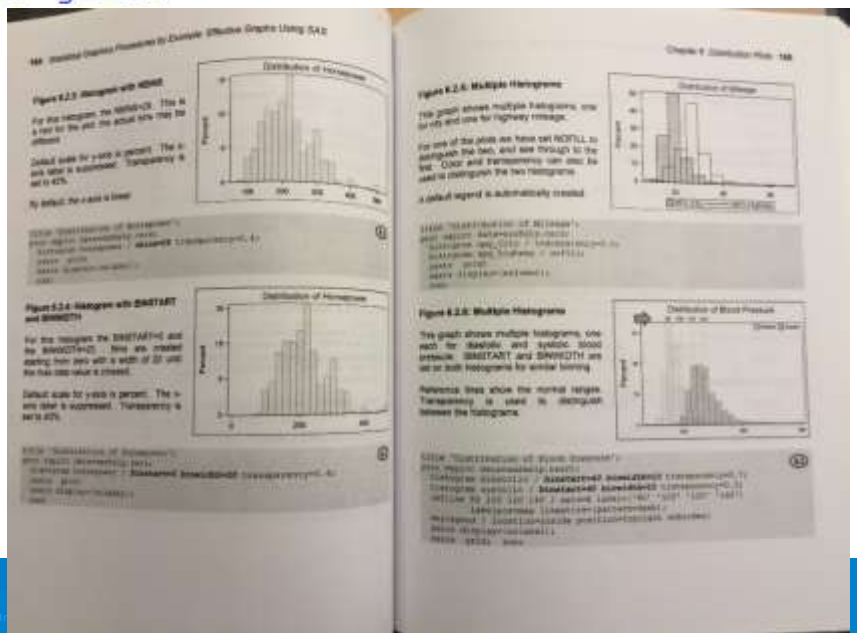
SANJAY MATANGE, DAN HEATH

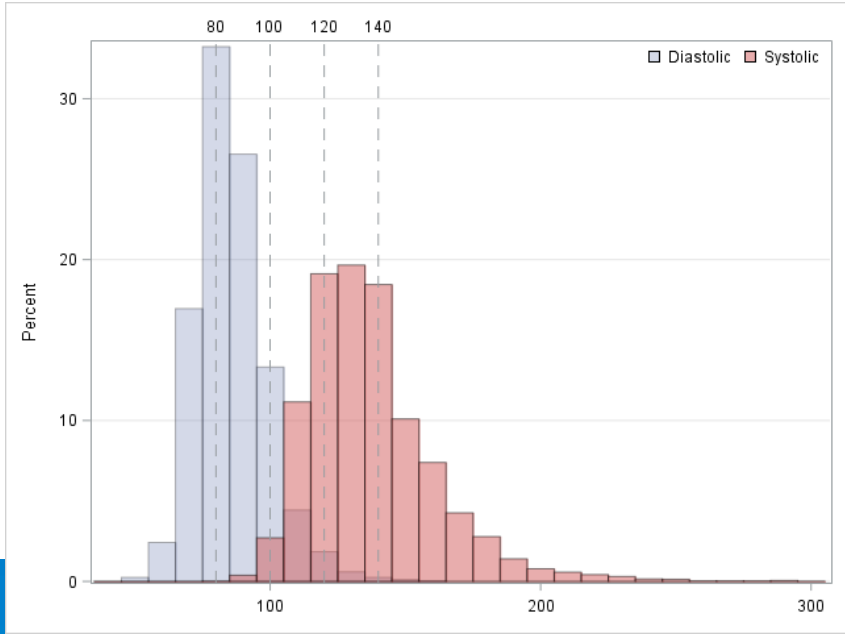
http://support.sas.com/publishing/authors/matange.html


```
title "Distribution of Blood Pressure";
proc sgplot data=sashelp.heart;
 histogram Diastolic / binstart=40 binwidth=10 transparency=0.7;
 histogram Systolic  / binstart=40 binwidth=10 transparency=0.5;

 refline 80 100 120 140 / axis=X label=('80' '100' '120' '140')
        labelpos=max lineattrs=(Pattern=dash);

 keylegend / location=inside position=topright noborder;
 xaxis display=(nolabel);
 yaxis grid;
run;
```



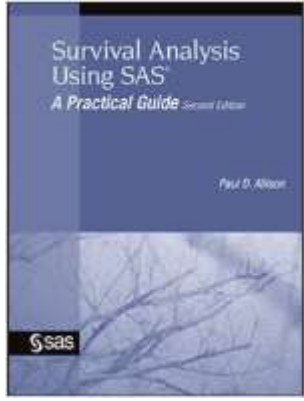


BUCH-TIPP #4

SURVIVAL ANALYSIS USING SAS®: A PRACTICAL GUIDE

PAUL D. ALLISON

http://support.sas.com/publishing/authors/allison.html



- Produktfokus: Survival Procedures in SAS®STAT:
  - Schätzen der Survival Kurven mit Proc LIFETEST (Kaplan Meier, Lifetable)
  - Parametrische Regressions Modelle mit Proc LIFEREG
  - Cox Regression mit Proc PHREG
  - Analyse zeitabhängiger Co-Variablen mit Proc LOGISITC

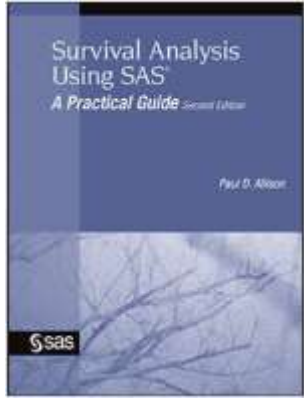
- Sehr gute Einführung in die Grundlagen der Survival-Analyse und die unterschiedlichen Ansätze und Analysemöglichkeiten
  - Diskussion von „Censored Data“ und von „Truncated Data“

BUCH-TIPP #4

SURVIVAL ANALYSIS USING SAS®: A PRACTICAL GUIDE

PAUL D. ALLISON

http://support.sas.com/publishing/authors/allison.html



Zeitangabe als Intervall

```
PROC PHREG DATA=stan2;
   MODEL (ageaccpt,agels)*dead(0)=surg ageaccpt / TIES=EFRON;
RUN;
```

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
surg	1	-1.04966	0.43934	5.7081	0.0169	0.350
ageaccpt	1	1.13190	0.27510	16.9285	<.0001	3.102

```
PROC PHREG DATA=stan2;
   MODEL (ageaccpt,agels)*dead(0)=plant surg ageaccpt / TIES=EFRON;
   IF agetrans>=agels OR agetrans=. THEN plant=0;
   ELSE plant=1;
RUN;
```

Berechnung einer abgeleiteten Variable in der Proc selbst

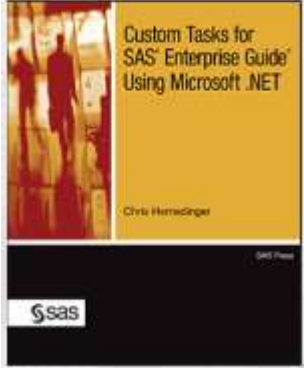
Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
plant	1	-0.47869	0.37353	1.6423	0.2000	0.620
surg	1	-1.07669	0.44177	5.9402	0.0148	0.341
ageaccpt	1	0.99391	0.28224	12.4012	0.0004	2.702

BUCH-TIPP #5

CUSTOM TASKS FOR SAS® ENTERPRISE GUIDE® USING MICROSOFT .NET

CHRIS HEMEDINGER

http://support.sas.com/publishing/authors/hemedinger.html
http://blogs.sas.com/sasdummy



- Chris Hemedinger (R&D Manager bei SAS für den SAS®Enterprise Guide) beschreibt wie Sie den SAS Enterprise Guide um sog. „Custom Tasks“ erweitern können.
  - Erscheinen im EXTRAS-Menü unter „Add-Ins“
  - Können in EG Prozessflüsse integriert und wie Standard Tasks verwendet werden
- Sie passen so den EG den Anforderungen der End-User an.

- Was benötigen Sie dafür:
  - Programmierung in SAS (Funktionalität des neuen Task)
  - Programmierung in .NET (Benutzeroberfläche und Einbettung in den SAS EG)

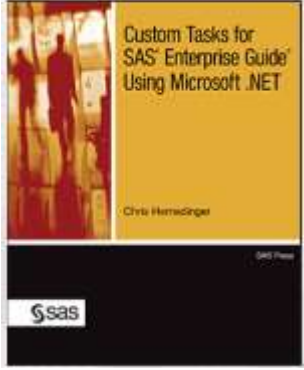
- Buch
  - Konzepte, Ideen und Werkzeuge
  - Beispiele für Custom Tasks
  - Tipps und nützliche Links dazu

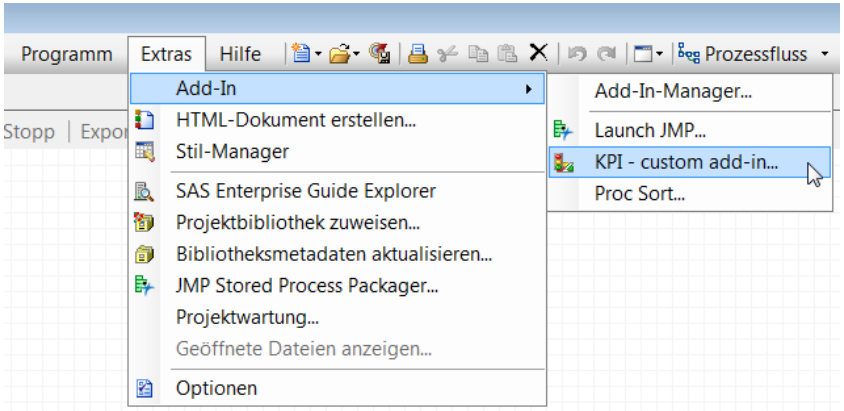
BUCH-TIPP #5

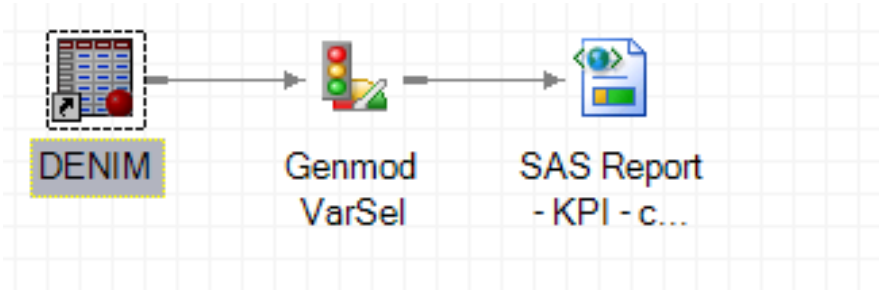
CUSTOM TASKS FOR SAS® ENTERPRISE GUIDE® USING MICROSOFT .NET

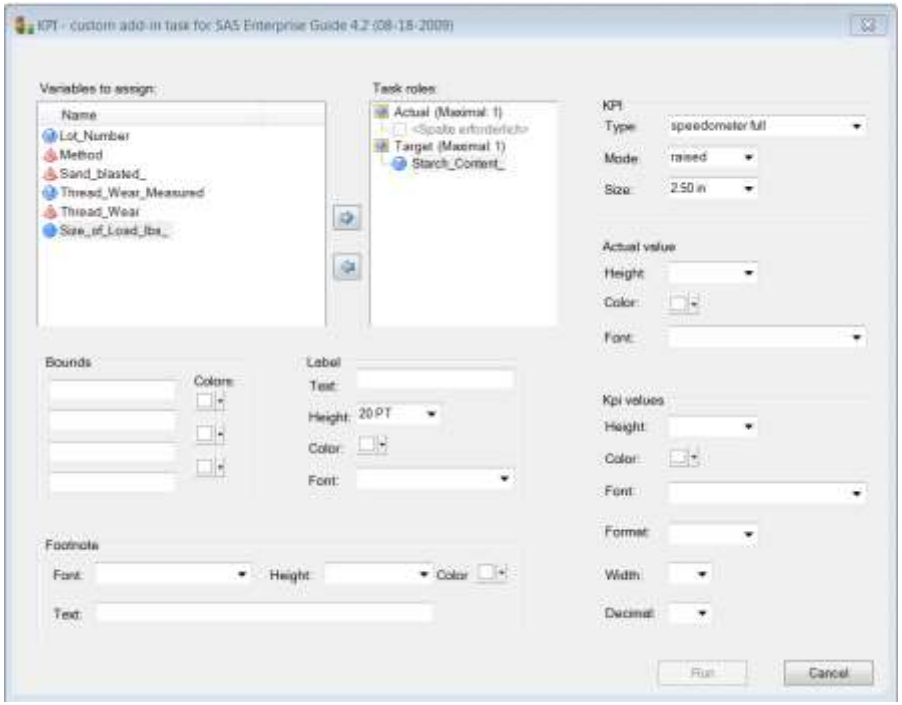
CHRIS HEMEDINGER

http://support.sas.com/publishing/authors/hemedinger.html







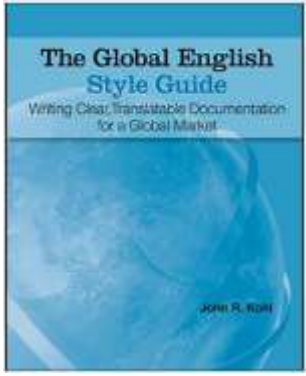


BUCH-TIPP #6

THE GLOBAL ENGLISH STYLE GUIDE: WRITING CLEAR, TRANSLATABLE DOCUMENTATION FOR A GLOBAL MARKET

JOHN R. KOHL

http://support.sas.com/publishing/authors/kohl.html



- Nicht nur für Autoren englischer Bücher!
- Wir schreiben immer mehr Texte: Emails, Conference Abstracts, Papers, Forum-Beiträge, …
- Das Buch ist voll mit „Do‘s“ und „Don‘t s“ Beispielen für die optimale Formulierung.

x If Chocolate Bits is set to NO, indicating that there are no chocolate bits in the sample batch of ice-cream, then the selection for Enough Bits and Size of Bits are grayed to present users from entering irrelevant data (40 words).

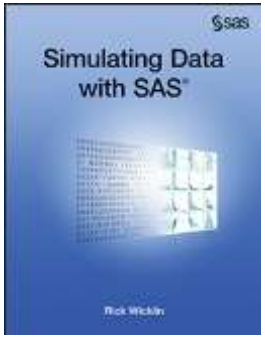
√ If Chocolate Bits is set to NO, then there are no chocolate bits in the sample batch of ice cream. **Therefore**, the selection for Enough Bits and Size of Bits are grayed to present users from entering irrelevant data (20 + 19 words).

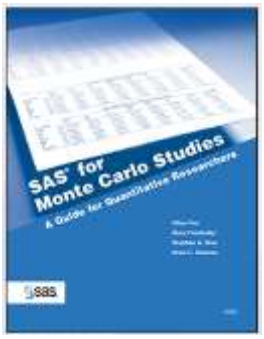
x The **import** of the data into MySQL is also very simple.

√ It is also very easy to **import** the data into MySQL.

WEITERE BÜCHER | BÜCHER, DIE IN DIESEM VORTRAG NICHT VORKOMMEN

Separater Vortrag
„Simulationen und Mathematische Programmierung mit SAS“





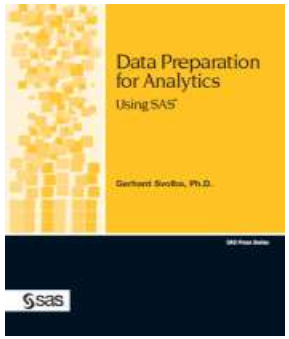
Xitao Fan, Ph.D.
Akos Felsovalyi, M.S.
Stephen A. Sivo, Ph.D.
Sean C. Keenan, Ph.D.
http://support.sas.com/publishing/authors/felsovalyi.html

Bücher von Rick Wicklin
http://support.sas.com/publishing/authors/wicklin.html

Gerhard Svolba

KSFE 2013: Data Quality for Analytics Using SAS
http://www.sascommunity.org/wiki/Data_Quality_for_Analytics

KSFE 2008: Data Preparation for Analytics Using SAS
http://www.sascommunity.org/wiki/Data_Preparation_for_Analytics



ÜBER DIE HEILENDE WIRKUNG VON BÜCHERVERBRENNUNGEN

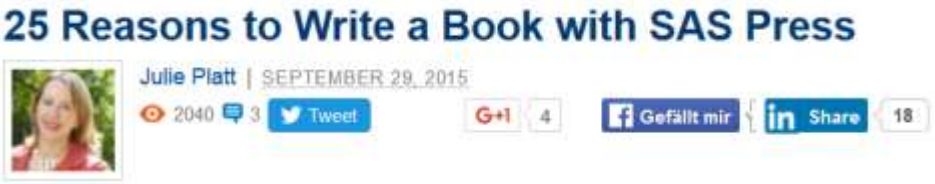
PUBLIZIEREN MIT SAS PRESS

WARUM SIE SELBST EIN BUCH SCHREIBEN SOLLTEN

- Detaillierte Auseinandersetzung mit einem für Sie wichtigen Thema
- Umwandlung von „impliziten Wissen“ in „explizites Wissen“
- Etwas Nachhaltiges schaffen, was ein Projekt, eine Auswertung „überlebt“
- Ihre eigenen SAS Programmier-Kenntnisse zu verbessern
- Sich in ihrem Fachgebiet mehr Sichtbarkeit und Bekanntheit zu verschaffen
- Sich daran zu freuen, Feedback, Anfragen, Fehlermeldungen aus aller Welt zu erhalten
- Weil Sie gerne auf English schreiben möchten

- ~~Weil Sie viel Geld verdienen möchten.~~

SAS Learning Post Blog: 25 reasons to write a book with SAS Press



PUBLIZIEREN MIT SAS PRESS

WIE SIEHT DER ABLAUF AUS

1. Sie haben eine Buch-Idee
2. Abklärung der Freigabe Ihres Arbeitgebers bzw. anderer vertraglicher Verpflichtungen
3. Einreichung der Buch-Idee an SAS Press
   a. Author's Questionnaire: Fakten über Sie, Ihr Umfeld und Ihre Buchidee
   b. Information Release Agreement: Darf SAS Press Ihre Idee zum Review weiterleiten?
   c. Outline: Beschreibung Ihrer Buch-Idee, Besonderheiten, Gliederung, ev. Sample Chapter
4. Feedback von SAS
5. Vertragliche Details, Definition eines Zeitplans
6. Sample Chapters und Review der Sample Chapters
7. Final Draft und Review Phase
8. Feedback zu den Reviews und Final Version
9. Copy Edit
10. Publikation

http://blogs.sas.com/content/sastraining/

PUBLIZIEREN MIT SAS PRESS

EIN DRITTES (LETZTES?) MAL

Applying Data Science:
Business Case Studies Using SAS

Data Science and Analytics helps you to solve your business questions

The SAS® Analytic Plattform is perfectly suited to perform these analyses

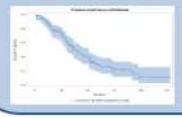
Eight Case Studies with Business Background, Results, Interpretation and SAS Code

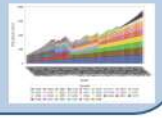
SAS Press (expected 2017)
http://www.sascommunity.org/wiki/Applying_Data_Science_-_Business_Case_Studies_Using_SAS

Gerhard Svolba

Performing Headcount Survival Analysis for Employee Retention

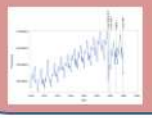
Can you make assumptions about the average length of time intervals, even if most of the endpoints have not yet been observed?

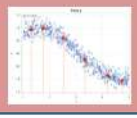




Detecting Outliers and Structural Changes in Longitudinal Data

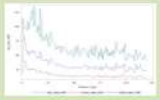
Can you automatically detect events and changes in the course of your data over time?

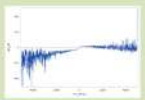




Explaining Deviations and Forecast Errors

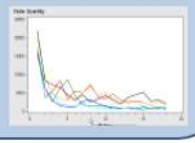
Do the demand planners really improve forecast accuracy with their manual overwrites?





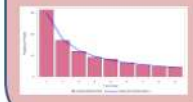
Forecasting the Demand for New Products

Can you assess the exptected demand of products that are introduced right now?



Checking the Alignment with Predefined Pattern

Which customers show a behaviour which is far from what you expected?

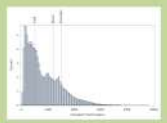


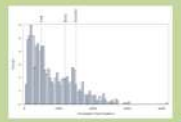
Listen to Your Data – Discover Unknown Relationships

Can your data tell you stories, even if you don't ask them?

Using Monte Carlo Simulations to Understand the Outcome Distribution

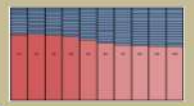
Will the Sales Manager keep his job (when you look at his sales pipeline)?





Studying Complex Systems – Simulate the Monopoly® Board Game

How can you simulate complex environments to get insight in the most frequent processes?



ZUSAMMENFASSUNG Links und Referenzen

- Ron Cody: http://support.sas.com/publishing/authors/cody.html
- Sanjay Matange: http://support.sas.com/publishing/authors/matange.html
- Paul Allison: http://support.sas.com/publishing/authors/allison.html
- STAT Users Guide: https://support.sas.com/documentation/onlinedoc/stat/index.html
- Chris Hemedinger: http://support.sas.com/publishing/authors/hemedinger.html
- John Kohl: http://support.sas.com/publishing/authors/kohl.html
- Rick Wicklin: http://support.sas.com/publishing/authors/wicklin.html

- SAS Learning Post Blog: 25 reasons to write a book with SAS Press
- SAS Bookshelf Blog: 4 ways to cure yourself after writing a book