

Eckard Moll, Thomas Stauber

Neues für den Anwender der SAS-Version 8 im Vergleich zu SAS 6.12

1 Einführung

Die meisten Anwender steigen derzeit von der Version SAS 6.12 auf die Version 8 um. Jede Version hat bisher mindestens kleine Änderungen und Erweiterungen mit sich gebracht. Der Vorteil der Version SAS 6.12 liegt aber darin, dass sie über mehrere Jahre hinweg stabil genutzt werden konnte. Einer Unruhe, die jede neue Version hervorruft, sollten auch für den gelegentlichen Nutzer akzeptable Erweiterungen gegenüber stehen. Die Version 8 begann mit SAS 8.0 und hat erst mit SAS 8.1 den Leistungsumfang von SAS 6.12 bzw. geht darüber hinaus. Neu ist mit der Version 8 die Oberfläche, die grafische Ausgabe in hochauflösende Grafiken als Standard, neue Prozeduren wie beispielsweise BOXPLOT und NLMIXED, interessante Erweiterungen in vorhandenen Prozeduren, das Output-Delivery-System, erweiterte Möglichkeiten des Datenim- und -exports und anderes. Es kann und soll nicht das Anliegen dieses Artikels sein, alle Änderungen und Neuerungen in SAS 8.1 im Vergleich zu SAS 6.12 aufzuzeigen, sondern dem Anwender den Übergang von SAS 6.12 auf SAS 8.1 bzw. SAS 8.2 unter MS-Windows 9.x, MS-Windows 2000 und NT 4.0 zu erleichtern und auf einige uns wesentlich erscheinende Veränderungen einzugehen.

1. Oberfläche und Hilfesystem

1.1 Neue Fenster und deren Eigenschaften

Beim Start der Version 8 von SAS fällt sofort auf, dass im Vergleich zum Window-Standard ein neues Fenster dazugekommen ist, das Explorer/Results Fenster. Dieses Fenster bietet den Zugriff auf SAS-Libraries, Catalogs, Datasets und deren Eigenschaften. Das Results-Fenster erlaubt ein besseres Verwalten des Outputs. Eine Besonderheit besteht darin, dass Results und Explorer zusammen in einem Fenster (docking Funktion) stehen können.

Die wichtigste Neuerung der Oberfläche ist der SAS-Programmeditor (Enhanced Editor). Dieser Editor ist syntaxsensitiv und die SAS-Schlüsselwörter werden farbig unterlegt. Dadurch wird die Eingabe syntaktisch richtiger SAS-Programme erleichtert. Es können - manchmal auch ungewollt - mehrere Editor-Fenster geöffnet werden. Die Farbskala für die Schlüsselwörter kann nutzerseitig beliebig verändert werden.

1.2 Das Hilfesystem

Das Hilfesystem der Version 8 von SAS ist wesentlich erweitert worden. Es existieren jetzt drei Hilfesysteme, die Standardhilfe, die SAS OnlineDoc im HTML- und die SAS OnlineDoc im PDF-Format. Alle Hilfen und Dokumentatio-

nen basieren auf HTML. SAS bevorzugt als Browser den MS-Internet Explorer. Der Netscape Navigator funktioniert leider nur eingeschränkt. Die SAS OnlineDoc im PDF-Format entspricht zu großen Teilen der gedruckten Form der SAS Handbücher. Sie gehört nicht zum normalen Lieferumfang. Der HTML-Index dieser Dokumentation weist auf eine Vielzahl von PDF-Dateien, die zum Nachlesen oder zum auszugsweisen Drucken geeignet sind. Zum Suchen von Begriffen, Prozeduren oder anderen Einträgen sind die Standardhilfe und die SAS OnlineDoc in HTML besser geeignet.

Neu ist auch, dass mit der F1-Taste wie in Windows üblich die Standardhilfe aufgerufen werden kann. Zu dem Begriff, auf dem der Cursor im Editor-Fenster steht, wird die Hilfe angezeigt.

1.3 Konventionen für Tabellen und Felder (Variablen)

Im Vergleich zu SAS 6.12 wurden Längen von Bezeichnern und dergleichen erweitert (Müller 1999, Ridder 2000):

Tabellennamen, Spaltennamen, Katalognamen, Katalogeinträge	32 Zeichen
Libraries, Formate, Filerefs	8 Zeichen
Länge von Character Variablen	max 32 kByte
Label	max. 256 Zeichen

Variablenamen dürfen keine Umlaute enthalten. Es bleiben dieselben Variablen, wenn ein oder mehrere Buchstaben bei verschiedenen Aufrufen klein oder groß geschrieben sind.

2. Neuerungen und Änderungen, die sich hauptsächlich auf die Berechnung statistischer Maßzahlen und auf statistische Verfahren beziehen

2.1 Zu „What’s New in Release 8.1“

Im Online-Hilfesystem sind unter „Inhalt“ ® „Help on SAS Software Products“ die SAS Produkte aufgelistet. So werden beispielsweise unter SAS/STAT Software die einzelnen Prozeduren aufgeführt. Zuerst steht aber die Rubrik „What’s New in Release 8.1“. Dort sollte man beim Umstieg auf die Version 8 einen Blick hinein werfen. Allerdings bauen die meisten Hinweise auf die Vorgänger-Version SAS 8.0 auf. Beispielsweise zu der Prozedur BOXPLOT, die dort bereits auftrat, fehlen die Ausführungen.

2.2 PROC BOXPLOT

Box-Wisker-Plots erfreuen sich zunehmender Beliebtheit, um von einer Datenmenge die Lage des Mittelwertes, die Quartile und Extremwerte zu zeichnen und bereits visuell Vergleiche zwischen der Lage der verschiedenen Schätzwerte vorzunehmen. Unter SAS 6.12 sind es zwei Wege, die allerdings nicht so verbreitet sind, Box-Plots darzustellen. Das ist einmal die SAS/GRAPH-Prozedur Gplot mit der Anweisung PLOT und unbedingt der Option I=BOXT in der SYMBOL-Anweisung. Das Ergebnis sind filigrane Box-Plots ohne Markierung des Mittelwertes. Andererseits führt die Anwendung der SAS/QC-Prozedur SHEWHART mit der Anweisung BOXCHART zu einer grafisch ansprechenden Lösung.

Nun steht neu unter SAS/STAT die Prozedur BOXPLOT zur Verfügung, die ausschließlich Box-Wisker-Plots liefert. Ein kleines Beispiel im Programm 1 zeigt die Anwendung dieser Prozedur für zufällig erzeugte Daten von drei verschiedenen Verteilungen. Das Ergebnis ist in der Abb. 1 zu sehen. Lässt man die Sortierung weg, entsteht zu jedem der Einzelpunkte eine auf einen Strich verkürzte Box.

Das Programm 1 zeigt die Neuerung, dass die Bezeichnung einer Variablen länger als die bisher üblichen 8 Zeichen lang sein kann. Bei der Ausgabe einschließlich der Grafik wird die erste auftretende Schreibweise, die gleichzeitig der Variablendefinition dient, gewählt. Im Programm 1 ist zu erkennen, dass die Variable Verteilung beim ersten Aufruf mit großem Anfangsbuchstaben geschrieben ist, beim zweiten Auftreten mit kleinem. In der Grafik (Abbildung 1) ist diese Variable folglich mit großem Anfangsbuchstaben aufgeführt.

Programm 1: Beispiel für die Anwendung von PROC BOXPLOT

```
data sas1;
  do i = 1 to 100;
    Verteilung = 'Normal- ';
    Wert = rannor(0);
    output;
    verteilung = 'Poisson- ';
    Wert = ranpoi(0,1);
    output;
    Verteilung = 'Binomial-';
    Wert = ranbin(0,10,0.6);
    output;
  end;

proc sort data=sas1;
  by Verteilung;

proc boxplot data=sas1;
  plot Wert*Verteilung / cboxes=black;
run;
quit;
```

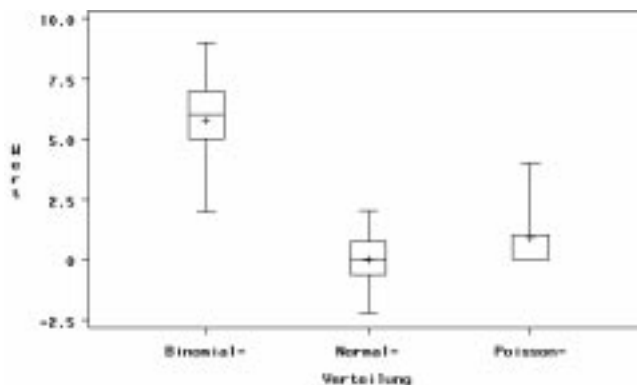


Abb. 1: Boxplots

Die Optionen der PLOT-Anweisung der Prozedur BOXPLOT, die nach dem Schrägstrich folgen, beziehen sich überwiegend auf die grafische Gestaltung und sind schnell abzuleiten. So kann beispielsweise die Breite der Boxen mit der Option BOXWIDTH=n verändert werden. Die Grafik kann in die Zwischenablage übernommen oder mit „File“ ® „Export as Image ...“ in verschiedenen Grafikformaten abgespeichert werden.

2.3 PROC UNIVARIATE

Um einen Vergleich des Lageparameters m mit einer Konstanten m_0 wurde die SAS/BASE-Prozedur UNIVARIATE erweitert, wobei standardseitig gilt: $m_0 = 0$. Die Vorgabe eines anderen Wertes für m_0 ist möglich. Beim Vergleich auf Normalverteilung sind weitere Tests hinzu gekommen. Zusätzlich ist es möglich, ein Histogramm der Daten zeichnen zu lassen, das von einer Darstellung der Normalverteilung überlagert werden kann. Um das zu demonstrieren, wird im ersten Teil des Programms 2, der DATA-Anweisung, eine zufällige Auswahl mit $N(3.3, 9)$ normalverteilter Daten vorgenommen. Zu beachten ist, dass die Variable Wert hier das erste und einzige Mal mit großem Anfangsbuchstaben geschrieben ist.

Programm 2: Beispiel für die Anwendung von PROC UNIVARIATE

```
data sas2;
  do i = 1 to 250;
    my = 3.3;
    sigma2 = 9;
    Wert = my + sqrt(sigma2)*rannor(0);
    output;
  end;

proc univariate data=sas2
  normal
  mu0=3.3;

  var wert;
  histogram/ normal ;
  INSET median mean;
run; quit;

proc univariate data=sas2 noprint;
  var wert;
  output out=sas3 mean=mean std=s gini=g
  mad=mad;
data sas4;
  set sas3;
  pi = 2*arcsin(1);
  rg = g*sqrt(pi)/2;
  rmad = 1.4826*MAD ;
```

```
proc print data=sas4 noobs;
  var mean g mad s rg rmad;
run;
```

Im ersten Aufruf der Prozedur UNIVARIATE (Programm 2) werden zusätzlich zur Berechnung bisher üblicher statistischer Maßzahlen und Tests durch die Option NORMAL verschiedene Tests auf Normalverteilung (Output 1) durchgeführt. Der c2-Anpassungstest zählt nicht dazu. Nur hingewiesen wird auf die Optionen zur Berechnung von Konfidenzintervallgrenzen CIBASIC für Mittelwert, Standardabweichung und Varianz unter der Annahme der Normalverteilung, CIPCTLDF für Quantile ohne Verteilungsannahme und CIPCTLNORMAL für Quantile unter der Annahme der Normalverteilung.

Durch die Option m0=3.3 wird die Nullhypothese H0: m = 3.3 (Output 2) getestet. Keiner der Tests auf Normalverteilung zeigt mit den Überschreitungswahrscheinlichkeiten an, dass die Nullhypothese - die Verteilung der Werte unterscheidet sich nicht von der einer Normalverteilung - verworfen werden muss. Auch die Vergleiche des Mittelwertes, der für diese zufällig ausgewählten Werte mit 3.17 angegeben wird, mit der Konstanten m0 = 3.3 liefern kein Ergebnis, das berechtigt, die Nullhypothese mit einer Irrtumswahrscheinlichkeit von 0.05 abzulehnen. Als Tests werden der t-Test, der Vorzeichenstest und der Vorzeichen-Rangsummen-Test herangezogen.

Zusätzlich zu ausgewählten geschätzten Quantilen werden nun auch ohne die Option NORMAL immer die unter der Normalverteilung erwarteten Quantile berechnet und nebeneinander aufgelistet.

Output 1: Ergebnisse der Tests auf Normalverteilung der Werte

Tests for Normality				
Test		--Statistic--		----p Value-----
Shapiro-Wilk	W	0.996395	Pr < W	0.8386
Kolmogorov-Smirnov	D	0.034422	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.043817	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.289428	Pr > A-Sq	>0.2500

Output 2: Ergebnisse der Tests der Nullhypothese H0: m = 3.3

Tests for Location: Mu0=3.3				
Test		-Statistic-		----p Value-----
Student's t	t	-0.69455	Pr > t	0.4880
Sign	M	1	Pr >= M	0.9496
Signed Rank	S	-732.5	Pr >= S	0.5232

Von Interesse wird die Anweisung HISTOGRAM sein. Es wird eine Klasseneinteilung der Werte vorgenommen und die Häufigkeiten in diesen Klassen werden als Balkendiagramm gezeichnet (Abb. 2). Zusätzlich können verschiedene Optionen angegeben werden. Erwähnt werden sollen hier nur die für solche Verteilungen wie die Beta-, Exponential-, Gamma-, Normal-, Lognormal- und Weibull-Verteilung sowie für nichtparametrische Kerndichteschätzungen, deren Parameter entweder geschätzt oder vorgegeben werden. Die Option NORMAL im Programm 2 bezieht sich also auf eine Normalverteilung mit den geschätzten Parameterwerten, die für das Beispiel mit N(3.17, 8.74) berechnet wurden. Diese Dichtefunktion wird in das Histogramm gezeichnet (Abb. 2). Die INSET-Anweisung gibt in einem Textfeld in der

Histogramm-Grafik für die aufgeführten Statistiken die Schätzwerte an. In dem Beispiel sind das die Schätzwerte für den Median und den Mittelwert (Abb. 2).

Auf die Anweisungen PROBLOT oder QQPLOT in der Prozedur UNIVARIATE, die ein Wahrscheinlichkeits- bzw. ein Quantile-Quantile-Plot erzeugen, soll hier nicht weiter eingegangen werden.

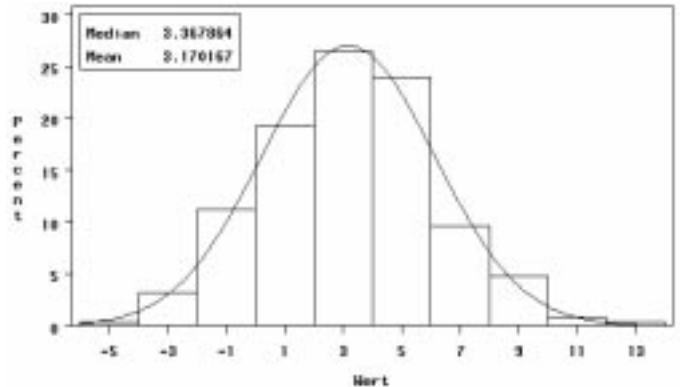


Abb. 2: Histogramm mit Normalverteilung

Der zweite Aufruf der Prozedur UNIVARIATE im Programm 2 nimmt Bezug auf zwei neue Maßzahlen, robuste Skalenschätzer (SAS Procedures Guide, Version 8 1999, S. 1403 f), die in der OUTPUT- und der INSET-Anweisung angegeben werden können:

$$\text{GINI : Gini's mean difference : } G = \frac{1}{\binom{n}{2}} \sum_{i < j} |x_i - x_j|$$

Der Erwartungswert für Daten aus normalverteilter Grundgesamtheit ist $E(G) = \frac{2\sigma}{\sqrt{\pi}}$. Für Daten aus normalverteilter

Grundgesamtheit ist $G\sqrt{\pi}/2$ ein unverzerrter Schätzer für die Standardabweichung σ (Variable rg im Programm 2) und MAD : median absolute deviation about the median:

MAD = $med_i(|x_i - med_j(x_j)|)$, wobei

- med_i : Median der n Beobachtungen und
- med_j : Median der Absolutwerte der Abweichungen vom Median.

MAD ist ein sehr robuster Schätzer, der für Daten aus normalverteilter Grundgesamtheit mit $1.4826 \cdot \text{MAD}$ herangezogen werden kann, um die Standardabweichung σ zu schätzen (Variable rmad im Programm 2).

Diese beiden Maßzahlen werden in der letzten DATA-Anweisung von Programm 2 aufbereitet. Die robusten Schätzer rg und rmad sind im Output 3 im Vergleich zu s, dem Schätzwert für die Standardabweichung σ angegeben, die mit $\sigma = 3$ zur zufälligen Auswahl der Werte im Programm 2 vorgegeben wurde.

Output 3: Die robusten Schätzwerte mit GINI und MAD

mean	g	mad	s	rg	rmad
3.17017	3.32758	1.88356	2.95565	2.94899	2.79257

2.4 PROC TTEST

Die bisher übliche Form des Aufrufs der SAS/STAT-Prozedur TTEST liefert einen veränderten Output. Bisher wurden für die beiden Stufen der CLASS-Variablen die Anzahl, der Mittelwert, die Standardabweichung und der Standardfehler des Mittelwertes ausgegeben. Hinzu kommen nun standardseitig die 0,95-Konfidenzintervalle für die Mittelwerte und die Standardabweichungen. Alle diese Maßzahlen werden zusätzlich auch für die Differenzen zwischen den beiden Stufen der CLASS-Variablen – jeweils in der vorliegenden Reihenfolge der Werte – gebildet. Dabei ist allerdings zu beachten, dass bei der Berechnung der Differenz ein Fehlwert durch den Mittelwert der entsprechenden Stufe ersetzt wird. Die anderen Ausgaben entsprechen der bisherigen Form des t-Testes für unabhängige Stichproben, d. h. für die Annahme gleicher Varianzen () wird für die beiden Stufen der CLASS-Variablen der t-Test und für die Annahme ungleicher Varianzen () der Welch-Test gerechnet.

Für alle Fragestellungen des t-Testes kann eine zusätzliche Konstante zum Testen der Nullhypothese(n) $H_0: m = m_0$ vorgegeben werden, die standardseitig Null ist. Dazu wird im Unterschied zur Prozedur UNIVARIATE die Option $H_0=$ genutzt. Die Wirkung dieser Option soll am Beispiel des t-Testes für ein Einstichprobenproblem gezeigt werden. Der Einstichproben-t-Test wird realisiert, indem die CLASS-Anweisung weggelassen wird. Mit dem Programm 3 werden die Mittelwerte der Variablen Wert1, Wert2 und Diff, die paarweise Differenz zwischen den beiden Variablen, mit der Konstanten 13 verglichen. Neben den genannten statistischen Maßzahlen sind die Testergebnisse im Output 4 aufgeführt. Sie zeigen, dass sich die Mittelwerte der Variablen Wert1 und Wert2 nicht signifikant von der Konstanten 13 unterscheiden, wenn von einer Irrtumswahrscheinlichkeit $\alpha=0.05$ zugrunde gelegt wird. Die mittlere Differenz hingegen ist signifikant von der Konstanten 13 verschieden.

Programm 3: t-Test für Einstichprobenprobleme

```
data sas5;
  input wert1 wert2 @@;
  diff = wert1-wert2;
lines;
  12 15   14 13   13 17   11 14   13 12   12
  16
;
proc ttest data=sas5 h0=13;
  var wert1 wert2 diff;
run;
```

Output 4: Einstichproben-t-Tests im Vergleich mit der Konstanten 13

The TTEST Procedure							
Statistics							
Variable Std Err	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev
wert1 0.4282	6	11.399	12.5	13.601	0.6547	1.0488	2.5723
wert2 0.7638	6	12.537	14.5	16.463	1.1678	1.8708	4.5884
diff 0.9661	6	-4.483	-2	0.4834	1.4771	2.3664	5.8039

T-Tests			
Variable	DF	t Value	Pr > t
wert1	5	-1.17	0.2956
wert2	5	1.96	0.1067
diff	5	-15.53	<.0001

Der paarweise t-Test muss nicht über die Differenzenbildung vorgenommen, sondern kann direkt realisiert werden. Dafür wird wie im Programm 4 ersichtlich die Anweisung PAIRED anstelle der Anweisung VAR genutzt. Die einfachste Form der Variablen-Liste der PAIRED-Anweisung ist die Verwendung des Doppelpunktes (Programm 4). Andere Formen sind möglich. Das im Programm 4 angegebene Beispiel bezieht sich auf den Datensatz SAS5 aus dem Programm 3. Das Testergebnis (Output 5) lautet, dass sich die mittlere Differenz zwischen den beiden Variablen Wert1 und Wert2 bei einer Irrtumswahrscheinlichkeit $\alpha=0.05$ nicht signifikant von Null unterscheidet.

Programm 4: t-Test für gepaarte Stichproben

```
proc ttest data=sas5;
  paired wert1 : wert2;
run;
```

Output 5: Ergebnis des paarweisen t-Testes

The TTEST Procedure							
Statistics							
Difference Std Err	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev
wert1 - wert2 0.9661	6	-4.483	-2	0.4834	1.4771	2.3664	5.8039

T-Tests			
Difference	DF	t Value	Pr > t
wert1 - wert2	5	-2.07	0.0932

2.5 PROC DATASETS

Eine andere Art der Statistik als die zur Analyse von Versuchsdaten nutzt die neue SAS/BASE-Prozedur DATASETS. Viele bisher nicht zugriffsfähige Informationen über eine Datei und ihre Variablen können in der Version 8 abgerufen und verändert werden. In der Regel wird man diese Eigenschaften für permanent abgespeicherte SAS-Dateien und nicht für SAS-Dateien, die nur während einer Sitzung und für eine Sitzung angelegt wurden, nutzen. Darüber hinaus können mit dieser mächtigen Prozedur nicht nur verschiedene Datei-Eigenschaften aufgelistet und verändert, sondern auch Dateien kopiert, umbenannt, repariert, gelöscht und angefügt werden. Am Beispiel der für den t-Test verwendeten Datei SAS5 wird im Programm 5 gezeigt, wie die Informationen über eine Datei und ihre Variablen im der Anweisung CONTENTS aufgelistet und in eine neue Datei CONT_SAS5

gespeichert werden. Da das einfach nachvollziehbar ist, wird auf weitere Details verzichtet.

Programm 5: PROC DATASETS für das Abspeichern der Eigenschaften der temporären Datei SAS5

```
proc datasets nolist;
  contents data=sas5 out=cont_sas5;
proc print data=cont_sas5 noobs;
run;
```

2.6 PROC NLMIXED

Neu ist die SAS/STAT-Prozedur NLMIXED. Ein einfaches Beispiel, das die Anwendung dieser Prozedur demonstriert, ist in der Sammlung von SAS Programmen und Applikationen zur Prozedur NLMIXED zu aufgeführt. Diese Sammlung findet man über

„Help“ @ „SAS System Help“ @ „Sample SAS Programs and Applications“ @ „? Sample SAS Programs and Applications“.

In dem Fenster „SAS System Library“ wählt man SAS/STAT und dort „Documentation Examples for PROC NLMIXED“. Zunächst werden diese Daten, es sind Wachstumsdaten der Umfänge von 5 Baumstämmen (SAS/STAT User’s Guide 1999), unter Verwendung der Prozedur NLIN an die nichtlineare Funktion

$$y = b_1 / (1 + \exp[-(\text{day} - b_2) / b_3]) + e$$

angepasst (Programm 6). Geschätzt werden die Koeffizienten mit $b_1 = 192.7$, $b_2 = 728.8$ und $b_3 = 353.5$.

Programm 6: Die Daten aus Beispiel 1 der SAS-Beispielssammlung zu PROC NLMIXED ausgewertet mit PROC NLIN

```
data tree;
  input tree day y @@;
  datalines;
1 118 30 1 484 58 1 664 87 1 1004 115 1 1231 120 1 1372 142 1 1582 145
2 118 33 2 484 69 2 664 111 2 1004 156 2 1231 172 2 1372 203 2 1582 203
3 118 30 3 484 51 3 664 75 3 1004 108 3 1231 115 3 1372 139 3 1582 140
4 118 32 4 484 62 4 664 112 4 1004 167 4 1231 179 4 1372 209 4 1582 214
5 118 30 5 484 49 5 664 81 5 1004 125 5 1231 142 5 1372 174 5 1582 177
;

proc nlin data=tree;
  parms b1=190 to 200 by 5
        b2=700 to 750 by 10
        b3=330 to 350 by 10;
  model y = b1 / (1 + exp(-(day-b2)/b3));
run;
```

Wird nun diese Wachstumsfunktion um eine weitere Zufallsvariable u_1 erweitert:

$$y = (b_1 + u_1) / (1 + \exp[-(\text{day} - b_2) / b_3]) + e = \text{num} / \text{den} + e ,$$

dann ist eine Anpassung mit der Prozedur NLMIXED möglich. Diese leistungsfähige Prozedur analysiert auf der Grundlage der Maximum-Likelihood-Schätzung gemischte Modelle, d. h. Modelle mit festen und zufälligen Effekten. Diese Effekte gehen nichtlinear in das Modell ein. Im Programm 7 (SAS-Beispielssammlung zu PROC NLMIXED) wird die Realisierung für die Beispieldaten gezeigt. Die neuartige Schreibweise der MODEL- und der RANDOM-Anweisungen fällt auf. Die Tilde ~ zeigt an, dass die Variable der darauf folgenden Verteilung folgt. Die Wachstumsdaten entstammen einer normalverteilten Grundgesamtheit mit dem Erwartungswert num/den und der Varianz $s2e = 60$ und die Zufallsvariable u_1 soll normalverteilt sein mit dem Erwartungswert Null und der Varianz $s2u = 1000$.

Programm 7: PROC NLMIXED aus Beispiel 1 der SAS-Beispielssammlung

```
proc nlmixed data=tree;
  parms b1=190 b2=700 b3=350 s2u=1000
  s2e=60;
  num = b1+u1;
  ex = exp(-(\text{day}-b2)/b3);
  den = 1 + ex;
  model y ~ normal(num/den,s2e);
  random u1 ~ normal(0,s2u) subject=tree;
run;
```

Die Schätzungen für die Koeffizienten sind nun $b_1 = 192.05$, $b_2 = 727.90$ und $b_3 = 348.07$ und für die Varianzen $s2e = 61.5139$ und $s2u = 999.88$. Berechnet werden auch die Statistiken Log-Likelihood, AIC, AICC und BIC zum Vergleich verschiedener linearer Modelle. Das Ergebnis ist grafisch in der Abbildung 3 veranschaulicht. Die dünn und gestrichelt gezeichneten Linien sind die Wachstumskurven der einzelnen Bäume und die dick gezeichnete Kurve ist die Funktion, die sich mit den geschätzten Parametern für b_1 , b_2 und b_3 ergibt.

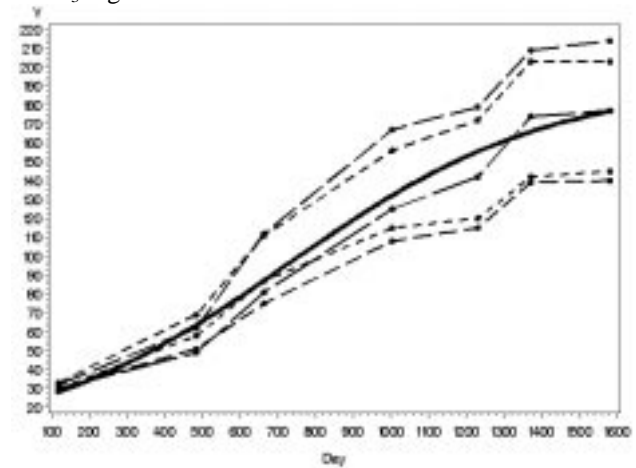


Abb. 3: Anpassung einer nichtlinearen, normalverteilten Funktion an Wachstumsdaten

Die Prozedur NLMIXED kann aber noch mehr. In der MODEL-Anweisung sind als Verteilungen wählbar:

- normal(m,v) Normalverteilung mit den Parametern Mittelwert m und Varianz v,
- binary(p) binäre Verteilung (Bernoulli-) mit dem Parameter Wahrscheinlichkeit p,
- binomial(n,p) Binomialverteilung mit den Parametern Anzahl n und Wahrscheinlichkeit p,
- gamma(a,b) Gammaverteilung mit den Parametern a und b,
- negbin(n,p) negative Binomialverteilung mit den Parametern Anzahl n und Wahrscheinlichkeit p,
- poisson(m) Poisson-Verteilung mit dem Parameter Mittelwert m,
- general(l1) allgemeine log-Likelihood-Funktion, die durch die SAS-Programmanweisungen definiert wird.

Der Nutzer kann auch programmseitig Verteilungen vorgeben, die in die Auswertung einbezogen werden.

2.7 PROC TREE

Die grafische Darstellung der Dendrogramme als Ergebnis einer Clusteranalyse war mit der SAS/STAT-Prozedur TREE unter SAS 6.12 kaum mehr als eine Notlösung, die dazu führte, von Nutzern geschriebene Macros zu verwenden. Auch PROC TREE ist von der Umstellung auf hochauflösende Grafiken betroffen. Folglich beziehen sich auch die meisten der Optionen auf die grafische Gestaltung. Wie das Programm 8 zeigt, ist die Verwendung einfach. Wer allerdings die standardseitig eingestellte Schriftart nicht verwenden will, muss aufpassen. Wenn beispielsweise die Schriftart Swiss verwendet wird, wird die Beschriftung der Ordinatenachse unterdrückt. Im Log-Fenster erscheint dann der Hinweis, dass nicht genügend Platz vorhanden ist, um für das verwendete average linkage-Verfahren die Bezeichnung „Average Distance Between Clusters“ entlang der Achse zu schreiben. Die Drehung durch die Winkelangabe in der Anweisung AXIS1 (Programm 8), die mit der Option VAXIS= der Prozedur TREE gekoppelt wird, schafft schnell Abhilfe (Abb. 4). Dass eine deutschsprachige Ersetzung einfach ist, lässt die kommentierte (*) AXIS1-Anweisung erkennen.

Programm 8: Beispiel für die Anwendung von PROC TREE

```
data clust;
  input Objekt $ m1-m15;
datalines;
Objekt1 0 1 1 2 0 1 1 2 1 1 1 1 2 0 1
Objekt2 0 3 1 2 0 7 2 1 1 1 1 2 1 0 1
Objekt3 0 3 0 1 1 0 1 1 1 2 1 1 2 0 2
Objekt4 0 1 2 1 1 1 2 3 1 1 0 2 1 1 2
Objekt5 0 2 1 0 1 1 2 2 1 3 1 2 3 1 0
Objekt6 1 0 0 0 0 0 1 1 1 0 6 2 1 0 1
;
proc cluster method=average outtree=dendro
standard;
  var m1-m15;
  id objekt;
run;
options htext=1.3 ftext=swiss;
axis1 label=(a=90);
*axis1 label=(a=90 "mittlerer Abstand");
proc tree data=dendro lines=(color=blue
width=5) vaxis=axis1;
  id objekt;
run;
quit;
```

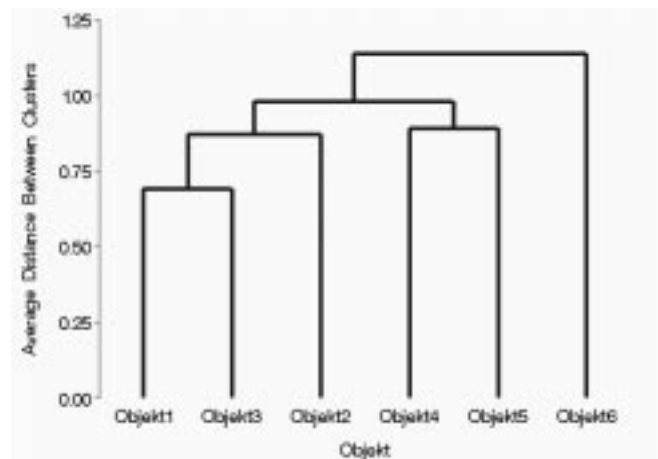


Abb. 4: Beispiel für die Dendrogramm-Darstellung

2.8 PROC MIXED

Die SAS/STAT-Prozedur MIXED ist eine relativ junge Prozedur und dient der Analyse linearer Modelle mit zufälligen und fixen Effekten. Mit dem durchgängigen Übergang auf das ODS (Output Delivery System) ist die MAKE-Anweisung als Überbleibsel noch vorhanden. Allerdings wird die Option NOPRINT nicht mehr unterstützt und Ihre Anwendung führt zum Abbruch der Prozedur.

Die METHOD=Option der PROC-Anweisung hat analog zur Prozedur GLM die drei zusätzlichen Wahlmöglichkeiten TYPE1, TYPE2 und TYPE3 der Quadratsummenzerlegung erhalten.

Die DDFM-Option in der MODEL-Anweisung wurde um die Wahlmöglichkeit KENWARDROGER erweitert. Diese verstärkt Anwendung findende Option zur Berechnung der Freiheitsgrade basiert auf der Methode von Kenward und Roger (1997) und umfasst zusätzlich eine Korrektur der Unterschätzung der Schätzfehler der festen Effekte.

Leider ist der besonders für unbalancierte, gemischte Modelle sehr zeitraubende Algorithmus zur Schätzung der Varianzkomponenten mit Hilfe der REML- (Restricted Maximum Likelihood) Methode noch nicht gegen einen schnelleren ausgetauscht worden.

2.9 PROC GENMOD

Die SAS/STAT-Prozedur GENMOD hat als zusätzliche Option RORDER=. Analog zur Option ORDER=, die sich auf die Reihenfolge der Stufen der CLASS-Variablen bezieht, bestimmt RORDER= die Reihenfolge der Stufen der Response-Variablen. Auch die DESCENDING-Option bezieht sich auf die Response-Variable und ordnet die Ausgabe deren Stufen von der höchsten zur niedrigsten.

Neu hinzugekommen sind die Anweisungen ESTIMATE, LSMEANS und OUTPUT. Ähnlich wie die CONTRAST-Anweisung testet ESTIMATE die Hypothese $L'b=0$, wobei L keine Matrix, sondern ein Zeilen-Vektor ist. Die Optionen sind ALPHA= für die Angabe des Konfidenzniveaus, E für die Ausgabe der Koeffizienten des Vektors L und EXP für die Berechnung von $\exp(L'b)$, dem Standardfehler und den

Konfidenzintervallgrenzen. Mit Hilfe der LSMEANS-Anweisung werden die adjustierten Mittelwerte (least squares means) für die Effekte des linearen Prediktors berechnet. Wie in den Prozeduren Glm oder Mixed können nach den Effekten getrennt durch einen Schrägstrich Optionen folgen. Diese sind ALPHA= für die Angabe des Konfidenzniveaus, CL berechnet die (1-a)-Konfidenzintervalle (Standard: 0.95-Konfidenzintervalle), CORR liefert die geschätzte Korrelationsmatrix und COV die geschätzte C für die LsMeans, E gibt die Koeffizienten der Matrix L für alle Effekte aus und DIFF berechnet alle Differenzen zwischen den LsMeans, dazu die Standardfehler und einen Wald-c2-Test. Mit der OUTPUT-Anweisung werden die Eingabedaten und optional verschiedene, durch Schlüsselwörter anzugebende Schätzwerte in eine SAS-Datei geschrieben.

Von den hinzugekommenen Optionen der MODEL-Anweisung sollen nur CL (Berechnung der Konfidenzintervalle), EXPECTED (nutzt die erwartete Fisher-Matrix für die Berechnung der Kovarianzen und verwandter Statistiken) und PREDICTED (gibt die Vorhersagewerte und weitere Statistiken aus) genannt werden. Bei den Verteilungsfunktionen (DIST=) können zusätzlich die multinomiale und die negative Binomialverteilung eingesetzt werden. Die Linkfunktionen (LINK=) wurden um die kumulative komplementäre Log-log-Funktion, die kumulative Logit- und die kumulative Probit-Funktion erweitert.

Wie bei PROC MIXED wird auch in der Prozedur GENMOD in der MAKE-Anweisung die Option NOPRINT nicht mehr unterstützt. Einige neue Optionen gibt es auch in der REPEATED-Anweisung, auf die aber nicht weiter eingegangen werden soll.

2.10 PROC FREQ

Um die TEST-Anweisung wurde die SAS/STAT-Prozedur FREQ erweitert. Mit ihr werden Abhängigkeits- und Assoziationsmaßzahlen (einfacher und gewichteter kappa-Koeffizient, Gamma, Kendall's tau-b, Pearson-Korrelationskoeffizient, Spearman-Korrelationskoeffizient, Somer's D (C|R), Somer's D (R|C) und Stuart's tau-c) berechnet und deren asymptotische Tests durchgeführt.

Als neue Optionen in der TABLE-Anweisung sind besonders AGREE(WT=FC) (Wichtung der kappa-Statistik nach Fleiss-Cohen), BINOMIAL (binomiale Proportion mit asymptotischem Test), CONTENTS= (Text für den Link der HTML-Seite), FORMAT= (formatierte Ausgabe je Zelle) und SCOROUT (Ausgabe der Zeilen- und Spalten-Scores) zu nennen.

Die Berechnung exakter Statistiken mit der Anweisung EXACT ist zeitaufwendig. Um das etwas abzufangen, wurden die Option MAXTIME=value zur Vorgabe der Maximalzeit in Sekunden für die Berechnung der exakten p-Werte und die Option MC zur Monte Carlo Schätzung der exakten p-Werte eingeführt. Mit der MC-Option korrespondieren darüber hinaus die Optionen ALPHA=p (Konfidenzniveau 1-p für die Schätzungen der Monte Carlo p-Werte [Standard: p = 0.01]), N=n (Anzahl der Stichproben n für die Monte Carlo

Schätzung) und SEED=n (Startwert für die Monte Carlo Schätzung).

2.11 PROC NPAR1WAY

Die SAS/STAT-Procedure NPAR1WAY, die nichtparametrische Testverfahren zum Vergleich von Mittelwerten umfasst, ist um eine Vielzahl von Verfahren, besonders um exakte Verfahren ergänzt worden. Diese Prozedur umfasst für Lageparameter und deren Differenzen folgende Tests auf der Grundlage von Scores: Wilcoxon-, Median-, van der Waerden-, Savage-, Siegel-Tukey-Scores-, Ansari-Bradley-, Klotz- und Mood-Test, für die sowohl der asymptotische als auch der exakte p-Wert mit der Anweisung EXACT ausgegeben werden können. Auch hier ist der Hinweis, dass die Berechnung der exakten Statistiken sehr zeitaufwendig ist, ernst zu nehmen. Wie bei der Prozedur FREQ sind deshalb auch die Optionen MAXTIME=value und MC - letztere mit korrespondierenden Optionen (s. o.) – eingeführt worden.

Weiterhin berechnet NPAR1WAY auch die Kolmogorov-Smirnov Statistik, die Cramer-von-Mises Statistik und wenn die Daten von genau zwei Stichproben vorliegen, die Kuiper Statistik. Hierfür werden nur die asymptotischen p-Werte berechnet.

2.12 Bemerkungen zu den Prozeduren TPSPLINE, GAM, MI und MIANALYSE

Die neue SAS/STAT-Procedure TPSPLINE paßt nichtparametrische Regressionsmodelle mit Straffunktion nach der Methode der kleinsten Quadrate an. Parametrische Modellannahmen sind nicht notwendig. Die Art der Glättung kann durch die GCV- (generalized cross validation) Funktion bestimmt werden.

Die SAS/STAT-Prozedur GAM ist in SAS 8.1 erst experimentell verfügbar. Details der experimentellen Dokumentation sind unter <http://www.sas.com/rnd/app/papers/gam.pdf> zu finden. Nach STRÜBY (2000) dient die Prozedur GAM der Anpassung verallgemeinerter additiver Modelle nach Hastie und Tibshirani (1990). Ihre Grundlagen sind die nichtlineare Regression und Glättungstechniken. Sie berechnet nichtparametrische Schätzungen für die additiven Modelle. Ihre Syntax ist mit der anderer Regressions-Prozeduren vergleichbar.

Auch die SAS/STAT-Prozedur MI ist erst experimentell vorhanden. Hinweise enthält die experimentelle Dokumentation <http://www.sas.com/rnd/app/papers/mi.pdf>. Die Zielstellung der Prozedur MI besteht nach Strüby (2000) darin, eine Ausgabedatei mit m multiplen Imputationen der Originaldatei für die Analyse mit fehlenden Werten zu erzeugen. Fehlende Werte werden in jeder Version durch berechnete (plausible) Werte ersetzt. Dafür werden verschiedene Algorithmen angeboten. Die einfachste Ersetzung fehlender Werte ist die durch den Mittelwert (Strüby 2001). Für die p-dimensionale Imputation sind die 3 Methoden Regressions-, Neigungs-Score- und Markovketten-Monte-Carlo-Methode (MCMC) möglich. „Für die Markovketten-Monte-Carlo-Methode (MCMC) kann ausgewählt werden, ob eine einzige Kette für alle m Imputationen oder eine separate Kette für

jede Imputation verwendet werden soll“ (Strüby 2000). Anfangsschätzungen für die MCMC-Methode können vorgegeben werden. Die imputierten Daten können mit Standard-Prozeduren analysiert werden. Die Ergebnisse können auch mit der SAS/STAT-Prozedur MIANALYZE kombiniert werden.

Die SAS/STAT-Prozedur MIANALYZE ist auch erst experimentell nutzbar. Details stehen unter <http://www.sas.com/rnd/app/papers/mianalyze.pdf>. Mit PROC MIANALYZE werden die Ergebnisse von Analysen auf m Imputationen kombiniert und die Qualität der Kombinationen (statistisch) bewertet. PROC MIANALYZE „nutzt als Input entweder eine speziell strukturierte Datei aus Parameterschätzungen und die Kovarianzmatrix jeder imputierten Datei oder ein Paar entsprechend strukturierter Dateien“ (STRÜBY 2000).

2.13 Zu SAS/INSIGHT

Die grafischen Elementen von SAS/INSIGHT umfassen unter SAS 6.12 die Darstellungsmöglichkeiten Histogramm / Bar Chart (Y), Box Plot / Mosaic Plot (Y), Line Plot (Y X) und Rotation Plot (Z Y X).

Hinzu gekommen ist das Contour Plot (Z Y X).

Von den multivariaten Verfahren [Multivariate (YX)] war bisher die Hauptkomponentenanalyse realisiert. Genutzt werden können nun noch:

- die kanonische Korrelationsanalyse,
- die Maximum-Redundanz-Analyse,
- die kanonische Diskriminanzanalyse und
- statistische Maßzahlen.

2.14 Zur Analyst Application

Die SAS Analyst Application ist eine interaktive Anwendung, die sowohl grafische als auch statistische Verfahren – sowohl zur Versuchsplanung als auch –auswertung – beinhaltet. Vorteilhaft ist, dass erzeugte SAS-Programme auch bereit gestellt werden und für weitere Auswertung angepasst werden können.

Die SAS Analyst Application ist zu finden unter „Solutions“ @ „Analysis“ @ „Analyst“. Einen prinzipiellen Unterschied zwischen der Anwendung unter SAS 6.12 und unter SAS 8.1 gibt es nicht. Das Zeichnen der Boxplots wird anstelle der bisher dafür eingesetzten SAS/QC-Prozedur SHEWHART mit der SAS/STAT-Prozedur BOXPLOT vorgenommen. Die grafischen Darstellungen (Graphs) sind um die Tortendiagramme (Pie Charts...) erweitert worden. Die statistischen Verfahren sind unter „Statistics“ vereint. Hinzugekommen sind unter „ANOVA“ Nonparametric One-Way ANOVA... berechnet die nichtparametrischen Statistiken und Tests für einfaktorielle Anlagen, wobei die Prozedur NPAR1WAY (s. o.) eingesetzt wird, Mixed Models... legt die Prozedur MIXED zugrunde.

Neu sind als multivariate Verfahren die Hauptkomponentenanalyse und die kanonische Korrelation und als Erhebungs-

verfahren die Sterbetafeln und das proportionale Ausfallmodell.

3 Neue Möglichkeiten des Zugriffs auf externe Datenquellen

3.1 SAS Import/Export Wizard

Der aus SAS 6.12 bekannte Import und Export der PC-Standardformate Excel, dBASE über einen Assistenten, den Import- bzw. Export-Wizard, ist um MS-Access erweitert worden. Diese Möglichkeit ist sehr einfache und kann über „File“ @ „Import Data...“ bzw. „File“ @ „Export Data...“ realisiert werden.

Allerdings werden dabei gerade beim Import viele Kopien hergestellt, für deren Aktualisierung wieder einiger Aufwand nötig ist. Wesentlich besser ist der direkte Zugriff auf die Daten per ODBC. Der Vorteil besteht darin, dass immer die aktuellen, originalen Daten vorliegen.

3.2 ODBC-Verbindung zu externen Dateien

Bereits in SAS 6.12 funktionierte die Technik des Zugriff auf externe Daten recht gut. Es konnten für externe Datenquellen sogenannte Views definiert werden, die wie SAS-Datasets verwendet werden konnten.

Verbessert in SAS 8.1 ist die mögliche dauerhafte Verbindung einer ODBC-Datenquelle mit einer SAS-Library. Die ODBC-Verbindung muss nur ein einziges Mal hergestellt werden. Import und Export sind nicht mehr nötig. Es kann dann mit den externen Daten wie mit SAS Datasets gearbeitet werden.

Dazu ist zuerst der ODBC Deskriptor auf die externe Datenquelle zu erstellen. Das passiert außerhalb von SAS unter „Systemsteuerung“ @ „ODBC-Datenquellen (32Bit)“. Im ODBC-Administrator wird „System-DSN“ und dort „Hinzufügen...“ (bzw. „Add...“) gewählt. In dem sich öffnenden Fenster „Neue Datenquelle erstellen“ wird der zutreffende Treiber, beispielsweise Microsoft Access-Treiber (*.mdb), ausgewählt, und im dann folgenden Fenster „ODBC Microsoft Access Setup“ der Name für die Datenquelle festgelegt. Eine Beschreibung kann hinzu gefügt werden. Nunmehr liegen der Name und der Treiber für die Datenquelle fest. Die Zuordnung zur Datenbank wird mit dem Button „Auswählen...“ vorgenommen. Das ist auch bereits der letzte Schritt der Definition eines ODBC Diskriptors.

Was noch bleibt ist die Verbindung des ODBC Diskriptors mit SAS. Dazu wird im SAS-Explorer das Icon Libraries mit der rechten Maustaste angeklickt. Von den beiden Möglichkeiten wird „New...“ gewählt. Bei „Name“ wird eine Bezeichnung für die SAS-Library eingetragen. „Engine“ ist ODBC. Mit dem Eintrag der Datenbank so wie sie im „ODBC Microsoft Access Setup“-Fenster für die Datenquelle (s. o.) festgelegt wurde, erfolgt die Verbindung zwischen SAS-Library und beispielsweise der Access-Datenbank. Dauerhaft wird sie, wenn das Häkchen bei „Enable at startup“ gesetzt wird. User-ID und Password können eingetragen werden.

Vorteilhaft ist, dass die Daten nur einmal vorliegen – beispielsweise im Access-Format – und sie SAS-seitig wie SAS-

DataSets behandelt werden. Um die Imports und Versionen braucht man sich jetzt nicht mehr zu kümmern. Aufgrund der SAS-Namenskonvention muss aber darauf geachtet werden, dass keine Umlaute in Feldnamen vorliegen. Die Tabellen sind in der ODBC Standardeinstellung schreibgeschützt aber löschtbar!

Mit der Version 8 von SAS ist es unter SAS/ACCESS auch möglich, durch Verwendung dynamischer Libname Engines, die wie Optionen in die LIBNAME-Anweisung geschrieben werden, die Verbindung zu relationalen Datenbanken wie DB2, Oracle, Informix, Sybase und anderen herzustellen (Müller 1999).

3.3 ODBC-Verbindung zu SAS-Daten

Die meisten SAS-Prozeduren enthalten die Möglichkeit, ihren Output in SAS-Dateien auszugeben.

Um diese Daten zu verarbeiten, kann es sinnvoll sein, mit anderen Anwendungen z. B. MS-Excel oder MS-Access auf diese SAS Daten zuzugreifen. Dazu verwendet man am besten eine ODBC-Verbindung zu diesen Daten. Die Vorgehensweise ist auch in der SAS-OnlineDoc (HTML) unter Index SAS ODBC Driver beschrieben

Es sind einige wichtige Vorarbeiten nötig. Wenn nicht schon geschehen, ist der SAS ODBC V8.1 Treiber von der SAS-CD zu installieren. Weiterhin muss ein lokaler tcp/ip Service für den ODBC Treiber definiert werden. Das erfolgt in der Datei „services“. In Windows 9x liegt diese Datei unter C:\WINDOWS\SERVICES und in Windows NT unter C:\WINNT\SYSTEM32\DRIVERS\ETC\SERVICES. Um den Service einzurichten, muss in der Datei die Zeile **sasodbc 5010/tcp #Local ODBC Server** hinzugefügt werden.

Anschließend wird, wie im Abschnitt 3.2 beschrieben, ein ODBC-Deskriptor eingerichtet. Die Menüfolge zum Erstellen des ODBC Deskriptors für die SAS Datenquelle lautet:

Systemsteuerung ® ODBC-Datenquellen (32Bit). Im ODBC-Administrator wird „System-DSN“ gewählt und als neue Datenquelle SAS gewählt. Bei den Server-Einstellungen muss der Servername sasodbc aus der Services-Datei verwendet werden. Zu beachten ist, dass bei „Konfigurieren...“ unter SAS Settings der richtige SAS-Ordner eingestellt wird. Standardseitig wird c:\programms\sas\sas.exe angenommen, was nicht immer zutreffend ist.

Die Libraries Einstellungen enthalten den Ordner der Librefs der SAS Installation des Nutzers:

z. B. Name: sasuser Host File c:\meinsas .

Nach erfolgreicher Erstellung des ODBC-Deskriptors für SAS-Dateien kann jetzt mit jeder ODBC fähigen Anwendung auf SAS Daten zugegriffen werden.

4 Die Ausgabe von SAS-Ergebnissen mit dem Output Delivery System (ODS)

4.1 Was ist das ODS?

Prinzipiell ist eine SAS-Nutzung ohne das ODS (Output Delivery System) möglich. Dann muss aber auf (fast) jegliche Gestaltung der Ausgabe verzichtet und das Unzulängliche, was an der bisherigen Ausgabeform kritisiert wurde, akzeptiert werden. Mit dem Output Delivery System (ODS) wurde eine Schale geschaffen, die um alle SAS-Prozeduren liegt. Sie dient der Gestaltung der Ausgabe von Listen, Tabellen, Grafiken und anderem. Es werden Ausgabeobjekte durch ODS erzeugt und verwaltet, mit denen auch die Ausgabe in Formate wie HTML, RTF und Postscript möglich wird. Dazu gekommen sind Gerätetreiber zur Erzeugung von Java und ActiveX zum Einbinden von „lebendigen“ SAS-Grafiken. Allerdings funktionieren die neuen Gerätetreiber unter SAS 8.1 leider noch nicht für alle Grafikprozeduren.

Mit der Version 8 von SAS wird die Ausgabe in einzelne Abschnitte gegliedert. Diese neuartige Strukturierung der Ausgabe ist im Result-Fenster sichtbar. Alle dort aufgeführten Ausgabeteile Listen, Tabellen, Grafiken, Views und anderes können direkt durch Anklicken erreicht werden. Das ist mit den ODS-Anweisungen programmseitig einsetzbar, so dass der Nutzer die von ihm gewünschte Ausgabeform steuern kann. Die ODS-Anweisungen werden vor die betreffende SAS-Prozedur gesetzt. Mit ihnen werden Ausgabemedium (Dateien, Ordner usw.), Format und Ausgabeobjekte definiert. Um die ODS-Anweisungen gezielt einsetzen zu können, müssen die für jede Prozedur unterschiedlichen Ausgabeabschnitte und damit die Namen der ODS-Tabellen bekannt sein.

Um die Namen der ODS-Tabellen heraus zu finden, stehen zwei Wege zur Verfügung. Einmal kann man über die Hilfe zu jeder SAS/STAT-Prozedur unter dem Punkt „Additional Topics“ die „ODS Table Names“ finden, wo alle Bezeichnungen der ODS-Tabellen und die der Variablen aufgelistet sind. Leider sucht man das für STAT/BASE-Prozeduren wie beispielsweise MEANS oder UNIVARIATE vergebens. Der andere Weg ist, sich die für jede Prozedur verschiedenen ODS-Objekte mit Hilfe von ODS-Anweisungen anzeigen zu lassen. Dazu werden um den eigentlichen Prozedur-Aufruf wie eine Klammer die Anweisungen ODS TRACE ON bzw. ODS TRACE OFF geschrieben. Um das zu demonstrieren, wird der Prozedur-Aufruf des Programms 3 (Programm 9) genutzt. Im Log-Fenster werden die Namen der ODS-Objekte und andere Informationen (Output 6) mitgeteilt. Wird der Anweisung ODS TRACE ON nach einem Schrägstrich die Option LISTING hinzugefügt, dann werden die im Output 6 aufgeführten Informationen ins Output-Fenster vor jedes der ODS-Objekte geschrieben. Für das kleine Beispiel (Programm 6) sind das die ODS-Objekte „Statistics“ und „Ttests“.

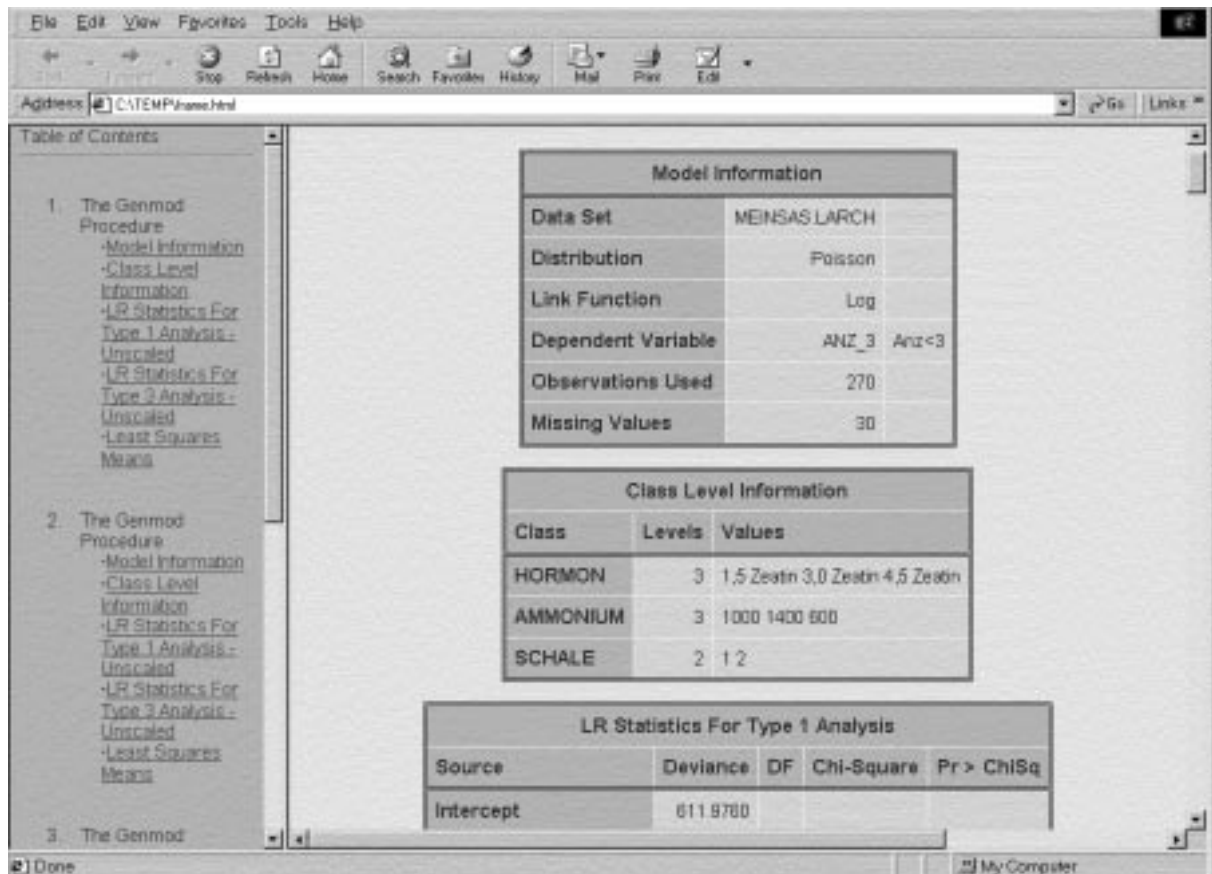


Abb. 5: Beispiel für eine HTML-Ausgabe

Programm 9: Beispiel für das Auflisten der ODS-Objekte

```
ods trace on;
proc ttest data=sas5 h0=13;
  var wert1 wert2 diff;
run;
ods trace off;
```

Output 6: Informationen zu den ODS-Objekten im Log-Fenster

```
Output Added:
-----
Name:      Statistics
Label:     Statistics
Template:  Stat.TTest.Statistics
Path:     Ttest.Statistics
-----

Output Added:
-----
Name:      TTests
Label:     T-Tests
Template:  Stat.TTest.TTests
Path:     Ttest.Ttests
-----
```

Auf die Templates soll hier nur aufmerksam gemacht werden. Dort befinden sich Standardeinstellungen zur Schriftart, -größe, Farben und anderes mehr für den betreffenden Ausgabeabschnitt. Sollten diese für eine Anwendung geändert werden müssen, steht die Prozedur TEMPLATE bereit, die, wenn überhaupt, von fortgeschrittenen SAS Programmierern eingesetzt werden wird (Schwarz 1999).

4.2 Die ODS-Anweisungen

Um für die verschiedenen Formen einer Ausgabe bessere Möglichkeiten nutzen zu können, stehen die in der Tabelle 1 aufgeführten ODS-Anweisungen zur Verfügung (Schwarz 1999, Base SAS Software). Anhand nachfolgender Beispiele sollen einige der Ausgabeformen demonstriert werden.

Tabelle 1: Liste der ODS-Anweisungen

Ziel	Ausgabemedium	ODS-Anweisung
Verbindung zu den Ausgabemedien	SAS-Output-Fenster	ODS Listing
	HTML-Datei	ODS HTML
	SAS-DataSet	ODS Output
	für den Drucker aufbereitete Datei	ODS Printer
	RTF-Format-Ausgabe	ODS RTF
Auswahl von ODS-Objekten	Ausschluss von Objekten	ODS Exclude
	Auswahl von Objekten	ODS Select
Kontrolle der ins Log-Fenster geschriebenen Informationen	Steueranweisung für LOG	ODS Show
		ODS Trace
		ODS Verify
Setzen des Suchpfades für Templates	allgemeine Steueranweisung	ODS Path
Unterdrücken der Title-Zeilen der Prozedur		ODS Proctitle

4.3 Anwendungsbeispiele für ODS

4.3.1 Ausgabe nach HTML in Frames und Anpassen des Outputs

In dem im Programm 10 aufgeführten Beispiel werden die ODS-Objekte

Modellinformation	ModelInfo
Stufen der Klassifizierungsvariablen	ClassLevels
globale Tests Type1 Type3	Type1 Type3
Differenzen der Mittelwerte	LSMeanDiffs

der Prozedur GENMOD ausgewählt, um die Ausgabe auf die interessierenden Teile zu begrenzen. Das geschieht mit ODS SELECT. Für die Ausgabe nach HTML wird die ODS HTML-Anweisung herangezogen. Die Option PARAMETERS dieser Anweisung ist nicht notwendig. Sie realisiert aber einen weißen Hintergrund. Definiert werden dort auch die Zieldateien für die ODS-Ausgabe. Das Programm 10 erzeugt 3 HTML Dateien, die für die Frame-Darstellung des Outputs benötigt werden.

Programm 10: Beispiel für Programmieren einer HTML-Ausgabe

```
filename odsout 'c:\temp\'; /*Verzeichnis für
alle HTML Dateien*/;
ods html
  parameters = ("COLORSCHEME" = "Bright"
"BACKCOLOR" = "#FFFFFF")
  body = 'body.html'
  frame = 'frame.html'
  contents='contents.html'
  path=odsout
  nogtitle;

ods select ModelInfo ClassLevels Type1 Type3
LSMeanDiffs;

proc genmod data = meinsas.larch;
  class hormon ammonium schale;
  model anz_3= schale hormon ammonium hor-
mon*ammonium
        /dist=poisson link=log type1
type3;
  lsmeans hormon*ammonium /diff e;
run;

ods html close;
ods listing;
```

Eine übersichtliche Aufbereitung der Ausgabe wird durch Öffnen der Datei frame.html (Abbildung 5) erreicht.

4.3.2 Ausgabe nach MS-WORD und anderen Texteditoren

Für die Ausgabe in eine MS-WORD-Datei wird das RTF-Format benötigt. Im Programm 11 wird deshalb die ODS RTF-Anweisung eingesetzt, mit der gleichzeitig die Ausgabedatei festgelegt wird. Die Ausgabe ist im Output 7 zu sehen. Die RTF-Datei kann dann „ganz normal“ mit MS-WORD und anderen Texteditoren wie beispielsweise Word-Perfect und StarWriter bearbeitet werden.

Programm 11: Beispiel für das Programmieren einer RTF-Ausgabe

```
ods rtf file="c:\temp\word1.rtf";
ods select ModelInfo ClassLevels Type1 Type3
LSMeanDiffs;

title "Anzahl > 3 abhängig von hormon ammi-
um hormon*ammonium Konzentration";
proc genmod data = meinsas.larch;
  class hormon ammonium schale;
  model anz_30= schale hormon ammonium hor-
mon*ammonium
        /dist=poisson link=log type1
type3;
  lsmeans hormon*ammonium /diff e;
run;

ods rtf close;
ods listing;
```

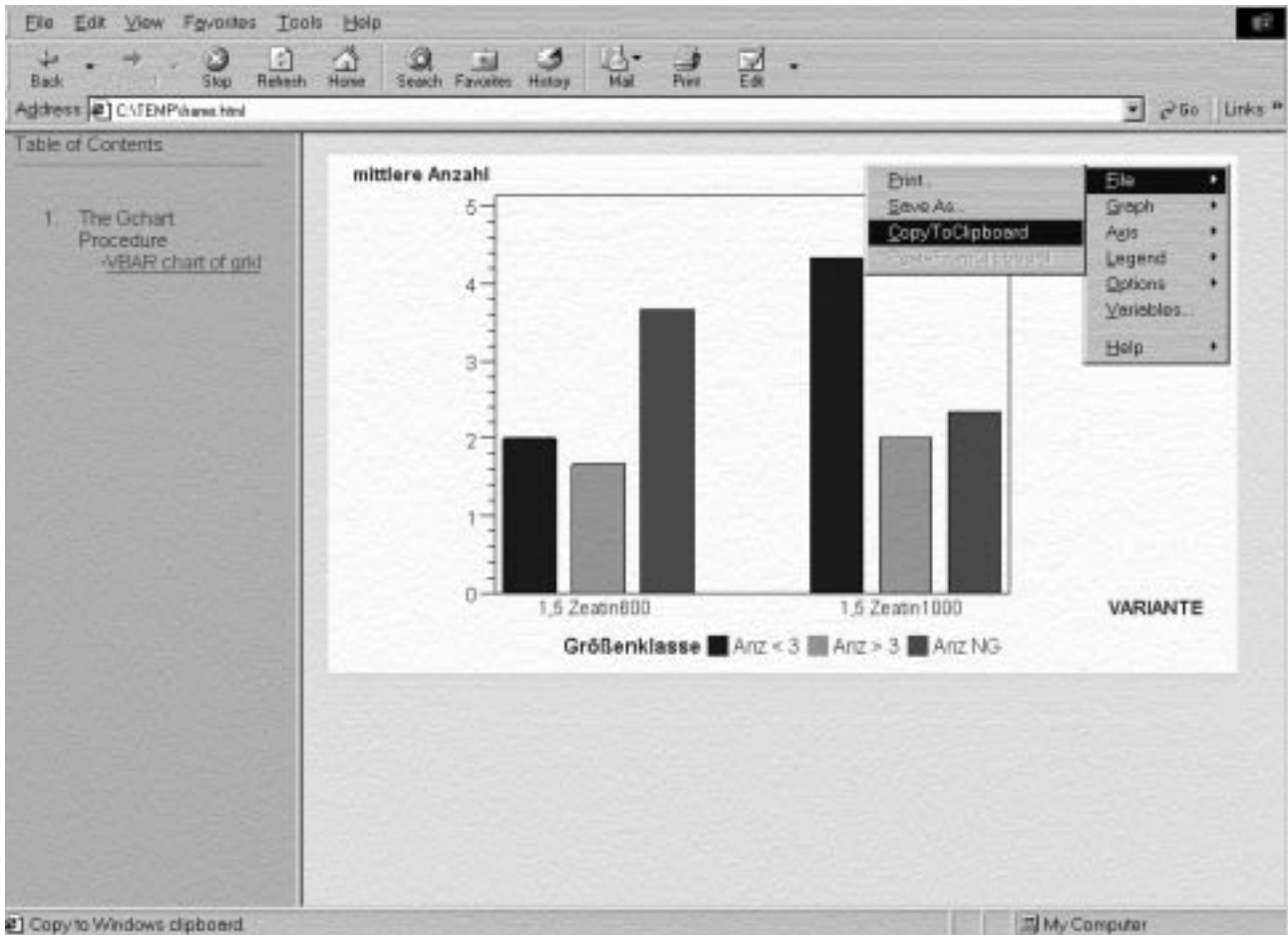


Abb. 6: Mit ActiveX unter SAS erzeugte Balkengrafik in HTML

4.3.3 Ausgabe in eine SAS-Datei

Mit der Anweisung ODS OUTPUT werden den ODS-Objekten SAS-Dateinamen zugeordnet. Diese Dateien nehmen dann die entsprechenden Informationen auf. Das Beispiel im Programm 12 zeigt leicht nachvollziehbar die Abspeicherung in SAS-Dateien und die Ausgabe der Mittelwertvergleiche auf gleicher Ammoniumstufe im RTF-Format. Das Ergebnis ist im Output 8 aufgeführt.

Programm 12: Beispiel für das Abspeichern in SAS-Dateien

```
ods select ModelInfo ClassLevels Type1 Type3
LSMeanDiffs;
ods output ModelInfo = modinfdat
ClassLevels = classldat
Type1 = typ1dat
Type3 = typ3dat
LSMeanDiffs = diffdat;

proc genmod data = meinsas.larch;
class hormon ammonium schale;
model anzng = schale hormon ammonium hor-
mon*ammonium
/dist=poisson link=log type1
type3;
lsmeans hormon*ammonium /diff e;
run;

ods rtf file="c:\temp\mdiff.rtf";
```

```
data amm;
set diffdat;
where Ammonium=_Ammonium;
proc sort data=amm;
by Ammonium _Ammonium;
run;

title "Vergleich der Differenzen auf gleicher
Ammoniumstufe";
proc print noobs;
var Effect Hormon Ammonium _Hormon Estimate
ProbChiSq;
run;

ods rtf close;
title;
```

Model Information		
Data Set	MEINSAS.LARCH	
Distribution	Poisson	
Link Function	Log	
Dependent Variable	ANZ_3	Anz<3
Observations Used	270	
Missing Values	30	

Class Level Information		
Class	Levels	Values
HORMON	3	1,5 Zeatin 3,0 Zeatin 4,5 Zeatin
AMMONIUM	3	1000 1400 600
SCHALE	2	1 2

LR Statistics For Type 1 Analysis				
Source	Deviance	DF	Chi-Square	Pr > ChiSq
Intercept	611.9760			
SCHALE	601.9320	1	10.04	0.0015
HORMON	596.9725	2	4.96	0.0838
AMMONIUM	568.6615	2	28.31	<.0001
HORMON*AMMONIUM	538.7603	4	29.90	<.0001

Output 7: Beispiel für eine RTF-Ausgabe: Auszug aus der Datei word.rtf

Output 8: Geschätzte Differenzen der mittleren Effekte (LsMeans) auf gleicher Ammoniumstufe im RTF Format (mdiff.rtf)

Vergleich der Differenzen auf gleicher Ammoniumstufe

Effect	HORMON	AMMONIUM	_HORMON	Estimate	ProbChiSq
HORMON*AMMONIUM	1,5 Zeatin	1000	3,0 Zeatin	0.9416	<.0001
HORMON*AMMONIUM	1,5 Zeatin	1000	4,5 Zeatin	0.5276	0.0013
HORMON*AMMONIUM	3,0 Zeatin	1000	4,5 Zeatin	-0.4140	0.0449
HORMON*AMMONIUM	1,5 Zeatin	1400	3,0 Zeatin	0.7084	<.0001
HORMON*AMMONIUM	1,5 Zeatin	1400	4,5 Zeatin	0.5008	0.0004
HORMON*AMMONIUM	3,0 Zeatin	1400	4,5 Zeatin	-0.2076	0.2137
HORMON*AMMONIUM	1,5 Zeatin	600	3,0 Zeatin	0.4873	0.0004
HORMON*AMMONIUM	1,5 Zeatin	600	4,5 Zeatin	-0.3416	0.0020
HORMON*AMMONIUM	3,0 Zeatin	600	4,5 Zeatin	-0.8289	<.0001

4.3.4 Ausgabe von Grafiken nach HTML

Die Ausgabe mit Hilfe der ODS-Anweisung erzeugt Grafiken, die sowohl in HTML als auch bei ihrer Abspeicherung in MS-WORD interaktiv bleiben, auch wenn SAS nicht mehr läuft. Das bedeutet, dass kleinere Änderungen wie Achsenbeschriftungen, Hintergrundfarben per Kontextmenü (rechte Maustaste) verändert werden können. Wenn die Maus über die Grafik bewegt wird, werden die Balkendaten mit einem PopUp angezeigt. Der im Programm 13 verwendete Treiber Activex erzeugt derart „lebendige“ Grafiken. Das Ergebnis, die Grafik in der HTML-Datei, zeigt die Abb. 6. Die Balkengrafik lässt sich leicht nach MS-WORD kopieren, indem die Maus auf dem Grafikrand positioniert und mit der rechten Maustaste „File“ @ „CopyToClipboard“ gewählt wird. In MS-WORD wird der Inhalt der Zwischenablage gespeichert.

Programm 13: Beispiel für das Programmieren einer Grafik in HTML-Ausgabe mit ActiveX

```
filename odsout 'c:\temp\'; /*Verzeichnis für
alle HTML Dateien*/;
ods html
  parameters = ("COLORSCHEME" = "Bright"
"BACKCOLOR" = "#FFFFFF")
  body      = 'body.html'
  frame    = 'frame.html'
  contents = 'contents.html'
  path     = odsout
  nogtitle;

goptions reset=all gunit=pct noborder
  dev=activex xpixels=700 ypixels=400
  htitle=3 htext=4;
  legend1 value=(h = 4 pct f=swissb) label=(h=4 pct f=swissb);

data graph;
  label grkl="Größenklasse"
  anz = "mittlere Anzahl";
  input VARIANTE $15. anz grkl $10.;
datalines;
1,5 Zeatin1000 3 Anz < 3
1,5 Zeatin1000 5 Anz < 3
1,5 Zeatin1000 5 Anz < 3
1,5 Zeatin1000 3 Anz > 3
1,5 Zeatin1000 2 Anz > 3
1,5 Zeatin1000 1 Anz > 3
1,5 Zeatin1000 3 Anz NG
1,5 Zeatin1000 2 Anz NG
1,5 Zeatin1000 2 Anz NG
1,5 Zeatin600 2 Anz < 3
1,5 Zeatin600 1 Anz < 3
1,5 Zeatin600 3 Anz < 3
1,5 Zeatin600 2 Anz > 3
1,5 Zeatin600 2 Anz > 3
1,5 Zeatin600 1 Anz > 3
1,5 Zeatin600 2 Anz NG
1,5 Zeatin600 5 Anz NG
1,5 Zeatin600 4 Anz NG
;
run;
proc gchart data=graph;
  vbar grkl/type=mean discrete
  sumvar=anz
  group =variante
  subgroup=grkl
  maxis=none raxis=axis2 gaxis=axis3
  space=0 gspace=8 width =6
  legend=legend1 ;
axis1 value=none label=(' ');
axis2 label=(h=4 pct f=swissb j=1);
axis3 value=( r=0 h= 4 pct f=swissb)
```

```
order=('1,5 Zeatin600' '1,5
Zeatin1000')
  label=(h=4 pct f=swissb j=1);
run; quit;

ods html close;
```

5 Literatur

Base SAS Software: What's New in Base SAS Software for Release 8.1, Online-Hilfe.

HASTIE, T.J. and Tibshirani, R.J. (1990): Generalized Additive Models. Monographs on Statistics and Applied Probability, 43, Chapman and Hall, London.

KENWARD, M. G. and Roger, J. H. (1997): Small sample inference for fixed effects from restricted maximum likelihood, Biometrics, 53, pp. 983-997.

MÜLLER, B. (1999): Neue Basisfunktionalitäten in der Version 8 des SAS Systems, Vortrag auf der DISK 1999, Ulm.

RIDDER, K. (2000): Neue Basisfunktionalitäten in der Version 8 des SAS Systems, Vortrag auf der 4. KSFE „Konferenz der SASÖ-Anwender in Forschung und Entwicklung“, Gießen.

SAS Procedures Guide, Version 8, SAS Institute Inc., Cary, NC, USA, 1999.

SAS/STAT User's Guide, SAS Online Doc Version 8 with PDF Files, SAS Institute Inc., Cary, NC, USA, 1999

SCHWARZ, H. (1999): Output Delivery System (ODS) - der schnelle Weg zu geschäftsfähigen Berichten, Vortrag auf der DISK 1999, Ulm.

STRÜBY, R. (2000): Neue Statistische Verfahren und interaktive Analysen in der Version 8e des SAS Systems Vortrag auf der Disk 2000, Berlin.

STRÜBY, R. (2001): Neue Statistische Verfahren und interaktive Analysen in der Version 8e des SAS Systems, Vortrag auf der 5. KSFE „Konferenz der SASÖ-Anwender in Forschung und Entwicklung“, Stuttgart.

Neues für den Anwender der SAS-Version 8 im Vergleich zu SAS 6.12 (E. Moll, T. Stauber)

Zusammenfassung

Die SAS-Version 8 löst die seit Jahren stabile Version 6.12 ab. Sie zeigt sich äußerlich durch eine neue Oberfläche und ein nutzerfreundlicheres Hilfesystem. Neue Prozeduren, Änderungen und Ergänzungen in bisherigen werden vorgestellt. Die in der SAS-Version 8 neuen Möglichkeiten des Zugriffs auf externe Datenquellen werden dargelegt. Das Output Delivery System (ODS) ist eine neue Form der Ausgabemöglichkeit, die sich an viele Ausgabemedien, so beispielsweise die Ausgabe in HTML- oder RTF-Format und in SAS-Dateien, unterstützt. Einige der früheren Probleme beim Im- und Export von Grafiken und Daten sind behoben.

Stichworte: SAS 8, ODBC, ODS

New instructions for users of the release SAS 8, compared to SAS 6.12 (E. Moll, T. Stauber)

Summary

Release SAS 8 is now replacing SAS 6.12, which has remained unchanged over several years. Externally, SAS 8 has a new surface and a more user-friendly help facility. New procedures, modifications and supplementary features to existing procedures are presented. SAS 8 offers new options for accessing external data sources. The Output Delivery System (ODS) is a new way of output which can use a number of output media, such as the HTML or the RTF format or SAS files. A number of old problems concerning the import and export of graphs and data have been solved.

Keywords: SAS 8, ODBC, ODS

Dr. Eckard Moll ist verantwortlich für das Sachgebiet Biometrie und SAS und arbeitet in der Zentralen EDV-Gruppe der Biologischen Bundesanstalt für Land- und Forstwirtschaft, Außenstelle Kleinmachnow (Stahnsdorfer Damm 81, 14532 Kleinmachnow, e-mail: e.moll@bba.de)

Thomas Stauber ist wissenschaftlicher Angestellter im Institut für Forstgenetik und Forstpflanzenzüchtung der Bundesforschungsanstalt für Forst- und Holzwirtschaft (Eberswalder Chaussee 3, 15377 Waldsiedersdorf, e-mail: stauber@holz.uni-hamburg.de)

Stellenangebote

Ankündigung

Doluschitz/Spilke (Hrsg.): Agrarinformatik.
UTB 2230 S, ISBN 3-8252-2230-6, Ulmer 2001.
Ca. 300 Seiten, ca. 36,80 DM, ca. Oktober.