

Neue Statistische Verfahren und interaktive Analysen in der Version 8e des SAS Systems

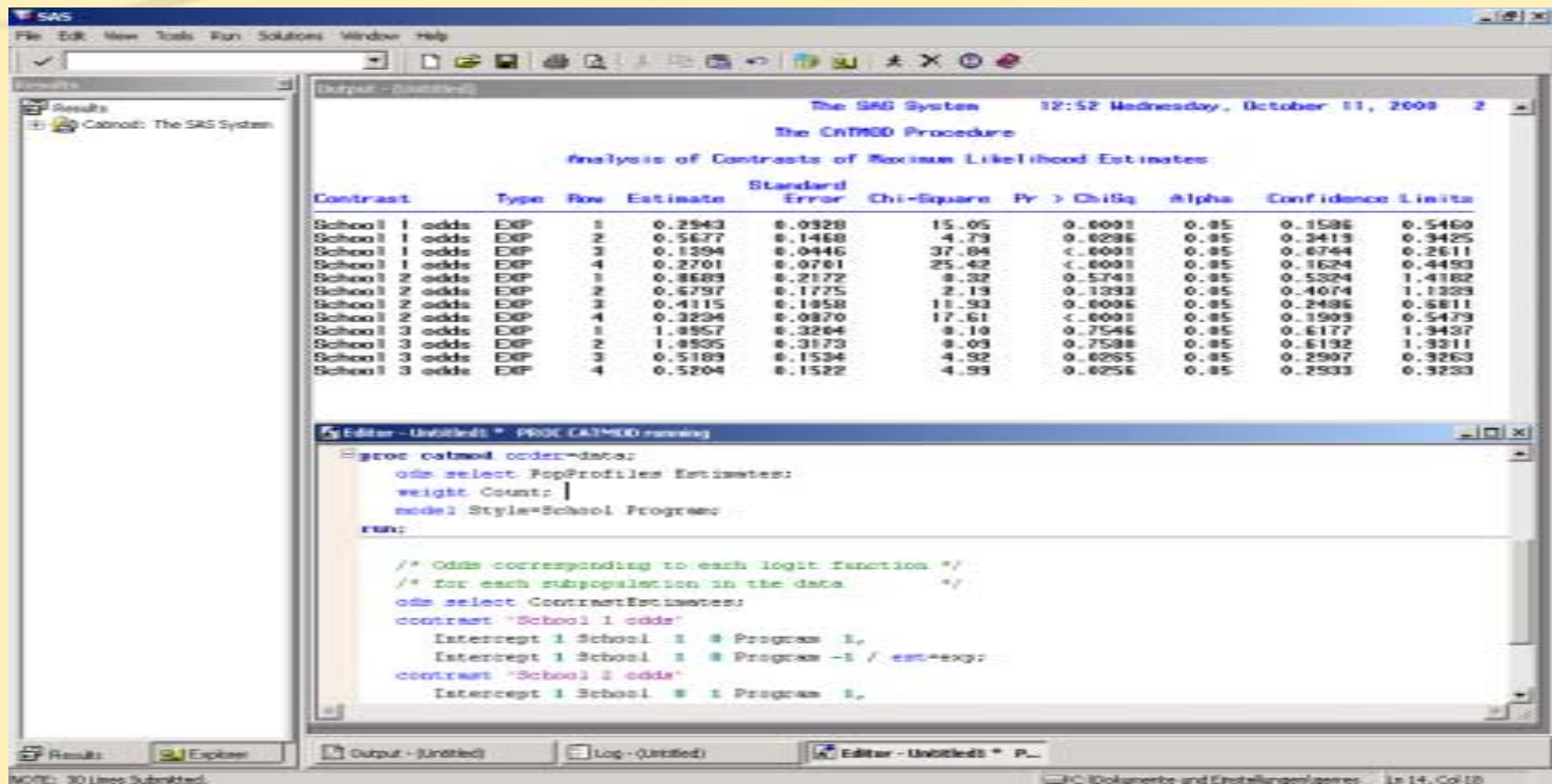
Dr. Reinhard Strüby

Business Competence Center

SAS Deutschland

The Power to Know.

SAS/STAT Software



The SAS System 12:52 Wednesday, October 11, 2000 2

The CATMOD Procedure

Analysis of Contrasts of Maximum Likelihood Estimates

Contrast	Type	Row	Estimate	Standard Error	Chi-Square	P > ChiSq	Alpha	Confidence Limits
School 1 odds	EXP	1	0.2943	0.0328	15.05	0.0001	0.05	0.1586 0.5460
School 1 odds	EXP	2	0.5677	0.1468	4.73	0.0286	0.05	0.2413 0.9425
School 1 odds	EXP	3	0.1394	0.0446	37.84	<.0001	0.05	0.6744 0.2611
School 1 odds	EXP	4	0.2701	0.0781	25.42	<.0001	0.05	0.1624 0.4453
School 2 odds	EXP	1	0.8683	0.2172	9.32	0.0241	0.05	0.5304 1.4182
School 2 odds	EXP	2	0.6797	0.1775	2.19	0.1393	0.05	-0.4074 1.1328
School 2 odds	EXP	3	0.4115	0.1858	11.93	0.0006	0.05	0.2486 0.6811
School 2 odds	EXP	4	0.3234	0.0870	17.61	<.0001	0.05	0.1593 0.5479
School 3 odds	EXP	1	1.8957	0.3264	9.10	0.0246	0.05	0.6177 1.9437
School 3 odds	EXP	2	1.8935	0.3173	9.09	0.0258	0.05	0.6132 1.9311
School 3 odds	EXP	3	0.5189	0.1534	4.92	0.0265	0.05	0.2507 0.9263
School 3 odds	EXP	4	0.5204	0.1522	4.99	0.0256	0.05	0.2533 0.9253

```

proc catmod order=data;
  ods select PopProfiles Estimates;
  weight Count;
  model Style=School Program;
run;

/* Odds corresponding to each logit function */
/* for each subpopulation in the data */
ods select ContrastEstimates;
contrast "School 1 odds"
  Intercept 1 School 1 * Program 1,
  Intercept 1 School 1 * Program -1 / est=exp;
contrast "School 2 odds"
  Intercept 1 School 2 * Program 1,

```

NOTE: 30 Lines Submitted.

Proc CATMOD

- Neu: Iterativer proportionaler Anpassungs-algorithmus (IPF) für hierarchische Log-lineare Modelle mit **einer** Population
 - Keine unabhängigen Variablen
 - Keine Populationsvariablen
 - Log-Likelihood G^2 und vorhergesagte Zellhäufigkeiten ohne aufwendige Parameter- und Kovarianzberechnungen

Proc CATMOD

- Ansteuerung des Iterative Proportional Fitting: ML=IPF
 - Verschiedene Optionen für Konvergenzkontrolle verfügbar
 - Varianten für die Berechnung der Freiheitsgrade von G^2

Proc FACTOR

- Verallgemeinerte Crawford-Ferguson Rotationen
 - Direkt Oblimin, Quartimin, weitere orthogonale und schiefwinklige Rotationen
- Kontrolle der Rotationszyklen und der Konvergenz
- Konfidenzintervalle für die Faktorladungen

Proc FREQ

- Neue TABLE Statements
 - CONTENTS= für HTML-Ausgabe
 - FORMAT= für Formatierung der Zelhäufigkeiten
 - OUTCUM Option liefert kumulative Häufigkeiten und Prozentwerte in der Ausgabedatei bei Einwegtabellen

Proc LOESS

- Automatische Schätzung des Glättungsparameters
- Mischungen verschiedener Ordnung für Modelle mit einem oder zwei Regressoren
- Tabelle mit Anpassungswerten für jeden Glättungsparameter

Proc LOGISTIC

- Exakte bedingte Inferenzen für binäre abhängige Variable
- Sinnvoll, wenn asymptotische Annahmen verletzt sind, z.B.
 - Bei kleinen Datenmengen
 - Bei schiefen Verteilungen

Beispiel Proc LOGISTIC

Proc logistic ;

 model y = x1 x2 x3 ;

 exact x1 x2 ;

Run ;

- Test für x1 aus exakter bedingter Verteilung der suffizienten Statistik bei gegebener Konstante, x2 und x3
- Test für x2 analog

Proc MIXED

- Übersichtlichere Tabellenausgabe von Modellbewertungskriterien
 - AIC (Akaike's Information Criterion)
 - BIC (Bayesian Information Criterion)
- Für kleine Stichproben korrigierte Version von AIC (AICC)

Burnham, K.P. and Anderson, D.R. (1998), *Model Selection and Inference: A Practical Information-Theoretic Approach*, New York: Springer-Verlag.

Proc MODECLUS

- Standard-Glättungsparameter verfügbar für Optionen DR=, CR=, R=, DK=, CK=, K=
 - DR= bestimmt den Support-Radius für Uniform-Kerndichteschätzungen
 - CR= bestimmt den Radius der Clusterungs-Umgebung
 - DK= bestimmt die Zahl der Nachbarn für die k th-Nearest-Neighbor Dichteschätzung
 - CK= bestimmt die Zahl der Nachbarn für die Clusterung

Proc MULTTEST

- 2 neue Korrekturmethode für p-Werte sind verfügbar
 - Fisher-Kombination
 - p-Wert für Test j: Maximum aller p-Werte der Kombinationstests, die j enthalten
 - Voraussetzung: p-Werte unabhängig
 - Hommel-Verfahren
 - Geeignet für unabhängige und positiv abhängige p-Werte

Proc NLMIXED

- 2 neue Verteilungen für die Modellierung nichtlinearer gemischter Modelle
 - Gammaverteilung **gamma(a,b)**
 - Negativ-Binomialverteilung **negbin(n,p)**
- AIC und BIC übersichtlicher aufbereitet und ausgegeben
- Für kleine Stichproben korrigierte Version von AIC (AICC)

Proc PHREG

- Analysen von Überlebensdaten
 - Daten oft unvollständig oder zensiert
 - Übliche parametrische Annahmen nicht gerechtfertigt
- Robuste Schätzung der Kovarianzmatrix nach Wei und Lin (1989)
- Anpassung eines Nullmodells durch Nichtangabe von erklärenden Variablen im MODEL Statement

Proc SURVEYMEANS

- Domainanalyse für Umfragedaten
 - Analysen auf Subpopulationen
 - Domäne unabhängig vom Umfrageplan
 - Deshalb Stichprobenumfänge der Subpopulationen zufällig
- Überarbeitete Berechnungsmethoden
 - Schätzung von Populationsmittelwerten
 - Varianzschätzungen
 - Konfidenzgrenzen

Proc GAM

- Experimentelle Prozedur zur Anpassung verallgemeinerter additiver Modelle nach Hastie und Tibshirani (1990)
- Basierend auf nichtparametrischer Regression und Glättungstechniken
 - Simultane Exploration mehrerer nichtparametrischer Beziehungen
 - Flexibilität der Verteilungen wie bei Nelder, Wedderburn (1972)

Proc GAM

- Liefert nichtparametrische Schätzungen für additive Modelle
- Unterstützt multidimensionale Daten
- Multiple SCORE Statements möglich
- Vorgabe von Freiheitsgrad oder Glättungsparameter ermöglicht Modellwahl

Proc MI

- Multiple Imputation für die Analyse von Daten mit fehlenden Werten
 - Ersetzen fehlender durch plausible Werte
 - Im Anschluß Standardanalysen
 - Glättung durch Kombination der Analysen
 - Inferenzen zur Beurteilung der Imputationen

Proc MI

- 3 Methoden für p-dimensionale Imputationen
 - Regressionsmethode
 - Neigungs-Score-Methode
 - Markovketten-Monte-Carlo-Methode (MCMC)
- Experimentell in Version 8.1

Proc MIANALYZE

- Kombiniert die Resultate von Analysen auf m Imputationen
- Erzeugt statistische Bewertungen der Qualität der Kombination
- Experimentell in Version 8.1

www.sas.com

> Service and Support

A screenshot of a Microsoft Internet Explorer browser window. The title bar reads 'SAS/STAT Software: MI Procedure - Microsoft Internet Explorer'. The address bar shows the URL 'http://ieprof.uro.sas.com/service/library/onlinehelp/v01/whatsnew/c1499716.htm'. The page content includes a navigation bar with 'Home', 'Contents', 'Index', and 'What's New' links. The main heading is 'MI Procedure'. The text describes multiple imputation as a strategy for handling missing data, mentioning the regression, propensity score, and MCMC methods. It also notes that the new MI procedure is experimental in Release 8.1 and offers three methods for creating imputed data sets. A second navigation bar is located below the text, and a copyright notice for SAS Institute Inc. is at the bottom of the page content area.

SAS/STAT Software: MI Procedure - Microsoft Internet Explorer

Home Contents Index

MI Procedure

Multiple imputation is a strategy for dealing with data sets that contain missing values. You replace each missing value with a set of plausible values that represent the uncertainty about the right value to impute. You create multiply imputed data sets, analyze them with standard analyses, and then combine the results. You produce valid statistical inferences that properly reflect the uncertainty due to the missing values.

The new MI procedure, which is experimental in Release 8.1, creates multiply imputed data sets for incomplete p -dimensional multivariate data. It offers three methods for creating the imputed data sets: the regression method, the propensity score method, and the Markov Chain Monte Carlo (MCMC) method. The procedure creates an output data set that contains m imputed versions of the original data. In each version, the missing values are replaced with imputed values. For the MCMC method, you can specify whether you want a single chain for all m imputations or a separate chain for each imputation. You can also specify the initial estimator for the MCMC method.

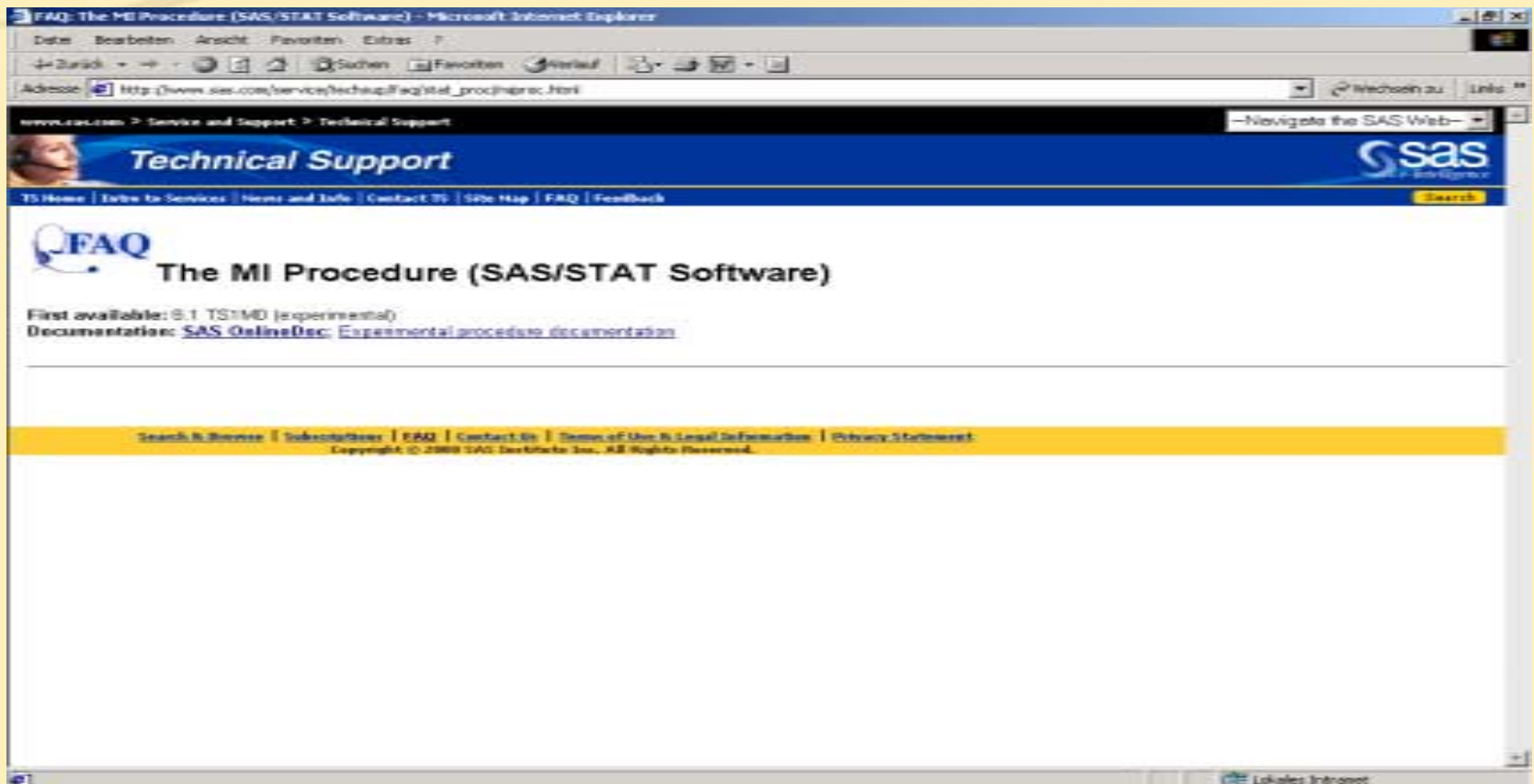
After analyzing your imputed data with standard procedures, you use the MIANALYZE procedure to combine the results.

Home Contents Index What's New

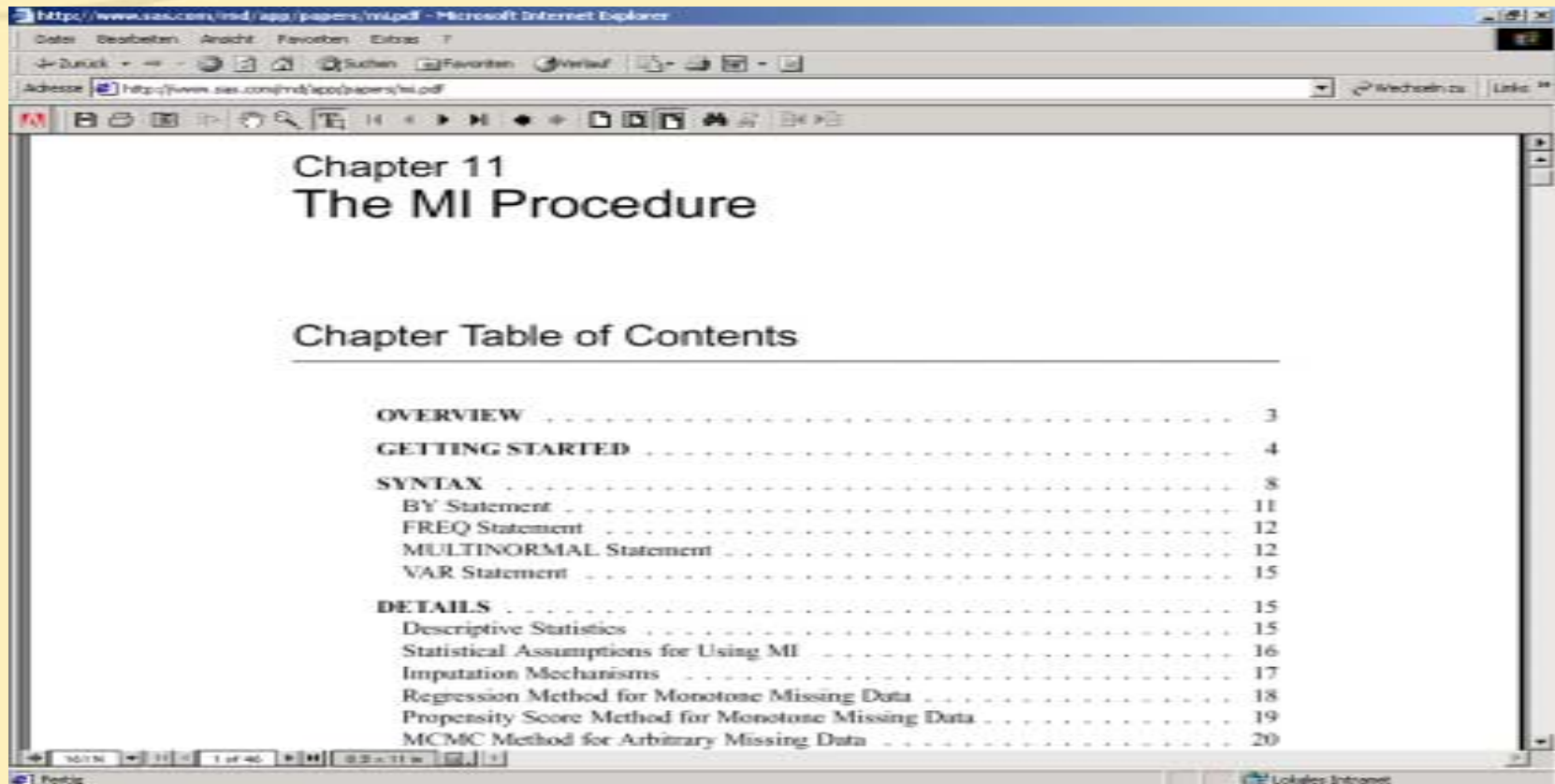
Copyright 2001 by SAS Institute Inc., Cary, NC, USA. All rights reserved.

Service and Support

> Technical Support



Proc MI - Experimentelle Dokumentation, 46 Seiten



Chapter 11
The MI Procedure

Chapter Table of Contents

OVERVIEW	3
GETTING STARTED	4
SYNTAX	8
BY Statement	11
FREQ Statement	12
MULTINORMAL Statement	12
VAR Statement	15
DETAILS	15
Descriptive Statistics	15
Statistical Assumptions for Using MI	16
Imputation Mechanisms	17
Regression Method for Monotone Missing Data	18
Propensity Score Method for Monotone Missing Data	19
MCMC Method for Arbitrary Missing Data	20

SAS Enterprise Guide

- Thin Client Lösung für Microsoft-Windows-Plattformen
 - Berichte
 - Grafiken
 - Statistische Analysen
- Grafische Benutzeroberfläche für
 - Mitarbeiter aus Fachabteilungen
 - Erfahrene SAS-Programmierer

Datenfilter

SAS Enterprise Guide

Query Builder

Name:

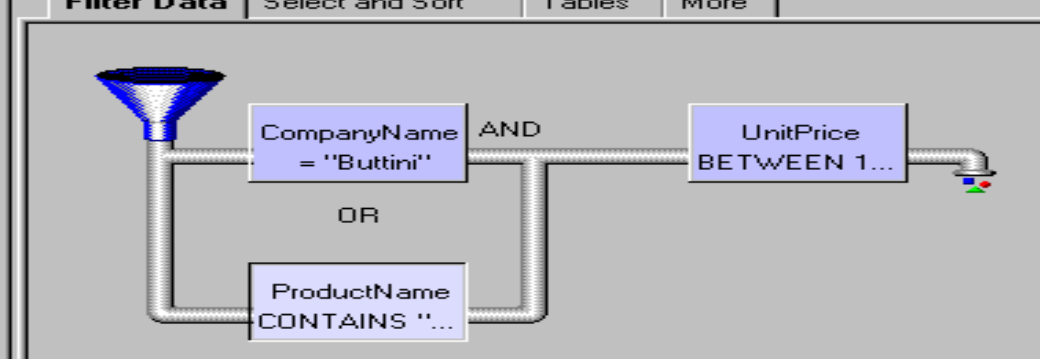
products

- ProductID
- ProductName
- SupplierID
- CategoryID
- QuantityPerUnit
- UnitsInStock
- UnitsOnOrder
- ReorderLevel
- Discontinued
- UnitPrice

Suppliers_Suppliers_...

- SupplierID
- CompanyName
- ContactName
- ContactTitle

Filter Data | Select and Sort | Tables | More



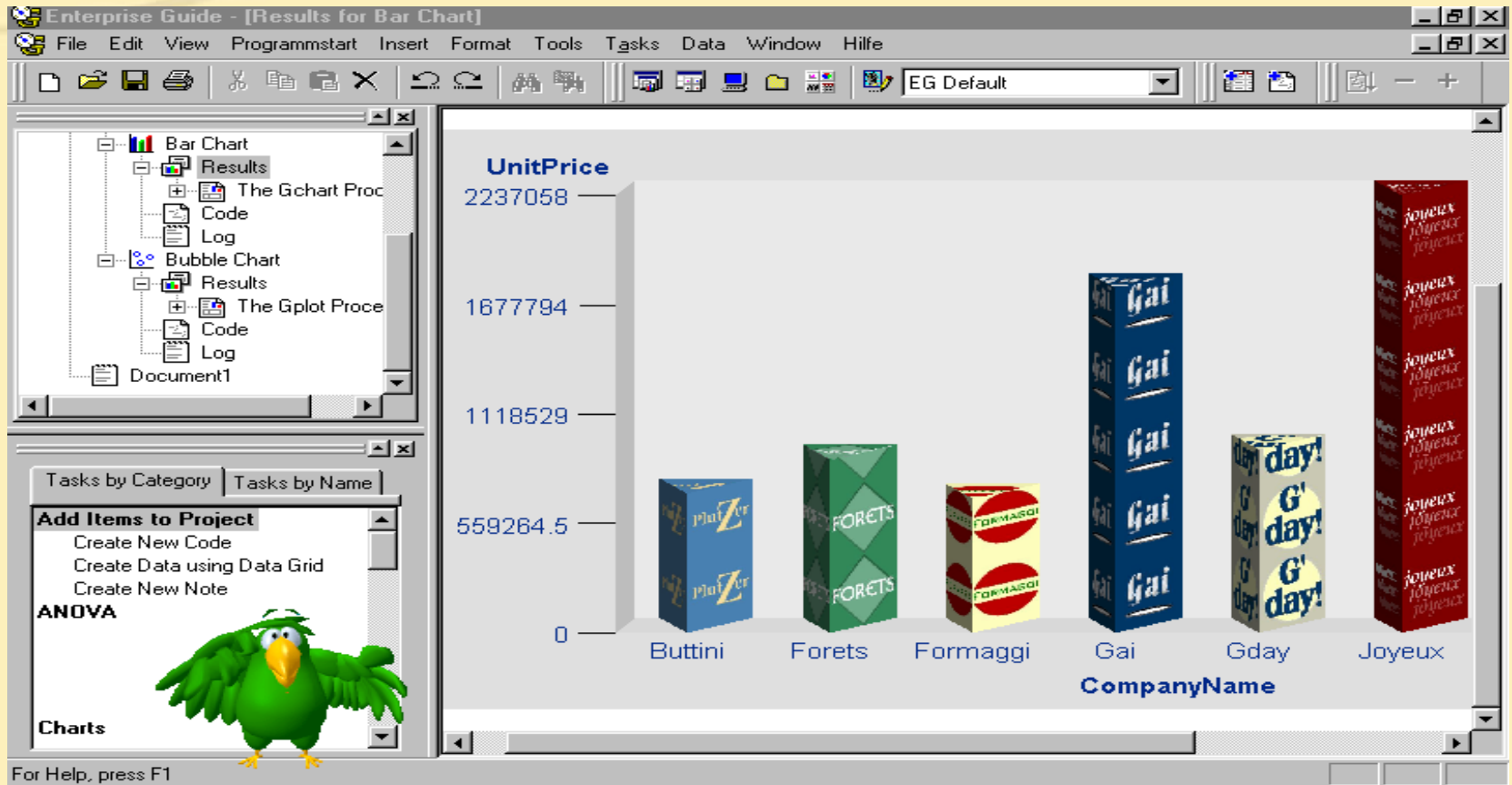
Preview the query results
 Preview the query code

Add Table | OK | Cancel | Help

Query Results for "Query Builder"

	ProductID	ProductName	SupplierID	CategoryID	QuantityPerUnit	
1	57	Ravioli Angelo	26	5	24 - 250 g pkgs.	
2	57	Ravioli Angelo	26	5	24 - 250 g pkgs.	
3	56	Gnocchi di nonn...	26	5	24 - 250 g pkgs.	
4	57	Ravioli Angelo	26	5	24 - 250 g pkgs.	
5	56	Gnocchi di nonn...	26	5	24 - 250 g pkgs.	

Geschäftsberichte SAS Enterprise Guide



Statistische Analysen

SAS Enterprise Guide

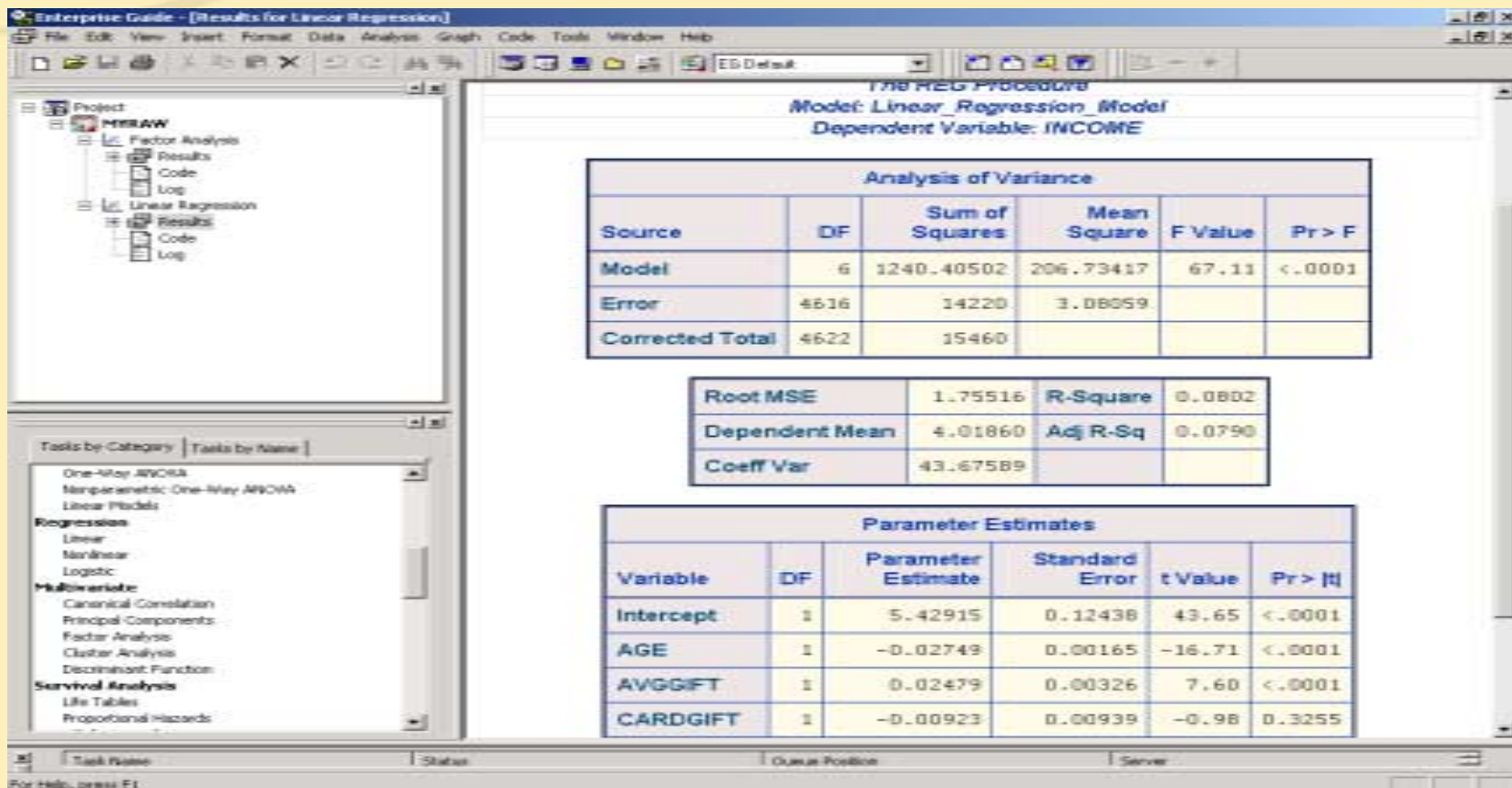
- Deskriptive Statistik
 - Verteilung, Korrelation, Häufigkeiten
- ANOVA
- Regression
 - Linear, Nichtlinear, Logistisch
- Multivariate Analysen
 - Kanon. Korrelation, PCA, Faktoranalyse, Clusteranalyse, Diskriminanzanalyse

Statistische Analysen

SAS Enterprise Guide

- Überlebenszeitanalysen
 - Kaplan-Meier, Cox-Modelle
- Qualitätskontrolle
 - Prozessfähigkeitsanalyse, Kontrollkarten, Pareto Charts
- Zeitreihenanalysen
 - Einfache Prognosen, ARIMA-Modelle, Regressionen mit Zeitreihen

HTML-Output SAS Enterprise Guide



The screenshot displays the SAS Enterprise Guide interface with the following components:

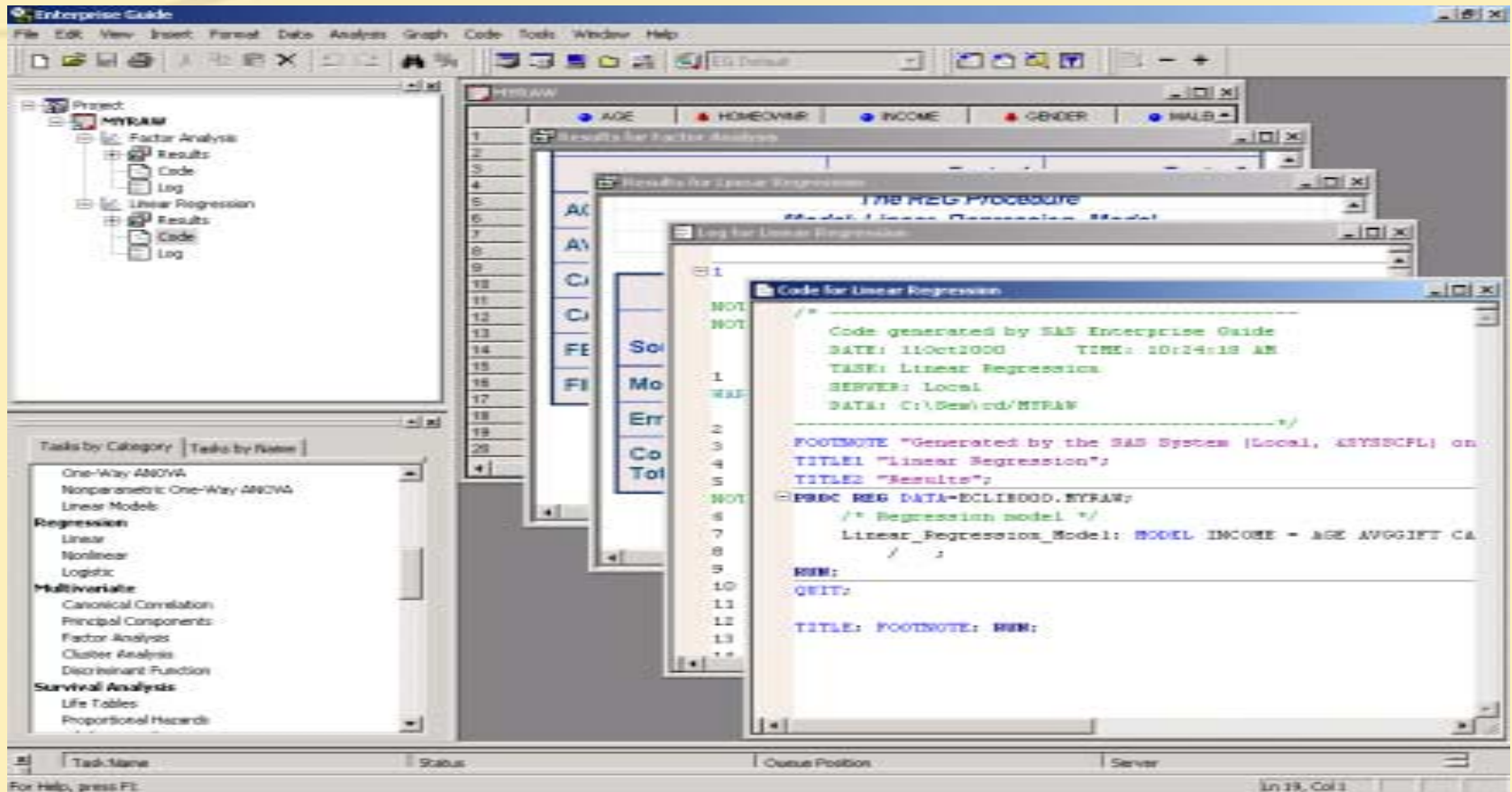
- Project Tree:** Shows a project named 'HEERAW' containing sub-projects for 'Factor Analysis' and 'Linear Regression', each with 'Results', 'Code', and 'Log' files.
- Task List:** A sidebar on the left lists various statistical tasks such as 'One-Way ANOVA', 'Regression', and 'Survival Analysis'.
- Main Output Window:** Displays the results for a linear regression model.
 - Model:** Linear_Regression_Model, Dependent Variable: INCOME
 - Analysis of Variance:**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1240.40502	206.73417	67.11	<.0001
Error	4616	14220	3.08059		
Corrected Total	4622	15460			
 - Summary Statistics:**

Root MSE	1.75516	R-Square	0.0802
Dependent Mean	4.01860	Adj R-Sq	0.0790
Coeff Var	43.67589		
 - Parameter Estimates:**

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.42915	0.12438	43.65	<.0001
AGE	1	-0.02749	0.00165	-16.71	<.0001
AVGGIFT	1	0.02479	0.00326	7.60	<.0001
CARDGIFT	1	-0.00923	0.00939	-0.98	0.3255

Projektverwaltung SAS Enterprise Guide



The screenshot displays the SAS Enterprise Guide interface. On the left, a 'Project' tree shows a folder named 'MYRAW' containing sub-folders for 'Factor Analysis' and 'Linear Regression', each with 'Results', 'Code', and 'Log' sub-items. Below the tree is a 'Tasks by Category' list including 'One-Way ANOVA', 'Regression', 'Multivariate', and 'Survival Analysis'. The main workspace is divided into several overlapping windows:

- Code for Linear Regression:** Contains SAS code for a linear regression model. The code includes a title, a regression model statement, and a quit command.


```
Code generated by SAS Enterprise Guide
DATE: 11Oct2000    TIME: 10:24:18 AM
TASK: Linear Regression
SERVER: Local
DATA: C:\Dev\cd\MYRAW

FOOTNOTE "Generated by the SAS System (Local, ASYSCPL) on
TITLE "Linear Regression";
TITLE "Results";
PRIC REG DATA=ECLIS00D.MYRAW;
/* Regression model */
Linear_Regression_Model: MODEL INCOME = AGE AVGGIFT CA
/;

RUN;
QUIT;

TITLE: FOOTNOTE: RUN;
```
- Log for Linear Regression:** Shows the execution log for the code above.
- Results for Linear Regression:** Displays the output of the regression analysis.
- Results for Factor Analysis:** Displays the output of a factor analysis.

The bottom status bar shows 'Task Name', 'Status', 'Queue Position', 'Server', and 'Ln 19, Col 1'. A note at the bottom left says 'For Help, press F1'.

SAS/STAT → ← SAS Enterprise Guide

■ SAS/STAT

- 33 Prozeduren in Rel. 6
- 43 Prozeduren in Rel. 6.12, Orlando II
- 53 Prozeduren in Rel. 8.0
 - PLS, KRIGE2D, VARIOGRAM, SURVEYSELECT, SURVEYMEANS, SURVEYREG, KDE, LOESS, TPSPLINE, NLMIXED
- 56 Prozeduren in Rel. 8e
 - GAM, MI, MIANALYZE

SAS/STAT → ← SAS Enterprise Guide

- SAS Enterprise Guide
 - Base SAS
 - SAS/STAT
 - SAS/ETS
 - SAS/QC
 - SAS/GRAPH

Entwicklung lebt von den Anregungen der
Anwender !

www.sas.com

www.sas.de

