

Innovation für Statistik und Datenanalyse in Orlando I/II

Reinhard Strüby
Technical Consultant
SAS Institute

Statistik und Datenanalyse mit dem SAS System

- Wurzeln: Computerprogramme für statistische Auswertungen
- Erweiterung der analytischen Themen
- Großes Methodenspektrum
- Vollständige Software-Infrastruktur
- Datenanalyse bleibt Kernbereich des SAS Systems !

Statistik und Datenanalyse mit dem SAS System

- Endbenutzersysteme
- Statistische Analyse und Modellierung
- Zeitreihenanalyse
- Qualitätskontrolle
- Operations Research
- Data Mining

Statistik und Datenanalyse mit dem SAS System

Endbenutzersysteme

- SAS Research Applikation
- Time Series FORECASTing System
- MARKET Research Application

Endbenutzersysteme

SAS Research Applikation

- Produktivitätstool
- benutzerfreundlich
- erweiterbar

Endbenutzersysteme

FORECAST Zeitreihenanalysen

- Orlando I: prod. + Orlando II: erweitert
- Diagnose: grafisch + analytisch
- Automatische Modellwahl
- Selbständige Dokumentation

Endbenutzersysteme

MARKET Marktforschung

- Conjoint Measurement
- Korrespondenzanalyse
- Multidimensionale Skalierung
- Multidimensionale Präferenzanalyse

Statistische Analyse und Modellierung

- Generalised Estimation Equations
- Exakte Methoden
- Multiples Testen
- Reliability Analyse: **SAS/QC**
- Differentialgleichungssysteme:
SAS/ETS
- und mehr...

Generalised Estimation Equations

- Implementation: Proc **GENMOD**
- Repeated Anweisung
- Varianzfunktionen: Normal, Poisson, Binomial, Gamma, invers Gauss + individuell erweiterbar
- Link-Funktionen für populäre Modelle: logit, probit, ...

Exakte Methoden

- Exakte Probs für lineare Rangstatistiken: Proc **NPAR1WAY**
- Exakte Probs und KIs für Statistiken: Proc **FREQ**
- wichtig für kleine Fallzahlen und nicht normal verteilte Daten

Exakte Tests

- Wilcoxon
- Median
- Van der Waerden
- Pearson Chi Square
- Likelihood Ratio Chi Square
- Mantel-Haenszel Chi Square
- McNemar's Test
- Cochran-Armitage Trend Test
- Jonckhere-Terpsta Test
- Pearson Correlation Coefficient
- Kendall's Tau

Multipltes Testen

- Permutationstest: Proc **MULTTEST**
- P-value Adjustments:
 - Bonferroni
 - Sidak
 - stepdown
 - Hochberg, Hochberg/Benjamini
 - bootstrap
 - permutation

Multipltes Testen

- Gewichtung von Tests nach STRATA Variablen
- Nichtkorrigierte P-Werte als Input möglich: ohne Originaldaten

Statistische Analyse und Modellierung

- Tests der Varianzhomogenität und allg. Welch Test: Proc **GLM**
- Zusätzliche Flexibilität bei LSMEANS: Proc **GLM** + **MIXED**
- Automatische Kalkulation von Ableitungen: Proc **NLIN**
- Output Anweisung: Proc **NPAR1WAY**

Zeitreihenanalyse

- Differentialgleichungen 1.Ordnung, Numerische Integration: Proc **MODEL**
- Moore-Penrose Generalized Inverse
- Chow Test für strukturelle Brüche
- Durbin-Watson Tests höherer Ordnung
- BY Processing erweitert

Zeitreihenanalyse

- Profile Likelihood KI: Proc **MODEL**
- 42 neue Transformationen
cumulative, moving, exp smoothing,...
Proc **EXPAND**
- Periodogramme mit vorbereiteten
Kernen: Proc **SPECTRA**
- White Noise + Stationary Tests:
FORECAST Menüsystem

Qualitätskontrolle

- Process Flow Diagram: prod.
- Multivariate Charts: Proc **SHEWHART**
- Proc **RELIABILITY**

RELIABILITY Analyse

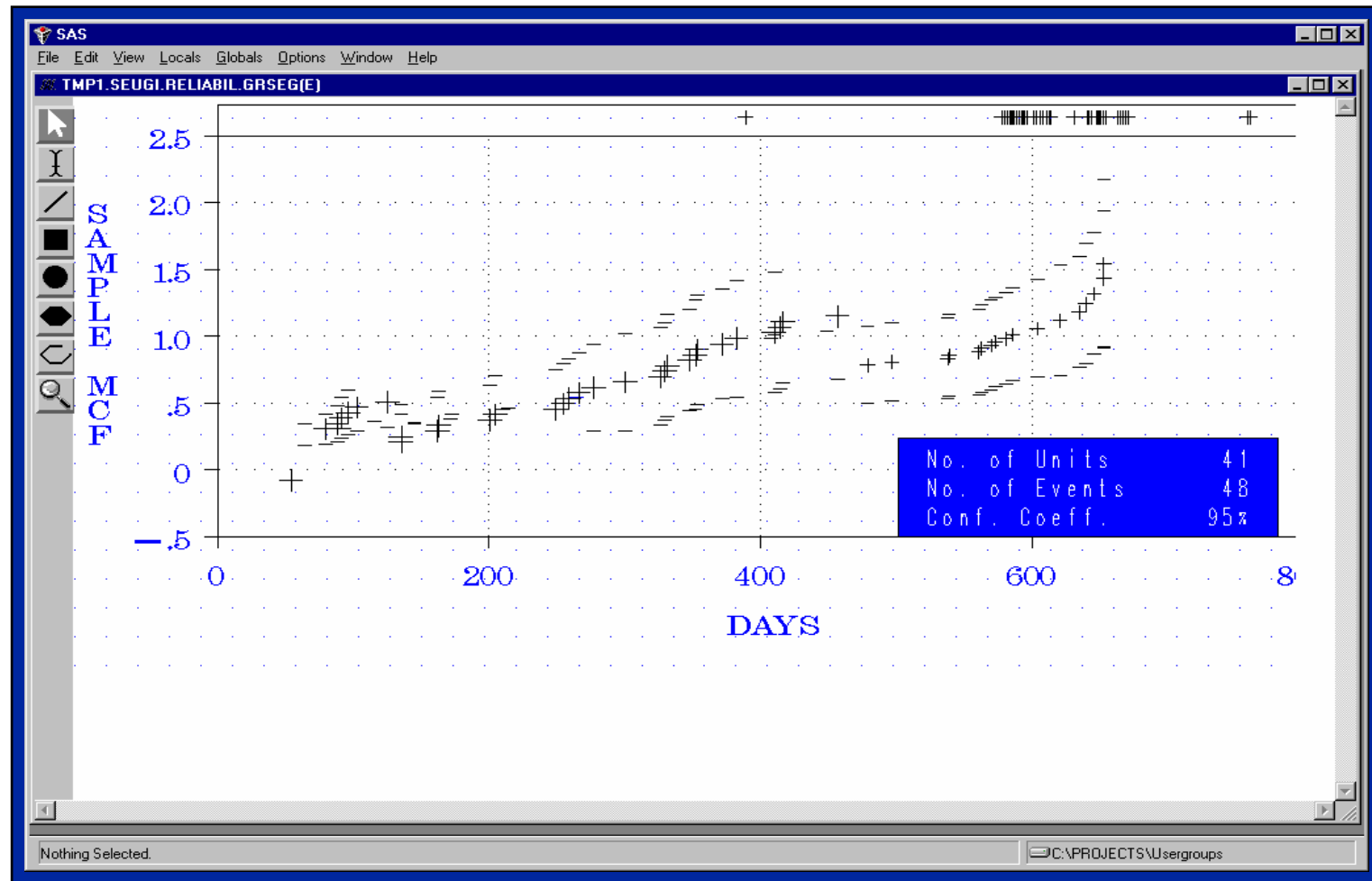
Relevanz für:

- Produktlebensdauer
- Systemreparatur
- Medizinische Forschung
- Pharma Forschung

RELIABILITY Analyse

- Wahrscheinlichkeitsplots und Parameterschätzungen für übliche Lebensdauervertelungen
- Maximum Likelihood Schätzungen
- AN- und LR-Konfidenzintervalle
- Plot von Stressfaktoren

MCF für Anzahl der Reparaturen



Operations Research

XPROJMAN Menüsystem (beta)

- Projektieren mit Zeit-, Reihenfolge- und Ressourcenrestriktionen
- Unterstützung von Projekthierarchien
- Interaktiver Editor für Projektnetz
- Neue CPM Funktionalität

Operations Research

QSIM Applikation (beta)

- GUI zum Aufsetzen und Simulieren von Warteschlangenmodellen
- Anwendungen für CPE, Telekommunikation, Transport, Produktionsoptimierung

Data Mining

Merkmale

- Zugang auf beliebige Datenquellen
- Umfassendes Methodenspektrum
- Technologie der Implementierung

Data Mining Technologie

SEMMA

S ampling	Stichprobenverfahren
E xploration	Visualisierungen
M odification	Datenaufbereitung
M odelling	Statistische Verfahren
A ssessment	Modellbewertung

Data Mining

Methoden

- Neuronale Netze: **NN** Applikation
- CHAID: **TREE** Applikation
- Exploration: **INSIGHT, SVIEW, GIS**
- Zeitreihenanalyse: **ETS**
- Traditionelle Statistik: **STAT**
- Marktforschungsmethoden: **MARKET**

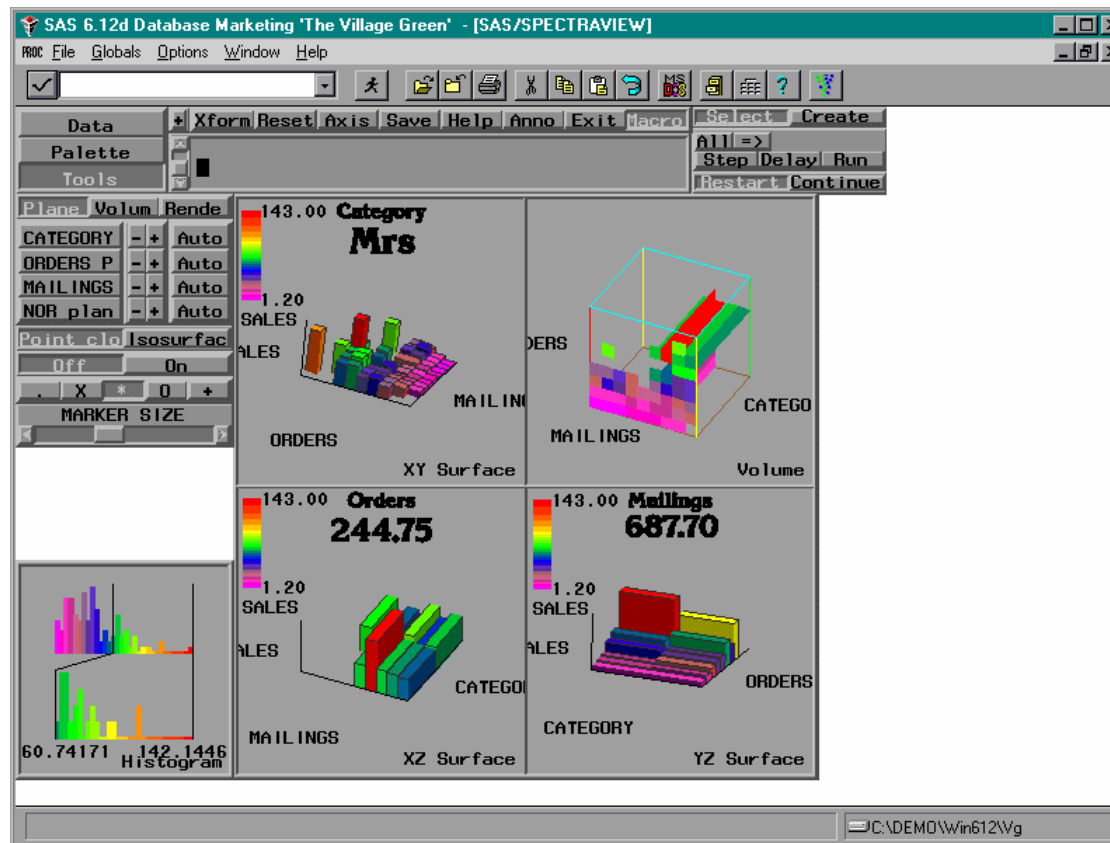
Data Mining

SAS/SPECTRAVIEW

- Erweiterte Integration: Ausgabe von Daten und Filtern in SAS-Dateien
- Kategorisieren von Werten
- Charaktervariable werden unterstützt

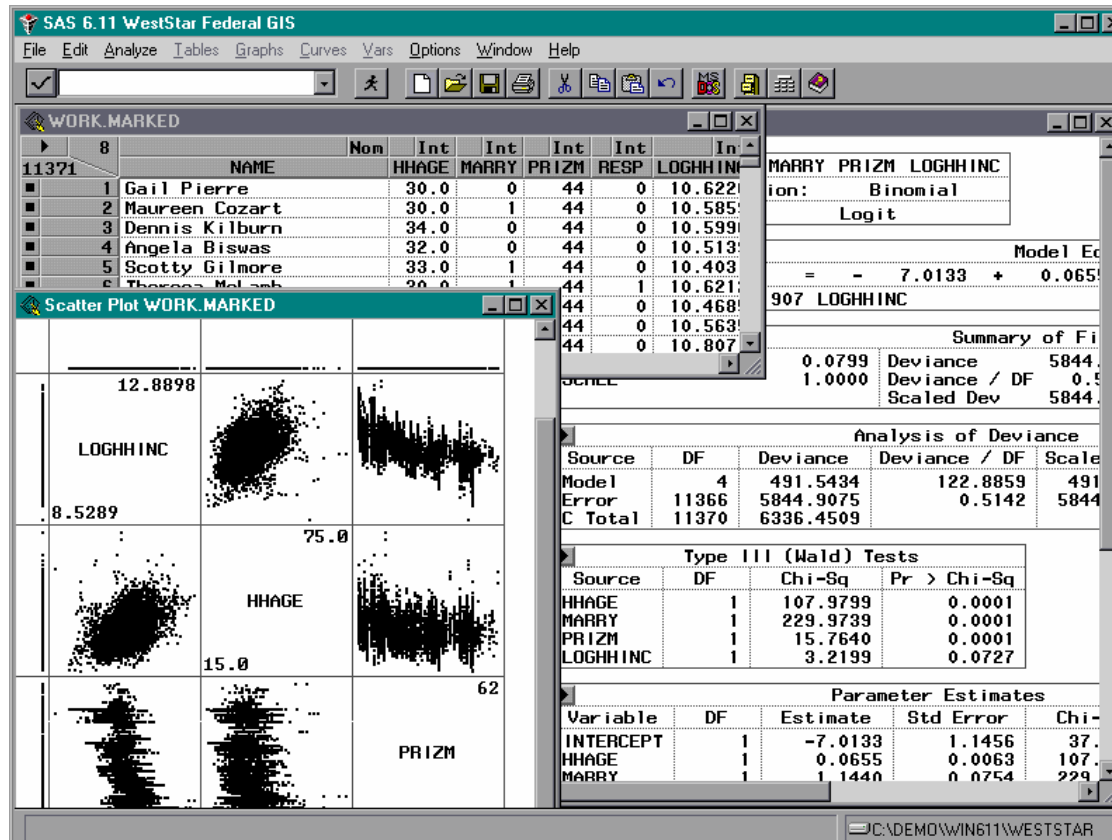
Data Mining Produkte

- SAS/SPECTRAVIEW



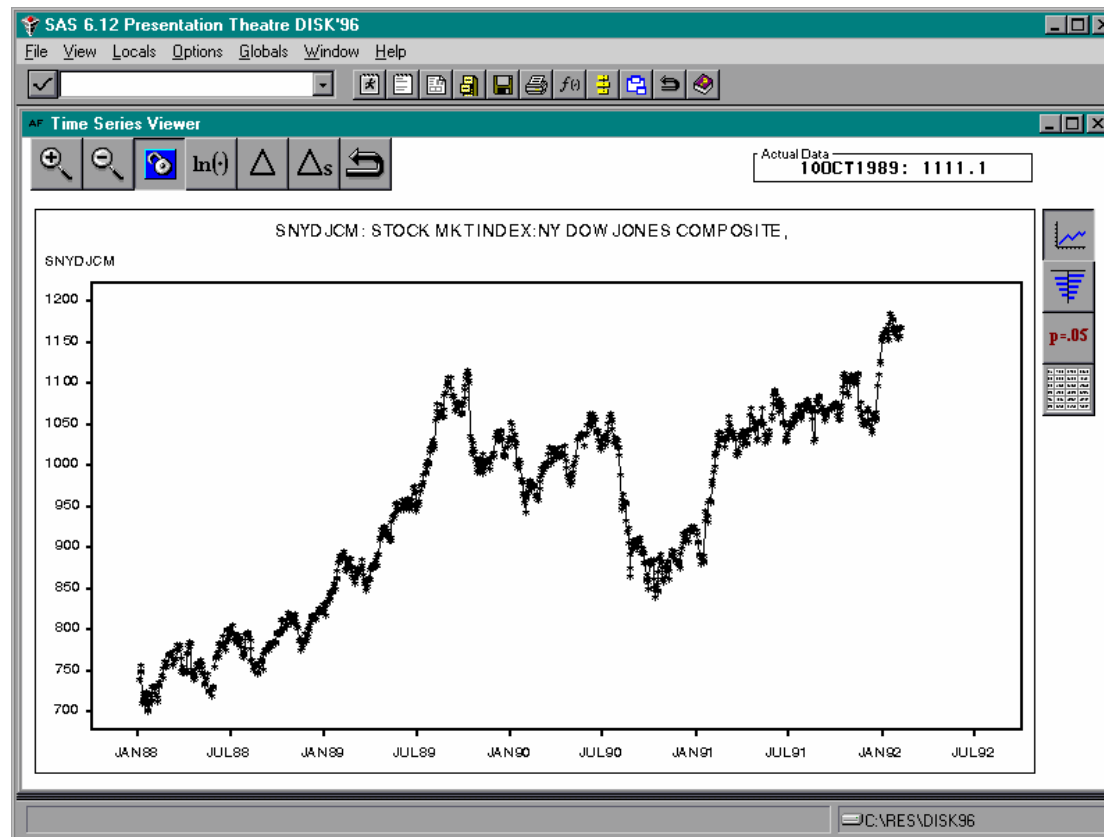
Data Mining Produkte

- SAS/INSIGHT



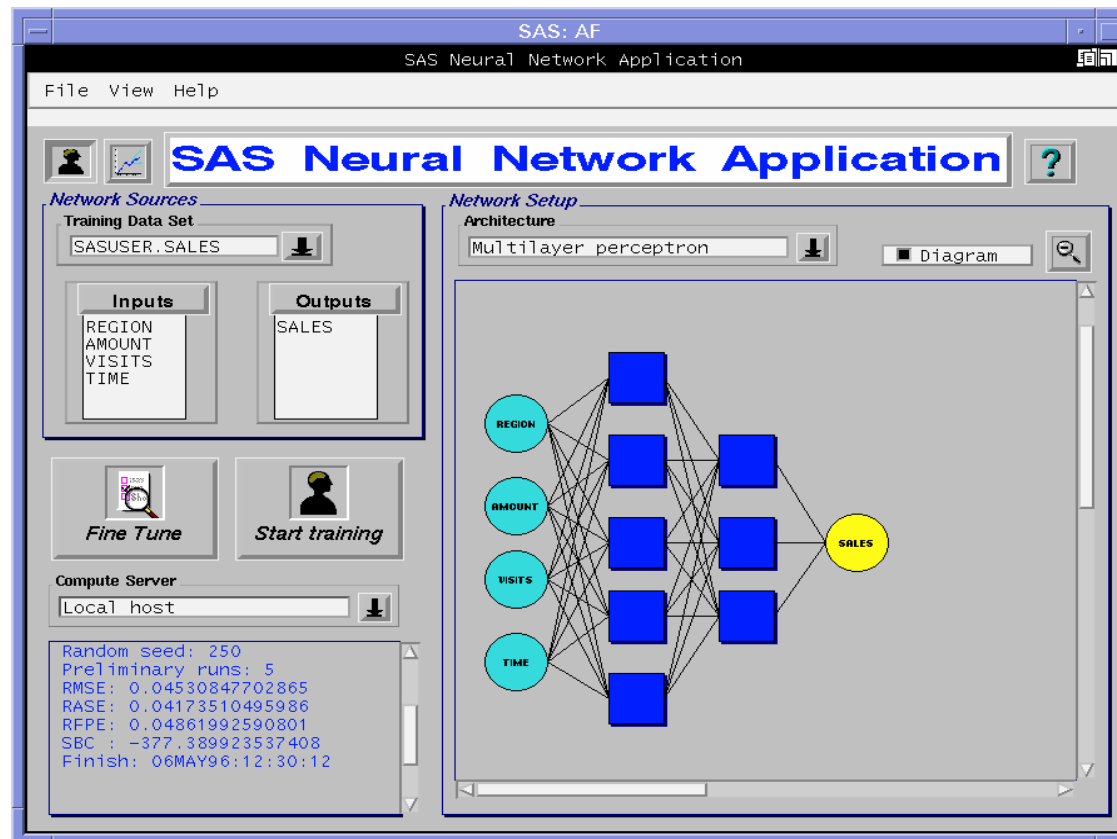
Data Mining Produkte

- SAS/ETS Forecast



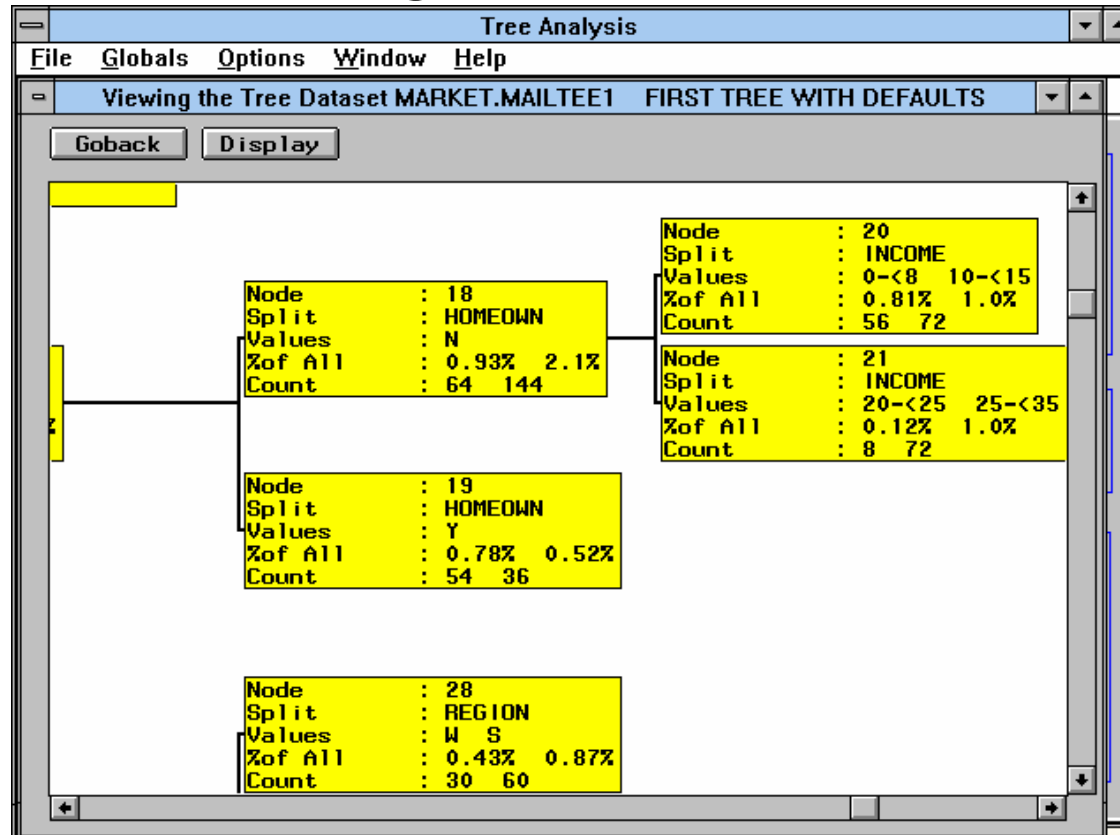
Data Mining Produkte

- Neuronale Netze



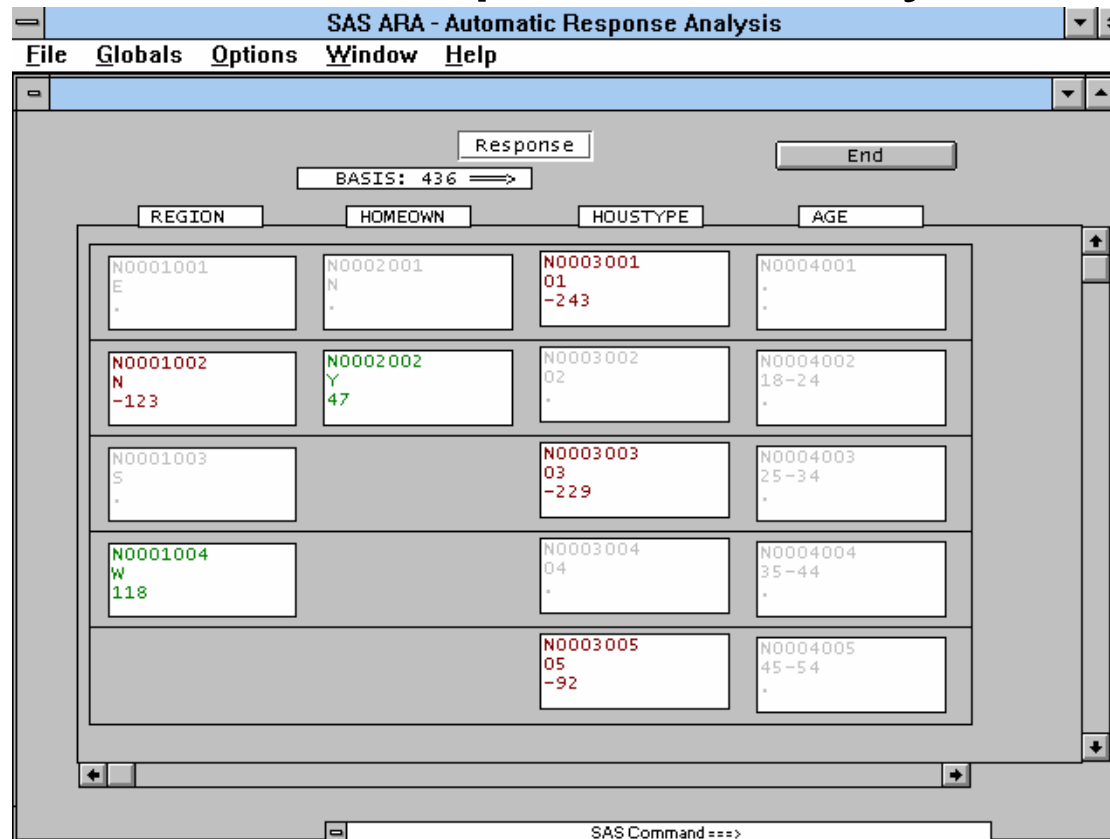
Data Mining Lösungen

- Entscheidungsbäume



Data Mining Lösungen

- Automatic Response Analysis



Innovation für Statistik und Datenanalyse in Orlando I/II

Fragen ?

Demonstration, Raum 2097:

Donnerstag 15.00 Uhr

Freitag 11.30 Uhr