AGENDA

• Mehr als linear oder logistisch? – Quantils Regressionen und Adaptive Splines in SAS  
  Dr. Mihai Paunescu

• Moderne SAS Analytik Architekturen  
  Mag. Gernot Engel, Ing. Phillip Manschek,  
  DI Rainer Sternecker, Dr. Gerhard Svolba

• Eine Liebeserklärung an den SAS Enterprise Guide  
  Mag. Bernadette Fabits

• SAS Analytics 4 U – Neue Möglichkeiten für Studierende und Dozenten  
  DI (FH) Bettina Brandl
MEHR ALS LINEAR ODER LOGISTISCH?
QUANTILS REGRESSIONEN UND ADAPTIVE SPLINES IN SAS

MIHAI PAUNESCU
**proc univariate data=dat;**
ods select moments quantiles;
var sales;
run;

<table>
<thead>
<tr>
<th>Location</th>
<th>Quantile (Definition 5)</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>241825.8</td>
<td>823581</td>
</tr>
<tr>
<td>Median</td>
<td>75875.1</td>
<td>6,78E+16</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
<td>18605198</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>168416</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>18605230.40</td>
</tr>
<tr>
<td>99%</td>
<td>2699730.30</td>
</tr>
<tr>
<td>95%</td>
<td>779877.70</td>
</tr>
<tr>
<td>90%</td>
<td>458327.90</td>
</tr>
<tr>
<td>75% Q3</td>
<td>198564.00</td>
</tr>
<tr>
<td>50% Median</td>
<td>75875.10</td>
</tr>
<tr>
<td>25% Q1</td>
<td>30147.60</td>
</tr>
<tr>
<td>10%</td>
<td>13607.10</td>
</tr>
<tr>
<td>5%</td>
<td>5598.60</td>
</tr>
<tr>
<td>1%</td>
<td>500.75</td>
</tr>
<tr>
<td>0% Min</td>
<td>32.55</td>
</tr>
</tbody>
</table>
ods graphics on / width=1000px height=700px;
proc sgplot data=dat;
  where sales < 2699730.30;
  histogram sales;
  refline 779877.70 458327.90 198564.00 / axis=x lineattrs=(color=bigb pattern=dot thickness=3);
  refline 30147.60 13607.10 5598.60 500.75 /axis=x lineattrs=(color=dapk pattern=dot thickness=3);
  refline 75875.10 /axis=x label='Median' lineattrs=(color=brown thickness=3);
  refline 241825.8 /axis=x label='Mean' lineattrs=(color=big thickness=3);
run;
**PROC QUANTREG UND MEDIAN**

```sas
proc quantreg data=dat;
  model Sales= / quantile=(0.5);
run;
```

**Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>75875.10</td>
<td>2.654.220</td>
<td>70.670.691</td>
<td>81.079.509</td>
<td>28.59</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Quantiles (Definition 5)**

<table>
<thead>
<tr>
<th>Level</th>
<th>Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>18605230.40</td>
</tr>
<tr>
<td>99%</td>
<td>2699730.30</td>
</tr>
<tr>
<td>95%</td>
<td>779877.70</td>
</tr>
<tr>
<td>90%</td>
<td>458327.90</td>
</tr>
<tr>
<td>75% Q3</td>
<td>198564.00</td>
</tr>
<tr>
<td>50% Median</td>
<td>75875.10</td>
</tr>
<tr>
<td>25% Q1</td>
<td>30147.60</td>
</tr>
<tr>
<td>10%</td>
<td>13607.10</td>
</tr>
<tr>
<td>5%</td>
<td>5598.60</td>
</tr>
<tr>
<td>1%</td>
<td>500.75</td>
</tr>
<tr>
<td>0% Min</td>
<td>32.55</td>
</tr>
</tbody>
</table>
proc quantreg data=dat ci=resampling;
    class cust_grp;
    model Sales=cust_grp / quantile=(0.5) seed=12345;
run;

| Parameter  | DF | Estimate | Standard Error | 95% Confidence Limits | t Value | Pr > |t| |
|------------|----|----------|----------------|------------------------|---------|------|---|
| Intercept  | 1  | 47335.60 | 3.071.371      | 41.313.237             | 53.357.963 | 15.41 | <.0001 |
| Cust_grp   | 1  | 88078.60 | 6.903.957      | 74.541.278             | 101615.92 | 12.76 | <.0001 |
| Cust_grp   | 0  | 21870.35 | 5.324.410      | 11.430.214             | 32.310.486 | 4.11  | <.0001 |
| Cust_grp   | 0  | 0.0000   | 0.0000         | 0.0000                 | 0.0000   | .     | .   |

proc means data=dat P1 median P90;
    var sales;
    class Cust_grp;
run;

<table>
<thead>
<tr>
<th>Analysis Variable : Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cust_grp</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>O</td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

47335.6 + 88078.6 = 135414.2
QUANTILS-REGRESSION
MEDIAN SCHÄTZUNG FÜR INTERVALVARIABLE

```sas
proc quantreg data=dat ci=resampling plots=(fitplot);
  where cust_grp='n';
  model Sales=Rabatt_Menge / quantile=(0.5) seed=12345;
run;
```

### Parameter Estimates

| Parameter    | DF | Estimate | Standard Error | 95% Confidence Limits | t Value | Pr > |t|
|--------------|----|----------|----------------|-----------------------|---------|------|
| Intercept    | 1  | 28552.1  | 5873.371       | 17026.581 40077.615   | 4.86    | <.0001|
| Rabatt_Menge | 1  | 9.1      | 0.4522         | 8.2082 9.9829         | 20.12   | <.0001|

**Fit at Quantile Level 0.5 for sales**
*With 95% Confidence Limits*
**QUANTILS-REGRESSION**

**QUANTILSCHÄTZUNG FÜR INTERVALVARIABLE**

```
proc quantreg data=dat ci=resampling plots=(fitplot);
ods select fitplot;
where cust_grp='n' and Rabatt_Menge < 200000;
model Sales=Rabatt_Menge / quantile=(0.1 0.2 0.5 0.7 0.9) seed=12345;
run;
```
QUANTILE PROZESS PLOT

```
proc quantreg data=dat ci=resampling;
ods select quantplot;
where cust_grp='n' and Rabatt_Menge < 200000;
model Sales=Rabatt_Menge /quantile=(0.1 to 0.9 by 0.05) plot=(quantplot /unpack ols) seed=1268;
run;
```
**PROC QUANTREG DATA=DAT CI=RESAMPLING;**

```sas
where cust_grp='n';
model Sales=Rabatt_Wachstum Rabatt_Menge /
quantile=(0.1 to 0.9 by 0.05)  plot=(quantplot /unpack ols) seed=1268 ;
run;
```

---

**Estimated Parameter by Quantile Level for Sales**

With 50% Confidence Limits

---

**Estimated Parameter by Quantile Level for Sales**

With 50% Confidence Limits
QUANTILS-REGRESSION

FAZIT

• Verwendung:
  • Zusammenhänge entdecken für extreme Bereiche der Zielvariable
  • Robuste Medianschätzung gegenüber Ausreißer, ohne Verteilungsannahmen

• Hinweise:
  • Ist nicht äquivalent zu linearen Regressionen für Segmente von Beobachtungen
  • Schneller: HPQUANTSELECT ab SAS/STAT 13.2
proc adaptivereg
plots=all
details=bases;
model y = x;
run;
### Variable Parameter Estimates

| Variable | Estimate | t Value | Pr > |t| |
|----------|----------|---------|------|---|
| Intercept| 1.02     | 52.04   | <.0001|
| Basis1   | -0.73    | -67.82  | <.0001|
| Basis2   | -0.21    | -3.31   | 0.001 |

#### Data
```plaintext
data ds2;
set ds;
Basis1 = max(x-1.8,0);
Basis2 = max(1.8-x,0);
run;

proc reg data=ds2;
  model y = basis1 basis2;
run;
```
### Regression Spline Model after Backward Selection

<table>
<thead>
<tr>
<th>Name</th>
<th>Coefficient</th>
<th>Parent Variable</th>
<th>Variable</th>
<th>Knot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis0</td>
<td>1.2339</td>
<td>Intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basis2</td>
<td>-0.5857</td>
<td>Basis0</td>
<td>x</td>
<td>1.7865</td>
</tr>
<tr>
<td>Basis3</td>
<td>0.3277</td>
<td>Basis0</td>
<td>x</td>
<td>4.1447</td>
</tr>
<tr>
<td>Basis7</td>
<td>0.3681</td>
<td>Basis0</td>
<td>x</td>
<td>3.3424</td>
</tr>
<tr>
<td>Basis9</td>
<td>0.897</td>
<td>Basis0</td>
<td>x</td>
<td>4.7489</td>
</tr>
<tr>
<td>Basis11</td>
<td>-0.5508</td>
<td>Basis0</td>
<td>x</td>
<td>2.5061</td>
</tr>
<tr>
<td>Basis15</td>
<td>-0.5206</td>
<td>Basis0</td>
<td>x</td>
<td>1.3345</td>
</tr>
</tbody>
</table>

**Fit for y**

- GCV = 0.038

**x**

- 1
- 2
- 3
- 4
- 5

**y**

- 1.5
- 1.0
- 0.5
- 0.0
- -0.5
- -1.0
- -1.5
**ADAPTIVEREG**

**geeignet für klassifikationsvariablen und missing values**

```
proc adaptivereg data=autompg plots=all details=bases;
  class cylinders year origin;
  model mpg = cylinders displacement horsepower
          weight acceleration year origin / additive;
run;
```

<table>
<thead>
<tr>
<th>Name</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis0</td>
<td>1</td>
</tr>
<tr>
<td>Basis1</td>
<td>Basis0*MAX(Weight - 3139,0)</td>
</tr>
<tr>
<td>Basis2</td>
<td>Basis0*MAX(3139 - Weight,0)</td>
</tr>
<tr>
<td>Basis3</td>
<td>Basis0*NOT(MISSING(HorsePower))</td>
</tr>
<tr>
<td>Basis4</td>
<td>Basis0*MISSING(HorsePower)</td>
</tr>
<tr>
<td>Basis5</td>
<td>Basis3*MAX(HorsePower - 102,0)</td>
</tr>
<tr>
<td>Basis6</td>
<td>Basis3*MAX(102 - HorsePower,0)</td>
</tr>
<tr>
<td>Basis7</td>
<td>Basis0*(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)</td>
</tr>
<tr>
<td>Basis8</td>
<td>Basis0*NOT(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)</td>
</tr>
<tr>
<td>Basis9</td>
<td>Basis0*MAX(Displacement - 85,0)</td>
</tr>
<tr>
<td>Basis10</td>
<td>Basis0*MAX(85 - Displacement,0)</td>
</tr>
<tr>
<td>Basis11</td>
<td>Basis0*MAX(Displacement - 97,0)</td>
</tr>
<tr>
<td>Basis12</td>
<td>Basis0*MAX(97 - Displacement,0)</td>
</tr>
<tr>
<td>Basis13</td>
<td>Basis0*MAX(Acceleration - 21,0)</td>
</tr>
<tr>
<td>Basis14</td>
<td>Basis0*MAX(21 - Acceleration,0)</td>
</tr>
<tr>
<td>Basis15</td>
<td>Basis3*MAX(Displacement - 105,0)</td>
</tr>
<tr>
<td>Basis16</td>
<td>Basis3*MAX(105 - Displacement,0)</td>
</tr>
</tbody>
</table>
### Basis Information

<table>
<thead>
<tr>
<th>Name</th>
<th>Coefficient</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis3</td>
<td>-4.03</td>
<td>Basis0*NOT(MISSING(HorsePower))</td>
</tr>
<tr>
<td>Basis4</td>
<td>0</td>
<td>Basis0*MISSING(HorsePower)</td>
</tr>
<tr>
<td>Basis5</td>
<td>-0.05</td>
<td>Basis3*MAX(HorsePower - 102,0)</td>
</tr>
<tr>
<td>Basis6</td>
<td>0.19</td>
<td>Basis3*MAX(102 - HorsePower,0)</td>
</tr>
<tr>
<td>Basis7</td>
<td>2.67</td>
<td>Basis0*(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)</td>
</tr>
<tr>
<td>Basis8</td>
<td>0</td>
<td>Basis0*NOT(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)</td>
</tr>
</tbody>
</table>

![Graph 1](image1)

![Graph 2](image2)
PROC Adaptivereg

Weitere Nicht-Parametrische Regressionen: PROC GAM, PROC LOESS
MODERNE SAS ANALYTIK ARCHITEKTUREN
MODERNE ARCHITEKTUREN
SAS IST “IN”

IN-tegration mit neuen Plattformen
IN-Database
IN-Memory
IN-Echtzeit
• SAS/ACCESS Interface to
  • PostGreSQL
  • Vertica
  • PI System
  • Impala
  • SAP Hana
  • Hadoop
INTEGRATION SAP HANA

- SAS/ACCESS Interface to SAP Hana
• SAS Scoring Accelerator for SAP HANA
• SAS High-Performance Predictive Modeling Workbench for SAP HANA
Historie:
- von Yahoo für die Entwicklung einer WWW-Suchmaschine aufgegriffen
- an die Apache Foundation übergeben, seitdem open-source (2009)

Technologie:
- Prinzip der horizontalen Skalierung auf kostengünstiger Hardware ("Scale out")
- Prinzip der Datenlokaltät: der Programmcode wird auf die Cluster-Nodes mit den zugehörigen Daten verteilt, dort verteilt ausgeführt und (Teil-)Ergebnisse wieder zusammengeführt
- Structure on read / late binding
- HDFS: Hadoop Distributed File System
- Map/Reduce: ein Verarbeitungsverfahren nach dem "Teile und herrsche"-Ansatz
DAS HADOOP ÖKOSYSTEM

HADOOP IST EINE PROJEKTPLATTFORM MIT HDFS UND MAP/REDUCE ALS KERNMODULEN

Hortonworks HDP

- PLATFORM SERVICES
  - Knox
  - Enterprise Readiness
  - High Availability
  - Disaster Recovery
  - Rolling Upgrades
  - Security and Governance

Hortonworks DATA PLATFORM (HDP)

- OS/VM
- Cloud
- Appliance

Cloudera CDH

NoSQL DB

- Cassandra
  - Column Database
- NEO4J
  - Graph Database
- CouchDB
  - Document Database
- Tamino
  - XML Database

Operational & Platform Services

- Ambari
  - Monitoring and admin.
- Oozie
  - Orchestration and workflow
- Zookeeper
  - Coordination
- Falcon
  - Data Lifecycle Mgt.
- KNOX
  - Security

Data Services

- Pig
  - Scripting
- Hive
  - Query
- HCatalog
  - Metadata
- HBase
  - Column Database
- Flume
  - Data collection
- Sqoop
  - Data Exchange
- Mahout
  - Machine Learning
- WebHDFS
  - Connectivity

Hadoop Core

- HDFS
  - Hadoop Distributed File System
- MapReduce
  - Distributed Processing

SAS Support
Tier 1: Cloudera, Hortonworks
Tier 2: MapR, Pivotal, IBM ..
VISION - WARUM UNTERSTÜTZT SAS HADOOP?

To be the Analytic and Data Management solution of choice for Hadoop.
SAS/Access to Hadoop Push some SAS processing from Hadoop into SAS

Embedded Process - Push SAS data processing to Hadoop with Map Reduce

In-Memory Analytics - Use Hadoop for Storage persistence and commodity computing.

SAS/Access to Hadoop
SAS/Access to Cloudera Impala
SAS DI Server

SAS/Scoring Accelerator for Hadoop
SAS Code Accelerator for Hadoop
SAS Data Quality Accelerator for Hadoop
SAS Data Loader (*EA @ 9.4M2)

SAS Visual Analytics
SAS Visual Statistics
SAS in memory Statistics for HADOOP
SAS HPA Produkt bundles
Base SAS
• Map Reduce + Pig Scripting + HDFS Kommandos

SAS Access to Hadoop
• Hive, Hive2 + eigene Metadaten („Information Maps für HDFS Dateien“)
• Pushdown von Procs: FREQ, RANK, REPORT, SORT, SUMMARY/MEANS & TABULATE

SAS Access to Impala (Cloudera)

SAS Data Integration Studio Transformationen
• Read/Write HDFS files
• Submit HiveQL code
• Execute Map/Reduce code
• Submit Pig Latin
• Transfer data to/from Hadoop using Hadoop utilities
• SQL transforms pushed down with Access to Hadoop engine
• Submit DQ and DS2 code with Code Accelerator

SAS Federation Server
• Datenvirtualisierung und Zugriffsschutz für Hadoop und andere Datenquellen

SAS Event Stream Processing Engine
• Hadoop Adapter für SAS ESP, um Data-Streams im HDFS zu speichern

/* SUBMIT SQL QUERY TO HIVE */
LIBNAME MYHDP HADOOP PORT=10000 SERVER=HDPSRV02;
PROC SQL;
   INSERT INTO MYHDP.CARS_HIVE
   SELECT MAKE, MODEL, MSRP FROM SASHELP.CARS;
QUIT;

/* SUBMIT PIG SCRIPT, HDFS COMMAND, MR JOB */
FILENAME CFG "C:\SAMPLE_DATA\HADOOP_CONFIG.XML";
FILENAME PIGCODE1 "C:\SAMPLE_DATA\PIG_CD.TXT";
PROC HADOOP OPTIONS=CFG;
   PIG CODE=PIGCODE1;
   HDFS DELETE="/USER/HADOOP/OUTPUT_MR1";
   MAPREDUCE INPUT="/..." OUTPUT="/..."
   JAR="/..." MAP="/..." REDUCE="/...";
RUN;

/* COPY FILE FROM HDFS TO LOCAL SAS */
FILENAME CFG "C:\SAMPLE_DATE\HADOOP_CONFIG.XML";
PROC HADOOP OPTIONS=CFG USERNAME="HADOOP"
   PASSWORD="XXXX";
   HDFS COPYTOLOCAL="/USER/HADOOP/TESTFOLDER"
   OUT="C:\SAMPLE_DATA" ;
RUN;
**SAS Scoring Accelerator for Hadoop**
Ausführen von Scoring Modellen aus EM und STAT Projekten

**SAS Data Quality Accelerator for Hadoop (EA in 9.4M2)**
Ausführen von DQ Routinen (Parse, Standardize, Gender Analysis, Identification, Match Code…)

**SAS Code Accelerator for Hadoop**
Ausführen von DataStep2 Code

**SAS Data Loader for Hadoop (EA in 9.4M2)**
Web-basiertes In-Database Tool für Endanwender (z.B. aus den Fachbereichen)

---

**Steckbrief SAS Data Loader 2.1**

- Point & Click Datenmanagement-Routinen, die in Hadoop ausgeführt werden (Abfragen, Tabellen anlegen, SAS Programme)
- In-Memory Beladung: Transfer von Hadoop-Daten in den LASR Server
- “Profile Data” für Hadoop-Daten
- HTML5 basiertes Interface
- Wird als vApp ausgeliefert (läuft in VMWare Player)

**Roadmap**

- Joins in Hadoop
- Sqoop Support
- Data Quality Verfahren
- Metadaten / Lineage
- Security
- Monitoring
SAS/Access to Hadoop Push some SAS processing from Hadoop into SAS

Embedded Process - Push SAS data processing to Hadoop with Map Reduce

In-Memory Analytics - Use Hadoop for Storage persistence and commodity computing.

Hive
Impala
SAS/Access to Hadoop
SAS/Access to Cloudera Impala
SAS DI Server

Score A
Code A
SAS/Scoring Accelerator for Hadoop
SAS Code Accelerator for Hadoop*
SAS Data Quality Accelerator for Hadoop*
SAS Data Loader(*EA @ 9.4M2)

HPA
LASR
SAS Visual Analytics
SAS Visual Statistics
SAS in memory Statistics for HADOOP
SAS HPA Produkt bundles
• Linear Regression
• Multiple Regression
• Logistic Regression
• Analysis of Variance
• K-means Clustering
• Decision Trees
• Random Woods
• Generalized Linear Models
• Density-based Spatial Clustering (DBSCAN)
• Text pre-processing
• Text parsing
• Singular Value Decomposition
• Topic analysis
• Associations
• Implicit and explicit recommendations
• Group By processing
HADOOP ANALYTICS FÜR SPEZIALTHEMEN
INPUT / ANREICHERUNG EINES EDW

Operational Data Sources

EDW

Data Mart

Analytic Mart

BI and Analytics

DATAWAREHOUSE ANWENDUNGS SZENARIEN
1. **Data Management**: SAS optimiert und erleichtert den Zugriff auf Daten in Hadoop.

2. **In-Memory Analytics**: SAS erweitert und beschleunigt Analytik auf Hadoop-Daten.

3. **In-Database Processing**: SAS verlagert (analytische) SAS Funktionalität in das Hadoop Cluster.
• SAS und HADOOP Informationen:
  Code Beispiele:

• Interessante White papers:
  BARC: Big data analytics in der DACH region:

• Tom White: Hadoop : The Definitive Guide (O'Reilly)
  http://shop.oreilly.com/product/0636920021773.do

• Edward Capriolo: Programming Hive (O'Reilly)
  http://shop.oreilly.com/product/0636920023555.do

• Next Steps, wie Starten?
  Test SAS&HADOOP - SAS Testlizenzen, Testumgebungen, Erfahrungsaustausch…
  Kontakt : gernot.engel@sas.com, rainer.sternecker@sas.com
IN-MEMORY ANALYTICS

SAS VISUAL ANALYTICS

DATA BUILDER
- Join data from multiple sources
- Create calculated and derived columns
- Load data

ADMINISTRATOR
- Monitor SAS® LASR™ Analytic server
- Load/unload data
- Manage security

EXPLORER
- Perform ad-hoc analysis and data discovery
- Apply advanced analytics

VISUAL STATISTICS
- Take VA analytics one step further
- Perform statistical modeling and classification

DESIGNER
- Create dashboard style reports for web or mobile

MOBILE BI
- Native iOS and Android applications that deliver interactive reports
• Demo Visual Analytics / Visual Statistics
IN-DATABASE ÜBERBLICK

SAS IN-DATABASE CAPABILITIES
- SQL Pushdown
- Reporting / OLAP
- ETL Integration
- Analytic Discovery
- Model Deployment & Management
- Data Manipulation/ Data Quality

DATABASE ENVIRONMENT

MANAGEMENT ADMINISTRATION SECURITY
Traditional data mining and statistical modeling → Scoring files generated → Models published using Embedded Process → Files passed through SAS/ACCESS® → Scoring programs inserted into DB → Models scored by users using stored procedure

- SAS® Enterprise Miner™
- Scoring Code XML File Formats
- SAS Program Editor (EM models)
- Or
- SAS® Model Manager (EM or STAT models)
- SAS/STAT®
<table>
<thead>
<tr>
<th>High-Performance Statistics</th>
<th>High-Performance Data Mining</th>
<th>High-Performance Text Mining</th>
<th>High-Performance Optimization</th>
<th>High-Performance Econometrics</th>
<th>High-Performance Forecasting</th>
</tr>
</thead>
<tbody>
<tr>
<td>• HPLOGISTIC</td>
<td>• HPREDUCE</td>
<td>• HPTMINE</td>
<td>• OPTLSO</td>
<td>• HPCOUNTREG</td>
<td>• HPFORECAST</td>
</tr>
<tr>
<td>• HPREG</td>
<td>• HPNEURAL</td>
<td>• HPTMSCORE</td>
<td>• Select features in</td>
<td>• HPSEVERITY</td>
<td></td>
</tr>
<tr>
<td>• HPLMIXED</td>
<td>• HPFOREST</td>
<td></td>
<td>• OPTMILP</td>
<td>• HPQLIM</td>
<td></td>
</tr>
<tr>
<td>• HPNLMOD</td>
<td>• HP4SCORE</td>
<td></td>
<td>• OPTLP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• HPSPLIT</td>
<td>• HPDECLARE</td>
<td></td>
<td>• OPTMODEL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• HPGENSELECT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Common procedures (HPDS2, HPDMD, HPSAMPLE, HPSUMMARY, HPIMPUTE, HPBIN, HPCORR)
SAS® ESP

- ermöglicht und erleichtert die Verarbeitung von Massendaten (Sensoren, Telematik, Netzwerkverkehr, …)
- kann Daten in Bewegung verarbeiten, ohne sie speichern zu müssen
- kann als Frontend für beliebige Systeme eingesetzt werden
- skaliert sehr gut (mehrere Mio. Events/Sek), bei extrem niedrigen Antwortzeiten (Millisekunden)
- arbeitet mit Datenflüssen (gerichteten Graphen) deren Knoten u.a. SQL-Operationen (wie aggregate, join, union, compute, filter, copy) und pattern matching repräsentieren
**Event Stream Processing**

- Publish / subscribe
- Inserts/updates/deletes
- Continuous queries
  - Aggregate (group by)
  - Correlate (join)
  - Compute
  - Filter
  - Procedural
  - User defined functions
  - Retention windows
  - Pattern matching
- Ad-hoc queries
- Command & control
- Security (Auth, Encrypt, AC)
- Persistence / recovery
- Fail-over
- Distributed services
- Model management

### Event Sources
- Streams
- Applications
- Databases
- Telemetry

### Processed Event Consumers
- Mobile BI / BAM Alerts
- Workflows
• Demo VA in Echtzeit
SAS VISUAL ANALYTICS

WORD CLOUDS – TOPICS & TERMS

Visualization 2

Topic: you, rogerfederer, i, the

A word cloud of body

be he back

if i think my feel on

see know

will

get in team hope

luck can

now your win

switzerland

The well

feed soon

what better

do not now

that not

Results Topics Documents Analysis

<table>
<thead>
<tr>
<th>Term</th>
<th>Topic Term Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>0.357</td>
</tr>
<tr>
<td>rogerfederer</td>
<td>0.254</td>
</tr>
</tbody>
</table>
WORD CLOUDS – SENTIMENT

SAS VISUAL ANALYTICS

Visualization 2

Topic: hint, reconciliation, tweet, picture

- you, the, i, rogerfederer
- make, against, time, will
- fitness, battle, cup, fitness, battle, davis
- federer injury, sweat, over, swiss sweat
- s, ever, first, davis cup win
- cloud, dream, report, davis cup dream
- down, play, after, spot
- hint, reconciliation, tweet, picture
- fall, until, v, out
- surprise, myself, thane, have, start
- doubt, say, djokovic, novak djokovic

Word cloud:
- wawrinka
- switzerland
- twitter
- roger
- reconcile
- stanislas
- world
- no2
- federer
- together

Topic term weight

Total Documents: 95

Body

- Roger Federer hints at reconciliation with Stanislas Wawrinka http://it.co/4jcg.WpN0QH8
  Sentiment: 0.60
  Relevance: 0.21
- Roger Federer hints at reconciliation with Stanislas Wawrinka | Sport | The Guardian http://it.co/4kxyKRmUcB
  Sentiment: 0.60
  Relevance: 0.21
<table>
<thead>
<tr>
<th>Term</th>
<th>Topic Term Weight</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>twitter</td>
<td>0.083</td>
<td>Proper noun with ambiguous classi...</td>
</tr>
<tr>
<td>via</td>
<td>0.081</td>
<td>Preposition</td>
</tr>
<tr>
<td>they</td>
<td>0.078</td>
<td>Pronoun</td>
</tr>
<tr>
<td>together</td>
<td>0.077</td>
<td>Adverb</td>
</tr>
<tr>
<td>federer</td>
<td>0.076</td>
<td>Proper noun with ambiguous classi...</td>
</tr>
<tr>
<td>switzerland</td>
<td>0.072</td>
<td>Geographical place</td>
</tr>
<tr>
<td>swap</td>
<td>0.065</td>
<td>Verb</td>
</tr>
</tbody>
</table>