

1 SAS-Hilfe

- **in SAS:** Menü Hilfe → SAS Hilfe und Dokumentation
dann beispielsweise: Inhalt → SAS-Produkte
 - Base SAS → SAS-Prozeduren → SAS-Prozeduren: CORR, FREQ, UNIVARIATE
 - SAS/STAT → SAS/STAT User's Guide: Übersicht über statistische Prozeduren
- **SAS-Online Dokumentation:**
zu finden über Internetsuche nach "sas online documentation"
 - Link: SAS OnlineDoc "version" for the Web
dann beispielsweise weiter mit Link: "SAS Procedures"
anschließend Link zu interessierender Prozedur
 - oder pdf-Dateien zum Herunterladen
- **Internetsuche:** erstes Stichwort SAS, anschließend z.B. Name der Prozedur zu empfehlende Suchergebnisse
 - SAS-Hilfeseiten von UCLA Academic Technology Services
(<http://www.ats.ucla.edu/stat/sas/...>):
grundlegende Information mit einfachen Beispielen - für Standardanwendungen
 - <http://www2.sas.com/proceedings/sugi...>:
Veröffentlichungen von SAS - sehr detaillierte, aber trotzdem kurze Information zu einem Thema - für eher spezielle Fragestellungen
- **Quick Reference:** verlinkt auf Homepage zur Veranstaltung
enthält unter anderem hilfreiche Befehle / Begriffe
 - zum Arbeiten mit Wörtern: Abschnitt Character (z.B. SUBSTR)
 - zu arithmetischen Funktionen: Abschnitt Arithmetic (z.B. arithmetische Funktionen wie ABS, CEIL, FLOOR, ROUND)
 - zu Wahrscheinlichkeitsfunktionen: Abschnitt Probability
 - zu Zufallszahlenfunktionen: Abschnitt Random Numbers
 - zu statistischen Funktionen: Abschnitt Statistical (z.B. MAX, MIN, MEAN)
 - zu Datums- und Zeitfunktionen: Abschnitt Dates and Times
 - zu Formaten: informats und formats

Einige dieser Befehle können innerhalb von Makros mit einem % vor dem Befehl auch als Makro-Befehle verwendet werden.

2 Das SAS-System

- SAS: “Statistical Analysis System”
- momentane Version im CIP-Pool: 9.4
- Vorteile:
 - umfangreich und flexibel
 - geeignet auch für sehr große Datenmengen
 - eigene Programmiermöglichkeiten
- Nachteile:
 - Präsentationsgrafiken aufwändig zu erstellen
 - teuer
- SAS Institute: Firmensitz in Cary, North Carolina
- <http://www.sas.com/offices/europe/germany> <http://www.sas.com/pubs>

2.1 Modularer Aufbau

Das SAS-System setzt sich aus mehr als 20 einzelnen Komponenten zusammen, den **SAS-Modulen**.

Ein Modul übernimmt spezielle Aufgaben und ist prinzipiell unabhängig von den anderen Modulen.

Jedes Modul setzt sich aus einer oder mehreren Prozeduren zusammen.

- SAS/BASE: Anlegen und Verwalten von Dateien, Prozeduren: PRINT, SORT, MEANS, ...
Grundlage für alle anderen Module
- SAS/STAT: Bündelung statistischer Prozeduren, Prozeduren: TTEST, REG, GLM, CORR, ...
- SAS/GRAPH: Erzeugen von Grafiken, Prozeduren: GCHART, GPLOT, ...
- SAS/INSIGHT: explorative Datenanalyse, nur eine Prozedur: INSIGHT
- SAS/IML: interaktive Programmiersprache
- SAS/ASSIST, SAS/ETS, SAS/FSP, SAS/AF, SAS/ACCESS, ...

2.2 Die SAS-Oberfläche

- Programmfenster (Enhanced/Program Editor):
Erstellen und Ausführen von Programmen
Öffnen und Speichern von Programmen erfolgt über das Menü, d.h.im Gegensatz zu anderen Programmpaketen wird über die Menüfolge *File*→*Open* **keine** Datendatei geöffnet.
optionale Zeilennummerierung (Kommando *num*)
- Protokollfenster (Log):
Protokollierung aller ausgeführten Programme
Hinweise, Warnungen, Status- und Fehlermeldungen
- Ergebnis-/Verzeichnisfenster (Results, Explorer):
Results: Verwaltung der erzeugten Outputs und Grafiken
Explorer: Verwaltung der SAS-Bibliotheken und ihrer SAS-Daten-Dateien
- Ausgabefenster (Output):
Ergebnisse der ausgeführten SAS-Programme in Tabellen, Listen und einfachen (zeichenbasierten) Grafiken → Listenoutput
Befindet sich zunächst im Hintergrund, rückt durch neue Ausgabe in den Vordergrund
Kumulative Ansammlung aller erzeugten Outputs → zwischendurch löschen
- Grafikfenster (Graph):
Darstellung von Präsentationsgrafiken
Öffnet sich erst nach Verwendung von Grafik-Prozeduren
- Kommandozeile
- Symbolleiste
- Menüleiste (→ Online-Hilfe)
- CIP-Pool: Hilfe über Start zu erreichen (Start → Programme → Statistik → SAS-Sytem → SASonlinedoc.doc)

2.3 SAS-Programme und SAS-Dateien

SAS-Datei (data set/data file, *.sas7bdat, *.sd7):

Organisation der Daten in Form einer Tabelle
(Zeilen: Beobachtungen, Spalten: Variablen)
Fehlende Werte werden mit “.” codiert.

SAS-Programme (*.sas):

werden im Programmfenster geschrieben und an das SAS-System übergeben (→ Protokollfenster, evtl. Output).

Man unterscheidet in einem SAS-Programm:

- Globale Anweisungen (z.B. OPTIONS...;)
- Datenschnitte (DATA steps, Einlesen und Bearbeiten von Daten)
- Prozedurschnitte (PROC steps, statistische Analysen, Grafiken)
- Kommentare (* ...; bzw. /* ... */)

2.4 Mögliche Struktur eines SAS-Programms

```
OPTIONS ...; /* Globale Einstellungen */
TITLE ...; /* Titelzeile */
... /* Weitere globale Anweisungen */

DATA ...; /* Datenschnitt */
...
RUN;

PROC ...; /* Prozedurschnitt */
...
RUN;

.....; /* Weitere Daten-und
... /* Prozedurschnitte */
RUN;
```

- Abarbeitung von oben nach unten
- keine Parallelverarbeitung
- Daten- und Prozedurschnitte bestehen i.d.R. aus mehreren Anweisungen, die jeweils mit einem Schlüsselwort beginnen und mit einem Semikolon enden.

2.5 Typischer Ablauf

- Schreiben/Öffnen eines Programms im Programmfenster
- Ausführen des SAS-Programms durch:
 - a) Icon (laufendes Männchen)
 - b) Taste F3
 - c) Menü (Run → Submit)
- Nachprüfen der Fehlermeldungen/Warnungen im LOG-Fenster
→ gegebenenfalls Modifizierung
- Betrachtung der Ergebnisse im Output-Fenster
→ gegebenenfalls Modifizierung

Durch Markieren einzelner Daten-/Prozedurschritte kann das SAS-Programm auch nur teilweise abgeschickt werden.

3 Anlegen von Arbeitsverzeichnissen

1. Definition der permanenten Bibliothek (Verzeichnis, siehe Explorer-Fenster)
Der Name der Bibliothek ist frei wählbar.
LIBNAME saskurs "u:\user";
→ Globale Anweisung (nicht innerhalb eines Datenschnitts !)
2. Verwendung von zweiteiligen Dateinamen im Datenschnitt
DATA saskurs.groesse;
SET groesse;
RUN;
DATA saskurs.groesse2;
SET saskurs.groesse;
...;
RUN;
3. Falls mit dem Datensatz Formate verknüpft sind, muss eine Bibliothek namens "library" angelegt werden, die auf den Ordner, in dem die Formate abgelegt sind, verweist. Diese Name ist nicht frei wählbar!
LIBNAME library "u:\user";
4. Um mit einem Datensatz zu arbeiten, empfiehlt es sich, diesen in die Work-Bibliothek zu legen. (Vermeidet Datenverlust und erhöht Rechengeschwindigkeit.)
DATA groesse;
SET saskurs.groesse;
RUN;

Es empfiehlt sich, den Pfad als Makro-Variable anzulegen:

```
%LET path=u:\user;
```

Der Pfad verweist auf den Ordner, in dem die Daten liegen bzw. abgelegt werden sollen.

Die Eingabe kann durch das %PUT-Statement überprüft werden:

```
%PUT &path.;
```

Vor dem Namen der Makro-Variable steht ein „&“, danach ein „.“.

Anschließend kann die Library über das Makro definiert werden:

```
LIBNAME saskurs "&path.";
```

Wichtig: doppelte Anführungszeichen!

Es sind verschiedene LIBNAME-Anweisungen möglich, die auch auf dasselbe Verzeichnis verweisen können. Es können auch mehrere Pfad-Makros definiert werden, z.B. ein Verweis auf den Ordner, in dem die Daten liegen namens „path.data“ und ein Verweis auf den Ordner, in dem die Ergebnisse abgelegt werden sollen namens „path.results“.