

Einführung in die SAS Enterprise Guide Software für SAS-Anwender

Carina Ortseifen
Universitätsrechenzentrum Heidelberg
Im Neuenheimer Feld 293
69120 Heidelberg
carina.ortseifen@urz.uni-heidelberg.de

Zusammenfassung

Mit SAS 9 steht (fast) automatisch jedem SAS-Anwender unter Windows die SAS/Enterprise Guide Software zur Verfügung. Diese benutzerfreundliche Menüoberfläche erlaubt dem gelegentlichen Anwender schneller zu statistischen Ergebnissen mit ansprechenden Tabellen und Grafiken zu kommen. Der routinierte SAS-Anwender, der schon jahrelang mit dem ‚alten‘ SAS arbeitet und mit den Eigenheiten der Standardfenster vertraut ist, fragt sich dagegen vielleicht, ob er auch einen Nutzen aus dieser Neuerung ziehen kann. In dieser Einführung werden zunächst die prinzipiellen Arbeitsschritte vorgestellt, die beim Arbeiten mit der SAS/Enterprise Guide Software notwendig sind. Anschließend wird für den erfahrenen SAS-Anwender gezeigt, wie Aktionen, die bislang mit dem Datenschnitt realisiert wurden, mit der SAS/EG Software umgesetzt werden, wie vorhandener bzw. vom SAS/EG erzeugter Programmcode verwendet werden kann und einiges andere mehr. Abschließend werden Stolperfallen aufgezeigt, die nicht nur Anfängern den Einstieg schwer machen könnten.

Schlüsselworte: Enterprise Guide, SAS Enterprise Guide, Projekt, Tutorium, Einführung, Installation, Hotfix.

1 Prinzipielle Schritte beim Arbeiten mit der SAS/EG Software

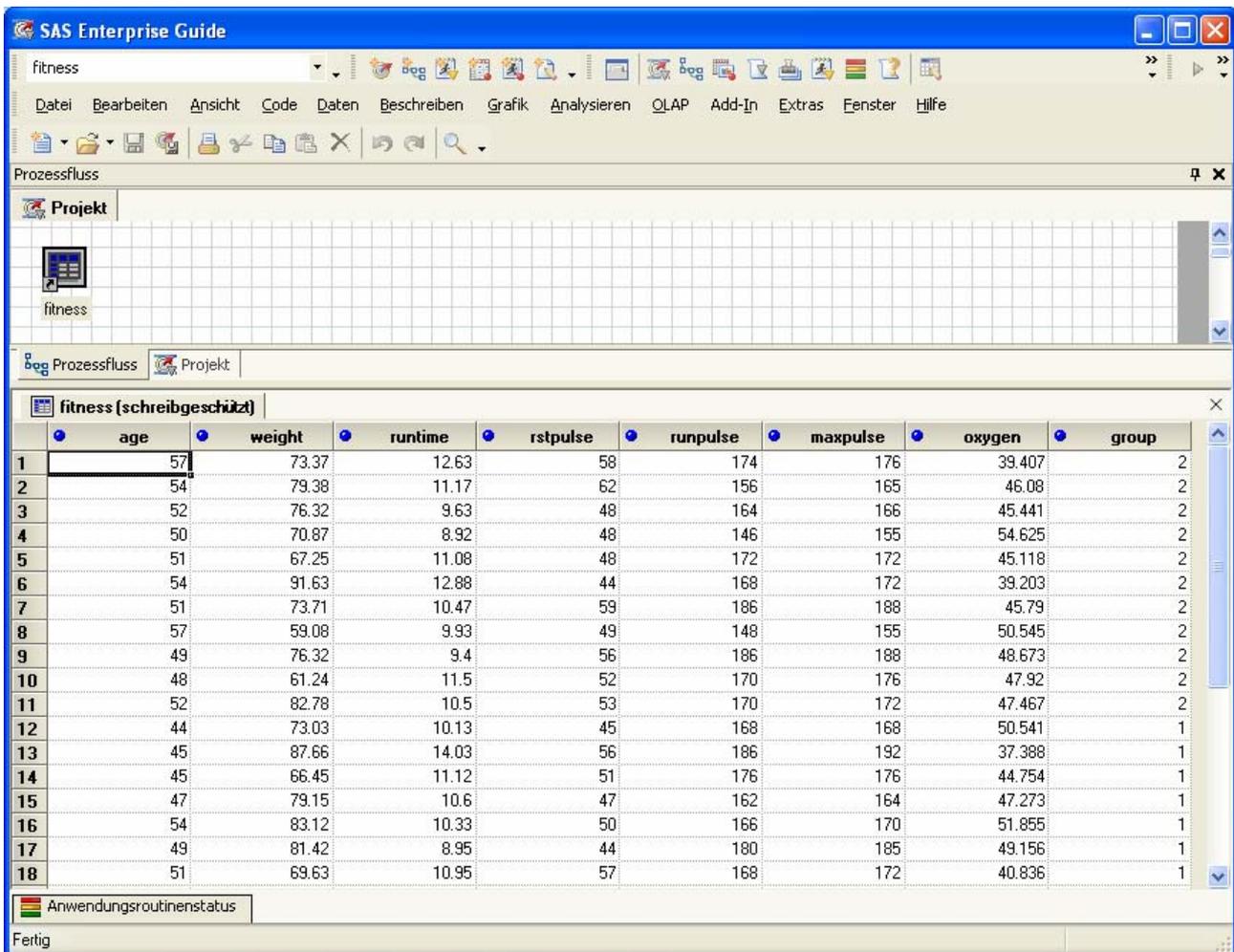
Die notwendigen Schritte, um Kennwerte aus einer SAS-Tabelle zu ermitteln, diese in einem Balkendiagramm darzustellen und für eine Veröffentlichung vorzubereiten, werden exemplarisch an der SAS-Tabelle Fitness erläutert. Diese Tabelle wird von SAS Institute mit der Software ausgeliefert und befindet sich im Enterprise Guide Systemordner Sample\Data.

C. Ortseifen

Als erstes wird ein neues Projekt angelegt (Menü *Datei* > *Neu* > *Projekt*). In dieses Projekt werden im Laufe der weiteren Arbeiten Tabellen eingetragen, die geöffnet bzw. neu definiert werden, sowie Analysen und Grafiken mitsamt ihrer Ergebnisse und die Protokolldateien der ausgeführten SAS-Anweisungen (analog zum Log-Fenster) abgelegt.

Nach dem Anlegen dieses neuen Projekts wird die Tabelle *Fitness.sas7bdat* über *Datei* > *Öffnen* > *Von meinem Computer* geöffnet. Dazu benötigt man keinen Bibliotheksnamen, sondern den Ordner, in dem die Tabelle gespeichert ist.

Die gewählte Tabelle wird in einem Datenfenster mit Schreibschutz abgelegt, um ungewollte Veränderungen zu verhindern. (Denn jede Änderung wird sofort wirksam, ohne Möglichkeit der Rückgängigmachung.)



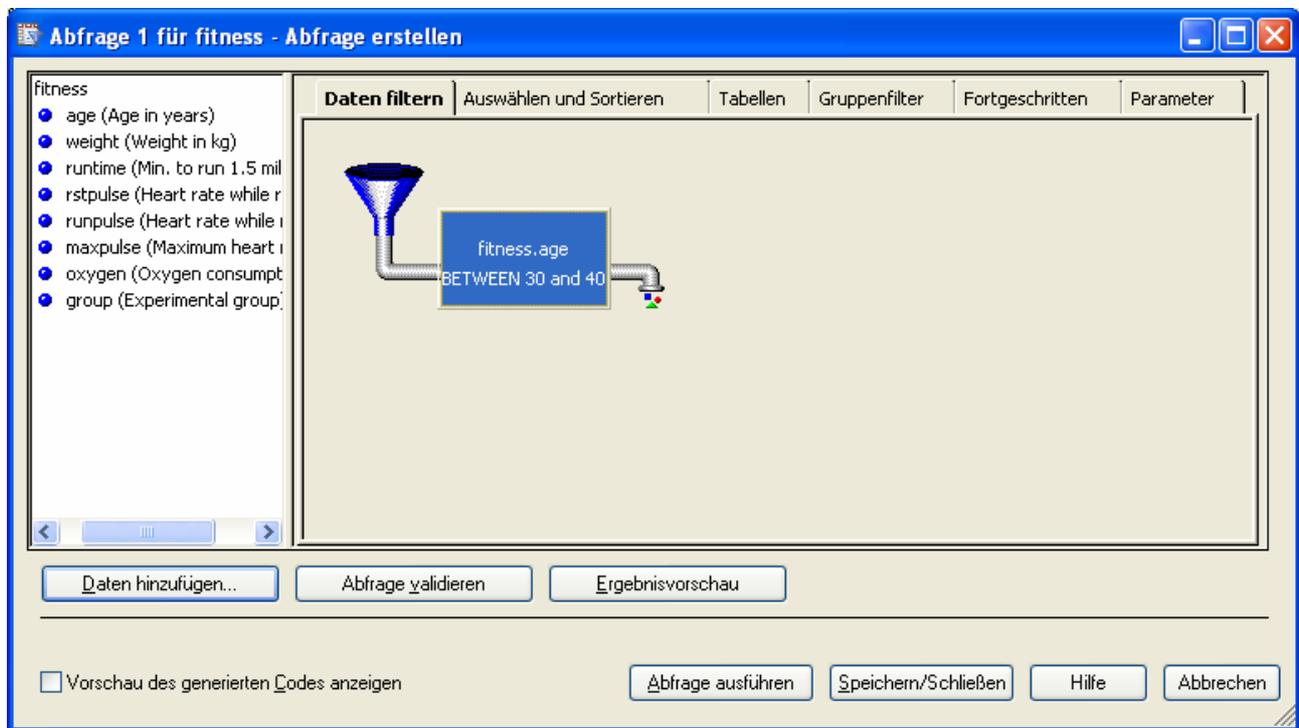
The screenshot displays the SAS Enterprise Guide interface. At the top, the title bar reads 'SAS Enterprise Guide'. Below it, the window title is 'fitness'. The menu bar includes 'Datei', 'Bearbeiten', 'Ansicht', 'Code', 'Daten', 'Beschreiben', 'Grafik', 'Analysieren', 'OLAP', 'Add-In', 'Extras', 'Fenster', and 'Hilfe'. The toolbar contains various icons for file operations and analysis. The main workspace is divided into two panes: 'Prozessfluss' (Process Flow) and 'Projekt' (Project). The 'Projekt' pane shows a tree view with a folder named 'fitness'. Below this, the 'fitness (schreibgeschützt)' data table is displayed in a grid view. The table has 18 rows and 10 columns: 'age', 'weight', 'runtime', 'rstpulse', 'runpulse', 'maxpulse', 'oxygen', and 'group'. The 'group' column has values 1 or 2. At the bottom of the window, there is a status bar with a traffic light icon and the text 'Anwendungsroutinenstatus' and 'Fertig'.

	age	weight	runtime	rstpulse	runpulse	maxpulse	oxygen	group
1	57	73.37	12.63	58	174	176	39.407	2
2	54	79.38	11.17	62	156	165	46.08	2
3	52	76.32	9.63	48	164	166	45.441	2
4	50	70.87	8.92	48	146	155	54.625	2
5	51	67.25	11.08	48	172	172	45.118	2
6	54	91.63	12.88	44	168	172	39.203	2
7	51	73.71	10.47	59	186	188	45.79	2
8	57	59.08	9.93	49	148	155	50.545	2
9	49	76.32	9.4	56	186	188	48.673	2
10	48	61.24	11.5	52	170	176	47.92	2
11	52	82.78	10.5	53	170	172	47.467	2
12	44	73.03	10.13	45	168	168	50.541	1
13	45	87.66	14.03	56	186	192	37.388	1
14	45	66.45	11.12	51	176	176	44.754	1
15	47	79.15	10.6	47	162	164	47.273	1
16	54	83.12	10.33	50	166	170	51.855	1
17	49	81.42	8.95	44	180	185	49.156	1
18	51	69.63	10.95	57	168	172	40.836	1

Der Schreibschutz lässt sich über *Daten > Schreibschutz* einfach entfernen, was für die folgenden Aktionen aber nicht erforderlich ist.

Das gesamte Projekt kann sowohl als Projektfluss oder als Projektbaum (Projekt) dargestellt werden.

Im nächsten, dem zweiten Schritt sollen nun bestimmte Fälle ausgewählt werden. Das Filtern von Beobachtungen führt man über *Daten > Aktive Daten filtern und auswählen* durch. Als Beispiel werden die Beobachtungen gewählt, deren Altersangabe zwischen 30 und 40 Jahren liegen.

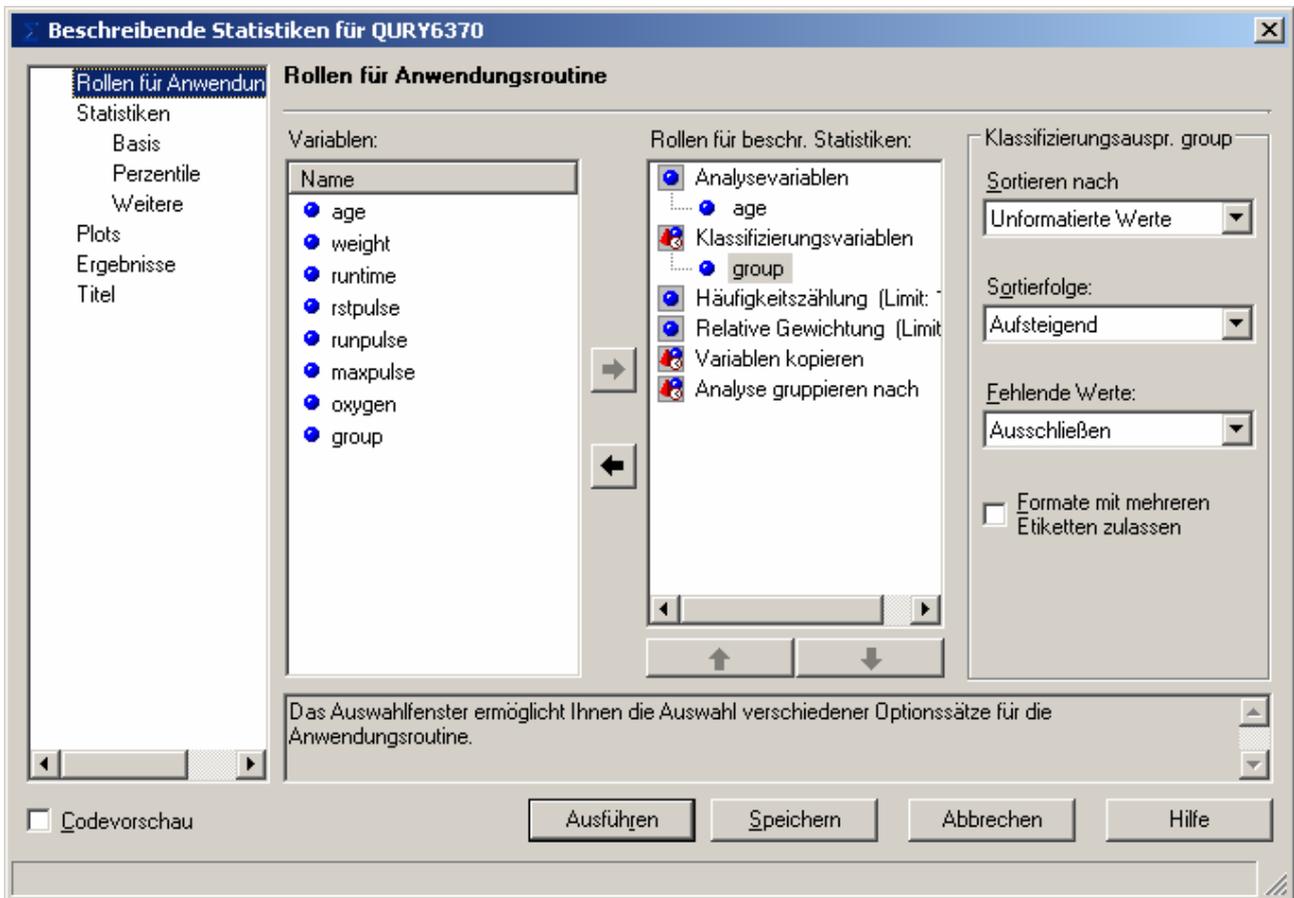


Mit der *Ergebnisvorschau* überprüft man die getroffene Auswahl. Mit *Abfrage ausführen* wird der Filtervorgang tatsächlich ausgeführt. Man erhält damit eine neue SAS-Tabelle mit den ausgewählten Beobachtungen.

Klickt man dagegen nur auf die Schaltfläche *Speichern/Schließen*, werden die Filteraktionen gespeichert, aber der Vorgang selbst erst zu einem späteren Zeitpunkt (mittels Klick mit der rechten Maustaste auf dem Projektbaum-Eintrag und *ausführen*) vollzogen.

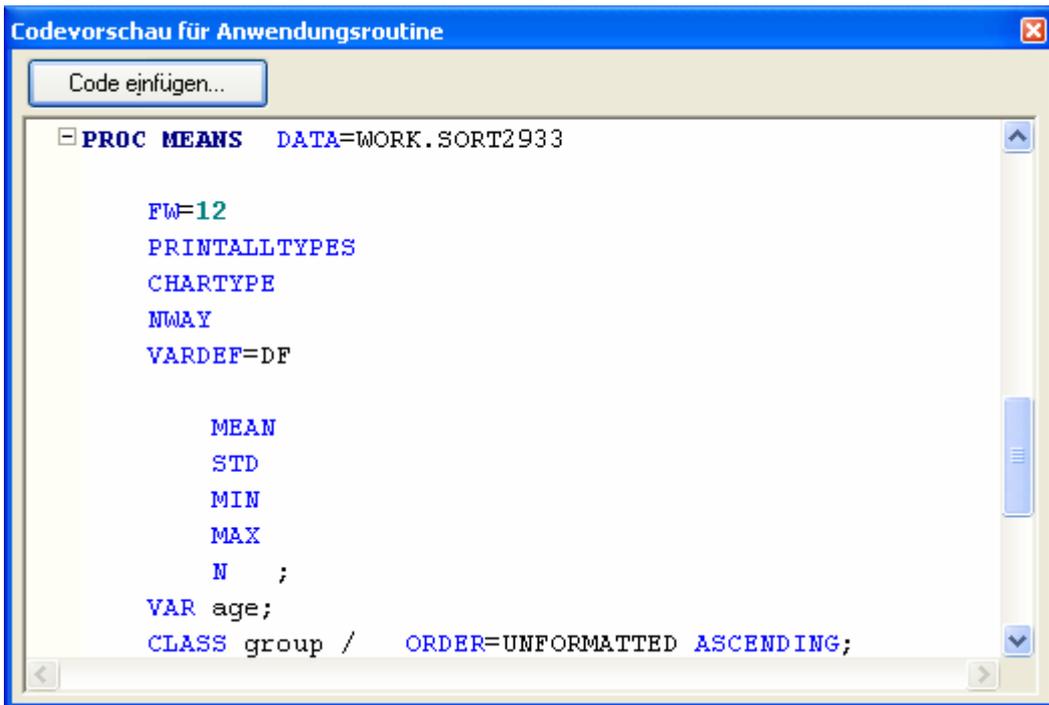
C. Ortseifen

Über das Menü *Beschreiben* > *Beschreibende Statistiken* werden als nächstes die Kennwerte (mit der Prozedur Means) berechnet.



Die Variable Age spielt dabei die Rolle der Analysevariablen, Group die der Klassifizierungsvariablen. Über Statistiken wählt man die Kennwerte, die berechnet werden sollen, und eventuell über Plots zusätzliche Grafiken.

Aktiviert man die Codevorschau, zeigt die SAS/EG Software das Programm an, das ausgeführt wird.



```

Codevorschau für Anwendungsroutine
Code einfügen...
PROC MEANS DATA=WORK.SORT2933

  FW=12
  PRINTALLTYPES
  CHARTYPE
  NWAY
  VARDEF=DF

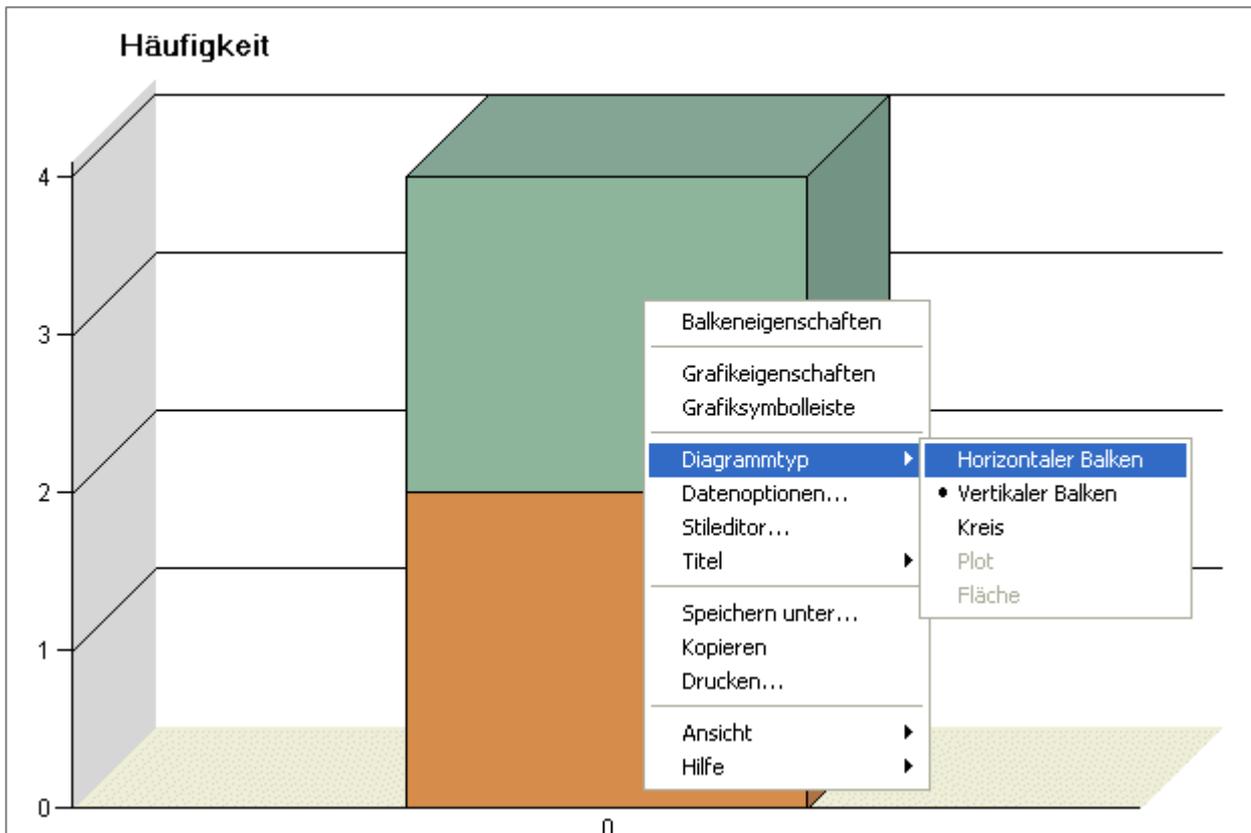
  MEAN
  STD
  MIN
  MAX
  N ;
VAR age;
CLASS group / ORDER=UNFORMATTED ASCENDING;

```

Ein stärker zu beeinflussendes Balkendiagramm erhält man über *Grafik > Balkendiagramm > Gestapelt > Vertikaler Balken*. Um die Analogie zu den Kennwerten zu erhalten, werden die Spalten für die Diagramme mit Age belegt und die Balken nach Group gestapelt. Die Grafik kann dann auch noch nachträglich mittels Klick mit rechter bzw. linker Maustaste in ihren Einstellungen geändert werden.

Das Standardformat der berechneten Kennwerte und Diagramme ist HTML. Klickt man im Projektbaum mit der rechten Maustaste auf ein HTML-Objekt, kann man über *Exportieren* bzw. *Senden an* dieses Objekt aus SAS exportieren. Mit dem „Dokumentersteller“ (*Extras > Dokument erstellen*) lassen sich mehrere Ausgaben auf einer HTML-Seite zusammenfassen.

Um stattdessen oder zusätzlich die Ergebnisse in RTF- oder PDF-Format zu erhalten, muss man vor Ausführung der Kennwert- bzw. Grafikmenüs über *Extras > Optionen > Ergebnisse* die gewünschten Ausgabeformate (Ziele oder Kanäle, wie diese im Output Delivery System (ODS) auch genannt werden) auswählen. Die weitere Verarbeitung geschieht dann analog, mit Ausnahme der Dokumenterstellung, die nur für das HTML-Format funktioniert.



Das Projekt kann schließlich abgespeichert und zu einem späteren Zeitpunkt wieder geöffnet werden. Physikalisch erhält die Datei die Endung *.epg. (In Vorgängerversionen lautete diese Dateierweiterung *.seg.) Die SAS-Tabellen, die im Projekt geöffnet und erzeugt wurden, werden lediglich als Verknüpfung gespeichert (Vorsicht beim Übertragen!). Die Programmcodes, Notizen, Analysen und Ausgabeformen dagegen sind im Projekt physikalisch abgelegt.

Das waren kurz umrissen die prinzipiellen Arbeitsschritte. Ausführlicher erläutert werden diese im SAS-Anwenderhandbuch im Netz (siehe Abschnitt 5).

2 SAS/EG für SAS-Anwender

Mit der Einführung in die wichtigsten Arbeitsschritte im ersten Abschnitt hat der mit der ‚normalen‘ SAS-Umgebung bereits vertraute Anwender vielleicht bereits gesehen, wo die SAS/Enterprise Guide Software ihre Vorzüge hat. Auf der anderen Seite interessiert ihn nun aber auch, in welcher Form ein Datenschnitt ausgeführt werden

kann, wie sich neue Variablen erzeugen lassen und andere Dinge mehr. Einige, der Autorin wichtig erscheinende Aktionen, sollen im vorliegenden Abschnitt näher betrachtet werden.

2.1 Der Datenschnitt

Der langjährige SAS-Anwender hat sich beim Lesen bis zu dieser Stelle vielleicht schon gefragt: Wo bleibt der Datenschnitt? Diesen wird man im SAS Enterprise Guide vergeblich suchen, denn alle Dateioperationen werden mit der Prozedur SQL umgesetzt. Durch die Formulierung der Menüpunkte *Daten > Aktive Daten filtern und abfragen* sowie *Daten > Neue leere Abfrage erstellen* wird dies offensichtlich, wenn man weiß, dass SQL die Abkürzung für Structured Query Language ist und Query als Abfrage übersetzt wird.

Anmerkung: Wenn auch die SAS Enterprise Guide Software nicht mit Datenschnitten arbeitet, können Sie diese trotzdem verwenden. Öffnen Sie ein neues Code-Fenster (*Datei > Neu > Code*) und tragen Sie den auszuführenden Datenschnitt ein.

2.2 Neue Variablen erzeugen

Um nun neue Variablen anzulegen, was im Datenschnitt mittels Zuweisung realisiert wird, stehen mit der Enterprise Guide Software zwei Möglichkeiten zur Verfügung:

- (1) Eine direkte Methode im Datenfenster, die allerdings keine Umkehr oder Wiederholung erlaubt und
- (2) eine Abfrage.

Für die direkte Methode fügt man im Datenfenster über das rechte Maustastenmenü eine Spalte ein (*RM Spalte einfügen*) und definiert deren Werte im Feld *Ausdruck*.

The screenshot shows a dialog box for creating a new variable. The fields are as follows:

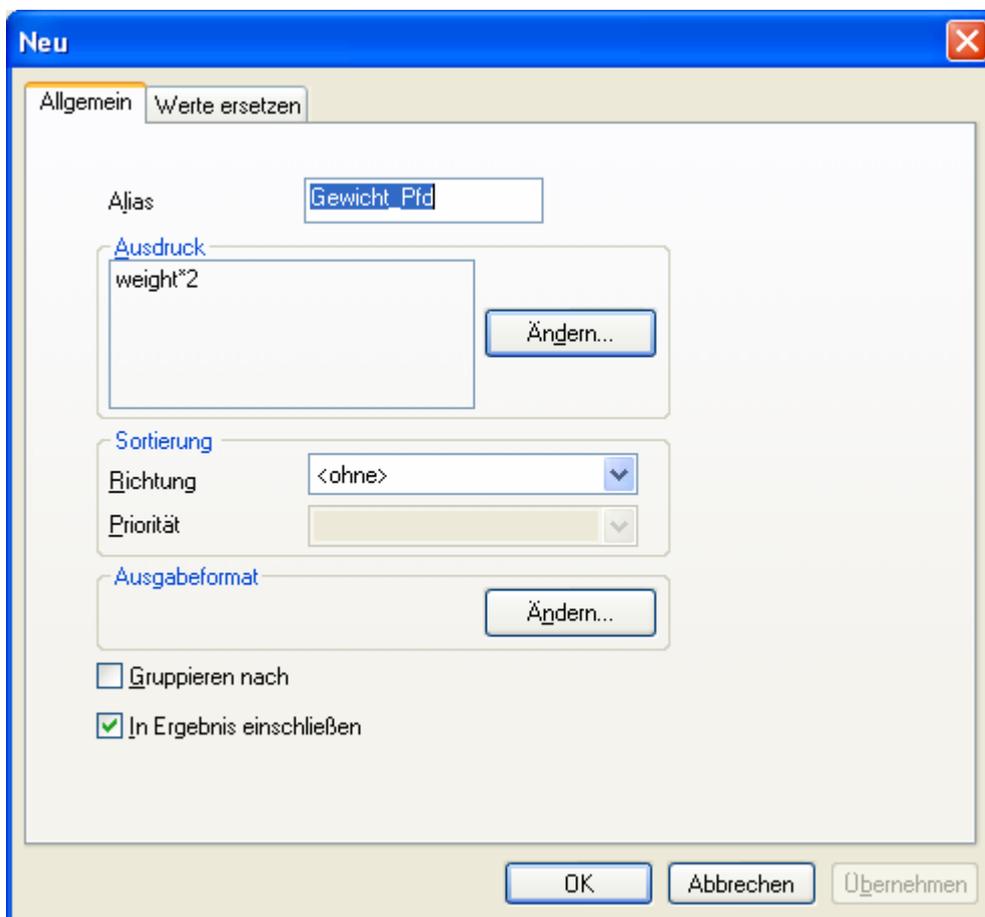
- Name:** Gewicht
- Etikett:** Gewicht in Pfund
- Typ:** Numerisch (dropdown menu)
- Länge (in Byte):** 8
- Gruppe:** Numerisch (dropdown menu)
- Ausdruck:** weight*2

A warning message is displayed: "Hinweis: Der Ausdruck wird nur einmal angewendet und nicht gespeichert."

C. Ortseifen

Nachteil dieses Vorgehens: Werden anschließend in die Tabelle neue Beobachtungen eingefügt, wird der Ausdruck für diese nicht berechnet, d.h. der Wert der neuen Variablen muss von Hand eingetragen werden oder der Vorgang (neue Variable einfügen) muss wiederholt werden. Auch lassen sich diese Änderungen nicht rückgängig machen. Die Tabelle bleibt verändert, es sei denn, man löscht die neue Spalte von Hand.

Die Abfrage ist dagegen wesentlich flexibler. Der Vorgang lässt sich jederzeit wiederholen, etwa wenn neue Beobachtungen an die Tabelle hinzugefügt oder Datenwerte verändert werden. Der entsprechende Menüpunkt dazu lautet *Daten > Aktive Daten filtern und abfragen*. Dort wählt man im Register *Auswählen und Sortieren* die Schaltfläche *Neu*.



2.3 Weitere Möglichkeiten mit Abfragen

Neben Filtern (Bildung von Teilmengen) und dem Erzeugen neuer Variablen können über den Menüpunkt *Daten > Aktive Daten filtern und abfragen* noch weitere Aktionen durchgeführt werden.

- Umkodieren von Werten und Klassenbildung (Register Auswählen und Sortieren, Schaltfläche Neu, Register Werte ersetzen)
- Vertauschen der Reihenfolge der Variablen (Register Auswählen und Sortieren, Variable anklicken und über die Schaltflächen Nach unten/Nach oben die gewünschte Position festlegen)
- Tabellen verknüpfen

2.4 Verknüpfen von Tabellen

Für das Verknüpfen von Tabellen gibt es zwei prinzipielle Möglichkeiten:

1. Untereinander, im Datenschnitt mit der Anweisung Set realisiert. Dabei wird eine Tabelle an eine andere Tabelle angehängt, die zusätzliche Beobachtungen enthält.
2. Nebeneinander, entsprechend der Anweisung Merge. An eine Tabelle wird dabei eine zweite angehängt mit weiteren Variablen.

Für die Verknüpfungsvariante 1 bietet die Enterprise Guide Software den Menüpunkt *Daten > Tabelle anhängen*. Ausgeführt wird ein Proc Sql-Schritt mit einem Outer Union Corr-Befehl.

Die Verknüpfungsvariante 2 wird über den Menüpunkt *Daten > Aktive Daten filtern und abfragen* und die Registerkarte *Tabellen* sowie die Schaltfläche *Daten hinzufügen* umgesetzt. Ausgeführt wird dabei wieder ein Proc Sql-Schritt, diesmal aber mit einem Inner/Left/Right oder Full Join-Befehl.

Die Ergebnisse von Abfragen sind per Default SAS-Tabellen, weil ein Create Table ... As -Befehl ausgeführt wird. Alternativ kann aber auch eine Datenansicht (Create View) oder ein Bericht (Sql ohne Create) erzeugt werden. Diese Einstellung legt man über das Menü *Extras > Optionen > Abfragen* fest.

2.5 Programmcode

Zu jeder Anwendungsroutine wird der verwendete SAS-Code im Projektfenster angezeigt. Dieser Code kann kopiert, modifiziert und ausgeführt werden (rechte Maustaste Öffnen und Editerversuch unternehmen).

C. Ortseifen

In diesem Programmcode stehen allerdings nicht alle verwendeten Anweisungen drin. Erst wenn man unter *Extras > Optionen > Anwendungsroutinen > Anw.routinen Allg.* die Option *Gesamten generierten SAS-Code in Ausgabe anzeigen* aktiviert, sieht man wirklich alle ausgeführten Anweisungen.

2.6 Bibliotheken

Die SAS/Enterprise Guide Software vergibt beim Öffnen von Tabellen über das Menü *Datei > Öffnen* interne Bibliotheksnamen ECLIB000, ECLIB001 usw. Diese Bibliotheksnamen und –anweisungen werden in einem neuen (bzw. editierbaren) Code-Fenster angezeigt, wenn man die Tabelle aus dem Projektfenster bei gedrückter rechter Maustaste in das Code-Fenster zieht. Beispielsweise erscheint für die Fitness-Tabelle folgende Angabe:

```
LIBNAME ECLIB000 "C:\Programme\SAS\Enterprise Guide 3.0\Sample\Data";  
/*Datenquelle: ECLIB000.fitness */
```

2.7 Excel und andere externe Dateien

Auf externe Dateien, wie Tabellen aus Excel, Access oder Textdateien kann direkt über das Menü *Datei > Öffnen* oder mittels *Datei > Daten importieren* unter Ausnutzung des Importassistenten zugegriffen werden. Im ersten Fall wird entweder lediglich lesend (ohne Schreibmöglichkeit) auf die Daten zugegriffen oder es wird, wie bei der zweiten Möglichkeit auf jeden Fall, eine Konvertierung der externen Tabelle in eine SAS-Tabelle vorgenommen.

3 Vorsicht – Stolperfallen

In dem Beitrag zu den diesjährigen Tipps und Tricks finden Sie eine ausführliche Beschreibung der möglichen Stolperfallen, die beim Arbeiten mit der SAS Enterprise Guide Software auftreten können, weshalb auf eine Wiederholung an dieser Stelle verzichtet werden soll.

4 Hinweise zur Installation

Die SAS Enterprise Guide Software ist eine reine Windows-Anwendung und kann daher auch nur unter Windows installiert werden. Als Server dient dabei entweder

eine lokale SAS für Windows-Installation, wie bei allen hier gezeigten Beispielen, oder eine SAS Version auf einem anderen Server, Unix, MVS ...

Zur Installation der aktuellen SAS Enterprise Guide Software Version 3.0 werden zwei CDs benötigt:

1. System Requirements Wizard for SAS Enterprise Guide 3.01
2. SAS Enterprise Guide 3.01

Mit der ersten CD werden die Systemanforderungen geprüft und bei Bedarf angepasst. Die zweite CD enthält die eigentliche Software. Im Laufe des Installationsprozesses wird eine mehrstellige Site-Nummer verlangt, die sich in der Regel auf der CD befindet. (Diese ist nicht mit dem Lizenzcode, dem Setinit, zu verwechseln.)

Der Standardinstallationsordner lautet: C:\Programme\SAS und nicht C:\Programme\SAS Institute\, wo SAS 9 per Default installiert wurde.

Angekündigt für den Monat März 2005 wurde bereits die Version 3.02 der SAS Enterprise Guide Software, die wieder eine deutschsprachige Online-Hilfe enthalten soll. (Die Version 3.01 verfügt nur über eine englischsprachige, die Vorgängerversionen 2.0 und 2.1 hatten eine deutschsprachige Hilfe integriert.)

Auf dem SAS Server befindet sich unter dem Link

http://ftp.sas.com/techsup/download/hotfix/ent_guide30.html

das Hotfix 30EG02 für die SAS/EG Version 3.0.

5 Literatur

Zur SAS Enterprise Guide Software gibt es bislang noch nicht viele Bücher bzw. Lehrbücher.

An erster Stelle muss natürlich die Literatur von SAS Institute selbst erwähnt werden:

Der Einstieg in die Enterprise Guide Software. Zweite Ausgabe.
SAS Institute, Cary NC, 2002. 15 Euro.

Auf 82 Seiten erfahren Sie in deutscher Sprache, wie Sie mit Projekten arbeiten, Berichte erstellen, Daten filtern und verknüpfen und einfache Summenstatistiken erstellen. Das ist für den Preis ein ordentliches Buch, aber für denjenigen, der konkrete Aufgaben zu bewältigen hat, zu wenig.

C. Ortseifen

Ein weiteres Buch beinhaltet einen kleinen Abschnitt (37 von 852 Seiten) über den SAS Enterprise Guide:

Ch. Schendera: Datenmanagement und Datenanalyse mit dem SAS System.
Oldenbourg Verlag, München, 2004. 59,80 Euro.

Leider werden nur die Oberfläche und die Menübäume erläutert, ohne dem Anwender konkrete Hilfen mitzugeben.

Die dritte Literaturempfehlung basiert auf den Kursunterlagen der Autorin, die mit Hilfe von zwei wissenschaftlichen Hilfskräften einen Beitrag zum SAS-Anwenderhandbuch im Netz erstellt hat:

C. Ortseifen, Ch. Schredle, Ch. Kögler: Einführung in die SAS Enterprise Guide Software. Version 1.0.
<http://www.urz.uni-heidelberg.de/statistik/sas-ah/>

In dieser Einführung finden Sie eine ausführliche Beschreibung all der Themen, die hier nur kurz angerissen wurden.