

# SENAT DER BUNDESFORSCHUNGSANSTALTEN DES BUNDESMINISTERIUMS FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND VERBRAUCHERSCHUTZ



## Biometrische Aus- und Fortbildung mit SAS im Geschäftsbereich des Bundesministeriums im Jahr 2006

Eckard Moll

### Ist 2006 etwas Besonders im Vergleich zu den anderen Jahren?

Bisher wurden jährlich etwa 4 bis 6 Fortbildungskurse im Auftrag des Senats der Bundesforschungsanstalten durchgeführt. Das Jahr 2006 ist mit 10 Kursen das bisher anspruchsvollste Jahr der biometrischen Fortbildung. Hinzu kommen ein biometrisches Kolloquium, eine in ihrem Umfang über Bisheriges weit hinausgehende Informationsveranstaltung gemeinsam mit SAS-Institute zu SAS Life Science und 2 Treffen der Biometriebeauftragten der Bundesforschungsanstalten und der WGL-Institute.

### Veranstaltungen 2006:

Nr.	Senatskurs / -veranstaltung	Dozent
148	Einführung in SAS	Herr Minkenberg
149	Einführung in die Datenanalyse mit dem SAS/Enterprise Guide	Frau Dr. Ortseifen
150	Biometrisches Kolloquium (15. Treffen der Biometriebeauftragten)	Frau Dr. Ring
151	Einführung in ArcGIS (Grundlagen)	Herr Torkler
152	Einführung in die Macroprogrammierung von SAS	Herr Dr. Bregenzer
153	Auswertung linearer Modelle mit SAS	Herr Prof. Dr. Piepho
154	Einführung in die Biometrie - III	Herr Dr. Schumacher
155	Clusteranalyse u. Diskriminanzanalyse mit SAS	Frau Dr. Zöllner
156	Aspekte der Versuchsplanung mit Unterstützung von CADEMO und SAS/ANALYST	Frau Dr. Kroschewski
157	Hauptkomponenten- und Faktorenanalyse mit SAS	Herr PD Dr. Marohn u. Herr Dr. René Michel
158	Informationsveranstaltung zu SAS/GENETICS (16. Treffen der Biometriebeauftragten)	Frau Dr. Nedbal (SAS Europa) u. Herr Dr. Strüby (SAS Deutschland)
159	Einführung in die Geostatistik mit SAS	Frau Dr. Marx
160	Einführung in die Biometrie - IV	Herr Prof. Dr. Rasch u. Herr Prof. Dr. Verdooren

### zu den Inhalten der Fortbildungskurse 2006:

#### Einführung in SAS

Ziel ist es, einen Einblick in das umfangreiche Softwarepaket SAS zu erhalten. So lernt der Teilnehmer anhand der Dateneingabe und -aufbereitung am Beispiel statistischer Verfahren der beschreibenden Statistik und einfacher Grafiken den Umgang mit SAS kennen.

Da SAS nicht nur den Zugang mit Hilfe eines selbstgeschriebenen Programms ermöglicht, werden auch die anderen Wege vermittelt.

Die Kenntnis einfacher Verfahren der beschreibenden Statistik ist von Vorteil.

#### Macroprogrammierung in SAS

Der Begriff Macro wird in SAS zweimal verwendet. Das ist einmal die Macrovariable, die als globale Variable genutzt wird und formale Ähnlichkeit mit Data-Step-Variablen hat, inhaltlich aber etwas anderes ist. Ihre Definitionsmöglichkeiten und Verwendung werden gezeigt.

Zum anderen lassen sich mit Anweisungen der Macroprogrammierung Programmschleifen (%DO-%END) oder Programmverzweigungen (%IF-%THEN-%ELSE) in der von Data-Steps bekannten Weise nutzen und mehrfach abarbeitbare (Unter-) Programme mit und ohne Parameterübergabe schreiben.

#### Auswertung linearer Modelle mit SAS

Lineare Modelle sind die Basis für die meisten statistischen Verfahren. Als Beispiele werden lineare Regression, Polynomregression, ein- und mehrfaktorielle Varianzanalyse und die Kovarianzanalyse genannt. Das lineare Modell wird weiter unterschieden in lineares gemischtes Modell (LMM), generalisiertes lineares Modell (GLM) und generalisiertes lineares gemischtes Modell (GLMM).

In der Agrarforschung haben wir es mit allen Formen des linearen Modells zu tun. Deshalb wird der Überblick über die Modelle und deren statistische Auswertungsverfahren gegeben. Die SAS-Umsetzung wird für verschiedenen Prozeduren und deren Optionen vorgestellt.

#### Aspekte der Versuchsplanung mit Unterstützung von CADEMO und SAS/ANALYST

Ausgehend von der Rolle der Versuchsplanung in der experimentellen Forschung werden für Aufgabenstellungen des Schätzens von Parametern und des Testens von Hypothesen über zwei und mehr Parameter die Grundlagen der Versuchsplanung, d.h. die Berechnung von Kenngrößen der Genauigkeit und der Sicherheit vorgestellt. Mit dem Schwerpunkt der Versuchsplanung werden die modular aufgebaute Software CADEMO und der sich mit der Versuchsplanung sowie der Berechnung des Risikos 2. Art beschäftigende Teil des SAS/ANALYST behandelt.

#### Einführung in die Datenanalyse mit dem SAS/Enterprise Guide

Anhand eines konkreten Falles wird gezeigt, wie Auswertungen angefangen vom Einlesen der Daten bis zur grafischen Darstellung in Word mit Hilfe der SAS/Enterprise Guide Software völlig ohne Programmierkenntnisse durchgeführt werden können.

Die Versuchsdaten werden hierzu in ein neues Enterprise Guide Projekt eingebunden und einfache bzw. bei Bedarf beispielhaft auch komplexere Analyseverfahren durchgeführt. Die wichtigsten grafischen Elemente wie Balken- und Streudiagramm sowie Box-Plot und der Export der Ergebnisse nach MS-Word bzw. ins Internet werden vorgestellt.

Der Einsatz der Enterprise Guide Software setzt keine Kenntnisse von der Programmierung mit SAS voraus. Die im Verlauf der Analysen erzeugten SAS-Programme werden aber im Projekt abgelegt und können so von erfahrenen bzw. fortgeschrittenen SAS-Anwendern eingesehen werden. Die Vorgänge sind damit dokumentiert und können bei Bedarf beliebig oft abgearbeitet werden.

#### Biometrie III • Einführung in die Biometrie - Teil 3

Die Biometrie ist eine breit gefächerte Wissenschaft, deren Vermittlung von Basiswissen bereits den Rahmen eines Kurses sprengen würde. Deshalb werden auf der Grundlage der Reihe "Einführung in die Biometrie" Grundkenntnisse vermittelt.

Behandelt wird das Heft 3 SCHUMACHER: "Vergleich von mehr als zwei Parametern".

Auftakt bildet die einfaktorielle Varianzanalyse. Zum Vergleich der fixen Effekte werden verschiedene multiple Testverfahren vorgestellt. Bei zufälligen Effekten werden die Varianzkomponenten geschätzt und getestet.

Die mehrfaktorielle Varianzanalyse wird zunächst für balancierte Daten behandelt, wobei das Hauptaugenmerk auf das Testen fixer Effekte unter Berücksichtigung der Analyse der Wechselwirkungen gelegt wird. Kreuz- und hierarchische Klassifikationen im fixen, zufälligen und gemischten Modell werden betrachtet.

Das Problem unbalancierter Daten wird behandelt. Dabei wird zwischen schwach und stark unbalancierten Daten unterschieden.

Den Abschluss bilden parameterfreie Testverfahren zum Testen von mehr als zwei Parametern, wobei besonders der Permutationstest behandelt wird.

#### Biometrie IV • Einführung in die Biometrie - Teil 4

Die Biometrie ist eine breit gefächerte Wissenschaft, deren Vermittlung von Basiswissen bereits den Rahmen eines Kurses sprengen würde. Deshalb werden auf der Grundlage der Reihe "Einführung in die Biometrie" Grundkenntnisse vermittelt.

Behandelt wird das Heft 4 RASCH & VERDOOREN: "Grundlagen der Korrelationsanalyse und der Regressionsanalyse".

Betrachtet werden Zusammenhänge und Abhängigkeiten von Merkmalen und darauf aufbauend werden die Grundlagen der Korrelations- und der Regressionsanalyse behandelt. Vorgestellt werden die einfache lineare, quadratische, kubische und eigentlich nichtlineare Regression mit Versuchsplanung, Parameterschätzung, Vorhersage und Tests. Auf die Modellwahl bei den verschiedenen Regressionen wird aufmerksam gemacht.

#### Einführung in ArcGIS

Dieser Kurs bietet eine Einführung in die Produktfamilie ArcGIS, also im Einzelnen ArcView 8, ArcEditor 8 und ArcInfo 8, die alle über dieselben Oberflächen bedient werden (ArcMap, ArcCatalog, ArcToolbox).

Der Kurs richtet sich vorwiegend an ArcViewGIS 3.1-Nutzer, aber auch an GIS-Neulinge. Daher werden im Kurs sowohl grundlegende Konzepte und Funktionen vermittelt als auch Fragen der Migration von ArcView 3-Projekten zu ArcGIS thematisiert.

Kernfragen werden sein:

"Wie mache ich das, was ich in ArcView3 mache, in ArcView 8 ?",  
"Was ist mit meinen 'alten' ArcView3-Projekten ?",  
"Was ist an ArcGIS besser ?"

Die Kursteilnehmer werden einerseits Daten erzeugen, bearbeiten, analysieren und darstellen, andererseits ArcView3-Projekte importieren. An Beispielen werden die neuen Funktionen geübt.

#### Clusteranalysen und Diskriminanzanalyse mit SAS

In beiden Verfahrengruppen spielen Abstandsmaße eine entscheidende Rolle. Bei einem Clusterverfahren entscheiden das zugrundegelegte Abstandsmaß als Ausdruck für die Ähnlichkeit der zu gruppierenden Objekte und der Gruppierungsalgorithmus über die Gruppierung der Objekte, d.h. über die Zugehörigkeit zu einem Cluster mit (möglichst) ähnlichen Objekten.

Aus der Vielzahl der Clusterverfahren werden für die gebräuchlichsten Abstandsmaße und Algorithmen die Verfahren vorgestellt.

Die Diskriminanzfunktion (oder Trennfunktion) ist entscheidend für die Zuordnung eines Objekts zu den unterschiedlichen Klassen. Sie zeigt auch die Bedeutung der in die Analyse einfließenden Merkmale für die Unterscheidung zwischen den Klassen.

Unterschiede und Zusammenhänge zwischen den Verfahren der Clusteranalyse und der Diskriminanzanalyse werden dargelegt.

#### Faktoren- und Hauptkomponentenanalyse mit SAS

Faktoren- und Hauptkomponentenanalyse sind mehrdimensionale Analyseverfahren. Beide Verfahrengruppen werden auch zu den Methoden der Strukturanalyse gezählt. Werden an den Versuchsobjekten viele Merkmale beobachtet, so berechnet man mit Hilfe der Hauptkomponentenanalyse aus der Merkmalsmenge der Objekte orthogonale Hauptachsen. Ziel der Faktorenanalyse ist es, die Vielzahl der Merkmale durch eine geringe Anzahl von Faktoren zu erklären. Wir erhalten also ein k-Faktoren-Modell.

#### Einführung in die Geostatistik mit SAS

Verfahren der Geostatistik haben ihren Ursprung, den Bergbau und in der Montan-geologie, längst verlassen. Betrachtet werden Zufallsfunktionen zur Beschreibung orts-abhängiger Variablen. Die Verfahren der Geostatistik beschäftigen sich einerseits mit der räumlichen (statistisch-gesetzmaßigen) Variation und andererseits mit Analyseverfahren unter Berücksichtigung georeferenzierter Daten. Vorgestellt werden unter anderem Variogramme und Krigeschätzungen.

Internetadresse: Biometrie und SAS-Anwendung im BMELV : <http://biometrie.bfn-inst2.de/>