

1. Konferenz für SAS-Benutzer in Forschung und Entwicklung (KSFE)
Humboldt-Universität, Berlin, 20.-21. Februar 1997

Die Nutzung von SAS im Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere

Paul Eberhard RUDOLPH, Günter HERRENDÖRFER

Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere Dummerstorf-Rostock
Wilhelm-Stahl-Allee 2, D-18196 Dummerstorf

1. Einleitung

In den 7 Forschungsbereichen des Forschungsinstituts für die Biologie landwirtschaftlicher Nutztiere wird SAS einerseits zur Datenanalyse mit Methoden der mathematischen Statistik (Varianzanalyse, lineare und nichtlineare Regressionsanalyse, verallgemeinerte lineare Modelle, statistische Tests usw.) eingesetzt und andererseits im Forschungsbereich Biometrie auch zur Modellierung biologischer Prozesse und zur Untersuchung statistischer Verfahren verwendet. Beispiele für letzteres sind Simulationsuntersuchungen zur Genauigkeit nur asymptotisch geltender Lösungen von Versuchsplanungsproblemen im Endlichen oder zur Beurteilung der Vorhersagegenauigkeit in gemischten linearen Modellen.

In der stark überarbeiteten Neuauflage der zweibändigen Verfahrensbibliothek: Versuchsplanung und -auswertung (Band I: R. Oldenbourg Verlag München 1996, Band II: Oldenbourgverlag München 1997), an der der Forschungsbereich Biometrie wesentlich beteiligt ist, sind für viele Verfahren die entsprechenden SAS-Programme nutzerfreundlich eingearbeitet worden.

Nachdem im Abschnitt 2 zunächst die Forschungsbereiche des FBN mit ihren Forschungsschwerpunkten vorgestellt werden, wird im dritten Abschnitt ein Versuchsplanungsproblem für ein nichtlineares Regressionsproblem zur Auswertung von Langzeitselektionsexperimenten und dessen asymptotische Lösung vorgestellt. Die Anwendbarkeit dieser asymptotischen Lösung wird durch eine mit SAS realisierte Simulationsstudie gezeigt.

Im Abschnitt 4 wird abschließend das Buchprojekt 'Verfahrensbibliothek Versuchsplanung und -auswertung' des Forschungsbereiches Biometrie mit der Einarbeitung von SAS in bestimmte Verfahren an einem Beispiel demonstriert.

2. Die Forschungsbereiche des FBN und ihre Forschungsschwerpunkte

Am 29. Januar 1993 wurde das Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere (FBN) mit Sitz in Dummerstorf und Rostock als gemeinnützige Stiftung des öffentlichen Rechts des Landes Mecklenburg-Vorpommern gegründet. Damit steht diese Institution auf agrarhistorischem Boden, denn schon 1939 wurde in Dummerstorf das renommierte Kaiser-Wilhelm-Institut für Tierzuchtforschung eingeweiht. 1952 setzte die Akademie der Landwirtschaftswissenschaften die Tradition mit der Etablierung des Instituts für Tierzuchtforschung fort; 1953 folgte das Oskar-Kellner-Institut für Tierernährung in Rostock. 1970 wurden diese Einrichtungen in dem Forschungszentrum für Tierproduktion Dummerstorf-Rostock vereint. Obwohl das FBN einen Großteil der historischen Liegenschaften, des vorhandenen Forschungspotentials und der technischen Einrichtungen nutzt, ist es kein Rechtsnachfolger dieser Institutionen. Es wird über die gemeinnützige

Stiftung zu gleichen Teilen durch den Bund und das Land Mecklenburg-Vorpommern getragen.

Satzungsgemäß ist der Zweck der Stiftung die Grundlagenforschung auf dem Gebiet der Biologie landwirtschaftlicher Nutztiere. Den hohen wissenschaftlichen Ansprüchen an ökologisch und ökonomisch nachhaltige, nutztierartgerechte Zucht und Haltung kann nur mit interdisziplinärer Forschung entsprochen werden. Das FBN hat es sich zum Ziel gesetzt, Erkenntnisse zu erarbeiten, die Innovationsschübe für die Landwirtschaft ermöglichen. Deshalb erforschen Wissenschaftler aus den Fachdisziplinen Agrarwissenschaften, Biologie, Tiermedizin, Chemie, Physik und Mathematik gemeinsam die Grundlagen von Genetik, Fortpflanzung, Ernährung, Verhalten, Wachstum und Entwicklung landwirtschaftlicher Nutztiere. Im Institut bestehen neben seinen Forschungseinrichtungen auch Stallkapazitäten zur Unterbringung von bis zu 450 Großvieheinheiten und eine Schlachtstätte.

Das Institut selbst ist in 7 Forschungsbereiche gegliedert und kooperiert mit 43 nationalen und 45 internationalen Forschungseinrichtungen in 17 Ländern, so daß sich Projekte in der Regel nicht ausschließlich einem Forschungsbereich zuordnen lassen, sondern vielmehr in kollegialer Zusammenarbeit innerhalb des FBN und in Kooperation mit externen Wissenschaftspartnern durchgeführt werden.

Im folgenden sind die 7 Forschungsbereiche mit ihren jeweiligen Forschungsschwerpunkten aufgelistet:

Forschungsbereich : **Populationsbiologie und Züchtungsforschung**

Forschungsschwerpunkte:

- Grundlagen biologischer Strukturen von Tierpopulationen und deren Wechselbeziehungen, Entwicklung und Charakterisierung von Markern zur markergestützten Selektion, populationsbiologische Auswirkungen biotechnischer Methoden
- Weiterentwicklung von Methoden zur Bewertung und Auswahl von Merkmalen für Vermehrung, Wachstum, Überleben und Anpassung
- Populationsbiologische Grundlagen der Pflege und Erhaltung tiergenetischer Ressourcen (experimentelle Prüfung von Kriterien zur Erhaltung tiergenetischer Vielfalt)

Forschungsbereich : **Molekularbiologie**

Forschungsschwerpunkte:

- Nachweis, Isolierung und Charakterisierung merkmalsgekoppelter polymorpher genetischer Marker bei Rind, Schwein und Maus
- Erstellung von chromosomen- bzw. chromosomenfragmentspezifischen Genkarten des Rindes unter besonderer Beachtung von Kopplungsgruppen, die tierzüchterisch relevante Gene einschließen sowie Darstellung und Analyse leistungsgekoppelter chromosomaler Veränderungen beim Nutztier
- Analyse der Struktur von Kandidatengenen für tierzüchterisch bedeutsame Leistungen (z.B. Laktoferringen - Mastitisanfälligkeit; Kaseingen - Milchleistung; Hitzeschockproteingene - Belastbarkeit)
- Untersuchungen zur Regulation der Expression von ausgewählten Genen, die mit tierischen Leistungen assoziiert sind (Fruchtbarkeit - Aromatasegen; Myogenese- Gene der Myo D-Familie)

Forschungsbereich : **Fortpflanzungsbiologie**

Forschungsschwerpunkte:

- Untersuchungen von Regulationsmechanismen der ovariellen Follikel- und Eizellentwicklung
- Nachweis von Zellrezeptoren und Charakterisierung von Rezeptor-Ligand-Interaktionen in ovariellen Zellen des weiblichen Reproduktionstraktes
- Aufklärung von Kern-Zytoplasma-Wechselwirkungen bei der Reprogrammierung der embryonalen Entwicklung
- Erforschung von Regulationsmechanismen der Interaktion zwischen Embryo und maternalem Milieu im präimplantiven Zeitraum

Forschungsbereich : **Physiologische Grundlagen der Tierhaltung**

Forschungsschwerpunkte:

- Neuronale Aktivität in limbischen Hirnarealen von Labor- und Nutztieren in Bezug zu dem Verhalten in unterschiedlichen Haltungswelten
- Ethophysiologie von Anpassungsreaktionen bei Haustieren
- Haltungsabhängige Entwicklung endokrinologischer, immunologischer und neurophysiologischer Parameter

Forschungsbereich : **Muskelbiologie und Wachstum**

Forschungsschwerpunkte:

- Zellphysiologische Grundlagen der Muskelfunktion und -beschaffenheit in Abhängigkeit von Tierarten, Entwicklungsstadien und Funktionskategorien
- Erforschung von Prozessen des Wachstums sowie des Stoffansatzes unter der Wirkung leistungsbeeinflussender Faktoren, Quantifizierung der Einflüsse und Beurteilung aus der Sicht der Humanernährung
- Auswirkungen zellphysiologischer Funktionen auf biologische Phänomene des Wachstums

Forschungsbereich : **Ernährungsphysiologie "Oskar Kellner"**

Forschungsschwerpunkte:

- Quantitative und regulatorische Aspekte des Protein- und Energieumsatzes landwirtschaftlicher Nutztiere unter Einbeziehung von Labortieren als Modelle
- Physiologie des Verdauungstraktes: Endogene Sekretion sowie Wirkung von Mikroben im Verdauungstrakt

Forschungsbereich : **Biometrie**

Forschungsschwerpunkte:

- Schätzung und Tests in gemischten linearen Modellen - Versuchsplanung und Robustheit
- Modellwahl für biologische Prozesse
- Methoden zur biologischen Wirkstoffprüfung

Mit Ausnahme des Bereiches Biometrie werden in allen Forschungsbereichen Experimente durchgeführt, deren Ergebnisse stochastischem Einfluß unterliegen und damit eine statistische Bearbeitung implizieren. Seit Gründung des Instituts werden deshalb insbesondere der statistische und der graphische Teil des Programmpaketes SAS als strategisches Werkzeug für die Datenanalyse mit anspruchsvollen statistischen Verfahren verwendet. Dazu gehören neben einer elementaren Datenaufbereitung die Verfahren der Varianzanalyse, lineare und nichtlineare Regressionsanalyse, Schätzungen und Tests in verallgemeinerten linearen Modellen u.a.

Darüber hinaus wird im Forschungsbereich Biometrie das SAS-System auch zur Untersuchung der Robustheit statistischer Verfahren und bei der Modellierung biologischer Prozesse eingesetzt.

In den folgenden beiden Abschnitten wird die intensive Nutzung von SAS insbesondere im Forschungsbereich Biometrie aufgezeigt.

3. Versuchsplanung für ein nichtlineares Regressionsproblem

Zur Beschreibung der Selektionsergebnisse in Abhängigkeit von der Generationsanzahl bei Langzeitselektionsexperimenten wurden wiederholt nichtlineare Regressionsmodelle

$$\underline{y}_i = f(x_i, \vartheta) + \underline{e}_i, \quad i = 0, 1, \dots, m \quad (3.1)$$

verwendet, wobei \underline{y}_i die zufälligen Beobachtungen für die Regressorwerte x_i (Generationen), f eine Regressionsfunktion, die nichtlinear ist für wenigstens eine Komponente des unbekanntem Parametervektors ϑ , und \underline{e}_i die zufälligen Fehler bezeichnen.

Für Langzeitselektionsexperimente mit Labormäusen in Dummerstorf hat sich die folgende Form einer 3-parametrischen Exponentialfunktion

$$f(x, \vartheta) = a - (a - c)e^{-\frac{bx}{a - c}} \quad \text{mit} \quad \vartheta = (a, b, c)' \quad (3.2)$$

als besonders günstig sowohl im Hinblick auf die Anpassung der Daten als auch die biologische Interpretation der Parameter erwiesen.

Für das Modell (3.1) mit der Regressionsfunktion (3.2) war u.a. folgendes Versuchsplanungsproblem zu lösen:

Es ist die notwendige (jeweils gleiche) Anzahl n von Beobachtungen je Generation derart zu bestimmen, daß die halbe erwartete Länge eines $(1-a)$ -Konfidenzintervalls für den Parameter a des Parametervektors $\vartheta = (a, b, c)'$ kleiner oder gleich einer vorgegebenen Konstanten d ist.

Eine auf der asymptotischen Kovarianzmatrix der MKQ-Schätzungen basierende Lösung dieses Versuchsplanungsproblems ist bei RUDOLPH, HERRENDÖRFER (1995) zu finden. Um eine praktische Anwendung dieser auch nur asymptotisch geltenden Lösung überhaupt sinnvoll zu machen, ist die Genauigkeit dieser Lösung für endliche Stichprobenumfänge zu überprüfen. Das ist durch entsprechende Simulationsuntersuchungen möglich. Unter Verwendung der SAS-Prozedur NLIN und der SAS-subroutine RANNOR wurde eine solche Simulationsstudie durchgeführt. Die Details dazu sind bei RUDOLPH (1995) zu finden.

4. SAS-Programme in der ‘Verfahrensbibliothek Versuchsplanung und -auswertung’

Die Nutzung der Ergebnisse der Mathematischen Statistik in den biologischen Wissenschaften hängt erheblich von einer zusammengefaßten und einfach verständlichen Darstellung ab, die auch die Nutzung der entsprechenden Software einschließt. Ein Arbeitsgebiet des Forschungsbereiches Biometrie ist es, dem Anwender sowohl die Ergebnisse der Mathematischen Statistik als auch die entsprechende Software in einer solchen Form anzubieten., daß ihre Nutzung möglichst einfach und selbständig erfolgen kann. Zur Lösung dieser Aufgabe werden Kurse durchgeführt und entsprechende Hefte in der FBN-Reihe erarbeitet. Ein größeres Projekt, an dem die Mitarbeiter des Forschungsbereiches Biometrie wesentlich beteiligt sind, ist die Erarbeitung der zweibändigen ‘Verfahrensbibliothek Versuchsplanung und -auswertung’. Der erste Band ist bereits 1996 im R. Oldenbourg Verlag München erschienen.



Der zweite Band wird voraussichtlich 1997 erscheinen.

Wie in der ‘Verfahrensbibliothek’ dem Leser SAS-Programme (auch auf Diskette) anwendungsfreundlich zur Verfügung gestellt werden, soll im folgenden an einem Beispiel-Verfahren aus Band I demonstriert werden.

1/61/2110 Varianztabelle für zweifache Kreuzklassifikation mit gleicher Klassenbesetzung ohne Wechselwirkungen

Problemstellung

Gegeben sei eine Versuchsanlage mit den Faktoren A und B einer zweifachen Kreuzklassifikation. Der Faktor A hat die Stufen A_1, A_2, \dots, A_a und der Faktor B die Stufen B_1, B_2, \dots, B_b . Zu jeder Faktorstufenkombination liegen n Beobachtungswerte vor.

Es sind die Formeln für die \underline{SQ} , FG und \underline{MQ} in Form einer Varianztabelle zusammenzustellen. Für Modell I, Modell II, das gemischte Modell und das Modell mit endlichen Stufengesamtheiten sind die Erwartungswerte der \underline{MQ} anzugeben. Für Modell I lautet die Modellgleichung:

$$\underline{y}_{ijk} = \mu + a_i + b_j + \epsilon_{ijk} \quad i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n.$$

Weitere Modellvoraussetzungen und Hinweise zur Symbolik (z.B. $\tilde{\sigma}$) findet man in 1/61/0000.

Lösungsweg

Siehe Übersichten:

Übersicht 1:

Variationsursache	\underline{SQ}	FG	\underline{MQ}
Zwischen den Stufen des Faktors A	$\underline{SQ}_A = \frac{1}{bn} \sum_i \underline{Y}_{i..}^2 - \frac{1}{N} \underline{Y}_{...}^2$	$a - 1$	$\underline{MQ}_A = \frac{\underline{SQ}_A}{a - 1}$
Zwischen den Stufen des Faktors B	$\underline{SQ}_B = \frac{1}{an} \sum_j \underline{Y}_{.j.}^2 - \frac{1}{N} \underline{Y}_{...}^2$	$b - 1$	$\underline{MQ}_B = \frac{\underline{SQ}_B}{b - 1}$
Rest	$\underline{SQ}_R = \sum_{i,j,k} \underline{y}_{ijk}^2 - \frac{1}{bn} \sum_i \underline{Y}_{i..}^2 - \frac{1}{an} \sum_j \underline{Y}_{.j.}^2 + \frac{1}{N} \sum_{i,j} \underline{Y}_{ij.}^2$	$N - a - b + 1$	$\underline{MQ}_R = \frac{\underline{SQ}_R}{N - a - b + 1}$
Gesamt	$\underline{SQ}_G = \sum_{i,j,k} \underline{y}_{ijk}^2 - \frac{1}{N} \underline{Y}_{...}^2$	$N - 1$	

Übersicht 2:

Variationsursache	$E(\underline{MQ})$		
	Modell I	Modell II und Modell mit endlichen Stufengesamtheiten	gemischtes Modell A fest, B zufällig
Zwischen den Stufen des Faktors A	$\sigma^2 + \frac{bn}{a-1} \sum_i a_i^2$	$\sigma^2 + bn\tilde{\sigma}_a^2$	$\sigma^2 + \frac{bn}{a-1} \sum_i a_i^2$
Zwischen den Stufen des Faktors B	$\sigma^2 + \frac{an}{b-1} \sum_j b_j^2$	$\sigma^2 + an\tilde{\sigma}_b^2$	$\sigma^2 + an\sigma_b^2$
Rest	σ^2	σ^2	σ^2

Literatur

Rasch, D. (1995)

Beispiel

Aus folgenden Versuchsergebnissen ist eine Varianztabelle aufzustellen:

B-Stufen	Beobachtung	A-Stufen				Y _j
		A ₁	A ₂	A ₃	A ₄	
B ₁	1	24	22	29	23	
	2	21	24	25	21	
	3	22	23	25	22	
	4	20	26	22	27	
	Y _{i1}	87	95	101	93	
B ₂	1	22	20	23	25	
	2	23	24	24	23	
	3	21	28	25	22	
	4	25	23	22	24	
	Y _{i2}	91	95	94	94	
B ₃	1	27	23	26	26	
	2	23	24	25	22	
	3	25	20	21	24	
	4	24	21	21	21	
	Y _{i3}	99	88	93	93	
Y _{i.}		277	278	288	280	Y _{...} = 1123

a = 4, b = 3, n = 4, N = 48

Jetzt werden folgende Werte berechnet:

$$\frac{1}{N} Y_{...}^2 = \frac{(1123)^2}{48} = 26273,52$$

$$\frac{1}{bn} \sum_i Y_{i.}^2 = \frac{1}{12} \left((277)^2 + \dots + (280)^2 \right) = 26279,75$$

$$\frac{1}{an} \sum_j Y_{.j}^2 = \frac{1}{16} \left((376)^2 + \dots + (373)^2 \right) = 26273,81$$

$$\sum y_{ijk}^2 = (24)^2 + \dots + (21)^2 = 26485,00$$

Mit diesen Werten läßt sich die Varianztabelle angeben:

Variationsursache	SQ	FG	MQ
Zwischen den A-Stufen	6,23	3	2,077
Zwischen den B-Stufen	0,29	2	0,145
Rest	204,96	42	4,880
Gesamt	211,48	47	

Zu dem gleichen Ergebnis gelangt man mit der SAS-Prozedur **GLM**. Dazu mögen die Daten in einer SAS-Datendatei **sas1** vorliegen, welche unter der Variablen **a** die Stufen des Faktors A, unter der Variablen **b** die des Faktors B und unter der Variablen **y** die jeweiligen Meßwerte enthält.

SAS-Programm

```
proc glm data=sas1;
  class  a b;
  model  y=a b / ss3;
run;
```

SAS-Ausgabe (gekürzt)

```
General Linear Models Procedure
Class Level Information
```

```
Class      Levels      Values
A           4          a1 a2 a3 a4
B           3          b1 b2 b3
```

```
Number of observations in data set = 48
```

```
General Linear Models Procedure
Dependent Variable: Y
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	6.52083333	1.30416667	0.27	0.9285
Error	42	204.95833333	4.87996032		
Corrected Total	47	211.47916667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	6.22916667	2.07638889	0.43	0.7357
B	2	0.29166667	0.14583333	0.03	0.9706

Literatur:

- RUDOLPH, P.E.: The Accuracy of an Asymptotic Solution of an Experimental Design Problem for a Nonlinear Regression Model Used in the Analysis of Long-term Selection Experiments. *Arch. Tierz.* **38** (1995) 3, 347-352
- RUDOLPH, P.E.; HERRENDÖRFER, G.: Optimal Experimental Design and Accuracy of Parameter Estimation for Nonlinear Regression Models Used in Long-term Selection. *Biom. J.* **37** (1995) 2, 183-190
- RASCH, D.; HERRENDÖRFER, G.; BOCK, J.; VICTOR, N.; GUIARD, V.: *Verfahrensbibliothek Versuchsplanung und -auswertung, Band I, Buch mit Diskette.* R. Oldenbourg Verlag München Wien 1996

Anschrift der Autoren:

Dr. PAUL EBERHARD RUDOLPH
 Prof. Dr. GÜNTER HERRENDÖRFER
 Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere
 Forschungsbereich Biometrie
 Wilhelm-Stahl-Allee 2
 D-18196 Dummerstorf
 e-mail: rudolph@fbn.uni-rostock.de
 www: http://www.fbn.uni-rostock.de