

Eine SAS-Anwendung für nichtlineare Regressionsprobleme

H.-P. Altenburg, Th. Floren und M. Prümmer

Universität Heidelberg
Fakultät für Klinische Medizin Mannheim
Med. Statistik, Biomathematik und Informationsverarbeitung
Theodor-Kutzer Ufer, D-68135 Mannheim

Zusammenfassung: Viele Fragestellungen der Pharmakokinetik und Pharmakodynamik erfordern die Anwendung von verallgemeinerten Kleinste-Quadrate-Schätzverfahren, weil neben den Regressionsparametern der funktionalen Beziehung noch Parameter der modellierten Varianz geschätzt werden müssen. Es wird eine SAS/AF Anwendung vorgestellt, die Parameterschätzung von verschiedenen Erwartungswert- als auch Varianzmodellen erlaubt.

1. Einleitung

Die Entwicklung neuer Therapieformen verläuft über verschiedene Stufen der Pharmakokinetik und Pharmakodynamik, wie z.B. Dosis → Konzentration-Beziehung, Konzentration → Effekt-Beziehung oder eine Effekt → Wirksamkeits-Beziehung. Die Zusammenhänge in all diesen Phasen der Entwicklung lassen sich durch entsprechende Prozesse modellieren, die nichtlineare Erwartungswertfunktionen und Varianzen bzgl. der Parameter aufweisen. Das Schätzen der Varianzparameter ist besonders bei heteroskedastischen Daten von Interesse. Zwar kann man sich für die Schätzung der Parameter oft behelfen, indem man durch eine geeignete Transformation die Beziehung „linear“ macht. Jedoch führt dies nicht in allen Situationen zu einer befriedigenden Lösung. Flexibler und auch besser sind deshalb meist nichtlineare funktionale Beziehungen und deren Parameterschätzungen.

Das Ziel der im folgenden vorgestellten SAS-Anwendung ist, eine einheitliche Oberfläche anzubieten, die es erlaubt, sowohl für verschiedene Erwartungswertmodelle als auch unterschiedliche zugehörige funktionale Beziehungen der Varianz die Parameter zu schätzen. Bereits über viele Jahre angesammelte und entwickelte SAS-Macros sollten in dieses System integriert werden. Neben der Parameterschätzung soll auch die graphische Darstellung der Ergebnisse berücksichtigt werden.

Zwei große Klassen von Modellen wurden realisiert:

- Dosis-Wirkungs-Beziehungen und
- Pharmakokinetische Modelle.

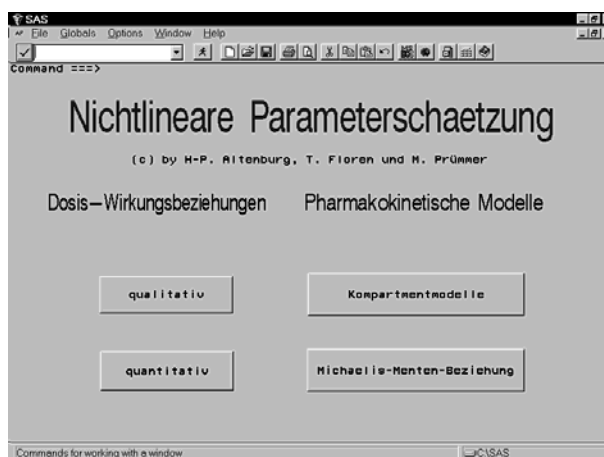


Abb.1: Hauptmenu

Im nächsten Abschnitt werden die realisierten Modelle näher erläutert.

2. Erwartungswert-Modelle

2.1 Dosis-Wirkungs-Beziehungen

Viele Fragestellungen in der biologisch-medizinischen Forschung betreffen die Beziehung zwischen einer Response und einem verabreichten Stimulus. Zwei große Klassen werden unterschieden, je nach Typ der abhängigen Variablen: qualitative und quantitative Dosis-Wirkungs-Beziehungen. Bei den qualitativen Beziehungen oder Alles-Oder-Nichts-Effekten ist die Zielgröße dichotom ausgeprägt, bei den quantitativen Beziehungen dagegen stetig verteilt. Beide Beziehungen haben ein sehr weites Anwendungsspektrum. Die Idee, die zu den quantalen Responsemodellen führte, ist folgende: Jedes Individuum hat unter kontrollierten Bedingungen ein bestimmtes „(Dosis-) Intensitätslevel“ bis zu dem kein Effekt sichtbar ist. Dieses Schwellen- oder Toleranzlevel variiert individuell. Bei einem niedrigen Dosislevel x ist deshalb der Anteil reagierender Individuen niedrig oder gar null ($p(x) \approx 0$), dagegen geht der Anteil reagierender Individuen mit wachsender Dosis ($x \rightarrow \infty$) gegen 1 ($p(x) \approx 1$). Es liegt deshalb nahe ein Modell der Art

$$p(x) = \Pr(\text{Toleranzwert} \leq x) = \begin{cases} G((\log x - \mu)/\sigma) & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

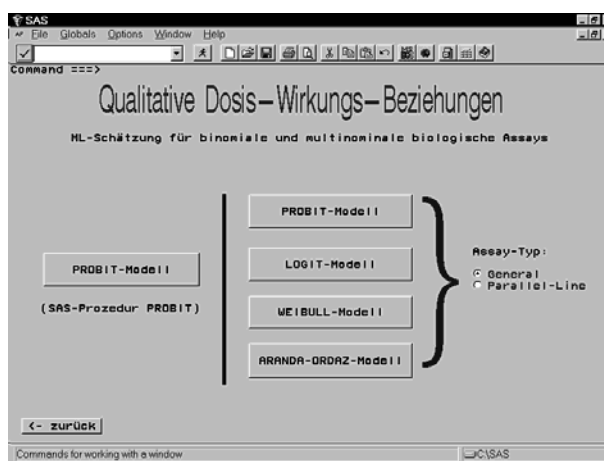


Abb. 2: Modellauswahl bei quantitativen Dosis-Wirkungs-Beziehungen

anzupassen. $G(z)$ ist dabei eine Verteilungsfunktion und μ bzw. σ sind Kurvenparameter, die die Steigung der Kurve und die Mitte beschreiben. Vier mehr oder weniger klassische Modelle stehen für die Verteilungsfunktion $G(\cdot)$ zur Verfügung:

Probit - Modell:
$$G(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-z^2/2) dz$$

Logit - Modell:
$$G(z) = 1/(1 + \exp(-z))$$

Weibull - Modell:
$$G(z) = 1 - \exp(-e^z)$$

Aranda - Ordaz - Modell:
$$p(x) = \begin{cases} 1 - (1 + \kappa \exp(\alpha + \beta x))^{-1/\kappa} & \text{falls } \kappa \exp(\alpha + \beta x) < -1 \\ 1 & \text{sonst} \end{cases}$$

Bei den quantitativen Dosis-Wirkungs-Beziehungen erhält man ähnliche symmetrisch-sigmoide Kurven. Der Hauptunterschied liegt darin, daß die beiden unteren und oberen Asymptoten noch zusätzlich geschätzt werden müssen. Die eigentliche Aufgabe besteht darin, den Effekt eines Stimulus durch ein quantitatives Merkmal zu schätzen, wie z.B. den freien oder gebundenen Anteil radioaktiver Counts eines Antigens oder Antikörpers bei einem RIA oder IRMA.

Während der quantalen Response ein binomiales Modell mit keinen zusätzlich zu schätzenden Varianzparametern zugrunde liegt, läßt sich ein quantitatives Response-Modell der obigen Form (RIA oder IRMA) nur durch einen überstreuten Poisson-Prozeß modellieren, wobei der Erwartungswert etwa die Form

$$\lambda = E[\text{Count} \mid \text{Dose} = x] = \delta_0 + \rho G(\beta - \gamma \log x)$$

und die (Standard-) Varianz etwa die Form

$$\Lambda = \lambda(1 + \kappa^* \lambda)$$

haben. x bezeichnet dabei wieder die verabreichte Dosis, $R = \delta_{\infty} - \delta_0$, den Range zwischen oberer und unterer Asymptote, β sowie γ die Kurvenparameter im engeren Sinne ähnlich wie beim quantalen Fall und $G(\cdot)$ wieder eine Verteilungsfunktion. Die Varianz enthält in diesem Standardmodell noch den zusätzlich zu schätzenden Parameter κ^* . In anderen experimentellen Situationen (wie z.B. ELISAs) sind u.U. auch andere Erwartungswert- bzw. Varianzmodelle sinnvoll. Unterschiedliche experimentelle Situationen erfordern meist auch Modifikationen dieses Modells hinsichtlich der Designwahl (Parallel-Line- oder General-Assay). Für Details hierzu sei auf [1] bzw. [2] verwiesen.

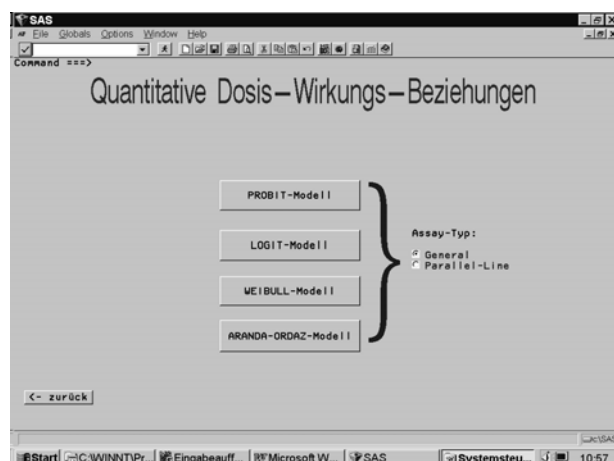


Abb. 3: Modell- und Designwahl

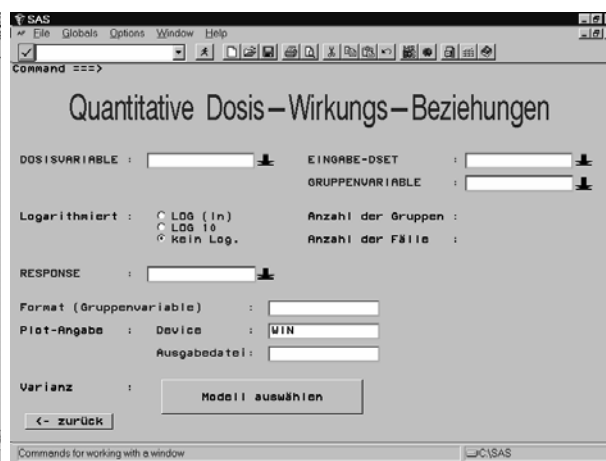


Abb. 4: Parametereingabe

2.2 Kompartimentmodelle und Michaelis-Menten-Beziehungen

Drei unterschiedliche Modellvarianten wurden für Kompartimentmodelle realisiert:

- Das offene Ein-Kompartimentmodell, dessen Regressionsbeziehung z.B. in der Form

$$E[y] = \lambda_i(t, \Theta) = (\theta_1 \theta_3 / (\theta_1 - \theta_2)) (\exp(-\theta_2 t) - \exp(-\theta_1 t))$$

geschrieben werden kann. Hierbei stellt $\lambda_i(t, \Theta)$ die erwartete Substanzkonzentration zur Zeit t dar, der Parameter θ_1 die Absorptionsrate, θ_2 die Eliminationsrate, und θ_3 kann als die physikalische Dichte ([Masse/Volumen]) interpretiert werden (oder auch als der Quotient aus aD und Volumen, wobei a den vom Gastrointestinaltrakt absorbierten Anteil der Substanz und D die zu Beginn verabreichte Dosis bezeichnet). Ein wichtiger Parameter ist hierbei der Bioverfügbarkeitsparameter AUC (area under the concentration curve), z.B. durch den Quotienten θ_3 / θ_2 gegeben.

- das dreiparametrische Ein-Kompartiment-Modell

$$\begin{aligned} E[y] &= \lambda(t, \Theta) \\ &= E[\text{Response bei Konzentration } x] \\ &= \theta_1 \exp(\theta_2 x) - \theta_3 \end{aligned}$$

und

- das vierparametrische Kompartiment-Modell.

$$\begin{aligned} E[y] &= \lambda(t, \Theta) \\ &= E[\text{Response bei Konzentration } x] \\ &= \theta_1 \exp(\theta_2 x) + \theta_3 \exp(\theta_4 x) \end{aligned}$$

Als Standardvarianzmodell bei Vorliegen von Datenheteroskedastizität wird hier meist das sog. Powerfunktions- oder ein polynomiales Varianzmodell gewählt.

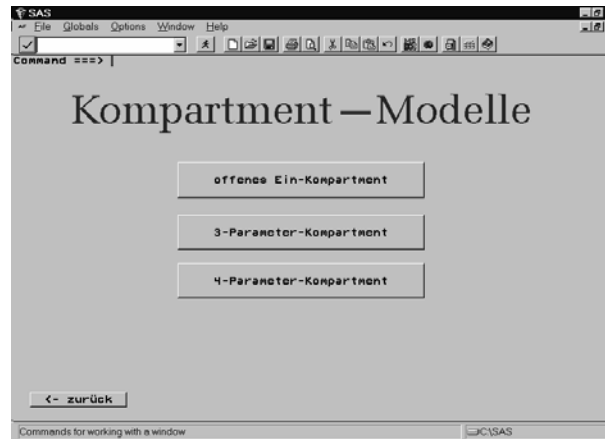


Abb. 5: Wahl des Kompartimentmodells

Außer diesen drei Modellen erlaubt das System auch das Schätzen der Parameter einer Michaelis-Menten-Beziehung

$$\begin{aligned} E[y] &= \lambda(t, \Theta) \\ &= E[\text{Response bei Konzentration } x] \\ &= \frac{\theta_1 x}{\theta_2 + x} \end{aligned}$$

mit oder ohne eine Varianzmodellierung.

3. Varianzmodelle

Um eine größere Flexibilität hinsichtlich der Datenanpassung zu erreichen, wurden die folgenden Varianzmodelle in das System integriert. Es bezeichnen dabei y die Response, λ den Erwartungswert und Λ die Varianz. Man beachte auch, daß die Wahl eines geeigneten Varianzmodelles u.U. eine schwierige Aufgabe ist, da dabei Aspekte der Modellierung des zugrundeliegenden Prozesses zu berücksichtigen sind. Auf keinen Fall sollte unter den zur Verfügung stehenden Varianzmodellen eines beliebig ausgewählt und in den Anpassungsalgorithmus eingebunden werden.

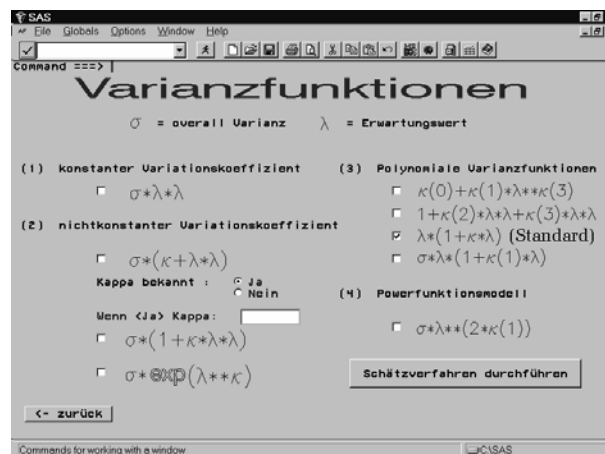


Abb. 6: Wahl der Varianzfunktion

Varianzfunktionen:

- Konstanter Variationskoeffizient:

$$\Lambda = \sigma_0^2 \lambda^2$$

σ_0 entspricht dann dem Variationskoeffizienten

- Varianten mit nichtkonstantem Variationskoeffizient:

- $\Lambda = \sigma_0^2(\kappa^* + \lambda^2)$

- κ^* bekannte Konstante oder auch zu schätzen

- $\Lambda = \sigma_0^2 (1 + \kappa^* \lambda^2)$

- $\Lambda = \sigma_0^2 \exp(\lambda^{\kappa^*})$

- $\kappa^* > 0$

- Powerfunktionsmodell:

$$\Lambda = \kappa_0^2 \lambda^{2\kappa_1}$$

Das Powerfunktionsmodell erlaubt verschiedene Interpretationsmöglichkeiten: $\kappa_0=1$, $\kappa_1=0.5$ entspricht dabei dem Standard-Poisson-Modell, $\kappa_1 > 0.5$ einem überstreuten Modell, $\kappa_1=1$ einem konstanten Variationskoeffizienten oder $\kappa_1=0$ einer konstanten Varianz.

- Polynomiale Varianzfunktionen:

$$\Lambda = \kappa_0 + \kappa_1 \lambda^{\kappa_3}$$

$$\Lambda = 1 + \kappa_2 \lambda^2 + \kappa_3 \lambda^3$$

$$\Lambda = \kappa_2 \lambda^2 + \kappa_3 \lambda^3$$

$$\Lambda = \sigma_0^2 \lambda (1 + \kappa^* \lambda)$$

$$\Lambda = \lambda (1 + \kappa^* \lambda)$$

Letztere Varianzfunktionen haben natürlich auch keinen konstanten Variationskoeffizienten mehr!

4. Schätzverfahren

Den qualitativen Dosis-Wirkungs-Beziehungen liegt eine Binomialverteilung zugrunde, was die Verwendung von Maximum-Likelihood-Schätzern impliziert. Das verwendete Schätzverfahren ist dabei ein iterativer Kleinste-Quadrate-Algorithmus (IRLS) mit der Varianz der Binomialverteilung als Gewicht.

Komplexer wird es bei den anderen Modellen, wenn Heteroskedastizität vorliegt und zusätzliche Parameter in der Varianzfunktion geschätzt werden müssen. In diesen Fällen ist als Schätzprozedur ein verallgemeinerter Kleinste-Quadrate-Algorithmus (GLS-Algorithmus) zu wählen. Die Grundlage dieses Algorithmus bilden die Quasi-Likelihood-Methoden von Wedderburn [5] und McCullagh [4]. Das GLS-Verfahren verläuft im wesentlichen in zwei Schritten, die einige Male zu iterieren sind:

- Schritt 0: Suche Startwerte für die Parameter der Erwartungswertfunktion und der Varianzfunktion. Dies kann in den meisten Fällen mit Hilfe der Methode der Kleinsten Quadrate geschehen.

Die folgenden beiden Schritte sind dann mindestens dreimal zu iterieren:

- Schritt 1: Schätze die Parameter der Varianzfunktion und halte dabei die Parameter des Erwartungswertes fest (gewichtete Schätzung).
- Schritt 2: Schätze die Parameter der Erwartungswertfunktion und halte dabei die Parameter der Varianz fest (ebenfalls gewichtete Schätzung, Gewichte: Varianzfunktion).

Modifikationen des GLS-Verfahrens wie z.B. robuste Varianten (Tukey, Huber oder L_1 -Norm) können in diesen Algorithmus leicht integriert werden. Realisiert wurden die Schätzverfahren über SAS-Macros, die hauptsächlich mit den SAS-Prozeduren NLIN (GLS) und NLP (L_1 -Norm) arbeiten. Bei den qualitativen Dosis-Wirkungsbeziehungen (vgl. Abb. 2) wurde zusätzlich noch die Möglichkeit von Parameterschätzungen mit Hilfe der SAS-Prozedur PROBIT implementiert, die die Anpassung eines sog. Schwellenwertmodells erlaubt. Alle Modelle erlauben natürlich auch graphische Ausgaben der angepaßten Modelle über die Prozedur GPLOT.

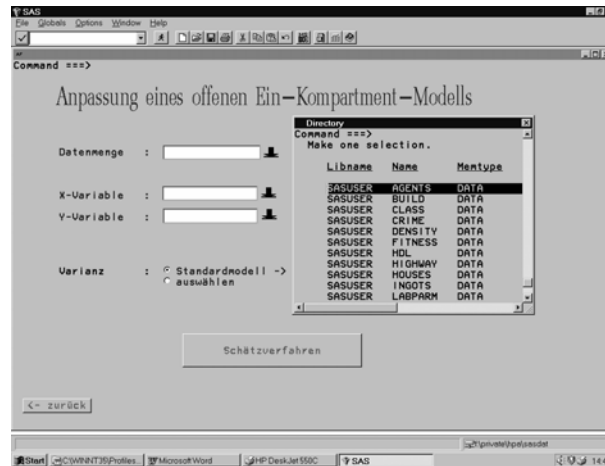


Abb. 7: Auswahl der Datenmenge für ein offenes Ein-Kompartimentmodell

Literatur

1. Altenburg, H.-P. (1994): Fitting Response Curves for Radioimmunoassays or Immunoradiometric Assays. Eur. J. Clin. Chem. Biochem. 32, 873-880
2. Altenburg, H.-P. and Rosenkranz, G. (1990): Analysis of Dose Resonse Data with SAS. In: F. Faulbaum et al. 1990: SoftStat '89 - Fortschritte der Statistik-Software 2, G. Fischer Stuttgart, 183-190
3. Altenburg, H.-P. (1997): Wie robust sind nichtlineare GLS- und L_p -Norm-Verfahren? Ein Vergleich für eine pharmakologische Fragestellung. In: Medizinische Informatik, Biometrie und Epidemiologie - GMDS '97. 42. Jahrestagung der GMDS Ulm, September 1997, MMV Medizin Verlag, München, 325-329
4. McCullagh, P. (1983): Quasi-likelihood functions. Annals of Statistics 11, 59-67
5. Wedderburn, R.W.M. (1974): Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61, 439-447
6. SAS/STAT: User's Guide. Volume 1 and 2. Version 6. SAS Institute, Cary, North Carolina