

## Auswertung von kategorialen Daten mit Messwiederholung in SAS

Norbert Mielenz  
Martin-Luther-Universität Halle-  
Wittenberg,  
Institut für Agrar- und  
Ernährungswissenschaften  
Adam-Kuckhoff-Straße 35  
Halle (Saale)  
norbert.mielenz@landw.uni-halle.de

Joachim Spilke  
Martin-Luther-Universität Halle-  
Wittenberg,  
Institut für Agrar- und  
Ernährungswissenschaften  
Ludwig-Wucherer-Str. 82-85  
Halle (Saale)  
joachim.spilke@landw.uni-halle.de

### Zusammenfassung

Zur mathematischen Beschreibung von geordneten kategorialen Daten, die sich durch wiederholte Merkmalerfassung pro Objekt auszeichnen, stehen zwei grundsätzliche Ansätze zur Verfügung. Liegt das Interesse nur auf dem Populationsmittel und den Effekten von erklärenden Variablen auf dieses Mittel, so kann das marginale Modell (population averaged model) zur Beschreibung der korrelierten Daten verwendet werden. Dieses Modell erweitert das auf unabhängige Daten anwendbare generalisierte lineare Modell (GLM) durch Einbeziehung einer Arbeitskorrelationsmatrix (working correlation matrix) auf korrelierte Daten. Sind dagegen Aussagen über die zufälligen Effekte oder über die zufälligen Regressionskoeffizienten pro Objekt zu treffen, so erfolgt die Auswertung zweckmäßig mit generalisierten linearen Modellen, die zusätzlich zufällige Effekte im linearen Prädiktor enthalten. Diese so genannten generalisierten linearen gemischten Modelle (GLMM) werden in der Literatur im Gegensatz zu den „population averaged“ Modellen auch als subjekt-spezifische Modelle bezeichnet. Die Parameterschätzung in marginalen Modellen und in GLM's mit zufälligen Effekten unter Verwendung der SAS-Prozeduren GENMOD, GLIMMIX und NLMIXED wird an zwei Beispielen aus dem Tierversuchswesen demonstriert.

**Schlüsselwörter:** korrelierte kategoriale Daten, subjekt-spezifisch, population-averaged, GENMOD, GLIMMIX, NLMIXED

## 1 Einleitung

Zur Auswertung von Merkmalen, deren Ausprägungen nur in verschiedenen Kategorien vorliegen, bedient man sich zunehmend generalisierter linearer Modelle. Diese Modelle verallgemeinern die Auswertung normalverteilter Merkmale mit linearen gemischten Modellen auf die Auswertung beispielsweise binomial- oder multinomial verteilter korrelierter Daten. Unter Verwendung von speziellen Linkfunktionen wird der Zusammenhang zwischen den Einflussfaktoren als Bestandteil des linearen Prädiktors und dem Erwartungswert des Zielmerkmals hergestellt.

## 2 Problemstellung

Beispiele für kategoriale Merkmale im Tierversuchswesen sind die Klauenbefunde von Milchkühen oder die Boniturnoten der Befiederungsschäden von Legehennen.

### Beispiel 1:

Die Klauenbefunde stellen ein binäres Merkmal mit den Ausprägungen 0=gesund und 1=Befund dar. Das so definierte Merkmal wurde an 451 Kühen mehrfach erfasst, wobei die Untersuchungstiere von 96 Bullen abstammen. Gesucht wird beispielsweise ein möglicher Einfluss der Laktationsnummer oder des Betriebes auf die Klauengesundheit.

### Beispiel 2:

Die Bonitierung von 365 Legehennen erfolgte an verschiedenen Körperpartien (wie beispielsweise Hals, Brust und Rücken) mit Noten von 0 bis 6. Hierbei steht 0 für „perfekt“ und 6 für „kahl“. Die Hennen standen in verschiedenen Haltungssystemen. Die Bonitierung erfolgte in der 45. und 70. Lebenswoche.

### 2.1 Das subjekt-spezifische Modell

Den genannten Beispielen ist gemeinsam, dass geordnete kategoriale Daten vorliegen, die pro Tier wiederholt erfasst wurden.

Zur Auswertung derartiger Daten stehen zwei grundsätzliche mathematische Ansätze zur Verfügung. Besonders in der Tierzucht gut eingeführt ist das subjekt-spezifische Modell. Bei diesem Ansatz werden bedingte Erwartungswerte bzw. Wahrscheinlichkeiten (WK) betrachtet. Die WK, dass eine Beobachtung von Tier  $i$  den Wert  $k$  besitzt, gegeben den zufälligen Effekt  $u_i$  von Tier  $i$ , wird über ein Schwellenwertmodell abgebildet. Bezeichne  $Y_{ij}$  die Zufallsgröße zugehörig zur  $j$ -ten Beobachtung von Tier  $i$ . Dann gilt:

$$(1) \quad \Pr(Y_{ij} = k | u_i) \Leftrightarrow \Pr(Z_{ij} \in (\theta_{k-1}, \theta_k]); \quad k = 1, \dots, C$$

mit  $-\infty = \theta_0 < \theta_1 < \dots < \theta_{C-1} < \theta_C = \infty$ ,  $\theta_k = \text{Stützpunkte}$

Die kategoriale Zufallsgröße  $Y_{ij}$  nimmt den Wert  $k$  an, falls eine stetige, nichtbeobachtbare Zufallsgröße  $Z_{ij}$  Werte in einem bestimmten Intervall annimmt. Sei  $\beta$  der Vektor der fixen Modelleffekte,  $u_i$  der Vektor der zufälligen Effekte von Tier  $i$  und  $x'_{ij}$  (bzw.  $z'_{ij}$ ) zugehörige Versuchsplanvektoren. Dann wird der Erwartungswert von  $Z_{ij}$  mit Hilfe des linearen Prädiktors wie folgt angegeben:

$$(2) \quad \eta_{ij} = E(Z_{ij} | u_i) = -[x'_{ij}\beta + z'_{ij}u_i]$$

Durch Verwendung des Minuszeichens lässt sich dann der Zusammenhang zwischen dem linearen Prädiktor und der kumulativen Wahrscheinlichkeit mit Hilfe der Linkfunktion  $g(\cdot)$  wie in SAS üblich angeben durch:  $g(\Pr(Y_{ij} \leq k | u_i)) = \theta_k + \eta_{ij}$ . Im Spezialfall lediglich einer kontinuierlichen Einflussgröße kann der lineare Prädiktor die folgende Gestalt besitzen.

$$(3) \quad \eta_{ij} = -[(\beta_0 + u_{0i}) + (\beta_1 + u_{1i}) \cdot t_j]$$

In obiger Darstellung werden die zufälligen tierspezifischen Regressionskoeffizienten  $u_{0i}$  und  $u_{1i}$  auch als „intercept“ und „slope“ bezeichnet.

Ist  $Z_{ij}$  normalverteilt, so lassen sich die kumulativen WK mit Hilfe der Probit-Linkfunktion ausdrücken durch:

$$(4) \quad \Phi^{-1}(\Pr(Y_{ij} \leq k | u_i)) = \theta_k + \eta_{ij}; \quad \Phi(\cdot) = \text{standardisierte Normalverteilung.}$$

Die Schätzung der Stützpunkte und der Modellparameter im linearen Prädiktor kann mit der Maximum-Likelihood-Methode erfolgen [9]. Bei zwei zufälligen Effekten pro Objekt sind jedoch bereits Doppelintegrale numerisch zu berechnen. Deshalb existieren Näherungsverfahren, die sich in die Klasse der Pseudo-Likelihood-Methoden einordnen lassen [13]. Das subjekt-spezifische Modell wird in der Literatur auch als generalisiertes lineares Modell mit Zufallseffekten oder kurz als bedingtes (conditional) Modell bezeichnet.

## 2.2 Das „population-averaged“ Modell

Neben den subjekt-spezifischen Modellen sind sogenannte population-averaged Modelle weit verbreitet. Bei diesen Modellen wird der marginale Erwartungswert als Funktion lediglich der fixen Modelleffekte formuliert. Liegen nur 2 Kategorien vor, so gilt:

$$(5) \quad \mu_{ij} = E(Y_{ij}) = 0 \cdot \Pr(Y_{ij} = 0) + 1 \cdot \Pr(Y_{ij} = 1) = \Pr(Y_{ij} = 1) = g^{-1}(\theta_1 + x'_{ij}\beta)$$

mit  $g(\cdot) = \text{Logit - oder Probitlinkfunktion}$

Im Fall einer Multinomialverteilung (mit mehr als zwei Kategorien) muss, um eine zu (5) analoge Darstellung zu erhalten, zuerst eine Transformation in eine multivariate binäre Zufallsgröße  $Z_{ij} = (Z_{ij,0}, \dots, Z_{ij,C-1})'$  erfolgen [5]. Folglich gilt:

$$(6) \quad \mu_{ijk} = \Pr(Y_{ij} \leq k) = E(Z_{ijk}) = g^{-1}(\theta_k + x'_{ij}\beta); \quad g^{-1}(\cdot) = \text{inverse Linkfunktion}$$

mit  $Z_{ijk} = 1$  für  $Y_{ij} \leq k$  und  $Z_{ijk} = 0$  für  $Y_{ij} > k$ ;  $k = 0, 1, \dots, C-1$

Die Matrix der Varianzen und Kovarianzen zwischen den binären Zufallsgrößen pro Objekt wird mit Hilfe der sogenannten „working correlation matrix“  $R(\alpha)$  abgebildet. Diese muss im Fall multivariater binärer Zufallsgrößen eine spezielle Struktur besitzen (vgl. [10])

$$(7) \quad \text{Sei } Z_i = (Z_{i1,0}, \dots, Z_{i1,C-1}, \dots, Z_{iT,0}, \dots, Z_{iT,C-1})'$$

$R(\alpha) = R^*(\alpha) \otimes S = \text{„working correlation matrix“}$

Hierbei ist  $R^*(\alpha)$  die Korrelationsmatrix zwischen den T Messzeitpunkten und S die Korrelationsmatrix zwischen den multivariaten binären Zufallsgrößen zu jedem Mess-

zeitpunkt. Die Schätzung der jetzt ausschließlich fixen Modellparameter erfolgt mit der „Generalized Estimating Equation“-Methode (GEE-Methode, vgl. [3], [7]). Die Schätzung der Elemente der Matrix  $R(\alpha)$  kann beispielsweise mit der Momentenmethode unter Verwendung der  $\mu_{ijk}$  aus Gleichung (6) erfolgen. Das „population-averaged model“ wird in der Literatur auch kurz als marginales Modell bezeichnet.

## 2.3 Interpretation der Modellparameter

In beiden Modellen besitzen die in Vektor  $\beta$  gesammelten fixen Effekte unterschiedliche Interpretation:

- im marginalen Modell erhält man Erwartungswerte bzw. WK bereits gemittelt über alle Werte der zufälligen Effekte,
- im bedingten Modell erhält man WK lediglich für gegebene Werte der zufälligen Effekte.

Nur in Spezialfällen lassen sich bedingte WK in marginale WK umrechnen ([1], [12], [15])

## 3 Auswertung mit SAS

Die klassischen SAS Prozeduren zur Anpassung von marginalen und bedingten Modellen an geordnete, korrelierte kategoriale Daten sind: GENMOD, GLIMMIX und NLMIXED.

- In GENMOD ist die GEE-Methode realisiert.
- Näherungsmethoden (wie Linearisierung oder auch äquivalent dazu Pseudo-Likelihood-Methoden) sind in GLIMMIX umgesetzt.
- Die ML-Methode (also numerische Integration) ist in NLMIXED enthalten und seit Version 9.2 Bestandteil von GLIMMIX.

Nachfolgend wird die Anwendung der drei Prozeduren an den zwei Beispielen aus Abschnitt 2 demonstriert.

### 3.1 Marginale Modelle: Prozedur GENMOD

#### Beispiel 1: Klauenbefunde (Umsetzung in SAS)

```
PROC GENMOD DATA=klauen desc;  
  CLASS vdate dimcl lanr kuh vat;  
  MODEL bef=vdate dimcl lanr / DIST=binomial LINK=logit;  
  REPEATED SUBJECT=kuh(vat) / type=cs corrw;  
  LSMEANS lanr;  
RUN;
```

Neben der Laktationsnummer (lanr) wurden im Prädiktor noch das Besuchsdatum (vdate) und das Stadium der Laktation (dimcl) als fixe Effekte berücksichtigt. Es liegen nur 2 Kategorien vor. Folglich kann mit der Binomialverteilung gearbeitet werden. Die Korrelationsstruktur zwischen den wiederholten Beobachtungen pro Tier wird mit der repeated-Anweisung berücksichtigt. Der zusätzliche zufällige Effekt des Vaters lässt sich bei Verwendung von GENMOD nicht im Modell berücksichtigen.

**Beispiel 2: Befiederungsschäden (Umsetzung in SAS)**

```
PROC GENMOD DATA=r_fieder;
  CLASS anim genotyp trt;
  MODEL breast=genotyp trt / DIST=multinomial LINK=cumprobit;
  REPEATED SUBJECT=anim / type=ind corrw;
  ESTIMATE 'LS at TRT=1 k=2' int 0 0 1 genotyp 1 trt 1;
  ODS OUTPUT ESTIMATES=est;
RUN;
```

Im Prädiktor wurden die Stufen der Prüffaktoren Haltungssystem (trt) und Rasse (bzw. Genotyp) als fixe Effekte berücksichtigt. Es liegen insgesamt 7 Kategorien vor. Deshalb muss die Multinomialverteilung zur Anwendung kommen. Für diese Verteilung lässt sich in GENMOD keine von der Einheitsmatrix abweichende Korrelationsmatrix auswählen. Für die Multinomialverteilung ist keine lsmeans-Anweisung verfügbar.

**Tabelle 1:** Vor- und Nachteile bei der Nutzung von Prozedur GENMOD

Vorteile	Nachteile
<ul style="list-style-type: none"> <li>• class- und lsmeans-Anweisung vorhanden</li> <li>• Schätzung von marginalen Erwartungswerten und Wahrscheinlichkeiten</li> <li>• für GLM ohne Messwiederholung fallen ML- und GEE-Methode „zusammen“</li> </ul>	<ul style="list-style-type: none"> <li>• Korrelationsstruktur wird ohne Verteilungsannahmen abgebildet</li> <li>• keine Vorhersagen für die zufälligen Effekte möglich</li> <li>• multinomial Verteilung nicht für GLM mit Messwiederholungen realisiert</li> <li>• repeated-Anweisung darf nur auf ein Subjekt, ausgedrückt durch mehrere Faktoren, verweisen</li> </ul>

**3.2 Bedingte Modelle: Prozedur GLIMMIX**

**Beispiel 1: Klauenbefunde (Umsetzung in SAS, METHOD=QUAD)**

```
PROC GLIMMIX DATA=klauen METHOD=quad;
  CLASS vdate dimcl lanr kuh vat;
  MODEL bef=vdate dimcl lanr / DIST=binary LINK=logit s;
  RANDOM int / SUBJECT=kuh(vat);
  LSMEANS lanr / cl ilink;
RUN;
```

Durch die Angabe METHOD=quad wird numerische Integration (Gaus-Hermite-Quadratur) also eine vollständige Maximum-Likelihood-Analyse zur Schätzung der Modellparameter durchgeführt. Bedingt durch die Schwierigkeiten bei der numerischen Integration ist die ML-Methode im Allgemeinen auf zwei zufällige Effekte pro Objekt (intercept und slope) begrenzt. Zufällige Vataeffekte lassen sich bei Anwendung der ML-Methode nicht berücksichtigen.

### Beispiel 1: Klauenbefunde (Umsetzung in SAS, METHOD=RSPL)

```
PROC GLIMMIX DATA=klauen METHOD=rspl;  
  CLASS vdate dimcl lanr kuh vat;  
  MODEL bef=vdate dimcl lanr / DIST=binary LINK=logit s;  
  RANDOM vat;  
  RANDOM int / SUBJECT=kuh(vat);  
  LSMEANS lanr / cl ilink;  
RUN;
```

Sollen im linearen Prädiktor die zufälligen Vateffekte berücksichtigt werden, so muss zur Schätzung der Modellparameter ein Näherungsverfahren wie die Pseudo-Likelihood-Methode zur Anwendung kommen. Im Fall der RSPL-Methode werden die Daten durch subjekt-spezifische Taylorentwicklung linearisiert ([13]).

### Beispiel 2: Befiederungsschäden (Umsetzung in SAS, METHOD=QUAD)

```
PROC GLIMMIX DATA=r_fieder METHOD=quad;  
  CLASS anim genotyp trt;  
  MODEL breast=genotyp trt / DIST=multinomial LINK=cumprobit;  
  RANDOM int / SUBJECT=anim;  
  ESIMATE 'LS at TRT=1 k=2' int 0 0 1 genotyp 1 trt 1;  
  ODS OUTPUT ESTIMATES=est;  
RUN;
```

Die Multinomialverteilung ist in GLIMMIX umgesetzt. Ab SAS-Version 9.2 kann numerische Integration zur Realisierung der ML-Methode angewendet werden. Für die Multinomialverteilung ist keine lsmeans-Anweisung verfügbar.

Betrachtet wird nachfolgend ein generalisiertes lineares (gemischtes) Modell in der folgenden allgemeinen Form:

$$(8) \quad \begin{aligned} \mu &= E(y | u) = g^{-1}(X\beta + Zu) \\ u &\sim N(0, G); \quad \text{Var}(y | u) \approx \phi \cdot A^{\frac{1}{2}} R(\alpha) A^{\frac{1}{2}} \text{ mit } A = \text{Diag}(v(\mu)) \end{aligned}$$

In Darstellung (8) bringt die Schreibweise  $v(\mu)$  zum Ausdruck, dass sich die Diagonalelemente von Matrix  $A$  in Abhängigkeit von der unterstellten Verteilung (gewählt aus der Familie der Exponentialverteilungen) als Funktion der Erwartungswerte schreiben lassen. Unter Verwendung von Darstellung (8) kann die Prozedur GLIMMIX hinsichtlich des Auftretens von zufälligen Modelleffekten im linearen Prädiktor wie folgt systematisiert werden.

**Tabelle 2:** Modellklassen umgesetzt in Prozedur GLIMMIX

Vektor u	zufällige Effekte		Modellklasse
	G-side	R-side	
No	No	$R(\alpha)$	population-averaged (marginal) model
No	No	$R(\alpha)=I$	generalized linear model (GLM <sup>1</sup> )
No	No	$R(\alpha)=\phi \cdot I$	GLM mit overdispersion
Yes	Yes	No	subject specific (conditional) model generalized linear mixed model (GLMM)

<sup>1</sup>)GLM als Obergriff für Modelle ohne Berücksichtigung von Messwiederholung

**Tabelle 3:** Vor- und Nachteile bei der Nutzung von Prozedur GLIMMIX

Vorteile	Nachteile
<ul style="list-style-type: none"> <li>• Pseudo-Likelihood-Methoden für sehr komplexe GLM mit zufälligen Effekten umgesetzt</li> <li>• volle ML-Analyse (Integration) für einfache subjektspezifische Modelle möglich</li> <li>• class-Anweisung vorhanden, Freiheitsgrad-Approximation analog zu Prozedur MIXED möglich</li> </ul>	<ul style="list-style-type: none"> <li>• Alternativen zu Pseudo-Likelihood-Methoden auf einfache subjektspezifische Modelle beschränkt</li> <li>• bei Laplace-Approximation keine komplexen G-side Effekte möglich</li> <li>• Leistungsfähigkeit der approximativen Methoden schwer einschätzbar!</li> </ul>

### 3.3 Bedingte Modelle: Prozedur NLMIXED

Subjekt-spezifische Modelle zur Auswertung der Befiederungsdaten lassen sich ebenfalls mit der Prozedur NLMIXED anpassen. Leider ist keine class-Anweisung vorhanden. Deshalb müssen Faktoren mit endlicher Stufenzahl unter Verwendung von Hilfsvariablen als Regressionsmodell abgebildet werden. Als Methode zur Parameterschätzung ist lediglich die numerische Integration umgesetzt. Die Anwendung von NLMIXED erfordert hohen Programmieraufwand. Die notwendigen Anweisungen sind in der Literatur ausführlich beschrieben (vgl. [1], [4], [11]).

- Innerhalb von NLMIXED lassen sich Felder vereinbaren. Weiterhin besteht die Möglichkeit beispielsweise für das „cumulative probit model“ bedingte in marginale WK exakt umzurechnen.
- Insbesondere lassen sich Differenzen von WK berechnen und mit Hilfe der ESTIMATE-Anweisung auf Signifikanz testen.
- Im Prädiktor können Modellparameter nichtlinear auftreten. Beispielsweise lässt sich mit NLMIXED das sogenannte „multiplicative cumulative probit model“ [8] gut anpassen.

Nachfolgender Auszug aus unserem SAS-Programm deutet die oben genannten Vorteile an. Das vollständige Programm liegt bei den Autoren vor.

```
PROC NLMIXED DATA=r_fieder METHOD=gauss;
    /* Programm bei den Autoren verfügbar */
```

```

/* Berechnung marginaler WK */
c=1.0/sqrt(1.0+s2u);
array qm[2,6];
do i=1 to 2;
  ii=i-1;
  do k=1 to 6;
    qm[i,k]=probnorm(c*(t[k]-(b0-b1*ii)));
  end;
end;
estimate 'qm1_trt1_LS' (qm[1,1]);
RUN;

```

**Tabelle 4:** Vor- und Nachteile bei der Nutzung von Prozedur NLMIXED

Vorteile	Nachteile
<ul style="list-style-type: none"> <li>Parameterschätzung durch numerische Integration (Anwendung der ML-Methode liefert Likelihood und abgeleitet AIC und BIC)</li> <li>Vorhersagen für die zufälligen Effekte möglich</li> <li>Schätzung der Differenz von Wahrscheinlichkeiten (inklusive Durchführung von Tests) möglich</li> </ul>	<ul style="list-style-type: none"> <li>keine class- und lsmeans-Anweisung</li> <li>Numerische Integration gekoppelt an relativ einfache Kovarianzstruktur (ein- bzw. zweidimensionale Integrale!)</li> <li>teilweise hoher Programmieraufwand</li> </ul>

## 4 Diskussion

Die bisher vorgestellten drei Prozeduren lassen sich im Vergleich zu Prozedur MIXED hinsichtlich Verteilungsannahmen, Modellansatz und Schätzverfahren wie folgt systematisieren.

**Tabelle 5:** Prozeduren zur Auswertung von wiederholten Beobachtungen pro Objekt

Prozedur	Verteilung	Modell	Methode	default
MIXED	normal	marginal, conditional	ML, REML	REML
GENMOD	(zero-inflated) poisson, (negativ) binomial, ...	marginal	GEE <sup>1</sup>	GEE
GLIMMIX	poisson, binomial, multinomial, ...	marginal, conditional	MMPL <sup>2</sup> , RSPL <sup>3</sup> , Laplace, ML(Quad <sup>4</sup> )	RSPL
NLMIXED	poisson, binomial, multinomial, ...	conditional	ML(Quad <sup>4</sup> )	ML

<sup>1</sup>) Schätzung der „working correlation matrix“ R nach Momentenmethode

<sup>2</sup>) MMPL=„maximum marginal expansions pseudo-likelihood“

<sup>3</sup>) RSPL=„restricted subject-specific expansions pseudo-likelihood“

<sup>4</sup>) Quad=„adaptive Gaussian quadrature“ mit Quad(qpoints=1)=Laplace  
PL und Laplace = approximative Verfahren (Verzerrung!)

## 5 Schlussfolgerungen

Für die Auswertung kategorialer (ordinalskalierter) Merkmale mit mehr als zwei Kategorien bei Vorliegen von korrelierten Daten lässt sich schlussfolgern:

- In SAS ist eine Vielzahl von Möglichkeiten vorhanden.
- Standardanwendungen sind jedoch die Prozeduren GLIMMIX und NLMIXED.
- Monte-Carlo-Markov-Chain (MCMC)-Methoden (umgesetzt in PROC MCMC) sind im Vormarsch! (In der Hilfe von SAS sind bisher nur wenige Beispiele vorhanden!)
- Die MCMC-Methodik (vgl [14]) ist ab Version 9.2 Bestandteil von GENMOD, aber ohne repeated-Anweisung!

Für die Auswertung korrelierter kategorialer Daten (sowohl ordinal- als auch nominalskaliert) existieren in SAS alternative Prozeduren wie beispielsweise LOGISTIC und CATMOD. Diese Prozeduren besitzen aber andere Hauptanwendungsgebiete und sind nicht zur Modellierung komplexer und speziell strukturierter Kovarianzmatrizen geeignet.

Durch Transformation der Ausgangsdaten lassen sich die (in diesem Artikel) vorgestellten Prozeduren auch zur Auswertung lediglich nominalskalierter korrelierter Daten nutzen. Hier sei jedoch lediglich auf die Literatur verwiesen (vgl. [2], [6]).

## Literatur

- [1] Agresti, A.: Categorical Data Analysis. 2nd ed. Wiley, New York (2002)
- [2] Chen, Z.; Kuo, L.: A note on the estimation of the multinomial logit model with random effects. *The American Statistician* 55 (2001) 89-95
- [3] Hardin, J.W.; Hilbe, J.M.: Generalized estimating equations. Chapman&Hall/CRC (2003)
- [4] Hartung, K.; Piepho, H.P.: A Threshold Model for Multiyear Genebank Data Based on Different Rating Scales. *Crop Science* 45 (2005) 1045-1051
- [5] Heagerty, P.J.; Zeger, S.L.: Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* 91 (1996) 1024-1036
- [6] Kuss, O.; McLerran, D.: A note on the estimation of the multinomial logistic model with correlated responses in SAS. *Computer Methods and Programs in Biomedicine* 87 (2007) 262-269.
- [7] Liang, K.-Y.; Zeger, S.L.: Longitudinal data analysis using generalized linear models. *Biometrika* 73 (1986) 13-22.

- [8] Lunn, D.J.; Wakefield, J. Racine-Poom, A.: Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores. *Statistics in Medicine* 20 (2001) 2261-2285
- [9] McCulloch, C.E.; Searle, S.R.: *Generalized, Linear and Mixed Models*. John Wiley & Sons, Inc. (2001)
- [10] Parsons, N.R.; Costa, M.L.; Achten, J.; Stallard, N.: Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. *Computational Statistics and Data Analysis* 53 (2009) 632-641
- [11] Piepho, H.P.: Auswertung von Bonituren mit der SAS Prozedur NLMIXED. *Zeitschrift für Agrarinformatik* 10 (2002) 30-41
- [12] Ritz, J.; Spiegelman, D.: Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research* 13 (2004) 309-323
- [13] Wolfinger, R.D.; O'Connell, M.: Generalized linear models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 48 (1993) 233-243
- [14] Zeger, S.L.; Karim, M.R.: Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association*, 86 (1991) 79-86.
- [15] Zeger, S. L., Liang, K. Y.; Albert, P. S.: Models for longitudinal data: A generalized estimating equation approach. *Biometrics* 44 (1988) 1049--1060.