Regressionsmodelle für diskrete Zielgrößen und korrelierte Beobachtungen

Oliver Kuß*; Uwe Hasenbein**;

*Institut für Medizinsche Epidemiologie, Biometrie und Informatik,
Martin-Luther-Universität Halle-Wittenberg, Halle(Saale)

** Institut für Neurologisch-Neurochirurgische Rehabilitationsforschung (INNRF), Magdeburg

Kolloquium Medizinische Biometrie, Informatik und Epidemiologie, Heidelberg, 4.7.2005

Programm

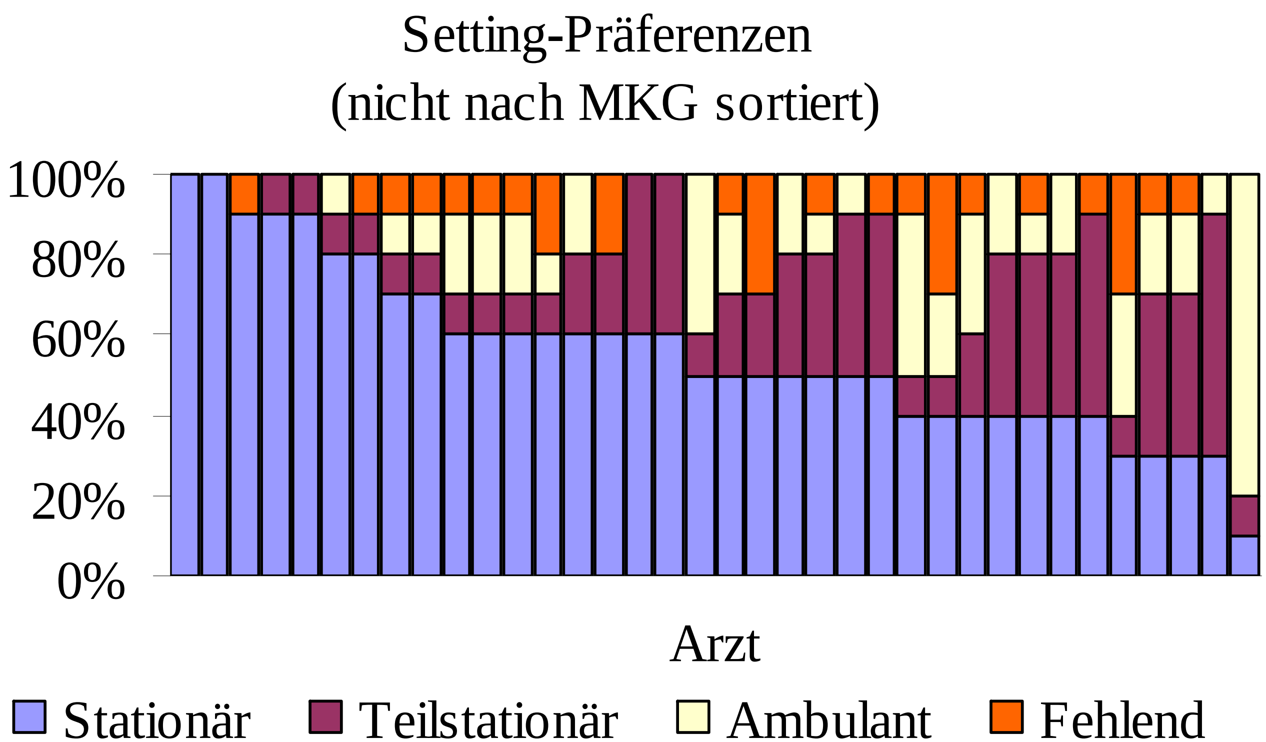
- Motivation
- Probleme
- Teil-Lösungen
- Das Modell
- Resultate
- Fazit

Motivation

36 Ärzte wurden um ihre Einschätzung zur optimalen rehabilitativen Versorgung von Patienten mit Schädel-Hirn-Trauma gebeten.

Jedem wurden 10 Musterkrankengeschichten vorgelegt und sie sollten das optimale Setting für den jeweiligen Patienten (stationäre, teilstationäre oder ambulante Rehabilitation) auswählen.

Welche Faktoren (sowohl die Ärzte als auch die Patienten betreffend) beeinflußen die Setting-Präferenzen?


Setting-Präferenzen
(nicht nach MKG sortiert)
100%
80%
60%
40%
20%
0%
Arzt
Stationär
Teilstationär
Ambulant
Fehlend

Probleme

1. Einschätzungen innerhalb eines Arztes werden korreliert sein.

2. Ist die Zielgröße ordinal oder nominal?

3. Original-Ton U. Hasenbein: „Liegt das teilstationäre Setting näher beim stationären oder beim ambulanten Setting?“

(Teil-)Lösungen I

Lösung Problem 1:

$\longrightarrow$ Mixed Models

Lösung Probleme 1+2:

$\longrightarrow$ Mixed Models für ordinale und nominale Zielgrößen

(**Ordinal:** Hedeker/Gibbons, 1994; Tutz/Hennevogl, 1996; Hartzel/Caffo/Agresti, 2001; **Nominal:** Hedeker, 2003; Hartzel/Caffo/Agresti, 2001; Skrondal/Rabe-Hesketh, 2003)

(Teil-)Lösungen II

Lösung Probleme 2+3

$\longrightarrow$ Stereotype Regression (Anderson, 1984)

Aber alle drei gleichzeitig???

Das Modell: Lösung Probleme 1+2+3

Erweitere Stereotype Modell zu Stereotype Modell mit gemischten Effekten.

Gegeben:

N	Cluster (Ärzte), $i = 1, \ldots, N$, jedes mit T_i Beobachtungen
Y_{ij}	Wert der Zielgröße der j-ten Beobachtung in Cluster i, $j = 1, \ldots, T_i$
π_{ijr}	Ereigniswahrscheinlichkeit $= P(Y_{ij} = r), r = 1, \ldots, R$
x_{ij}	Spaltenvektor von (festen) Kovariablen

Modellsukzession (ohne Korrelation)

- $R = 2$

  Logistische Regression:

  $$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \theta + \beta x_i'$$

- $R > 2$

  Multinomiale Logistische Regression:

  $$\log\left(\frac{\pi_{ir}}{\pi_{i1}}\right) = \theta_r + \beta_r x_i', r = 1, \ldots, R$$

  mit $\theta_1 = 0, \beta_1 = 0$

- $R > 2$, Setze $\beta_r = \phi_r \beta$, $r = 1, \ldots, R$

  Stereotype Regression:

$$\log\left(\frac{\pi_{ir}}{\pi_{i1}}\right) = \theta_r + \phi_r \beta x_i', r = 1, \ldots, R$$

  mit $\theta_1 = 0, \beta_1 = 0, \phi_1 = 0$ und $\phi_R = 1$

Modellsukzession (mit Korrelation)

- $R = 2$

  Logistische Regression mit zufälligem Intercept:

  $$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \theta + \beta x_i' + u_i$$

- $R > 2$

  Multinomiale Logistische Regression mit zufälligem Intercept:

  $$\log\left(\frac{\pi_{ijr}}{\pi_{ij1}}\right) = \theta_r + \beta_r x_{ij}' + u_{ir}, \quad r = 1, \ldots, R$$

- $R > 2$, Setze $\beta_r = \phi_r \beta$, $r = 1, \ldots, R$

  Stereotype Regression mit zufälligem Intercept:

$$\log\left(\frac{\pi_{ijr}}{\pi_{ij1}}\right) = \theta_r + \phi_r \beta x'_{ij} + u_{ir}, \quad r = 1, \ldots, R$$

  mit $\theta_1 = 0, \beta_1 = 0, \phi_1 = 0$ und $\phi_R = 1$

Annahme

Die zufälligen Effekte sind multivariat normalverteilt mit unstrukturierter Kovarianzmatrix $\mathbf{\Sigma}$, d.h. für $u_i' = (u_{i2}, \ldots, u_{i,R})$ gilt

$$u_i \sim N((0, \ldots, 0)', \mathbf{\Sigma}) \text{ mit}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{22}^2 & \sigma_{23}^2 & \cdots & \sigma_{2,R}^2 \\ & \sigma_{33}^2 & \cdots & \sigma_{3,R}^2 \\ & & \ddots & \vdots \\ & & & \sigma_{R,R}^2 \end{pmatrix}$$

Eigenschaften I

- Das Stereotype Modell geht aus dem multinomialen Modell durch die Restriktion $\beta_r = \phi_r \beta$, $r = 1, \ldots, R$, hervor.

- Die ϕ induzieren eine Metrik der Kovariablen:

  Beispiel: Betrachte multinomiale Parameter (R=4)
  $\beta_1 = 0$ $\quad \beta_2 = 0.5$ $\quad \beta_3 = 1$ $\quad \beta_4 = 1.5$

  Diese werden im Stereotype Modell zu:
  $\beta_1 = \phi_1 \beta$ $\quad \beta_2 = \phi_2 \beta$ $\quad \beta_3 = \phi_3 \beta$ $\quad \beta_4 = \phi_4 \beta$
  $\beta_1 = 0 * 1.5$ $\quad \beta_2 = \frac{1}{3} * 1.5$ $\quad \beta_3 = \frac{2}{3} * 1.5$ $\quad \beta_4 = 1 * 1.5$

Also: $\phi_2 = \frac{1}{3}, \ , \phi_2 = \frac{2}{3}, \ \beta = 1.5.$

Das Stereotype Modell hat i.a. weniger Parameter als das multinomiale Modell.

Die Anordnung der ϕ ist keine Modellannahme, sondern kann getestet werden.

Eigenschaften II

- Nichtlinearität der Parameter

- Einführung der zufälligen Effekte in die Modellgleichung modelliert gleichzeitig Heterogenität zwischen und Korrelation innerhalb der Cluster.

- Mehrere zufällige Effekte sind möglich.

- Zu maximierende Likelihoodfunktion ist ein Produkt von N nicht geschlossen lösbaren Integralen.

Schätzmethoden

Adaptive Gauss-Hermite Quadrature: Numerische Integration, SAS PROC NLMIXED, liefert 'exakte' ML-Schätzer

Weitere: MCMC (WinBUGS), Marginale Modelle (GEE, SAS %NLINMIX)

Die R-Funktion nlme (Pinheiro/Bates) liefert nur asymptotische ML-Schätzer (und deren Funktionalität hat SAS %NLINMIX schon seit 1994 ;-))).

Resultate I

Zur Erinnerung:

- **Zielgröße:** Optimales Reha-Setting (stationär [Referenz], teilstationär [TS], ambulant [AB])

- **Kovariablen:** alle binär
  - Arzt: Neurologe, Facharzt
  - Patienten: Zeit seit SHT, Schwere der Beeinträchtigung

- **Fallzahlen:** $N = 36$, $\sum_i^N T_i = 331$

Resultate II

Multinomiales Modell mit zufälligem Intercept (PROC NLMIXED)

Kovariable	$\hat{\beta}_{TS}(SE)$	$\hat{\beta}_{AB}(SE)$
Neurologe	0.25 (0.73)	1.32 (0.92)
Facharzt	-0.46 (0.71)	0.28 (0.86)
Zeit seit SHT	2.37 (0.41)	4.14 (0.59)
Schwere d.B.	-1.60 (0.48)	-2.59 (0.53)

$$\hat{\Sigma} = \begin{pmatrix} 1.65(0.80) & 1.89(0.86) \\ & 2.61(1.20) \end{pmatrix}$$

AIC: 487.18, BIC: 507.76

Resultate III

Stereotype Modell mit zufälligem Intercept (PROC NLMIXED)

Kovariable	$\hat{\beta}(SE)$
Neurologe	1.36 (0.93)
Facharzt	0.40 (0.88)
Zeit seit SHT	4.23 (0.58)
Schwere d.B.	-2.60 (0.53)
ϕ_2	0.55 (0.09)

$$\hat{\Sigma} = \begin{pmatrix} 2.01(0.96) & 1.94(0.93) \\ & 2.62(1.20) \end{pmatrix}$$

AIC: 487.45, BIC: 503.29

Resultate IV

Standard Stereotype Modell (PROC NLMIXED)

Kovariable	$\hat{\beta}(SE)$
Neurologe	0.89 (0.46)
Facharzt	0.19 (0.43)
Zeit seit SHT	3.26 (0.45)
Schwere d.B.	-2.00 (0.43)
ϕ_2	0.50 (0.10)

AIC: 517.30, BIC: 543.91

Resultate V

Stereotype Modell mit zufälligem Intercept (WinBUGS, Nicht-informative Prior-Verteilungen, Läufe: 20.000, Burn-In: 10.000)

Kovariable	$\widehat{\beta}(SE)$
Neurologe	1.39 (1.03)
Facharzt	0.48 (0.95)
Zeit seit SHT	4.38 (0.60)
Schwere d.B.	-2.66 (0.54)
ϕ_2	0.54 (0.09)

$$\hat{\Sigma} = \begin{pmatrix} 2.28(1.15) & 2.22(1.16) \\ & 3.35(1.60) \end{pmatrix}$$

SAS PROC NLMIXED

```
proc nlmixed data=sht;
     parms u2=-1.5 u3=-3 phi2=0.5
           b_facharzt=1.5 ...
           sd2=2 sd3=2 cov23=2;

     eta  = b_facharzt*facharzt + b_neurologe*neurologe +
            b_zeit*zeit + b_schwere*schwere;

     p1=                 1 /(1 + exp(u2 + phi2*eta) + exp(u3 + eta));
     p2= exp(u2 + phi2*eta)/(1 + exp(u2 + phi2*eta) + exp(u3 + eta));
     p3= exp(u3 +      eta)/(1 + exp(u2 + phi2*eta) + exp(u3 + eta));

     if resp1 then loglike=log(p1);  else
     if resp2 then loglike=log(p2);  else
                   loglike=log(p3);

     model setting ~ general(loglike);
     random u2 u3 ~ normal([a2,a3],[sd2,cov23,sd3])
                    subject=arzt;
run;
```

WinBUGS

```
model;
   {for (i in 1:I){
         covariates[i]<- b_facharzt*facharzt[i]+b_neurologe*neurologe[i]+
                         b_zeit*zeit[i]+b_schwere*schwere[i];
         nenner[i]<-(1+ exp(a2+u[experte[i],1]+ phi2*covariates[i])+
                        exp(a3+u[experte[i],2]+      covariates[i]));
         p[i,1] <- 1/nenner[i];
         p[i,2] <- exp(a2+u[experte[i],1] + phi2*covariates[i])/nenner[i];
         p[i,3] <- exp(a3+u[experte[i],2] +      covariates[i])/nenner[i];
         setting[i] ~ dcat(p[i,1:3])
                  }

 # Prior-Verteilungen fuer die Parameter
   a2~dnorm(0,0.00001); a3~dnorm(0,0.00001);
   phi2~dnorm(0,0.00001);b_neurologe~dnorm(0,0.00001); ...

 # Multivariate Normalverteilung fuer den Prior der RE-Verteilung
   for (k in 1:K) {u[k,1:2]~dmnorm(mu[],R[,]);}

 # (Hyper-)Prior-Verteilung fuer die (inverse) Kovarianzmatrix der REs
   R[1:2,1:2]~dwish(V[,],2);
   }
```

Fazit

- Faktoren, die die ärztliche Empfehlung beeinflussen, konnten identifiziert werden.

- Zusätzliche Information über die Abstände der Kategorien der Zielgröße konnte gewonnen werden.

- Schätzung mit Standard-Software ist möglich.

Literatur

Anderson JA, (1984): Regression and Ordered Categorical Variables. Journal of the Royal Statistical Society, Series B, 46, 1-30.

Hasenbein U, Kuß O, Bäumer M, Schert C, Schneider H, Wallesch CW, (2003): Physicians' preferences and expectations in traumatic brain injury rehabilitation - results of a case-based questionnaire survey. Disability and Rehabilitation 25, 136-42.

Hartzel J, Agresti A, Caffo B, (2001): Multinomial Logit Random Effects Models. Statistical Modelling 1, 81-102.

Hedeker D, (2003): A mixed-effects multinomial logistic regression model. Statistics in Medicine 22, 1433-46.

Hedeker D, Gibbons RD, (1994). A random-effects ordinal regression model for multilevel analysis. Biometrics 50, 933-44.

Skrondal A, Rabe-Hesketh S, (2003): Multilevel logistic regression for polytomous data and rankings. Psychometrika 68, 267-287.

Tutz G, Hennevogl W, (1996): Random effects in ordinal regression models. Comput. Stat. Data Anal. 22, 537-557.