

2 Regression mit binärem Response

Aufgabe 16

Die Zusammenhangsstruktur dreier binärer Variablen X_A , X_B und X_C lasse sich durch das Modell $X_A X_B / X_A X_C / X_B X_C$ beschreiben. Welcher Zusammenhang besteht dann zwischen dem zugehörigen loglinearen Modell und einem Logit-Modell mit X_B als abhängiger Variable und X_A und X_C als Einflussgrößen?

Aufgabe 17

Zur Untersuchung der Erklärungskraft von Kovariablen auf einen binären Response mit den Werten 0 bzw. 1 dient u.a. folgendes nach Gini benanntes Maß:

$$G = \frac{\sum_{i=1}^n \hat{\pi}_i^2 - n\bar{p}^2}{n\bar{p}(1-\bar{p})}$$

Dabei steht \bar{p} für den Anteil an Einsern bei der Zielgröße in der Stichprobe. Zeigen Sie, dass G dem Bestimmtheitsmaß R^2 im Falle einer binären Zielgröße entspricht.

Aufgabe 18

Zur Modellierung von Überdispersion in einem Binomial-Modell besteht ein sinnvoller Ansatz darin, die zusätzliche Variabilität durch eine latente Variable zu erklären. Anstatt einer Binomialverteilung für die Anzahl der Erfolge in einer Gruppe $i = 1, \dots, N$

$$y_i \sim B(n_i, \pi_i)$$

wird nun folgendes Modell angenommen:

- Für die latente Variable $D_i \in [0; 1]$ gelte:

$$\begin{aligned} E(D_i) &= \pi_i \\ \text{Var}(D_i) &= \delta \pi_i (1 - \pi_i) \quad \text{mit} \quad \delta > 0 \end{aligned}$$

- Gegeben $D_i = \vartheta_i$ gelte:

$$y_i | D_i = \vartheta_i \sim B(n_i, \vartheta_i)$$

- (a) Zeigen Sie, dass dann für die marginale Verteilung von y_i gilt:

$$\begin{aligned} E(y_i) &= n_i \pi_i \\ \text{Var}(y_i) &= n_i \pi_i (1 - \pi_i) \phi_i \quad \text{mit} \quad \phi_i = 1 + (n_i - 1) \delta \end{aligned}$$

- (b) Im Falle von $D_i \sim Be(a_i, b_i)$ spricht man vom sog. Beta-Binomial-Modell. Wie lautet die entsprechende Parametrisierung über π_i und δ ? Überlegen Sie sich hierzu ein adäquates Beispiel.

Aufgabe 19

Es sollen die Samen der Sommerwurz (Fachbegriff *orobanche*; auch Unkraut genannt) untersucht werden. Betrachten Sie hierzu den Datensatz *orob2* aus dem Paket `aod`, der ursprünglich von Crowder (1978) analysiert wurde. Berechnen und vergleichen Sie die in den entsprechenden Teilaufgaben genannten Modelle mit folgenden Kovariablenstrukturen:

- (i) `seed`
- (ii) `seed + root`
- (iii) `seed * root`

Die Zielgröße soll dabei jeweils *Samen gekeimt ja/nein* sein. Des Weiteren soll stets der Logit-Link verwendet werden.

- (a) Binomialmodelle mittels der Funktion `glm()`.
- (b) Beta-Binomial-Modelle mittels der Funktion `betabin()`.
- (c) Quasi-Binomial-Modelle mittels der Funktion `glm()`.
- (d) Vergleichen Sie die einzelnen Modell-Outputs. Was fällt Ihnen auf? Diskutieren Sie Möglichkeiten des Modellvergleichs.