Statistik IV für Nebenfachstudierende

4.2. Responseverteilung Mehrkategorielle Regression

Prof. Dr. Andreas Mayr

Institut für Statistik, LMU München

Sommersemester 2017

Mehrkategorielle Regression

- Generelles Setting:
- Ungeordnete Kategorien:
- Ordinale Kategorien:

Schreibweise

Statt $y \in \{1, ..., k\}$ betrachte Vektor $\underline{y}^\top = (y_1, ..., y_k)$

wobei

Für eine Beobachtung erhält man:

Response - Verteilung

Mit $\pi_r =$

lässt sich definieren

Multinomialverteilung

$\underline{y}^\top = (y_1, ..., y_k)$ ist multinomialverteilt

$\underline{y} \sim M(n, \underline{\pi}^\top = (\pi_1, ..., \pi_k))$ und

$$\mathbb{P}(\underline{y}^\top = (n_1, ..., n_k)) = \frac{n!}{n_1! \cdot ... \cdot n_k!} \pi_1^{n_1} \cdot ... \cdot \pi_k^{n_k}$$

mit $\sum_{r=1}^{k} n_r = n$ und $n_r \in \{0, 1, ..., n\}$.

Bemerkung:

Statt

$$\mathbb{P}(\underline{y}^\top = (n_1, ..., n_k)) = \frac{n!}{n_1! \cdot ... \cdot n_k!}\ \pi_1^{n_1} \cdot ... \cdot \pi_k^{n_k}$$

wird auch oft

$$\mathbb{P}(\underline{y}^\top = (y_1, ..., y_k)) = \frac{n!}{y_1! \cdot ... \cdot y_k!}\ \pi_1^{y_1} \cdot ... \cdot \pi_k^{y_k}$$

verwendet.

Motivation

Urnenmodell:

Spezialfall $k = 2$

Binärer Fall:

Eigenschaften

Für $\underline{y}^\top = (y_1, ..., y_k) \; \sim \; M(n, \; \underline{\pi}^\top = (\pi_1, ..., \pi_k))$ gilt:

$\text{Var}(y_r) =$

$\text{Cov}(y_r, y_s) =$

$\mathbb{E}(y_r) =$

Skalierte Multinomialverteilung

$$\underline{y}^\top = (y_1, ..., y_k) \;\sim\; M(n,\; \underline{\pi}^\top = (\pi_1, ..., \pi_k))$$

$\underline{y}/n$ ist *skaliert* Multinomoalverteilt:

Bemerkung

Wir haben beobachtet:

$\underline{y}^\top = (y_1, ..., y_k)$ bei k Alternativen

Es reicht aber prinzipiell

$\underline{y}^\top = (y_1, ..., y_q)$ mit $q = l - 1$ zu betrachten, weil: