Vorlesung: Lineare Modelle

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2015

5 Metrische Einflußgrößen: Polynomiale Regression, Trigonometrische Polynome, Regressionssplines, Transformationen.

8 Variablenselektion

9 Das allgemeine lineare Modell: Gewichtete KQ-Methode, Autokorrelierte und heteroskedastische Störterme

10 Das logistische Regressionsmodell

11 Das gemischte lineare Regressionsmodell („Linear mixed Model“)

Behandlung von metrischen Einflussgrößen I

1. Einfach linear:

$$E(Y) = \beta_0 + \beta_1 x$$

2. Transformiert:

$$E(Y) = \beta_0 + \beta_1 T(x)$$

Beachte: Andere Interpretation von β_1 z.B.:

Logarithmisch:	$T(x)$	$=$	$\ln(x)$
Logarithmisch mit Nullpunkt-Erhaltung:	$T(x)$	$=$	$\ln(1 + x)$
Exponentiell mit bekanntem c:	$T(x)$	$=$	x^c

Behandlung von metrischen Einflussgrößen II

3. Als Polynom:

   $E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots \beta_k x^k$

   Problem: Bestimmung von k

   Mögliche Lösung: Tests auf $\beta_l = \ldots = \beta_k = 0$

   Verwende Sequentielle Quadratsummen

Behandlung von metrischen Einflussgrößen III

4. Stückweise konstante Funktion

$$E(Y) = \begin{cases} \beta_0 & \text{für } x \leq x_0 \\ \beta_1 & \text{für } x_0 < x < x_1 \\ \vdots & \\ \beta_h & \text{für } x > x_{h-1} \end{cases}$$

Dies entspricht der Kategorisierung der x-Variablen.

Bei x-Variablen, die nur wenige Werte haben, kann das Modell zur Überprüfung der Linearität benutzt werden.

Behandlung von metrischen Einflussgrößen IV

5. Stückweise linear

   $$E(Y) = \beta_0 + \beta_1 x + \beta_2 (x - g_1)_+ + \beta_3 (x - g_2)_+ + \ldots \beta_h (x - g_k)_+$$

   mit bekannten Bruchpunkten (Knoten) g_k und $t_+ = max(t, 0)$.

6. Regressionsspline

   $$E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - g_1)_+^3 + \beta_5 (x - g_2)_+^3$$

   Polynom 3. Grades 2 x stetig differenzierbar, da x^3 2 x stetig differenzierbar in 0. Bekannte Knoten g_k.

Behandlung von metrischen Einflussgrößen V

7. Fraktionale Polynome

   Beispiel: Grad (-2,2)

   $$E(Y) = \beta_0 + \beta_1 x^{-2} + \beta_2 x^{-1} + \beta_3 x^{-0.5} + \beta_4 \ln(x) + \beta_5 x^{0.5} + \beta_6 x + \beta_7 x^2$$

   Flexible Darstellung mit Strategien zur Selektion (Sauerbrei et al., Uni Freiburg)

Behandlung von metrischen Einflussgrößen VI

8. Trigonometrische Polynome zur Modellierung von periodischen Termen (Saisonfigur)

   Beispiel:

   $$E(Y) = \beta_0 + \beta_1 \sin(\frac{2\pi}{T} \cdot x) + \beta_2 \cos(\frac{2\pi}{T} \cdot x) + \beta_3 \sin(\frac{2\pi}{T} \cdot 2x) + \beta_4 \cos(\frac{2\pi}{T} \cdot 2x)$$

   T: Periodenlänge, x: Zeit
   Alternative: Saison- Dummy (Indikator) Variablen
   Beachte:

   $$A_1 * cos(x) + A_2 * sin(x) = A_3 * sin(x + \phi)$$

Allgemeiner Ansatz mit Basisfunktionen

Bekannte Basisfunktionen: $B_1, B_2, B_3, \ldots$

$$E(Y) = \beta_0 + \beta_1 B_1(x) + \beta_2 B_2(x) + \beta_3 B_3(x)$$

Obige Ansätze haben obige Darstellung

Interaktionen bei metrischen Variablen

Zwei metrische Variablen x_1 und x_2:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2$$

Interpretation durch die Darsetellung

$$E(Y) = \beta_0 + \beta_1 x_1 + (\beta_2 + \beta_3 x_1) x_2$$

Basisfunktionsansatz auch mehrdimensional erweiterbar durch Basisfunktionen $B(x_1, x_2)$

Beispiel: Trendmodell für die Populationsgröße von Füchsen in Baden Württemberg

Gegeben:
So genannte Jagdstrecken Y = Anzahl der geschossenen Füchse als Indikator für die Populationsgröße

Modelle:

$$
\begin{aligned}
ln(Y) &= \beta_0 + \beta_1 t + \beta_2 t^2 \\
ln(Y) &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-70)_+^3 + \beta_5 (t-85)_+^3 \\
etc.
\end{aligned}
$$

Versuchen Sie eine Modellierung von ln (Hase) !!