Vorlesung: Lineare Modelle

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2014

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN Institut für Statistik STABLAB Statistisches Beratungs Labor

Behandlung von metrischen Einflussgrößen I

1. Einfach linear:
   $y = \beta_0 + \beta_1 x + \varepsilon$

2. Transformiert:
   $y = \beta_0 + \beta_1 T(x) + \varepsilon$

   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:
   $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots \beta_k x^k$

   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

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

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

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

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

6. Regressionsspline

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

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. Trigonometrische Polynome zur Modellierung von periodischen Termen (Saisonfigur)

Beispiel:

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

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

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

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) !!