



Vorlesung
Wahrscheinlichkeit und Regression
WS 2005/06

Kapitel 11
Bedingte nichtlineare Regression

Prof. Dr. Rolf Steyer

**Lehrstuhl für Methodenlehre
und Evaluationsforschung**



Bedingte nichtlineare Regression

Worum geht es in dieser Sitzung?

Beispiel: Das Verhältnismodell für geometrisch-optische Täuschungen III

Bedingte lineare Quasi-Regression

Eigenschaften des Residuums

Parametrisierungen der bedingten linearen Quasi-Regression

Parametrisierungen der bedingten nichtlinearen Regression

Beispiel: Das Verhältnismodell für geometrisch-optische Täuschungen IV

Prüfung der bedingten Linearität

Logistische Regression



Beispiel: Das Verhältnismodell für geometrisch-optische Täuschungen III

Wie kann man das Verhältnismodell für geometrisch-optische Täuschungen, das ja die Linearität der Regression des logarithmierten Urteils ($\ln Y$) auf den logarithmierten Serienreiz ($\ln X$) bei konstantem Kontext-Serienreiz-Verhältnis $Z = z$ behauptet, empirisch überprüfen? Gilt also

$$E(\ln Y \mid \ln X, Z) = g_0(Z) + g_1(Z) \cdot \ln X ?$$

Bevor wir dieses praktische Vorhaben durchführen, sollen aber die dazu notwendigen begrifflichen Voraussetzungen, insbesondere also die Begriffe der bedingten linearen Quasi-Regression und der bedingten nichtlinearen Regression, eingeführt werden.



Bedingte lineare Quasi-Regression: Definition

Definition 1. Seien X , Y und Z Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum und $g_0(Z)$, $g_1(Z)$ zwei numerische Funktionen von Z . Die Funktion

$$Q(Y | X, Z) = g_0(Z) + g_1(Z) \cdot X$$

heißt die *bzgl. Z bedingte lineare Quasi-Regression* von Y auf X , wenn für jeden Wert z von Z gilt, dass die Funktion

$$LS_z[f_0(z), f_1(z)] = E_{Z=z}[(Y - [f_0(z) + f_1(z) \cdot X])^2]$$

für $f_0(z) = g_0(z)$ und $f_1(z) = g_1(z)$ ihr Minimum hat, wobei $g_0(z)$ und $g_1(z)$ die Werte von $g_0(Z)$ bzw. $g_1(Z)$ sind.



Eigenschaften des Residuums

Das Residuum

$$v := Y - g_0(Z) + g_1(Z) \cdot X$$

(sprich: v) hat also für jeden Wert z von Z die Eigenschaften

$$E_{Z=z}(v) = E(v | Z = z) = 0,$$

$$\text{Cov}_{Z=z}(v, X) = \text{Cov}(v, X | Z = z) = 0,$$

$$E(v) = 0, \text{ und } \text{Cov}(v, X) = 0,$$

Aber im Allgemeinen gilt

$$E_{Z=z}(\varepsilon | X) = 0,$$

nicht.



Parametrisierungen der bedingten linearen Quasi-Regression

Alle im letzten Kapitel besprochenen Möglichkeiten, die Funktionen $g_0(Z)$ und $g_1(Z)$ zu parametrisieren, kann man in gleicher Weise auch für die Funktionen $g_0(Z)$ und $g_1(Z)$ der bedingten linearen Quasi-Regression vornehmen.

Ebenfalls identisch sind die Formeln für den *Determinationskoeffizienten der bedingten linearen Quasi-Regression*

$$Q_{Y|X,Z}^2 = \frac{\text{Var}[g_0(Z)] + \text{Var}[g_1(Z) \cdot X] + 2 \cdot \text{Cov}[g_0(Z), g_1(Z) \cdot X]}{\text{Var}(Y)}.$$



Bedingte nichtlineare Regression I

Wir verwenden wieder Indikatorvariablen um die bedingte nichtlineare Regression zu parametrisieren (*Zellenmittelwertmodell*). Dabei wird für jede Wertekombination von X und Z eine eigene Indikatorvariable gebildet,

$$I_{ij} = \begin{cases} 1, & \text{falls } X = x_i \text{ und } Z = z_j \\ 0, & \text{andernfalls} \end{cases}, \quad i = 1, \dots, n, \quad j = 1, \dots, k.$$

Die folgende Parametrisierung als Zellenmittelwertmodell liefert nun eine saturierte Parametrisierung:

$$E(Y | X, Z) = \mu_{11} \cdot I_{11} + \mu_{12} \cdot I_{12} + \dots + \mu_{ij} \cdot I_{ij} + \dots + \mu_{nk} \cdot I_{nk}.$$



Bedingte nichtlineare Regression II

Die Verwendung der Symbole μ_{ij} als Koeffizienten der I_{ij} ist auch hier mit Bedacht gewählt, da es sich bei diesen Parametern tatsächlich um die bedingten Erwartungswerte handelt, d. h.

$$E(Y \mid X = x_1, Z = z_1) = \mu_{11} \cdot 1 + \mu_{21} \cdot 0 + \dots + \mu_{nk} \cdot 0 = \mu_{11},$$

$$E(Y \mid X = x_2, Z = z_1) = \mu_{11} \cdot 0 + \mu_{21} \cdot 1 + \dots + \mu_{nk} \cdot 0 = \mu_{21},$$

$$\vdots$$

$$E(Y \mid X = x_n, Z = z_k) = \mu_{11} \cdot 0 + \mu_{21} \cdot 0 + \dots + \mu_{nk} \cdot 1 = \mu_{nk}.$$



Beispiel: Das Verhältnismodell für geometrisch-optische Täuschungen IV

Die wesentliche Aussage des Verhältnismodells geometrisch-optischer Täuschungen ist, dass die logarithmierte Urteilslinie ($\ln Y$) bei konstantem Kontext-Serienreizverhältnis $Z = z$ *linear regressiv* von der logarithmierten Reizlinie ($\ln X$) abhängt. In der nun eingeführten Terminologie heißt das, dass die bzgl. Z bedingte lineare Quasi-Regression

$$Q(\ln Y \mid \ln X, Z) = \beta_0 + \beta_1 I_{1/2} + \beta_2 I_{1/1} + \gamma_0 \ln X + \gamma_1 I_{1/2} \ln X + \gamma_2 I_{1/1} \ln X$$

der logarithmierten Urteilslinie auf die logarithmierte Reizlinie zugleich auch die echte bzgl. Z bedingte lineare Regression von $\ln Y$ auf $\ln X$ ist



Prüfung der bedingten Linearität

Man berechnet mit einem Programm zur multiplen linearen Regression zunächst eine bzgl. Z bedingte lineare Quasi-Regression und den zugehörigen Quasi-Determinationskoeffizienten $Q_{Y|X,Z}^2$. Danach kann man eine saturierte Parametrisierung für die Regression $E(Y | X, Z)$ und den Determinationskoeffizienten $R_{Y|X,Z}^2$ berechnen. Die Prüfung der bedingten Linearität der Regression geschieht nun über den Vergleich von $Q_{Y|X,Z}^2$ und $R_{Y|X,Z}^2$ mittels folgender Prüfstatistik:

$$F = \frac{(\hat{R}_{Y|X,Z}^2 - \hat{Q}_{Y|X,Z}^2)/(n-p)}{(1 - \hat{R}_{Y|X,Z}^2)/(N-n)}.$$

Dabei sind n die Anzahl der Parameter im saturierten Modell, p die Anzahl der Parameter im eingeschränkten bzgl. Z bedingten linearen Modell und N die Stichprobengröße.



Logistische Regression

Logistische lineare Parametrisierung

$$P(Y = 1 \mid X, Z) = \frac{\exp[g_0(Z) + g_1(Z) \cdot X]}{1 + \exp[g_0(Z) + g_1(Z) \cdot X]}$$

Logistische polynomiale Parametrisierung

$$P(Y = 1 \mid X, Z) = \frac{\exp[g_0(Z) + g_1(Z) \cdot X + g_2(Z) \cdot X^2 + \dots]}{1 + \exp[g_0(Z) + g_1(Z) \cdot X + g_2(Z) \cdot X^2 + \dots]}$$

Saturierte Parametrisierung

$$P(Y = 1 \mid X, Z) = \frac{\exp(\lambda_{11}I_{11} + \lambda_{21}I_{21} + \dots + \lambda_{nk}I_{nk})}{1 + \exp(\lambda_{11}I_{11} + \lambda_{21}I_{21} + \dots + \lambda_{nk}I_{nk})}$$



Bedingte logistisch lineare Regression: Abbildung

$$P(Y = 1 | X, Z)$$

