



Vorlesung
Wahrscheinlichkeit und Regression
WS 2005/06

Kapitel 8
Einfache nichtlineare Regression

Prof. Dr. Rolf Steyer

**Lehrstuhl für Methodenlehre
und Evaluationsforschung**



Einfache nichtlineare Regression

Worum geht es in dieser Sitzung?

Beispiel: Das Stevenssche Potenzgesetz III

Lineare Quasi-Regression

Beispiel: Das Stevenssche Potenzgesetz IV

Einfache nichtlineare Regression

Parametrisierung als polynomiales Regressionsmodell

Parametrisierung als Zellenmittelwertemodell

Prüfung der Linearität einer Regression

Logistische Regression



Müller-Lyer- und Baldwin-Figur

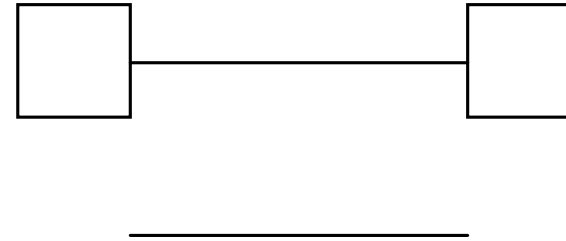
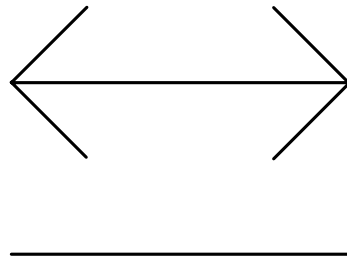


Abbildung 1. Müller-Lyer-Figur (links) und Baldwin-Figur (rechts).
Unterhalb der Figuren sind zum direkten Vergleich jeweils Linien gleicher Länge ohne die entsprechenden Kontextreize abgebildet.



Das Stevenssche Potenzgesetz III

- Wie kann das Stevenssche Potenzgesetz empirisch überprüft werden, d. h. wie kann getestet werden ob die Abhängigkeit tatsächlich linear regressiv ist?
- Wie hängt die Länge der Urteilslinie von der Länge der Reizlinie ab?

Bevor wir diese Anwendung weiter verfolgen, werden in den nächsten Abschnitten weitere regressionstheoretische Konzepte eingeführt: die lineare Quasi-Regression und die einfache nichtlineare Regression.



Lineare Quasi-Regression: Definition

Definition 8.1

Seien X und Y numerische Zufallsvariablen mit endlichen Erwartungswerten und Varianzen auf dem gleichen Wahrscheinlichkeitsraum. Als *lineare Quasi-Regression* wird diejenige lineare Funktion

$Q(Y | X) := \alpha_0 + \alpha_1 \cdot X$ von X bezeichnet, für deren Residuum

$$\nu := Y - (\alpha_0 + \alpha_1 \cdot X), \quad \alpha_0, \alpha_1 \in \mathbb{R},$$

die folgenden beiden Gleichungen gelten:

$$E(\nu) = 0,$$

$$\text{Cov}(\nu, X) = 0.$$



Lineare Quasi-Regression : Bemerkungen

Die Variable Y wird auch hier als Summe einer linearen Funktion der Variablen X und der Fehlervariablen ν dargestellt, wie man aus der Umstellung der Gleichung für das Residuum $\nu := Y - (\alpha_0 + \alpha_1 \cdot X)$ ersehen kann:

$$Y = \alpha_0 + \alpha_1 \cdot X + \nu.$$



Lineare Quasi-Regression : Eigenschaften des Residuums

Definitionsgemäß gelten zwar immer

$$E(\nu) = 0$$

und

$$\text{Cov}(\nu, X) = 0.$$

Jedoch gilt

$$E(\nu | X) = 0$$

nicht immer.



Lineare Quasi-Regression: Alternative Definition

Definition 8.2.

Unter den gleichen Voraussetzungen wie in Definition 8.1 kann die *lineare Quasi-Regression von Y auf X* auch als diejenige lineare Funktion von X definiert werden, welche die folgende Funktion der reellen Zahlen a_0 und a_1 minimiert:

$$LS(a_0, a_1) = E[(Y - (a_0 + a_1 \cdot X))^2]. \quad \text{Kleinst-Quadrat-Kriterium}$$

Diejenigen Zahlen a_0 und a_1 , für welche die Funktion $LS(a_0, a_1)$ ein Minimum annimmt, seien mit α_0 und α_1 bezeichnet. Die lineare Quasi-Regression ist dann definiert durch:

$$Q(Y | X) := \alpha_0 + \alpha_1 \cdot X.$$



Lineare Quasi-Regression: Identifikation der Koeffizienten

Aus (beiden Arten) der Definition der Quasi-Regression kann man übrigens, unter Verwendung einiger elementarer Rechenregeln für Erwartungswerte und Kovarianzen, die beiden folgenden Gleichungen ableiten:

$$\alpha_0 = E(Y) - \alpha_1 \cdot E(X),$$

$$\alpha_1 = \text{Cov}(X, Y) / \text{Var}(X).$$

Diese Formeln sind identisch mit denen der linearen Regression. Auch die Formeln für den Determinationskoeffizienten sind identisch:

$$Q_{Y|X}^2 := \frac{\alpha_1^2 \text{Var}(X)}{\text{Var}(Y)} = \text{Kor}(X, Y)^2.$$

Dieser ist nur dann auch ein echter Determinationskoeffizient, wenn $E(Y | X)$ tatsächlich eine *lineare* Funktion von X ist. In diesem Fall sind dann auch die Regression $E(Y | X)$ und die lineare Quasi-Regression $Q(Y | X)$ identisch.



Lineare Quasi Regression: Illustration

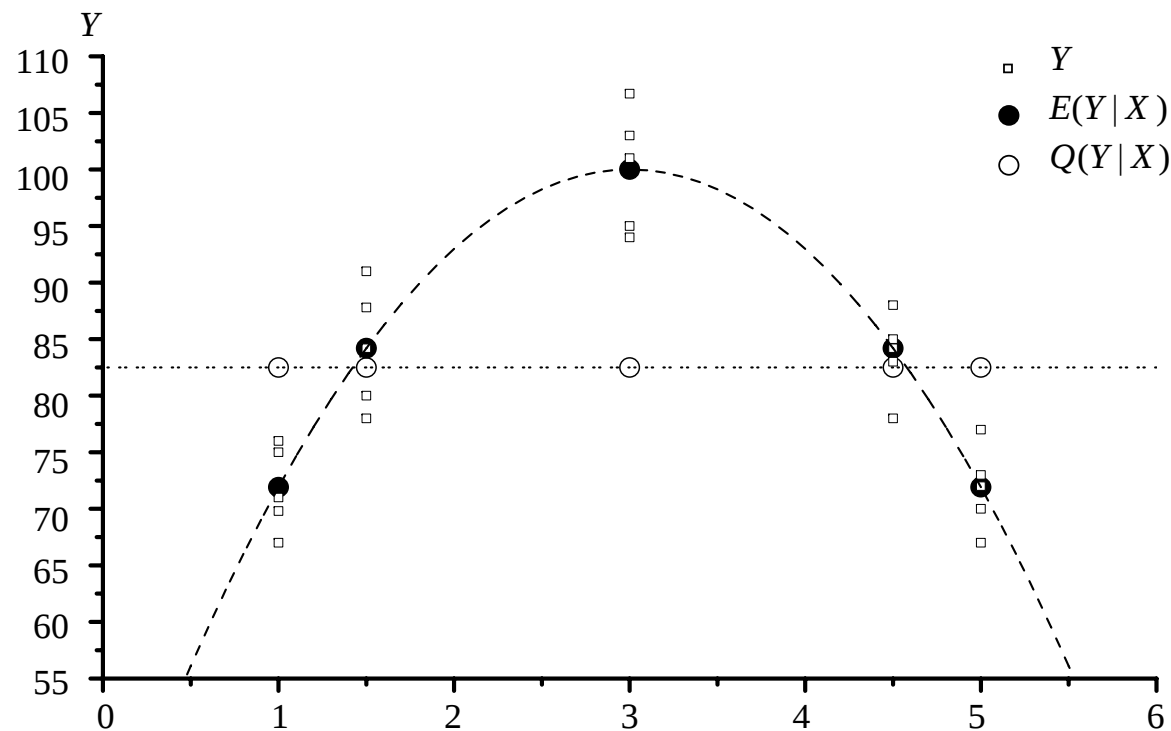


Abbildung 8.2. Regression und lineare Quasi-Regression bei parabolischer regressiver Abhängigkeit einer Variablen Y von einer Variablen X . Die (echte) Regression ist die Parabel, die bedingten Erwartungswerte sind die ausgefüllten Punkte. Die lineare Quasi-Regression wird durch die parallel zur X -Achse verlaufende Gerade mit den nicht ausgefüllten Punkten gekennzeichnet. Die Quadrate sind die Wertepaare von (X, Y) .



Einfache nichtlineare Regression: Parametrisierung als polynomiales Regressionsmodell

In Abbildung 8.2 haben wir oben bereits einen Fall kennen gelernt, in dem eine einfache Regression $E(Y | X)$ keine lineare Funktion, sondern eine *quadratische Funktion*

$$E(Y | X) = \alpha_0 + \alpha_1 \cdot X + \alpha_2 \cdot X^2,$$

von X ist. Prinzipiell gibt es natürlich auch Fälle in denen die Regression auch eine kubische Funktion von X oder gar ein Polynom noch höherer Ordnung ist:

$$E(Y | X) = \alpha_0 + \alpha_1 \cdot X + \alpha_2 \cdot X^2 + \dots + \alpha_{n-1} \cdot X^{n-1}.$$

Dabei muss der Regressor X allerdings mindestens n verschiedene Werte annehmen können, denn durch n Punkte kann man immer ein Polynom $(n - 1)$ -ter Ordnung legen.



Einfache nichtlineare Regression: Parametrisierung als Zellenmittelwertemodell I

Wenn der Regressor X einer Regression $E(Y | X)$ nur n verschiedene Werte $x_1, \dots, x_n$ annimmt, können wir die Regression $E(Y | X)$ auch als Zellenmittelwertemodell parametrisieren. Dazu werden zunächst die folgenden n Indikatorvariablen eingeführt:

$$I_i = \begin{cases} 1, & \text{falls } X = x_i \\ 0, & \text{andernfalls} \end{cases}, \quad i = 1, \dots, n.$$

Die Variablen I_i , die oft auch als *Dummy-Variablen* bezeichnet werden, zeigen jeweils mit ihrem Wert 1 an, ob der Regressor X den Wert x_i annimmt. Man beachte, dass diese Indikatorvariablen I_i Funktionen von X sind und dass die $I_1, \dots, I_n$ zusammen exakt die gleiche Information beinhalten, wie der Regressor X . Es gilt daher $E(Y | X) = E(Y | I_1, \dots, I_n)$, so dass wir uns beider Notationen für diese Regression bedienen können.



Einfache nichtlineare Regression: Parametrisierung als Zellenmittelwertemodell II

Die folgende Parametrisierung als Zellenmittelwertemodell liefert nun ein saturiertes Modell:

$$E(Y | X) = \mu_1 \cdot I_1 + \dots + \mu_n \cdot I_n.$$

Die Verwendung der Symbole $\mu_1, \dots, \mu_n$ als Koeffizienten der I_i ist mit Bedacht gewählt, da es sich bei diesen Parametern tatsächlich um die bedingten Erwartungswerte handelt, d. h.

$$\mu_i = E(Y | X = x_i), \quad \text{for } i = 1, \dots, n.$$

Dies erkennt man sofort, wenn man die bedingten Erwartungswerte $E(Y | X = x_i)$ für alle n Werte von X aus der Modellgleichung ausrechnet:

$$E(Y | X = x_1) = \mu_1 \cdot 1 + \mu_2 \cdot 0 + \dots + \mu_n \cdot 0 = \mu_1,$$

$$E(Y | X = x_2) = \mu_1 \cdot 0 + \mu_2 \cdot 1 + \mu_3 \cdot 0 + \dots + \mu_n \cdot 0 = \mu_2,$$

$$\vdots$$

$$E(Y | X = x_n) = \mu_1 \cdot 0 + \mu_2 \cdot 0 + \dots + \mu_n \cdot 1 = \mu_n.$$



Prüfung der Linearität einer Regression I

Man kann die Linearität einer Regression $E(Y | X)$ dadurch prüfen, dass man mit einem Programm zur multiplen linearen Regression zunächst eine lineare Quasi-Regression und den zugehörigen Quasi-Determinationskoeffizienten $Q_{Y|X}^2 = \alpha_1^2 \text{Var}(X) / \text{Var}(Y)$ berechnet. Danach kann man eine saturierte Parametrisierung für die Regression $E(Y | X)$ und den Determinationskoeffizienten $R_{Y|X}^2$ berechnen. Die Prüfung der Linearität der Regression geschieht nun über den Vergleich von $Q_{Y|X}^2$ und $R_{Y|X}^2$. Ist die Regression $E(Y | X)$ wirklich linear, dann gilt:

$$H_0: R_{Y|X}^2 - Q_{Y|X}^2 = 0$$

und die lineare Quasi-Regression ist auch die echte Regression.



Prüfung der Linearität einer Regression II

Nun können Sie einen Signifikanztest für die Nullhypothese

$$H_0: R_{Y|X}^2 - Q_{Y|X}^2 = 0$$

durchführen, dass die Regression linear ist. Diese geschieht über die Teststatistik:

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

Dabei sind n die Anzahl der Parameter in der saturierten Parametrisierung (hier: die 5 verschiedenen Reizlinien), 2 die Anzahl der Parameter in der eingeschränkten, linearen Quasi-Regression und N die Stichprobengröße (hier: die Gesamtzahl der Urteilslinien). Diese Teststatistik ist mit $n - 2$ Zähler- und $N - n$ Nennerfreiheitsgraden F -verteilt.



Logistische Regression I

Ist Y ein dichotomer Regressand mit Werten 0 und 1, dann ist die Regression $E(Y | X)$ zugleich auch die bedingte Wahrscheinlichkeitsfunktion $P(Y = 1 | X)$. Ist X nicht auch ein dichotome Variable, sondern kontinuierlich, dann kann $E(Y | X)$ nicht als lineare Regression parametrisiert werden, da eine Gerade mit einer Steigung ungleich 0 inkonsistent mit dem Wertebereich $[0, 1]$ einer (bedingten) Wahrscheinlichkeit ist. In diesem Fall wäre eine *lineare* Regression logisch widersprüchlich und man benutzt die logistisch lineare Parametrisierung

$$P(Y = 1 | X) = \frac{\exp(\gamma_0 + \gamma_1 X)}{1 + \exp(\gamma_0 + \gamma_1 X)}$$

der Regression $E(Y | X)$ bzw. $P(Y = 1 | X)$, die prinzipiell auch dann gelten kann, wenn X kontinuierlich ist.



Logistische Regression II

Im linearen Fall kann der *logit* wie folgt geschrieben werden:

$$\ln \frac{P(Y = 1 | X)}{P(Y = 0 | X)} = \gamma_0 + \gamma_1 X$$



Logistische Regression III

Bei einem kontinuierlichen Regressor X und einem dichotomen Regressanden Y muss die Regression $E(Y | X)$ bzw. $P(Y = 1 | X)$ natürlich nicht unbedingt durch eine logistisch lineare Parametrisierung darstellbar sein. Der allgemeinere Fall wäre der einer logistisch polynomialen Parametrisierung:

$$P(Y = 1 | X) = \frac{\exp(\gamma_0 + \gamma_1 X + \gamma_2 X^2 + \dots + \gamma_{n-1} X^{n-1})}{1 + \exp(\gamma_0 + \gamma_1 X + \gamma_2 X^2 + \dots + \gamma_{n-1} X^{n-1})}.$$



Logistische Regression IV

Aus den gleichen Gründen, die wir in den Abschnitten zuvor schon genannt haben, sollte man auch die Anwendung einer Parametrisierung erwägen, die analog zu der des Zellenmittelwertmodells ist. Das heißt, man verwendet wieder die Indikatorvariablen für die Werte des Regressors und erhält mit

$$P(Y = 1 \mid X) = \frac{\exp(\lambda_1 I_1 + \lambda_2 I_2 + \dots + \lambda_n I_n)}{1 + \exp(\lambda_1 I_1 + \lambda_2 I_2 + \dots + \lambda_n I_n)}$$

eine saturierte Parametrisierung, falls der Regressor X nur n verschiedene Werte $x_1, \dots, x_n$ annehmen kann.