

1 Nichtlineare Regression

Unterlagen von Andreas Ruckstuhl, Zürcher Hochschule Winterthur,
überarbeitet von Werner Stahel, ETHZ, Jan. 2000 bis 2004

1.1 Das Modell

- a Regression studiert den Zusammenhang zwischen einer **Zielgrösse** Y und einer oder mehreren **Ausgangs-Variablen** $x^{(j)}$. Das allgemeine Modell lautet

$$Y_i = h\langle x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}; \theta_1, \theta_2, \dots, \theta_p \rangle + E_i.$$

Dabei ist h eine geeignete Funktion, die von den Ausgangs-Variablen und von Parametern abhängt, die wir zu Vektoren zusammenfassen wollen, $\underline{x} = [x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}]$ und $\underline{\theta} = [\theta_1, \theta_2, \dots, \theta_p]$.

- b In der (multiplen) **linearen Regression** werden Funktionen h betrachtet, die linear sind in den Parametern θ_j ,

$$h\langle x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}; \theta_1, \theta_2, \dots, \theta_p \rangle = \theta_1 \tilde{x}_i^{(1)} + \theta_2 \tilde{x}_i^{(2)} + \dots + \theta_p \tilde{x}_i^{(p)},$$

wobei die $\tilde{x}^{(j)}$ beliebige Funktionen der ursprünglichen Ausgangs-Variablen $x^{(j)}$ sein können. (Die Parameter werden dort üblicherweise mit β_j statt θ_j bezeichnet.)

- c In der **nichtlinearen Regression** werden Funktionen h untersucht, die sich nicht als lineare Funktionen in den Parametern schreiben lassen. Oft wird eine solche Funktion aus der Theorie abgeleitet. Es bestehen im Prinzip unbeschränkte Möglichkeiten, den deterministischen Teil des Modells anzusetzen. Wie wir sehen werden, wird diese Flexibilität erkauft durch einen grösseren Aufwand, statistische Aussagen zu gewinnen.

Die Voraussetzungen für den zufälligen Teil, der ja nur aus den **zufälligen Abweichungen** oder Fehlern E_i besteht, sind die gleichen wie bei der linearen Regression:

$$E_i \sim \mathcal{N}\langle 0, \sigma^2 \rangle, \quad \text{unabhängig.}$$

- d \triangleright **Beispiel Puromycin**. Die Geschwindigkeit, mit der eine enzymatische Reaktion abläuft, hängt von der Konzentration eines Substrates ab. Gemäss den Angaben von Bates and Watts (1988) wurde untersucht, wie eine Behandlung des Enzyms mit einer weiteren Substanz namens Puromycin diese Reaktionsgeschwindigkeit beeinflusst. Als Zielvariable wurde die Anfangsgeschwindigkeit der Reaktion gewählt, welche über Radioaktivität gemessen wird. (Die Einheit der Zielvariablen ist Anzahl/min²; die Anzahl Anschläge in einem Geigerzähler pro Zeiteinheit misst ja die Quantität der vorhandenen Substanz, und die Reaktionsgeschwindigkeit ist proportional zu deren Veränderung pro Zeiteinheit)

Der Zusammenhang der Zielgrösse mit der Substrat-Konzentration x (in ppm) wird beschrieben durch die *Michaelis-Menten-Funktion*

$$h\langle x; \underline{\theta} \rangle = \frac{\theta_1 x}{\theta_2 + x}.$$

Für unendlich grosse Substratkonzentration ($x \rightarrow \infty$) ergibt sich die „asymptotische“ Geschwindigkeit θ_1 . Es wurde vermutet, dass diese Grösse durch das Hinzufügen von Puromycin beeinflusst wird. Das Experiment wurde deshalb einmal mit dem mit Puromycin behandelten Enzym und einmal mit dem unbehandelten Enzym durchgeführt. Figur 1.1.d zeigt das Resultat. In diesem Abschnitt werden die Daten des behandelten Enzyms benutzt.

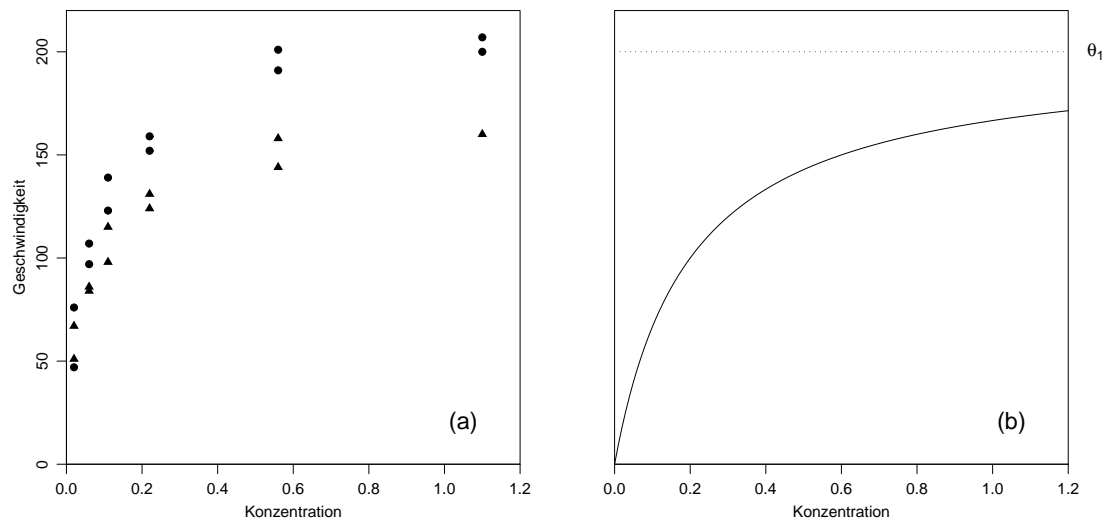


Abbildung 1.1.d: **Beispiel Puromycin.** (a) Daten. • behandeltes Enzym; \triangle unbehandelt) und (b) typischer Verlauf der Regressionsfunktion

◁

- e ▷ **Beispiel Sauerstoffverbrauch.** Um den biochemischen Sauerstoffverbrauch zu bestimmen, werden Flusswasserproben mit gelösten organischen Nährstoffen, mit anorganischen Materialien und mit gelöstem Sauerstoff angereichert und in verschiedene Flaschen abgefüllt (Marske, 1967, siehe Bates and Watts, 1988). Jede Flasche wird dann mit einer Mischkultur von Mikroorganismen geimpft und verschlossen in eine Klimakammer mit konstanter Temperatur gestellt. Die Flaschen werden periodisch geöffnet und nach gelöstem Sauerstoffgehalt analysiert. Daraus wird der biochemische Sauerstoffverbrauch [mg/l] berechnet. Das verwendete Modell, das den kumulierten biochemischen Sauerstoffverbrauch Y mit der Inkubationszeit x in Verbindung bringt, basiert auf exponentiellem Abfall der Zuwächse, was zu

$$h(x, \underline{\theta}) = \theta_1 \left(1 - e^{-\theta_2 x}\right)$$

führt. Figur 1.1.e zeigt die Daten und die zu verwendende Regressionsfunktion.

◁

- f ▷ **Beispiel aus der Membrantrenn-Technologie** (Rapold-Nydegger, 1994). Das Verhältnis von protonierten zu deprotonierten Carboxylgruppen in den Poren von Cellulosemembranen ist vom pH-Wert x der Aussenlösung abhängig. Die Protonierung des Carboxylkohlenstoffatoms kann mit ^{13}C -NMR erfasst werden. Wir nehmen an, dass der Zusammenhang mit der erweiterten „Henderson-Hasselbach-Gleichung“ für Polyelektrolyte

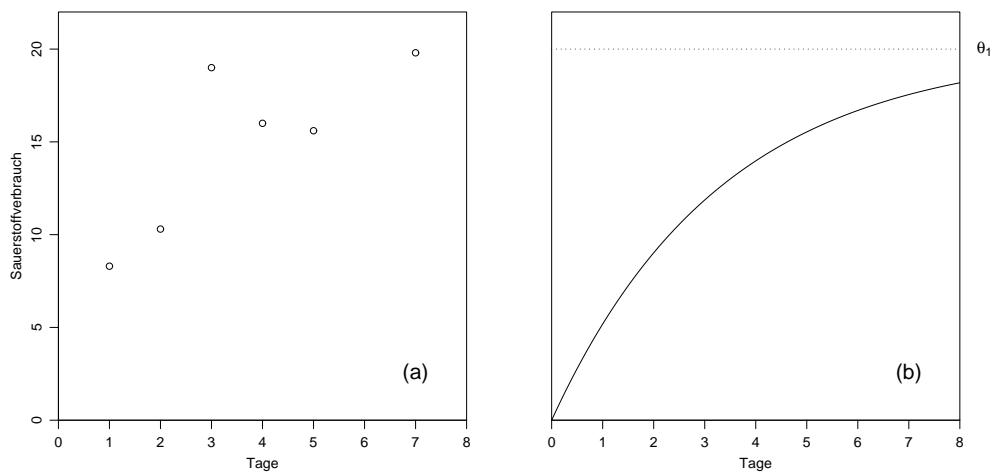


Abbildung 1.1.e: Beispiel Sauerstoffverbrauch. (a) Daten und (b) typischer Verlauf der Regressionsfunktion

beschrieben werden kann,

$$\log_{10} \left\langle \frac{\theta_1 - y}{y - \theta_2} \right\rangle = \theta_3 + \theta_4 x,$$

wobei die unbekannten Parameter θ_1, θ_2 und $\theta_3 > 0$ und $\theta_4 < 0$ sind. Auflösung nach y führt zu

$$h(x; \underline{\theta}) = \frac{\theta_1 + \theta_2 10^{\theta_3 + \theta_4 x}}{1 + 10^{\theta_3 + \theta_4 x}}.$$

Die Regressionsfunktion $h(x; \underline{\theta})$ für ein sinnvoll gewähltes $\underline{\theta}$ ist in Figur 1.1.f neben den Daten dargestellt.

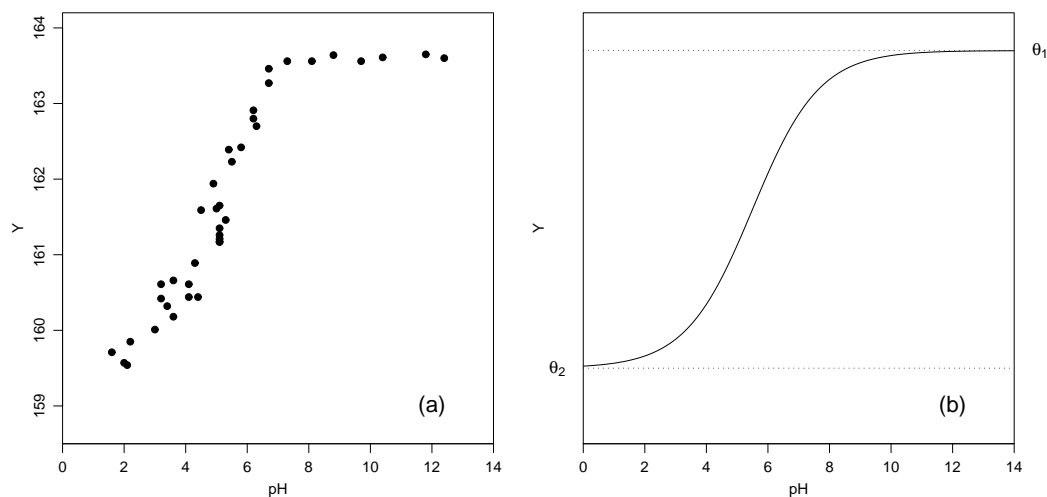


Abbildung 1.1.f: Beispiel Membrantrenn-Technologie. (a) Daten und (b) ein typischer Verlauf der Regressionsfunktion.

◁

g Einige weitere **Beispiele für nichtlineare Regressionsfunktionen:**

- Hill-Modell (Enzymkinetik):

$$h\langle x; \underline{\theta} \rangle = \theta_1 x^{\theta_3} / (\theta_2 + x^{\theta_3}) .$$

Für $\theta_3 = 1$ ist dies auch unter dem Namen Michaelis-Menten-Modell bekannt (1.1.d).

- Die Mitscherlich-Funktion wird in der Wachstumsanalyse verwendet,

$$h\langle x; \underline{\theta} \rangle = \theta_1 + \theta_2 \exp\langle \theta_3 x \rangle .$$

- Aus der Kinetik (Chemie) stammt die Funktion

$$h\langle x^{(1)}, x^{(2)}; \underline{\theta} \rangle = \exp \left\langle -\theta_1 x^{(1)} \exp \left\langle -\theta_2 / x^{(2)} \right\rangle \right\rangle .$$

- Die Produktions-Funktion von Cobbs und Douglas Cobbs-Douglas-Modell lautet

$$h \left\langle x^{(1)}, x^{(2)}; \underline{\theta} \right\rangle = \theta_1 \left(x^{(1)} \right)^{\theta_2} \left(x^{(2)} \right)^{\theta_3} .$$

Da die nützlichen nichtlinearen Regressionsfunktionen aus der Theorie des jeweiligen Anwendungsgebietes hergeleitet werden, ist eine allgemeine Übersicht von beschränktem Nutzen. Eine Zusammenstellung von Funktionen aus Publikationen findet man in Anhang 7 von Bates and Watts (1988).

- h Einige nichtlineare Regressionsfunktionen lassen sich durch Transformationen der Zielgrösse und der Ausgangs-Variablen **linearisieren**.
Beispielsweise lässt sich eine Potenzfunktion,

$$h\langle x; \underline{\theta} \rangle = \theta_1 x^{\theta_2}$$

zu einer (in den Parametern) linearen Funktion transformieren,

$$\ln\langle h\langle x; \underline{\theta} \rangle \rangle = \ln\langle \theta_1 \rangle + \theta_2 \ln\langle x \rangle = \beta_0 + \beta_1 \tilde{x} = \tilde{h}\langle \tilde{x}; \underline{\beta} \rangle ,$$

wobei $\beta_0 = \ln\langle \theta_1 \rangle$, $\beta_1 = \theta_2$ und $\tilde{x} = \ln\langle x \rangle$ ist. Die Regressionsfunktion h nennen wir **linearisierbar**, wenn sie durch Transformationen der Argumente und eine monotone Transformation des Resultats in eine in den Parametern lineare Funktion verwandelt werden kann.

Hier einige weitere linearisierbare Funktionen (siehe auch Daniel and Wood (1980)):

$$\begin{aligned} y = \theta_1 x / (\theta_2 + x) &\longleftrightarrow 1/y = 1/\theta_1 + \theta_2/\theta_1 \frac{1}{x} \\ y = \exp\langle -\theta_1 x^{(1)} \exp\langle -\theta_2 / x^{(2)} \rangle \rangle &\longleftrightarrow \ln\langle \ln\langle y \rangle \rangle = \ln\langle -\theta_1 \rangle + \ln\langle x^{(1)} \rangle - \theta_2 / x^{(2)} \\ y = \theta_1 \left(x^{(1)} \right)^{\theta_2} \left(x^{(2)} \right)^{\theta_3} &\longleftrightarrow \ln\langle y \rangle = \ln\langle \theta_1 \rangle + \theta_2 \ln\langle x^{(1)} \rangle + \theta_3 \ln\langle x^{(2)} \rangle . \end{aligned}$$

Das letzte ist das Cobbs-Douglas-Modell aus 1.1.g.

- i Eine lineare Regression mit der linearisierten Regressionsfunktion beruht im genannten Beispiel auf dem Modell

$$\ln\langle Y_i \rangle = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_i + E_i ,$$

wobei die Zufallsfehler E_i alle der gleichen Normalverteilung folgen. Transformieren wir dieses Modell zurück, so erhalten wir

$$Y_i = \theta_1 x^{\theta_2} \tilde{E}_i$$

mit $\tilde{E}_i = \exp\langle E_i \rangle$. Die Fehler \tilde{E}_i , $i = 1, \dots, n$ wirken nun multiplikativ und sind lognormal verteilt! Die Annahmen über die Zufallsabweichungen verändern sich also recht drastisch gegenüber einem Modell, das direkt auf h basiert,

$$Y_i = \theta_1 x^{\theta_2} + E_i^*$$

mit Zufallsabweichungen E_i^* , die wie üblich einer einzigen Normalverteilung folgen.

Eine Linearisierung der Regressionsfunktion ist deshalb nur dann angebracht, wenn damit auch die Annahmen über die Zufallsabweichungen besser befriedigt werden können – im Beispiel, falls tatsächlich die Fehler eher multiplikativ als additiv wirken und lognormal statt normal verteilt sind. Diese Annahmen müssen mit Residuen-Analyse geprüft werden.

- j Diese Überlegung kann umgekehrt auch dazu führen, dass man aus einem linearen Regressionsmodell ein nichtlineares macht.

▷ **Beispiel Schadstoffe im Tunnel.** (Text aus der Einführung in die lineare Regression.) Die Schadstoffe, die vom motorisierten Verkehr ausgestossen werden, bilden einen wesentlichen Bestandteil der Belastung der Luft. Um die Grösse dieser Belastung zu schätzen, werden für die Fahrzeuge so genannte **Emissionsfaktoren** bestimmt. Dies kann einerseits auf dem Prüfstand geschehen, auf dem die Strasse mit Rollen simuliert wird. Der Widerstand der Rollen wird dabei variiert, so dass ein typischer „Fahrzyklus“ durchgespielt werden kann. – Andererseits eignen sich Strassentunnels mit Ein-Richtungs-Verkehr für Messungen unter realen Bedingungen. Misst man Schadstoff-Konzentrationen am Anfang und am Schluss des Tunnels und zählt, wie viele Fahrzeuge durch den Tunnel fahren, so kann man ebenfalls Emissionsfaktoren ausrechnen. Allerdings erhält man zunächst nur einen gemittelten Faktor für jeden gemessenen Schadstoff, und dieser lässt sich nicht ohne zusätzliche Erkenntnisse auf andere Strassenabschnitte übertragen. Wenn man die Anzahl der Fahrzeuge nach Fahrzeug-Kategorien aufteilen kann, dann kann man immerhin mit Regressionsrechnung zu einem Emissionsfaktor für jede Fahrzeug-Kategorie kommen.

Während einer Woche im September 1993 wurden in der Südröhre des Gubrist-Tunnels nördlich von Zürich solche Messungen durchgeführt. Die Schadstoff-Konzentrationen am Anfang und am Ende wurden gemessen und die Luftströmung erfasst. Daraus lässt sich die Schadstoff-Emission Y pro Kilometer für alle durchgefahrenen Fahrzeuge zusammen berechnen. Von einem Schlaufen-Detektor im Strassenbelag wurden die Fahrzeuge in zwei Kategorien gezählt: Auf Grund des Abstands von Vorder- und Hinterachse wurden die Lastwagen von den übrigen Fahrzeugen getrennt. Es bezeichne $x^{(1)}$ die Anzahl „Nicht-Lastwagen“ und $x^{(2)}$ die Anzahl Lastwagen. Die gesamten Emissionen in der Zeitperiode i setzen sich zusammen gemäss

$$Y_i = \theta_1 x_i^{(1)} + \theta_2 x_i^{(2)} + E_i ,$$

wobei θ_1 die durchschnittliche Emission pro Nicht-Lastwagen und θ_2 diejenige pro Lastwagen bedeutet – also die Grössen, an denen wir in der Studie primär interessiert sind.

Die „Zufallsfehler“ E_i entstehen durch Variationen in Bauart und Zustand der Fahrzeuge, durch zeitliche Abgrenzungs-Schwierigkeiten und durch Mess-ungenauigkeiten.

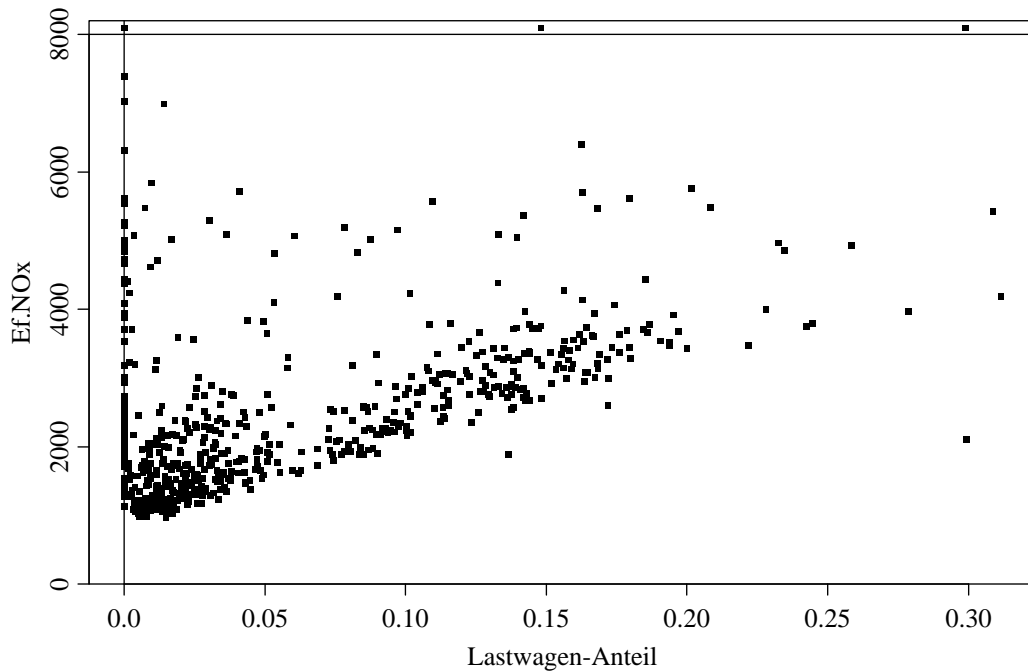


Abbildung 1.1.j: Emissionsfaktor für NO_x und Lastwagen-Anteil, gemittelt über jeweils 15 Minuten, im Beispiel der Schadstoffe im Tunnel. Drei extrem hohe Y -Werte sind im Bildrand dargestellt.

Die Formel lässt sich in eine üblichere und vielleicht noch einfachere Form bringen: Wir dividieren Y_i , $x_i^{(1)}$ und $x_i^{(2)}$ durch die gesamte Anzahl Fahrzeuge $x_i^{(1)} + x_i^{(2)}$ und erhalten $\tilde{Y}_i = \theta_1 \tilde{x}_i^{(1)} + \theta_2 \tilde{x}_i^{(2)} + \tilde{E}_i$, wobei \tilde{Y}_i der „mittlere Emissionsfaktor“ für die Zeitperiode i und $\tilde{x}_i^{(1)}$ und $\tilde{x}_i^{(2)}$ die Anteile der Nicht-Lastwagen und der Lastwagen bedeuten. Da $\tilde{x}_i^{(1)} = 1 - \tilde{x}_i^{(2)}$ ist, gilt

$$\tilde{Y}_i = \theta_1 + (\theta_2 - \theta_1) \tilde{x}_i^{(2)} + \tilde{E}_i.$$

◁

- k ▷ Die einfache und plausible Überlegung, wie die Schadstoffe zusammenkommen, hat also auf ein einfaches lineares Regressionsmodell geführt,

$$Y_i = \beta_0 + \beta_1 x_i + E_i$$

(mit $\beta_0 = \theta_1$, $\beta_1 = \theta_2 - \theta_1$ und $x_i = \tilde{x}_i^{(2)}$). In Figur 1.1.j zeigt sich als Tendenz in der Tat eine lineare Zunahme des mittleren Emissionsfaktors für NO_x mit zunehmendem Lastwagen-Anteil. Es wird aber auch klar, dass die Zufallsabweichungen eine schiefe Verteilung haben.

Bei solchen Verteilungen wurde angeraten, die Zielgröße zu transformieren, und dabei zuerst die Logarithmus-Transformation zu versuchen. Wenn wir dies aber in der üblichen Weise tun, wird die logarithmierte Zielgröße als lineare Funktion der gegebenen Ausgangs-Variablen angesetzt, also $\ln(Y_i) = \beta_0 + \beta_1 x_i + E_i$. Das widerspricht der Überlegung, die

zum Modell geführt hat. Da die Regressionsfunktion durch diese Überlegung festgelegt ist, müssen wir beide Seiten transformieren, $\tilde{h}\langle x; \tilde{\beta} \rangle = \ln\langle \beta_0 + \beta_1 x \rangle$ und erhalten als Modell

$$\ln\langle Y_i \rangle = \ln\langle \beta_0 + \beta_1 x_i \rangle + E_i .$$

◁

Hier haben wir also die lineare Regressionsfunktion in eine nichtlineare verwandelt, um einen geeigneten additiven Fehlerterm E_i zu erhalten.

- 1 Soweit die einführenden Beispiele.

Wir haben fast ausschliesslich von Regressionfunktionen gesprochen, die nur von einer Ausgangs-Variablen abhängen. Dies geschah vor allem, weil dann eine Grafik das Modell umfassend veranschaulichen kann. Die nun folgende Theorie funktioniert ebenso gut für Regressionfunktionen $h\langle \underline{x}; \underline{\theta} \rangle$, die von mehreren Ausgangs-Variablen $\underline{x} = [x^{(1)}, x^{(2)}, \dots, x^{(m)}]$ abhängen.

1.2 Methodik zur Schätzung der Parameter

- a Um Schätzungen für die Parameter $\underline{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T$ zu erhalten, wendet man, wie in der linearen Regression, das **Prinzip der Kleinsten Quadrate** an. Die Summe der quadrierten Abweichungen

$$S\langle \underline{\theta} \rangle := \sum_{i=1}^n (y_i - \eta_i\langle \underline{\theta} \rangle)^2 \quad \text{mit } \eta_i\langle \underline{\theta} \rangle := h\langle x_i; \underline{\theta} \rangle$$

soll also minimiert werden. Die Schreibweise, die $h\langle x_i; \underline{\theta} \rangle$ durch $\eta_i\langle \underline{\theta} \rangle$ ersetzt, ist sinnvoll, weil nach der Messung oder Beobachtung die Daten $[x_i, y_i]$ gegeben sind und nun die Parameter $\underline{\theta}$ zu bestimmen bleiben.

Leider lassen sich das Minimum der Quadratsumme und damit die Schätzungen nicht wie in der linearen Regression explizit angeben. **Iterative numerische Verfahren** helfen weiter. Die Grundidee, die hinter dem üblichen Algorithmus steckt, soll hier skizziert werden. Sie bildet auch die Basis für die einfachste Art, Tests und Vertrauensbereiche herzuleiten.

- b **Geometrische Veranschaulichung.** Die beobachteten Werte $\underline{Y} = [Y_1, Y_2, \dots, Y_n]^T$ legen einen Punkt im n -dimensionalen Raum fest. Das Gleiche gilt für die „Modellwerte“ $\underline{\eta}(\underline{\theta}) = [\eta_1\langle \underline{\theta} \rangle, \dots, \eta_n\langle \underline{\theta} \rangle]^T$ für gegebenes $\underline{\theta}$.

Achtung! Die übliche geometrische Veranschaulichung von Daten, die beispielsweise in der Multivariaten Statistik grundlegend ist, betrachtet die Beobachtungen, die durch m Variable $X^{(j)}$, $j = 1, 2, \dots, m$, festgelegt sind, als Punkte im m -dimensionalen Raum. Hier betrachten wir die Y - und η -Werte aller n Beobachtungen als Punkte im n -dimensionalen Raum.

Leider hört unsere Anschauung bei drei Dimensionen, also bei drei Beobachtungen auf. Versuchen wir es also für ein solches Minibeispiel.

- c ▷ Im **Beispiel des Sauerstoffverbrauchs** sind die Messungen für die Tage 1, 4 und 7 – also $\underline{x} = [1, 4, 7]^T$ gleich $\underline{Y} = [8.3, 16, 19.8]^T$. Die beiden Vektoren legen je einen Punkt im dreidimensionalen Raum fest.

Obwohl es für diese Daten wenig nützt, wollen wir zunächst die Anpassung einer Geraden, also eine **einfache lineare Regression**, betrachten. Für gegebene Parameter $\beta_0 = 5$ und $\beta_1 = 1$ können wir die Modellwerte $\eta_i(\underline{\beta}) = \beta_0 + \beta_1 x_i$ ausrechnen und den entsprechenden Vektor $\underline{\eta}(\underline{\beta}) = \beta_0 \underline{1} + \beta_1 \underline{x}$ ebenfalls als Punkt darstellen (Figur 1.2.c). Nun fragen wir, wo alle Punkte liegen, die durch Variation der Parameter erreicht werden können. Sie sind die möglichen Linearkombinationen (Summen von Vielfachen) der beiden Vektoren $\underline{1}$ und \underline{x} und bilden deshalb die Ebene die „durch $\underline{1}$ und \underline{x} aufgespannt“ wird. Figur 1.2.c hält das grafisch fest.

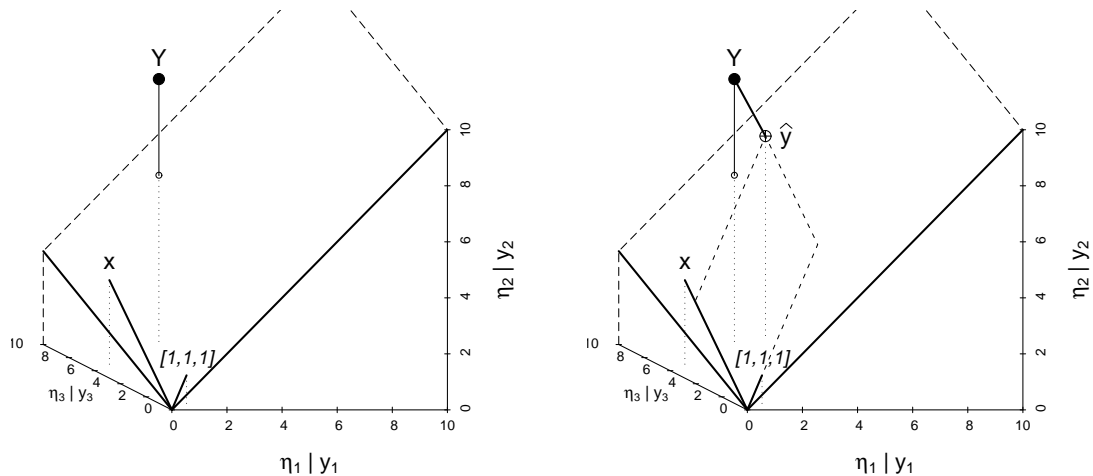


Abbildung 1.2.c: Geometrische Veranschaulichung der einfachen linearen Regression. Die Werte von $\underline{\eta}(\underline{\beta}) = \beta_0 + \beta_1 \underline{x}$ für variierende Parameter $[\beta_0, \beta_1]$ führen zu einer Ebene im Raum. Rechts ist zusätzlich der Punkt auf der Ebene eingezeichnet, der dem Punkt $\underline{Y} = [Y_1, Y_2, Y_3]$ am nächsten liegt. Er stellt die angepassten Werte $\hat{\underline{y}}$ dar und legt die geschätzten Parameterwerte $\hat{\underline{\beta}}$ fest.

Zurück zum Problem der Schätzung der Parameter. Das Kriterium $S(\underline{\beta})$ ist, geometrisch ausgedrückt, der quadrierte Abstand zwischen \underline{Y} und $\underline{\eta}(\underline{\beta})$. Gesucht ist also der Punkt auf der Ebene, der den kleinsten Abstand zu \underline{Y} hat. Dieser wird auch die **Projektion** von \underline{Y} auf die Ebene genannt. Die Parameterwerte, die diesem Punkt $\hat{\underline{\eta}}$ entsprechen, sind dann die geschätzten Parameterwerte $\hat{\underline{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]^T$.

◁

- d Nun soll die nichtlineare Funktion $h(\underline{x}; \underline{\theta}) = \theta_1 \exp(1 - \theta_2 x)$ an die gleichen drei Beobachtungen angepasst werden. Für $\theta_1 = 16$ und $\theta_2 = 0.4$ erhält man $\underline{\eta}(\underline{\theta}) = h(\underline{x}; \underline{\theta}) = [5.275, 12.770, 15.027]^T$. Verändert man die beiden Parameter, so erhält man eine zweidimensionale, gekrümmte Fläche im dreidimensionalen Raum, siehe Figur 1.2.d.
- e Das **Schätzproblem** besteht wieder darin, den Punkt $\hat{\underline{\eta}}$ auf der Fläche zu finden, der \underline{Y} am nächsten liegt. In Figur 1.2.e sieht man, dass der entsprechende Wert $\hat{\theta}_1$ etwas kleiner als 21 und $\hat{\theta}_2$ etwas grösser als 0.6 ist. Die genaue Lösung ist $\hat{\underline{\theta}} = [20.82, 0.6103]^T$.

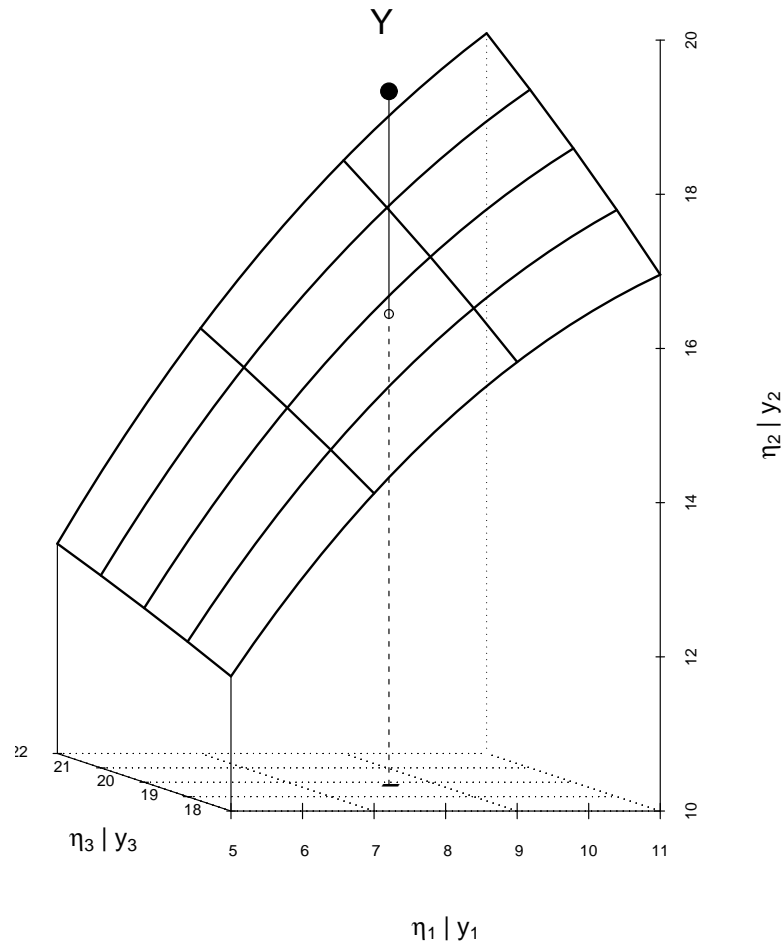


Abbildung 1.2.d: Geometrische Veranschaulichung der nichtlinearen Regression. Die Werte von $\underline{\eta}(\underline{\theta}) = h(\underline{x}; \theta_1, \theta_2)$ für variierende Parameter $[\theta_1, \theta_2]$ führen zu einer zweidimensionalen „Modellfläche“ im dreidimensionalen Raum. Die Linien auf der Fläche entsprechen konstantem η_1 respektive η_3 .

- f Die Hauptidee des üblichen **Algorithmus**‘ läuft wie folgt: Wenn ein vorläufig bester Wert $\underline{\theta}^{(\ell)}$ vorliegt, approximiert man die Modellfläche durch die Ebene, die die Fläche im Punkt $\underline{\eta}(\underline{\theta}^{(\ell)}) = h(\underline{x}; \underline{\theta}^{(\ell)})$ berührt. Nun sucht man den Punkt in dieser Ebene, der am nächsten bei \underline{Y} liegt. Das läuft auf die Schätzung in einem linearen Regressionsproblem hinaus. Dieser neue Punkt liegt auf der Ebene, aber nicht auf der Fläche, die dem nichtlinearen Problem entspricht. Er legt aber einen Parametervektor $\underline{\theta}^{(\ell+1)}$ fest, und mit diesem geht man in die nächste Iterations-Runde.
- g Um die approximierende Ebene zu bestimmen, brauchen wir die partiellen Ableitungen

$$A_i^{(j)}(\underline{\theta}) := \frac{\partial \eta_i(\underline{\theta})}{\partial \theta_j},$$

die wir zu einer $n \times p$ -Matrix \mathbf{A} zusammenfassen können. Die Approximation der Modellfläche $\underline{\eta}(\underline{\theta})$ durch die „Tangentialebene“ in einem Parameterwert $\underline{\theta}^*$ lautet

$$\eta_i(\underline{\theta}) \approx \eta_i(\underline{\theta}^*) + A_i^{(1)}(\underline{\theta}^*) (\theta_1 - \theta_1^*) + \dots + A_i^{(p)}(\underline{\theta}^*) (\theta_p - \theta_p^*)$$

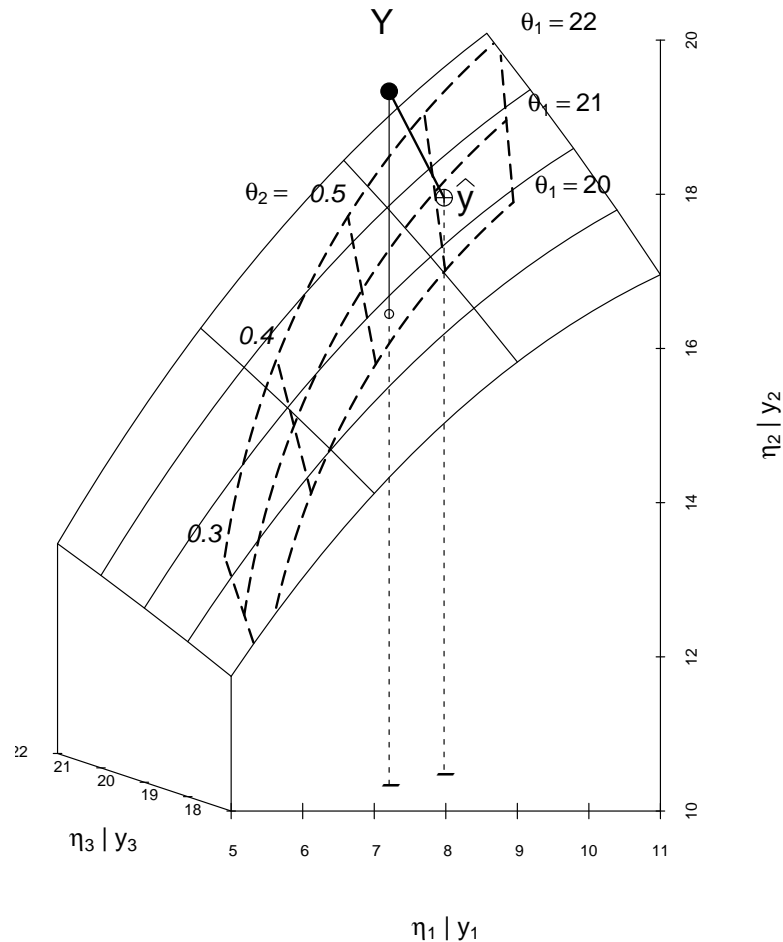


Abbildung 1.2.e: Geometrische Veranschaulichung der nichtlinearen Regression. Es sind zusätzlich Linien konstanter Parameterwerte θ_1 respektive θ_2 eingezeichnet. Der Vektor der geschätzten Modellwerte $\hat{\underline{y}} = h(\underline{x}; \hat{\underline{\theta}})$ ist der Punkt auf der Fläche, der dem Punkt \underline{Y} am nächsten liegt.

oder, in Matrixschreibweise,

$$\underline{\eta}(\underline{\theta}) \approx \underline{\eta}(\underline{\theta}^*) + \underline{A}(\underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*).$$

Wenn wir nun den Zufallsfehler wieder hinzufügen, erhalten wir ein lineares Regressionsmodell

$$\underline{Y} - \underline{\eta}(\underline{\theta}^*) = \underline{A}(\underline{\theta}^*) \underline{\beta} + \underline{E}$$

mit den „vorläufigen Residuen“ $Y_i - \eta_i(\underline{\theta}^*)$ als Zielgröße, den Spalten von \underline{A} als Regressoren und den Koeffizienten $\beta_j = \theta_j - \theta_j^*$ (ein Modell ohne Achsenabschnitt β_0).

- h Der **Gauss-Newton-Algorithmus** besteht darin, ausgehend von einem Startwert $\underline{\theta}^{(0)}$ für $\underline{\theta}$ das gerade eingeführte lineare Regressionsproblem mit $\underline{\theta}^* = \underline{\theta}^{(0)}$ zu lösen, um eine Korrektur $\hat{\underline{\beta}}$ und daraus einen verbesserten Wert $\underline{\theta}^{(1)} = \underline{\theta}^{(0)} + \hat{\underline{\beta}}$ zu erhalten. Für diesen wird wieder das approximierende Modell ausgerechnet, also die Residuen $\underline{Y} - \underline{\eta}(\underline{\theta}^{(1)})$ und die partiellen Ableitungen $\underline{A}(\underline{\theta}^{(1)})$ bestimmt, und daraus ergibt sich $\underline{\theta}^{(2)}$. Mit diesem Iterationsschritt wird so lange weitergefahren, bis die Korrektur $\underline{\beta}$ vernachlässigbar wird.

Es kann nicht garantiert werden, dass dieses Verfahren tatsächlich das Minimum der Quadratsumme findet. Die Chancen dafür stehen besser, je besser sich die p -dimensionale Modellfläche im Minimum $\hat{\underline{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^T$ durch eine p -dimensionale „Ebene“ lokal approximieren lässt, und je näher der Startwert $\underline{\theta}^{(0)}$ zur gesuchten Lösung ist.

- i* Komfortable Algorithmen bestimmen die Ableitungsmatrix \mathbf{A} numerisch. In komplexeren Problemen kann die numerische Näherung ungenügend sein und Konvergenzprobleme verursachen. Dann ist es von Vorteil, wenn analytische Ausdrücke für die ersten partiellen Ableitungen angegeben werden können. Damit kann die Ableitungsmatrix numerisch zuverlässiger bestimmt werden und das Verfahren konvergiert eher (siehe jedoch auch Abschnitt 1.7).

1.3 Schätzung der Parameter: Startwerte und Beispiele

- a Ein iteratives Verfahren benötigt **Startwerte**, damit es überhaupt angewandt werden kann. Gute Startwerte helfen, dass das iterative Verfahren schneller und sicherer die Lösung findet. Einige Möglichkeiten, diese mehr oder weniger einfach zu gewinnen, werden hier kurz vorgestellt.
- b Wie schon in der Einleitung bemerkt, stammen die nichtlinearen Modelle vielfach aus theoretischen Überlegungen in der jeweiligen Substanzwissenschaft. Bestehen schon **Vorkenntnisse** aus ähnlichen Experimenten, so können diese verwendet werden, um Startwerte zu gewinnen. Um sicher zu gehen, dass der gewählte Startwert passt, ist es ratsam, die Regressionsfunktion $h(x; \underline{\theta})$ für verschiedene mögliche Startwerte $\underline{\theta} = \underline{\theta}^0$ graphisch mit den Daten zusammen darzustellen (z. B. so wie in Abbildung 1.3.c, rechts).
- c Manchmal ist man wegen der Verteilung der Fehler gezwungen, in Modellen mit **linearisierbaren Regressionsfunktionen** bei der nichtlinearen Form zu verbleiben. Im Beispiel der Schadstoffe im Tunnel (1.1.j) war sogar der Ausgangspunkt ein lineares Modell, das wegen der Verteilungsannahmen in ein nichtlineares verwandelt wurde. Das lineare Modell kann aber Startwerte liefern.
- ▷ Im **Beispiel Puromycin** ist die Regressionsfunktion linearisierbar: Die Kehrwerte der beiden Variablen erfüllen

$$\tilde{y} = \frac{1}{y} \approx \frac{1}{h(x; \underline{\theta})} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \frac{1}{x} = \beta_0 + \beta_1 \tilde{x}.$$

Die Kleinste-Quadrate-Lösung für dieses modifizierte Problem ist $\hat{\underline{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]^T = (0.00511, 0.000247)^T$ (Figur 1.3.c (a)). Das liefert die Startwerte

$$\theta_1^{(0)} = 1/\hat{\beta}_0 = 196, \quad \theta_2^{(0)} = \hat{\beta}_1/\hat{\beta}_0 = 0.048.$$

◁

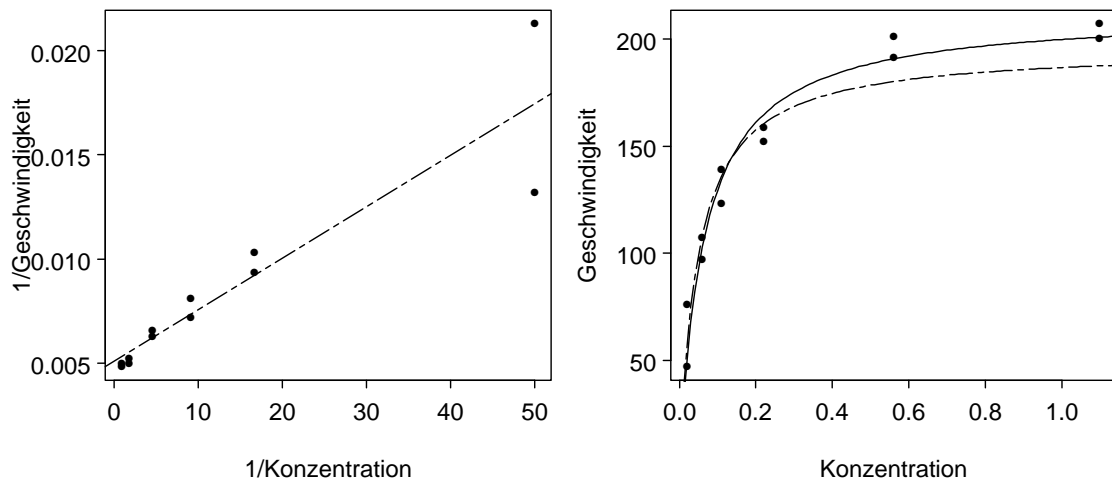


Abbildung 1.3.c: Beispiel Puromycin. (a) Regressionsgerade im linearisierten Problem. (b) Regressionsfunktion $h(x; \underline{\theta})$ für die Startwerte $\underline{\theta}^{(0)}$ (-----) und für die Kleinst-Quadrat-Schätzung $\hat{\underline{\theta}}$ (—).

- d ▷ Im **Beispiel Puromycin** können wir auch noch auf eine andere, instructive Art zu Startwerten gelangen: θ_1 ist der y -Wert für $x = \infty$. Da die Regressionsfunktion monoton steigend ist, können wir den maximalen y_i -Wert oder einen visuell bestimmten „asymptotischen Wert“ $\theta_1^0 = 207$ als Startwert für θ_1 benutzen. Der Parameter θ_2 ist der x -Wert, bei dem y die Hälfte des asymptotischen Wertes θ_1 erreicht. Das ergibt $\theta_2^0 = 0.06$.

◁

Die Startwerte ergeben sich also aus der geometrischen Bedeutung der Parameter und einer groben Bestimmung der entsprechenden Aspekte von einer von Auge eingepassten“ Kurve.

- e ▷ Lassen wir im **Beispiel aus der Membrantrenn-Technologie** $x \rightarrow \infty$ gehen, so geht $h(x; \underline{\theta}) \rightarrow \theta_1$ (da $\theta_4 < 0$); für $x \rightarrow -\infty$ geht $h(x; \underline{\theta}) \rightarrow \theta_2$. Aus Figur 1.1.f (a) und den Daten geht hervor, dass $\theta_1 \approx 163.7$ und $\theta_2 \approx 159.5$ ist. Sind θ_1 und θ_2 bekannt, so kann man die Regressionsfunktion linearisieren durch

$$\tilde{y} := \log_{10} \left\langle \frac{\theta_1^{(0)} - y}{y - \theta_2^{(0)}} \right\rangle = \theta_3 + \theta_4 x.$$

Man spricht von einer **bedingt linearisierbaren** Funktion. Die lineare Regression führt zu den Startwerten $\theta_3^{(0)} = 1.83$ und $\theta_4^{(0)} = -0.36$.

Mit diesen Startwerten konvergiert der Algorithmus zur Lösung $\hat{\theta}_1 = 163.7$, $\hat{\theta}_2 = 159.8$, $\hat{\theta}_3 = 2.67$ und $\hat{\theta}_4 = -0.512$. Die Funktionen $h(\cdot; \underline{\theta}^{(0)})$ und $h(\cdot; \hat{\underline{\theta}})$ sind in Figur 1.3.e (b) dargestellt.

* Die Eigenschaft der bedingten Linearität von Funktionen kann auch dazu benutzt werden, einen dieser Situation speziell angepassten Algorithmus zu entwickeln (siehe z. B. Bates and Watts (1988)).

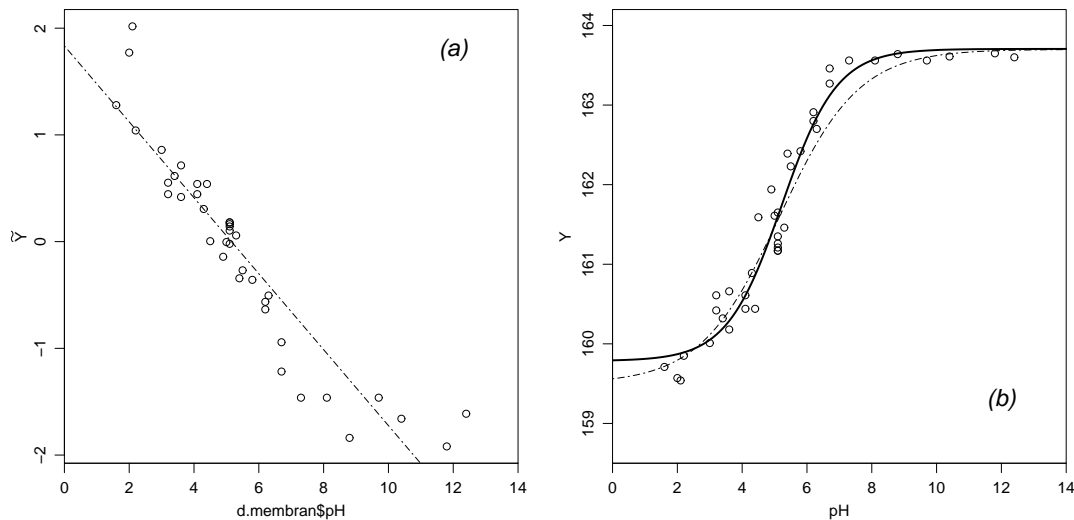


Abbildung 1.3.e: Beispiel aus der Membrantrenn-Technologie. (a) Regressionsgerade, die zur Bestimmung der Startwerte für θ_3 und θ_4 gebraucht wird. (b) Daten und Regressionsfunktion $h(x; \underline{\theta})$ für die Startwerte $\underline{\theta} = \underline{\theta}^{(0)}$ (-----) und für die der Kleinste-Quadrate-Schätzung $\underline{\theta} = \hat{\underline{\theta}}$ (—).

1.4 Genäherte Tests und Vertrauensbereiche

◁

- a Die Schätzung $\hat{\underline{\theta}}$ liefert den Wert von $\underline{\theta}$, der optimal zu den Daten passt. Nun fragen wir, *welche Parameterwerte $\underline{\theta}$ mit den Beobachtungen verträglich sind*. Der **Vertrauensbereich** ist die Menge all dieser Werte. Für einen einzelnen Parameter θ_j wird der Vertrauensbereich zum **Vertrauensintervall** oder **Konfidenzintervall**.

Die Resultate, die nun folgen, beruhen darauf, dass die Schätzung $\hat{\underline{\theta}}$ asymptotisch multivariat normalverteilt ist. Für einen einzelnen Parameter führt das zu einem „z-Test“ und zum entsprechenden Vertrauensintervall; für mehrere Parameter kommt der entsprechende Chiquadrat-Test zum Zug und liefert elliptische Vertrauensbereiche.

- b Die **asymptotischen Eigenschaften** der Schätzung können aus der linearen Approximation hergeleitet werden. Das Problem der nichtlinearen Regression ist ja näherungsweise gleich dem in 1.2.g erwähnten linearen Regressionsproblem, wenn der Parametervektor θ^* , der für die Linearisierung verwendet wird, nahe bei der Lösung liegt. Im Lösungspunkt $\hat{\underline{\theta}}$ ist die Lösung für $\underline{\beta}$ im linearen Problem exakt $= \underline{0}$ – sonst wäre es nicht die Lösung. Die Standardfehler der Koeffizienten $\underline{\beta}$ – und allgemeiner die Kovarianzmatrix von $\hat{\underline{\beta}}$ – geben aber näherungsweise die entsprechenden Größen für $\hat{\underline{\theta}}$ wieder.

* Etwas genauer: Die Standardfehler geben ja die Unsicherheiten wieder, die durch die Zufallsschwankungen der Daten erzeugt werden. Die vorliegenden Daten haben zum Schätzwert $\hat{\underline{\theta}}$ geführt. Wären die Daten etwas anders ausgefallen, dann wäre $\hat{\underline{\theta}}$ immer noch ungefähr richtig, also – so nehmen wir an – gut genug für die Linearisierung. Die Schätzung von $\underline{\beta}$ für den neuen Datensatz würde also so weit vom Schätzwert für den vorliegenden Daten weg liegen, wie es der Verteilung der Parameter im linearisierten Problem entspricht.

- c Aus dieser Überlegung folgt: Asymptotisch ist die Kleinste-Quadrate-Schätzung $\hat{\underline{\theta}}$ normalverteilt (und konsistent) und deshalb

$$\hat{\underline{\theta}} \approx \sim \mathcal{N}(\underline{\theta}, \mathbf{V}(\underline{\theta})/n) .$$

mit asymptotischer Kovarianzmatrix $\mathbf{V}(\underline{\theta}) = \sigma^2 (\mathbf{A}(\underline{\theta})^T \mathbf{A}(\underline{\theta}))^{-1}$, wobei $\mathbf{A}(\underline{\theta})$ die $n \times p$ Matrix der partiellen Ableitungen ist (1.2.g).

Um die Kovarianzmatrix explizit zu bestimmen, wird $\mathbf{A}(\underline{\theta})$ an der Stelle $\hat{\underline{\theta}}$ berechnet, und für die Fehlervarianz σ^2 wird die übliche Schätzung eingesetzt,

$$\widehat{\mathbf{V}(\underline{\theta})} = \hat{\sigma}^2 \left(\mathbf{A}(\hat{\underline{\theta}})^T \mathbf{A}(\hat{\underline{\theta}}) \right)^{-1}, \quad \hat{\sigma}^2 = \frac{1}{n-p} S(\hat{\underline{\theta}}) = \frac{1}{n-p} \sum_{i=1}^n (y_i - h(x_i; \hat{\underline{\theta}}))^2 .$$

Damit ist die **Verteilung der geschätzten Parameter** näherungsweise bestimmt, und daraus lassen sich wie in der linearen Regression Standardfehler und Vertrauensintervalle herleiten, ebenso Vertrauens-Ellipsen (oder -Ellipsoide), wenn mehrere Parameter gemeinsam betrachtet werden.

Der Nenner $n - p$ in $\hat{\sigma}^2$ wurde in der linearen Regression eingeführt, um die Schätzung erwartungstreu zu machen. Tests und Vertrauensintervalle wurden nicht mit der Normal- und Chiquadrat-Verteilung bestimmt, sondern mit der **t- und F-Verteilung**. Damit wurde berücksichtigt, dass die Schätzung von σ^2 eine zusätzliche Zufallsschwankung bewirkt. Auch wenn die Verteilungen nicht mehr exakt gelten, so werden die Näherungen doch genauer, wenn man dies bei der nichtlinearen Regression ebenfalls tut. Asymptotisch geht der Unterschied gegen null.

- d ▷ Eine Computer-Ausgabe für das **Beispiel aus der Membrantrenn-Technologie** zeigt Tabelle 1.4.d. Die Schätzungen der Parameter stehen in der Kolonne „Value“, gefolgt von den geschätzten approximativen Standardfehler und den Teststatistiken („t value“), die approximativ t_{n-p} -verteilt sind. In der letzten Zeile wird die geschätzte Standardabweichung $\hat{\sigma}$ der Zufallsfehler E_i angegeben.

Parameters:

	Value	Std. Error	t value
T1	163.706	0.1262	1297.21
T2	159.784	0.1595	1002.03
T3	2.675	0.3813	7.02
T4	-0.512	0.0703	-7.28

Residual standard error: 0.2931 on 35 degrees of freedom

Tabelle 1.4.d: Computer-Ausgabe für das Beispiel aus der Membrantrenn-Technologie

Aus diesen Angaben können wie in der linearen Regression die Vertrauensintervalle für die Parameter bestimmt werden: Das approximative 95%-Vertrauensintervall für den Parameter θ_1 ist $163.706 \pm q_{0.975}^{t_{35}} \cdot 0.1262 = 163.706 \pm 0.256$.

◁

- e ▷ **Beispiel Puromycin.** Zur Überprüfung eines Einflusses der Behandlung des Enzyms mit Puromycin von der postulierten Form (1.1.d) kann ein gemeinsames Modell für die Daten mit und ohne Behandlung folgendermassen formuliert werden:

$$Y_i = \frac{(\theta_1 + \theta_3 z_i) x_i}{\theta_2 + \theta_4 z_i + x_i} + E_i.$$

Dabei ist z die Indikatorenvariable für die Behandlung ($z_i = 1$, wenn behandelt, sonst $=0$).

Parameters:

	Value	Std. Error	t value
T1	160.286	6.8964	23.24
T2	0.048	0.0083	5.76
T3	52.398	9.5513	5.49
T4	0.016	0.0114	1.44

Residual standard error: 10.4 on 19 degrees of freedom

Tabelle 1.4.e: Computer-Ausgabe für das Beispiel Puromycin

Tabelle 1.4.e zeigt, dass der Parameter θ_4 nicht signifikant von 0 verschieden ist, denn der t-Wert von 1.44 ist kleiner als die kritische Grenze $q_{0.975}^{t_{19}} = 2.09$. Die Behandlung hat aber einen eindeutigen Einfluss, der sich durch θ_3 ausdrückt; das 95% Vertrauensintervall überdeckt $52.398 \pm 9.5513 \cdot 2.09 = [32.4, 72.4]$.

◁

- f Neben den Parametern ist oft der Funktionswert $h(\underline{x}_0, \underline{\theta})$ für ein beliebiges \underline{x}_0 von Interesse. In der linearen Regression wird der Funktionswert $h(\underline{x}_0, \underline{\beta}) = \underline{x}_0^T \underline{\beta}$ durch $\underline{x}_0^T \hat{\underline{\beta}}$ geschätzt, und das $(1 - \alpha)$ -Vertrauensintervall dafür ist

$$\underline{x}_0^T \hat{\underline{\beta}} \pm \hat{\sigma} \sqrt{\underline{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \underline{x}_0} q_{1-\alpha/2}^{t_{n-p}}.$$

Durch analoge Überlegungen und asymptotische Näherung kann man Vertrauensintervalle für den Funktionswerte $h(\underline{x}_0; \underline{\theta})$ für nicht lineare h angeben. Wird die Funktion $\eta_0(\underline{\theta}) := h(\underline{x}_0, \underline{\theta})$ an der Stelle $\underline{\theta}$ linear approximiert, so erhält man

$$\eta_0(\hat{\underline{\theta}}) \approx \eta_0(\underline{\theta}) + \underline{a}_0^T (\hat{\underline{\theta}} - \underline{\theta}) \quad \text{mit } \underline{a}_0 = \frac{\partial h(\underline{x}_0, \underline{\theta})}{\partial \underline{\theta}}.$$

(Wenn \underline{x}_0 gleich einem beobachteten \underline{x}_i ist, dann ist \underline{a}_0 gleich der entsprechenden Zeile der Matrix \mathbf{A} aus 1.2.g.) Das Vertrauensintervall für den Funktionswert $h(\underline{x}_0, \underline{\theta})$ ist dann approximativ

$$h(\underline{x}_0, \hat{\underline{\theta}}) \pm q_{1-\alpha/2}^{t_{n-p}} \hat{\sigma}_{x_0} \quad \text{mit } \hat{\sigma}_{x_0} = \hat{\sigma} \sqrt{\hat{\underline{a}}_0^T (\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} \hat{\underline{a}}_0}.$$

In dieser Formel wurden wieder die unbekannten Grössen durch ihre Schätzungen ersetzt.

- g Der Ausdruck für das Vertrauensintervall für $h(\underline{x}_0, \underline{\theta})$ gilt für beliebiges \underline{x}_0 . Es ist wie in der linearen Regression naheliegend, die Grenzen dieses Intervalls als Funktion von \underline{x}_0 als „**Vertrauensband**“ aufzuzeichnen, wie dies Figur 1.4.g für die beiden Beispiele Puromycin und Sauerstoffverbrauch zeigt.

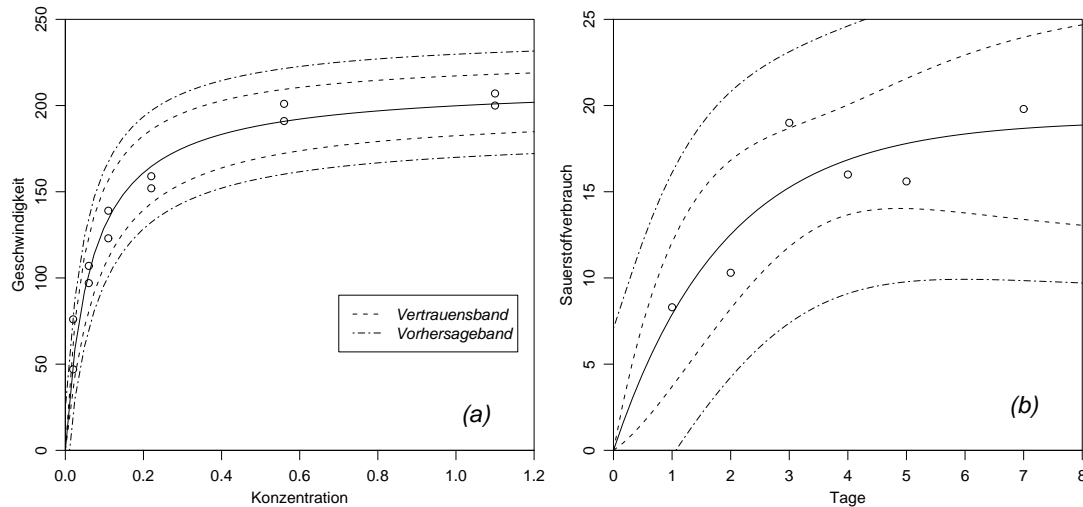


Abbildung 1.4.g: Vertrauensband für den Funktionswert h und Vorhersageband, (a) Beispiel Puromycin, (b) Beispiel Sauerstoffverbrauch.

Vertrauensbänder für lineare und nichtlineare Regressionsfunktionen verhalten sich verschieden: Bei linearen Funktionen ist das Vertrauensband beim Schwerpunkt der Ausgangs-Variablen am engsten und wird gegen aussen allmählich breiter. Im nichtlinearen Fall können die Bänder beliebiger sein. Weil die Funktionen in den Beispielen durch den Nullpunkt gehen müssen, schrumpft dort das Intervall zu einem Punkt. Beide Modelle haben eine horizontale Asymptote und deshalb wird das Band für grosse x eine konstante Breite erreichen.

- h Das betrachtete Vertrauensband gibt an, wo die **idealen Funktionswerte** $h(\underline{x}; \underline{\theta})$, also die Erwartungswerte von Y bei gegebenen x , liegen. Die Frage, in welchem Bereich **künftige Beobachtungen** Y_0 für vorgegebenes \underline{x}_0 zu liegen kommen, ist damit nicht beantwortet. Sie ist aber oft interessanter als die Frage nach dem idealen Funktionswert; man möchte beispielsweise wissen, in welchem Bereich der zu messende Wert des Sauerstoffverbrauches für eine Inkubationszeit von 6 Tagen liegen wird.

Eine solche Angabe ist eine Aussage über eine **Zufallsvariable** und ist prinzipiell zu unterscheiden von einem Vertrauensintervall, das über einen **Parameter**, also eine feste, aber unbekannte Zahl, etwas aussagt. Entsprechend der Fragestellung nennen wir den gesuchten Bereich **Vorhersage-Intervall** oder **Prognose-Intervall**.

Wie im linearen Fall ist dieses Intervall eng mit dem Vertrauensintervall für den Funktionswert verknüpft; man muss lediglich $\hat{\sigma}_{x_0}$ in der obigen Formel durch $\sqrt{\hat{\sigma}^2 + \hat{\sigma}_{x_0}^2}$ ersetzen. Die entsprechenden Bänder sind in Figur 1.4.g ebenfalls eingezeichnet.

1.5 Genauere Tests und Vertrauensintervalle

- a Die Qualität der approximativen Vertrauensbereiche ist stark von der Qualität der linearen Approximation abhängig. Ebenfalls werden die Konvergenzeigenschaften der Optimierungsalgorithmen durch die Qualität der linearen Approximation beeinflusst. Mit grösserem Rechenaufwand lässt sich die Linearität grafisch überprüfen, und gleichzeitig erhält man genauere Vertrauensintervalle.
- b Um eine Nullhypothese $\underline{\theta} = \underline{\theta}^*$ für den ganzen Parametervektor oder auch $\theta_j = \theta_j^*$ für eine einzelne Komponente zu testen, kann man, wie in der linearen Regression, den **F-Test zum Vergleich von Modellen** verwenden. Man vergleicht dabei die Quadratsumme $S\langle \underline{\theta}^* \rangle$, die sich unter der Nullhypothese ergibt, mit der Quadratsumme $S\langle \hat{\underline{\theta}} \rangle$. (Für $n \rightarrow \infty$ stimmt der F-Test mit dem so genannten Likelihood-Ratio-Test überein, und die Quadratsumme ist, bis auf eine Konstante, gleich der Log-Likelihood.)
- c Zunächst wollen wir eine Nullhypothese $\underline{\theta} = \underline{\theta}^*$ über den **ganzen Parameter** betrachten. Die Teststatistik ist

$$T = \frac{n-p}{p} \cdot \frac{S\langle \underline{\theta}^* \rangle - S\langle \hat{\underline{\theta}} \rangle}{S\langle \hat{\underline{\theta}} \rangle} \sim F_{p, n-p}.$$

Daraus erhält man als Vertrauensbereich

$$\left\{ \underline{\theta} \mid S\langle \underline{\theta} \rangle \leq S\langle \hat{\underline{\theta}} \rangle \left(1 + \frac{p}{n-p} q \right) \right\}$$

wobei $q = q_{1-\alpha}^{F_{p, n-p}}$ das $(1-\alpha)$ -Quantil der F-Verteilung mit p und $n-p$ Freiheitsgraden ist.

In der linearen Regression erhält man genau den gleichen Vertrauensbereich, wenn man die (multivariate) Normalverteilung der Schätzung $\hat{\underline{\beta}}$ benützt. Im nichtlinearen Fall sind die Ergebnisse verschieden. Der Bereich, der auf dem F-Test beruht, benützt die lineare Approximation des nichtlinearen Problems nicht und ist deshalb (viel) exakter.

- d Falls $p = 2$ ist, können wir den **exakten Bereich** finden, indem wir $S\langle \underline{\theta} \rangle$ auf einem Gitter von $\underline{\theta}$ -Werten berechnen und durch Interpolation die Grenzen des Vertrauensbereichs bestimmen, wie das für Kontur-Plots geläufig ist. In Figur 1.5.d sind die Konturen zusammen mit den elliptischen Bereichen, die sich aus der linearen Approximation ergeben, für die Beispiele **Puromycin** (links) und **Sauerstoffverbrauch** (rechts) wiedergegeben.

Für $p > 2$ gibt es keine Kontur-Plots. Wir werden im nächsten Abschnitt grafische Hilfsmittel kennenlernen, die auch für höhere Dimensionen funktionieren. Sie beruhen auf den folgenden Überlegungen.

- e Es soll geprüft werden, ob ein **einzelner Parameter** θ_k gleich einem bestimmten Wert θ_k^* sein kann. Über die übrigen Parameter macht eine solche Nullhypothese keine Aussage. Das Modell, das der Nullhypothese entspricht und am besten zu den Daten passt, ist durch eine Kleinste-Quadrate-Schätzung der übrigen Parameter bei festem $\theta_k = \theta_k^*$ bestimmt. Es wird also $S\langle \theta_1, \dots, \theta_k^*, \dots, \theta_p \rangle$ minimiert in Bezug auf alle θ_j , $j \neq k$. Das Minimum bezeichnen wir mit \tilde{S}_k und die Werte θ_j , die zu ihm führen, mit $\tilde{\theta}_j$. Beide Grössen hängen von θ_k^* ab. Wir schreiben deshalb $\tilde{S}_k\langle \theta_k^* \rangle$ und $\tilde{\theta}_j\langle \theta_k^* \rangle$.

Die Teststatistik für den F-Test ist

$$\tilde{T}_k = (n-p) \frac{\tilde{S}_k\langle \theta_k^* \rangle - S\langle \hat{\underline{\theta}} \rangle}{S\langle \hat{\underline{\theta}} \rangle}.$$

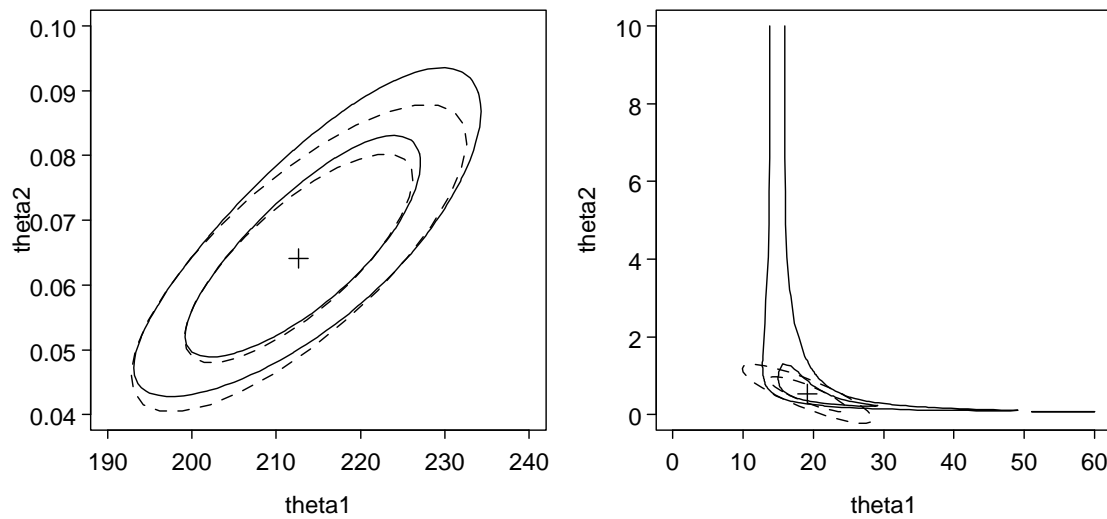


Abbildung 1.5.d: Nominale 80% und 95% Likelihood-Konturen (—) und die Vertrauensellipsen aus der asymptotischen Approximation (---). Der Punkt + zeigt die Kleinst-Quadrat Lösung. Im Beispiel Pyromycin (links) ist die Übereinstimmung gut, im Beispiel Sauerstoffverbrauch (rechts) dagegen schlecht.

Sie hat (genähert) eine $F_{1,n-p}$ -Verteilung.

Ein Vertrauensintervall erhält man daraus, indem man die Gleichung $\tilde{T}_k = q_{0.95}^{F_{1,n-p}}$ numerisch nach θ_k auflöst. Sie hat eine Lösung, die kleiner als $\hat{\theta}_k$ ist, und eine, die grösser ist.

- f In der linearen Regression und im vorhergehenden Abschnitt haben wir Tests und Vertrauensintervalle aus einer Testgrösse ausgerechnet, die einer t-Verteilung folgt (t-Test für die Koeffizienten). Ist das ein anderer Test?

Es stellt sich heraus, dass die Teststatistik des t-Tests in der linearen Regression in die Teststatistik des F-Tests übergeht, wenn man sie quadriert, und die beiden Tests sind äquivalent. In der nichtlinearen Regression ist der F-Test nicht äquivalent mit dem im letzten Abschnitt besprochenen t-Test (1.4.d). Aber wir können den F-Test in einen t-Test verwandeln, der genauer ist als der des letzten Abschnitts:

Aus der Teststatistik des F-Tests ziehen wir die Wurzel und versehen diese mit dem Vorzeichen von $\hat{\theta}_k - \theta_k^*$,

$$T_k\langle\theta_k^*\rangle := \text{sign}\langle\hat{\theta}_k - \theta_k^*\rangle \frac{\sqrt{\tilde{S}_k\langle\theta_k^*\rangle - S\langle\hat{\theta}\rangle}}{\hat{\sigma}}.$$

($\text{sign}\langle a \rangle$ bezeichnet das Vorzeichen von a , und es ist $\hat{\sigma}^2 = S\langle\hat{\theta}\rangle/(n-p)$.) Diese Teststatistik ist (genähert) t_{n-p} -verteilt.

Im linearen Regressionsmodell ist T_k , wie erwähnt, gleich der Teststatistik des üblichen t-Tests,

$$T_k\langle\theta_k^*\rangle = \frac{\hat{\theta}_k - \theta_k^*}{\text{se}(\hat{\theta}_k)}.$$

g^* Wir können auch mit dieser Technik ein Vertrauensintervall für einen Funktionswert an einer Stelle x_0 bestimmen. Dazu reparametrisieren wir das ursprüngliche Problem so, dass ein Parameter, sagen wir ϕ_1 , den Funktionswert $h(x_0)$ repräsentiert und gehen dann wie besprochen vor.

1.6 Profil-t-Plot und Profilsuren

- a Die grafischen Hilfsmittel zur Überprüfung der linearen Approximation beruhen auf dem gerade besprochenen t-Test, der ja eben diese Näherung nicht benützt. Wir betrachten die Teststatistik T_k (1.5.e) als Funktion ihres Argumentes θ_k und nennen sie **Profil-t-Funktion**. Für die lineare Regression erhält man, wie in 1.5.e erwähnt, eine Gerade, während für die nichtlineare Regression eine monoton steigende Funktion herauskommt. Den grafischen Vergleich von $T_k(\theta_k)$ mit einer Geraden ermöglicht der so genannte **Profil-t-Plot**. Es ist üblich, auf der horizontalen Achse nicht θ_k , sondern die auf Grund der linearen Approximation bestimmte standardisierte Version

$$\delta_k(\theta_k) := \frac{\theta_k - \hat{\theta}_k}{\text{se}(\hat{\theta}_k)}$$

zu verwenden. Die Vergleichsgerade wird dann die „Diagonale“, also die Gerade mit Steigung 1 und Achsenabschnitt 0.

- b Je stärker die Profil-t-Funktion gekrümmt ist, desto stärker ist die Nichtlinearität in einer Umgebung von θ_k . Folglich zeigt diese Darstellung, wie gut die lineare Approximation in einer Umgebung von $\hat{\theta}_k$ ist. (Die Umgebung, die für die Statistik wichtig ist, ist etwa durch $|\delta_k(\theta_k)| \leq 2.5$ bestimmt.) In Figur 1.6.b zeigt sich, dass im Beispiel Puromycin die Nichtlinearität minim, im Beispiel Sauerstoffverbrauch dagegen gross ist.

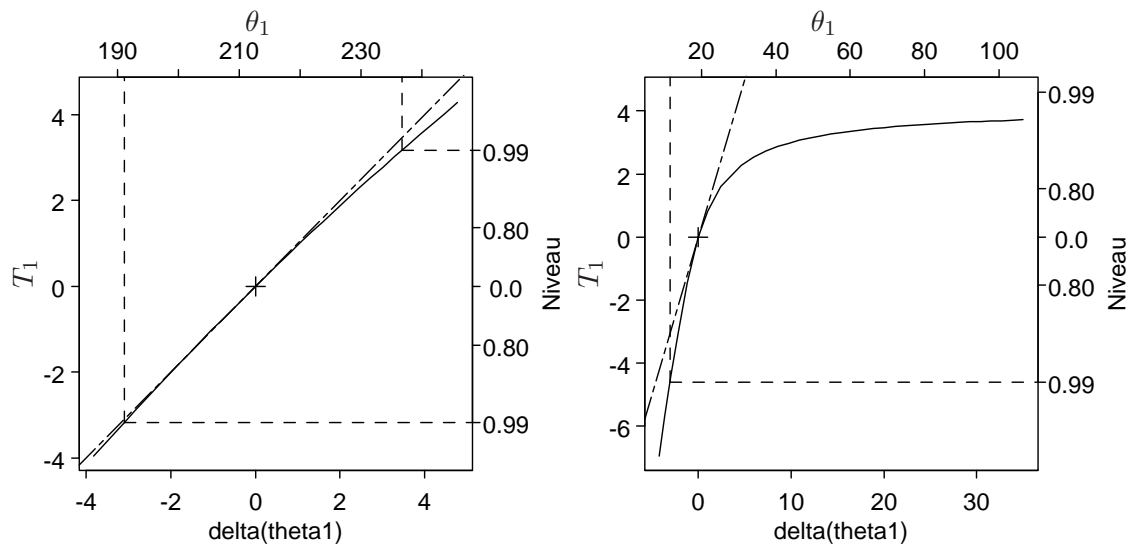
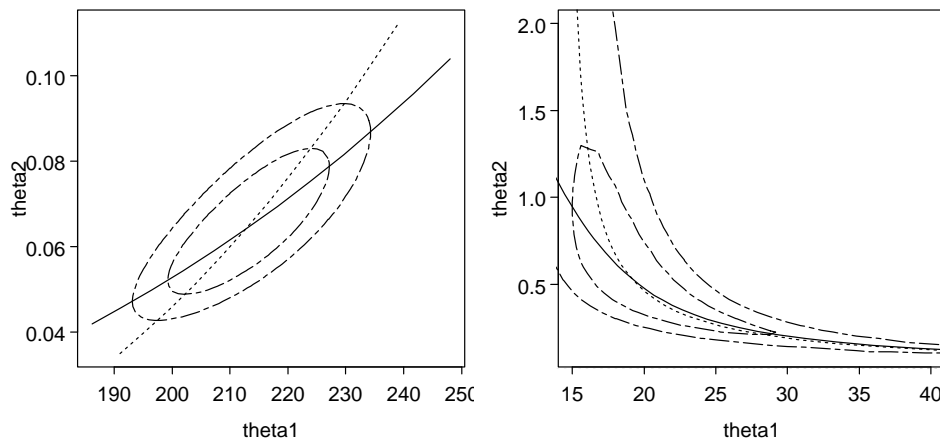


Abbildung 1.6.b: Profil-t-Plot für die ersten Parameter der Beispiele Puromycin und Sauerstoffverbrauch. Die gestrichelten Linien zeigen die verwendete lineare Approximation und die gepunktete Linie die Konstruktion des 99% Vertrauensintervalls mit Hilfe von $T_1(\theta_1)$.

- c Aus den Darstellungen kann man die Vertrauensintervalle gemäss 1.5.e ablesen. Der Bequemlichkeit halber sind auf der rechten vertikalen Achse die Wahrscheinlichkeiten $P\{T_k \leq t\}$ gemäss der t-Verteilung markiert. Im Beispiel des Sauerstoff-Verbrauchs ergibt sich ein Vertrauensintervall ohne obere Grenze!
- d Ein anderes nützliches Hilfsmittel sind die **Likelihood-Profilspuren** (likelihood profile traces). Hier werden die geschätzten Parameter $\tilde{\theta}_j$, $j \neq k$ bei festgehaltenem θ_k (siehe 1.5.e) als Funktionen $\tilde{\theta}_j^{(k)}(\theta_k)$ dieses Wertes betrachtet.

Die grafischen Darstellungen dieser Funktionen würden eine ganze Matrix von Diagrammen füllen, ohne Diagonale allerdings. Es lohnt sich, die „gegenüberliegenden“ Diagramme dieser Matrix zu kombinieren: Über die Darstellung von $\tilde{\theta}_j^{(k)}(\theta_k)$ wird $\tilde{\theta}_k^{(j)}(\theta_j)$ gelegt – in gespiegelter Form, damit die Achsen für beide Funktionen die gleiche Bedeutung haben.



Abbildungung 1.6.d: Likelihood-Profilspuren für θ_1 gegen θ_2 für die Beispiele Puromycin und Sauerstoffverbrauch, mit 80% und 95% Vertrauensbereichen (-----)

▷ In Figur 1.6.d ist je eines dieser Diagramme für unsere beiden **Beispiele** gezeigt. Zusätzlich wurden Konturen von Vertrauensbereichen für $[\theta_1, \theta_2]$ eingezeichnet. Man sieht, dass die Profilsuren die Konturen bei Berührungspunkten der horizontalen, respektive vertikalen Tangenten schneiden.

◁

- e Die Darstellung zeigt nicht nur Nichtlinearitäten, sie enthält nützliche Hinweise, **wie sich die Parameter gegenseitig beeinflussen**. Um diese zu verstehen, betrachten wir zuerst den Fall einer linearen Regressionsfunktion. Die Profilsuren in den einzelnen Diagrammen bestehen dann aus zwei Geraden, die sich im Nullpunkt schneiden. Standardisiert man die Parameter, so kann man zeigen, dass die Steigung der Spur $\tilde{\theta}_j^{(k)}(\theta_k)$ gleich dem Korrelationskoeffizienten c_{kj} der geschätzten Koeffizienten $\hat{\theta}_j$ und $\hat{\theta}_k$ ist. Die „Umkehrspur“ $\tilde{\theta}_k^{(j)}(\theta_j)$ weist dann gegenüber der horizontalen Achse eine Steigung von $1/c_{kj}$ auf. Der Winkel, den die Geraden einschliessen, ist also eine monotone Funktion dieser Korrelation. Er misst damit die **Kollinearität** zwischen den beiden Ausgangs-Variablen. Wenn die Korrelation zwischen den Parameterschätzungen null ist, dann stehen die Spuren senkrecht aufeinander.

Bei einer nichtlinearen Regressionsfunktion sind die beiden Spuren gekrümmt. Der Winkel zwischen ihnen zeigt aber immer noch, wie stark die beiden Parameter θ_j und θ_k zusammenhängen, also ihre Schätzungen korreliert sind.

- f Alle Profil-t-Plots und Profilsuren können zu einer Dreiecks-Matrix von Diagrammen zusammengestellt werden, wie sie Figur 1.6.f für das **Beispiel aus der Membrantrenn-Technologie** zeigt.

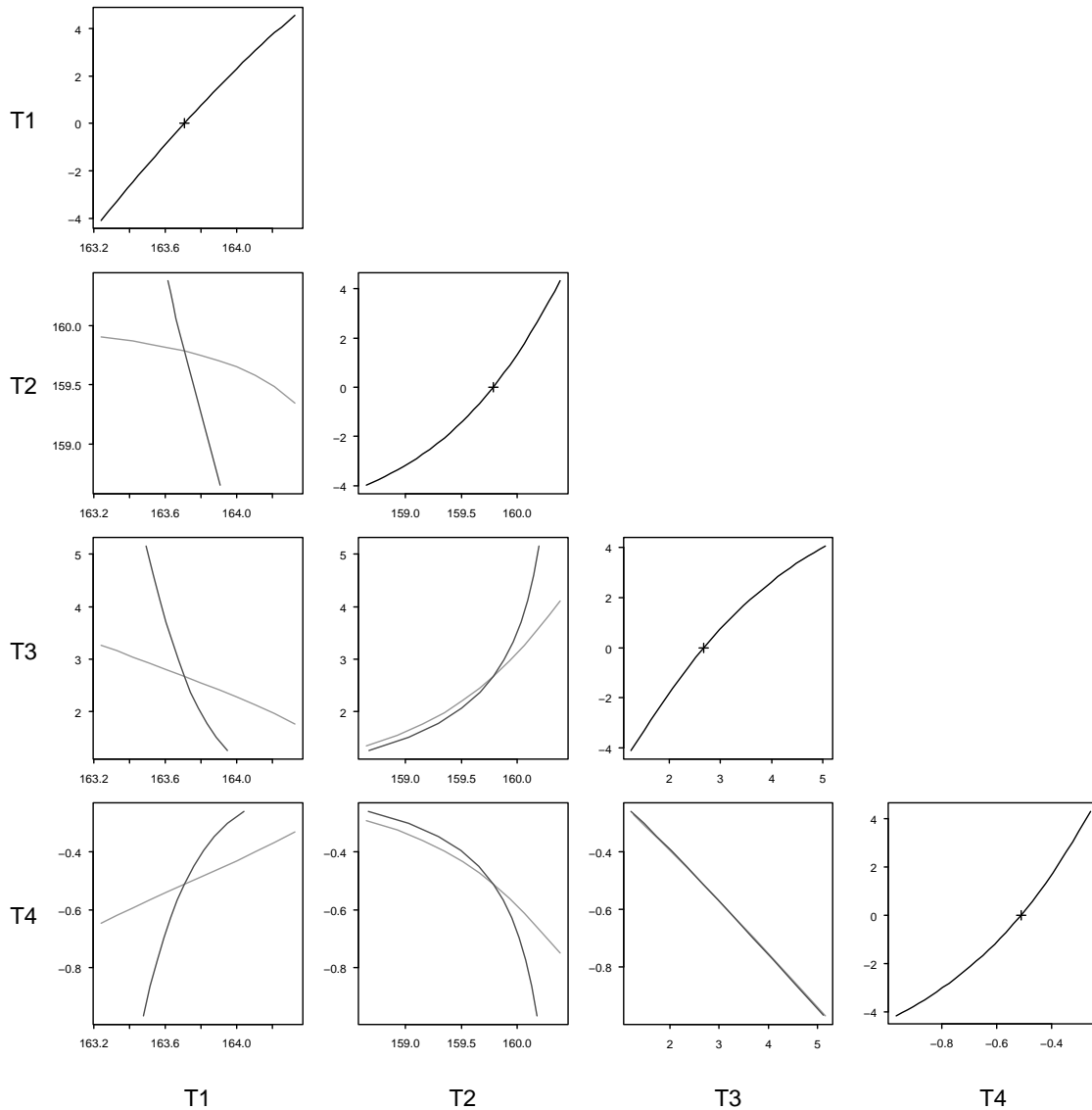


Abbildung 1.6.f: Profile-t-plot und Profilsuren für das Beispiel aus der Membrantrenn-Technologie. Das + im profile-t-plot bezeichnet die Kleinste-Quadrate Lösung.

Die meisten Profilsuren sind stark gekrümmt, d.h. die Regressionsfunktion neigt zu einer starken Nichtlinearität in der Nähe des geschätzten Parameterwertes. Obwohl die Profilsuren für θ_3 und θ_4 gerade sind, zeigt sich ein weiteres Problem: Die Profilsuren liegen aufeinander! Dies bedeutet, dass die Parameter θ_3 und θ_4 extrem stark kollinear sind. Parameter θ_2 ist mit θ_3 und θ_4 ebenfalls kollinear, wenn auch schwächer.

g^* Die Profilsuren können benutzt werden, um sehr genaue Approximationen für zweidimensionale Likelihood-Konturen zu konstruieren (siehe Bates and Watts (1988)). Ihre Berechnung ist rechnerisch weniger aufwändig als jene für die entsprechenden exakten Likelihood-Konturen.

1.7 Parameter-Transformationen

- a Transformationen der Parameter werden vor allem benutzt, um die lineare Approximation und damit das Konvergenzverhalten und die **Qualität der Vertrauensintervalle zu verbessern**.

Es ist hier ausdrücklich festzuhalten, dass Parameter-Transformationen, im Gegensatz zu Transformationen der Zielgrösse (vergleiche 1.1.h), den stochastischen Teil des Modells *nicht* verändern. Sie nützen also nichts, wenn die Annahmen über die Verteilung der Zufallsabweichungen verletzt sind. Es ist die Qualität der linearen Approximation und der darauf beruhenden statistischen Aussagen, die damit geändert werden.

Manchmal sind die transformierten Parameter für die Anwendung schlecht **interpretierbar**. Die wichtigen Fragestellungen betreffen oft einzelne Parameter – und zwar die ursprünglichen. Trotzdem kann man mit Transformationen arbeiten: Man leitet genauere Vertrauensbereiche für die transformierten Parameter her und transformiert diese zurück, um Resultate für die ursprünglichen Parameter zu erhalten.

- b Oft ist der **zulässige Bereich** eines Parameters eingeschränkt, beispielsweise, weil die Regressionsfunktion nur für positive Werte eines Parameters überhaupt definiert ist. Üblicherweise wird eine solche Nebenbedingung zuerst einmal ignoriert und abgewartet, ob und wohin der Algorithmus konvergiert. Erfahrungsgemäss landet die Parameterschätzung in einem sinnvollen Bereich, wenn das Modell die Daten gut beschreibt und die Daten umgekehrt für die Bestimmung der Parameter genügend Information liefern.

Manchmal treten aber im Lauf der Berechnungen Probleme auf, besonders wenn der Parameterwert, der am besten zu den Daten passt, nahe beim Rand des zulässigen Bereichs liegt. Die einfachste Art, mit solchen Problemen fertig zu werden, führt über die Transformation des Parameters.

Beispiele:

- Der Parameter θ soll positiv sein. Durch die Transformation $\theta \rightarrow \phi = \ln(\theta)$ ist $\theta = \exp(\phi)$ immer positiv für alle mögliche Werte von $\phi \in \mathbb{R}$.
- Der Parameter soll im Intervall (a, b) liegen. Mit der logistischen Transformation $\theta = a + (b - a) / (1 + \exp(-\phi))$ kann θ für beliebiges ϕ nur Werte in (a, b) annehmen.
- Im Modell

$$h(x, \underline{\theta}) = \theta_1 \exp(-\theta_2 x) + \theta_3 \exp(-\theta_4 x)$$

mit $\theta_2, \theta_4 > 0$ sind die Parameterpaare $[\theta_1, \theta_2]$ und $[\theta_3, \theta_4]$ austauschbar, d.h. $h(x, \underline{\theta})$ ändert sich beim Vertauschen nicht. Dies kann unangenehme Optimierungsprobleme erzeugen, weil u.a. die Lösungen nicht eindeutig ist. Die Nebenbedingung $0 < \theta_2 < \theta_4$, die die Eindeutigkeit sicherstellt, wird durch die Transformationen $\theta_2 = \exp(\phi_2)$ und $\theta_4 = \exp(\phi_2) (1 + \exp(\phi_4))$ erreicht. Die Funktion lautet nun

$$h(x, (\theta_1, \phi_2, \theta_3, \phi_4)) = \theta_1 \exp(-\exp(\phi_2) x) + \theta_3 \exp(-\exp(\phi_2) (1 + \exp(\phi_4)) x).$$

- c Eine simultane Variablen- und Parameter-Transformation kann hilfreich sein, um **Kollinearität** in den partiellen Ableitungsvektoren abzuschwächen. So hat z.B. das Modell $h(x, \underline{\theta}) = \theta_1 \exp(-\theta_2 x)$ die Ableitungen

$$\frac{\partial h}{\partial \theta_1} = \exp(-\theta_2 x), \quad \frac{\partial h}{\partial \theta_2} = -\theta_1 x \exp(-\theta_2 x)$$

Falls alle x -Werte positiv sind, neigen die beiden Vektoren

$$\begin{aligned}\underline{a}_1 &:= (\exp\langle -\theta_2 x_1 \rangle, \dots, \exp\langle -\theta_2 x_n \rangle)^T \\ \underline{a}_2 &:= (-\theta_1 x_1 \exp\langle -\theta_2 x_1 \rangle, \dots, -\theta_1 x_n \exp\langle -\theta_2 x_n \rangle)^T\end{aligned}$$

zu störender Kollinearität. Diese Kollinearität kann durch **Zentrieren** vermieden werden. Das Modell lässt sich als $h\langle x, \underline{\theta} \rangle = \theta_1 \exp\langle -\theta_2(x - x_0 + x_0) \rangle$ schreiben. Mit der Reparametrisierung $\phi_1 := \theta_1 \exp\langle -\theta_2 x_0 \rangle$ und $\phi_2 := \theta_2$ erhalten wir

$$h\langle x, \underline{\phi} \rangle = \phi_1 \exp\langle -\phi_2(x - x_0) \rangle .$$

Die Ableitungsvektoren werden ungefähr orthogonal, wenn für x_0 der Mittelwert der x_i gewählt wird.

- d ▷ Im **Beispiel aus der Membrantrenn-Technologie** ist aus der approximativen Korrelationsmatrix (Tabelle 1.7.d, linke Hälfte) ersichtlich, dass die Parameter θ_3 und θ_4 stark korreliert sind. (Diese Erkenntnis haben wir schon in 1.6.f aus den Profilsuren gewonnen).

	T1	T2	T3		T1	T2	TT3
T2	-0.256			T2	-0.256		
T3	-0.434	0.771		TT3	0.323	0.679	
T4	0.515	-0.708	-0.989	T4	0.515	-0.708	-0.312

Tabelle 1.7.d: Korrelationsmatrizen für das Beispiel aus der Membrantrenn-Technologie, für die ursprünglichen Parameter (links) und mit transformiertem Parameter $\text{TT3} = \tilde{\theta}_3$ (rechts).

Wenn das Modell reparametrisiert wird zu

$$y_i = \frac{\theta_1 + \theta_2 10^{\tilde{\theta}_3 + \theta_4(x_i - \text{med}\langle x_j \rangle)}}{1 + 10^{\tilde{\theta}_3 + \theta_4(x_i - \text{med}\langle x_j \rangle)}} + E_i, \quad i = 1 \dots n$$

mit $\tilde{\theta}_3 = \theta_3 + \theta_4 \text{med}\langle x_j \rangle$, wird eine Verbesserung erreicht (rechte Hälfte von Tabelle 1.7.d).

◁

- e In Abschnitt 1.6 haben wir Mittel zur graphischen Beurteilung der **linearen Approximation** vorgestellt. Falls die Approximation als ungenügend betrachtet wird, möchten wir diese gerne verbessern. Eine geeignete Reparametrisierung kann dazu viel beitragen.
- f ▷ **Beispiel aus der Membrantrenn-Technologie.** Die in 1.7.d angegebene Parameter-Transformation führt zu einem befriedigendem Resultat, was die Korrelation betrifft. Betrachtet man die Likelihood-Konturen oder den profile-t-plot und die Profilsuren, so ist die Parametrisierung immer noch nicht zufriedenstellend.

Eine intensive Suche nach weiteren Verbesserungen führte zu den folgenden Transformationen, für die die Profilsuren befriedigend ausfallen (Figur 1.7.f):

$$\begin{aligned}\phi_1 &:= \frac{\theta_1 + \theta_2 10^{\phi_3}}{10^{\phi_3} + 1}, & \phi_2 &:= \log_{10} \left(\frac{\theta_1 - \theta_2}{10^{\phi_3} + 1} 10^{\phi_3} \right) \\ \phi_3 &:= \theta_3 + \theta_4 \text{med}\langle x_j \rangle, & \phi_4 &:= 10^{\theta_4}.\end{aligned}$$

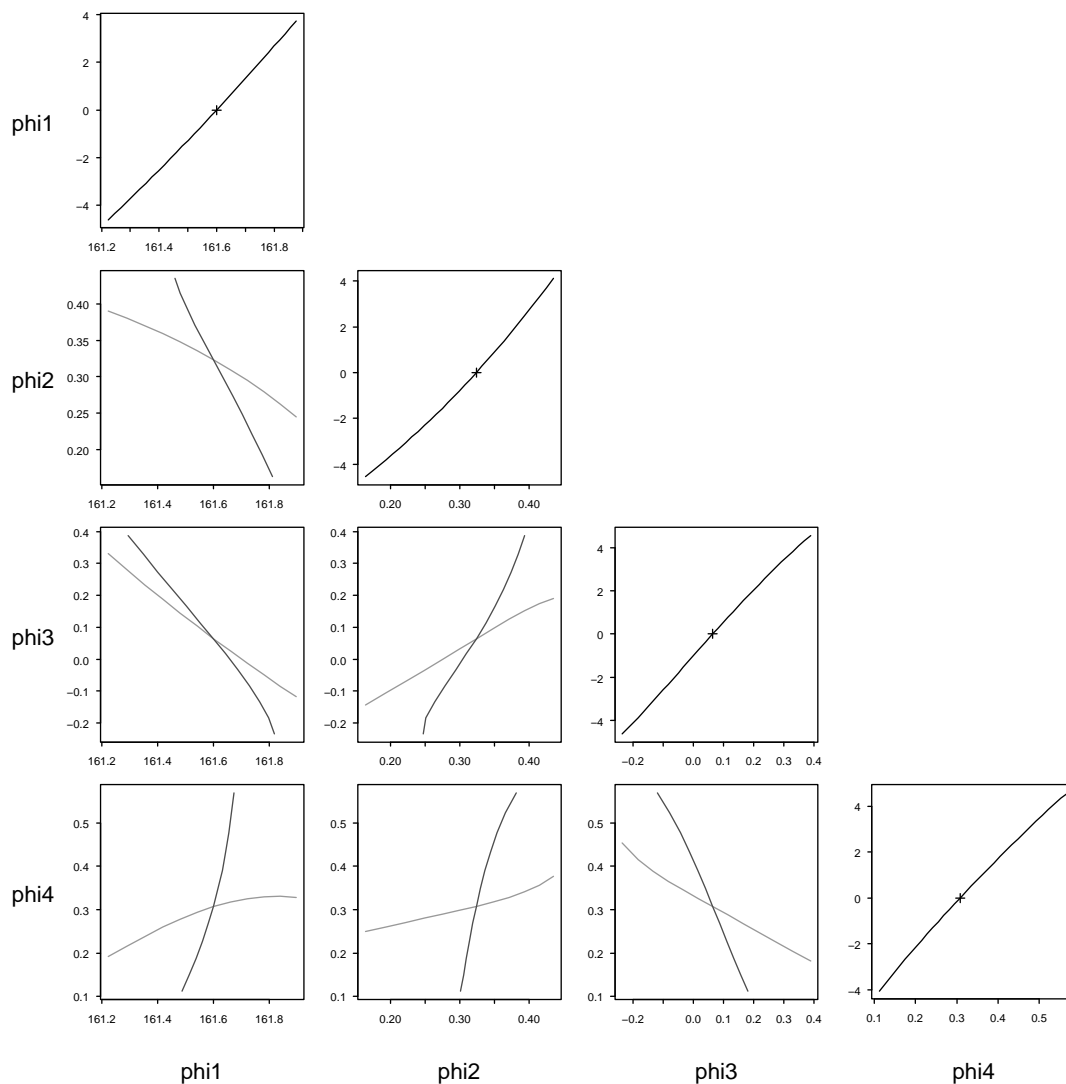


Abbildung 1.7.f: Profile-t-plot und Profilsuren für das Beispiel aus der Membrantrenntechnologie nach den angegebenen Transformationen.

Das Modell lautet dann

$$Y_i = \phi_1 + 10^{\phi_2} \frac{1 - \phi_4^{(x_i - \text{med}(x_j))}}{1 + 10^{\phi_3} \phi_4^{(x_i - \text{med}(x_j))}} + E_i,$$

und man erhält die in Tabelle 1.7.f gezeigten Resultate.

◁

g Es zeigt sich, dass eine erfolgreiche **Reparametrisierung vom Datensatz abhängt**, unter anderem, da die Nichtlinearitäten und Korrelationen zwischen geschätzten Parametern vom (geschätzten) Parametervektor selber abhängen. Deshalb können keine allgemein gültigen Rezepte angegeben werden, was die Suche nach geeigneten Reparametrisierungen oft sehr mühsam macht.

h Obwohl eine Parameter-Transformation uns hilft, Schwierigkeiten mit dem Konvergenzverhalten des Algorithmus oder der Qualität der Vertrauensintervalle zu umgehen, haben doch oft die **ursprünglichen Parameter** eine einfachere physikalische Interpretation. Nehmen wir das einfache Transformationsbeispiel $\theta \longrightarrow \phi = \ln(\theta)$ aus 1.7.b. Die Anpassung des


```

Formula: delta ~ phi1 + 10^phi2 * (1 - phi4^(I(pH - t.x0)))
          / (1 + 10^phi3 * phi4^(I(pH - t.x0)))

Parameters:
      Estimate Std. Error t value Pr(>|t|)
phi1 161.6001      0.0739 2187.12 < 2e-16 ***
phi2  0.3234       0.0313  10.32 3.7e-12 ***
phi3  0.0644       0.0595   1.08  0.29
phi4  0.3077       0.0498   6.18 4.5e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.293 on 35 degrees of freedom

Correlation of Parameter Estimates:
      phi1  phi2  phi3
phi2 -0.561
phi3 -0.766 0.641
phi4  0.151 0.354 -0.312

```

Tabelle 1.7.f: Computer-Ausgabe für das Beispiel aus der Membrantrenn-Technologie nach Parameter-Transformation

Models mündet in eine Schätzung $\hat{\phi}$ mit geschätztem Standardfehler $\hat{\sigma}_{\hat{\phi}}$. Eine naheliegende Schätzung für θ ist dann $\hat{\theta} = \exp\langle \hat{\phi} \rangle$. Das zurücktransformierte Vertrauensintervall ist von der Form „Schätzwert mal/durch Faktor“,

$$\exp\langle \hat{\phi} \rangle \times / \tau \quad \text{mit} \quad \tau = \exp\langle \hat{\sigma}_{\hat{\phi}} q_{0.975}^{t_{n-p}} \rangle .$$

- i ▷ Warum haben wir so viele Schwierigkeiten mit dem **Beispiel Sauerstoffverbrauch**? Betrachten wir die Abbildung 1.1.e und erinnern uns, dass der Parameter θ_1 den erwarteten Sauerstoffverbrauch bei unendlicher Inkubationszeit repräsentiert, so ist klar, dass θ_1 schwierig zu schätzen ist, weil die horizontale Asymptote durch die Daten schlecht bestimmt ist. Hätten wir noch weitere Beobachtungen mit längeren Inkubationszeiten, so hätten wir die Schwierigkeiten mit der Qualität des Vertrauensintervalles von θ vermeiden können. ◁

Gerade auch bei nichtlinearen Modellen ist eine gute (statistische) **Versuchsplanung** (experimental design) unerlässlich. Der Informationsgehalt der Daten wird durch die Wahl der Versuchsbedingungen festgelegt, und kein (statistisches) Verfahren ist in der Lage, Informationen über das Modell zu liefern, welche in den Daten nicht enthalten sind.

1.8 Literatur

- a Diese Unterlagen beruhen vor allem auf dem Buch von Bates and Watts (1988). Eine mathematischere Diskussion über die statistischen und numerischen Methoden in der nichtlinearen Regression findet sich in Seber and Wild (1989). Das Buch von Ratkowsky (1989) zählt zahlreiche mögliche nichtlineare Funktionen $h(\cdot)$ auf, die vor allem in biologischen Bereichen ihre Anwendung finden.
- b Seit einiger Zeit wird zur Bestimmung von Vertrauensbereichen auch der **Bootstrap** benutzt, siehe Huet, Bouvier, Gruet and Jolivet (1996). In diesem Buch wird auch der Fall mit **nichtkonstanter Varianz** (heteroskedastic models) besprochen. Dazu lohnt auch ein Blick in das Buch von Carroll and Ruppert (1988).
- c Heutzutage enthalten die meisten **Statistik-Pakete** eine Prozedur, welche nichtlineare Modelle anpassen und asymptotische Vertrauensintervalle für die Parameter berechnen kann. Prinzipiell ist es dann auch möglich, „t-Profile“ und Profilsuren zu berechnen, weil sie auch auf dem Anpassen von nichtlinearen Modellen, jedoch mit einem Parameter weniger, beruhen.

In den beiden Implementationen **S-Plus** und **R** der Statistik-Sprache **S** ist die Funktion **nls** verfügbar, die auf der Arbeit von Bates and Watts (1988) basiert. Zusammenfassungen der Methode finden sich deshalb in Chambers and Hastie (1992, Kapitel 10), in den **S-Plus-Manuals** oder in Venables and Ripley (1994). Die „Library“ **nlme** enthält **S-Funktionen**, die nichtlineare Regressionsmodelle mit korrelierten Fehlern (**gnls**) und zufälligen Effekten (**nlme**) an Daten anpassen können. Diese Implementationen basieren auf dem Buch „Mixed Effects Models in S and S-Plus“ von Pinheiro and Bates (2000).

Literaturverzeichnis

- Bard, Y. (1974). *Nonlinear parameter estimation*, Academic Press, N.Y.
- Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*, Wiley, N.Y.
- Bennett, J. H. (ed.) (1971-74). *Collected Papers of R. A. Fisher. 5 Volumes*, Univ. Adelaide, Australia.
- Boen, J. R. and Zahn, D. A. (1982). *The Human Side of Statistical Consulting*, Wadsworth, Belmont, Cal.
- Bortz, J. (2005). *Statistik für Sozialwissenschaftler*, 6. Aufl., Springer, Berlin.
- Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*, Wiley Series in Probability and Mathematical Statistics, Wiley, N.Y.
- Box, G. E. P., Hunter, W. G. and Hunter, J. S. (2005). *Statistics for Experimenters*, 2nd edn, Wiley, Hoboken, N.J.
- Brown, P. J. (1993). *Measurement, Regression, and Calibration*, Clarendon Press, Oxford, UK.
- Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Wiley, New York.
- Chambers, J. M. and Hastie, T. J. (1992). *Statistical Models in S*, Wadsworth & Brooks/Cole, Pacific Grove, Cal.
- Chatfield, C. (1996). *The Analysis of Time Series. An Introduction*, 5th edn, Chapman and Hall, London.
- Chatterjee, S. and Price, B. (2000). *Regression Analysis By Example*, 3rd edn, Wiley, N.Y.
- Constantinides, A. and Mostoufi, N. (1999). *Numerical Methods for Chemical Engineers with Matlab Applications*, Prentice Hall.
- Cook, R. D. and Weisberg, S. (1999). *Applied regression including computing and graphics*, Wiley, N.Y.
- Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*, Wiley, N.Y.
- Daniel, C. and Wood, F. S. (1980). *Fitting Equations to Data*, 2nd edn, Wiley, N.Y.
- Devore, J. L. (2004). *Probability and Statistics for Engineering and the Sciences*, 6th edn, Duxbury Press, Belmont, California.
- Draper, N. and Smith, H. (1998). *Applied Regression Analysis*, 3rd edn, Wiley, N.Y.
- Englezos, P. and Kalogerakis, N. (2001). *Applied parameter estimation for chemical engineers*, Marcel Dekker, N.Y.
- Federer, W. T. (1972, 1991). *Statistics and Society: Data Collection and Interpretation*, Statistics: Textbooks and Monographs, Vol.117, 2nd edn, Marcel Dekker, N.Y.
- Ferraris, G. B. and Donati, G. (1971). Analysis of the kinetic models for the reaction of synthesis of methanol, *Ing. Chim. Ital.* **7**: 53–64.

- Ferraris, G. B., Donati, G., Rejna, F. and Caprà, S. (1974). An investigation on kinetic models for ammonia synthesis, *Chemical Engineering Science* **29**: 1621–1627.
- Fisher, R. A. (1925-62). *Collected Papers*, siehe Bennet, 1971-74.
- Fox, J. (2002). *An R and S-Plus companion to applied regression*, Sage, Thousand Oaks, CA.
- Haaland, P. D. (1989). *Experimental Design in Biotechnology*, Marcel Dekker, N.Y.
- Harman, H. H. (1960, 1976). *Modern Factor Analysis*, 3rd edn, University of Chicago Press, Chicago.
- Harrell, F. E. J. (2002). *Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer Series in Statistics, Springer, NY. Corrected second printing
- Hartung, J., Elpelt, B. und Klösener, K. (2002). *Statistik. Lehr- und Handbuch der angewandten Statistik*, 13. Aufl., Oldenbourg, München.
- Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (eds) (1991). *Fundamentals of Exploratory Analysis of Variance*, Wiley Series in Probability and Mathematical Statistics, Wiley, N.Y.
- Hocking, R. R. (1996). *Methods and Applications of Linear Models; Regression and the Analysis of Variance*, Wiley Series in Probability and Statistics, Wiley, N.Y.
- Hogg, R. V. and Ledolter, J. (1992). *Applied Statistics for Engineers and Physical Scientists*, 2nd edn, Maxwell Macmillan International Editions.
- Huet, S., Bouvier, A., Gruet, M.-A. and Jolivet, E. (1996). *Statistical Tools for Nonlinear Regression: A Practical Guide with S-Plus Examples*, Springer-Verlag, New York.
- Lawley, D. N. and Maxwell, A. E. (1963, 1967). *Factor Analysis as a Statistical Method*, Butterworths Mathematical Texts, Butterworths, London.
- Linder, A. und Berchtold, W. (1982). *Statistische Methoden II: Varianzanalyse und Regressionsrechnung*, Birkhäuser, Basel.
- Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics, Theory and Methods*, Wiley Series in Probability and Statistics, Wiley, Chichester, England.
- Mead, R. (1988). *The design of experiments*, Cambridge University Press, Cambridge.
- Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading, Massachusetts.
- Myers, R. H. and Montgomery, D. C. (1995). *Response Surface Methodology; Process and Product Optimization Using Designed Experiments*, Wiley Series in Probability and Statistics, Wiley, NY.
- Petersen, R. G. (1985). *Design and Analysis of Experiments*, Statistics Textbooks and Monographs, Marcel Dekker, N.Y.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus*, Statistics and Computing, Springer, N.Y.
- Pokropp, F. (1994). *Lineare Regression und Varianzanalyse*, Oldenbourg.
- Rapold-Nydegger, I. (1994). *Untersuchungen zum Diffusionsverhalten von Anionen in carboxylierten Cellulosemembranen*, PhD thesis, ETH Zurich.
- Rasch, D., Guiard, V. und Nürnberg, G. (1992). *Statistische Versuchsplanung: Einführung in die Methoden und Anwendung des Dialogsystems CADEMO*, Gustav Fischer, Stuttgart.
- Ratkowsky, D. A. (1989). *Handbook of Nonlinear Regression Models*, Marcel Dekker, New York.
- Reichert, P. (1994). Aquasim – a tool for simulation and data analysis of aquatic systems, *Water Science Tech.* **30**(2): 21–30.

- Renner, R. M. (1993). The resolution of a compositional data set into mixtures of fixed source compositions, *Applied Statistics — Journal of the Royal Statistical Society C* **42**: 615–631.
- Rice, J. A. (2007). *Mathematical Statistics and Data Analysis*, 3rd edn, Duxbury Press, Belmont, California.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression & Outlier Detection*, Wiley, N.Y.
- Ryan, T. P. (1997). *Modern Regression Methods*, Series in Probability and Statistics, Wiley, N.Y. includes disk
- Sachs, L. (2004). *Angewandte Statistik*, 11. Aufl., Springer, Berlin.
- Scheffé, H. (1959). *The Analysis of Variance*, Wiley, N.Y.
- Schittkowski, K. (1994). Parameter estimation in systems of nonlinear equations, *Numerische Mathematik* **68**: 129–142.
- Schlittgen, R. (2003). *Einführung in die Statistik. Analyse und Modellierung von Daten*, 10. Aufl., Oldenbourg, München. *schoen, inkl. Sensitivity und breakdown, einfache regr mit resanal*
- Seber, G. and Wild, C. (1989). *Nonlinear regression*, Wiley, New York.
- Sen, A. and Srivastava, M. (1990). *Regression Analysis; Theory, Methods, and Applications*, Springer-Verlag, N.Y.
- Stahel, W. A. (2000). *Statistische Datenanalyse: Eine Einführung für Naturwissenschaftler*, 3. Aufl., Vieweg, Wiesbaden.
- Swinbourne, E. S. (1971). *Analysis of Kinetic Data*, Nelson, London.
- Venables, W. N. and Ripley, B. (1994). *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.
- Weisberg, S. (2005). *Applied Linear Regression*, 3rd edn, Wiley, N.Y.
- Wetherill, G. (1986). *Regression Analysis with Applications*, number 27 in *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.