# Aufgaben zur Regressionsrechnung

## 1 Lineare Regression

Von 4 Kfz sind das Alter und die Bremswege bei einer Vollbremsung von 100 km/h zum Stillstand gegeben:

| Alter (Jahre) | 4 | 7 | 11 | 2 |
|---|---|---|---|---|
| Bremsweg (m) | 50 | 80 | 70 | 45 |

(a) Bestimmen Sie die Koeffizienten der linearen Regression

(b) Extrapolieren Sie den erwarteten mittleren Bremsweg für 15 Jahre alte Fahrzeuge

(c) Zeichnen Sie ein Streudiagramm der Daten und zeichnen Sie die Regression ein.

## 2 Quasilineare Regression

Von 4 Kfz-Haltern sind im letzten Jahr die Kfz-Haltungskosten pro km und die Kilometerleistung bekannt:

| Haltungskosten ( € /km) | 0.5 | 0.2 | 0.35 | 0.4 |
|---|---|---|---|---|
| Fahrleistung (km) | 14000 | 42000 | 14000 | 25000 |

(a) Bestimmen Sie die Koeffizienten der Regression $\hat{y}(x) = a + b/x$

(b) Bestimmen Sie die mittleren Fixkosten, die variablen Kosten sowie die Preiselastizität bei einer Fahrleistung von 20000 km

(c) Zeichnen Sie die Daten und die Regressionsfunktion in ein Diagramm ein.

## 3 Nichtlineare Regression

Der Weltenergieverbrauch in Exajoule (EJ, $10^{18}$ J) veränderte sich wie folgt:

| Jahr | Weltenergieverbrauch (EJ) |
|---|---|
| 1900 | 15 |
| 1920 | 30 |
| 1940 | 50 |
| 1960 | 130 |
| 1972 | 230 |

(a) Führen Sie die nichtlineare Regression mit der Funktion

$$\hat{y}(x) = ae^{bx}$$

durch.

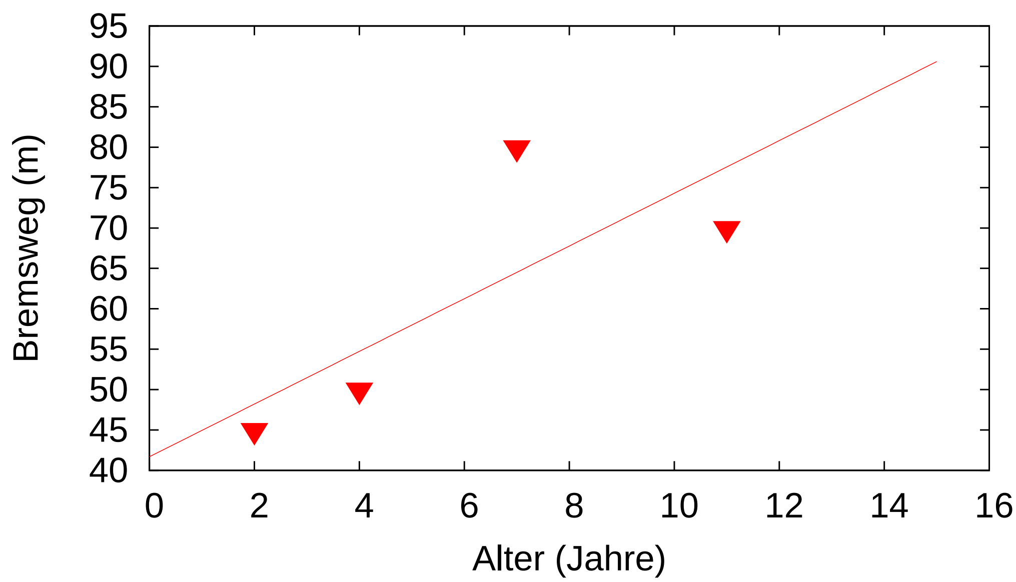
(b) Wie hoch ist der mit diesen Zahlen prognostizierte Verbrauch 1980 und 2020?

# Lineare Regression, Lösungen

(a) $\hat{y} = a + bx, \;\; a = 41.6848$ m, $b = 3.26087$ m/Jahr

(b) Erwarteter Bremsweg: $\hat{y}(15) = a + 15b = 90.6$ m

(c) Scatterplot mit Fitgeraden:

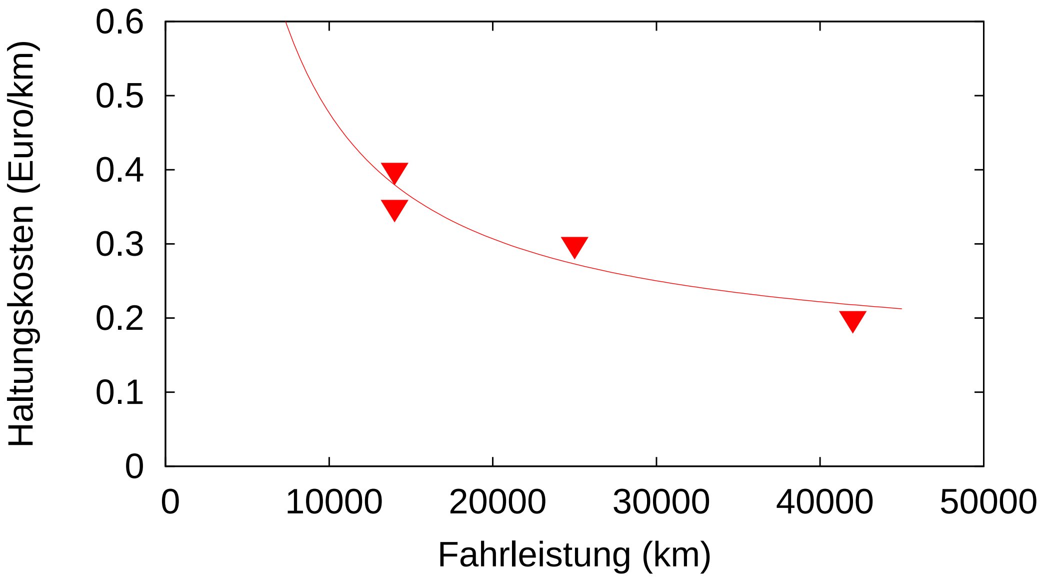


## Quasilineare Regression, Lösungen

(a) $\hat{y} = a + b/x$, $a = 0.137$ €, $b = 3396$ € km

(b) Fixkosten: 3396 €; Variablen Kosten: 0.137 €/km

(c) Scatterplot mit Fitgeraden:

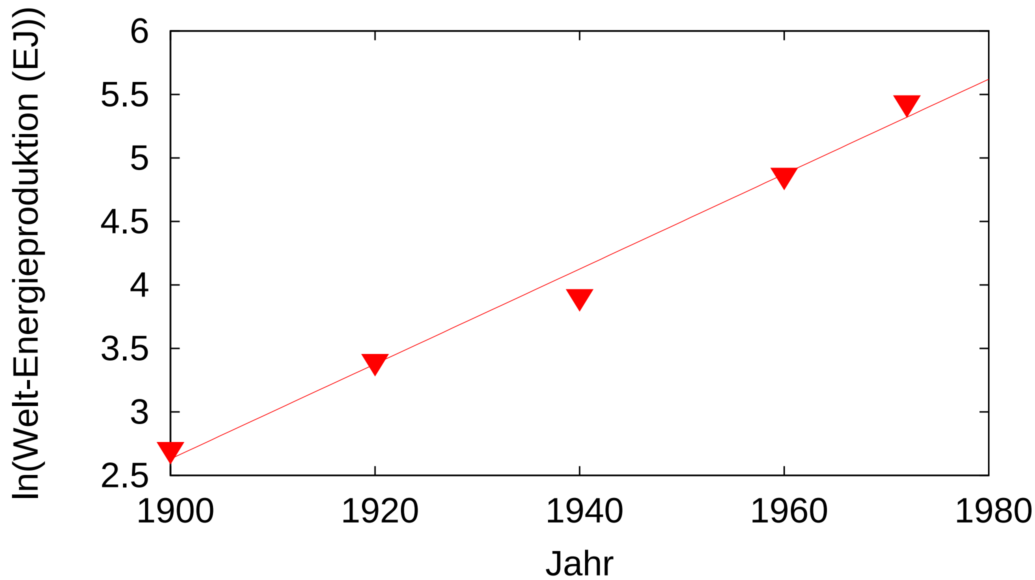


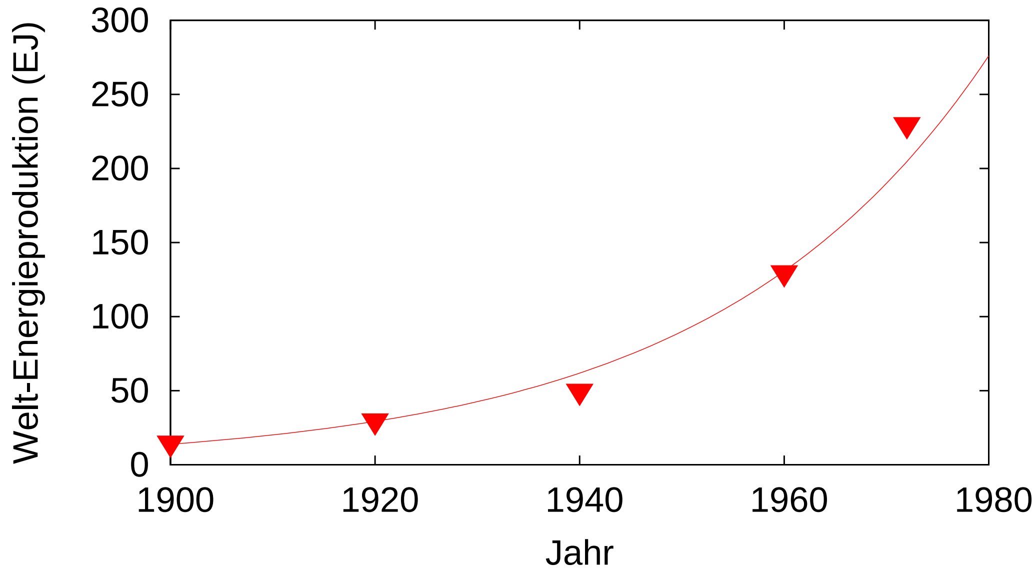
## Nichtlineare Regression, Lösungen

(a) $\hat{y}(x) = ae^{b(x-1900)}, \quad a = 13.9 \text{ EJ}, \; b = 0.037/\text{Jahr}$

(b) $\hat{y}(1980) = ae^{80b} = 13.9 * 19.3 = 268 \text{ EJ}$

(c) Scatterplot mit Fitgeraden:



# 4 Hintergrundinformation: Regression der Originaldaten vs. Regression transformierter Daten (vgl. Kap. 13.3)

Anhand zweier Regressionsfunktionen zeige ich für allgemeine Daten, dass

- eine *quasilineare* Regression der Originaldaten *äquivalent* zu einer linearen Regression von geeignet transformierten Daten ist,

- eine *auch in den Koeffizienten nichtlineare* Regression der Originaldaten i.A. *nicht* äquivalent zu einer linearen Regression von geeignet transformierten Daten ist.

## 4.1 Beispiel 1: Quasilineare Regression

Es sollen als Streudiagramm

$$\{(x_i, y_i), i = 1, \cdots, n\}$$

gegebene Daten durch die hyperbolische Regressionsfunktion

$$\hat{y}(x) = a + \frac{b}{x}$$

angeglichen werden.

### 4.1.1 Lineare Regression der transformierten Daten

Mit folgender Transformation der unabhängigen Variablen:

$$x = \frac{1}{z}$$

wird die Regressionsfunktion $\hat{y}(z) = a + bz$ linear und man erhält *in der neuen unabhängigen Variable $z$* die üblichen Ausdrücke für die Koeffizienten:

$$a = \bar{y} - b\bar{z}, \quad b = \frac{s_{zy}}{s_z^2}.$$

Nach Rücktransformation $z = \frac{1}{x}$ erhält man für $a$ und $b$ folgende Ausdrücke:

$$a = \bar{y} - \frac{1}{n}\left(\sum_{i=1}^{n} \frac{1}{x_i}\right) b, \tag{1}$$

$$b = \frac{\sum_{i=1}^{n} \frac{y_i}{x_i} - \bar{y}\sum_{i=1}^{n} \frac{1}{x_i}}{\sum_{i=1}^{n} \frac{1}{x_i^2} - \frac{1}{n}\left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^2} \tag{2}$$

### 4.1.2 Direkte Regression

Da $\hat{y}(x)$ quasilinear ist, ist dies ohne Probleme möglich:

$$F = \sum_{i=1}^{n} \left( y_i - a - \frac{b}{x_i} \right)^2 \stackrel{!}{=} \min!$$

$$\Rightarrow$$

$$\begin{aligned}
\frac{\partial F}{\partial a} = 2 \sum_{i=1}^{n} \left( y_i - a - \frac{b}{x_i} \right) (-1) &= 0, \\
\frac{\partial F}{\partial b} = 2 \sum_{i=1}^{n} \left( y_i - a - \frac{b}{x_i} \right) \left( \frac{-1}{x_i} \right) &= 0.
\end{aligned}$$

Daraus ein lineares Gleichungssystem für $a$ und $b$:

$$\begin{aligned}
na + \left( \sum_{i=1}^{n} \frac{1}{x_i} \right) b. &= \sum_{i=1}^{n} y_i, \\
\left( \sum_{i=1}^{n} \frac{1}{x_i} \right) a + \left( \sum_{i=1}^{n} \frac{1}{x_i^2} \right) b &= \sum_{i=1}^{n} \frac{y_i}{x_i}.
\end{aligned}$$

Daraus als Ergebnis die Koeffizienten der direkten Regression:

$$a = \bar{y} - \frac{1}{n} \left( \sum_{i=1}^{n} \frac{1}{x_i} \right) b, \tag{3}$$

$$b = \frac{\sum_{i=1}^{n} \frac{y_i}{x_i} - \bar{y} \sum_{i=1}^{n} \frac{1}{x_i}}{\sum_{i=1}^{n} \frac{1}{x_i^2} - \frac{1}{n} \left( \sum_{i=1}^{n} \frac{1}{x_i} \right) \left( \sum_{i=1}^{n} \frac{1}{x_i} \right)} \tag{4}$$

Die Ausdrücke (3) und (1) für den Koeffizienten $a$ sowie (4) und (2) für $b$ sind identisch! Dies ist insofern anschaulich, als

- einerseits bei der Regressionsrechnung ja unsymmetrisch nur die Fehlerquadratsumme $F$ bezüglich der *abhängigen* Variablen $y$ minimiert wird,

- andererseits eine quasilineare Regressionsfunktion durch Transformation ausschließlich der *unabhängigen* Variablen zur linearen Funktion wird. Im quasilinearen Fall wird die Fehlerquadratsumme $F$ also durch die Transformation gar nicht beeinflusst!

## 4.2 Beispiel 2: In $y$ nichtlineare Regression

Es sollen ein Streudiagramm

$$\{(x_i, y_i), i = 1, \cdots, n\}$$

durch die exponentielle Regressionsfunktion

$$\hat{y}(x) = ae^{bx}$$

angeglichen werden.

### 4.2.1 Lineare Regression der transformierten Daten

Mit der Transformation

$$y = e^z$$

wird die Regressionsfunktion $e^{\hat{z}} = ae^{bx}$ nach Logarithmierung linear:

$$\hat{z}(x) = a' + bx,$$

mit der neuen Regressionskonstante $a' = \ln(a)$.
Mit der neuen *abhängigen* Variablen $z$ erhält man die üblichen Ausdrücke für die Koeffizienten:

$$a' = \bar{z} - b\bar{x}, \quad b = \frac{s_{xz}}{s_x^2}.$$

Nach Rücktransformation $z = \ln y$ erhält man für $a = e^{a'}$ und $b$ folgende Ausdrücke:

$$a = e^{a'} = \frac{\exp\left[\frac{1}{n}\sum_{i=1}^{n} \ln y_i\right]}{\exp(b\bar{x})}, \tag{5}$$

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \left(\ln y_i - \frac{1}{n}\sum_{i=1}^{n} \ln y_i\right)}{\sum_{i=1}^{n} \left(\ln y_i - \frac{1}{n}\sum_{i=1}^{n} \ln y_i\right)^2} \tag{6}$$

### 4.2.2 Direkte nichtlineare Regression

$$F = \sum_{i=1}^{n} \left(y_i - ae^{bx}\right)^2 \stackrel{!}{=} \min!$$

$$\Rightarrow$$

$$\frac{\partial F}{\partial a} = 2\sum_{i=1}^{n} \left(y_i - ae^{bx_i}\right)\left(-e^{bx_i}\right) = 0, \tag{7}$$

$$\frac{\partial F}{\partial b} = 2\sum_{i=1}^{n} \left(y_i - ae^{bx_i}\right)\left(-ae^{bx_i}\right)(x_i) = 0. \tag{8}$$

Dieses nichtlineare Gleichungssystem ist nicht analytisch lösbar, aber man kann an Beispielen sehen, dass die Lösungen $a$ und $b$ i.A. nicht äquivalent zu den Gln (5) und (6) sind: Setzen Sie z.B. die sich für das Streudiagramm $\{(-1, e^{-1}), (0, 1), (1, e^4)\}$ aus (5) und (6) sich ergebenden Parameter $a = e^1$, $b = \frac{5}{14}$ in die Bedingung (7) und (8) ein, erhalten Sie Widersprüche.

**<u>Fazit:</u>** Bei einer nichtlinearen Regressionfunktion, die auch nichtlinear in den Parametern ist, sind die beiden Ansätze i.A. nicht äquivalent; je nach konkreter Situation ist die eine oder die andere Variante "besser"; im Zweifel ist die lineare Regression der transformierten Daten vorzuziehen (vgl. Regression der Langzeit-Dow-Jones-Daten).