Stochastik für Info SoSe 2023
Lineare Regression - Fortsetzung

Hanno Gottschalk

July 5, 2023

Inhaltsverzeichnis Vorlesung

Residuenplots

Residuen

Def. Es seien x_i und y_i die beobachteten Daten (x_i Einflussgröße, y_i Zielgröße) und

$$f(x_i) = \hat{\beta} x_i + \hat{\alpha}$$

die optimale Ausgleichsgrade. Dann heißen die folgenden Werte ($i = 1, \ldots, n$) Residuen der Ausgleichsgrade

$$\epsilon_i = y_i - f(x_i) = \text{Diff. zwischen Vorhersage und Beobachtung} \tag{1}$$

Residuen

Def. Es seien x_i und y_i die beobachteten Daten (x_i Einflussgröße, y_i Zielgröße) und

$$f(x_i) = \hat{\beta} x_i + \hat{\alpha}$$

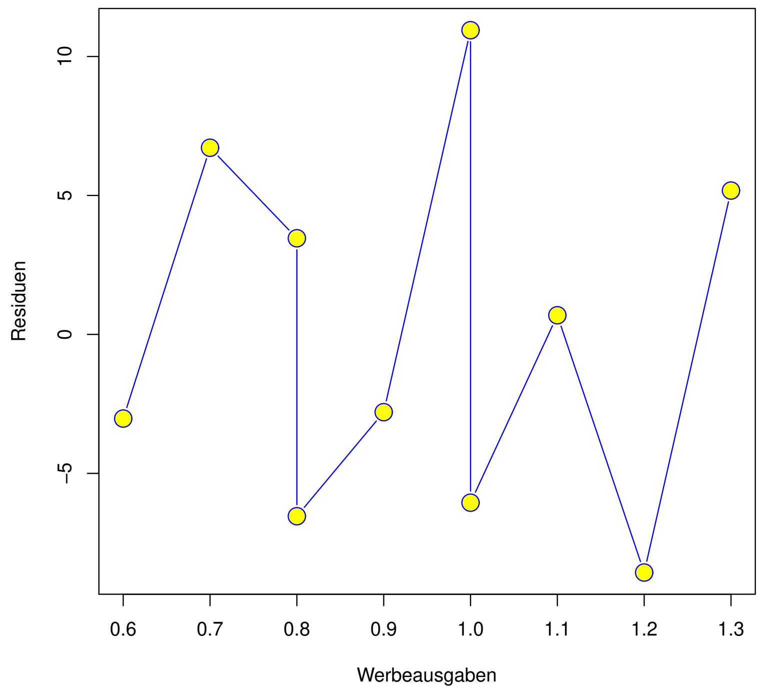
die optimale Ausgleichsgrade. Dann heißen die folgenden Werte ($i = 1, \ldots, n$) Residuen der Ausgleichsgrade

$$\epsilon_i = y_i - f(x_i) = \text{Diff. zwischen Vorhersage und Beobachtung} \quad (1)$$

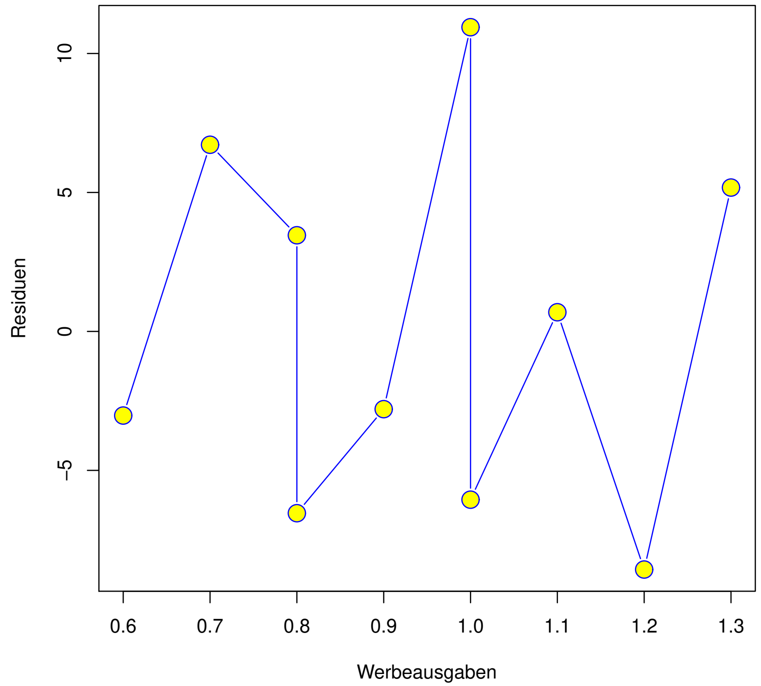
Beispiel: Verkäufe vs Werbeausgaben

Monat	1	2	3	4	5	6	7	8	9	10
Werbeausgaben	1.2	0.8	1.0	1.3	0.7	0.8	1.0	0.6	0.9	1.1
Verkäufe	101	92	110	120	90	82	93	75	91	105
Modellvorhersage=$f(x_i)$	109.56	88.54	99.05	114.82	83.28378	88.54	99.054	78.027	93.79	104.31
Residuum $y_i - f(x_i)$	-8.56	3.45	10.94	5.17	6.71	-6.54	-6.054	-3.02	-2.79	0.68

Der naive Residuenplot

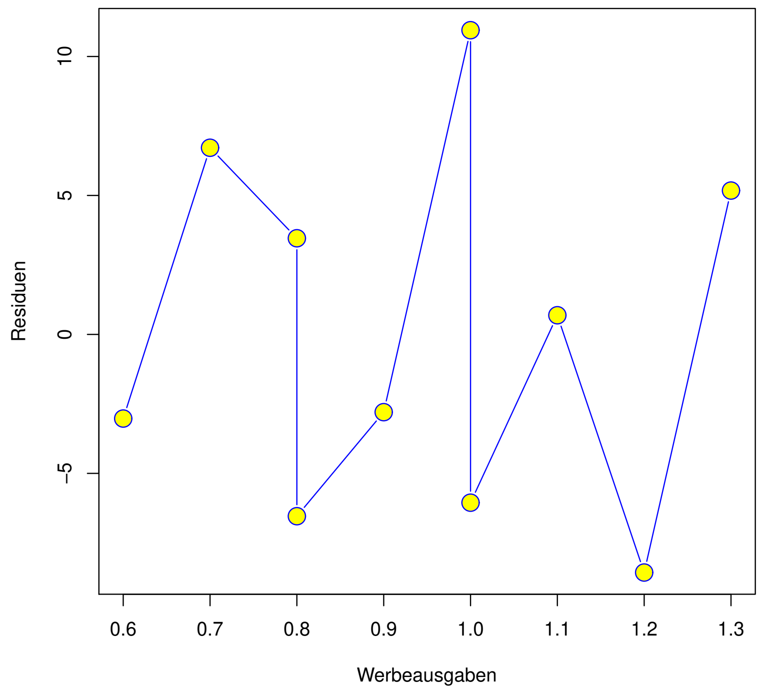


Der naive Residuenplot



Der naive Residuenplot stellt die Residuen über dem Regressor (x-Werte) dar.

Der naive Residuenplot



Der naive Residuenplot stellt die Residuen über dem Regressor (x-Werte) dar.

Hauptaufgabe des Rediduenplots ist das Aufspüren von Trends

Residuals over fitted

Problem beim naiven Residuenplot: Kann nur für einen Regressor (eine erklärende Größe x_i) eingesetzt werden.

Residuals over fitted

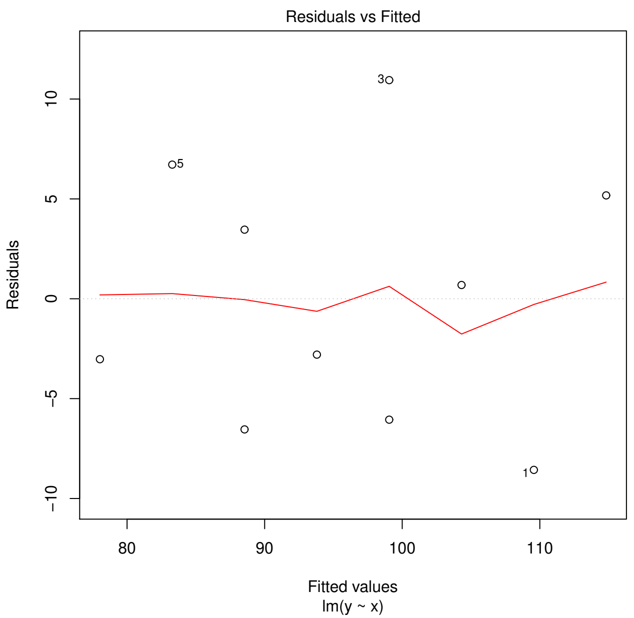
Problem beim naiven Residuenplot: Kann nur für einen Regressor (eine erklärende Größe x_i) eingesetzt werden.

Deshalb wird beim 'residuals over fitted'-plot das Residuum über dem Vorhersagewert dargestellt:
Man berechnet die Residuen $\epsilon_i = y_i - f(x_i)$ und stellt die Punktepaare $(f(x_i), \epsilon_i)$ in einem Streudiagramm dar.

Residuals over fitted

Problem beim naiven Residuenplot: Kann nur für einen Regressor (eine erklärende Größe x_i) eingesetzt werden.

Deshalb wird beim 'residuals over fitted'-plot das Residuum über dem Vorhersagewert dargestellt:
Man berechnet die Residuen $\epsilon_i = y_i - f(x_i)$ und stellt die Punktepaare $(f(x_i), \epsilon_i)$ in einem Streudiagramm dar.



Die rote Trendlinie ist ein nichtparametrischer moving average Trendschätzer:

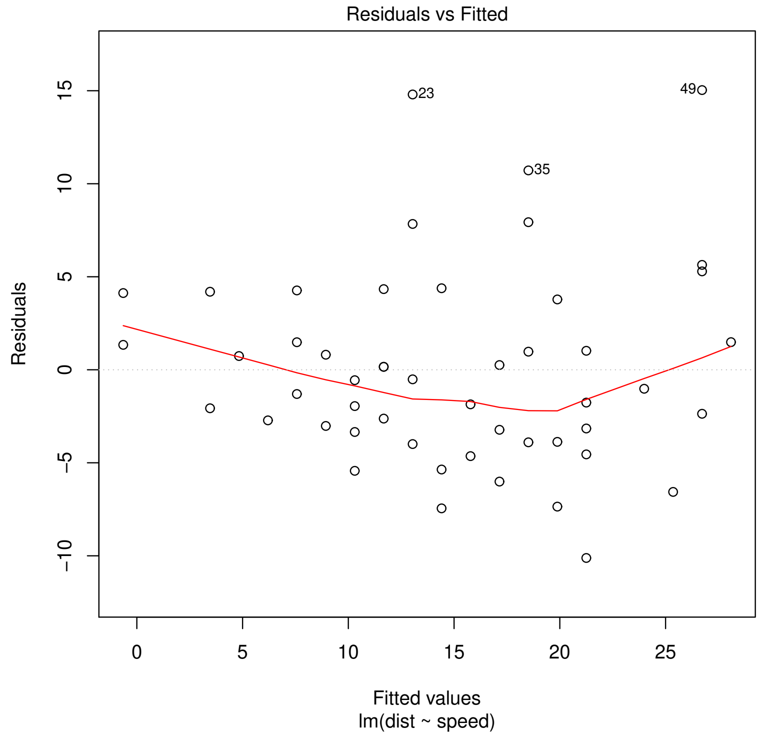
Hier ist kein klarer Trend erkennbar!

Residuals over fitted: Bremsweg

Bei den Bremswegdaten ist im Residuals over fitted Plot ein klarer Trend zu erkennen

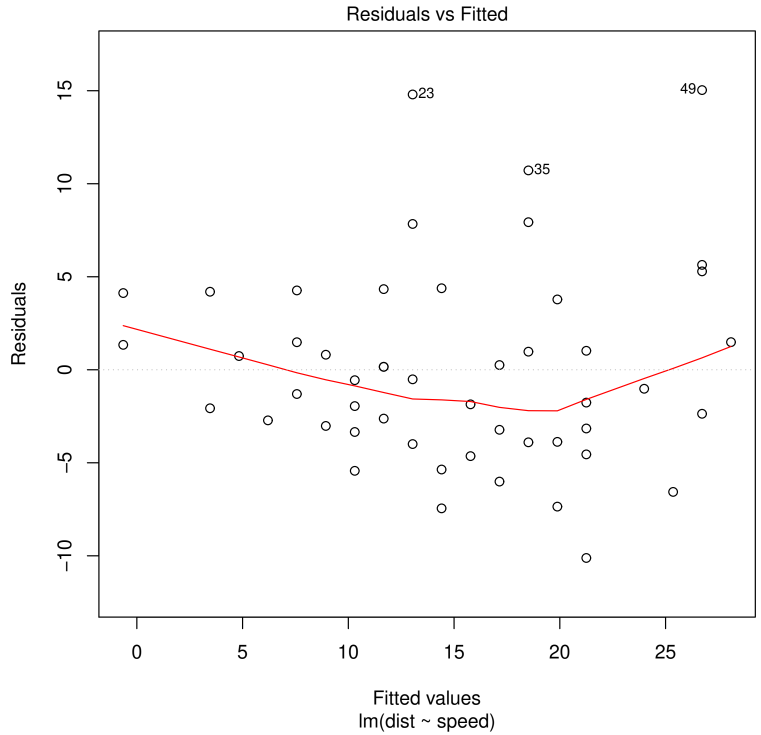
Residuals over fitted: Bremsweg

Bei den Bremswegdaten ist im Residuals over fitted Plot ein klarer Trend zu erkennen



Residuals over fitted: Bremsweg

Bei den Bremswegdaten ist im Residuals over fitted Plot ein klarer Trend zu erkennen



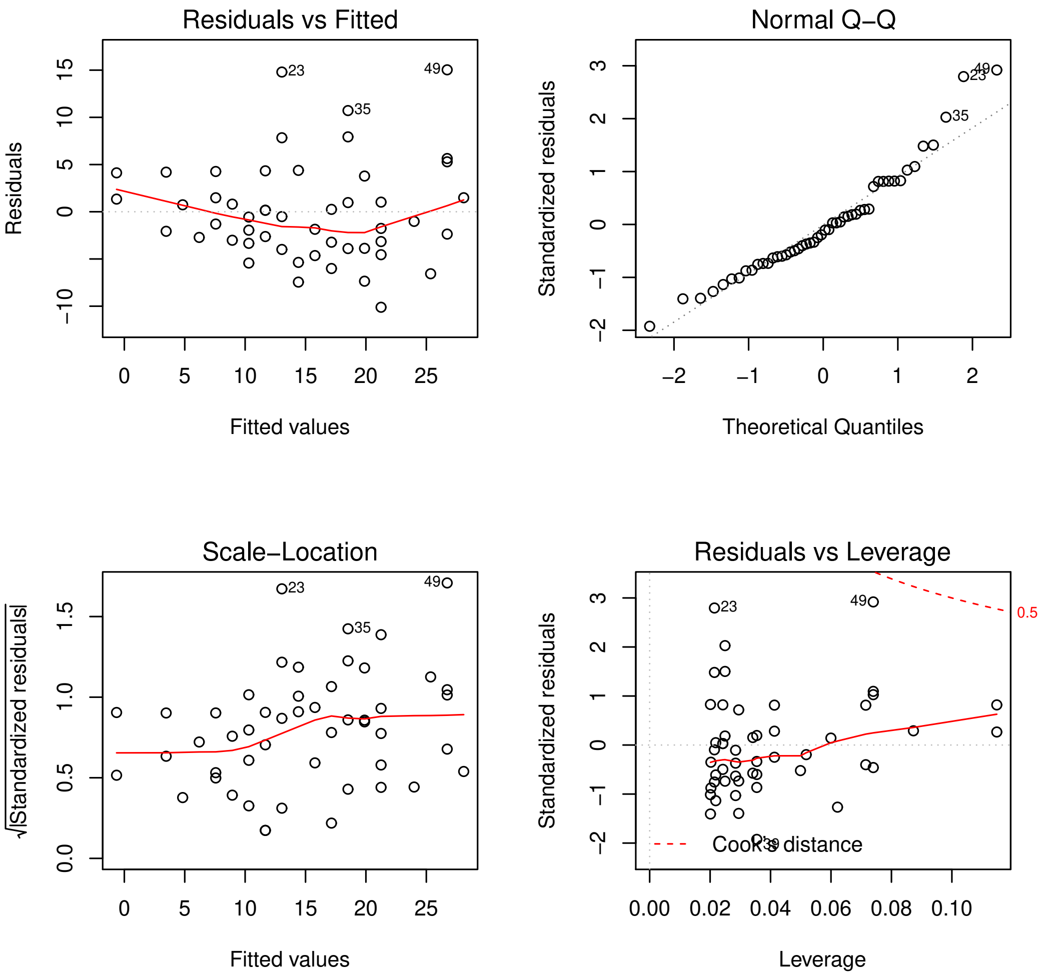
Erklärung: Kin. Energie $E = \frac{1}{2}mv^2$, die durch Reibung verlorene Energie ist jedoch proportional zum Weg $\Rightarrow$ quadr. Abhängigkeit des Bremsweg von Geschwindigkeit!

Andere diagnostische Plots

Professionelle Statistiksoftware offeriert mehrere diagnostische Plots:

Andere diagnostische Plots

Professionelle Statistiksoftware offeriert mehrere diagnostische Plots:



Streuzerlegung

Erklärte und Residuenstreuung

Def.: Die <u>Gesamtstreuung</u> des Merkmals Y ist

$$SQT = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{2}$$

Erklärte und Residuenstreuung

Def.: Die Gesamtstreuung des Merkmals Y ist

$$SQT = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{2}$$

Die Residuenstreuung ist – gegeben den funktionalen Zusammenhang $f(X)$ zur Vorhersage von Y

$$SQR = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{3}$$

Erklärte und Residuenstreuung

Def.: Die Gesamtstreuung des Merkmals Y ist

$$SQT = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{2}$$

Die Residuenstreuung ist – gegeben den funktionalen Zusammenhang $f(X)$ zur Vorhersage von Y

$$SQR = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{3}$$

Die Erklärte Streuung ist

$$SQE = \sum_{i=1}^{n} (\bar{y} - f(x_i))^2 \tag{4}$$

Streuzerlegungssatz

Satz (Streuzerlegung): Sei $f(x) = \hat{\beta}x + \hat{\alpha}$ mit $\hat{\beta}, \hat{\alpha}$ berechnet aus den Wertepaaren x_i, y_i, $i = 1, \ldots, n$. Dann

$$SQT = SQE + SQR \tag{5}$$

Streuzerlegungssatz

Satz (Streuzerlegung): Sei $f(x) = \hat{\beta}x + \hat{\alpha}$ mit $\hat{\beta}, \hat{\alpha}$ berechnet aus den Wertepaaren x_i, y_i, $i = 1, \ldots, n$. Dann

$$SQT = SQE + SQR \tag{5}$$

Denn: (Hier $\mathrm{Var}(X)$ für $\hat{\sigma}^2_X$ und $\mathrm{Cov}(X,Y)$ für $\hat{\sigma}_{X,Y}$)

$$\begin{aligned} SQT &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\ &= \sum_{i=1}^{n}(y_i - f(x_i) + f(x_i) - \bar{y})^2 \\ &= SQE + SQR + 2(n-1)\mathrm{Cov}(Y - f(X), f(X) - \bar{y}) \end{aligned}$$

Streuzerlegungssatz

Satz (Streuzerlegung): Sei $f(x) = \hat{\beta}x + \hat{\alpha}$ mit $\hat{\beta}, \hat{\alpha}$ berechnet aus den Wertepaaren x_i, y_i, $i = 1, \ldots, n$. Dann

$$SQT = SQE + SQR \tag{5}$$

Denn: (Hier $\mathrm{Var}(X)$ für $\hat{\sigma}_X^2$ und $\mathrm{Cov}(X, Y)$ für $\hat{\sigma}_{X,Y}$)

$$\begin{aligned} SQT &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\ &= \sum_{i=1}^{n} (y_i - f(x_i) + f(x_i) - \bar{y})^2 \\ &= SQE + SQR + 2(n-1)\mathrm{Cov}(Y - f(X), f(X) - \bar{y}) \end{aligned}$$

Zu Zeigen: $\mathrm{Cov}(Y - f(X), f(X) - \bar{y}) = 0$

Streuzerlegung – Beweis

$$
\begin{aligned}
\mathrm{Cov}(Y - f(X), f(X) - \bar{y}) &= \mathrm{Cov}(Y - \hat{\beta}X - \hat{\alpha}, \hat{\beta}X - \hat{\alpha} - \bar{y}) \\
&= \hat{\beta}\mathrm{Cov}(Y - \hat{\beta}X, X) \\
&= \hat{\beta}\mathrm{Cov}(Y, X) - \hat{\beta}^2\mathrm{Var}(X) \\
&= \hat{\beta}(\mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, X)\hat{\sigma}_X^2/\hat{\sigma}_X^2) \\
&= 0
\end{aligned}
$$

Streuzerlegung – Beweis

$$\begin{aligned}
\mathrm{Cov}(Y - f(X), f(X) - \bar{y}) &= \mathrm{Cov}(Y - \hat{\beta}X - \hat{\alpha}, \hat{\beta}X - \hat{\alpha} - \bar{y}) \\
&= \hat{\beta}\mathrm{Cov}(Y - \hat{\beta}X, X) \\
&= \hat{\beta}\mathrm{Cov}(Y, X) - \hat{\beta}^2\mathrm{Var}(X) \\
&= \hat{\beta}(\mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, X)\hat{\sigma}_X^2/\hat{\sigma}_X^2) \\
&= 0
\end{aligned}$$

qed.

R^2 - der erklärte Anteil

Def.: Der durch das Modell f erklärte Anteil R^2 ist gegeben durch

$$R^2 = SQE/SQT = 1 - SQR/SQT \tag{6}$$

R^2 - der erklärte Anteil

Def.: Der durch das Modell f erklärte Anteil R^2 ist gegeben durch

$$R^2 = SQE/SQT = 1 - SQR/SQT \tag{6}$$

Es gilt: Ist $f(x) = \hat{\beta}x + \hat{\alpha}$, dann $R^2 = \hat{r}^2_{X,Y}$

R^2 - der erklärte Anteil

Def.: Der durch das Modell f erklärte Anteil R^2 ist gegeben durch

$$R^2 = SQE/SQT = 1 - SQR/SQT \tag{6}$$

Es gilt: Ist $f(x) = \hat{\beta}x + \hat{\alpha}$, dann $R^2 = \hat{r}^2_{X,Y}$

Denn:

$$\begin{aligned}
R^2 &= \frac{\sum_i (f(x_i) - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i (\hat{\alpha} + \hat{\beta}x_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \\
&\overset{\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}}{=} \frac{\hat{\beta}^2 \sum_i (x_i - \bar{x})^2}{\sum_i (y_i - \bar{y})^2} \overset{\hat{\beta}=\hat{\sigma}^2_{X,Y}/\hat{\sigma}^2_X}{=} \frac{(\hat{\sigma}^2_{X,Y})^2 \hat{\sigma}^2_X}{(\hat{\sigma}^2_X)^2 \hat{\sigma}^2_Y} \\
&= \left(\frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y}\right)^2 = \hat{r}^2_{X,Y}
\end{aligned}$$

Eigenschaften von R^2

R^2 ist für beliebige funktionale Abhängigkeiten erklärt

Eigenschaften von R^2

R^2 ist für beliebige funktionale Abhängigkeiten erklärt

R^2 liegt zwischen Null und Eins

$$0 \leq R^2 \leq 1 \tag{7}$$

Eigenschaften von R^2

R^2 ist für beliebige funktionale Abhängigkeiten erklärt

R^2 liegt zwischen Null und Eins

$$0 \leq R^2 \leq 1 \tag{7}$$

Ist $R^2 = 1 \Rightarrow SQR = 0 \Rightarrow y_i = f(x_i)$ für $i = 1, \dots n$, d.h. die funktionale Abhängigkeit f erklärt die <u>beobachteten</u> Variationen von y_i genau.

Eigenschaften von R^2

R^2 ist für beliebige funktionale Abhängigkeiten erklärt

R^2 liegt zwischen Null und Eins

$$0 \leq R^2 \leq 1 \tag{7}$$

Ist $R^2 = 1 \Rightarrow SQR = 0 \Rightarrow y_i = f(x_i)$ für $i = 1, \ldots, n$, d.h. die funktionale Abhängigkeit f erklärt die <u>beobachteten</u> Variationen von y_i genau.

Dies impliziert jedoch NICHT dass eine funktionale Abhängigkeit mit größerem R^2 auch <u>nicht beobachtete</u> Werte besser vorhersagt (wegen potentiellem *overfitting*)!

Eigenschaften von R^2

R^2 ist für beliebige funktionale Abhängigkeiten erklärt

R^2 liegt zwischen Null und Eins

$$0 \leq R^2 \leq 1 \tag{7}$$

Ist $R^2 = 1 \Rightarrow SQR = 0 \Rightarrow y_i = f(x_i)$ für $i = 1, \ldots n$, d.h. die funktionale Abhängigkeit f erklärt die beobachteten Variationen von y_i genau.

Dies impliziert jedoch NICHT dass eine funktionale Abhängigkeit mit größerem R^2 auch nicht beobachtete Werte besser vorhersagt (wegen potentiellem *overfitting*)!

Wir kommen darauf zurück...

Transformationen und lin. Reg

Skalengesetze

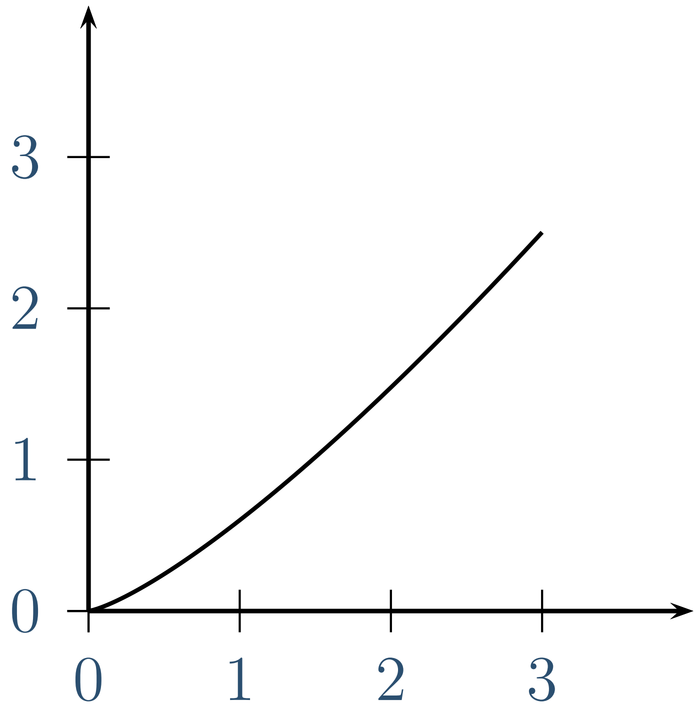
Bei Urliste $(t_1, z_1), \ldots, (t_n, z_n)$, $t_i > 0$, $z_i > 0$ die Skalengesetz folgt$\longrightarrow$ funktionale Abhängigkeit $z_i = f(x_i) \times \exp($stat. Schwankungen$)$:

$$f(t) = at^\beta \tag{8}$$

Skalengesetze

Bei Urliste $(t_1, z_1), \ldots, (t_n, z_n)$, $t_i > 0$, $z_i > 0$ die Skalengesetz folgt$\longrightarrow$ funktionale Abhängigkeit $z_i = f(x_i) \times \exp($stat. Schwankungen$)$:

$$f(t) = at^\beta \tag{8}$$



Doppellogarithmische Substitution

Setze $y_i = \log(z_i)$ und $x_i = \log(t_i) \Rightarrow$

$$\begin{aligned} y_i &= \log(z_i) = \log(a) + \beta \log(t_i) + \text{stat. Schwankungen} \\ &= \log(a) + \beta x_i + \text{stat. Schwankungen} \end{aligned} \tag{9}$$

Doppellogarithmische Substitution

Setze $y_i = \log(z_i)$ und $x_i = \log(t_i) \Rightarrow$

$$\begin{aligned} y_i &= \log(z_i) = \log(a) + \beta \log(t_i) + \text{stat. Schwankungen} \\ &= \log(a) + \beta x_i + \text{stat. Schwankungen} \end{aligned} \tag{9}$$

Die doppellogarithmische Transformation führt Skalengesetze auf lineare Gesetze zurück

Doppellogarithmische Substitution

Setze $y_i = \log(z_i)$ und $x_i = \log(t_i) \Rightarrow$

$$\begin{aligned} y_i &= \log(z_i) = \log(a) + \beta \log(t_i) + \text{stat. Schwankungen} \\ &= \log(a) + \beta x_i + \text{stat. Schwankungen} \end{aligned} \tag{9}$$

Die doppellogarithmische Transformation führt Skalengesetze auf lineare Gesetze zurück

Nach dopplellogarithmischer Transformation der Daten kann lineare Regression angewendet werden!

Doppellogarithmische Substitution

Setze $y_i = \log(z_i)$ und $x_i = \log(t_i) \Rightarrow$

$$\begin{aligned} y_i &= \log(z_i) = \log(a) + \beta \log(t_i) + \text{stat. Schwankungen} \\ &= \log(a) + \beta x_i + \text{stat. Schwankungen} \end{aligned} \tag{9}$$

Die doppellogarithmische Transformation führt Skalengesetze auf lineare Gesetze zurück

Nach dopplellogarithmischer Transformation der Daten kann lineare Regression angewendet werden!

Beachte jedoch, dass lineare Regression nur dann sinnvoll ist, wenn die Residuen NACH der Transformation über den Vorhersagebereich ungefähr dieselbe Größenordnung haben – sonst ist evtl. ein nichtlinearer Least Squares Fit ohne Transformation vorzuziehen.

Skalengesetze – Lösung

Die durch doppellogarithmische Transformation gewonnene Lösung ist

$$f(t) = e^{\hat{\alpha}} t^{\hat{\beta}} \qquad (10)$$

Skalengesetze – Lösung

Die durch doppellogarithmische Transformation gewonnene Lösung ist

$$f(t) = e^{\hat{\alpha}} t^{\hat{\beta}} \tag{10}$$

mit

$$\hat{\beta} = \hat{\sigma}_{\log(T),\log(Z)} / \hat{\sigma}^2_{\log(T)} \quad \hat{\alpha} = \overline{\log(z)} - \hat{\beta}\overline{\log(t)} \tag{11}$$

Skalengesetze – Lösung

Die durch doppellogarithmische Transformation gewonnene Lösung ist

$$f(t) = e^{\hat{\alpha}} t^{\hat{\beta}} \tag{10}$$

mit

$$\hat{\beta} = \hat{\sigma}_{\log(T),\log(Z)} / \hat{\sigma}^2_{\log(T)} \quad \hat{\alpha} = \overline{\log(z)} - \hat{\beta}\overline{\log(t)} \tag{11}$$

Dies ist natürlich nicht der funktionale Zusammenhang, der die Fehlerquadrate minimiert!

Skalengesetze – Lösung

Die durch doppellogarithmische Transformation gewonnene Lösung ist

$$f(t) = e^{\hat{\alpha}} t^{\hat{\beta}} \tag{10}$$

mit

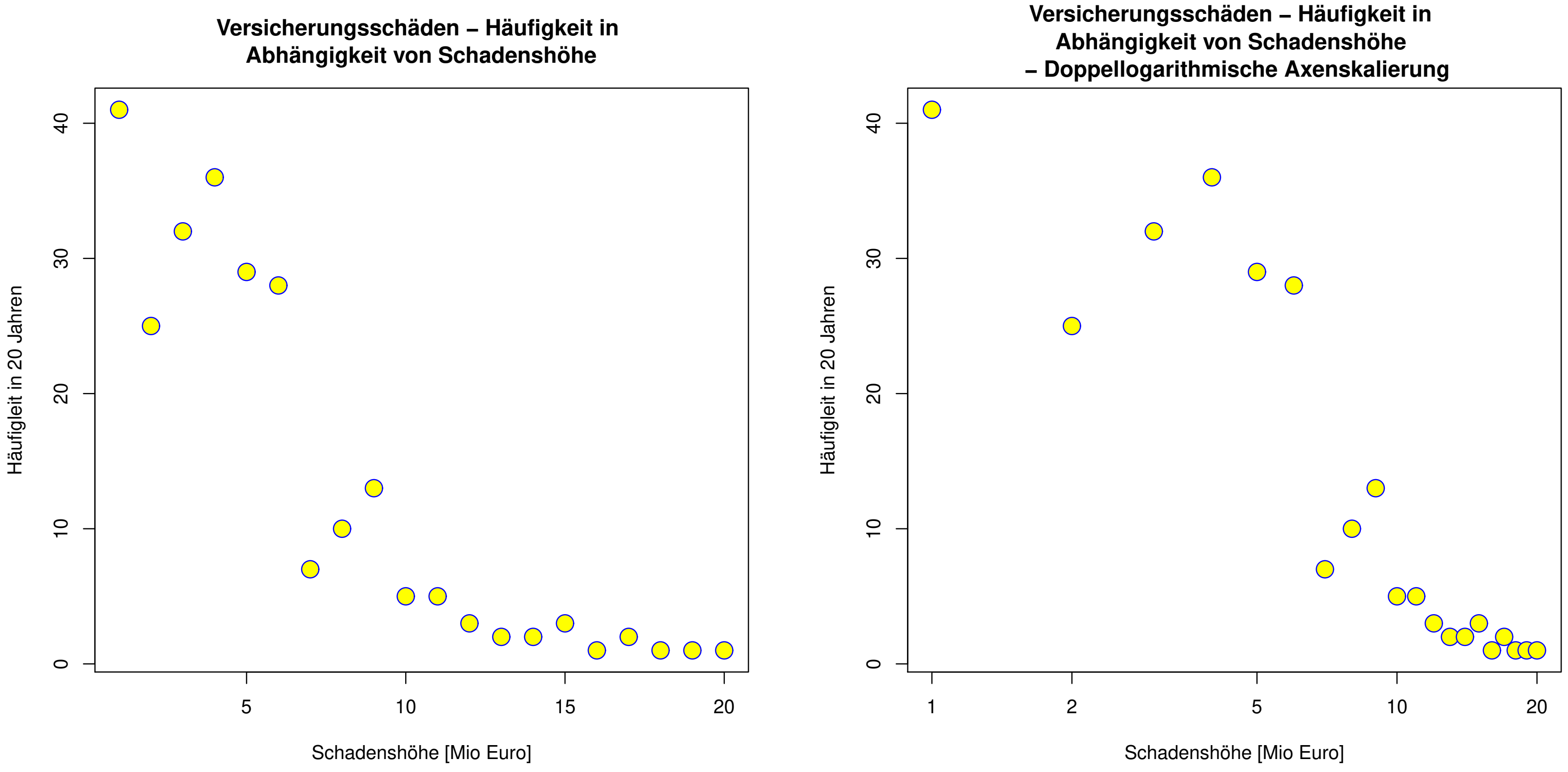
$$\hat{\beta} = \hat{\sigma}_{\log(T),\log(Z)} / \hat{\sigma}^2_{\log(T)} \quad \hat{\alpha} = \overline{\log(z)} - \hat{\beta}\overline{\log(t)} \tag{11}$$

Dies ist natürlich nicht der funktionale Zusammenhang, der die Fehlerquadrate minimiert!

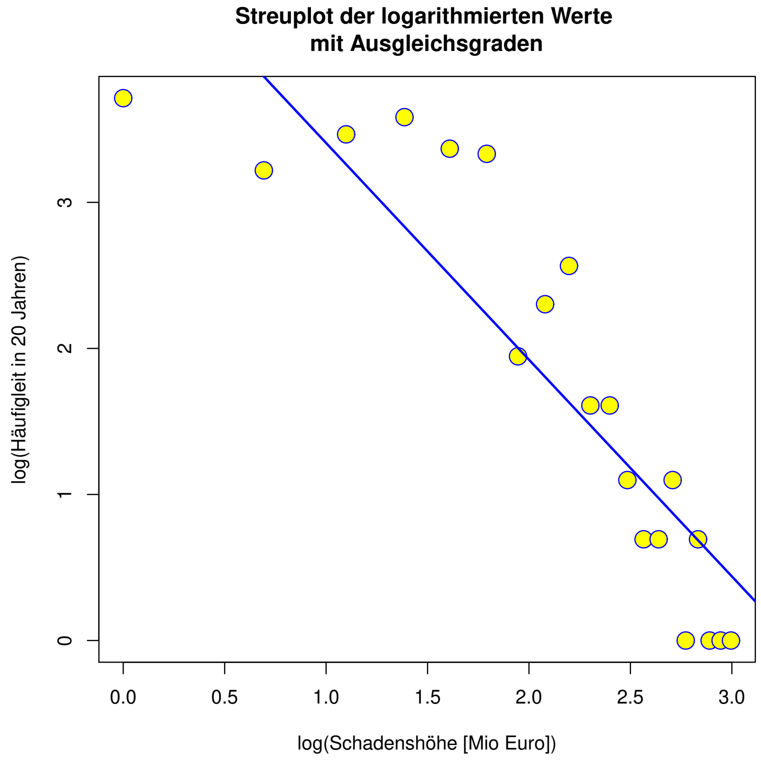
Vorgehen mittels doppellogarithmischer Transformation ist sinnvoll, wenn der zu erwartede Fehler größenmäßig proportional zum gemessenen Wert ist.

Skalengesetze - Beispiel

Schadenshäufigkeit in Abhängigkeit von Schadenshöhe bei Versicherung gegen Ernteausfälle (fiktive Daten)



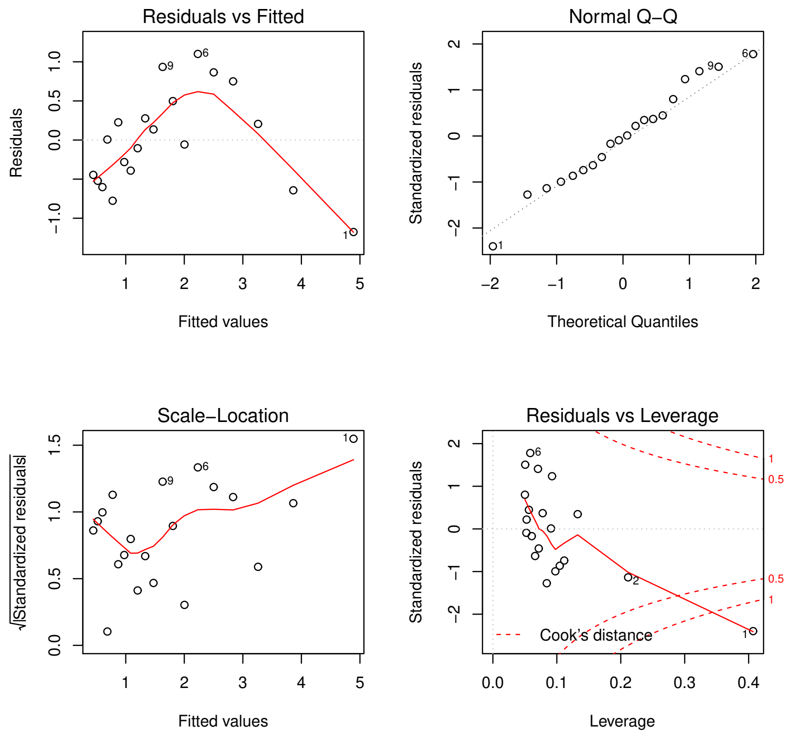
Skalengesetz - Beispiel II



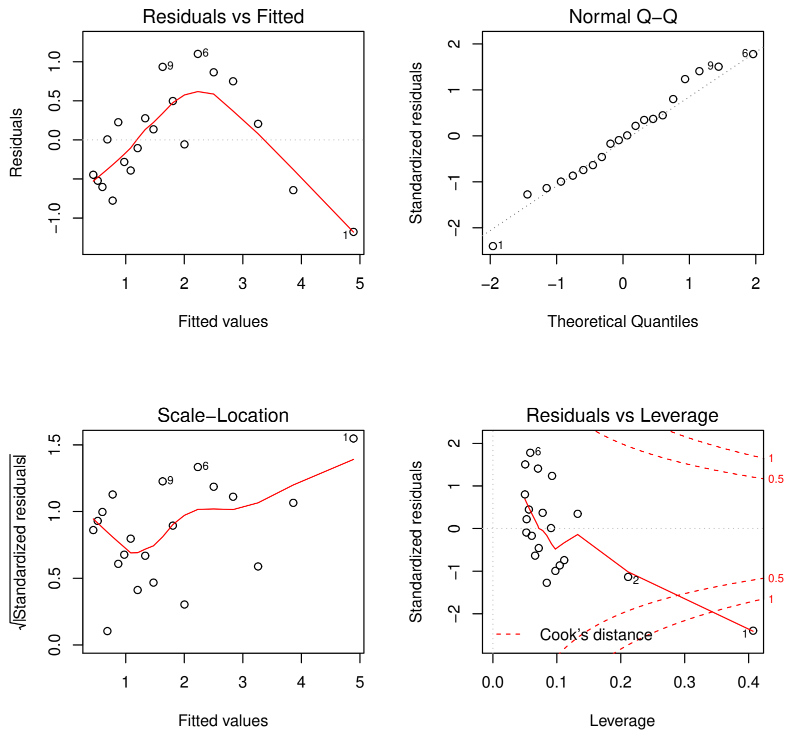
Berechnete Werte:

$$\hat{\beta} = -1.4836, \quad \hat{\alpha} = 4.4890$$

Skalengesetz - Beispiel III



Skalengesetz - Beispiel III



Diagnostische Plots nicht wirklich prickelnd. . .

Skalengesetz - Beispiel IV

Die ersten vier Werte (korrespondierend zu kleinen Schadenshöhen) passen in den Diagnostischen plots nicht gut ins Bild.

Skalengesetz - Beispiel IV

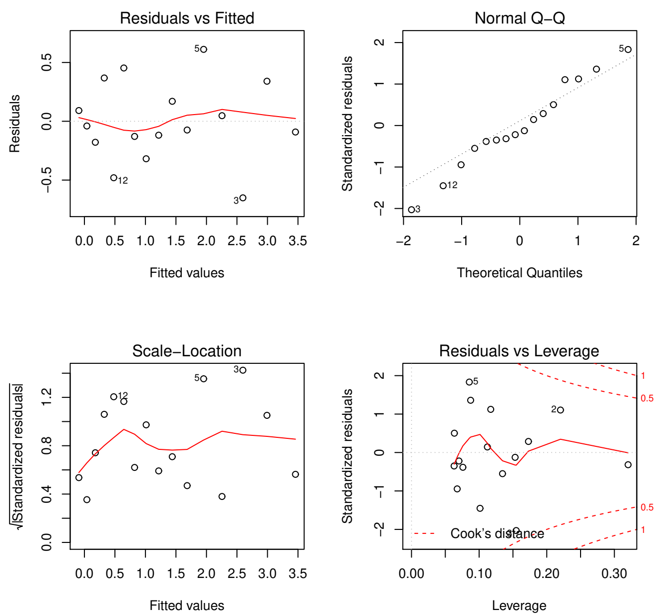
Die ersten vier Werte (korrespondierend zu kleinen Schadenshöhen) passen in den Diagnostischen plots nicht gut ins Bild.

Es wird argumentiert, dass kleine Schadenshöhen nicht wichtig sind in dieser Betrachtung und sie werden herausgenommen. . .

Skalengesetz - Beispiel IV

Die ersten vier Werte (korrespondierend zu kleinen Schadenshöhen) passen in den Diagnostischen plots nicht gut ins Bild.

Es wird argumentiert, dass kleine Schadenshöhen nicht wichtig sind in dieser Betrachtung und sie werden herausgenommen. . .



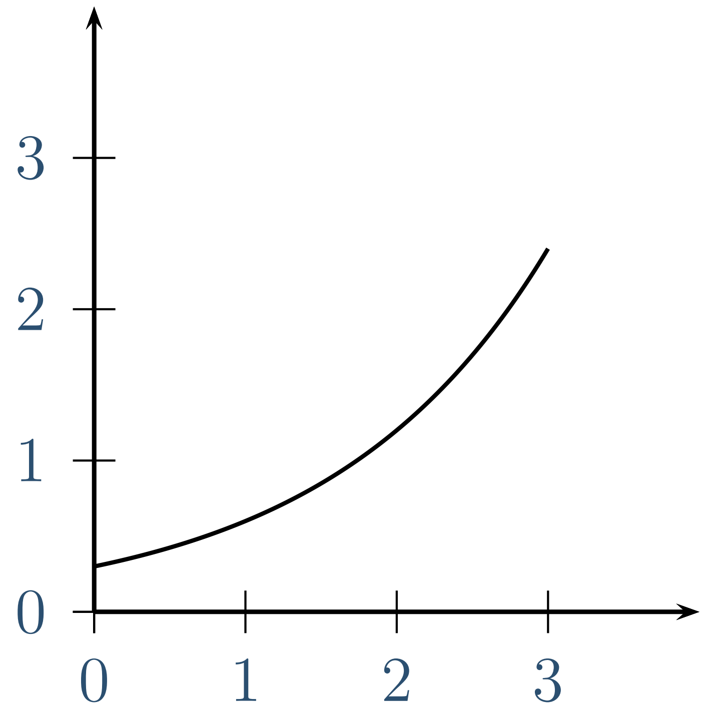
Die Diagnostischen Plots sehen nun viel besser aus, doch die Werte $\hat{\alpha} = 7.579$ und $\hat{\beta} = -2.561$ unterscheiden sich stark. . .

Welche Vorgehensweise ist gerechtfertigt **?**

Exponentialgesetz

Urliste enthalte Wertepaare $(x_1, z_1), \ldots (x_n, z_n)$ die einer exponentiellen funktionalen Abhängigkeit entsprechen:

$$f(x) = ae^{\beta x} \tag{12}$$



Ein.-logarithmische Substitution

Setze $y_i = \log(z_i)$ und $\alpha = \log(a)$, dann

$$y_i = \log(z_i) = \alpha + \beta x_i + \text{stat. Schwankungen} \qquad (13)$$

Ein.-logarithmische Substitution

Setze $y_i = \log(z_i)$ und $\alpha = \log(a)$, dann

$$y_i = \log(z_i) = \alpha + \beta x_i + \text{stat. Schwankungen} \tag{13}$$

Die lin. Regression nach einfachlogarithmischer Transfrormation ergibt

$$f(x) = e^{\hat{\alpha}} e^{\hat{\beta} x} \text{ mit } \hat{\beta} = \frac{\hat{\sigma}_{X,\log(Z)}}{\hat{\sigma}_X^2} \text{ und } \hat{\alpha} = \overline{\log(z)} - \hat{\beta}\bar{x} \tag{14}$$

Ein.-logarithmische Substitution

Setze $y_i = \log(z_i)$ und $\alpha = \log(a)$, dann

$$y_i = \log(z_i) = \alpha + \beta x_i + \text{stat. Schwankungen} \tag{13}$$

Die lin. Regression nach einfachlogarithmischer Transfrormation ergibt

$$f(x) = e^{\hat{\alpha}} e^{\hat{\beta} x} \text{ mit } \hat{\beta} = \frac{\hat{\sigma}_{X,\log(Z)}}{\hat{\sigma}_X^2} \text{ und } \hat{\alpha} = \overline{\log(z)} - \hat{\beta}\bar{x} \tag{14}$$

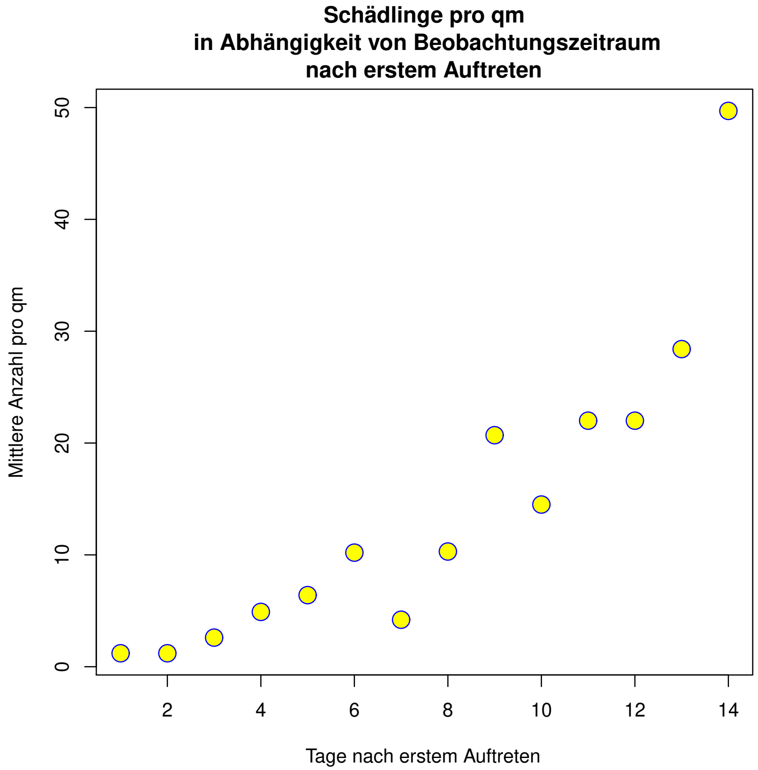
Die Bemerkungen über Residuen gelten analog zum doppellogarithmischen Fall

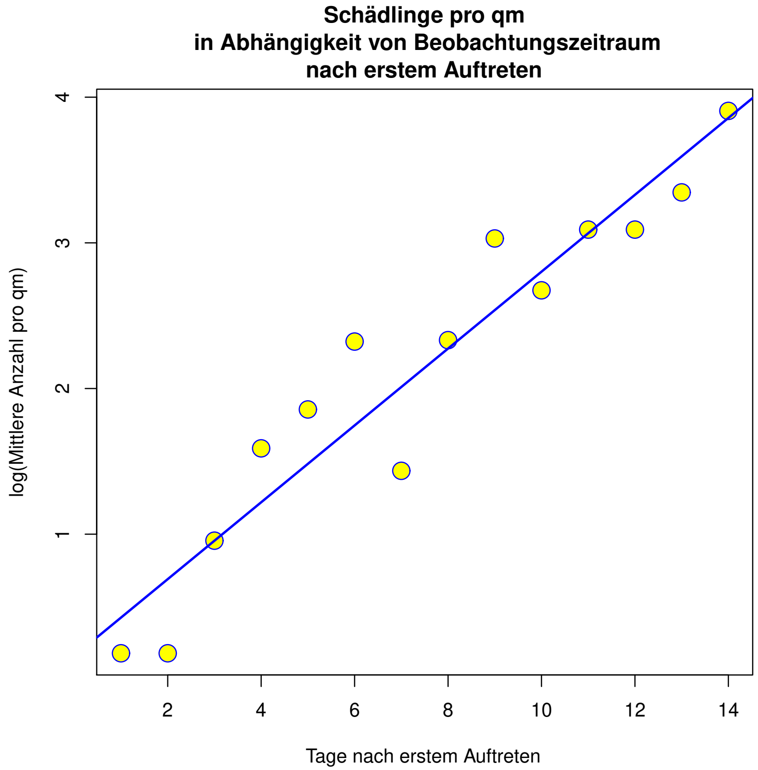
Exponentialgesetz - Beispiel

Ungebremste Vermehrung von Schädlingen

Exponentialgesetz - Beispiel

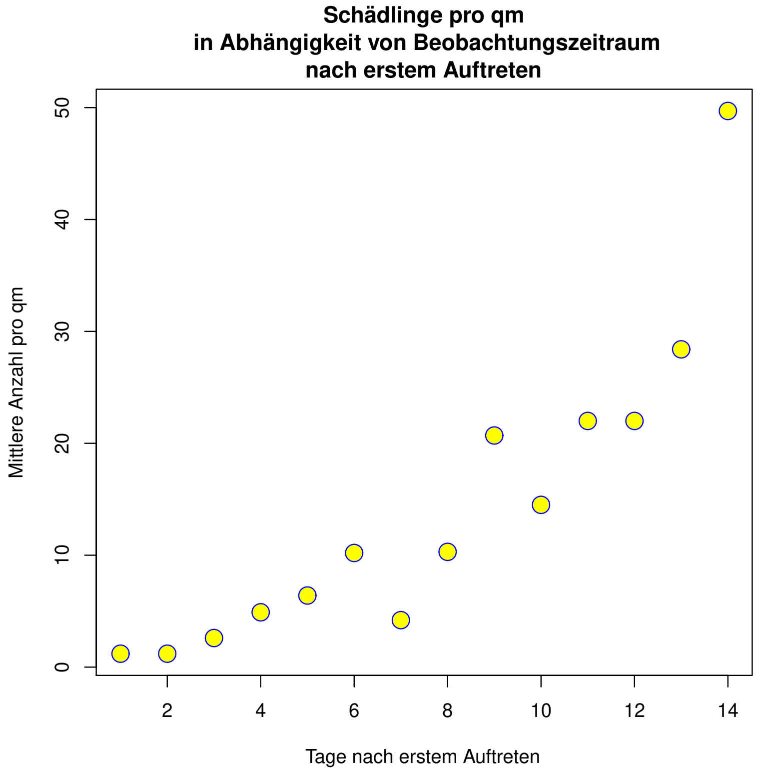
Ungebremste Vermehrung von Schädlingen

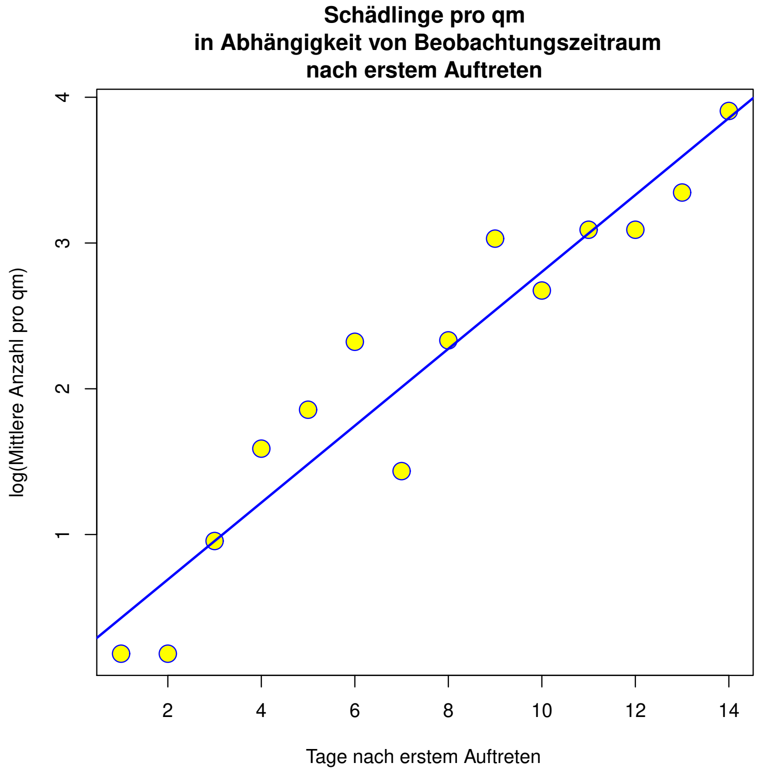




Exponentialgesetz - Beispiel

Ungebremste Vermehrung von Schädlingen

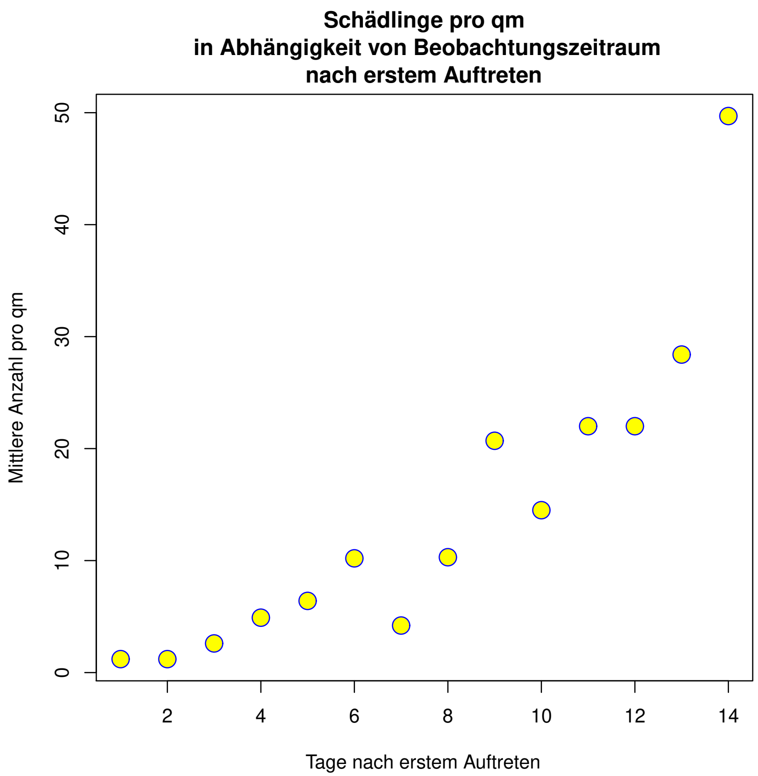


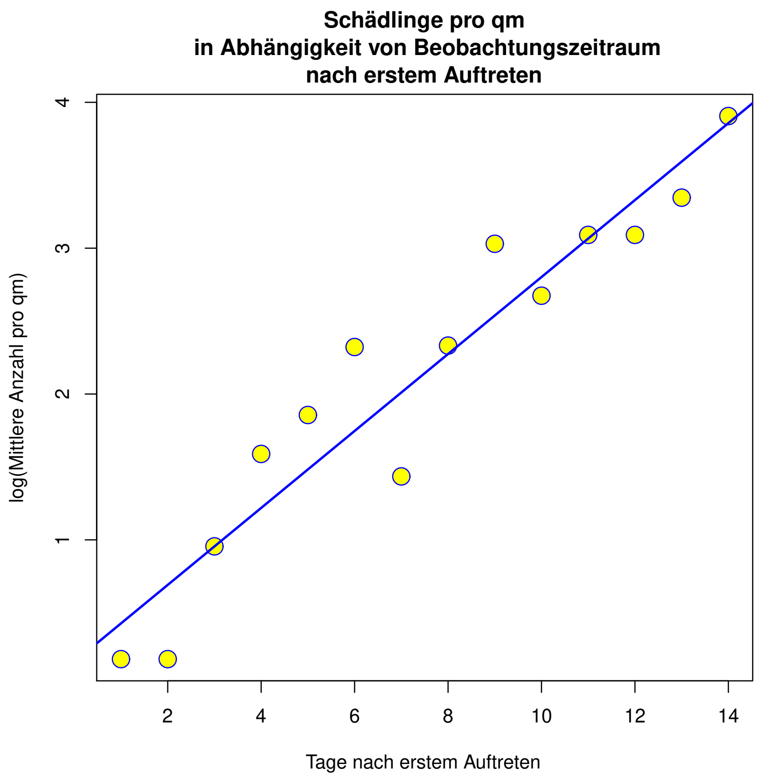


$$\hat{\alpha} = 0.1635 \quad \hat{\beta} = 0.2639 \quad f(x) = e^{0.1635} e^{0.2639x}$$

Exponentialgesetz - Beispiel

Ungebremste Vermehrung von Schädlingen





$$\hat{\alpha} = 0.1635 \quad \hat{\beta} = 0.2639 \quad f(x) = e^{0.1635} e^{0.2639x}$$

Z.B. Prognose für 15. Tag: $f(15) = e^{0.1635} e^{0.2639 \times 15} = 61.64$

Exponentialgesetz -Beispiel