

MSc. Research Methods - Statistikteil Lösungen 2019

Juergen Dengler

September 2019

Musterlösung Aufgabe 2.1: Regression

Übungsaufgabe (hier so ausführlich formuliert, wie dies auch in der Klausur der Fall sein wird)

- Laden Sie den Datensatz decay.csv. Dieser enthält die Zahl radioaktiver Zerfälle pro Zeiteinheit (amount) für Zeitpunkte (time) nach dem Start des Experimentes.
- **Ermitteln Sie ein statistisches Modell, dass die Zerfallshäufigkeit in Abhängigkeit von der Zeit beschreibt.**
- Bitte erklären und begründen Sie die einzelnen Schritte, die Sie unternehmen, um zu diesem Ergebnis zu kommen. Dazu erstellen Sie bitte ein Word-Dokument, in das Sie Schritt für Schritt den verwendeten R-Code, die dazu gehörigen Ausgaben von R, Ihre Interpretation derselben und die sich ergebenden Schlussfolgerungen für das weitere Vorgehen dokumentieren.
- Dieser Ablauf sollte insbesondere beinhalten:
 - Überprüfen der Datenstruktur nach dem Einlesen, welches sind die abhängige(n) und welches die unabhängige(n) Variablen
 - Explorative Datenanalyse, um zu sehen, ob evtl. Dateneingabefehler vorliegen oder Datentransformationen vorgenommen werden sollten
 - Auswahl und Begründung eines statistischen Verfahrens (es gibt hier mehrere statistisch korrekte Möglichkeiten!)
 - Ermittlung eines Modells
 - Durchführen der Modelldiagnostik für das gewählte Modell
 - Generieren aller Zahlen, Statistiken und Tabellen, die für eine wiss. Ergebnisdarstellung benötigt werden
 - Formulieren Sie abschliessend einen Methoden- und Ergebnisteil (ggf. incl. adäquaten Abbildungen) zu dieser Untersuchung in der Form einer wissenschaftlichen Arbeit (ausformulierte schriftliche Zusammenfassung, mit je einem Absatz von ca. 60-100 Worten, resp. 3-8 Sätzen für den Methoden- und Ergebnisteil). D. h. alle wichtigen Informationen sollten enthalten sein, unnötige Redundanz dagegen vermieden werden.
 - **Abzugeben sind am Ende (a) Ein lauffähiges R-Skript; (b) begründeter Lösungsweg (Kombination aus R-Code, R Output und dessen Interpretation) und (c) ausformulierter Methoden- und Ergebnisteil (für eine wiss. Arbeit).**

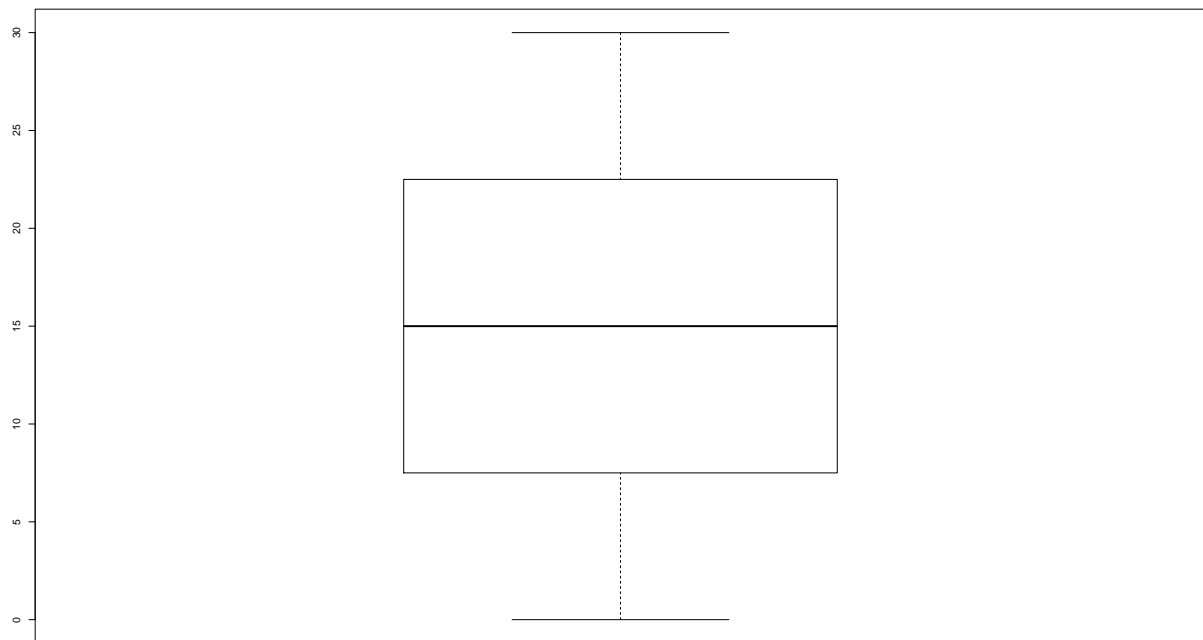
Loesung – Skript

Uebung 2.1 - Regressionsanalyse

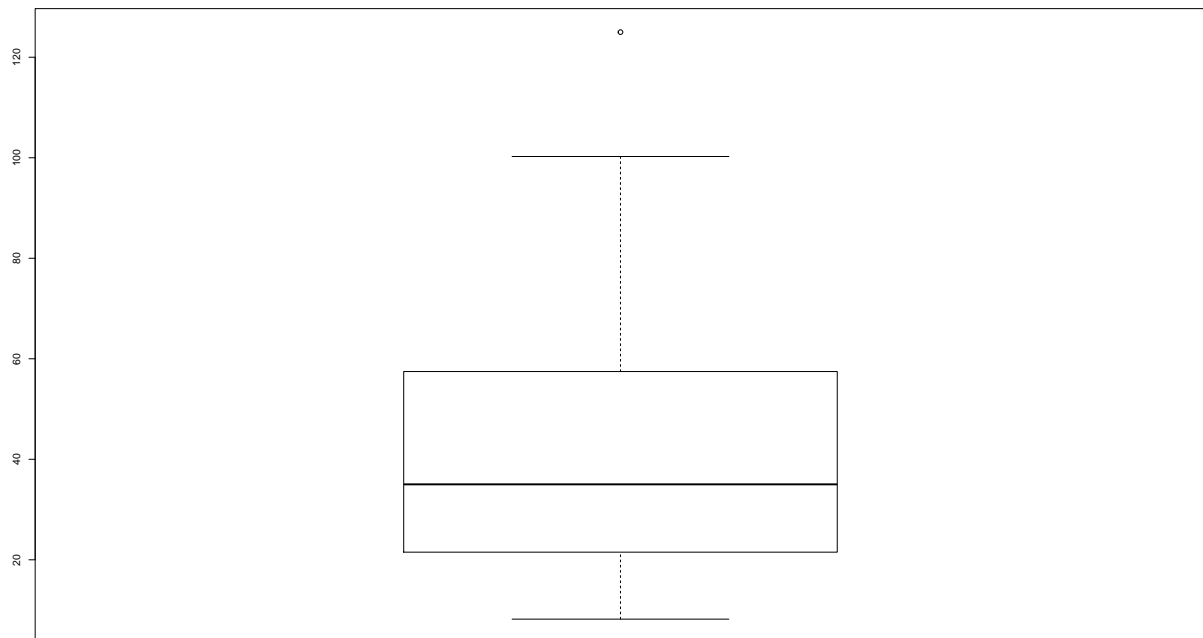
```
data <-read.csv("data/decay.csv")
data
attach(data)
summary(data)
str(data)
```

Explorative Datenanalyse

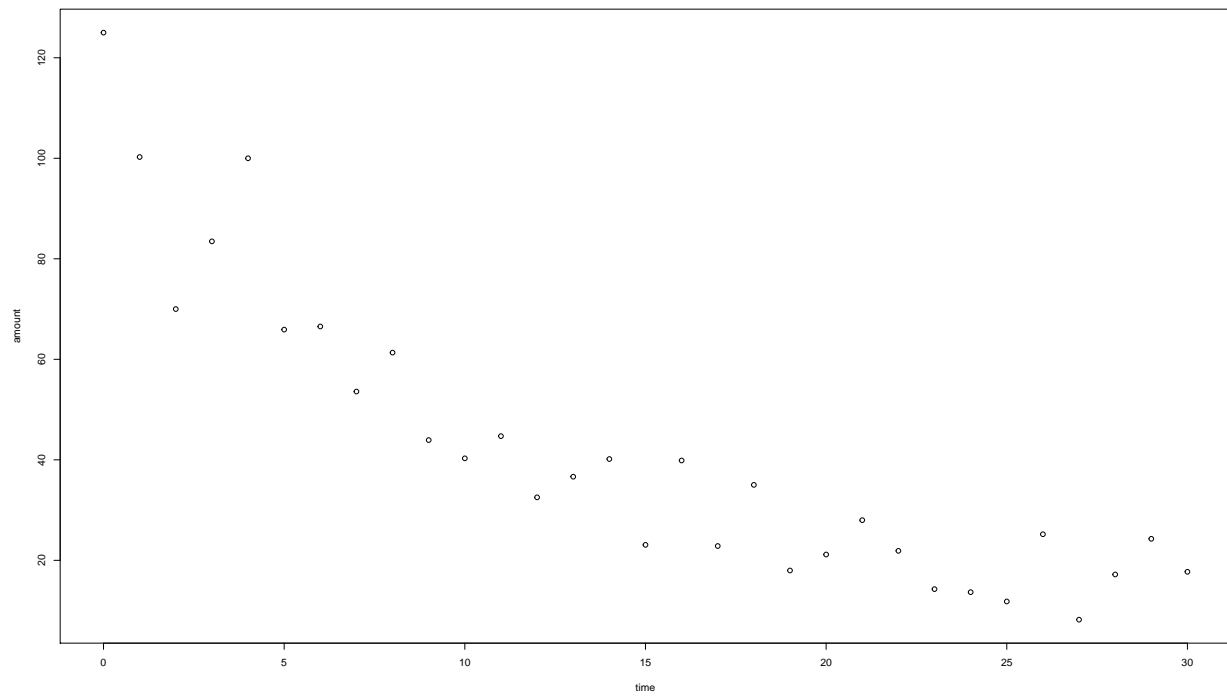
```
boxplot(time)
```



```
boxplot(amount)
```



```
plot(amount~time)
```

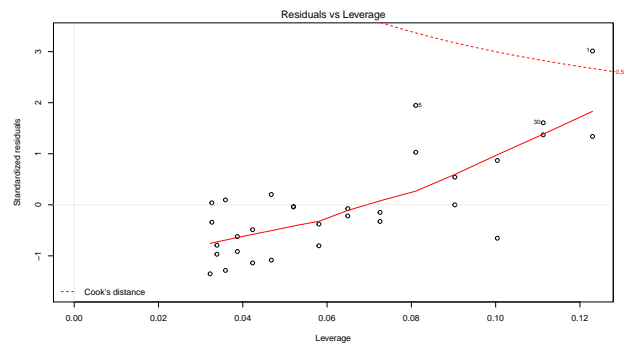
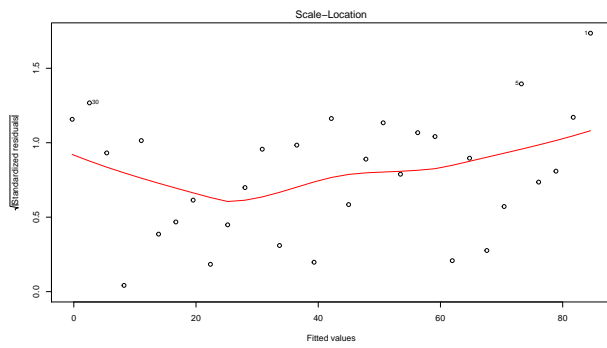
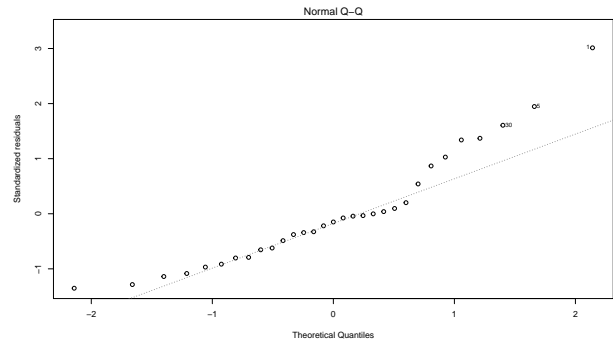
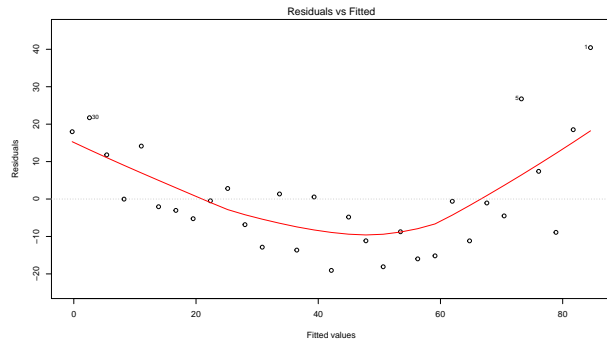


Einfaches lineares Modell

```
model1<-lm(amount~time)
summary(model1)
```

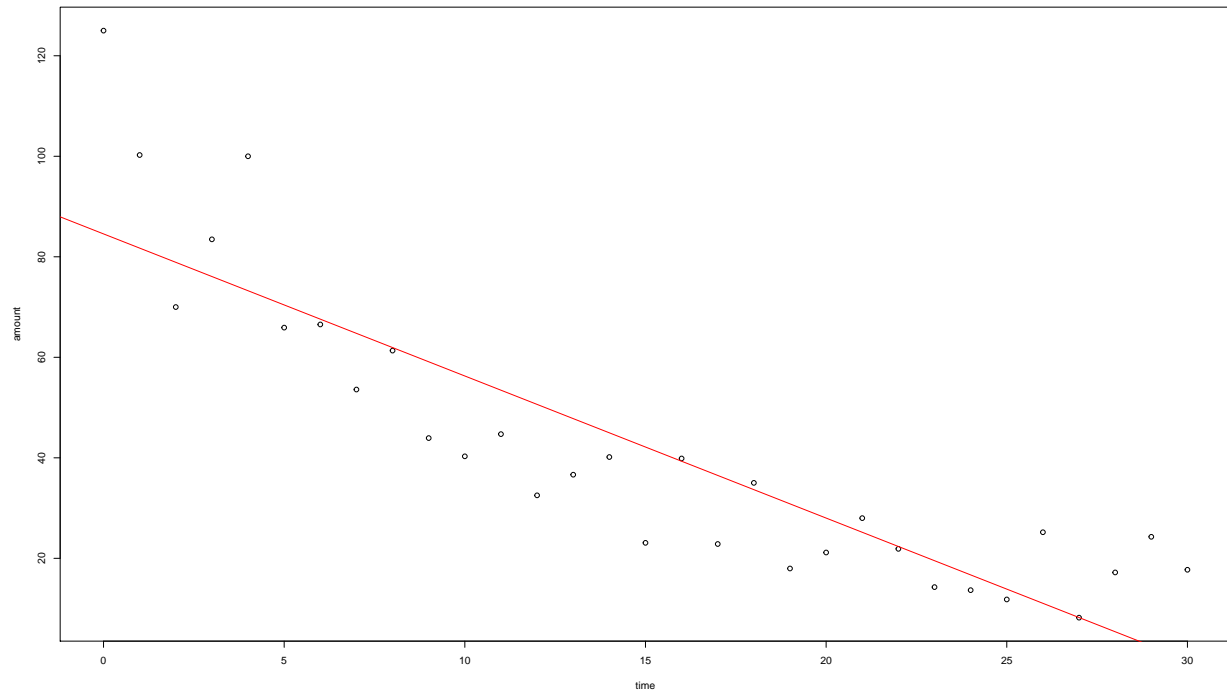
Modelldiagnostik

```
par(mfrow=c(2,2))
plot(model1)
```



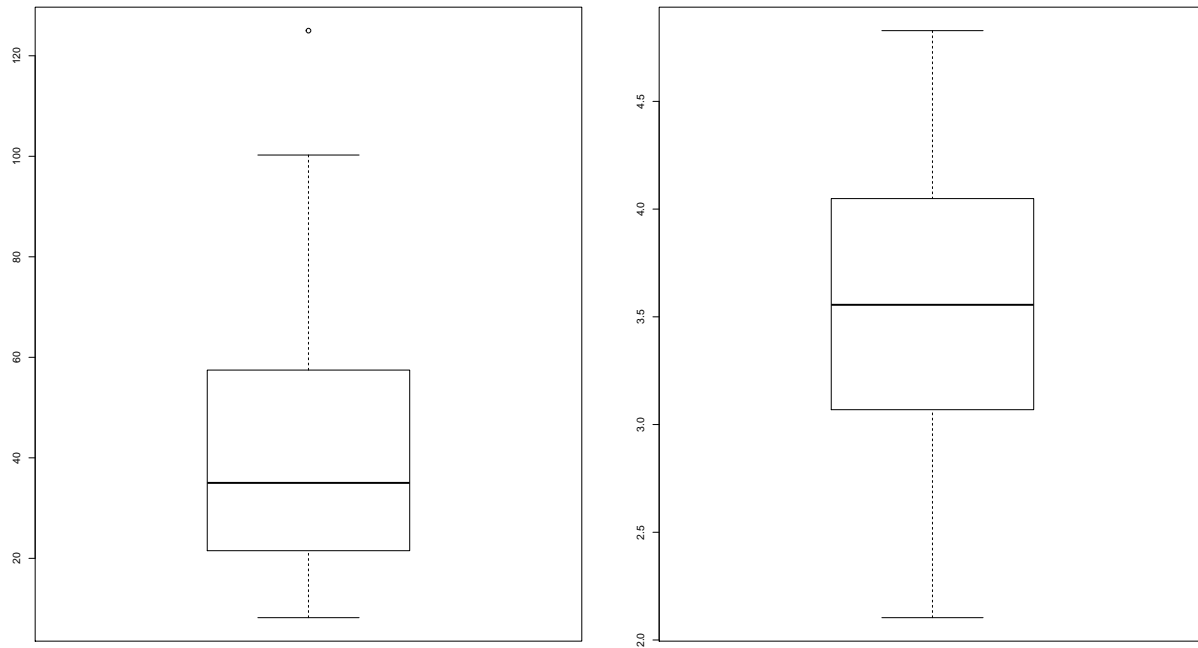
Ergebnisplot

```
par(mfrow=c(1,1))
plot(time,amount)
abline(lm(amount~time),col="red")
```

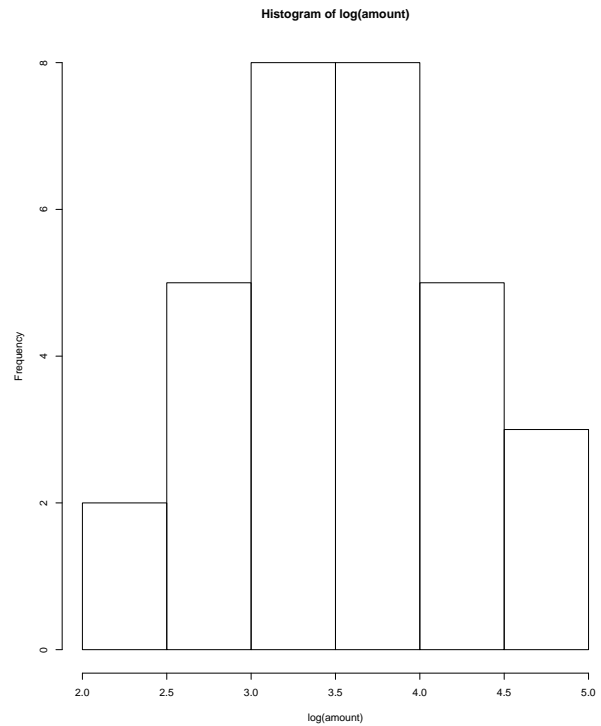
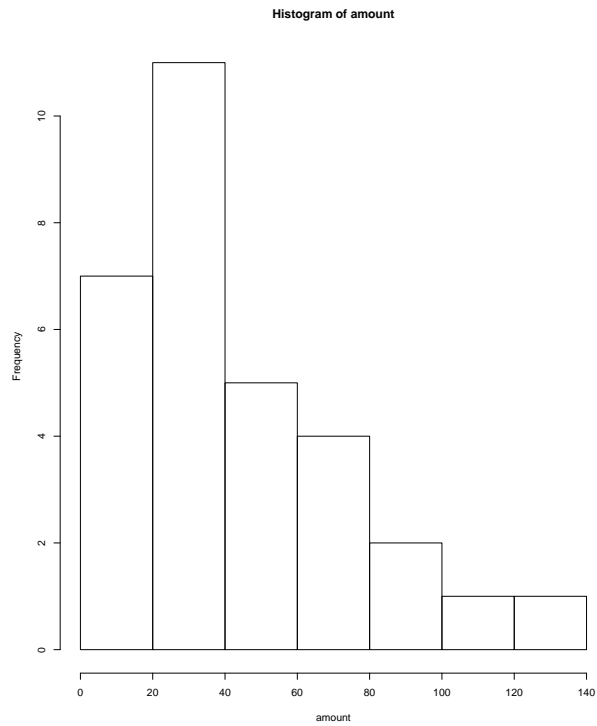


Loesung 1: log-Transformation der abhaengigen Variablen

```
par(mfrow=c(1,2))  
boxplot(amount)  
boxplot(log(amount))
```



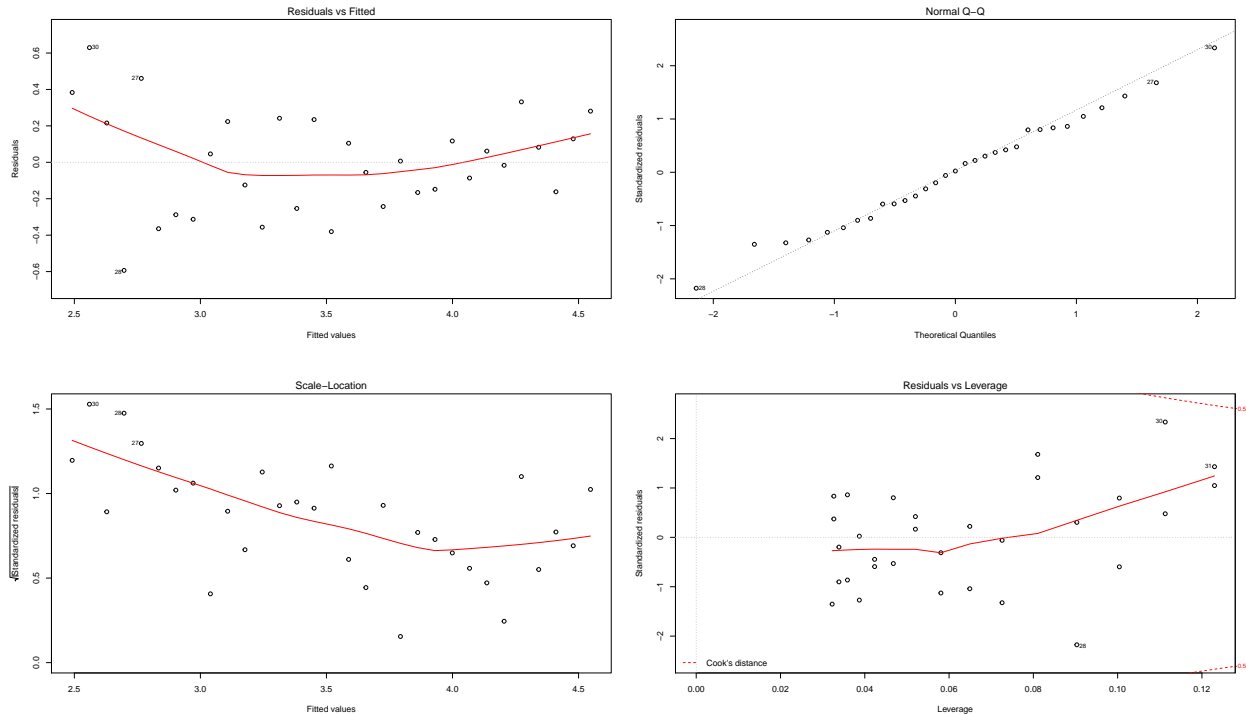
```
hist(amount)  
hist(log(amount))
```



```
model2<-lm(log(amount)~time)
summary(model2)
```

Modelldiagnostik

```
par(mfrow=c(2,2))
plot(model2)
```



Loesung 2: quadratische Regression (kam erst in Statistik 3; koente fuer die Datenverteilung passen, entspricht aber nicht der physikalischen)

Gesetzmaessigkeit

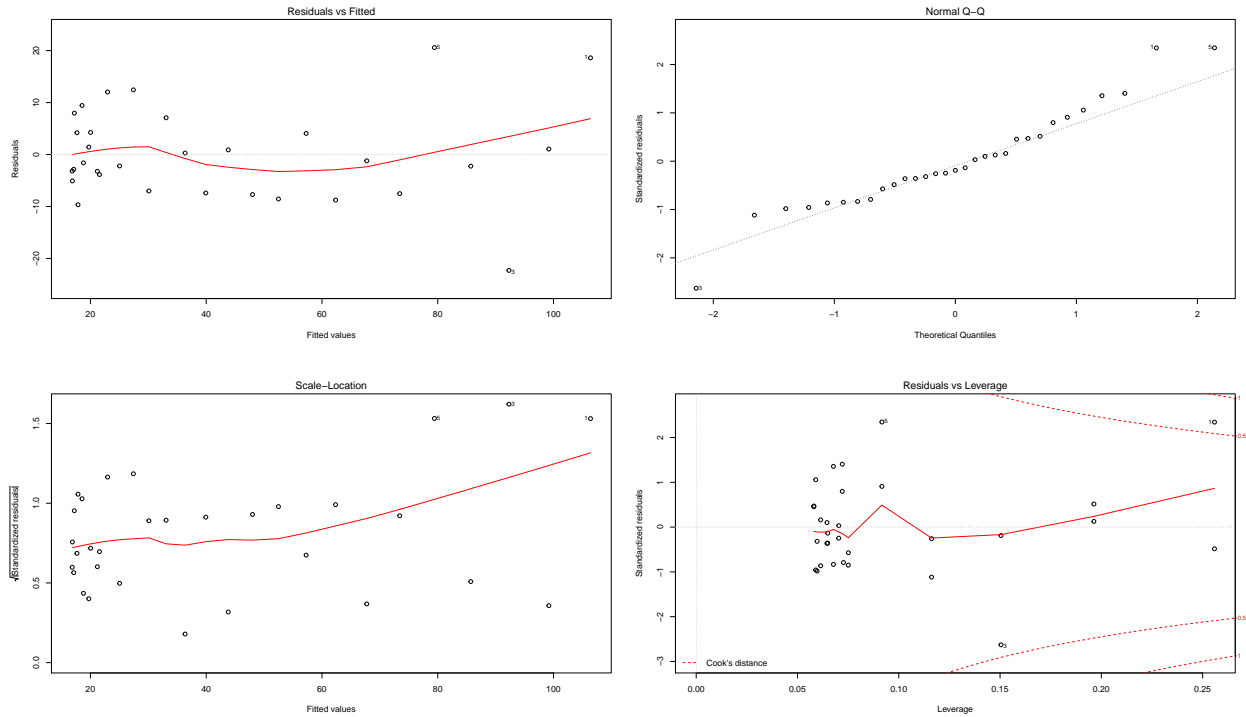
```
model.quad<-lm(amount~time+I(time^2))
summary(model.quad)
```

Vergleich mit dem einfachen Modell mittels ANOVA (es ginge auch AICc)

```
anova(model1,model.quad)
```

Modelldiagnostik

```
par(mfrow=c(2,2))
plot(model.quad)
```



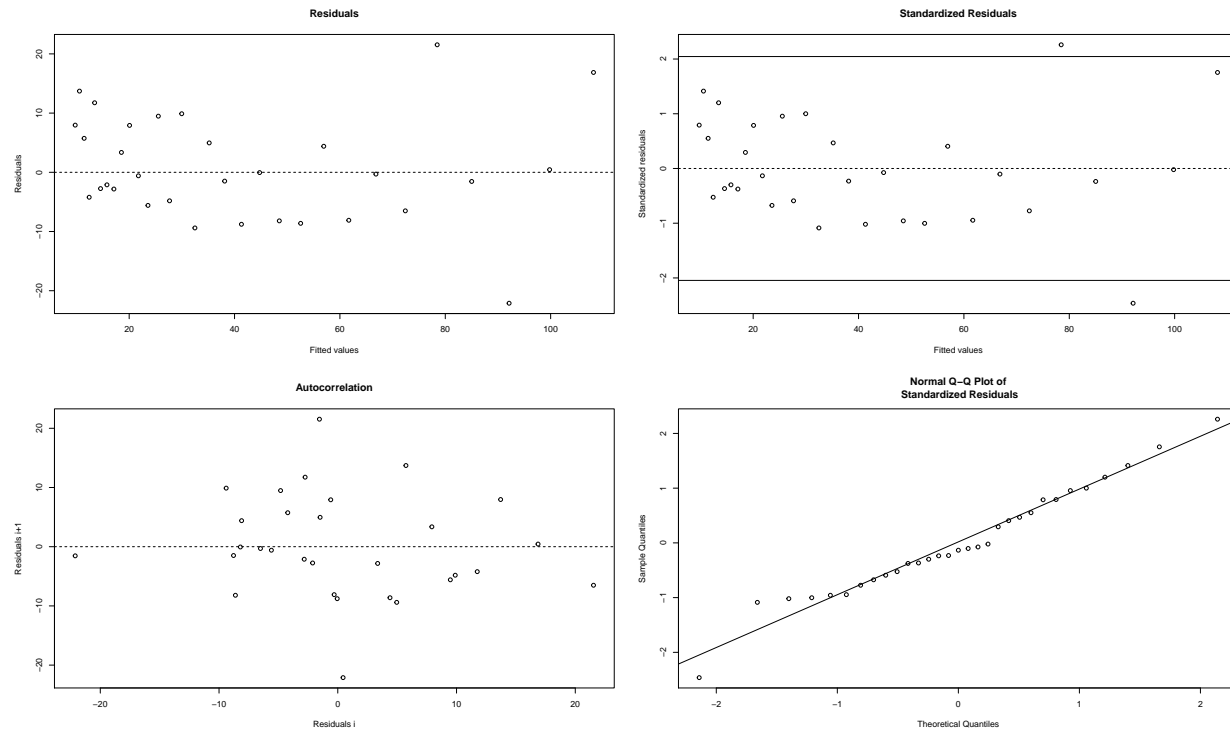
Loesung 3 (kam erst in Statistik 4; methodisch beste Loesung;

mit Startwerten muss man ggf. ausprobieren)

```
model.nls<-nls(amount~a*exp(-b*time),start=(list(a=100,b=1)))
summary(model.nls)
```

Modelldiagnostik

```
library(nlstools)
residuals.nls <- nlsResiduals(model.nls)
plot(residuals.nls)
```

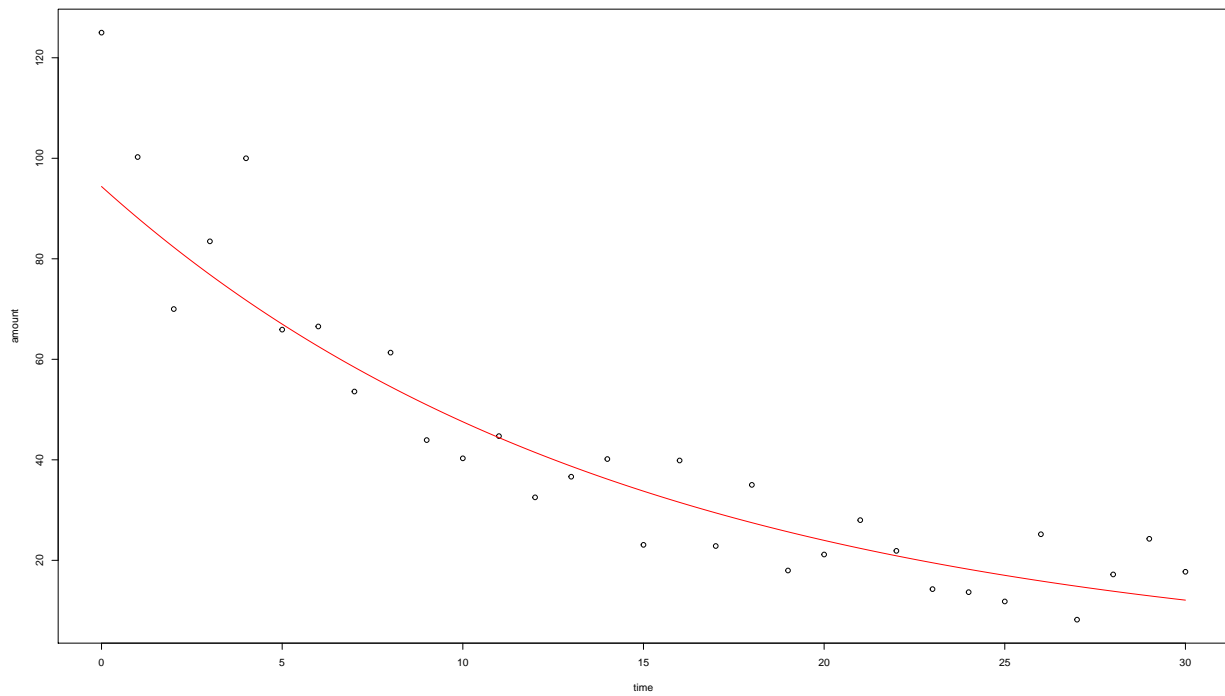



Ergebnisplots

```
par(mfrow=c(1,1))
xv<-seq(0,30,0.1)
```

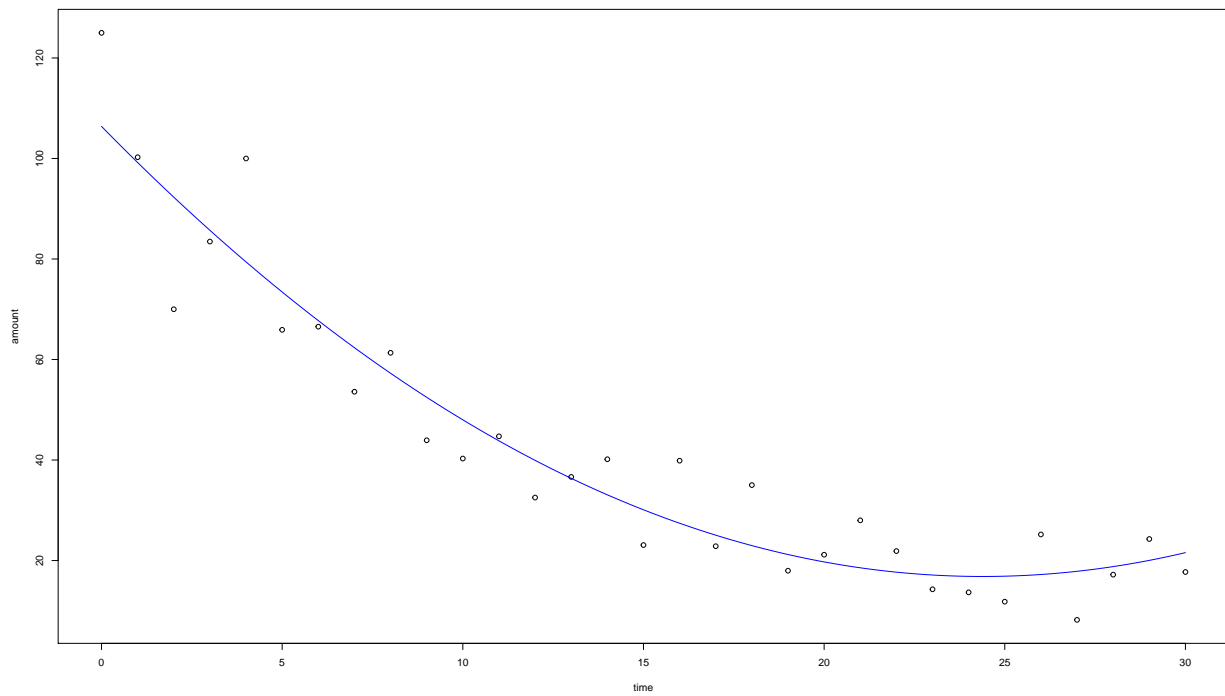
1. lineares Modell mit log-transformierter Abhaengiger

```
plot(time,amount)
yv1<-exp(predict(model2,list(time=xv)))
lines(xv,yv1,col="red")
```



2. quadratisches Modell

```
plot(time,amount)
yv2<-predict(model.quad,list(time=xv))
lines(xv,yv2,col="blue")
```



3. nicht-lineares Modell

```
plot(time,amount)
yv3<-predict(model.nls,list(time=xv))
lines(xv,yv3,col="green")
```

