

# lineare Regression: Diagnose in R - Normalverteilte Fehler

Benjamin Schlegel

27. Juni 2016

Um die Annahme der normalverteilten Fehler zu prüfen eignen sich Grafiken. In diesem Artikel werden zwei verschiedenen graphische Methoden vorgestellt. Eine allgemeine theoretische Erklärung der verschiedenen Diagnosemöglichkeiten zeigt der Artikel [lineare Regression: Diagnose](#).

Als ersten wird der Datensatz World UNDP Data 2014 eingelesen, welcher unter [Data](#) heruntergeladen werden kann. Der Datensatz erhält unter anderem Daten zum Bruttonationaleinkommen pro Kopf (gnipc) und Human Development Index (hdi) für jedes Land.

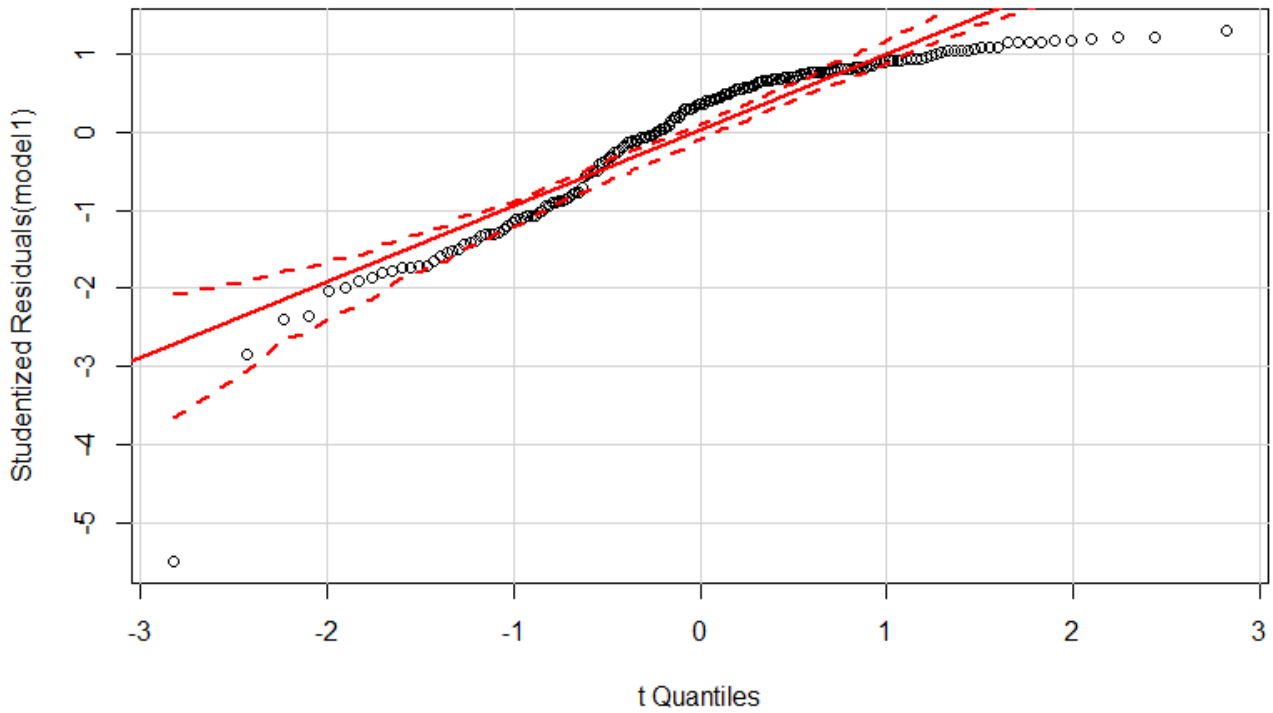
```
world= read.csv2("world_undp.csv", stringsAsFactors = F)
```

Anschliessend wird der HDI auf das Bruttonationaleinkommen regressiert und auf das logarithmierte Bruttonationaleinkommen.

```
modell1 = lm(hdi ~ gnipc, data=world)
modell2 = lm(hdi ~ log(gnipc), data=world)
```

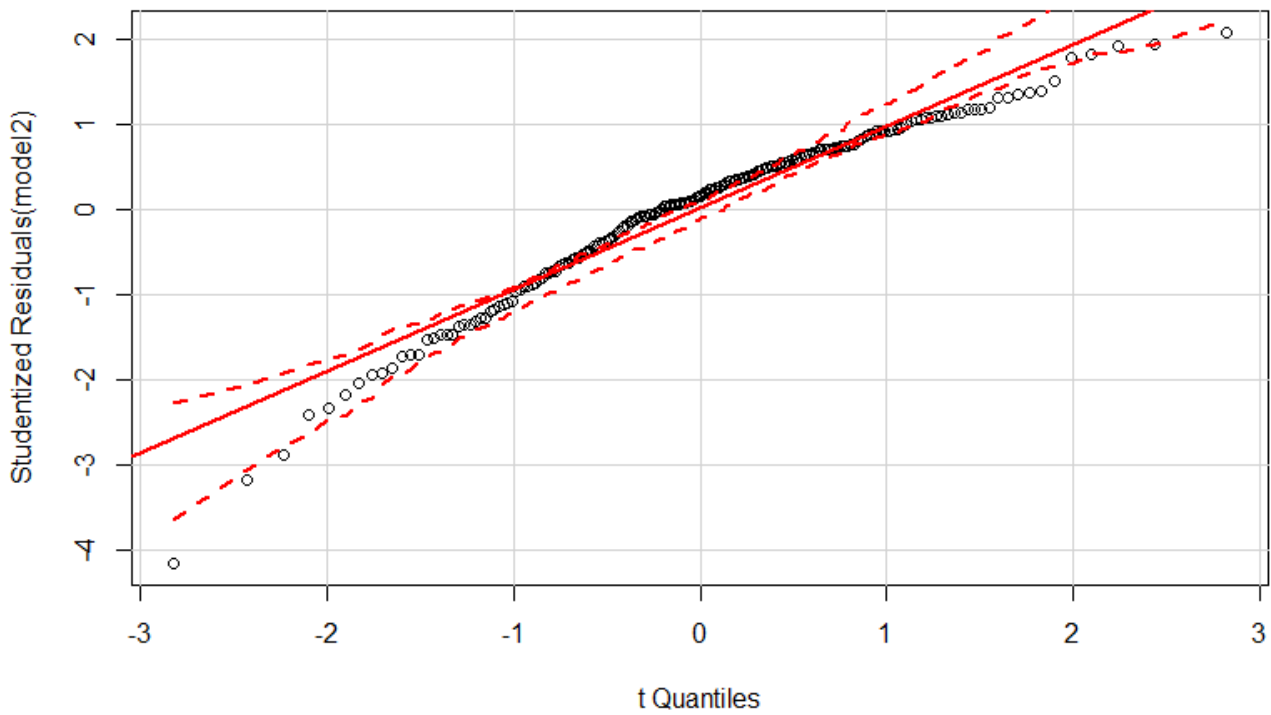
Mithilfe der Funktion `qqPlot` aus der Bibliothek `car` kann die Annahme der normalverteilten Fehler überprüft werden. Dabei sollten die Punkte möglichst alle im Konfidenzintervall liegen.

```
library(car)
qqPlot(modell1)
```



Es ist klar ersichtlich, dass die Punkte stark abweichen. Die Annahme ist deshalb verletzt.

`qqPlot(model2)`



Mithilfe des Logarithmus trifft die Annahme nun einigermaßen zu.

Die zweite Methode schaut sich die standardisierten Residuen an. Die Funktion `studres` aus der Bibliothek `MASS` berechnet die standardisierten Residuen. Diese können in einem Histogramm dargestellt werden. Wird die Normalverteilung darübergelegt, wird ersichtlich, wie stark die Residuen von der Normalverteilung abweichen.

```
library(MASS)
sresidl
```