

# lineare Regression: Diagnose in R - Homoskedastizität

Benjamin Schlegel

07. Juli 2016

Um die Annahme der Homoskedastizität zu prüfen gibt es mehrere Möglichkeiten. In diesem Artikel werden drei Tests 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.

```
modell = lm(hdi ~ gnipc, data=world)
model2 = lm(hdi ~ log(gnipc), data=world)
```

## Goldfeld-Quandt Test

Beim Goldfeld-Quandt Test werden die Varianzen zweier Stichprobenhälften miteinander verglichen. Der Test kann mit der Funktion `gqtest` aus der Bibliothek `lmtest` durchgeführt werden.

```
library(lmtest)
gqtest(modell)
```

```
Goldfeld-Quandt test
```

```
data: modell
GQ = 2.3307, df1 = 92, df2 = 92, p-value =
3.315e-05
```

```
gqtest(model2)
```

```
Goldfeld-Quandt test
```

```
data: model2
GQ = 2.2904, df1 = 92, df2 =
92, p-value = 4.632e-05
```

In beiden Fällen kann die Nullhypothese verworfen werden, dass Homoskedastizität herrscht. Wir haben deshalb Heteroskedastizität.

## Breusch-Pagan Test

Beim Breusch-Pagan Test wird die Beziehung zwischen standardisierten Residuen und den unabhängigen Variablen getestet. Der Test kann mit der Funktion `bptest` aus der Bibliothek `lmtest` durchgeführt werden.

```
library(lmtest)
bptest(modell1)
```

```
studentized Breusch-Pagan test
```

```
data: modell1
BP = 21.412, df = 1, p-value =
3.704e-06
```

```
bptest(modell2)
```

```
studentized Breusch-Pagan test
```

```
data: model2
BP = 0.16647, df = 1, p-value =
0.6833
```

Dieser Test stellt nur noch bei Modell 1 (ohne Logarithmierung) Heteroskedastizität fest, nicht jedoch bei Modell 2.

## White Test

Der White Test ist ein Spezialfall des Breusch-Pagan Tests und reagiert weniger sensibel auf Verletzung der Annahme normalverteilter Fehler. Er kann erst ab zwei unabhängigen Variablen durchgeführt werden. Deshalb wird zuerst ein drittes Modell aufgestellt.

```
modell3 = lm(hdi ~ log(gnipc) + educ_years, data=world)
```

Es wird die gleiche Funktion wie beim Breusch-Pagan Test verwendet. Der Unterschied liegt darin, dass als Formel die Interaktion bei unabhängigen Variablen und deren Quadratterme genommen wird.

```
library(lmtest)
bptest(modell3, ~ log(gnipc)*educ_years + I(log(gnipc)^2)
+ I(educ_years^2), data = world)
```

```
studentized Breusch-Pagan test
```

```
data:  model3  
BP = 9.351, df = 5, p-value = 0.09586
```

Die Nullhypothese kann nicht verworfen werden, weshalb von Homoskedastizität ausgegangen werden kann.

## Robuste Standardfehler nach Huber-White

Nun ist es war schön und recht, wenn man Heteroskedastizität feststellen kann. Doch was muss man tun, wenn es diagnostiziert wird? Da wegen der Heteroskedastizität die Standardfehler zu gross geschätzt werden, müssen diese korrigiert werden. Dies kann mithilfe eine Methode von Huber-White getan werden. Dazu wird die Funktion `vcovHC` aus der Bibliothek `sandwich` verwendet. Mit den üblichen Formel kann aus der gewonnen Varianz-Covarianzmatrix die p-Werte berechnet werden.

```
library(sandwich)  
r.vcov = vcovHC(modell)  
r.se = sqrt(diag(r.vcov))  
r.t = coef(modell) / r.se  
r.p = 2*pnorm(-abs(r.t))  
r.p
```

```
(Intercept)          gnipc  
4.630001e-261  1.386331e-08
```

Mit den berechneten p-Werten kann eine aufgestellte Hypothese geprüft werden.