

Multikollinearität

Benjamin Schlegel

24. Oktober 2016

Dieser Artikel erklärt, wie die Regression auf Multikollinearität geprüft werden kann. 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)
world$educ_years2 = world$educ_years
```

Anschliessend wird der HDI aufdas logarithmierte Bruttonationaleinkommen und Anzahl Bildungsjahre (2x) regressiert.

```
model3 = lm(hdi ~ log(gnipc) + educ_years + educ_years2, data=world)
```

Da perfekte Multikollinearität vorherrscht setzt R den einen Koeffizienten automatisch auf NA.

Call:

```
lm(formula = hdi ~ log(gnipc) + educ_years + educ_years2, data = world)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.101760	-0.018325	-0.001262	0.021403	0.073430

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.310849	0.019339	-16.07	2e-16 ***
log(gnipc)	0.076540	0.003315	23.09	2e-16 ***
educ_years	0.023590	0.001382	17.07	2e-16 ***
educ_years2	NA	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03232 on 185 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9564

F-statistic: 2050 on 2 and 185 DF, p-value: < 2.2e-16

Wenn keine perfekte Multikollinearität vorherrscht, wird das von R nicht angezeigt. Mithilfe der Funktion [vif](#) aus der Bibliothek [car](#) kann Multikollinearität festgestellt werden. Wenn die Wurzel des Wertes grösser als 2 ist, ist eine starke Multikollinearität vorhanden.

```
model3 = lm(hdi ~ log(gnipc) + educ_years, data=world)
```

```
library(car)  
sqrt(vif(model3)) > 2
```

```
log(gnipc) educ_years  
FALSE FALSE
```

Es besteht damit kein Problem der Multikollinearität. Wenn das Problem auftritt, sollte man drüber nachdenken einer der Prädiktoren im Modell wegzulassen.