

Lagemasse und Streuung in R

Benjamin Schlegel

14. März 2016

Mit R können mit mehr oder weniger Aufwand die verschiedenen [Lagemasse und Streuungen](#) berechnet werden. Einige der Funktionen sind standardmässig verfügbar. Für andere müssen Bibliotheken geladen werden. Und für ein paar wenige muss selber ein Funktion in R geschrieben werden.

Damit die Beispiele ausgeführt werden können, muss zuerst der Datensatz World Data in R geladen werden, welcher unter [Data](#) heruntergeladen werden kann. Wie ein Datensatz eingelesen werden kann, wird in [R Grundlagen](#) erklärt.

```
world = read.csv2("world_data.csv",stringsAsFactors=F)
world$fertility_2012 = as.numeric(world$fertility_2012)
```

Lagemasse

arithmetische Mittel

Das arithmetische Mittel kann in R mit der Funktion `mean()` berechnet werden.

```
mean(world$fertility_2012,na.rm=T)
```

```
[1] 2.900959
```

Die Länder haben eine durchschnittliche Geburtenrate von 2.9. Das `na.rm=T` wurde gesetzt, damit die NA ignoriert werden. Da es in diesem Datensatz keine NA gibt, wäre das `na.rm=T` nicht nötig gewesen. Wird das `na.rm=T` bei einer Variable weggelassen, welche NAs enthält, gibt die Funktion NA zurück. Das hat damit zu tun, dass die Funktion den Wert nicht berechnen kann, da nicht alle Werte bekannt sind. Mit `na.rm=T` werden die NAs ignoriert und man erhält in einen Wert zurück, auch wenn einzelne Werte fehlen.

geometrische Mittel

Das geometrische Mittel kann mit den Grundfunktionen vor R nicht direkt berechnet werden. Die Bibliothek `psych` enthält die Funktion `geometric.mean()`. Die Bibliothek enthält auch eine Funktion um das harmonische Mittel zu berechnen (`harmonic.mean()`). Im Beispiel wird der HDI von einem fiktiven Land berechnet.

```
library(psych)
x = c(0.75,0.8,0.3)
geometric.mean(x)
```

```
[1] 0.5646216
```

Median

Der Median kann mit der Funktion `median()` berechnet werden. Die Werte müssen numerisch sein, d.h. das ordinale Variablen zuerst in eine numerische Variable recodiert werden müssen.

```
median(world$fertility_2012,na.rm=T)
```

```
[1] 2.405
```

Die Hälfte der Länder hat eine Geburtenrate von unter 2.405.

Modus

In R gibt es keine Funktion um den Modus zu berechnen. Man kann jedoch selber eine Funktion definieren.

```
Mode = function(x,na.rm=FALSE) {  
  if(na.rm){  
    x = na.omit(x)  
  }  
  ux = unique(x)  
  return(ux[which.max(tabulate(match(x, ux)))] )  
}  
Mode(world$educ_2012,na.rm=T)
```

```
[1] 9
```

Neun Jahre obligatorische Schulzeit kommt am meisten vor.

Quantile

Quantile können in R mit der Funktion `quantile()` berechnet werden. Beim Parameter `probs` werden die Prozentwerte angegeben, von denen man die Werte berechnet haben will. 0.5 ist dabei der Median. Zuerst werden die Quartile berechnet, anschliessend die Quintile.

```
quantile(world$fertility_2012, probs=c(0.25,0.5,0.75),na.rm=T)
```

```
   25%   50%   75%  
1.79225 2.40500 3.84250
```

```
quantile(world$fertility_2012, probs=c(0.2,0.4,0.6,0.8),na.rm=T)
```

```
   20%   40%   60%   80%  
1.6120 2.0972 2.7114 4.4574
```

Streuung

Bandbreite

Um die Bandbreite in R zu berechnen, wird das `min()` vom `max()` abgezogen. Die in R enthaltene Funktion `range()` berechnet nicht die Bandbreite, sondern nur das Minimum und das Maximum.

```
max(world$fertility_2012, na.rm=T) -  
  min(world$fertility_2012, na.rm=T)
```

```
[1] 6.314
```

IQR

Der Interquartilsabstand kann in R mit der Funktion `IQR()` berechnet werden. Mit dem `type` kann die Berechnungsmethode abgegeben werden.

```
IQR(world$fertility_2012,na.rm=T,type=6)
```

```
[1] 2.0785
```

Varianz und Standardabweichung

Sowohl zur Berechnung der Varianz (`var()`) als auch der Standardabweichung (`sd()`) stehen Funktionen in R zur Verfügung.

```
var(world$fertility_2012,na.rm=T)
```

```
[1] 2.1236
```

```
sd(world$fertility_2012, na.rm=T)
```

```
[1] 1.457258
```

In R kann auch gut gezeigt werden, dass die Varianz das Quadrat der Standardabweichung ist.

```
sd(world$fertility_2012, na.rm=T)^2 == var(world$fertility_2012,na.rm=T)
```

```
[1] TRUE
```

Schiefe

Um die Schiefe zu berechnen, kann die Funktion `skewness()` aus der Bibliothek `moments` verwendet werden.

```
library(moments)  
skewness(world$fertility_2012,na.rm=T)
```

```
[1] 0.9748212
```

Wölbung

Um die Wölbung zu berechnen, kann die Funktion `kurtosis()` aus der Bibliothek `moments` verwendet werden. Will man den Exzess berechnen, zieht man 3 vom Resultat ab.

```
library(moments)
kurtosis(world$fertility_2012,na.rm=T)
```

```
[1] 2.983598
```

```
library(moments)
kurtosis(world$fertility_2012,na.rm=T) - 3
```

```
[1] -0.01640233
```

Variationsverhältnis

Um das Variationsverhältnis zu berechnen, muss eine eigene Funktion geschrieben werden. Die Funktion mach dabei gebraucht von der weiter oben definierten `Mode()` Funktion.

```
variation.ratio = function(x,na.rm=FALSE){
  if(na.rm){
    x = na.omit(x)
  }
  return(1 - sum(x==Mode(x,na.rm))/length(x))
}
variation.ratio(world$educ_2012)
```

```
[1] 0.7197802
```

Bei der Bildung haben etwa 72 Prozent der Länder nicht neun obligatorische Schuljahre (also mehr oder weniger).