

## Statistische Software (R)

Paul Fink, M.Sc.

Institut für Statistik  
Ludwig-Maximilians-Universität München

*Datenbeispiel*



## Aufgaben

1. Laden Sie den Datensatz `erwerbstaetige.txt` von der Veranstaltungs-Homepage ([http://www.statistik.lmu.de/institut/ag/agmg/lehre/2014\\_SoSe/StatSoft/material/erwerbstaetige.txt](http://www.statistik.lmu.de/institut/ag/agmg/lehre/2014_SoSe/StatSoft/material/erwerbstaetige.txt)) herunter und lesen Sie ihn in R ein und bringen Sie alle Variablen in ein sinnvolles Format.

Tipp: Konvertieren Sie Anzahlen nicht gleich in Faktoren.

Tipp: Die 1000er Trennzeichen könnten Probleme machen.

2. Erstellen Sie 2 neue Variablen des Datensatzes (`erwerbstaetige` und `erwerbspersonen`), welche die Gesamtzahl der Erwerbstätigen (Arbeitnehmer und Selbstständige) sowie Erwerbspersonen (Erwerbstätige und Erwerbslose) je Jahr enthalten.

Tipp: `?transform`

## Datenbeispiel

Das Statistische Bundesamt veröffentlicht jedes Jahr die Zahl der Erwerbstätigen Inländer in Deutschland (<https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Arbeitsmarkt/Erwerbstaetigkeit/Erwerbstaetigkeit.html#Tabellen>).

Die Daten können auch für die Jahre von 1970–2013 eingesehen werden. Als Variablen gibt es unter anderem die Anzahl Gesamtbevölkerung, der Arbeitnehmer, der Selbstständigen und Erwerbslosen (je in 1000).

Weil die Datenaufbereitung im Fokus steht, sind die Daten bereits als Textdatei auf der Veranstaltungs-Homepage hinterlegt.

Fink: Statistische Software (R) SoSe 2014

1

## Aufgaben

3. Erstellen Sie eine weitere Variable, die den Anteil der Nicht-Erwerbspersonen an der Gesamtbevölkerung enthält.

Lassen Sie sich eine statistische Kurzzusammenfassung dafür ausgeben und bestimmen Sie die Jahre, in denen dieser am höchsten bzw. niedrigsten war.

Tipp: Die Funktionen `which.min` und `which.max` werden Ihnen weiterhelfen.

## Aufgaben

---

### 4. Visualisieren Sie

- (a) die empirische Dichtefunktion des Arbeitnehmeranteils an der Bevölkerung und
- (b) den zeitlichen Verlauf des Ausländeranteils im Verlauf der Jahre.

Tipp: Die Funktion `plot(x,y)` zeichnet einen Scatterplot für 2 gleich lange Vektoren `x` und `y`, wobei sich ein Punkt aus  $(x_i, y_i)$  ergibt.

## Aufgaben

---

6. Fassen Sie die in Aufgabe 5 gebildeten Variablen zu einem neuen Datensatz `aenderung` zusammen und fügen Sie diesem noch eine Variable `jahr` hinzu, die für jede Änderung das Jahr angibt mit dem das Vorjahr verglichen wird, also die Werte 1971 bis 2013 enthält.

## Aufgaben

---

5. Jetzt wollen wir wissen, wie hoch die jeweilige Änderung der Anteile an der Gesamtbevölkerung von Jahr zu Jahr sind. Erstellen Sie dazu 3 neue Variablen (nicht als Teils des Datensatzes `erwerbstaetige`).

- `diff_arbeitnehmer` für die Änderung des Arbeitnehmeranteils,
- `diff_selbststaendige` für die Änderung des Selbstständigenanteils und
- `diff_erwerbslose` für die Änderung des Erwerbslosenanteils.

Tipp: `?diff`

Tipp: Mit `with` sparen Sie sich etwas Schreibarbeit.