

12 Statistik mit R

Aufgabe 1:

Der Datensatz `occupationalStatus` enthält den Beschäftigungsstatus von Vätern und Söhnen. Untersuchen Sie, ob es einen Zusammenhang gibt, wenn Sie den Status als nominales und als ordinales Merkmal auffassen.

Aufgabe 2:

Der Datensatz `alkfos` aus dem Paket **ISwR** enthält wiederholte Messungen von Alkaline Phosphatase in einem randomisierten Experiment von Tamoxifen Behandlung von Brustkrebspatientinnen.

- Visualisieren Sie die Daten geeignet.
- Untersuchen Sie, ob es Unterschiede zwischen den beiden Gruppen zum Zeitpunkt 0 gibt.
- Führt die Behandlung mit Tamoxifen zu unterschiedlichen Alkaline Phosphatasewerte bei den Patientinnen?

Nach dem Installieren des Pakets (z.B. mit `install.packages("ISwR")`) können Sie die Daten laden mit

```
> data("alkfos", package = "ISwR")
```

Aufgabe 3:

Es sollen bivariate normalverteilte Zufallszahlen mit vorgegebene Korrelation und jeweiliger Varianz erzeugt werden. Schreiben Sie dazu folgende Funktion:

```
> rbinorm <- function(n, mean = rep(0, 2), sd = rep(1, 2), rho = 0) {  
+   ...  
+ }
```

Dabei ist `sd` der Vektor der Standardabweichungen und `rho` gibt die Korrelation an.

Die Funktion soll folgendermaßen vorgehen:

- Erzeugen Sie 2 `n` unabhängig standardnormalverteilte Zufallsvariablen und speichern Sie sie in einer Matrix mit `n` Zeilen und 2 Spalten ab.
- Transformieren Sie die Werte in der zweiten Spalte x_2 folgendermaßen:

$$x_2 = \text{rho} \cdot x_1 + \sqrt{1 - \text{rho}^2} \cdot x_2$$

Dadurch bekommt man die gewünschte Korrelation.

- Transformieren Sie jetzt die Spalten noch, um die gewünschte Standardabweichungen und Mittelwerte zu bekommen.

Aufgabe 4:

Schreiben Sie analog zu `weighted.mean` eine Funktion `weighted.sd`, die die gewichtete Standardabweichung berechnet.

Aufgabe 5:

Laden Sie den Datensatz `eusilc` aus dem Paket `laeken`. Der Datensatz ist ein künstlicher Datensatz, der jedoch von der Struktur dem Original-Datensatz von EU-SILC (EU Statistics on Income and Living Conditions) entspricht.

Jeder Beobachtung im Datensatz ist ein Personengewicht (`rb050`) und ein Haushaltsgewicht (`db090`) zugeordnet. Diese drücken aus wie viele Personen / Haushalte in der Gesamtpopulation durch diese Beobachtung repräsentiert werden.

Bestimmen Sie mit und ohne die Gewichte folgendes:

- Durchschnittliche Haushaltsgröße und deren Standardabweichung
- Bevölkerungsanzahl der Bundesländer
- Geschlechtsverteilung

Interpretieren Sie die Ergebnisse.