

Aufgabenblatt zur Einführung in R

Freiwillige Abgabe. Besprechung am **25.05.2010 um 13:30** im SRA.

1 Datenstrukturen

Aufgabe 1.1. Die Teilnehmer eines Hochschulsport-Tanzkurses wurden nach Alter und Geschlecht befragt. Dabei ergaben sich folgende Daten:

Person	1	2	3	4	5	6	7	8	9	10
Geschlecht	w	w	m	w	m	m	w	w	m	m
Alter	23	24	26	27	23	28	21	20	25	26

Erstellen Sie aus diesem Datensatz eine Datentabelle `tanz` in R und untersuchen Sie:

- Wie alt sind der älteste und der jüngste Teilnehmer in der gesamten Gruppe? Wie alt sind der (die) älteste und der (die) jüngste bei den Männern (Frauen)?
- Benutzen Sie den Befehl `attach` um direkt auf die Spalte `Alter` zugreifen zu können und ordnen Sie diesen Vektor der Größe nach, einmal aufsteigend und einmal absteigend.
- Wie alt sind die Teilnehmer zusammen? Wie alt sind sie durchschnittlich?
- Welche Teilnehmer sind 23 Jahre alt? Welche sind älter als 21 und jünger als 27? Welche sind mindestens 25 Jahre alt oder genau 23 Jahre alt?
- Erstellen Sie mit Hilfe von `subset` einen Datenvektor `Alter.f`, der nur aus dem Alter der Studentinnen besteht. Wie alt sind die Studentinnen durchschnittlich?

Aufgabe 1.2. In R gibt es eingebaute Datensätze, die bei Eingabe ihres Namens aufgerufen werden und die im Hilfetext näher beschrieben werden. Einer davon ist der Datensatz `Titanic`.

- Beschreiben Sie die Struktur des Datensatzes `Titanic`: In welcher Form liegen die Daten vor und wie viele Dimensionen werden belegt? Wie viele Merkmale werden unterschieden und in welchen Dimensionen sind die entsprechenden Werte eingetragen?
- Reduzieren Sie die im Datensatz enthaltenen Informationen durch Summation, so dass lediglich eine Matrix übrigbleibt, die aus den Merkmalen `Sex` und `Survived` besteht. Bestimmen Sie die Anzahl aller Männer und aller Frauen sowie die jeweiligen Überlebensquoten.

Aufgabe 1.3. Mit dem Befehl `rnorm(16,mean=5,sd=9)` erhalten Sie 16 Zufallszahlen, die gemäß einer Normalverteilung mit Erwartungswert 5 und Standardabweichung 9 gesampelt wurden.

- (a) Runden Sie diese Zahlen auf eine Stelle nach dem Komma und tragen Sie sie zeilenweise in eine 4×4 Matrix `M` ein.
- (b) An welcher Position der Matrix steht das Minimum der Einträge? An welcher das Maximum?
- (c) Gibt es einen Eintrag, der im Intervall $[0, 1]$ liegt? Liegen alle Einträge zwischen -2 und 12 ?
- (d) Bilden Sie, ohne Verwendung der Funktion `colSums`, die Spaltensummen von `M` und schreiben Sie diese in einen Vektor `sp.sums`. In welcher Spalte ist die Summe am höchsten?

2 Programmierung, Schleifen und Abfragen

Aufgabe 2.1. (a) Erzeugen Sie den Vektor

$$v = (1, 2, 2, 3, 3, 3, \dots, \underbrace{100, \dots, 100}_{100 \text{ mal}})$$

mit einer `for` Schleife und ohne Verwendung der Funktion `rep`. Wie könnte man v mit `rep` erzeugen?

(b) Erzeugen Sie den Vektor

$$w = (1, 1.2, 1.4, 1.6, \dots, 99.6, 99.8, 100)$$

mit einer `while` Schleife und ohne Verwendung der Funktion `seq`. Wie könnte man w mit `seq` erzeugen?

Aufgabe 2.2. Die Folge $a = (a_n)_{n \in \mathbb{N}}$ sei rekursiv definiert durch $a_1 = a_2 = 1$ und

$$a_n = n + a_{n-1} + a_{n-2}$$

für $n \geq 3$.

- (a) Berechnen Sie die ersten 20 Folgenglieder von a .
- (b) Berechnen Sie die Folgenglieder, die ≤ 100000 sind. Wieviele sind es?
- (c) Wieviele davon sind auch ≥ 5000 ?

Aufgabe 2.3. Auf dem Wochenmarkt hat der örtliche Lakritz- und Weingummihändler folgendes Angebot:

- Bei jedem Kauf zahlt man 50 Cent „Bedienungsgebühr“ und muss ihm mindestens 250g seiner Ware abnehmen.
- Bei einem Kauf von bis zu 500g Süßigkeiten zahlt man pro 100g 0,60 Euro.
- Bei mehr als 500g und höchstens 1000g zahlt man 0,50 Euro pro 100g.
- Bei mehr als einem Kilo 40 Cent pro 100g.

Stellen Sie in R den unter diesen Bedingungen entstehenden Preis als Funktion der gekauften Menge dar und berechnen Sie damit den Preis für 384g, 496g, 502g, 980g und 1107g.

Nützlich ist die Funktion `stop("text")`. Sie bricht eine Funktion ab und liefert die Fehlermeldung "text".

3 Plots

Aufgabe 3.1. Bei der letzten Landtagswahl in NRW (im Mai 2010) ergab sich folgende Stimmverteilung beim Merkmal *Zweitstimme* (bei 7.759.325 gültigen Stimmen):

CDU	SPD	FDP	Grüne	Linkspartei	Andere
2.681.736	2.675.536	522.437	940.770	434.846	504.000

- Geben Sie die Daten als Vektor `wahl` ein, weisen Sie mit `names(wahl)` die Parteien den Zahlen zu und ordnen Sie den Vektor absteigend.
- Berechnen Sie die zugehörigen prozentualen Anteile an den abgegebenen (und gültigen) Stimmen auf eine Nachkommastelle genau.
- Erzeugen Sie mit den Daten ein, mit den Parteinamen und den zugehörigen Prozentzahlen beschriftetes, Kreisdiagramm (in den entsprechenden Parteifarben). Verwenden Sie dazu die Funktion `pie`.
- Erstellen Sie ein geordnetes Säulendiagramm in den entsprechenden Parteifarben. Verwenden Sie dazu die Funktion `barplot`.

Aufgabe 3.2. Zeichnen Sie die Dichten der $\mathcal{N}(0, 1)$ -Verteilung und der $\mathcal{C}(0, 1)$ -Verteilung auf dem Intervall $[-4, 4]$ in eine Grafik, beschriften Sie diese und fügen Sie eine Legende ein.

Hier wird mit $\mathcal{C}(a, b)$ eine Cauchyverteilung mit Verschiebungsparameter a und Skalierungsparameter b bezeichnet. An die Dichten der Verteilungen gelangen sie mit `dnorm(x)` bzw. `dcauchy(x)`. Für das Plotten in R empfiehlt sich die Funktion `curve`. Das Einfügen einer Legende in einen Plot geschieht mit dem Befehl `legend`.