

## Übungen

Abgabetermin: Mittwoch, 29.04.2009; 09:15 Uhr, BK 41

### Aufgabe 1. (5 Punkte)

Mit den Daten der Enzyklopädie *The Marshall Illustrated Encyclopedia of Animals* lässt sich die folgende Tabelle erstellen:

Spezies	Familie	max. Größe	Tragezeit	Schwanz
<i>Leontopithecus rosalia</i> (Golden Lion Tamarin)	Callitrichidae	22cm	133T	ja
<i>Papio ursinus</i> (Chacma Baboon)	Cercopithecidae	100cm	184T	ja
<i>Mandrillus sphinx</i> (Mandrill)	Cercopithecidae	95cm	225T	ja
<i>Hylobates concolor</i> (Black Gibbon)	Hylobatidae	63cm	225T	nein
<i>Pan paniscus</i> (Bonobo)	Hominidae	60cm	230T	nein
<i>Pan troglodytes</i> (Chimpanzee)	Hominidae	94cm	230T	nein

- (a) Bestimmen Sie das Skalenniveau der Merkmale.
- (b) Erzeugen Sie aus den Spalten der obigen Tabelle die Vektoren `species`, `family`, `body`, `gestation`, `tail`.  
Lassen Sie sich anschließend die Ausprägungen der nominalskalierten Merkmale ausgeben.
- (c) Erzeugen Sie eine Datentabelle, dessen Spalten aus den Vektoren `species`, `family`, `body`, `gestation`, `tail` besteht, und lassen Sie die Tabelle ausgeben.

### Aufgabe 2. (5 Punkte)

Eine Folge unabhängiger, auf  $(0, 1)$  gleichverteilter Zufallsgrößen realisiert die Zahlen

0,495 0,166 0,550 0,461 0,390 0,706 0,430 0,138 0,406 0,637.

- (a) Geben Sie die Daten als Vektor ein und weisen Sie sie der Variablen `Unif` zu.
- (b) Erzeugen Sie einen zweiten Vektor, indem Sie die Komponenten des Vektors auf 2 Nachkommastellen runden und sortieren Sie anschließend die Komponenten absteigend.
- (c) Bilden Sie einen neuen Vektor, indem Sie die Funktion  $x \mapsto -\log x$  auf die Komponenten des Vektors `Unif` anwenden und die Komponenten anschließend auf 3 Nachkommastellen runden. Transformieren Sie den entstehenden Vektor mit der Abbildung  $(x_1, \dots, x_n) \mapsto (x_1, x_1 + x_2, \dots, x_1 + \dots + x_n)$  und weisen Sie den neuen Vektor der Variablen `CumulExp` zu.

(Bitte wenden!)

- (d) Erzeugen Sie die Variable `CumulExp2`, indem Sie `CumulExp` um eine 0 im ersten Argument ergänzen. Erstellen Sie eine Graphik, indem Sie die Komponenten von `CumulExp2` auf der  $y$ -Achse gegen die natürlichen Zahlen (beginnend bei 0) auf der  $x$ -Achse abtragen. Beschriften Sie die  $x$ -Achse mit  $n$  und die  $y$ -Achse mit  $S_n$ . Die Beschriftung der Ordinate werde orthogonal zur Ordinate vorgenommen.
- (e) Die Komponenten des Vektors `CumulExp` können als Wartezeiten auf aufeinanderfolgende Ereignisse interpretiert werden. Schreiben Sie eine Funktion, die einer reellen Zahl  $t \in [0, 9]$  die Zahl  $N_t$  der bis zum Zeitpunkt  $t$  eingetretenen Ereignisse zuweist. Verwenden Sie dabei eine (sinnvolle) `for`-Schleife.

### Aufgabe 3. (5 Punkte)

Erzeugen Sie vier Bilder in einer Graphik, die jeweils (unterschiedliche) Approximationen von Pfaden einer zweidimensionalen (Standard-)Brownschen Bewegung darstellen. Lassen Sie dazu viermal die Werte einer zweidimensionalen Brownschen Bewegung in den Punkten  $k/n$  mit  $n = 10.000$  und  $k = 1, \dots, n$  von R randomisieren und interpolieren Sie dazwischen linear. Beschriften Sie die Achsen mit  $B_t^{(1)}$  bzw.  $B_t^{(2)}$ , wobei Sie wie üblich die Beschriftung der Ordinate orthogonal zur Ordinate vornehmen. Überschreiben Sie die Bilder mit Pfad 1 bis Pfad 4.

### Aufgabe 4. (5 Punkte)

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`.

- (a) 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?
- (b) 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.