Aufgabenblatt

Die Aufgaben können in **2er Gruppen** bearbeitet werden. Für jede Aufgabe sollen Sie ein R Skript erstellen, das Sie als „`Aufgabennr.Vorname1.Vorname2.R`" speichern, also z.B. „`3.2.Bernd.Ute.R`" für das R Skript zur Aufgabe 3.2.
Abgabe: **16.02.2010 bis 18:00 Uhr** per Email. Besprechung: **17.02.2010 um 13:30** im SRA.

1 Datenstrukturen

Aufgabe 1.1. (5 Punkte)
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:

(a) 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)?

(b) Benutzen Sie den Befehle `attach` um direkt auf die Spalte `Alter` zugreifen zu können und ordnen Sie diesen Vektor der Größe nach, einmal aufsteigend und einmal absteigend.

(c) Wie alt sind die Teilnehmer zusammen? Wie alt sind sie durchschnittlich?

(d) 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?

(e) 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. (4 Punkte)
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?

Aufgabe 1.3. (5 Punkte)
Herr Weishaupt misst jeden Mittag um 12:00 Uhr die Temperatur in Grad Celsius auf seinem Balkon und notiert sie in seinem Kalender. Für den Januar 2007 hat er sich folgende Daten eingetragen:

1. Woche:	-2	-4	-3.5	-3	-7	-6	-4
2. Woche:	-1	0	0.5	2	1	2	2.5
3. Woche:	0	-1	0	1.5	3	3	2
4. Woche:	4	3	1	-2	-3	-4	0

Erzeugen Sie mittels `ts` eine Zeitreihe `mess`. Achten Sie dabei darauf, einen Startwert und die richtige `frequency` anzugeben.

(a) Welche Werte sind größer als 0?

(b) An welchen Tagen hat es gefroren (Temperatur negativ)? An welchen Tagen lag die Temperatur bei genau 0 Grad Celsius?

Herr Weishaupts Nachbar misst ebenfalls um 12:00 Uhr die Temperatur und notiert diese. Für den gleichen Zeitraum erhält er die folgenden Werte:

1. Woche:	-1.5	-4	-3	-3.5	-7	-4	-3
2. Woche:	0	-1	1	2	1.5	2	3
3. Woche:	0	-1	0	-0.5	3	2.5	2.5
4. Woche:	4.5	3	1.5	-2	-3	-4	0.5

Erzeugen Sie aus diesen Daten eine Zeitreihe `mess2`.

(c) Wieviele Übereinstimmungen mit der ersten Messung gibt es?

(d) Was ist die größte Abweichung, und wann fand sie statt?

(e) Benutzen sie den Befehl `ts.union`, um aus den beiden Zeitreihen eine einzelne Zeitreihe `Messung` zu erzeugen.

2 Programmierung, Schleifen und Abfragen

Aufgabe 2.1. (5 Punkte)

(a) Erzeugen Sie den Vektor

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

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

(b) Erzeugen Sie den Vektor

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

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

Aufgabe 2.2. (4 Punkte)
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. (5 Punkte)
Auf dem Wochenmarkt hat ein 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. (5 Punkte)
Bei der letzten Bundestagswahl in Deutschland (im September 2009) ergab sich folgende Stimmverteilung beim Merkmal *Zweitstimme* (bei 43.371.190 gültigen Stimmen):

CDU/CSU	SPD	FDP	Grüne	Linkspartei	Andere
14.658.515	9.990.488	6.316.080	4.643.272	5.155.933	2.606.902

(a) Geben Sie die Daten als Vektor `wahl` ein, weisen Sie mit `names(wahl)` die Parteien den Zahlen zu und ordnen Sie den Vektor absteigend.

(b) Berechnen Sie die zugehörigen prozentualen Anteile an den abgegebenen (und gültigen) Stimmen auf eine Nachkommastelle genau.

(c) Erzeugen Sie mit den Daten ein, mit den Parteinamen und den zugehörigen Prozentzahlen beschriftetes, Kreissektorendiagramm (in den entsprechenden Parteifarben). Verwenden Sie dazu die Funktion `pie`.

(d) Erstellen Sie ein geordnetes Säulendiagramm in den entsprechenden Parteifarben. Verwenden Sie dazu die Funktion `barplot`.

Aufgabe 3.2. (4 Punkte)
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 0 und Skalierungsparameter 1 bezeichnet. An die Dichten der Verteilungen gelangen sie mit `dnorm(x)` bzw. `dcauchy(x)`. Für das Plotten in R empfielt sich die Funktion `curve`. Das Einfügen einer Legende in einen Plot geschieht mit dem Befehl `legend`.

Aufgabe 3.3. (5 Punkte)
Laden Sie mit `source("Pfadname")` Ihr Skript der Aufgabe 1.3. Auf die Variable `Messung` können Sie nun zugreifen. Überzeugen Sie sich davon, indem Sie mit `ls()` alle im Workspace definierten Variablen anzeigen lassen.

(a) Führen Sie den Befehl `plot(Messung)` aus. Wie können Sie erreichen, dass beide Zeitreihen in eine Graphik geplottet werden? Färben Sie dabei beide Zeitreihen unterschiedlich, erstellen Sie mit dem für `plot` optionalen Argument `main` eine Überschrift des Plots und beschriften Sie die x- und y-Achse passend (Argumente `xlab` und `ylab`).

(b) Fügen Sie eine Legende ein und zeichnen Sie mit `abline` die x-Achse (in Grau).

(c) Löschen Sie mit `rm` die Variable `Messung` wieder aus dem Workspace.

4 Zeitreihen ohne Saisonkomponente

Aufgabe 4.1. (8 Punkte)
Lesen Sie mit dem Befehl `source` die Datei `temperaturen.R` ein. Sie enthält in der Zeitreihe `Global.annual` die globalen jährlichen Durchschnittstemperaturen von 1856 bis 2005, als Abweichungen vom Durchschnitt der Jahre 1961-1990.

(a) Auf Zeitreiheneinträge können Sie wie auf Vektoreinträge zugreifen. Schreiben Sie eine Funktion, die den einfachen gleitenden Durchschnitt über 3 (5,7,9) Jahre berechnet, und plotten Sie diese.

(b) Verwenden Sie nun die R-Funktion `filter` dafür.

(c) Führen Sie (ohne `HoltWinters`) zu verwenden, eine exponentielle Glättung mit Parametern $\alpha \in \{0.1, 0.3, 0.5, 0.7\}$ durch.

(d) Wählen Sie einen Zeitraum aus, in welchem ein linearer Trend zu erkennen ist. Weisen Sie (mittels des Befehls `window`) einer neuen Zeitreihe die Temperaturdaten aus dem gewählten Zeitraum zu. Berechnen Sie mittels `lm` Achsenabschnitt und Steigung eines linearen Trends.

5 Zeitreihen mit Saisonkomponente

Aufgabe 5.1. (10 Punkte)
Lesen Sie mit dem Befehl `source` die Datei `steuer.R` ein. Sie enthält die deutschen Steuereinnahmen der Jahre 1991 bis 2008 in Euro (frühere Werte umgerechnet) als Quartalsdaten.

(a) Offensichtlich liegen hier saisonale Schwankungen vor. Warum?

(b) Berechnen Sie die gefilterte Zeitreihe der Veränderungen der Steuereinnahmen im Vergleich zum Vorjahresquartal.

(c) Zerlegen Sie die (ursprüngliche) Zeitreihe in Trend (gl. Durchschnitt) und Saisonkomponente, und plotten Sie diese.

(d) Bestimmen Sie mit der Methode der kleinsten Quadrate (also in einem linearen Modell) einen linearen Trend, sowie (additive) Saisonkomponenten. Zentrieren Sie Zeitreihe und die Zeit, um vernünftige Werte zu erhalten.

(e) Betätigen Sie sich als Steuerschätzer, indem Sie ein HoltWinters-Modell für diese Zeitreihe berechnen. Schätzen Sie (unter Annahme unveränderter Wirtschaftslage und Steuergesetzgebung) die Steuereinnahmen für 2009 und 2010, quartalsweise.

(*) Die Funktion `harmonic` aus dem Paket `TSA` erzeugt, gegeben eine Zeitreihe, eine Matrix mit harmonischen Funktionen (sinus und cosinus), mit denen die Saisonkomponente in einem linearen Modell durch Regression approximiert werden kann. Führen Sie dies durch.