

## Arbeitsblatt 1

1) Installieren Sie das Programm R. Es gibt Versionen für alle gängigen Betriebssysteme (Linux, Windows, Mac) auf <http://www.r-project.org/>. Diverse Einführungen finden sich unter <http://cran.r-project.org/manuals.html> und unter <http://www.r-project.org/other-docs.html>

R ist ein Interpreter, d.h. es führt Befehle nach Eingabe aus und präsentiert das Ergebnis. Daher kann man es wie einen Taschenrechner benutzen: Tippen Sie

```
> 2+2
```

antwortet R mit

```
[1] 4
```

Hilfe erhält man mit

```
> ?"+"
```

```
> ?matrix
```

```
> help(log)
```

etc. Die HTML-Version kann mit

```
> options(htmlhelp=TRUE)
```

```
> help.start()
```

gestartet werden. Alternativ kann die HTML-Hilfe direkt in einem Browser geöffnet werden.

Sie können R mit

```
> q()
```

verlassen. Sie werden dann gefragt, ob Sie den sogenannten "workspace" speichern wollen. I.d.R. sollten Sie mit **n** (no) antworten.

2) Die meisten Statistikprogramme benutzen zur Berechnung numerischer Ausdrücke „double precision reals“. Dabei werden Zahlen durch 64 Bits dargestellt: 1 Bit für das Vorzeichen, 52 Bits für die Mantisse und 11 Bits für den Exponenten, also als  $\pm 1.f\!f\!f\dots f \times 2^n$ , wobei  $f\!f\!f\dots f$  für die Mantisse steht und  $n$  für den Exponenten. Dadurch können Zahlen im Bereich zwischen  $-2 \times 10^{308}$  und  $2 \times 10^{308}$  dargestellt werden. Berechnungen in diesem „Zahlensystem“ führen immer zu Rundungs- und Abschneidefehlern. Die maximale Genauigkeit dieser Arithmetik entspricht etwa 15 Dezimalstellen. Die Druckdarstellung (nicht aber die Rechengenauigkeit) der Zahlen kann in R durch `options(digits=5)` verändert werden (7 ist die Voreinstellung).

a) Berechnen Sie  $2 + 3 * (4 + 5 * (6 + 7 * (8 + 9)))$ ,  $2^{11}$ ,  $\log_e(2)$ , und  $\sqrt{333}$ .

b) Ist für das Programm  $1 - 0.8 = 0.2$ ? (Logische Gleichheit wird durch `a == b` abgefragt)

c) Berechnen Sie

$$\frac{1335}{4}y^6 + x^2(11x^2y^2 - y^6 - 121y^4 - 2) + \frac{11}{2}y^8$$

für  $x = 77617$  und  $y = 33096$  (die korrekte Lösung ist -2).

3) Um mit R zu arbeiten, wird ein Editor benötigt, in dem man Programme korrigieren und editieren kann. Eine Liste solcher Editoren ist unter [http://www.sciviews.org/\\_rgui/](http://www.sciviews.org/_rgui/) erhältlich. Ich schlage vor, mit TINN-R zu arbeiten (<http://www.sciviews.org/Tinn-R/>). Installieren Sie TINN-R, legen Sie eine Datei mit dem Namen „test.R“ in einem Arbeitsverzeichnis an und schreiben Sie die folgenden Kommandos in die Datei:

```
a <- c(1,3,5); b <- c(a,1,a)
```

```
c <- 2*a+b+1
```

```
sin(c)
```

Speichern Sie die Datei und versuchen Sie, den Inhalt der Datei aus TINN-R heraus an R zu senden.

4) Wir werden mit den Daten der PSID (Panel Study of Income Dynamics) arbeiten, genauer: mit dem Cross-National Equivalent File (CNEF), 1980–2005. Es handelt sich um eine große Längsschnitt-Befragung in den USA.

Datensatz und Codebücher sind über <http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/cnef.cfm> zugänglich.

Wir benötigen das 'PSID-CNEF codebook', Teil 1 und 2; sowie den Datensatz: `pspsequiv.zip`. Speichern Sie alle drei Dateien in ein separates Verzeichnis. Der Datensatz muss nun entpackt werden, das Ergebnis sind 21 `.sav`-Dateien (SPSS Dateiformat).

In R lässt sich der Datensatz folgendermaßen einlesen:

```
library(foreign)
```

```
X <- read.spss("D:/Stattech/daten/pequiv_2003.sav", to.data.frame=T,  
use=F)
```

Die Variablenamen stehen zur Verfügung:

```
names(X)
```

```
X$X11101LL
```