

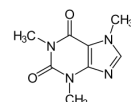
# Einführung in die statistische Programmierung mit R

## für epidemiologische (und andere) Fragestellungen

### Übungen

*Alle Aufgaben sind mit den bisher gelernten Konstrukten lösbar. Natürlich dürft ihr euch zusammensetzen, und eure Lösungsansätze diskutieren. Der Quelltext und die Kommentare sollten grundsätzlich auf Englisch geschrieben werden. Besonders schwere Aufgaben werden von wilden Tieren bewacht und können erstmal übersprungen werden. Die Einnahme leistungssteigernder Psychopharmaka ist gestattet. Um ein Auslösen der Rauchmelder zu vermeiden, wird von zu allzu verbissenem Nachdenken abgeraten.*

*Viel Spaß!*



### Einfache Arithmetik

**0** Mache dich mit den Befehlen der Kommandozeile, die wir heute gelernt haben, soweit vertraut, dass du sie benutzen kannst. Für den Editor sollte eine Zusammenfassung der Shortcuts in einem Ordner in eurem home-Verzeichnis auf den virtuellen Systemen zu finden sein.

**1** Das Thermometer zeigt 32 °C. Mit wieviel K muss der Thermodynamiker rechnen?

**2** Wieviel ml Ethanol sind ungefähr in einer Doppelmaß Märzen mit 4.8 Vol.-% enthalten?

**3** Wieviel ha sind 1001 Morgen Ackerland?

**4** 2223 extrem nervige Kinder streiten sich der Geburtstagsfeier eines verwöhnten Bengels reicher Eltern um 7777 Kuchenstücke. Wieviele Kuchenstücke bleiben bei Gleichverteilung ganzer Kuchenstücke an alle Blagen übrig? Wieviele Stücke erhält jedes Kind von den verteilten Stücken? Ein Kuchenstück wiegt 150 g. Wieviel Kuchen futtert jede Rotznase bei sozialistischer Gleichverteilung (nur das Geburtstagskind ist noch gleicher als alle anderen und verputzt 11 Stücke allein.)

**5** Der Reproduktionsfaktor einer infektiösen Erkrankung beträgt 3. Wieviel Personen sind durch einen einzelnen Quellfall, verursacht durch die Nichteinhaltung von Hygienemaßnahmen eines illegalen Kindergeburtstages, nach 3 Ansteckungsgenerationen infiziert, wenn jede Person jeweils gleichviel weitere Personen ansteckt?

### Variablen, Assignments

**6** Gib einer Zahl  $n$  einen beliebigen Wert  $n > (10000)_{10}$  und berechne jeweils die Zahl, die bei Aneinanderreihung der ersten 3 und letzten 3 Ziffern entsteht, wenn  $n$  im Dezimalsystem dargestellt wird. (Für Informatiker: Berechne die gleichen Zahlen nach der Konvertierung der Darstellung von  $n$  ins Oktalsystem)

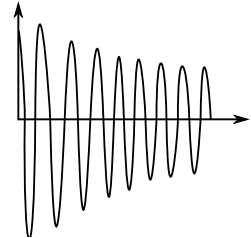
7 Wo schneidet die Parabel  $p(x)$  mit  $\mathbb{R} \mapsto \mathbb{R} \mid p(x) = 5x^2 + 6x - 7$  die x-Achse?

8 Der Reproduktionsfaktor einer infektiösen Erkrankung beträgt 3. Wieviel Personen sind durch einen einzelnen Quellfall nach 4 Ansteckungsgenerationen infiziert, wenn jede Person jeweils gleichviel weitere Personen ansteckt, sich aber der Reproduktionsfaktor sich in der zweiten Generation aufgrund einer Mutation bei einem Fall verändert und diese Person bzw. ihre Folgefälle jeweils 5 Personen ansteckt?

### Numerische Funktionen

9  $N(0) = 1.2 \times 10^5$  Viruskopien infizieren ein Gewebe. Die Wachstumskonstante  $\lambda$  des exponentiellen Wachstums  $N(t) = N(0) \cdot e^{\lambda t}$  beträgt  $\lambda = 0.3 \text{ h}^{-1}$ . Wieviele Viruskopien sind nach 3.5 h, wieviele nach 7.5 h ungefähr entstanden?

10 Die Auslenkung  $I$  einer Feder wird als gedämpfte Schwingung durch  $I(t) = I(0) \cdot \cos(\omega \cdot t) \cdot e^{dt}$  beschrieben. Die Dämpfungskonstante  $d$  beträgt  $d = 0.25 \text{ cm}$  und die Schwingungsfrequenz  $f = 25 \text{ Hz}$ . In welcher Position befindet sich die Feder nach exakt 5 s bei einer Auslenkung von  $I(0) = 4 \text{ cm}$ ? Wie weit muss die Feder ausgelenkt werden, so dass das System nach 4 min noch einen Abst von minimaler und maximaler Auslenkung  $|\Delta I|(t)$  von mindestens 6 min cm aufweist?



Tipps:

- Die Kreisfrequenz  $\omega$  beträgt  $\omega = 2\pi f$
- Wir setzen gehen davon aus, dass der maximale Auslenkungszustand und den unmittelbar folgenden minimale Zustand den gleichen Betrag haben, d.h.  $I(k2\pi f) \approx -I((k2 + 1)\pi f)$ ,  $k \in \mathbb{N}$
- Daher können wir die maximale Auslenkung als  $|\Delta I|(t) \approx 2 \cdot I(0) \cdot e^{dt}$  annähern.
- Die vergangene Zeit  $t$  lässt sich durch  $t = kf^{-1}$  beschreiben.

**print()**

### Kommentare

#### Vector: Access-Operatoren, Subsets

11 Wo liegt der Schwerpunkt (für ganz genaue „Massenmittelpunkt“) eines Objektes, das sich aus punktförmigen Massen  $m_i$  an den Orten  $\vec{p}_i$  zusammensetzt?

$$m_1 = 23 \mu\text{g}, \quad m_2 = 77 \mu\text{g}, \quad m_3 = 56 \mu\text{g},$$

$$\vec{p}_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \vec{p}_2 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \vec{p}_3 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

**12** Mache dich mit der Funktion `runif(n, min = 0, max = 1)` vertraut. Erstelle dann einen **Vector** namens `alpha` mit 5 Random-Werte zwischen 1 und 100 und einen **Vector** namens `beta` mit 100 Random-Werten des gleichen Intervalls. Vertausche anschließend die Inhalte der beiden Objekt, ohne irgendeine spezifische swap-Funktion zu verwenden.

**13** Erstelle einen **Vector** der Länge 1000, bei dem sich die Zahlen 0 und 1 abwechseln.

**14** Erstelle einen **Vector**

**15** Wieviele Zahlen zwischen 111111 und 77777777 enden mit einer 6 und sind durch 23 teilbar?

**list**

**dataframe**

**dataframe-Manipulation (Subsetting, Merging, Renaming)**

**Skalentypen (Theorie)**

**factor Kategorien)**

**Summen, Varianzen, Standardabweichung(en)**

**character, String-Manipulation, Substrings, Konkatenation**

**\*.csv-Dateien lesen**

**logical**

**Funktionales Ifelse(test, yes, no)**

**Funktionen**

**\*.csv-Dateien schreiben**

**Textdateien schreiben**

**Quantile, Median, Mittelwerte**

**R-base-Plots: dots, lines, boxplots**

**Lineare Regression**

**Residuenanalyse**

**Korrelation**

**Funktionelle Programmierung**

(Example: Solve eqs of type  $(x^2+m)=t$ )

**Minima, Maxima**

**Vektorisierung**

**pmin, pmax**

**(x)apply-Schleifen (I)**

**Achsentransformationen**

**Guter Coding-Stil**

**Type Casting**

**matrix und array**

**Grundlegende lineare Algebra**

**Kontingenztafeln**

**(x)apply-Schleifen (II)**

**Imperative Verzweigung (I), if-else**

**Imperative Schleifen (I), for-while**

**16 SEIR-Modell**

**Imperative Verzweigung (II), switch-case**

**Imperative Schleifen (II), break-continue**

**flags**

**Packages**

**Locals**

**Reshape2**

**Binomialkoeffizienten**

**t-test**

**Logistische Regression**

**Wilcoxon-Rangsumme**

**ANOVA**

**Imperative Schleifen (III), foreach**

**Tiles, pdf printing**

**Code-Strukturierung**

**Package-Zitation**

**Packages erstellen**

**magrittr**

**Tydiverse - tibble**

**Tydiverse - dplyr**

**Tydiverse - dplyr2**

**R to markdown**

**Bioinformatik - sequence alignments, Needleman-Wunsch / Smith Waterman, logits, blosum matrix, dynamic programming, multiple sequence alignments, pedigrees**

**knitr**

**RShiny**

**SQL-Interfaces**