Einführung in das Statistikprogramm R

Übungen

1. *R-Grundlagen*

a. Verwenden Sie R als Taschenrechner und berechnen Sie die folgenden Aufgaben:

```
13 + 78 - 54 =
23 * 16 =
45 / 6 =
( 44 + 11 ) * 10 =
( 44 + 11 ) / 10 =
```

$\log_{10}(10^2)=$

```
exp(1)=
log(exp(1)) =
```

b. Speichern Sie ein Teilergebnis in einem Objekt ab und verwenden Sie es nacheinander für weitere Berechnungen.

```
x <- c(1,2,3,4)
3 * x =
y <- x/2
y =
```

c. Führen Sie in R die folgenden Befehle nacheinander aus und kommentieren Sie die Befehle bzw. die Ausgabe von R.

```
x <- c(4,8,12)
x
x/2
x**2
x^2
sqrt(x)
x^(1/3)
x[2]
x[-1]
x-1
y<-c(-1,1,0.5)
z<-x*y
z
```

2. *Datenmanagement*

Die Datei `Blutdruck.txt` enthält folgende Variable:

```
ID    Patienten-ID
sys   systolischer Blutdruck (mm Hg)
dia   diastolischer Blutdruck (mm Hg)
```

a) Lesen Sie die Datei Blutdruck.txt mit R ein.

b) Fassen Sie die Daten mit geeigneten deskriptiven Kenngrößen zusammen.

c) Stellen Sie die Blutdruckdaten grafisch dar.

d) Enthalten die Daten verdächtige Werte, die möglicherweise falsch sind?

3. *Korrelation*

In einer Stichprobe von 15 Personen mit Psoriasis (Schuppenflechte) wurde die Epidermisdicke der Plaques und die Interleukinexpression relativ zu einer Referenz gemessen. Lesen Sie die Datei `Psorias.txt` mit folgenden Variablen ein.

```
ID    Patienten-ID
IL    Interleukin (relativ)
EPD   Epidermisdicke (mm)
```

Stellen Sie die Daten der Datei in einem Streudiagramm dar und berechnen Sie den Korrelationskoeffizient *r* nach Pearson.

4. *t-Test für verbundene Stichproben und Wilcoxon-Vorzeichen-Rangtest*

Lesen Sie die Datei Epidermisdicke.txt mit folgenden Variablen ein:

```
ID    Patienten-ID
pla   Epidermisdicke (mm) in läsionaler Haut
npla  Epidermisdicke (mm) in nicht-läsionaler Haut
```

a) Wie groß sind die Mittelwerte und die Standardabweichungen?

b) Wie groß ist der Mittelwert und die Standardabweichung der Änderung der Epidermisdicke?

c) Stellen Sie die Änderung in einem Boxplot dar.

d) Prüfen Sie mit einem gepaarten t-Test bzw. einem Wilcoxon-Vorzeichen-Rangtest, ob der Unterschied signifikant ist.

5. *Kontingenztafeln, Risikokennzahlen, Chi-Quadrat-Test*

In einer Studie wurden 166 Fälle von Lungenkrebs bei 100 000 Rauchern beobachtet, während nur 7 Lungenkrebsfälle in 100 000 Nichtrauchern auftraten.

a) Stellen Sie die Daten in einer Vierfeldertafel und mit Hilfe von R in einer Matrix dar.

b) Erzeugen Sie ein Balkendiagramm.

c) Wie groß ist das Risiko an Lungenkrebs zu erkranken bei Rauchern und bei Nichtrauchern.

d) Wie groß ist das relative Risiko und die Odds Ratio.

e) Führen Sie einen Chi-Quadrat-Unabhängigkeitstest durch, um den Zusammenhang der Merkmale Rauchverhalten und Erkrankung zu untersuchen.

6. *Einfache lineare Regression*

Lesen Sie wiederum die Datei `Psorias.txt` mit folgenden Variablen ein.

```
ID    Patienten-ID
IL    Interleukin (relativ)
EPD   Epidermisdicke (mm)
```

a) Stellen Sie ein einfaches lineares Regressionsmodell mit der Epidermisdicke EPD als abhängige Variable und der Interleukinexpression IL als unabhängige Variable auf.

$$EPD = b_0 + b_1 * IL$$

Verwenden Sie dazu die R-Funktion `lm( )`.

a) Wie groß ist der Achsenabschnitt b_0 und die Steigung b_1.

b) Stellen Sie die gemessenen Daten und die Regressionsgerade in einem gemeinsamen Diagramm dar.

c) Stellen Sie die Residuen (Differenzen zwischen gemessenen und berechneten Werten) in einem Diagramm dar.

d) Welche mittlere Epidermisdicke erwartet man bei IL = 0,1?

7. *Einfaktorielle Varianzanalyse*

Der Einfluss von Koffein auf motorische Fähigkeiten wurde in einer kleinen Gruppe von Freiwilligen untersucht. Lesen Sie die Datei Finger.txt mit folgenden Variablen ein:

```
taps  Fingerbewegungen
group Dosisgruppe (0, 1 oder 2)
```

a) Beschreiben Sie die Variablen deskriptiv.

b) Erzeugen Sie einen Boxplot.

c) Führen Sie eine einfaktorielle Varianzanalyse durch.

d) Prüfen Sie, ob die die Voraussetzung der Varianzanalyse erfüllt ist (Varianzhomogenität).

e) Unterscheiden sich die Mittelwerte der drei Gruppen signifikant ?

8. *Ereigniszeitanalyse*

In zwei Gruppen von je 17 Mäusen wurde in einem Zeitraum von 21 Tagen die Zeit bis zum Eintreten einer Paralyse nach einer Infektion untersucht. Es wurden die folgenden Ereigniszeiten (Tage) beobachtet:

Gruppe A
14, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*

Gruppe B
6, 7, 7, 8, 8, 8, 11, 11, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*, 21*

*zensierte Beobachtungen.

a) Beschreiben Sie die Überlebenszeitkurven deskriptiv.
b) Stellen Sie die Ereigniszeitdaten in einem Kaplan-Meier-Diagramm auf.
c) Testen Sie mithilfe des Log-Rang-Tests, ob sich die Überlebenszeitkurven unterscheiden.