

## Übung 1

1. Der Geysir Old Faithful im Yellowstone National Park ist eine der bekanntesten heißen Quellen. Für die Zuschauer und den Nationalparkdienst ist die Zeitspanne zwischen zwei Ausbrüchen und die Eruptionsdauer von grossem Interesse.

Im File `/afs/ethz.ch/users/s/sfs/datasets/geysir.dat` sind die Messungen vom 1.8.1978-8.8.1978 in 3 Kolonnen abgelegt: "Tag", "Zeitspanne" und "Eruptionsdauer".

- a) Zeichne Histogramme von der Zeitspanne zwischen zwei Ausbrüchen:

```
> geysir <- read.table("/afs/ethz.ch/users/s/sfs/datasets/geysir.dat", header=T)
# Datensatz einlesen

> hist(geysir[, "Zeitspanne"])
> hist(geysir[, "Zeitspanne"], nclass=10)
> hist(geysir[, "Zeitspanne"], breaks=seq(41,96,by=11))
# Histogramme zeichnen
```

Was fällt auf? Was ist der Unterschied zwischen diesen drei Histogrammen?

- b) Zeichne Histogramme (Anzahl Klassen variieren) von der Eruptionsdauer.

```
> hist(geysir[, "Eruptionsdauer"], ...)
```

Was fällt auf? Vergleiche mit der ersten Teilaufgabe.

Für weitere R-Anweisungen vergleiche die Einführungsunterlagen.

2. Im File `/afs/ethz.ch/users/s/sfs/datasets/lawinentote.dat` befinden sich die Anzahlen Lawinentoter in der Schweiz in den Wintern 1936/37 - 1998/99, getrennt in drei Kategorien: Lawinentote im freien Gelände, auf Verkehrswegen und in Gebäuden.

- a) Wir schauen uns zuerst einmal die Daten an. Dazu tragen wir die Anzahl Lawinentoter pro Kategorie und Jahr gegen die Jahreszahl auf.

R-Anleitung:

```
> daten <- read.table("/afs/ethz.ch/users/s/sfs/datasets/lawinentote.dat", header =
T) # Daten einlesen
> winter <- daten[,1] # Jahreszahl
> ltote <- as.matrix(daten[,2:4]) # Lawinentote pro Jahr und Kategorie
> matplot(winter, ltote, pch = 1:3, col=1:3) # jede Kategorie mit eigenem Symbol und
eigener Farbe zeichnen
> legend(1970, 70, colnames(ltote), pch = 1:3, col=1:3)
```

Wie unterscheiden sich die drei Kategorien?

- b) Wir wollen nun noch schauen, ob sich die Gesamtanzahl Lawinentoter über die Zeit verändert hat.

R-Anleitung:

```
> ltote.total <- ltote[,1] + ltote[,2] + ltote[,3] # total Tote pro Jahr
> plot(winter, ltote.total, type = 'l') # als Linie gezeichnet
```

Nimmt die Zahl der Toten eher zu oder ab? Kannst Du aus der Grafik eine Entwicklung über die letzten siebzig Jahre herauslesen?

- c) Da es in b) fast unmöglich ist eine vernünftige Aussage über die Entwicklung der Anzahl Lawinentoter herauszulesen, legen wir zuerst eine glatte Kurve durch unsere Beobachtungen und interpretieren dann diese. Wie sieht es nun mit der Entwicklung aus?

R-Anleitung:

```
> plot(winter, ltote.total)
> ltote.total.glatt <- lowess(winter, ltote.total, f=1/6) # Anzahlen geglättet
> lines(ltote.total.glatt)
```

3. In untenstehender Tabelle sind die Farbintensitäten von Äpfeln bei verschiedenen Lagerungstemperaturen (in °C) und Lagerungsdauern (in Tagen) aufgeführt.

Lagerungsdauer	Lagerungstemperatur			
	1	10	18	27
0	40.98	41.14	41.08	40.56
7	40.84	41.17	42.02	40.29
14	41.14	42.60	43.82	40.56
21	40.70	44.41	45.22	41.11

- a) Welcher Intensitätswert wird von ziemlich genau 25% aller Beobachtungen unterschritten? Wieviele Prozent der Beobachtungen weisen eine Intensität über 43 auf? Beantworte beide Fragen mittels einer geeigneten graphischen Darstellung.
- b) Erstelle ein Stamm- und Blattdiagramm der Farbintensitäten.
- c) Zeichne die Farbintensität gegen die Lagerungsdauer für die vier verschiedenen Temperaturen auf. Wähle für jede Temperaturstufe einen anderen Linientyp. Vergleiche, wie sich bei den verschiedenen Temperaturstufen die Farbintensität mit der Lagerungsdauer verändert. Gibt es Unterschiede?

R-Anleitung: Im File `/afs/ethz.ch/users/s/sfs/datasets/lightness.apples.dat` sind die Farbintensitäten aus obiger Tabelle in den Spalten "temp", "time", und "intensity" gespeichert (plus einige Spalten, die wir nicht brauchen).

```
a) > apfelfarben <- read.table("/afs/ethz.ch/users/s/sfs/datasets/lightness.apples.dat",
  header = T)[,c(1,2,7)]           # nur die erste, zweite und siebte Kolonne des Datensatzes
> apfelfarben                       # Datensatz anschauen
> hist(apfelfarben["intensity"],nclass=30,xlab="intensity")
                                                                    # Histogramm zeichnen

b) > stem(apfelfarben["intensity"])                                     # Stamm-&Blattdiagramm

c) > matplot(c(0,7,14,21),matrix(apfelfarben["intensity"],nrow=4,byrow=T),
  type="l")                         # zeichnet die Spalten der obigen Matrix gegen die vier Lagerungsdauern
> legend(0, 45,c("temp=1","temp=10","temp=18","temp=27"), lty=1:4, col=1:4)
                                                                    # erzeugt Legende
```

4. Berechne für die untenstehende Stichprobe den Mittelwert, die empirische Varianz, die empirische Standardabweichung, den Median, die Quartile, sowie die Medianabweichung (MAD).

4, 5, 2, 16, 8, 3, 7, 2, 4, 9, 6

Wieviele Standardabweichungen ist der Ausreisser vom Mittelwert entfernt?

Wieviele Medianabweichungen ist der Ausreisser vom Median entfernt?

Wieviele Daten liegen innerhalb  $\text{Median} \pm \frac{3}{2}\text{MAD}$ ?

**Abgabe:** Montag, 3. April, in der Pause der Vorlesungsstunde