Übung 02 - Datenanalyse und Statistik WS 2008/2009

```
# Starten Sie JGR!

# Tutorial: Skalenniveaus in R

#- Zahlen und Zeichen
1:4               # Zahlenfolge
"Hallo"           # Zeichenfolge
a <- c("W","E")   # Alle elementaren Datentypen können zusammengefasst werden
a
sort(a)           # Der Umgang mit den Vektoren hängt nicht vom Datentyp ab


#- Datenvektoren kategorieller Variablen

a <- factor( c("W","W","E") ) # Eine dichotome Variable
a
levels(a)                     # Zeigt die auftretenden Kategorien

tapply(1:3,a,sum)   # Anwendung einer Funktion auf Teildatensätze der ersten
tapply(1:3,a,mean)  # Variable, entsprechend den Kategorien der weiten ergeben

factor( c(1,1,2) )  # Reine Zahlen sind als Gruppenbezeichnungen möglich


#- Klassen
# In welcher Form ein Datensatz gespeichert ist, bzw. von welchem Typ er ist,
# lässt sich mit der Funktion class() anzeigen. (Dies ist zwar nicht ganz
# korrekt, erfüllt hier aber seinen Zweck)
class( c("W","E") )               # Zeichen
class( 1:4 )                      # Zahlen
class( factor( c("W","E") ) )     # Kategorien


#- Datenmatrix
# In einer Datenmatrix können Variablen mit verschiedenen Skalen
# aufgenommen werden.
x <- data.frame(Name=c("Bert","Ernie","Samson"),Alter=12:10,Gruppe=factor(c(1,1,2)))
x
class(x)            # data.frame := Datenmatrix

class(x[,1])        # Obacht: Zeichen werden in Datenmatrizen (data.frame)
                    # automatisch als Kategorien interpretiert !

# Der Zugriff kann über Namen oder Indices erfolgen
x$Alter             # Spalten
x[,2]
x[,"Alter"]
x[c(1,3),]          # Zeilen
x[2,3]              # einzelnes Element
```

```
# Erweiterungen können mittels rbind(), cbind() oder dem Zuweisungs-
# operator <- bzw. =  erfolgen, über den - wie gehabt - auch
# Ersetzungen erfolgen können:
cbind(x, Platz = 1:3)       # Hinzufügen einer Variablen
rbind(x, c("Bert",18,1) )   # Hinzufügen eines Faktors (Zeile)
rbind(x, c("Oscar",10,3) )  # Beachte: Es dürfen dabei keine neuen
   # Kategorien zu einer kategoriellen Variable hinzugefügt werden
rbind(x, data.frame(Name="Oscar",Alter=10,Gruppe=factor(3)) )

a <- x
a[c(1,3),1] <- c("Ernie","Ernie")
a
```

Aufgabe 1: Aufgabe: Die an einem Tag in einer bestimmten Miene in drei Schichten erbrachten Fördermengen an Cu und Ag seien mit 10, 13 und 9 Tonnen bzw. 1, 7 und 4 Kilogramm notiert. Erstellen Sie zu diesen Daten eine geeignete Datenmatrix (mit zwei Variablen kategorieller Skala) und berechnen Sie daraus die jeweils geförderten Tagesmengen der beiden Erze! Bügeln Sie den Fehler eines Kollegen aus, der in der zweiten Schicht ein Kilogramm Silber übersah, fügen Sie eine vierte Schicht mit beliebigen Fördermengen ein und löschen Sie dafür die erste Schicht (nicht notwendigerweise alles in einem Schritt)!

Aufgabe 2: Ein Versuch mit einem Schlafmittel
In einem klinischen Versuch sollte die Wirksamkeit eines Schlafmittels getestet werden. Dazu wurden von den Patienten, die in einer Klinik mit Schlafstörungen behandelt werden, zufällig 10 Patienten herausgesucht, die ein neuartigs Schlafmittel erhielten (Behandlungsgruppe) und 10 Patienten die eine wie Schlafmittel aussehende und schmeckende wirkungslose Pille erhielten (Kontrollgruppe). Um die Wirksamkeit zu bestimmen wurden jeweils zunächst die Schlafdauer in der Nacht vor der ersten Einnahme bestimmt. Vor der zweiten Nacht wurde den Patienten das Schlafmittel oder eine nach Schlafmittel aussehende wirkungslose Pille gegeben. Die Schlafdauer der zweiten Nacht wurde ebenfalls ermittelt.

Als Daten haben wir die Information, welches Mittel der Patient erhalten hat (2 = Schlafmittel, 1 = Wirkungsloses Placebo), und die Information, wieviel länger die Patienten in der zweiten Nacht geschlaften haben (in Stunden). Ziel der Untersuchung ist es nachzuweisen, dass das Schlafmittel bei den in der Klinik behandelten Schlafstörungen wirkt, aber so weit kommen wir heute noch nicht.

Laden Sie den Datensatz sleep!

```
data(sleep)
```

(1) Wie liegen die Daten vor?
Tip: z.B. als Datenmatrix, Datentafel, unvorbereitet,...

---

(2) Welche Variablen gibt es und was bedeuten Sie?
(reden Sie mit Ihrem Nachbarn darüber)

(3) **Welche Skala haben die einzelnen Variablen?**
(Liegen diese auch so in R vor ?)

Tip:z.B. ordinal

(4) **Stichprobe und/oder Zufallsexperiment?**

---

(5) **Wofür könnten diese Daten repräsentativ sein?**

Tip: Was könnte die Grundgesamtheit oder das Zufallsexperiment sein?

(6) **Welche Annahme müssen wir über die Datenerhebung treffen, damit die Daten dafür repräsentativ sind ?**

Tip: Naja, nach Vorlesung...

(7) **Sind die Daten laut Beschreibung für diese Grundgesamtheit / dieses Zufallsexperiment repräsentativ? Warum ?**

---

(8) Worüber kann man mit diesem Versuchsaufbau statistische Aussagen treffen?

- **O** Über die Qualität der medizinischen Versorgung in dem Krankenhaus?
- **O** Über die Wirksamkeit des Medikaments bei den Patienten in der Schlafstation?
- **O** Über die Wirkung des Medikaments auf alle Menschen?
- **O** Über den Schlafmittelverbrauch auf der Schlafstation.
- **O** Über gar nichts, da die Daten nicht repräsentiv sind.
- **O** Über gar nichts, da die Daten repräsentiv sind.
- **O** Über gar nichts, da die Daten zufällig sind.

Nachfolgend finden Sie eine Reihe von Befehlen zur Erzeugung von statistischen Graphiken. Verwenden Sie diese, um die weiteren Fragen zu beantworten!

```
# Punktdiagramm
stripchart(sleep$extra)
stripchart(sleep$extra,method="stack")     # gestapelt
stripchart(sleep$extra,method="jitter")    # verzittert
stripchart(sleep$extra ~ sleep$group)      # parallele
stripchart(sleep$extra ~ sleep$group, main="rel. Schlafdauer", col=2:3)
```

(9) Existieren in dem Datensatz Bindungen? Wenn ja, wieviele? Wieviele gibt es in den einzelnen Gruppen?

```
# Boxplot
boxplot( sleep$extra )
boxplot( c(sleep$extra,20) )               # mit Ausrei"ser
boxplot(sleep[,1] ~ sleep[,2])             # parallele
boxplot(extra ~ group, data = sleep)       # parallele
boxplot(extra ~ group, data = sleep, xlab="Gruppe",ylab="Schlafdauer")
boxplot(sleep$extra ~ sleep$group, notch=TRUE) # gekerbte
```

(10) Gibt es Ausreißer? Was können Sie aus einem Boxplot über die Verteilung der Variablen folgern?

```
# Balkendiagramm
barplot(table(sleep$group))                # table listet die Anzahlen auf

# Histogramm
hist(sleep$extra)
hist(sleep$extra,breaks=3)                 # Anzahl an Klassengrenzen
hist(sleep$extra,breaks=seq(-10,10,by=1))  # Definierte Klassengrenzen
hist(sleep$extra,breaks=seq(-10,10,by=1),density=30)
hist(sleep$extra,breaks=seq(-10,10,by=1),density=30,ylim=c(0,7))
```

(11) Gibt es auffällige Häufungen in der Schlafdauer?

Tip: Was ist das Skalenniveau? Was ist die Frage? Was ist also die Graphik?

(12) Beschreiben Sie die Verteilungsform der Extra-Schlafdauer !

Tip: Was ist das Skalenniveau? Was ist die Frage? Was ist also die Graphik?

```
# Quantils-Quantils-Plot gegen die Normalverteilung
qqnorm(sleep$extra[sleep$group==1])        # erste Gruppe
qqnorm(sleep$extra[sleep$group==2])        # zweite Gruppe
```

(13) Für die in der Orginalpublikation verwendeten Verfahren benötigt man Normalverteilung der Daten in der Behandlungsgruppe. Sind sie es? Woraus schließen Sie das?

---

(14) Aus welcher Graphik können Sie ersehen, ob die Verteilung der Schlafdauer bimodal ist ?

---

(15) Welche statistische Graphik würde sich eignen, um die Behandlungserfolge mit und ohne Medikament zu vergleichen und eventuelle Ausreißer zu erkennen? Können Sie aus dieser Graphik begründet auf unterschliedliche Behandlungserfolge schließen?

Tip: Derselbe wie auch schon weiter oben!

(16) Ist etwas an der Graphik ungewöhnlich? Wenn ja was?

Tip:Es muß nichts ungewöhnlich sein. Sie sollten sich diese Frage einfach immer stellen.

(17) Entsprechend die Beobachtungen dem, was man inhaltlich erwarten würde?

Tip: Also Schlafmittel, ...

(18) Beschreiben Sie eine Situation in der jemand in Ihrem angestrebten Beruf einen ähnlichen Datensatz erheben könnte.

Tip:Den gibt es bestimmt. Vielleicht sähe der Datensatz aber etwas anders aus. Seien Sie kreativ.

Auch gut zu kennen:

```
# Mehrfachanzeigen
# mfrow = Multi Frame by Row
par(mfrow=c(2,2)) # 4 Graphiken in einem Fenster: c(horizontal, vertikal)
stripchart(sleep$extra)
boxplot(extra ~ group, data = sleep)
hist(sleep$extra,breaks=seq(-10,10,by=1),density=30)
barplot(table(sleep$group))
```