Übung 1

Das Ziel dieser Übung ist es, dass Sie verschiedene Funktionen des Statistikpaketes S-Plus ausprobieren und sich weiter mit der Computer-Umgebung der Sun Workstations vertraut machen.

Die angegebenen S-Plus Befehle gelten grundsätzlich auch für R, Abweichungen werden in Klammern und *Kursivschrift* angegeben.

Starten Sie Splus, d.h. öffnen Sie mit `e` einen Emacs und tippen Sie dann **M-x S**.
(*Falls sie mit R arbeiten möchten, starten sie R mit* `M-x R`).

Bei Problemen mit dem login oder unix/emacs, konsultieren sie zuerst das Tutorial. Falls das nicht weiterhilft, scheuen sie sich nicht, den Assistenten zu kontaktieren oder kommen Sie am Donnerstag in die Präsenzstunde.

>**library(MASS)**	macht in S-Plus Funktionen und Datensätze der "MASS"–library verfügbar. (*In R werden mit diesem Befehl nur die Funktionen zur Verfügung gestellt. Um den Datensatz survey zur Verfügung zu haben, muss zusätzlich der Befehl* `data(survey)` *getippt werden.*)
>**attach(survey)**	setzt die Datei "survey" auf Position 2 in der Liste der verfügbaren Directories und Data frames.
>**search()**	zeigt diese Liste
>**ls(pos=2)**	listet alle Objekte im Data frame "survey"

Die Datei "survey" in der library "mass" enthält Antworten von und Messungen an 237 StatistikstudentInnen an der Universität Adelaide. Es sind folgende Variablen vorhanden:

Sex	Geschlecht
Wr.Hnd	Spannweite in cm (von Daumen bis kleinem Finger) der Schreibhand
NW.Hnd	Spannweite in cm (von Daumen bis kleinem Finger) der Nicht-Schreibhand
W.Hnd	Schreibhand
Fold	Verschränken Sie die Arme. Welcher ist oben?
Pulse	Anzahl Schläge pro Minute
Clap	Klatschen Sie in die Hände. Welche ist oben?
Exer	Wie oft treiben Sie Sport?
Smoke	Wieviel rauchen Sie?
Height	Körpergrösse umgerechnet in cm
M.I	Bevorzugen Sie metrische (cm/m) oder imperiale (feet/inches) Längenmasse?
Age	Alter in Jahren

Nützliche Funktionen für einen ersten Überblick:
summary(survey), **table(Sex)**, **table(Sex,Smoke)**

Zum Umgang mit fehlenden Werten (NA):

> **mean (Pulse)**	Resultat ist NA
> **mean(Pulse,na.rm=T)**	fehlende Werte werden bei der Berechnung weggelassen
> **var(na.omit(Pulse))**	die Funktion "var" verfügt nicht über das Argument na.rm. na.omit(Pulse) streicht alle NA's weg. (*In R verfügt die Funktion var wie die Funktion mean über das Argument na.rm. Der korrekte Funktionsaufruf heisst deshalb in R* `var(Pulse,na.rm=T)`. *Der Befehl var(na.omit(Pulse)) dagegen ergibt in R eine Fehlermeldung.*)

Nützliche Funktionen für Graphiken:

> **motif()**	öffnet ein Graphik-Fenster in Splus
(> `x11()`	*öffnet ein Graphik-Fenster in R*)
> **hist(Height)**	Histogramm
> **boxplot(Height)**	Boxplot
> **boxplot(split(Height,Sex))**	Boxplots
> **boxplot(Height[Sex=="Female"],Height[Sex=="Male"])**	Boxplots
> **plot(Wr.Hnd,NW.Hnd)**	Scatterplot
> **plot(Sex,Height)**	?

> **detach("survey")** löst die Verbindung zum Data frame in S-Plus.
(*in R geht dies ohne Anführungszeichen mit* `detach(survey)`.)

Nun können die Spalten nicht mehr mit dem Namen, sondern müssen via $ oder [.] angesprochen werden:
> **plot(survey$Wr.Hnd,survey$NW.Hnd)** oder **plot(survey[,2],survey[,3])**.

Auswahl von Beobachtungen, z. B. nur die ersten 50 Beobachtungen:
> **plot(survey[1:50,2],survey[1:50,3])**

Nicht zu vergessen die Online-Hilfe:
> **help(survey)**
> **help(plot)**
...

Analysieren Sie nun die Daten in "survey" mit deskriptiven Methoden. Erstellen Sie dazu Tabellen und Kreuztabellen für die kategoriellen Variablen und berechnen Sie Lage- und Streuungsmasse für die stetigen Variablen. Machen Sie geeignete graphische Darstellungen. Wie sehen die Verteilungen aus? Gibt es Ausreisser?

Beantworten Sie die folgenden Fragen:

a) Ist die Spannweite der Schreibhand im Allgemeinen grösser als die Spannweite der Nicht-Schreibhand?

b) Rauchen die beiden ältesten Studierenden?

c) Welche Faktoren könnten einen Einfluss auf den Puls haben?