


**STATISTIK 2 (107.325) WS 2006**  
**COMPUTERSTATISTIK (107.258) WS 2006**

**Übung 1****9. Oktober 2006****Dutter**

Analysieren Sie die Daten “Werner Chemistry Data“ mit dem Computer Program System . Das Originalfile *b.2d.inp* (siehe <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>) enthält als Kopfzeilen die Steuerbefehle für das Programm BMDP, was nicht weiter wesentlich ist, aber die Bezeichnung der Variablen sind enthalten. Das File *werner\_bcd.txt* enthält nur die nackten Zahlen (allerdings mit fehlenden Werten kodiert durch „\*“).

Das Importieren der Daten könnte mit den R-Befehlen innerhalb einer Funktion so ausschauen:

```
"f.ueb1" <-
function ()
{
# werner_bcd<-read.table("werner_bcd.txt")
  werner_bcd<-matrix(scan("werner_bcd.txt",na.strings="*"),ncol=9,byrow=T)
  dimnames(werner_bcd)<-list(werner_bcd[,1],
    c("ID","AGE","HEIGHT","WEIGHT","BRTHPILL","CHOLSTRL","ALBUMIN",
      "CALCIUM","URICACID"))
}
```

Dabei werden schon Namen der Variablen (Spalten) und Zeilen (Patientennummern) vergeben.

1. Wählen Sie eine Variable aus den “Blutkenngößen“ und zeichnen Sie ein Histogramm mit einem “eindimensionalen Streuungsdiagramm“.
2. Berechnen Sie Schätzungen für Lokation (Ort), Streuung, Form (Schiefe und Kurtosis) und beschreiben Sie verbal die Datenverteilung.
3. Berechnen Sie den Korrelationskoeffizient mit den anderen Variablen. Wählen Sie eine (sinnvolle) zweite Variable, zeichnen Sie ein Streuungsdiagramm (xyplot), und schätzen Sie eine Regressionsgerade. Kommentar!
4. Stellen Sie eine Stamm- und Blattdarstellung „vernünftig“ dar und diskutieren Sie darüber.
5. Das Gleiche mit einer Boxplot-Darstellung, über möglichst alle Variablen.

Bitte senden Sie die Ausarbeitung in Form eines pdf-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Listing des Programmkodes (Funktion) an

R.Dutter@tuwien.ac.at

bis 11. bzw 18. Oktober 2006.

Bitte den Namen des pdf-File folgendermaßen:

**name\_exer\_1.pdf**

wobei 'name' für den Familiennamen steht.

Empfehlenswert ist es, den Bericht mit „Sweave“ zu erstellen. (Vorlage siehe <http://www.statistik.tuwien.ac.at/public/dutt/vorles/> .)