

STATISTIK 2 (107.325) WS 2011
COMPUTERSTATISTIK (107.258) WS 2011

Übung 6

30. November 2011
Dutter

21. Analysieren Sie die Daten “Werner Chemistry Data“ mit dem Computer Program System \mathbb{R} , wobei gleich zu Beginn die beiden Ausreisser in der Variablen „CHOLSTRL“ ausgeschlossen werden sollten (d.h. man nehme nur jene Zeilen der Matrix, bei denen CHOLSTRL > 150 und < 400 ist).

Betrachten Sie 4 Altersgruppen der Patientendaten “Werner Blood Chemistry Data“ wie im Übungsbeispiel 19: (18,25], (25,32], (32,42], (42,55]. (Hinweis: Die Funktion ‘cut’ ist recht nützlich zum Zerschneiden des Vektors ‘AGE’. Die Anzahl pro Gruppe sollte dann sein: 50, 48, 42, 46.)

Untersuchen Sie nun die Abhängigkeit von den Altergruppen und gleichzeitig von der Variablen BRTHPILL. Stellen Sie möglich Interaktionen zwischen den beiden Variablen grafisch dar (“interaction.plot”).

22. Untersuchen Sie die Cholesterindaten abhängig von Altersgruppen und BRTHPILL wie vorhin. Nun aber mit ev. Wechselwirkungen. Diskutieren Sie das Ergebnis.
23. Betrachten Sie die Daten des Vienna City Marathons in den letzten 10 Jahren 2002 bis 2011 mit dem Computer Program System \mathbb{R} .

Betrachten Sie auch die letzten (wenn es mehrere gibt) Zwischenzeiten. Dazu sollten Sie die Funktion `f.mara_read` geringfügig verändern, womit Sie auch Zwischenzeiten, wenn vorhanden, verwenden können:

```
> f.mara_read <- function(dat) {
+   chdat <- paste("http://www.statistik.tuwien.ac.at/public/dutt/vorles/data/",
+   dat, sep = "")
+   mara <- read.table(chdat, sep = ":")
+   if (ncol(mara) == 10)
+     colnames(mara) <- c("Kl-Sex", "1.Zw.Zeit_h", "1.Zw.Zeit_m",
+     "1.Zw.Zeit_s", "2.Zw.Zeit_h", "2.Zw.Zeit_m", "2.Zw.Zeit_s",
+     "Endzeit_h", "Endzeit_m", "Endzeit_s")
+   else if (ncol(mara) == 7)
+     colnames(mara) <- c("Kl-Sex", "Zw.Zeit_h", "Zw.Zeit_m",
+     "Zw.Zeit_s", "Endzeit_h", "Endzeit_m", "Endzeit_s")
+   else colnames(mara) <- c("Kl-Sex", "Endzeit_h", "Endzeit_m",
+   "Endzeit_s")
+   mara[, 1] <- as.character(mara[, 1])
+   if (ncol(mara) == 10) {
+     dmara <- mara[, 8] + mara[, 9]/60 + mara[, 10]/3600
+     zmara <- mara[, 5] + mara[, 6]/60 + mara[, 7]/3600
+   }
+   else if (ncol(mara) == 7) {
```

```

+       dmara <- mara[, 5] + mara[, 6]/60 + mara[, 7]/3600
+       zmara <- mara[, 2] + mara[, 3]/60 + mara[, 4]/3600
+     }
+     else {
+       dmara <- mara[, 2] + mara[, 3]/60 + mara[, 4]/3600
+       zmara <- rep(NA, length(dmara))
+     }
+     kmara <- mara[, 1]
+     datasex <- substr(kmara, 1, 1)
+     datakl <- substr(kmara, 3, 5)
+     return(list(mara = mara, zmara = zmara, dmara = dmara, kmara = kmara,
+               datasex = datasex, datakl = datakl))
+ }

```

In einer for-Schleife können die Laufzeiten pro Jahr (mit den entsprechenden Kenngrößen) eingelesen und die resultierenden Listenelemente aus `f.mara_read` zu einem `data.frame` mit 64.518 Zeilen zusammengefügt werden.

Stellen Sie nun 10 Boxplots der 10 Datensätze nebeneinander dar. Diskutieren Sie das Bild.

24. Untersuchen Sie nun die Endzeiten in Abhängigkeiten der verschiedenen Variablen.

- Betrachten Sie zunächst nur Altersklassen (ohne Unterscheidung der Geschlechter) durchgehend über die Jahre 2002 bis 2011. Verwenden Sie dabei wie früher nur die Klassen `c("H", "30", "35", "40", "45", "50", "55", "60", "65", "70")`
- Betrachten Sie nun ein komplizierteres Modell mit der (linearen) Abhängigkeit der durchschnittlichen Endzeiten von Klasse und Jahr.
- Betrachten Sie zusätzlich die Abhängigkeit vom Geschlecht.
- Verwenden Sie in den vorigen beiden Modellen zusätzlich noch mögliche Wechselwirkungen.
- Diskutieren Sie die verschiedenen Ergebnisse.

Bitte, stellen Sie die Ausarbeitung in Form eines pdf-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Kurz-Listing des Programmcodes (Funktion) in die TUWEL-Seite

<https://tuwel.tuwien.ac.at/course/view.php?idnumber=107258-2011W>
bis zum 5. Dezember 2011.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_6.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/>.)

Kreuzen Sie außerdem bitte im Internet jede Nummer des Übungsbeispiels an, das Sie dann in der Übungsstunde an der Tafel (mit Beamer-Unterstützung) vorrechnen wollen und können. Der Termin ist üblicherweise 2 Stunden vor der Übung, d.h. Mi., 12:00. Siehe <http://www.statistik.tuwien.ac.at/cgi-bin/uebb/bspstart.cgi> .