

**STATISTIK 2 (107.325) WS 2010**  
**COMPUTERSTATISTIK (107.258) WS 2010**

**Übung 5**

**9. November 2010**  
**Dutter**

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

17. Betrachten Sie 4 Altersgruppen der Patientendaten "Werner Blood Chemistry Data": (18,25], (25,32], (32,42], (42,55]. Die Anzahl pro Gruppe sollte sein: 50, 48, 42, 46.

Testen Sie auf Einfluss des Alters (der Altersgruppen) auf Cholesterinwerte.

Erklären Sie die Voraussetzungen zu diesem Test. Kommentieren Sie das Resultat.

18. Verwenden Sie von den Altersgruppen der Patientendaten nur die ersten 3 bzw. die mittleren 2. Wie ändern sich die Signifikanzen? Diskutieren Sie Ihr Resultat.

19. Analysieren Sie die Daten des Vienna City Marathons 2010 mit dem Computer Program System  $\mathbb{R}$ . Sie stehen (anonymisiert) auf <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>

**Bemerkung:** Nachdem die Marathon-Daten (hier 1. und 2. Zwischenzeit sowie Endzeit) alle im Format 'Stunden:Minuten:Sekunden' gespeichert sind, ist es vorteilhaft, folgende Lese- und Umwandlungsbefehle (beispielhaft für das Jahr 2010, d.h. Datensatz 'mara\_t\_10', wobei hier nur die 1. Zwischenzeiten und Endzeiten gespeichert werden) zu verwenden:

```
> f.mara_read <- function() {
+   mara10 <-< read.table("mara_t_10", sep = ":", header = TRUE)
+   mara10[, 1] <-< as.character(mara10[, 1])
+   dmar10 <-< mara10[, 8] + mara10[, 9]/60 + mara10[, 10]/3600
+   dmar10 <-< cbind(mara10[, 2] + mara10[, 3]/60 + mara10[,
+     4]/3600, dmar10)
+   kmar10 <-< mara10[, 1]
+   dimnames(dmar10) <-< list(kmar10, c("1. Zwischenz.", "Endzeit"))
+   datasex10 <-< substr(kmar10, 1, 1)
+   datacl10 <-< substr(kmar10, 3, 5)
+ }
```

(wobei hier auch angegeben ist, wie man die Kennzeichnung 'Geschlecht' und 'Altersklasse' herauslesen kann).

- (a) Man zeichne eine Dichtekurve der 'Endzeiten'. Weiters plotte man darüber in rot die 'Endzeiten' der Männer und in grün die der Frauen.

Bitte um Kommentare.

(b) Man führe einen Test auf Gleichheit der Varianzen der Laufzeiten von Männern und der von Frauen durch.

Sind Männer im Schnitt etwa um eine halbe Stunde oder um .4 oder .3 Stunden schneller als Frauen in einem Laufmarathon?

Bitte um Kommentare.

20. Man betrachte nur die Altersklassen H, 30, 35, 40, 45, 50, 55, 60, 65 und 70. Haben diese Klassen einen signifikanten Einfluss auf die Laufzeiten der Männer? Das Gleiche bei den Frauen?

Gibt es einen signifikanten Unterschied der Gesamt-Laufzeitmittel der Hauptklasse „H“-Männer und der Klasse „50“ der Männer, d.h. Alter zwischen 50 und 55?

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 Programmkodes (Funktion) in die TUWEL-Seite

<https://tuwel.tuwien.ac.at/course/view.php?id=2604>

bis incl. 15. November 2010.

Bitte den Namen des pdf-File folgendermaßen:

**name\_exer\_5.pdf**

wobei 'name' für den Familiennamen steht. Bitte im pdf-File Name, Datum und Seitennumerierung nicht vergessen!

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