


STATISTIK 2 (107.325) WS 2008
COMPUTERSTATISTIK (107.258) WS 2008

Übung 5

22. November 2008
Dutter

17. Analysieren Sie die Daten des Vienna City Marathons 2008 mit dem Computer Program System . Sie stehen (anonymisiert) auf <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>

Bemerkung: Nachdem die Marathon-Daten (Laufzeiten) alle im Format ‘Stunden:Minuten:Sekunden’ gespeichert sind, ist es vorteilhaft, folgende Lese- und Umwandlungsbefehle (beispielhaft für das Jahr 2008, d.h. Datensatz ‘mara_t_08’, wobei hier die 2. Zwischen- und Endzeiten gespeichert sind) zu verwenden:

```
f.mara_read <- function() {
# Einlesen von 10 Spalten
  mara08 <- read.table("mara_t_08",sep=":",header=T)
  mara08[,1] <- as.character(mara08[,1])
# Errechnung der Marathon-Endzeiten in Stunden
  dmar08 <- mara08[,8]+mara08[,9]/60+mara08[,10]/3600
# Errechnung der 1. Marathon-Zwischenzeiten in Stunden und Verknuepfung
# mit den Endzeiten
  dmar08 <- cbind(mara08[,2]+mara08[,3]/60+mara08[,4]/3600,dmar08)
# Geschlecht- und Klassenkennzeichnung
  kmara08 <- mara08[,1]
# Zuweisung von Zeilen- und Spaltennamen der Zahlenmatrix
  dimnames(dmar08) <- list(kmara08,c("1. Zwischenz.", "Endzeit"))
# Herauslesung von ‘Geschlecht’
  datalsex08 <- substr(kmara08,1,1)
# Herauslesung von ‘Altersklasse’
  datakl08 <- substr(kmara08,3,5)
}
```

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

Man zeichne eine Dichtekurve der ‘Endzeiten’. (1) plote man darüber in rot die ‘Endzeiten’ der Männer und (2) in grün die der Frauen.

Bitte um Kommentare.

18. Man führe einen Test auf Gleichheit der Varianzen der Laufzeiten von Männern und der von Frauen durch.

Bitte um Kommentare.

19. Simulieren Sie einen 2-dimensionalen Datensatz (mit 2 normalverteilten Zufallszahlen), z.B.

```
require(MASS)
set.seed(seed=1)
x1 <- mvrnorm(100,mu=c(0,0),Sigma=matrix(c(1,.9,.9,1),ncol=2))
x2 <- mvrnorm(100,mu=c(.1,-.1),Sigma=matrix(c(1,.9,.9,1),ncol=2))
```

(Variieren Sie eventuell den Mittelwertsvektor von `x2`.) Plotten Sie die beiden Datensätze in der gleichen Zeichnung mit verschiedenen Farben.

Kann man Unterschiede entdecken? Kommentare.

20. Führen Sie 2 univariate t-Tests auf Verschiedenheit der Mittelwerte durch und zusätzlich eine Hotelling's- T^2 -Test.

Kommentare!

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 zum Dienstag-Abend, 25. November 2008.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_5.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/> .)