


STATISTIK 2 (107.325) WS 2009
COMPUTERSTATISTIK (107.258) WS 2009

Übung 5

23. November 2009
Dutter

17. Analysieren Sie die Daten des Vienna City Marathons 2009 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 2009, d.h. Datensatz ‘mara_t_09’, wobei hier die 2. Zwischen- und Endzeiten gespeichert sind) zu verwenden:

```
> f.mara_read <- function() {
+   mara09 <- read.table("mara_t_09", sep = ":", header = T)
+   mara09[, 1] <- as.character(mara09[, 1])
+   dmar09 <- mara09[, 8] + mara09[, 9]/60 + mara09[, 10]/3600
+   dmar09 <- cbind(mara09[, 2] + mara09[, 3]/60 + mara09[,
+     4]/3600, dmar09)
+   kmar09 <- mara09[, 1]
+   dimnames(dmar09) <- list(kmar09, c("1. Zwischenz.", "Endzeit"))
+   datasex09 <- substr(kmar09, 1, 1)
+   datakl09 <- substr(kmar09, 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.

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

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-T²-Test (Funktion steht auf der LVA-Homepage).
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 Samstag-Abend, 28. November 2009.

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/> .)