


STATISTIK 2 (107.325) WS 2014
COMPUTERSTATISTIK (107.258) WS 2014

Übung 3

28. Oktober 2014

Dutter

9. Analysieren Sie die Daten des Vienna City Marathons 2014 mit dem Computer Program System . Sie stehen (anonymisiert) auf <http://www.statistik.tuwien.ac.at/public/dutt/vorles/> (ganz unten!).

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 2014, d.h. Datensatz ‘mara_t_14’, wobei hier die 2. Zwischen- und Endzeiten gespeichert sind) zu verwenden:

```
> f.mara_read <- function() {
+ # Einlesen von 10 Spalten
+   mara14 <-< read.table(
+     "http://www.statistik.tuwien.ac.at/public/dutt/vorles/data/mara_t_14",
+     sep=":",header=T)
+   mara14[,1] <-< as.character(mara14[,1])
+ # Errechnung der Marathon-Endzeiten in Stunden
+   dmar14 <-< mara14[,8]+mara14[,9]/60+mara14[,10]/3600
+ # Errechnung der 1. Marathon-Zwischenzeiten in Stunden und Verknuepfung
+ # mit den Endzeiten
+   dmar14 <-< cbind(mara14[,2]+mara14[,3]/60+mara14[,4]/3600,dmar14)
+ # Geschlecht- und Klassenkennzeichnung
+   kmara14 <-< mara14[,1]
+ # Zuweisung von Zeilen- und Spaltennamen der Zahlenmatrix
+   dimnames(dmar14) <-< list(kmara14,c("1. Zwischenz.", "Endzeit"))
+ # Herauslesung von ‘Geschlecht’
+   datasex14 <-< substr(kmara14,1,1)
+ # Herauslesung von ‘Altersklasse’
+   datakl14 <-< substr(kmara14,3,5)
+ }
```

(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.

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

Testen Sie, ob Männer im Schnitt 25 Minuten schneller laufen als Frauen.

Bitte um Kommentare.

11. Simulieren Sie einen 2-dimensionalen Datensatz (aus 2 normalverteilten Zufallsgrößen), z.B. durch

```
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 x_2 .) Plotten Sie die beiden Datensätze in der gleichen Zeichnung mit verschiedenen Farben.

Kann man visuell Unterschiede entdecken? Kommentare!

12. Führen Sie univariate t-Tests auf Verschiedenheit der Mittelwerte durch und zusätzlich einen Hotelling's-T²-Test (Funktion steht z.B. auf der LVA-Homepage).

Kommentare!

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?idnumber=107258-2014W>

bis zum 7. November 2014, 23:45 Uhr.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_3.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 3 Stunden vor der Übung, d.h. Mo., 11:00. Siehe <https://tuwel.tuwien.ac.at/mod/checkmark/submissions.php?id=196792> .