


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

**Übung 7**

**24. November 2010**  
**Dutter**

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

**Bemerkung:** Um die Ähnlichkeit der Formate der Marathondaten über die Jahre hinweg zu verbessern, wurde die Kopfzeile des Datensheets auskommentiert. Dadurch ändert sich die Einlesefunktion im Aufruf von `read.table`:


```
> f.mara_read <- function() {
+   mara10 <- read.table("mara_t_10", sep = ":", header = FALSE)
+   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"))
+   datsex10 <- substr(kmar10, 1, 1)
+   datakl10 <- substr(kmar10, 3, 5)
+ }
> f.mara_read()
```

Man betrachte nur die Altersklassen 35, 40 und 45. Haben diese Klassen einen signifikanten Einfluss auf die Laufzeiten der Männer? Das Gleiche bei den Frauen?

Wiederholen Sie diese Analyse mit einer „Stichprobe“ der Größe 1000 aus dem Gesamtdatensatz durch zufälliges Auswählen (ohne zurücklegen). Was ist anders?

26. Man betrachte die Laufzeiten der Altersklassen 35, 40 und 45 und der beiden Geschlechter. Analysieren Sie diese Aufgabe mittels 2-Weg-Varianzanalyse ohne Wechselwirkung. Bitte um Kommentare zu den Resultaten.

Wiederholen Sie diese Analyse mit der vorigen „Stichprobe“ der Größe 1000 aus dem Gesamtdatensatz. Was ist anders?

27. Analysieren Sie die Daten des Vienna City Marathons in den letzten 10 Jahren 2001 bis 2010 mit dem Computer Program System .

Lesen Sie alle 10 Datensätze ein und merken Sie sich die Endzeiten.

Hinweis: Mit einer For-Schleife natürlich lösbar, aber aufpassen auf verschiedene Anzahlen von Zwischenzeiten in den verschiedenen Jahren. Die Leseroutine könnte so aussehen:

```
> 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
+   else if (ncol(mara) == 7)
+       dmara <- mara[, 5] + mara[, 6]/60 + mara[, 7]/3600
+   else dmara <- mara[, 2] + mara[, 3]/60 + mara[, 4]/3600
+   dmara <- cbind(mara[, 2] + mara[, 3]/60 + mara[, 4]/3600,
+       dmara)
+   kmara <- mara[, 1]
+   dimnames(dmara) <- list(kmara, c("1. Zwischenz.", "Endzeit"))
+   datasex <- substr(kmara, 1, 1)
+   datakl <- substr(kmara, 3, 5)
+   return(list(mara = mara, dmara = dmara, kmara = kmara, datasex = datasex,
+       datakl = datakl))
+ }
> endzeiten <- matrix(0, ncol = 2, nrow = 0)
> for (year in 1:10) {
+   chmara <- paste("mara_t_0", year, sep = "")
+   if (year == 10)
+       chmara <- "mara_t_10"
+   marathon <- f.mara_read(chmara)
+   e <- cbind(marathon$dmara[, 2], year)
+   endzeiten <- rbind(endzeiten, e)
+ }

```

(Vorsicht: Die Matrix `endzeiten` hat nun 64.518 Zeilen!)

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

28. Führen Sie einen Test auf Gleichheit der jährlichen Mittel der Laufzeiten durch.

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?id=2604>

bis incl. 29. November 2010.

Bitte den Namen des pdf-File folgendermaßen:

**name\_exer\_7.pdf**

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