

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

Übung 8

1. Dezember 2010
Dutter

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

Betrachten Sie auch die letzten (wenn es mehrere gibt) Zwischenzeiten. Dazu sollten Sie die Funktion `f.mara_read` vom letzten Mal geringfügig verändern, womit Sie auch Zwischenzeiten, wenn vorhanden, verwenden können:

```
> 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
+     zmara <- mara[, 5] + mara[, 6]/60 + mara[, 7]/3600
+   }
+   else if (ncol(mara) == 7) {
+     dmara <- mara[, 5] + mara[, 6]/60 + mara[, 7]/3600
+     zmara <- mara[, 2] + mara[, 3]/60 + mara[, 4]/3600
+   }
+   else {
+     dmara <- mara[, 2] + mara[, 3]/60 + mara[, 4]/3600
+     zmara <- NA
+   }
+   kmara <- mara[, 1]
+   datasex <- substr(kmara, 1, 1)
+   datakl <- substr(kmara, 3, 5)
+   return(list(mara = mara, zmara = zmara, dmara = dmara, kmara = kmara,
+     datasex = datasex, datakl = datakl))
+ }
```

Man bemerke, dass der allgemeine Mittelwert der Männerlaufzeiten über die 10 Jahre 3.990418 Stunden und der der Frauen 3.994557 beträgt, also die Mittelwerte

nur im Tausendstelbereich an Stunden bewegt.

Untersuchen Sie nun die Endzeiten in Abhängigkeiten der Altersklassen, des Geschlechts und der Jahre 2001 bis 2010. Verwenden Sie dabei wie früher nur die Altersklassen `c("H", "30", "35", "40", "45", "50", "55", "60", "65", "70")`

Führen Sie die gleiche Untersuchung mit einer zufälligen Untermenge der „Beobachtungen“ von nur 1000 Werten durch. Vor der „zufälligen“ Auswahl mit der Funktion

```
sample
```

setzen Sie bitte aber den Startwert des Zufallszahlengenerators auf einen Standardwert, nämlich mit

```
set.seed(5).
```

30. Man betrachte die Studie über die Wirksamkeit von Schmerztabletten (ähnlich dem Beispiel im Skriptum): Insgesamt testen 20 Personen 4 verschiedene Schmerzmittel. Die Daten sind in Tabelle 1 dargestellt. Die Personen „sind in den Medikamenten geschachtelt“, weil jede Person nur ein Medikament testen kann. Man prüfe die unterschiedliche Wirkung der Medikamente mit Hilfe des „geschachtelten Versuchsplanes“. Die Daten stehen unter dem Namen „schmerz.txt“ auf der Vorlesunghomepage <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>.

		P1	P2	P3	P4	P5
M1	1	8.1	10.6	7.5	14.8	11.0
	2	7.5	11.5	12.2	11.7	9.1
	3	14.5	11.2	8.1	3.4	13.4
M2	1	10.9	11.0	13.8	13.5	12.8
	2	13.8	13.3	11.2	5.0	12.9
	3	10.8	10.5	6.6	9.6	12.3
M3	1	16.1	11.7	13.2	11.8	7.9
	2	10.8	10.8	11.8	15.3	14.3
	3	11.5	11.2	14.1	13.7	9.9
M4	1	10.9	14.1	15.3	12.7	15.6
	2	14.2	11.2	14.0	9.6	17.3
	3	18.9	11.9	9.9	14.7	12.6

Tabelle 1: Wirkungsdauer von Schmerztabletten

31. Analysieren Sie die „Werner Blood Chemistry“ Daten (`werner_bcd`) in der folgenden Art:

(a) Entfernen Sie die beiden Ausreisser im `CHOLSTRL` und die Personen mit fehlenden Werten mit `na.omit`, sodass 180 Beobachtungen übrig bleiben.

(b) Teilen Sie sowohl die Altersvariable in Gruppen mit `AGE_fac <- cut(werner_bcd_corr[, 'AGE'], c(18, 25, 32, 42, 55))` als auch die Gewichte mit

```
q <- quantile(werner_bcd_corr$WEIGHT)
q[1] <- q[1] - 1
werner_bcd_corr$WEIGHT_fac <- cut(werner_bcd_corr$WEIGHT, q)
```

(c) Führen Sie eine Varianzanalyse durch mit `CHOLSTRL` abhängig von den Faktoren `AGE_fac` und `WEIGHT_fac` ohne und mit Wechselwirkungen.

32. Nehmen Sie das Beispiel 31. und dann an, dass die Patientinnen zufällig aus einem Kollektiv gewählt wurden, wodurch sich die Annahme ergibt, dass die Effekte von sowohl Faktor **Alter** als auch von **Gewicht** als zufällig angenommen werden können. Führen Sie wieder eine Varianzanalyse durch. (Bem.: Mit `Error(WEIGHT_fac + AGE_fac)` im Modell verwenden Sie die Annahme der zufälligen Effekte. Analog mit Wechselwirkung.)

Wie schon im Skriptum zu sehen, werden mit \mathcal{R} weder F-Werte noch p-Werte gedruckt. Versuchen Sie händisch bzw. besser mit \mathcal{R} die fehlenden Ergebnisse einzutragen, sodass z.B. so etwas Ähnliches wie

```
Error: WEIGHT_fac
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  3 8074.8  2691.6  0.5879 0.4619
```

entsteht. Studieren Sie das Ergebnis im Vergleich mit der Annahme der festen Effekte.

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. 12. Dezember 2010.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_8.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/>.)