

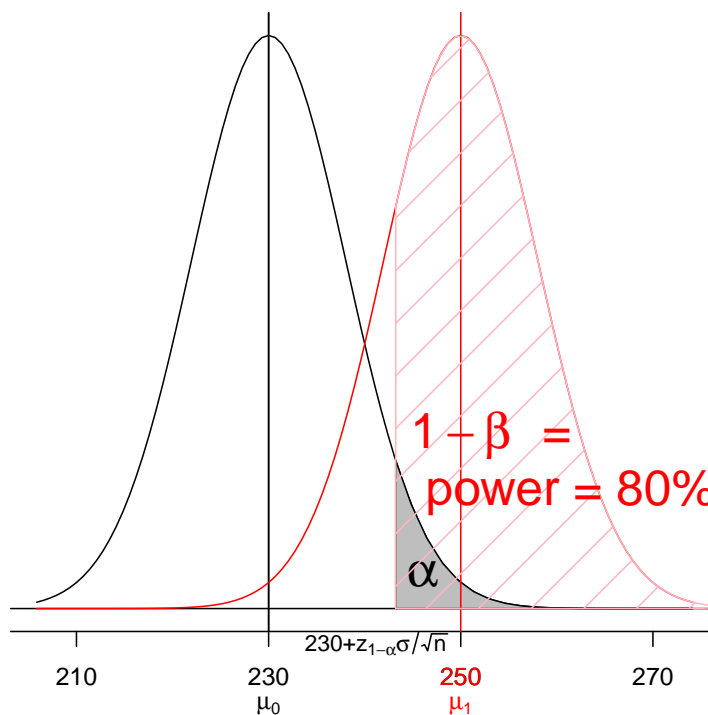
STATISTIK 2 (107.325) WS 2013
COMPUTERSTATISTIK (107.258) WS 2013

Übung 6

22. November 2013
Dutter

21. Betrachten wir die Werte von Cholesterin im Datensatz „werner.bcd“. Nach der Entfernung der beiden Ausreißer wird die Hypothese, dass der Mittelwert aller möglichen Cholesterinwerte gleich $\mu_0 = 230$ ist, aufgestellt.

Wie groß muss die Stichprobe sein, damit mit Wahrscheinlichkeit 80% (Schärfe, power) der wahre Wert $\mu_1 = 250$ entdeckt wird? (Als Approximation der Varianz nehme man die empirische aus dem obigen Datensatz.)



22. Betrachten Sie die Daten des Vienna City Marathons in den letzten 10 Jahren 2004 bis 2013 mit dem Computer Program System .

Betrachten Sie auch die letzten Zwischenzeiten. Dazu sollten Sie die Funktion `f.mara_read` geringfügig ändern, womit Sie auch Zwischenzeiten, wenn vorhanden, verwenden können:

```
f.mara_read <- function(dat) {
  # dat sollte den Datennamen darstellen (z.B. 'mara_t_13')
  # Einlesen von 10 Spalten
  chdat <- paste("http://www.statistik.tuwien.ac.at/public/dutt/vorles/data/"
    ,
```

```

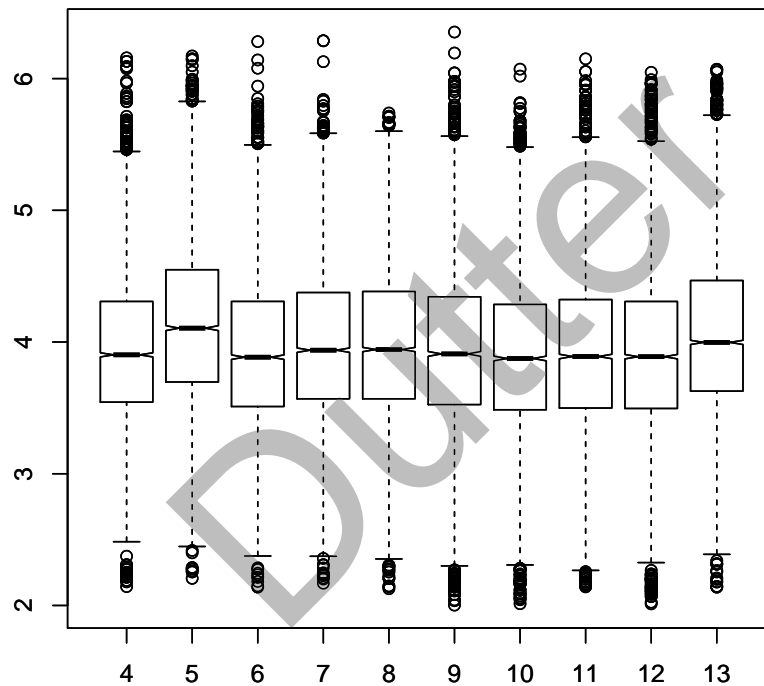
    dat, sep="")
mara <- read.table(chdat, sep=":")
if ( ncol(mara) == 10 )
  colnames(mara) <- c("Kl-Sex", "1.Zw.Zeit_h", "1.Zw.Zeit_m",
    "1.Zw.Zeit_s", "2.Zw.Zeit_h", "2.Zw.Zeit_m", "2.Zw.Zeit_s",
    "Endzeit_h", "Endzeit_m", "Endzeit_s")
else if ( ncol(mara) == 7 )
  colnames(mara) <- c("Kl-Sex",
    "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])
# Errechnung der Marathon-Endzeiten (und ev. der letzten Zwischenzeiten)
# in Stunden
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 <- rep(NA, length(dmara))
}
# Geschlecht- und Klassenkennzeichnung
kmara <- mara[,1]
# Herauslesung von `Geschlecht`
datasex <- substr(kmara,1,1)
# Herauslesung von `Altersklasse`
datakl <- substr(kmara,3,5)
return(list(mara=mara, zmara=zmara, dmara=dmara, kmara=kmara, datasex=datasex,
  datakl=datakl))
}

```

In einer for-Schleife können die Gesamtlaufzeiten pro Jahr (mit den entsprechenden Kenngrößen) eingelesen und die resultierenden Listenelemente aus `f.mara_read` zu einem data.frame mit 57 548 Zeilen zusammengefügt werden.

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

Bemerkung: Sollten sich große Schwierigkeiten mit Speicherplatz und/oder Rechenzeit ergeben, kann man auch eine (zufällige !!) Auswahl von z.B. 1000 Teilnehmern pro Jahr verwenden.



23. Untersuchen Sie nun die Endzeiten in Abhängigkeiten der verschiedenen Variablen:

- Betrachten Sie zunächst nur Altersklassen (ohne Unterscheidung der Geschlechter) durchgehend über die Jahre 2004 bis 2013. Verwenden Sie dabei nur die Klassen `c("H", "30", "35", "40", "45", "50", "55")`.
- Verwenden Sie nur die männlichen Teilnehmer. Ist das Ergebnis anders?
- Gilt Ähnliches für die Frauen?
- Bitte um Kommentare, auch bezüglich anderer Werte als den p-Werten, z.B. Teststatistiken, Quadratsummen ...

24. Stellen Sie eine „Versuchsplanmatrix“ \mathbf{X} für das Modell „CHOLSTRL hängt linear von den Klassen von `bmi` ab“, beschrieben im Übungsblatt 5. Präsentieren Sie (zumindest auszugsweise) die Matrix.

Hinweis: Die Funktion `model.matrix` kann hilfreich sein.

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?idnumber=107258-2013W>
bis zum 29. November 2013, 23:45 Uhr.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_6.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 <http://www.statistik.tuwien.ac.at/cgi-bin/uebbbsp/bspstart.cgi> .