

# Statistik-Praktikum/WiMa-Praktikum II - Übungsblatt 1

Vorstellung der Ergebnisse in der Übung am 23.04.2015

## Aufgabe 1

- (a) Lies die Datei 'olympics.csv' ein und erstelle aus ihr einen Data Frame. Hinweis: Mit dem Befehl `read.csv` kann man csv-Dateien einlesen, und mit der Option `sep=" "` kann man einen beliebigen Spaltentrenner eingeben.
- (b) Gib folgende Informationen aus dem erstellten Data Frame aus:
  - i) alle Städte, in denen olympische Spiele stattgefunden haben,
  - ii) Anzahl der Teilnehmer, die 1960 an den olympischen Spielen mitgewirkt haben,
  - iii) alle Jahre, in denen mehr als 100 Nationen teilgenommen haben.
- (c) Überschreibe den Data Frame mit einem neuen Data Frame, der alle Spalten bis auf 'Sports' und 'Events' enthält.
- (d) Erstelle eine neue Spalte 'WRatio', fülle sie mit der Größe 'Women/Participants' und binde sie in den Data Frame ein.
- (e) Schreibe den neuen Data Frame in eine Datei 'olympicsratio.txt' (nicht csv).
- (f) Bestimme den maximalen Anteil an weiblichen Sportlerinnen. In welchem Jahr wurde dieser Anteil erreicht?
- (g) Plote den Verlauf des Anteils an weiblichen Sportlerinnen von 1896-2008 und speichere ihn in einer pdf-Datei.

## Aufgabe 2

- (a) Erstelle eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x \exp(-|x|)$ .
- (b) Simuliere 10.000 standardnormalverteilte Zufallsvariablen und wende die Funktion  $f$  auf sie an. Gib den geschätzten Mittelwert und die geschätzte Varianz der resultierenden Werte aus.

## Aufgabe 3

Das Ergebnis der Bundestagswahl 2009 lautet wie folgt:

	CDU/CSU	SPD	FDP	Grüne	Linkspartei	Andere
Stimmanteile	33.80 %	23.03 %	14.56 %	10.71 %	11.89 %	6.01 %
Gewinn / Verlust	-1.37 %	-11.22 %	4.73 %	2.59 %	3.18 %	2.09 %

Stelle das Ergebnis mit zwei Balkendiagrammen nebeneinander in einem Schaubild dar. Die Stimmenteile sollen waagrecht und die Gewinne bzw. Verluste senkrecht liegen. Färbe die Balken mit den typischen Farben der Parteien ein.

#### **Aufgabe 4**

Das R-Paket MASS enthält den Datensatz `geyser`, der die Dauer von 299 Ausbrüchen des Geysirs ‘Old Faithful’, im Yellowstone Nationalpark und die Wartezeiten nach den Ausbrüchen (jeweils in Minuten) angibt. Nimm an, dass der erste Ausbruch genau zur Minute 0 stattgefunden hat. Stelle die Dauer der Ausbrüche als Kurve in Abhängigkeit von der Zeit des Ausbruchsbeginns dar. Achte dabei auf eine sinnvolle Zeiteinheit. Zeichne die mittlere Ausbruchsdauer mit einer farbigen Linie in die Grafik. Welche Aussagekraft besitzt dieser Mittelwert über die Dauer der Ausbrüche?

*Hinweis:* Folgender R-Code könnte nützlich sein:

```
s <- rep(0, length(x))
s[1] <- x[1]
for(i in 2:length(s)) {
s[i] <- x[i] + s[i-1]
}
```

Hiermit wird zunächst ein neuer Vektor `s` mit derselben Länge wie der bestehende Vektor `x` angelegt und mit Nullen gefüllt, bevor ihm in einer Schleife die Partialsummen von `x` zugewiesen werden.