

Blatt 1

Aufgabe 1: (2+2+2=6 Punkte)

Laden Sie den Datensatz *alterBev.csv* aus Moodle herunter und öffnen Sie den Datensatz in **R**. Der Datensatz enthält die im Zensus 2011 ermittelte Bevölkerungszahl von allen Bundesländern unterteilt in verschiedenen Altersgruppen.

```
setwd("path") #path ist der Pfad zum Verzeichnis der Datei alterBev.txt.  
alter<-read.table("alterBev.csv", header=T, sep=",")
```

- Berechnen Sie die Größe der verschiedenen Altersgruppen für Gesamtdeutschland.
- Berechnen Sie für jedes Bundesland das mittlere Alter und bestimmen Sie das Bundesland mit dem höchsten bzw. mit dem niedrigsten mittlerem Alter. Verwenden Sie hierbei für jede Altersgruppe die „goldene Mitte“, d.h. 2.5 Jahre für die erste Gruppe, 7.5 Jahre für die zweite usw., für die Gruppe der über 90-Jährigen verwenden Sie das Alter 92.5 Jahre.
- Erstellen Sie für das Bundesland mit dem höchsten mittlerem Alter und das mit dem niedrigsten mittlerem Alter einen Graphen, der als y-Achse den relativen Anteil der Altersgruppen an der Bevölkerung besitzt und als x-Achse die „goldene Mitte“ der Altersgruppen. Zeichnen Sie beide Graphen in einen Plot. Verwenden Sie zwei unterschiedliche Farben, beschriften Sie die Achsen und geben Sie dem Plot eine Überschrift und eine Legende.

Hinweis: Verwenden Sie den Befehl `help(plot)`, um mehr über die Möglichkeiten einen Plot zu beschriften zu erfahren. Um beide Graphen in einen Plot zu zeichnen, können Sie zum Beispiel den ersten Graphen mit `plot` zeichnen und den zweiten mit `points`. Mit Befehlen wie `colSums`, `rowSums`, `whichMax` und `whichMin` können Sie sich viele Schleifen sparen.

Aufgabe 2: (2+2+2=6 Punkte)

Wir werfen eine Münze mit den Seiten Kopf und Zahl, solange bis zum ersten Mal Kopf erscheint. Bei jedem Wurf fällt Kopf mit Wahrscheinlichkeit $p \in (0, 1)$. Sei G die Anzahl der Würfe mit Ergebnis Zahl bis zum ersten Mal Kopf fällt. Wir sind an der Verteilung von G interessiert.

- Schreiben Sie mithilfe von `while` und `sample` eine Funktion `muenze(p)`, die die Folge von Münzwürfen bis zum ersten Erfolg simuliert und den von G angenommenen Wert zurückgibt.

- b) Simulieren Sie G genau 10^4 mal für $p = 0.6$.
- c) Berechnen Sie für $n = 0, \dots, 10$ die relative Häufigkeit h_n mit der G den Wert n in Ihren Simulationen aus **b)** annimmt. Plotten Sie $h = (h_0, \dots, h_{10})$ zusammen mit n mithilfe von `plot`, beschriften Sie die Achsen sinnvoll und geben Sie dem Plot die Überschrift **Geometrische Verteilung**.

Aufgabe 3: (2 Punkte)

Im Laufe dieses Semesters wollen wir verschiedene stochastische Modelle simulieren. Ein besonders nützlicher Befehl hierbei ist der Befehl `sample`, den wir in dieser Aufgabe besser kennenlernen wollen. Definieren Sie in **R** die Vektoren:

```
S<-c("H", "e", "l", "l", "o", "_", "W", "o", "r", "l", "d")
P<-c(900,5,5,5,10,10,400,5,10,10,300)
```

Probieren Sie nun folgende Befehle aus:

- | | |
|---|--|
| 1. <code>sample(S)</code> | 5. <code>sample(S,100,TRUE,prob = P)</code> |
| 2. <code>sample(S,size=5)</code> | 6. <code>sample(S,size=3,prob = P)</code> |
| 3. <code>sample(S,15)</code> | 7. <code>sample(S,size=3,prob = P/sum(P))</code> |
| 4. <code>sample(S,15,replace=TRUE)</code> | 8. <code>?sample</code> |

Aufgabe 4: (2 Punkte)

Sie werden auf den Übungsblättern häufig mit den Datentyp `vector` arbeiten müssen. Die folgenden Befehlszeilen sollen Ihnen helfen sich mit diesem Datentyp vertraut zu machen. Probieren Sie diese aus und versuchen Sie zu verstehen, was die Befehle bewirken.

- | | | |
|--------------------------------------|---------------------------|----------------------------------|
| 1. <code>x<-1:5; x</code> | 7. <code>sum(x)</code> | 13. <code>x/y</code> |
| 2. <code>y<-6:10+1; y</code> | 8. <code>cumsum(x)</code> | 14. <code>x[1:3]</code> |
| 3. <code>z<-6:(10+1); z</code> | 9. <code>x+y</code> | 15. <code>x[-(1:3)]</code> |
| 4. <code>w<-seq(1,10,2); w</code> | 10. <code>x+z</code> | 16. <code>x%%2</code> |
| 5. <code>n<-numeric(5); n</code> | 11. <code>x*2</code> | 17. <code>x[x%%2==0]</code> |
| 6. <code>c(x,z)</code> | 12. <code>x**2</code> | 18. <code>sum(x[x%%2==0])</code> |

Aufgabe 5: (2 Punkte) Lesen Sie die Daten aus `m.txt` und `w.txt` bspw. mit dem Befehl `daten <- scan(m.txt)` ein. Es handelt sich um einen (repräsentativ generierten) Datensatz mit den Lebensaltern von 100 000 Frauen und Männern, die in Deutschland in den letzten Jahren verstorben sind. Analysieren Sie beide Datensätze folgendermaßen:

1. Bestimmen Sie mit `mean` den Mittelwert der Stichprobe.
2. Bestimmen Sie mit `median` den Median der Stichprobe.
3. Bestimmen Sie mit `sd` die Standardabweichung der Stichprobe.
4. Bestimmen Sie mit `var` die Varianz der Stichprobe.
5. Erzeugen Sie mit `boxplot` einen Boxplot.
6. Testen Sie den Befehl `range` an diesen Daten.
7. Testen Sie den Befehl `summary` an diesen Daten.

Erstellen Sie außerdem einen Plot, der Boxplots für beide Datensätze nebeneinander zeigt. (Hinweis: Der Befehl `par(mfrow=c(1,2))` könnte hilfreich sein.)