

Einführung in R

5. Aufgabenblatt

Präsenzaufgabe 1

Wir betrachten einen fiktiven Datensatz der Notenverteilung einer Klasse, den wir wie folgt erstellen:

```
set.seed(123)
Klasse <- sample(1:6, size=50, replace=TRUE,
                 prob=c(0.05,0.2,0.5,0.2,0.03,0.02))
```

- Erstellen Sie in R Tabellen, welche die absoluten, relativen und die kumulativen Häufigkeiten der Notenverteilung darstellen.
- Stellen Sie die Verteilung der Häufigkeiten in einem Tortendiagramm (`pie()`) und in einem Balkendiagramm (`barplot()`) dar.
- Modifizieren Sie die Standard-Voreinstellungen der Optionen für das Erstellen der Balkendiagramme (*space, names.arg, main, ylab, col*) und Tortendiagramme (*labels, radius, clockwise, col, main*).
- Definieren Sie die Matrix:

$$\text{mat1} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 0 \end{pmatrix}$$

Führen Sie dann die folgenden R-Codes aus und diskutieren Sie über deren Bedeutung.

```
layout(mat1)
barplot(table(Klasse))
pie(table(Klasse))
```

- optional*: Erstellen Sie ähnliche Grafiken wie in b) und c) mittels `ggplot2`.

Präsenzaufgabe 2

In einer mikrobiologischen Untersuchung sollten die Eigenschaften von Mikroorganismen in der Luft untersucht werden. Dazu wurde ein Nährboden auf einer runden Agarplatte 30 Minuten bei Zimmertemperatur offen im Raum stehen gelassen. Nach Inkubation über Nacht waren 40 Pilz- bzw. Bakterienkolonien gewachsen. Von diesen 40 Kolonien wurden der Durchmesser, die Farbe sowie die Antibiotikaresistenz (auf einer fünfstufigen Skala) bestimmt. Die erfassten Merkmale sind in der untenstehenden Tabelle erklärt.

Merkmal	Erläuterungen
durchmesser	Durchmesser in mm
resistenz	Antibiotikaresistenz mit den Ausprägungen: 1 = sehr sensitiv; 2 = sensitiv; 3 = intermediär; 4 = resistent; 5 = sehr resistent
farbe	Farbe mit den Ausprägungen: 1 = gelb; 2 = weißlich; 3 = braun; 4 = orange; 5 = farblos; 6 = rosa; 7 = grün

- Laden Sie den Datensatz `bakterien.txt` und erzeugen Sie jeweils für die qualitativen Merkmale `resistenz` und `farbe` die Tabelle der absoluten und der relativen Häufigkeiten.
- Stellen Sie Ihre Ergebnisse mit Hilfe von Kreis- und Balkendiagrammen dar (4 einzelne Grafiken).
- Speichern Sie die Grafiken aus b) jeweils als Bilddatei ab, z.B. PNG (`png()`).
- Stellen Sie die 4 Grafiken zusammen in einem gemeinsamen Plot dar.

Präsenzaufgabe 3

Der Datensatz `Titanic` aus dem Paket `datasets` enthält Daten zu Geschlecht, Alter, Fahrgastklasse und Überleben der Passagiere der *Titanic*. Die Daten liegen in Form eines 4-dimensionalen *array* vor.

- Erstellen Sie mit Hilfe der Funktion `structable()` aus dem Package `vcd` 1-dimensionale Tabellen für alle 4 Variablen, also z.B.:
`tab <- structable(~ Sex, data = Titanic)`
- Erstellen Sie mit Hilfe der Funktionen `fable()` und `structable()` Tabellen, in den Sie 2, bzw. 3 bzw. 4 verschiedene Merkmale auf einmal darstellen. Beispiel:
`tab1 <- structable(Survived ~ Sex, data = Titanic)`
`tab2 <- structable(Sex+Age ~ Class+Survived, data = Titanic)`
`tab3 <- ftable(Titanic, row.vars = "Class", col.vars = "Survived")`
`tab4 <- ftable(Titanic, row.vars = 1:2, col.vars = 3:4)`
- Stellen Sie die Daten mit Hilfe Kreisdiagrammen, Balkendiagrammen und gruppierten Balkendiagramme dar. Beispiel zu gruppierten Balkendiagrammen:
`barplot(ftable(Titanic, row.vars=1, col.vars=4),
 legend=dimnames(Titanic)$Class,
 names.arg=dimnames(Titanic)$Survived,
 beside=TRUE, xlab="Survived")`

Aufgabe 1 (12 Punkte)

Der Datensatz `Arthritis` aus dem Package `vcd` enthält die Studienergebnisse zur Wirkung einer Behandlungsmethode bei Arthritis-Patienten. Der Datensatz enthält die Patientenummer (`ID`), die Behandlungsart (`Treatment`), das Geschlecht (`Sex`), das Alter (`Age`) und die Wirkung der Behandlung (`Improved`).

- Erstellen Sie 1-, 2- und 3- dimensionale Kontingenztafeln für die nominal-skalierten Merkmale `Treatment`, `Sex` und `Improved`.
- Stellen Sie die Verteilung des Merkmals `Improved` nach `Treatment` für Männer und für Frauen getrennt mit Hilfe von Kreisdiagrammen dar und speichern Sie alle Grafiken in einer *pdf-Datei*. Benutzen Sie die Funktion `layout`, um Ihre Grafikelemente in einer geeigneten Reihenfolge und Größe zusammenzustellen.
- Veranschaulichen Sie nun die Verteilung des Merkmals `Improved` nach `Treatment` mit Hilfe von gruppierten Balkendiagrammen.

Aufgabe 2 (8 Punkte)

Den Datensatz `m11survey` im Paket `tigerstats` (Datensatz auch in StudIP vorhanden) enthält die Daten von StudentInnen des `Georgetown College`.

- Welche Variablen in diesem Datensatz sind qualitative Merkmale?
- Erstellen Sie mit Hilfe der Funktionen `structable()` und `fable()` informative Tabellen in Bezug auf die qualitativen Merkmale aus diesem Datensatz.
- Stellen Sie die Verteilung der qualitativen Merkmalen graphisch dar. Sie können dazu z.B. Tortendiagramme und (einfache oder gruppierte) Balkendiagramme zurückgreifen.

Zusatzaufgabe: Titanic-Daten (3 Punkte)

Wenden Sie die Funktion `mosaicplot(...)` auf den `Titanic`-Datensatz an, um informative Grafiken zu erstellen. Interpretieren und kommentieren Sie die Grafiken.

Abgabe der Lösungen: bis Montag 18.11.2019,

maendle@uni-bremen.de