

Aufgabenblatt 7

Statistik für Wirtschaftsinformatiker, Übung, HTW Berlin

Martin Spott, Michael Heimann

Stand: 29.05.2023

Wiederholung

- Wofür werden Kontingenztabelle (Kontingenztafeln) verwendet? Wie sind sie aufgebaut?
- Was versteht man unter Randsummen/Randverteilungen?
- Was ist eine bedingte relative Häufigkeit?

Aufgabe 7.1

Die folgende Kontingenztabelle enthält Daten zu landwirtschaftlichen Betrieben nach Führung und Größe in Hektar.

```
library(knitr)
options(knitr.kable.NA='') # zeige NAs als Leerzeichen

betriebe <- matrix(c(NA,64,41,NA,487,131,41,659,203,NA,NA,389,54,91,17,162,
                     46,112,18,176,1429,551,NA,NA),
                  nrow=6,ncol=4,byrow=T,
                  dimnames=list(c("[0,50)", "[50, 180)",
                                   "[180, 500)", "[500, 1000)", ">= 1000", "Summe"),
                                c("Vollzeit", "Nebenerwerb", "Pacht", "Summe")))
kable(betriebe)
```

	Vollzeit	Nebenerwerb	Pacht	Summe
[0,50)		64	41	
[50, 180)	487	131	41	659
[180, 500)	203			389
[500, 1000)	54	91	17	162
>= 1000	46	112	18	176
Summe	1429	551		

- Ergänze die Tabelle um die sechs fehlenden Einträge.
- Wie viele Betriebe haben weniger als 50 Hektar?
- Wie viele Betriebe werden von einem Pächter geführt?
- Wie viele Betriebe werden im Nebenerwerb betrieben und haben zwischen 500 und 1000 Hektar?
- Wie viele Betriebe werden nicht im Vollerwerb betrieben?
- Wie viele Pachtbetriebe haben 180 Hektar oder mehr?
- Berechne die Kontingenztabelle mit den relativen Häufigkeiten. Runde auf die dritte Stelle nach dem Komma.

Aufgabe 7.2

Von 70 Studienanfängern wurden die Abiturnoten in Mathematik und Englisch erfasst. Die Tabelle zeigt die Noten von sechs Studierenden, der Datensatz enthält die Noten von allen.

```
noten <- read.csv("noten_mathematik_englisch.csv")
kable(head(noten))
```

Mathematik	Englisch
1	1
1	2
1	2
1	3
1	3
1	3

- Erstelle eine Kontingenztafel mit den zugehörigen Randverteilungen.
- Erstelle eine zweite Tabelle, die die relativen Häufigkeiten enthält. Runde auf die dritte Stelle nach dem Komma.
- Erstelle zu den Daten einen Mosaikplot.

Aufgabe 7.3

Gegeben sei die folgende zweidimensionale Häufigkeitstabelle der beiden Merkmale X und Y für insgesamt 10 Beobachtungen bzw. Beobachtungspaare.

	x_1	x_2	Σ
y_1		3	
y_2			
Σ			

Es ist weiter bekannt, dass $f(y_2|x_1) = 0,5$ und $f(y_1|x_2) = 0,5$ ist. Bestimme die fehlenden absoluten Häufigkeiten und trage diese in die Häufigkeitstabelle ein.

Aufgabe 7.4

Im folgenden sind die Ergebnisse einer (fiktiven)¹ medizinischen Untersuchung zur Wirkung einer Hautsalbe zur Behandlung von Hautausschlägen. Die Frage ist, ob nach Anwendung der Salbe der Ausschlag besser oder schlimmer wird.

```
ausschlag <- matrix(c(223, 75, 107, 21), nrow=2, byrow=T,
                     dimnames=list(c("mit Salbe", "ohne Salbe"), c("besser", "schlimmer")))
kable(ausschlag)
```

	besser	schlimmer
mit Salbe	223	75
ohne Salbe	107	21

- Ohne nachzurechnen, hilft die Salbe oder schadet sie eher?
- Ergänze die Tabelle um die Randsummen.
- Fertige eine Tabelle für die relativen Häufigkeiten (für die einzelnen Zellen) inklusive der Randsummen an.
- Bestimme die bedingten relativen Häufigkeiten für die Zeilen und gib deren formale Darstellung $f(\dots)$ an.

- e) Erstelle (auf Papier) einen relativen Häufigkeitsbaum zu den Daten sowie in R den zum Baum äquivalenten Mosaikplot. Mache dir den Zusammenhang der beiden Darstellungen klar.
- f) Ist es aufgrund der Datenlage empfehlenswert, die Salbe gegen einen Hautausschlag zu verwenden?

¹ aus einem Experiment präsentiert in dem Paper *Motivated Numeracy and Enlightened Self-Government* von D. M. Kahan u.a., http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319992