

# Aufgabenzettel 02

Gruppe 01

29.4.2020

## Aufgabe 04 Mit Zettel und Bleistift - Wahrscheinlichkeiten

- a) Verallgemeinern Sie den Additionssatz  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  für drei einander nicht ausschließende Ereignisse  $P(A \cup B \cup C)$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- b) Eintrittswahrscheinlichkeit von A für  $P(A | B)$  für die Fälle

- A und B sind unabhängig voneinander:  $P(A | B) = (P(A \cap B))/P(B)$
- A und B schließen sich einander aus:  $P(A | B) = (P(A \cap B))/P(B)$

$$P(A \cap B) = P(A) * P(B)$$

Daraus folgt:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A)$$

A und B schließen einander aus

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

- c) Multiplikationssatz:  $P(A \cap B \cap C) = P(A) * P(B | A) * P(C | A \cap B) = 11/22 * 10/21 * 9/20 = 0,1071$

$P(A)$ ,  $P(B)$  und  $P(C)$  sind jeweils die Wahrscheinlichkeiten dafür, dass ein Spieler eine rote Karte bekommt. Nachdem ein, bzw. zwei Spieler eine rote Karte bekommen haben, verringert sich jeweils die Grundgesamtheit und die Anzahl an günstigen Fällen um eins.

(Hedderich and Sachs 2018)

## Aufgabe 05 Der Übungsdatensatz

```
#Daten laden
data <- read.delim("data/yingtan_20_v1.csv",
                  header = TRUE,
                  sep = ";",
                  dec = ",")
```

- a) In was für einer Datenstruktur wurde die geladene Datei in R abgelegt? Wie viele Beobachtungen und Variablen enthält sie? Von welchem Datentyp ist die Variable 'Ca\_exch'?

```
#Datenstruktur
str(data)
```

```
## 'data.frame': 337 obs. of 9 variables:
## $ OBJECTID: int 1 2 3 4 5 6 7 8 9 10 ...
## $ SAMPLING: chr "regular" "regular" "regular" "regular" ...
## $ EAST : num 490591 490591 490591 490741 490741 ...
## $ NORTH : num 3123240 3123390 3123540 3123390 3123540 ...
## $ C : num 0.599 0.647 0.527 0.812 0.756 ...
## $ Ca_exch : num 12.38 13.13 3.77 31.69 24 ...
## $ Mg_exch : num 3.2 4.77 1.45 7.72 8.33 ...
## $ K_exch : num 3.509 0.992 4.046 2.655 4.02 ...
## $ Na_exch : num 0.452 0.592 0.122 0.687 0.244 0.383 0.244 0 0 0.435 ...
```

```
str(data$Ca_exch)
```

```
## num [1:337] 12.38 13.13 3.77 31.69 24 ...
```

- b) Sie haben sich mit dem Aufbau des Datensatzes vertraut gemacht und bemerkt, dass für zwei Standorte keine Messwerte vorliegen. Erstellen Sie ein neues Objekt `ljz`, in dem diese Proben nicht mehr auftauchen.

```
#NA definieren
is.na(data) <- data == -9999.000
#Entfernung von NA-Werten
ljz <- na.omit(data)
```

- c) Beim Transfer der Laborergebnisse in die Messtabelle scheint bei den austauschbaren Na-Ionen ein Missgeschick passiert zu sein. Identifizieren Sie den fehlerhaften Wert und ersetzen Sie ihn durch die in R gängige NoData-Bezeichnung 'NA'. Fallen Ihnen zusätzliche Werte auf, die bei möglichen Rechenoperationen zu Problemen führen könnten? Begründen Sie ihre Wahl und beheben Sie das Problem auf angemessene Weise.

```
#Spalte mit Na-Ionen
ljz$Na_exch
```

```
## [1] 0.452 0.592 0.122 0.687 0.244 0.383 0.244 0.000 0.000 0.435
## [11] 0.165 0.000 0.374 0.113 0.592 0.522 0.122 0.174 0.183 0.104
## [21] 0.452 0.339 0.505 0.505 0.400 0.548 0.800 1.018 0.130 0.130
## [31] 0.365 0.104 0.426 0.322 0.365 0.635 0.078 0.043 0.035 0.052
## [41] 0.070 0.400 0.487 0.418 0.104 0.278 0.391 0.644 0.200 0.217
## [51] 0.191 0.270 0.391 0.592 0.557 0.070 0.287 0.374 0.000 0.017
## [61] 0.313 0.261 0.000 0.531 0.017 0.017 0.261 0.844 0.670 0.731
## [71] 0.122 0.322 0.600 0.278 0.565 0.313 2.262 0.809 0.444 0.496
## [81] 0.522 0.661 0.357 1.644 0.261 0.052 0.670 0.087 0.252 0.365
## [91] 0.887 0.452 0.304 0.148 0.478 -0.496 0.217 0.452 0.035 0.130
## [101] 0.209 0.374 0.557 0.078 0.244 0.339 0.070 0.322 0.070 0.348
## [111] 0.339 0.157 0.331 0.304 0.522 0.322 0.557 0.009 0.035 0.278
## [121] 0.078 0.583 0.539 0.183 0.304 2.166 0.052 0.026 0.061 0.078
## [131] 0.209 0.304 0.696 0.426 0.522 0.522 0.574 0.644 0.696 0.757
## [141] 0.539 0.435 0.496 0.531 0.574 1.114 0.679 0.400 0.766 0.357
## [151] 0.383 0.478 0.670 0.661 1.053 0.313 0.209 0.339 0.444 0.278
## [161] 0.435 0.635 0.096 0.644 0.487 0.522 0.383 0.635 0.296 3.810
## [171] 0.365 0.461 0.792 0.348 0.383 0.861 0.470 0.505 0.809 0.374
## [181] 0.409 0.426 0.531 0.348 0.713 0.644 0.461 0.548 0.487 0.383
## [191] 0.774 0.792 0.722 0.505 0.522 0.557 0.887 0.679 0.357 0.565
## [201] 0.774 0.731 0.644 0.435 0.478 0.487 0.461 0.104 0.783 0.409
## [211] 0.122 0.130 0.565 0.531 0.670 0.731 0.505 0.513 0.565 0.496
## [221] 0.487 0.774 0.600 0.635 0.539 0.478 0.626 0.826 0.418 0.322
## [231] 0.583 0.400 0.383 0.635 0.661 0.600 0.670 0.548 0.766 0.574
## [241] 0.444 0.383 0.357 0.539 0.409 0.383 0.505 0.522 0.835 0.896
```

```
## [251] 1.305 1.140 1.522 1.087 0.774 0.618 0.522 0.583 0.679 0.722
## [261] 0.766 1.018 1.235 0.679 0.618 0.652 0.783 0.905 1.027 0.922
## [271] 0.679 0.757 0.879 0.687 1.174 1.244 0.731 1.270 0.966 1.244
## [281] 1.035 1.114 1.148 0.618 0.539 0.679 0.600 0.809 0.670 0.609
## [291] 0.644 0.696 0.644 0.853 0.418 0.522 0.435 0.426 0.592 0.426
## [301] 0.609 0.574 0.539 0.400 0.409 0.618 0.635 0.635 0.548 0.409
## [311] 0.374 0.670 0.609 0.696 0.583 0.618 0.696 0.826 0.496 0.766
## [321] 0.757 0.487 0.574 0.539 0.487 0.435 0.565 0.496 0.270 0.357
## [331] 0.435 0.322 0.217 1.018 0.174
```

```
#In Zeile 96 ist ein negativer Wert
#Diesen Wert als NA setzen
is.na(ljz[, "Na_exch"]) <- ljz$Na_exch < 0
#Außerdem gibt es in derselben Spalte einige Werte mit einer 0
#Da man nicht durch 0 teilen oder die Wurzel aus 0 berechnen kann,
#kann dies zu Problemen führen
is.na(ljz[, "Na_exch"]) <- ljz$Na_exch == 0
ljz$Na_exch
```

```
## [1] 0.452 0.592 0.122 0.687 0.244 0.383 0.244 NA NA 0.435 0.165 NA
## [13] 0.374 0.113 0.592 0.522 0.122 0.174 0.183 0.104 0.452 0.339 0.505 0.505
## [25] 0.400 0.548 0.800 1.018 0.130 0.130 0.365 0.104 0.426 0.322 0.365 0.635
## [37] 0.078 0.043 0.035 0.052 0.070 0.400 0.487 0.418 0.104 0.278 0.391 0.644
## [49] 0.200 0.217 0.191 0.270 0.391 0.592 0.557 0.070 0.287 0.374 NA 0.017
## [61] 0.313 0.261 NA 0.531 0.017 0.017 0.261 0.844 0.670 0.731 0.122 0.322
## [73] 0.600 0.278 0.565 0.313 2.262 0.809 0.444 0.496 0.522 0.661 0.357 1.644
## [85] 0.261 0.052 0.670 0.087 0.252 0.365 0.887 0.452 0.304 0.148 0.478 NA
## [97] 0.217 0.452 0.035 0.130 0.209 0.374 0.557 0.078 0.244 0.339 0.070 0.322
## [109] 0.070 0.348 0.339 0.157 0.331 0.304 0.522 0.322 0.557 0.009 0.035 0.278
## [121] 0.078 0.583 0.539 0.183 0.304 2.166 0.052 0.026 0.061 0.078 0.209 0.304
## [133] 0.696 0.426 0.522 0.522 0.574 0.644 0.696 0.757 0.539 0.435 0.496 0.531
## [145] 0.574 1.114 0.679 0.400 0.766 0.357 0.383 0.478 0.670 0.661 1.053 0.313
## [157] 0.209 0.339 0.444 0.278 0.435 0.635 0.096 0.644 0.487 0.522 0.383 0.635
## [169] 0.296 3.810 0.365 0.461 0.792 0.348 0.383 0.861 0.470 0.505 0.809 0.374
## [181] 0.409 0.426 0.531 0.348 0.713 0.644 0.461 0.548 0.487 0.383 0.774 0.792
## [193] 0.722 0.505 0.522 0.557 0.887 0.679 0.357 0.565 0.774 0.731 0.644 0.435
## [205] 0.478 0.487 0.461 0.104 0.783 0.409 0.122 0.130 0.565 0.531 0.670 0.731
## [217] 0.505 0.513 0.565 0.496 0.487 0.774 0.600 0.635 0.539 0.478 0.626 0.826
## [229] 0.418 0.322 0.583 0.400 0.383 0.635 0.661 0.600 0.670 0.548 0.766 0.574
## [241] 0.444 0.383 0.357 0.539 0.409 0.383 0.505 0.522 0.835 0.896 1.305 1.140
## [253] 1.522 1.087 0.774 0.618 0.522 0.583 0.679 0.722 0.766 1.018 1.235 0.679
## [265] 0.618 0.652 0.783 0.905 1.027 0.922 0.679 0.757 0.879 0.687 1.174 1.244
## [277] 0.731 1.270 0.966 1.244 1.035 1.114 1.148 0.618 0.539 0.679 0.600 0.809
## [289] 0.670 0.609 0.644 0.696 0.644 0.853 0.418 0.522 0.435 0.426 0.592 0.426
## [301] 0.609 0.574 0.539 0.400 0.409 0.618 0.635 0.635 0.548 0.409 0.374 0.670
## [313] 0.609 0.696 0.583 0.618 0.696 0.826 0.496 0.766 0.757 0.487 0.574 0.539
## [325] 0.487 0.435 0.565 0.496 0.270 0.357 0.435 0.322 0.217 1.018 0.174
```

## Aufgabe 06 Tinn R

- In R lassen sich Arbeitsschritte prima automatisieren. Stellen Sie sich vor, ihr Chef, der völlig zu Recht von R begeistert ist, möchte, dass Sie für ein neues Projekt aus dem ursprünglichen Datensatz nur die Standorte der Catena A extrahieren. Er verlangt von Ihnen ein R-Skript, das außerdem ...
- ... die Reihenfolge der Standorte nach abnehmender Ca-Ionenkonzentration sortiert.

- c) ... das Ergebnis in eine sinnvoll benannte CSV-Datei schreibt. Führen Sie die Schritte 6 a bis c in einer Pipe aus und binden Sie diese in ihr Protokoll ein. Speichern Sie die Pipe zusätzlich als .R-Skript in ihrem Kursordner ab und geben Sie im Protokoll den Befehl an, mit dem es sich in der Konsole ausführen lässt.

```
#Aus data Standorte der Catena A  
#Reihenfolge nach abnehmender Ca-Ionenkonzentration  
#als csv schreiben  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(tibble)  
catA <- dplyr::filter(data,  
                      SAMPLING == "catA") %>%  
  arrange(desc(Ca_exch)) %>%  
  write.table("data/catA.csv",  
             sep = ";",  
             dec = ",",
```

Öffnen des RSkripts mit der Funktion:

```
source(file = "catA.R")
```

## Literatur

Hedderich, Jürgen, and Lothar Sachs. 2018. *Angewandte Statistik*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-56657-2>.