Statistik Aufgabensammlung

Sommersemester 2016

Prof. Dr. Stefan Etschberger – Hochschule Augsburg

Anmerkungen zu den Übungsaufgaben:

- Nach der Vorlesung finden Sie jeweils in der Aufgabensammlung die für die jeweilige Woche zu bearbeitenden Aufgaben; besprochen und gelöst werden die Aufgaben in der darauf folgenden Woche in den Übungsgruppen.
- Grundlagen in R sind ein wichtiger, obligatorisch zu erlernender Bestandteil des Kurses; alle in den Übungsaufgaben behandelten Lösungen in R sind prüfungsrelevant und müssen auch bei veränderter Aufgabenstellung (ohne Rechner) gelöst werden können.
- Es gibt für die Klausur keine Einschränkung auf nur eine Aufgabe mit R. Klausuraufgaben mit R könnten in der Prüfung bei verschiedenen Themen als Teilaufgabe oder als separate Aufgabe eingebaut sein. R-Teile in der Klausur können, müssen aber nicht als single choice formuliert sein.
- Es gibt *kein* vorgefertigtes „cheat-sheet“ mit den wichtigsten Funktionen in R für die Klausur; bitte schreiben Sie sich die wichtigsten Funktionen inkl. Parametern auf Ihre selbsterstellte Formelsammlung. Vorausgesetzt werden für die Klausur alle in den Lösungshinweisen der Übungsaufgaben verwendeten Funktionen.
- (Teil)aufgaben, deren Lösungen (auch) in R erarbeitet werden sollen, sind am Rand mit dem Symbol R gekennzeichnet.

Inhalt

Aufgaben zu R Grundlagen

Aufgabe 1

Installation und Kennenlernen von R und RStudio

R

(Sofern Sie über keinen eigenen Rechner verfügen, können Sie im Rechnerraum im W-Gebäude arbeiten; dort sind R und Rstudio installiert)

a) Installieren Sie R von `http://goo.gl/ALaUXu` (für Windows) bzw. von `http://cran.r-project.org/bin/` für andere Plattformen.

R ist das Statistikprogramm, das Daten verarbeitet und die Ergebnisse ausgibt; es ist in der Rohfassung nicht sehr komfortabel zu bedienen. Deswegen arbeiten wir in diesem Kurs mit RStudio, einer sehr komfortablen und mächtigen integrierten Entwicklungsumgebung.

b) Installieren Sie RStudio von `http://goo.gl/RX11dj`.
c) Öffnen Sie RStudio. Klicken Sie in den linken unteren Bereich des Fensters („Console“), tippen Sie

```
1 + 2
```

und schließen Sie die Eingabe mit Enter ab.

In der Kommandozeile der Konsole werden alle Anweisungen eingegeben und Textrückmeldungen des Programms ausgegeben; dazu gehören Ergebnisse, aber auch Hinweise, Warnungen und Fehlermeldungen, falls etwas nicht geklappt hat. Die Kommandzeile eignet sich auch prima als Taschenrechner. Kennt man die Bedeutung einer Funktion nicht, kann man ein Fragezeichen voranstellen und bekommt eine Erklärung (rechts im Hilfebereich).

Bei Rechenoperationen gelten die Vorrangregeln der Mathematik (Potenz vor Punkt vor Strich). Der Dezimaltrenner ist ein Punkt (kein Komma). Exponential-, Logarithmus- bzw. Quadratwurzeln berechnet man über Funktionsaufrufe, das Argument steht in runden Klammern. Groß- und Kleinschreibung macht einen Unterschied. Stellt man einer Zeile ein #-Zeichen voran, wird die Zeile von R nicht ausgeführt.

d) Geben Sie folgende Ausdrücke ein und erklären Sie jeweils das Ergebnis

```
2 + 3 * 4
(2 + 3) * 4
0.2 * 3 - 1.1
0,2 * 3
2^3^2
(2^3)^2
exp(1)
?exp
log(exp(1))
```

```
sqrt(16)
16^(1/2)
Sqrt(16)
# Das ist ein Kommentar.
```

Lösungshinweis:

```
2 + 3 * 4

## [1] 14

(2 + 3) * 4

## [1] 20

0.2 * 3 - 1.1

## [1] -0,5

# 0,2 * 3 # Fehler, ',' wird nicht als Dezimalkomma
# akzeptiert
2^3^2

## [1] 512

(2^3)^2

## [1] 64

exp(1)

## [1] 2,7183

log(exp(1))

## [1] 1

sqrt(16)

## [1] 4

16^(1/2)

## [1] 4

Sqrt(16)  # Fehler, sqrt() schreibt man mit kleinem 's'


## Error in eval(expr, envir, enclos): konnte Funktion "Sqrt" nicht finden
```

Aufgabe 2

R

Variablen, Zuweisungen und Funktionen

Zahlen (und andere Objekte) können in R in Variablen gespeichert werden. Dazu kann der Zuweisungsoperator `=` *oder alternativ* `<-` *beispielsweise folgendermaßen verwendet werden:*

```
x <- 3.5
x2 <- 1.5   # funktioniert genauso mit x2 = 1.5
```

Mit diesen Variablen kann dann weitergerechnet werden. In Variablennamen dürfen Buchstaben, Ziffern (nicht als erstes Zeichen), Punkte und Unterstriche (_) vorkommen. Diese Bezeichner dürfen keine Leerzeichen enthalten. Auch hier ist Groß- und Kleinschreibung zu beachten.

a) Weisen Sie der Variablen `x` den Wert 4 zu. Weisen Sie dann der Variablen `x.2` den folgenden Wert zu:

$$\sqrt{3x^2 + \ln\left(\frac{1}{e^x}\right) + 5}\,.$$

Funktionsaufrufe schreibt man in R mit einem Funktionsbezeichner, auf den direkt (keine Leerstelle!) ein Paar runder Klammern folgt. Innerhalb der runden Klammern können ein oder mehrere Argumente oder Parameter der Funktion stehen. Funktionen kann man auch verschachtelt aufrufen.
Die Funktion `ls()` *gibt die in der aktuellen Sitzung definierten Objekte aus. Mit* `rm(<Var>)` *kann man eine Variable löschen, wenn man ihren Bezeichner anstatt* `<Var>` *in die runden Klammern schreibt.*

b) Überlegen Sie was folgende Zeilen ausgeben und führen Sie diese dann in R aus, um Ihr Ergebnis zu überprüfen.

```
x
x.2
X
x + x.2
x.Produkt <- x * x.2
x.Produkt
x.Produkt <- x.Produkt * x
ls()
rm(x)
x
ls()
```

Außer Zahlen kann R auch mit Zeichenketten umgehen. Diese können in Objekten gespeichert werden, indem man die Zeichenkette in Anführungsstriche setzt. Zeichenketten, die Zahlen beinhalten werden nicht als Zahlen interpretiert. Man kann mit ihnen also nicht rechnen.

c) Welche Ausgabe bewirken folgende Zeilen? Überlegen Sie, bevor Sie die Eingabe in R ausprobieren.

```
tubby.1 <- "Tinky-Winky"
tubby.2 <- "Dipsy"
Zahl <- 10
keine.Zahl <- "10"
Zahl + 1
keine.Zahl + 1
```

Tricks zur Ein- und Ausgabe:

- *Ist eine Eingabe in einer Zeile nicht vollständig, kann R das mit einem „+“-Zeichen anzeigen; die Eingabe kann dann vervollständigt werden.*
- *Sofortige Hilfe bei der Eingabe einer Funktion erhält man, wenn man nach Eingabe der ersten Buchstaben des Funktionsbezeichners die Tabulator-Taste betätigt. Die möglichen Funktionen werden dann zur Auswahl aufgelistet und können dann ausgewählt werden.*
- *Mit der ↑-Taste auf der Tastatur kann der letzte (oder bei zweimaligem Drücken der vorletzte usw.) Befehl wieder sichtbar gemacht und dann nochmals ausgeführt oder verändert werden.*
- *Im RStudio-Fenster finden Sie (meistens rechts oben) einen Reiter History. Auch dort werden alle eingegebenen Befehle abgespeichert.*
- *Im Reiter Environment werden alle Objekte der aktuellen Sitzung aufgelistet.*

d) Probieren Sie die angesprochenen Tricks zur Ein- und Ausgabe aus.

Lösungshinweis:

```
x <- 4
x.2 <- sqrt(3 * x^2 + log(1/exp(x)) + 5)
x.2

## [1] 7

X


## Error in eval(expr, envir, enclos): Objekt 'X' nicht gefunden

x + x.2

## [1] 11

x.Produkt <- x * x.2
x.Produkt

## [1] 28

x.Produkt <- x.Produkt * x
rm(x)
x  # Fehler: x gibt's ja nicht mehr, kann deswegen auch nicht ausgegeben werden


## Error in eval(expr, envir, enclos): Objekt 'x' nicht gefunden
```

```
tubby.1 <- "Tinky-Winky"
tubby.2 <- "Dipsy"
Zahl <- 10
keine.Zahl <- "10"
Zahl + 1  # Ergebnis: 11

## [1] 11

keine.Zahl + 1  # Fehler: das geht nicht...


## Error in keine.Zahl + 1: nicht-numerisches Argument für binären Operator
```

Aufgabe 3

Daten: Vektoren

R

Eine Urliste von Daten eines Merkmals wird in R durch einen Vektor repräsentiert. Zur Erzeugung eines Vektors dient die Funktion `c()`. die Einträge der Urliste werden dann zum Beispiel als Argumente von `c()` durch Kommata getrennt angegeben. Als Ausprägungen sind Zahlen oder Zeichenketten möglich. R versucht dann durch die Art der Argumente automatisch zu entscheiden, ob es sich um ein nominales oder ein metrisches Markmal handelt.

a) Legen Sie eine Urliste für das Merkmal `x` an, das die Werte 1, 4, 2, 1.5 enthält. Geben Sie `x` aus. Legen Sie ein weiteres Merkmal `Geschlecht` mit den Werten `Mann`, `Frau`, `Frau`, `Frau` an. Geben Sie auch `Geschlecht` aus. Das dritte Merkmal `z` soll die Werte `1, 2, 1, "1"` enthalten. Ist `z` für R nominal oder metrisch? Überprüfen Sie Ihre Entscheidung.

Vektoren aufeinanderfolgender ganzer Zahlen werden mit dem Doppelpunkt-Operator gebildet. 2:5 steht zum Beispiel für den Vektor mit den Zahlen 2, 3, 4, 5. Mit der Funktion `seq()` kann man genauer Vektoren als Folgen von Zahlen erzeugen. `seq(from=2, to=3, by=0.2)` erzeugt zum Beispiel den Vektor (2, 2.2, 2.4, 2.6, 2.8, 3). Mit `rep()` werden Werte oder ganze Vektoren vervielfacht als Vektor ausgegeben. Zum Beispiel ergibt `rep(c(1,2), 3)` den Vektor (1,2,1,2,1,2). Die Hilfe-Seiten (Aufruf über `?seq` bzw. `?rep`) erklären die Details.

b) Erzeugen Sie folgende Vektoren in R:

```
## [1] 5 6 7 8 9
##  [1] 10  9  8  7  6  5  4  3  2  1
## [1] -0,10 -0,08 -0,06 -0,04 -0,02  0,00
## [1] 10000 12500 15000 17500 20000
##  [1] -3 -2 -1  0  1  2 -3 -2 -1  0  1  2 -3 -2 -1  0  1
## [18]  2
##  [1]  5,0  6,0  7,0  8,0  9,0 10,0 10,1 10,2 10,3 10,4
## [11] 10,5
```

Rechenoperationen können zwischen (numerischen) Vektoren elementweise ausgeführt werden. Hat ein Vektor weniger Elemente als ein anderer, werden die Elemente vom Beginn des kürzeren Vektors einfach solange wiederholt, bis die Länge der beiden Vektoren gleich ist. Die Länge eines Vektors kann mir der Funktion `length()` ausgelesen werden. Die Summe aller Elemente eines Vektors wird mit `sum()` errechnet. Beispielsweise ergibt mit `x=1:5` und `y = c(10.1, 10.5)` die Summe `x+y` den Vektor `(11.1, 12.5, 13.1, 14.5, 15.1)`. Analog funktioniert `-, *, /`.

c) Gegeben sind die Vektoren

```r
x <- 4:2
y <- seq(from = 0.1, to = 0.5, by = 0.1)
```

Erklären Sie, was folgende Ausdrücke ergeben und überprüfen Sie Ihr Ergebnis in R:

```r
x + y
x * y
x^3 + 1
2 * x - 3 * y
n <- length(x + y)
sum(x + y)/n
```

Teile oder einzelne Elemente eines Vektors können mit der Angabe der Indexwerte in eckigen Klammern ausgegeben werden. Auch Bedingungen mit Vergleichsoperatoren (z.B. < *für* kleiner als *oder* == *für ist gleich) sind möglich in eckigen Klammern. Verknüpfungen zwischen Vergleichen sind mit logisch UND (&) beziehungsweise ODER (|) möglich.*

d) Gegeben sind die Vektoren

```
x <- seq(from = 0, to = 100, by = 2)
y <- 100:1
```

Schreiben Sie die Ergebnisse folgender Ausdrücke auf und überprüfen Sie anschließend Ihr Ergebnis in R:

```
x[3]
y[c(1, 3, 10)]
x[1:4]
x[x > 91]
x[x > 20 & x <= 30]
y[y == 5 | y > 95 | y < 3]
```

Anmerkung: Die Ausgabe von Relationen wie $x < y$ auf Vektoren in R sind Vektoren mit den Ausprägungen `TRUE` beziehungsweise `FALSE`. Diese sogenannten logischen Vektoren können zur Indizierung von Vektoren verwendet werden; Elemente mit einem Index von `TRUE` werden ausgewählt, die mit Wert `FALSE` werden übergangen.

e) Was ergeben folgende Zeilen in R:

```
x <- seq(from = 0.2, to = 2, by = 0.3)
y <- -3:3
x < y
x^2 < x
Index <- x^2 < x
x[Index]
y[Index]
```

Lösungshinweis:

a)

```
x <- c(1, 4, 2, 1.5)  # Anlegen eines metrischen Merkmals x
# mit Ausprägungen für 4 Objekte
x  # Ausgabe

## [1] 1,0 4,0 2,0 1,5

Geschlecht <- c("Mann", "Frau", "Frau", "Frau")
Geschlecht  # nominales Merkmal, auch von 4 Objekten

## [1] "Mann" "Frau" "Frau" "Frau"

z <- c(1, 2, 1, "1")  # z ist für R nominal, da der letzte Wert
# als Zeichenkette eingegeben wurde
```

```
z
## [1] "1" "2" "1" "1"
```

b)
```
5:9
## [1] 5 6 7 8 9
10:1
##  [1] 10  9  8  7  6  5  4  3  2  1
seq(from = -0.1, to = 0, by = 0.02)
## [1] -0,10 -0,08 -0,06 -0,04 -0,02  0,00
seq(from = 10000, to = 20000, length.out = 5)
## [1] 10000 12500 15000 17500 20000
rep(-3:2, 3)
##  [1] -3 -2 -1  0  1  2 -3 -2 -1  0  1  2 -3 -2 -1  0  1
## [18]  2
c(5:10, seq(from = 10.1, by = 0.1, to = 10.5))
##  [1]  5,0  6,0  7,0  8,0  9,0 10,0 10,1 10,2 10,3 10,4
## [11] 10,5
```

c)
```
x <- 4:2
y <- seq(from = 0.1, to = 0.5, by = 0.1)
```

```
x + y
## [1] 4,1 3,2 2,3 4,4 3,5
x * y
## [1] 0,4 0,6 0,6 1,6 1,5
x^3 + 1
## [1] 65 28  9
2 * x - 3 * y
## [1] 7,7 5,4 3,1 6,8 4,5
n <- length(x + y)
sum(x + y)/n
## [1] 3,5
```

d)
```
x <- seq(from = 0, to = 100, by = 2)
y <- 100:1
```

```
x[3]
## [1] 4
y[c(1, 3, 10)]
```

```
## [1] 100  98  91
x[1:4]
## [1] 0 2 4 6
x[x > 91]
## [1]  92  94  96  98 100
x[x > 20 & x <= 30]
## [1] 22 24 26 28 30
y[y == 5 | y > 95 | y < 3]
## [1] 100  99  98  97  96   5   2   1
```

e)
```
x <- seq(from = 0.2, to = 2, by = 0.3)
y <- -3:3
x < y
## [1] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
x^2 < x
## [1]  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
Index <- x^2 < x
x[Index]
## [1] 0,2 0,5 0,8
y[Index]
## [1] -3 -2 -1
```

Aufgabe 4

Daten mit mehreren Merkmalen als Data Frames

R

Urlisten mit mehr als einem Merkmal kann man sinnvoll in einer Tabelle abbilden. Die Zeilen entsprechen den Objekten, die Spalten den Merkmalen. In R heißen solche Datenmatrizen `Data Frames`; diese können mit der Funktion `data.frame()` gebildet werden. Ein Beispiel: Es soll ein Data Frame erstellt werden, der aus 6 Studenten, deren Alter, den Matrikelnummern, einer Klausurnote und der Information, ob diese Klausur bestanden wurde besteht:

```
namen <- c("Arno", "Bert", "Carl", "Doro", "Edda", "Fred")
alter <- c(19, 21, 20, 22, 20, 27)
matrnr <- c(101010, 101007, 200000, 123456, 654321, 111111)
note <- c(1, 5, 2.3, 2.7, 1.3, 4)
bestanden <- (note < 4.3)
Studenten <- data.frame(Name = namen, Alter = alter, Matrikelnummer = matrnr,
    Note = note, Bestanden = bestanden)
Studenten

##   Name Alter Matrikelnummer Note Bestanden
## 1 Arno    19         101010  1,0      TRUE
## 2 Bert    21         101007  5,0     FALSE
## 3 Carl    20         200000  2,3      TRUE
## 4 Doro    22         123456  2,7      TRUE
## 5 Edda    20         654321  1,3      TRUE
## 6 Fred    27         111111  4,0      TRUE
```

Um auf einzelne Merkmale bzw. Objekte eines Data Frames zuzugreifen, kann der Zeilen- und Spaltenindex durch ein Komma getrennt in eckigen Klammern angegeben werden. Dabei können auch wie bei Vektoren mehrere Zeilen- bzw. Spaltenindexwerte angegeben werden, bzw. über Auswahloperatoren gebildet werden. Um auf ein Merkmal zuzugreifen, kann der Bezeichner des Merkmals mit einem `$`-Zeichen an den Bezeichner des dataframes angehängt werden.

a) Was gibt R jeweils nach folgenden Zeilen aus? Überprüfen Sie Ihre Antwort in R.

```
Studenten[1, 3]
Studenten[1:3, c(1, 4)]
Studenten[2, ]
Studenten[, 4]
Studenten$Note
Studenten[, 3:5]
Studenten[note < 4, ]
```

b) Erzeugen Sie einen Data Frame der folgenden 7 Kinder gemäß der Datentabelle

Alter	Geschlecht	Taschengeld (in Euro)	besitzt Fahrrad
6	männlich	12	ja
7	männlich	18	ja
6	weiblich	14	nein
7	weiblich	20	ja
8	männlich	26	ja
7	weiblich	20	ja
8	weiblich	20	nein

c) Generieren Sie in R einen Data Frame des Alters und des Taschengeldes aller Kinder mit höchstens 7 Jahren.

d) Berechnen Sie in R die Summe des Taschengelds aller Fahrradbesitzer.

Lösungshinweis:

a)
```
Studenten[1, 3]  # vom 1. Student aus der Liste den Wert des 3. Merkmals (M.Nr.)

## [1] 101010

Studenten[1:3, c(1, 4)]  # Studenten 1, 2, und 3, davon 1. und 4. Merkmal

##   Name Note
## 1 Arno  1,0
## 2 Bert  5,0
## 3 Carl  2,3

Studenten[2, ]  # vom 2. Student alle Merkmale

##   Name Alter Matrikelnummer Note Bestanden
## 2 Bert    21         101007    5     FALSE

Studenten[, 4]  # alle Studenten, nur das 4. Merkmal

## [1] 1,0 5,0 2,3 2,7 1,3 4,0

Studenten$Note  # dito

## [1] 1,0 5,0 2,3 2,7 1,3 4,0

Studenten[, 3:5]  # Alle Studenten, Merkmale 3, 4, 5

##   Matrikelnummer Note Bestanden
## 1         101010  1,0      TRUE
## 2         101007  5,0     FALSE
## 3         200000  2,3      TRUE
## 4         123456  2,7      TRUE
## 5         654321  1,3      TRUE
## 6         111111  4,0      TRUE

Studenten[note < 2.5, ]  # Alle Studenten mit Note besser 2.5
```

```
##   Name Alter Matrikelnummer Note Bestanden
## 1 Arno    19         101010  1,0      TRUE
## 3 Carl    20         200000  2,3      TRUE
## 5 Edda    20         654321  1,3      TRUE
```

b)
```
Kinder <- data.frame(Alter = c(6, 7, 6, 7, 8, 7, 8), Geschlecht = c("m",
    "m", "w", "w", "m", "w", "w"), Taschengeld = c(12, 18,
    14, 20, 26, 20, 20), Fahrrad = c("ja", "ja", "nein",
    "ja", "ja", "ja", "nein"))
```

c)
```
Kinder[Kinder$Alter <= 7, c("Alter", "Taschengeld")]

##   Alter Taschengeld
## 1     6          12
## 2     7          18
## 3     6          14
## 4     7          20
## 6     7          20
```

d)
```
sum(Kinder[Kinder$Fahrrad == "ja", "Taschengeld"])

## [1] 96
```

Aufgabe 5

Datentypen und Umgang mit Data Frames

R

R kann zwischen metrischen, ordinalen und nominalen Merkmalen unterscheiden. Das Skalenniveau eines Merkmals kann über die Funktion `str()` abgefragt werden. Um explizit ein bestimmtes Skalenniveau abzufragen, können die Funktionen `is.numeric()` für metrisches, `is.ordered()` für ordinales sowie `is.factor()` für nominale Merkmale verwendet werden.
Metrische Merkmale werden typischerweise automatisch von R erkannt, wenn gültige Zahlen eingelesen werden. Falls metrische Ausprägungen als Zeichenketten eingegeben wurden, können sie mittels `as.numeric()` umgewandelt werden.
Nominale Merkmale können mit der Funktion `factor()` angelegt, bzw. mit `as.factor()` umgewandelt werden.
Ordinale Merkmale sind in R auch factors, haben aber zusätzlich eine Rangfolge der Ausprägungen hinterlegt. Zum Anlegen eines solchen Merkmals wird `ordered()`, zum umwandeln `as.ordered()` verwendet.

a) Lesen Sie die Hilfe-Seiten zu `factor()` und und entscheiden Sie dann, was die folgenden Zeilen in R bewirken. Überprüfen Sie Ihre Lösung mit R.

```
Merkmal1 <- c(1, 2, 3, 2, 1, 2, 3)
Merkmal2 <- c("2.1", 1, 3, "-3e2", 2000, "-0.2", "-.3")
Merkmal3 <- c("gut", "gut", "katastrophal", "mittel", "katastrophal",
    "gut", "mittel")
is.numeric(Merkmal1)
is.numeric(Merkmal2)
Merkmal2 <- as.numeric(Merkmal2)
Merkmal2
is.numeric(Merkmal2)
is.factor(Merkmal3)
Merkmal3 <- factor(Merkmal3)
is.factor(Merkmal3)
Merkmal3
Merkmal3 <- ordered(Merkmal3)
Merkmal3
Merkmal3 <- ordered(Merkmal3, levels = c("katastrophal",
    "mittel", "gut"))
Merkmal3
```

b) Gegeben ist der folgende Data Frame „`Studenten`“:

```
namen <- c("Arno", "Bert", "Carl", "Doro", "Edda", "Franz")
geschlecht <- c(0, 0, 0, 1, 1, 0)
alter <- c(19, 21, 20, 22, 20, 27)
note <- c("sehr gut", "durchgefallen", "gut", "gut", "sehr gut",
    "gut")
Studenten <- data.frame(Name = namen, Alter = alter, Geschlecht = geschlecht,
    Note = note)
```

Kodieren Sie in R den Datentyp der jeweiligen Merkmale richtig (metrisch, nominal, ordinal).

c) Transformieren Sie die Ausprägungen des Merkmals `Geschlecht` von den Originalwerten `0, 1` in die Werte `Mann` bzw. `Frau`. Benutzen Sie dazu den `labels` Parameter der Funktion `factor()`.

In einem Data Frame kann die Anzahl der Objekte mit `nrow()`, die Anzahl der Merkmale mit `ncol()` ausgegeben werden. Dient ein Merkmal lediglich als eindeutiger Bezeichner für die Objekte des Data Frames (z.B. Name oder Personalnummer) werden diese Bezeichner den Objekten üblicherweise mit der Funktion `row.names()` über die Zuweisung `row.names() <- bezeichner` zugewiesen. Möchte man bestimmte Merkmale eines Data Frames löschen, kann man ihnen `NULL` zuweisen. Sortiert wird ein Data Frame nach den Ausprägungen eines bestimmten Merkmals über die Funktion `order()`. Umgedreht wird die Reihenfolge im Aufruf von `order()` mit dem Parameter `decreasing=TRUE`.

d) Wie ist das Ergebnis folgender Anweisungen? Überprüfen Sie Ihr Ergebnis mit R.

```
nrow(Studenten)
ncol(Studenten)
row.names(Studenten) <- Studenten$Name
Studenten
Studenten$Name <- NULL
Studenten
Studenten[order(Studenten$Alter), ]
Studenten[order(Studenten$Note), ]
Studenten[order(Studenten$Note, decreasing = TRUE), ]
```

Weitere Objekte werden einem Data Frame mit `rbind()` hinzugefügt, mit `cbind()` erweitert man Data Frames um weitere Merkmale.

e) Erweitern Sie den Datensatz `Studenten` um die 26-jährige Studentin Gerlinde, die in der Prüfung sehr gut abgeschnitten hat.

Lösungshinweis:

a)
```
Merkmal1 <- c(1, 2, 3, 2, 1, 2, 3)
Merkmal2 <- c("2.1", 1, 3, "-3e2", 2000, "-0.2", "-.3")
Merkmal3 <- c("gut", "gut", "katastrophal", "mittel", "katastrophal",
    "gut", "mittel")
is.numeric(Merkmal1)

## [1] TRUE

is.numeric(Merkmal2)

## [1] FALSE

Merkmal2 <- as.numeric(Merkmal2)
Merkmal2

## [1]    2,1    1,0    3,0 -300,0 2000,0   -0,2   -0,3
```

```r
is.numeric(Merkmal2)

## [1] TRUE

is.factor(Merkmal3)

## [1] FALSE

Merkmal3 <- factor(Merkmal3)
is.factor(Merkmal3)

## [1] TRUE

Merkmal3

## [1] gut          gut          katastrophal mittel
## [5] katastrophal gut          mittel
## Levels: gut katastrophal mittel

Merkmal3 <- ordered(Merkmal3)
Merkmal3

## [1] gut          gut          katastrophal mittel
## [5] katastrophal gut          mittel
## Levels: gut < katastrophal < mittel

Merkmal3 <- ordered(Merkmal3, levels = c("katastrophal",
    "mittel", "gut"))
Merkmal3

## [1] gut          gut          katastrophal mittel
## [5] katastrophal gut          mittel
## Levels: katastrophal < mittel < gut
```

b)
```r
str(Studenten)  # Name und Alter sind OK, Geschlecht und Note nicht

## 'data.frame': 6 obs. of  4 variables:
##  $ Name      : Factor w/ 6 levels "Arno","Bert",..: 1 2 3 4 5 6
##  $ Alter     : num  19 21 20 22 20 27
##  $ Geschlecht: num  0 0 0 1 1 0
##  $ Note      : Factor w/ 3 levels "durchgefallen",..: 3 1 2 2 3 2

Studenten$Geschlecht <- factor(Studenten$Geschlecht)
Studenten$Note <- ordered(Studenten$Note, levels = c("durchgefallen",
    "gut", "sehr gut"))
Studenten$Note  # jetzt stimmt die Reihenfolge

## [1] sehr gut      durchgefallen gut
## [4] gut           sehr gut      gut
## Levels: durchgefallen < gut < sehr gut

str(Studenten)

## 'data.frame': 6 obs. of  4 variables:
##  $ Name      : Factor w/ 6 levels "Arno","Bert",..: 1 2 3 4 5 6
##  $ Alter     : num  19 21 20 22 20 27
##  $ Geschlecht: Factor w/ 2 levels "0","1": 1 1 1 2 2 1
##  $ Note      : Ord.factor w/ 3 levels "durchgefallen"<..: 3 1 2 2 3 2
```

c)
```r
Studenten
```
```
##    Name Alter Geschlecht         Note
## 1  Arno    19          0     sehr gut
## 2  Bert    21          0 durchgefallen
## 3  Carl    20          0          gut
## 4  Doro    22          1          gut
## 5  Edda    20          1     sehr gut
## 6 Franz    27          0          gut
```
```r
Studenten$Geschlecht <- factor(Studenten$Geschlecht, labels = c("Mann",
    "Frau"))
Studenten
```
```
##    Name Alter Geschlecht         Note
## 1  Arno    19       Mann     sehr gut
## 2  Bert    21       Mann durchgefallen
## 3  Carl    20       Mann          gut
## 4  Doro    22       Frau          gut
## 5  Edda    20       Frau     sehr gut
## 6 Franz    27       Mann          gut
```

d)
```r
nrow(Studenten)
```
```
## [1] 6
```
```r
ncol(Studenten)
```
```
## [1] 4
```
```r
row.names(Studenten) <- Studenten$Name
Studenten
```
```
##        Name Alter Geschlecht         Note
## Arno   Arno    19       Mann     sehr gut
## Bert   Bert    21       Mann durchgefallen
## Carl   Carl    20       Mann          gut
## Doro   Doro    22       Frau          gut
## Edda   Edda    20       Frau     sehr gut
## Franz Franz    27       Mann          gut
```
```r
Studenten$Name <- NULL
Studenten
```
```
##       Alter Geschlecht         Note
## Arno     19       Mann     sehr gut
## Bert     21       Mann durchgefallen
## Carl     20       Mann          gut
## Doro     22       Frau          gut
## Edda     20       Frau     sehr gut
## Franz    27       Mann          gut
```
```r
Studenten[order(Studenten$Alter), ]
```
```
##       Alter Geschlecht         Note
## Arno     19       Mann     sehr gut
## Carl     20       Mann          gut
## Edda     20       Frau     sehr gut
```

```
## Bert      21        Mann durchgefallen
## Doro      22        Frau            gut
## Franz     27        Mann            gut
```

```
Studenten[order(Studenten$Note), ]
```

```
##       Alter Geschlecht          Note
## Bert     21       Mann durchgefallen
## Carl     20       Mann           gut
## Doro     22       Frau           gut
## Franz    27       Mann           gut
## Arno     19       Mann      sehr gut
## Edda     20       Frau      sehr gut
```

```
Studenten[order(Studenten$Note, decreasing = TRUE), ]
```

```
##       Alter Geschlecht          Note
## Arno     19       Mann      sehr gut
## Edda     20       Frau      sehr gut
## Carl     20       Mann           gut
## Doro     22       Frau           gut
## Franz    27       Mann           gut
## Bert     21       Mann durchgefallen
```

e)

```
Studenten <- rbind(Studenten, data.frame(Alter = 26, Geschlecht = "Frau",
    Note = "sehr gut", row.names = "Gerlinde"))
Studenten
```

```
##          Alter Geschlecht          Note
## Arno        19       Mann      sehr gut
## Bert        21       Mann durchgefallen
## Carl        20       Mann           gut
## Doro        22       Frau           gut
## Edda        20       Frau      sehr gut
## Franz       27       Mann           gut
## Gerlinde    26       Frau      sehr gut
```

Aufgabe 6

Einlesen von Daten aus Textdateien

R

Wenn Daten als Tabelle in einer Textdatei vorliegen können sie leicht in R eingelesen werden; dazu ist es gut, wenn in der Tabelle die Zeilen den Objekten und die Spalten den Merkmalen entsprechen. Über die graphische Benutzeroberfläche in RStudio funktioniert das Importieren einfach über den Knopf „`Import Dataset`“ im Reiter `Environment`. Dabei kann eine lokale Datei ausgewählt werden oder eine Datei, die über das Netzwerk via URL erreichbar ist. Klickt man auf `Import` werden Die Daten als Data Frame unter dem angegebenen Bezeichner zum Beispiel mit der Funktion `read.csv()` eingelesen.

a) Importieren Sie mit der graphischen Benutzeroberfläche von RStudio die Umfragedaten aus der Vorlesung von der Adresse `http://goo.gl/Mg6kmj` und speichern Sie den Data Frame unter der Bezeichnung „`Umfrage`“.
b) Importieren Sie die Daten nochmals, diesmal aber indem Sie eine lokale Kopie der csv-Datei auf Ihrer Festplatte anlegen und das Einlesen über die Funktion `read.csv()` bewerkstelligen.

Die Funktion `head()` zeigt die ersten Objekte eines Data Frames an, mit `summary()` bekommt man einen Überblick über die Verteilung der Ausprägungen in den einzelnen Merkmalen.

c) Verschaffen Sie sich einen Überblick über die Daten, indem Sie sich die Struktur des Data Frames mit `str()`, die ersten Objekte mit `head()` und die Verteilung der Ausprägungen in den einzelnen Merkmalen mit `summary()` ansehen.

Lösungshinweis:

a)
```
# Klicken auf Import Dataset, Angabe der URL, Ändern des
# Bezeichners Ändern des Dezimaltrenners auf ','
Umfrage <- read.csv("http://goo.gl/Mg6kmj", sep = ";", dec = ",")
```

b)
```
Umfrage <- read.csv("c:/Verzeichnis/Umfrage_HSA_2015_03.csv",
    sep = ";", dec = ",")
```

c)
```
str(Umfrage)

## 'data.frame': 377 obs. of  17 variables:
##  $ Jahrgang       : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
##  $ Alter          : int  20 25 19 21 25 20 25 20 23 21 ...
##  $ Groesse        : int  174 157 163 185 178 170 165 175 180 161 ...
##  $ Geschlecht     : Factor w/ 2 levels "Frau","Mann": 1 1 1 2 2 1 1 2 2 1 ...
##  $ AlterV         : int  55 54 51 52 60 50 60 52 56 70 ...
##  $ AlterM         : int  53 61 49 50 63 55 60 49 50 55 ...
```

```
##  $ GroesseV       : int  187 185 178 183 170 183 185 175 175 180 ...
##  $ GroesseM       : int  169 160 168 165 160 160 170 169 170 165 ...
##  $ Geschwister    : int  3 1 1 4 2 2 4 1 1 2 ...
##  $ Farbe          : Factor w/ 6 levels "blau","gelb",..: 4 6 4 4 1 6 1 6 4 4 ...
##  $ AusgKomm       : num  240 119 270 40 550 ...
##  $ AnzSchuhe      : int  25 30 25 6 5 65 10 7 10 22 ...
##  $ AusgSchuhe     : int  450 300 100 100 80 250 150 400 150 300 ...
##  $ Essgewohnheiten: Factor w/ 6 levels "","carnivor",..: 2 2 2 2 2 2 6 2 2 2 ...
##  $ Raucher        : Factor w/ 3 levels "","ja","nein": 1 3 3 3 2 3 3 3 3 2 ...
##  $ NoteMathe      : num  2,3 3,3 1,7 2,0 4,0 4,0 3,3 2,7 3,7 3,3 ...
##  $ MatheZufr      : Factor w/ 5 levels "","geht so","sehr zufrieden",..: 2 2 2 2 2 2 2 2 2
```

```
head(Umfrage)
```

```
##   Jahrgang Alter Groesse Geschlecht AlterV AlterM GroesseV GroesseM
## 1     2015    20     174       Frau     55     53      187      169
## 2     2015    25     157       Frau     54     61      185      160
## 3     2015    19     163       Frau     51     49      178      168
## 4     2015    21     185       Mann     52     50      183      165
## 5     2015    25     178       Mann     60     63      170      160
## 6     2015    20     170       Frau     50     55      183      160
##   Geschwister   Farbe AusgKomm AnzSchuhe AusgSchuhe Essgewohnheiten
## 1           3 schwarz    240,0        25        450        carnivor
## 2           1   weiss    119,4        30        300        carnivor
## 3           1 schwarz    270,0        25        100        carnivor
## 4           4 schwarz     40,0         6        100        carnivor
## 5           2    blau    550,0         5         80        carnivor
## 6           2   weiss    420,0        65        250        carnivor
##   Raucher NoteMathe MatheZufr
## 1               2,3   geht so
## 2    nein       3,3   geht so
## 3    nein       1,7   geht so
## 4    nein       2,0   geht so
## 5      ja       4,0   geht so
## 6    nein       4,0   geht so
```

```
summary(Umfrage)
```

```
##     Jahrgang         Alter        Groesse     Geschlecht     AlterV
##  Min.   :2014   Min.   :18   Min.   :151   Frau:250   Min.   :38,0
##  1st Qu.:2014   1st Qu.:20   1st Qu.:165   Mann:127   1st Qu.:50,0
##  Median :2014   Median :21   Median :170              Median :54,0
##  Mean   :2014   Mean   :22   Mean   :172              Mean   :54,1
##  3rd Qu.:2015   3rd Qu.:23   3rd Qu.:178              3rd Qu.:57,0
##  Max.   :2015   Max.   :36   Max.   :197              Max.   :77,0
##
##      AlterM         GroesseV        GroesseM     Geschwister
##  Min.   :37,0   Min.   :160   Min.   : 76   Min.   :0,00
##  1st Qu.:49,0   1st Qu.:175   1st Qu.:162   1st Qu.:1,00
##  Median :51,0   Median :180   Median :165   Median :1,00
##  Mean   :51,6   Mean   :179   Mean   :166   Mean   :1,49
##  3rd Qu.:54,0   3rd Qu.:183   3rd Qu.:170   3rd Qu.:2,00
##  Max.   :68,0   Max.   :202   Max.   :192   Max.   :9,00
##
```

```
##      Farbe         AusgKomm       AnzSchuhe       AusgSchuhe
##  blau   : 21   Min.   :   0   Min.   :  2,0   Min.   :   0
##  gelb   :  4   1st Qu.: 200   1st Qu.: 10,0   1st Qu.: 130
##  rot    : 19   Median : 350   Median : 20,0   Median : 200
##  schwarz:176   Mean   : 409   Mean   : 22,2   Mean   : 271
##  silber : 36   3rd Qu.: 540   3rd Qu.: 30,0   3rd Qu.: 350
##  weiss  :121   Max.   :1900   Max.   :275,0   Max.   :2000
##                NA's   :1
##      Essgewohnheiten Raucher      NoteMathe             MatheZufr
##              :205        :208   Min.   :1,00                 : 56
##  carnivor    :157    ja  : 21   1st Qu.:2,30   geht so       : 81
##  fruktarisch :  1    nein:148   Median :3,30   sehr zufrieden: 43
##  pescetarisch:  6               Mean   :3,28   unzufrieden   :129
##  vegan       :  1               3rd Qu.:4,00   zufrieden     : 68
##  vegetarisch :  7               Max.   :5,00
##                                 NA's   :73
```

Aufgabe 7

R-Skripten: Führen eines Logbuches der eigenen Analysen

R

In vielen Fällen besteht eine statistische Untersuchung aus mehr als einem Schritt. Meistens werden Daten eingelesen, bereinigt, aufbereitet, verdichtet, graphisch dargestellt usw. Um diesen Ablauf zu dokumentieren kann man eine Textdatei mit der Endung `.R`, ein sogenanntes `R`-Skript erstellen und alle Kommandos dort ablegen, mit Kommentaren dokumentieren und für spätere Wiederverwendung abspeichern.

a) Legen Sie eine `.R`-Datei mit dem Bezeichner `Statistik-Uebung.R` an (in RStudio über `File -> New -> R-Script`) und schreiben Sie in diese Datei in die ersten Zeilen als Kommentar (`#-Zeichen voranstellen`) Ihren Namen, das Datum sowie eine Anmerkung, dass diese Datei alle `R`-Lösungen der Übungsaufgaben enthält.

b) Fügen Sie für jede bis hierher bearbeitete Aufgabe nach einem entsprechenden Kommentar den jeweiligen R-Code in diese Datei ein und schreiben Sie zu möglichst vielen Zeilen einen Kommentar, in dem Sie eine Anmerkung schreiben was die Zeile bewirkt.

Um eine Zeile aus einem `R`-Skript in `R` auszuführen, kann der Cursor in die entsprechende Zeile platziert werden; durch die Tastenkombination `Strg-Enter` (auf englischsprachigen Tastaturen `Ctrl-Enter`) wird die Zeile in die Console kopiert und ausgeführt; danach springt der Cursor in die nächste Zeile des Skripts. Wiederholt man das mehrmals, werden der Reihe nach alle Zeilen ausgeführt (Kommentarzeilen werden übergangen). Möchte man mehr als eine Zeile ausführen, kann man den entsprechenden Teil des Skripts mit der Maus markieren und mit `Strg-Enter` ausführen.

c) Führen Sie die Befehle der ersten R-Aufgabe zunächst zeilenweise aus und beobachten Sie die Ein- und Ausgaben in der Console, danach markieren Sie die komplette Aufgabe und wiederholen die Ausführung.

Lösungshinweis:

a)
```
# ------------------------------------------------------- 22.3.2016,
# Max Maier R-Skript zur Statistik Übung im SS 2016
# -------------------------------------------------------
```

b)
```
# ------------------------------------------------------- 22.3.2016,
# Max Maier R-Skript zur Statistik Übung im SS 2016
# -------------------------------------------------------
# Aufgabe 1
2 + 3 * 4  # hier gilt Punkt vor Strich
(2 + 3) * 4  # Klammer zuerst
0.2 * 3 - 1.1
# 0,2 * 3 # Fehler, ',' wird nicht als Dezimalkomma akzeptiert
```

```r
2^3^2  # entspricht 2^(3^2)
(2^3)^2
exp(1)  # das ist e^1
log(exp(1))  # log() entspricht dem ln; e^x und ln heben sich auf
sqrt(16)  # Quadratwurzel
16^(1/2)  # auch QW
# Sqrt(16) # Fehler, sqrt() schreibt man mit kleinem 's'
# Aufgabe 2 ...
```

c)

```r
# -------------------------------------------------------- 22.3.2016,
# Max Maier R-Skript zur Statistik Übung im SS 2016
# --------------------------------------------------------
# Aufgabe 1
2 + 3 * 4  # hier gilt Punkt vor Strich

## [1] 14

(2 + 3) * 4  # Klammer zuerst

## [1] 20

0.2 * 3 - 1.1

## [1] -0,5

# 0,2 * 3 # Fehler, ',' wird nicht als Dezimalkomma akzeptiert
2^3^2  # entspricht 2^(3^2)

## [1] 512

(2^3)^2

## [1] 64

exp(1)  # das ist e^1

## [1] 2,7183

log(exp(1))  # log() entspricht dem ln; e^x und ln heben sich auf

## [1] 1

sqrt(16)  # Quadratwurzel

## [1] 4

16^(1/2)  # auch QW

## [1] 4

# Sqrt(16) # Fehler, sqrt() schreibt man mit kleinem 's'
# Aufgabe 2 ...
```

Aufgabe 8

Häufigkeiten in R und Umgang mit fehlenden Werten

R

Häufigkeitsauszählungen können in `R` mit `table()` erstellt werden. `cumsum()` bildet kumulierte Summen. `mean()` berechnet das arithmetische Mittel, `median()` den Median

a) Lesen Sie von `http://goo.gl/dZkICg` die Daten der Vorlesungsumfrage ein und bilden Sie jeweils eine Tabelle der absoluten und relativen sowie der absoluten kumulierten und relativen kumulierten Häufigkeiten des Merkmals Alter.

b) Wandeln Sie das Merkmal `MatheZufr` (Antwort auf „Waren Sie zufrieden mit Ihrer Leistung in der Matheklausur) in ein ordinales Merkmal mit sinnvoller Reihenfolge um.

c) Berechnen Sie den Median und das arithmetische Mittel aller metrischen Merkmale der eingelesenen Daten.

Ausprägungen fehlender Werte werden in R mit `NA` (Not Available) dargestellt, Objekte mit fehlenden Werten können mittels `na.omit()` gelöscht werden. Die Funktion `sort()` gibt einen metrisch oder ordinal skalierten Vektor in aufsteigender Reihenfolge zurück.

d) Für ordinale Merkmale ist der Median zwar definiert, mit der eingebauten Funktion in `R` erhält man aber eine Fehlermeldung. Lösen Sie das Problem und berechnen Sie den Median aller (vorhandenen) Ausprägungen des Merkmals `MatheZufr`.

Lösungshinweis:

a)
```
# Umfragedaten einlesen
Umfrage <- read.csv("http://goo.gl/yMeyJp", sep = ";", dec = ",")
```

```
T = table(Umfrage$Alter)   # absolute Häufigkeiten
n = length(Umfrage$Alter)  # Anzahl der Objekte
T

##
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
## 24 58 71 53 43 40 25 11 11  8 12  7  2  2  5  2  1  1  1

cumsum(T)                  # kumuliert

##  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34
##  24  82 153 206 249 289 314 325 336 344 356 363 365 367 372 374 375
##  35  36
## 376 377

T/n                        # relative Häufigkeiten

##
##         18         19         20         21         22         23          24
```

```
## 0,0636605 0,1538462 0,1883289 0,1405836 0,1140584 0,1061008 0,0663130
##        25        26        27        28        29        30        31
## 0,0291777 0,0291777 0,0212202 0,0318302 0,0185676 0,0053050 0,0053050
##        32        33        34        35        36
## 0,0132626 0,0053050 0,0026525 0,0026525 0,0026525
```

```
round(T/n, 3)             # auf drei Kommastellen gerundet
```

```
##
##    18    19    20    21    22    23    24    25    26    27    28
## 0,064 0,154 0,188 0,141 0,114 0,106 0,066 0,029 0,029 0,021 0,032
##    29    30    31    32    33    34    35    36
## 0,019 0,005 0,005 0,013 0,005 0,003 0,003 0,003
```

```
round(cumsum(T)/n, 3)     # kumulierte rel. Häufigkeiten
```

```
##    18    19    20    21    22    23    24    25    26    27    28
## 0,064 0,218 0,406 0,546 0,660 0,767 0,833 0,862 0,891 0,912 0,944
##    29    30    31    32    33    34    35    36
## 0,963 0,968 0,973 0,987 0,992 0,995 0,997 1,000
```

```
cbind(as.data.frame(T),
      as.data.frame(cumsum(T)),
      as.data.frame(T/n),
      as.data.frame(cumsum(T)/n)
      )
```

```
##    Var1 Freq cumsum(T) Var1      Freq cumsum(T)/n
## 18   18   24        24   18 0,0636605     0,06366
## 19   19   58        82   19 0,1538462     0,21751
## 20   20   71       153   20 0,1883289     0,40584
## 21   21   53       206   21 0,1405836     0,54642
## 22   22   43       249   22 0,1140584     0,66048
## 23   23   40       289   23 0,1061008     0,76658
## 24   24   25       314   24 0,0663130     0,83289
## 25   25   11       325   25 0,0291777     0,86207
## 26   26   11       336   26 0,0291777     0,89125
## 27   27    8       344   27 0,0212202     0,91247
## 28   28   12       356   28 0,0318302     0,94430
## 29   29    7       363   29 0,0185676     0,96286
## 30   30    2       365   30 0,0053050     0,96817
## 31   31    2       367   31 0,0053050     0,97347
## 32   32    5       372   32 0,0132626     0,98674
## 33   33    2       374   33 0,0053050     0,99204
## 34   34    1       375   34 0,0026525     0,99469
## 35   35    1       376   35 0,0026525     0,99735
## 36   36    1       377   36 0,0026525     1,00000
```

b)

```
Umfrage$MatheZufr <- ordered(Umfrage$MatheZufr, levels = c("nicht", "geht so",
    "zufrieden", "sehr"))
Mathe.sortiert <- sort(Umfrage$MatheZufr)
n <- length(Mathe.sortiert)
is.integer(n/2)   # FALSE, also n ungerade
```

```
## [1] FALSE
```

```
Mathe.Median <- Mathe.sortiert[(n + 1)/2]
Mathe.Median

## [1] geht so
## Levels: nicht < geht so < zufrieden < sehr
```

Aufgabe 9

Grafiken

R

Grafiken kann man in R mit `plot()` erzeugen. `plot` versucht anhand der übergebenen Merkmalstypen aus dem Kontext zu entscheiden, welche Art von Grafik erzeugt werden soll:

a) Probieren Sie folgende plots in R aus:

```
# Umfragedaten einlesen
Daten <- read.csv("http://goo.gl/yMeyJp", sep = ";", dec = ",")
```

```
# plot mit einem nominalen Merkmal als Argument zeigt
# ein Balkendiagramm
plot(Daten$Farbe)
# Farben können als optionales Argument angegeben werden
# Ausrichtung der Beschriftung über las
plot(Daten$Farbe, col = c("blue", "yellow", "red", "black",
    "grey", "white"), las = 2)
# zwei nominale Merkmale werden als 'spineplot'
# gegenübergestellt
plot(Daten$Farbe, Daten$Geschlecht)
# ein metrisches und ein nominales Merkmal wird als
# Liste von Boxplots gezeichnet
plot(Daten$Geschlecht, Daten$Alter)
# zwei metrische Merkmale verarbeitet plot() als
# Streuplot
plot(Daten$AlterV, Daten$AlterM)
# mehr als zwei metrische Merkmale führen zu einer
# Streuplotmatrix:
plot(Daten[, c("Alter", "AlterM", "AlterV")])
```

`plot()` versteht je nach Kontext viele Parameter zur Anpassung der Grafiken. Die Liste aller Grafikeinstellungen von `plot()` und deren möglichen Werte findet man unter `?par`. Einige wichtige Parameter sind:

- *`main`: Überschrift der Grafik*
- *`xlab, ylab`: Beschriftung der Abszisse bzw. Ordinate*
- *`pch, col, cex`: ein Wert oder ein Vektor, der das Symbol, das gezeichnet wird, beinflusst; dabei steht `pch` für die Form des Symbols, `col` für die Farbe und `cex` für die Größe des Symbols.*

Mit der Funktion `rgb()` kann man beliebige Farben mischen und deren Transparenz festlegen.

b) Probieren Sie folgende Ausdrücke in R aus und experimentieren Sie mit Änderungen der Parameter

```
# Speichere zunächst 2 Farben
# rgb(): "r"ot, "g"rün, "b"lau;
```

```
# Argumente 1  bis 3: Anteile an rot, grün, blau
# 4. Argument: Transparenz (1 ist undurchsichtig, 0 ist komplett transparent)
Farben = c(rgb(1, 0, 0, 0.2), rgb(0, 0, 1, 0.2))
Symbole = c(20,18) # Kreis und Raute

plot(Daten$AnzSchuhe, Daten$AusgSchuhe,
     col=Farben[Daten$Geschlecht],  # farbliche Markierung über Geschlecht
     pch=Symbole[Daten$Geschlecht], # unterschiedliche Symbole je nach Geschlecht
     xlab="Wieviele Paar Schuhe besitzen Sie?", # Beschriftung der Abszisse
     ylab="Ausgaben in den letzten 12 Monaten für Schuhe") # Ordinate

plot(Daten[,c("Alter", "AlterV", "AlterM")],
     col=Farben[Daten$Geschlecht],  # farbliche Markierung über Geschlecht
     pch=Symbole[Daten$Geschlecht]) # unterschiedliche Symbole je nach Geschlecht
```

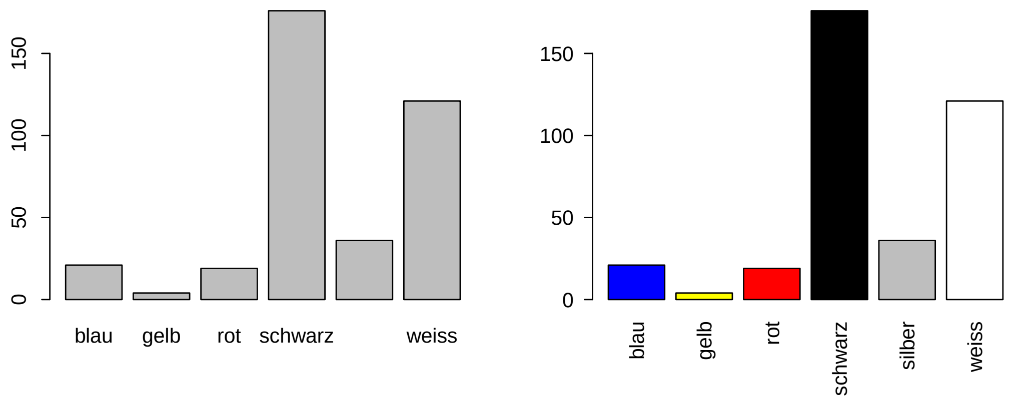
```
# Alternative zu spineplot: Mosaikplot
mosaicplot(Daten$Farbe ~ Daten$Geschlecht,
           shade=TRUE, # farbige Markierung
           las=2, # Ausrichtung der Achsenbeschriftung
           xlab="Geschlecht",
           ylab="Wunschfarbe", main="")
```

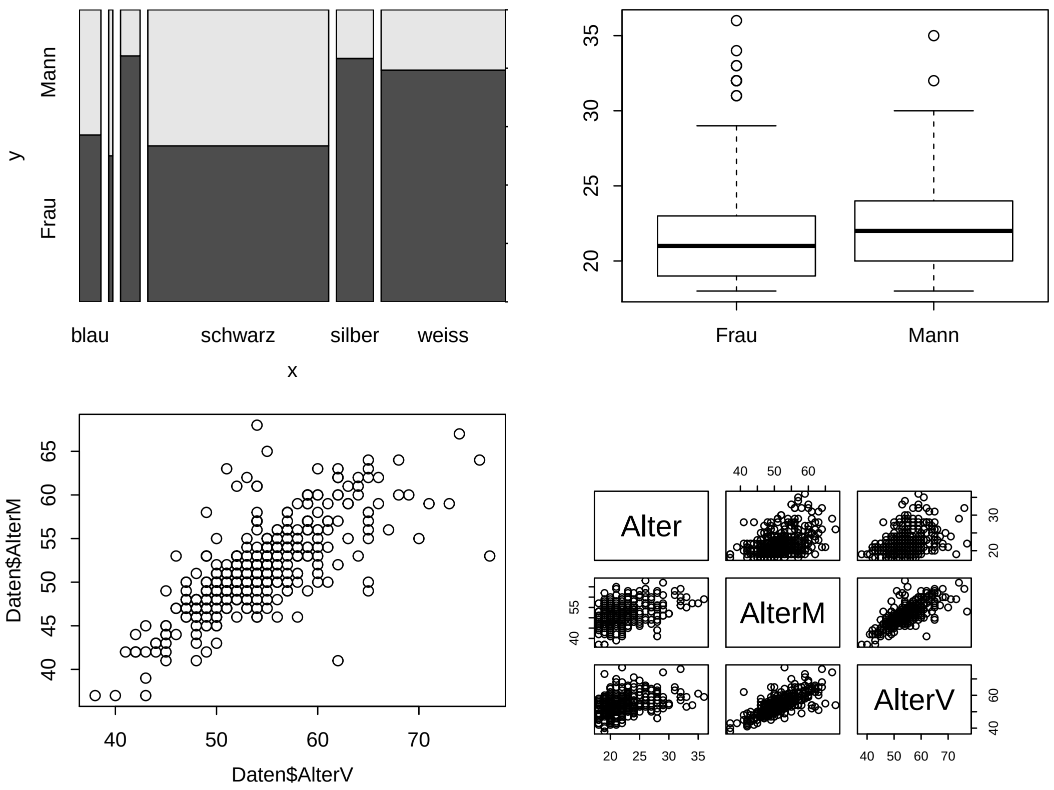
```
# Boxplot, horizontale Ausrichtung, eingefärbt
plot(Daten$Geschlecht, Daten$Alter, col = rgb(0, 0, 1, 0.2),
    horizontal = TRUE, las = 2, lwd = 1.5, pch = 20)
```

c) Generieren Sie einen Streuplot, in dem die Ausgaben für Schuhe und die Anzahl der Schuhe gegenübergestellt werden. Trennen Sie die Farben nach dem Merkmal „Farbe“ und färben Sie mit der jeweils in der Umfrage genannten Lieblingsfarbe ein, jeweils mit 30% Transparenz. Zeichnen Sie für die Frauen ausgefüllte Dreiecke und für die Männer ausgefüllte Quadrate.

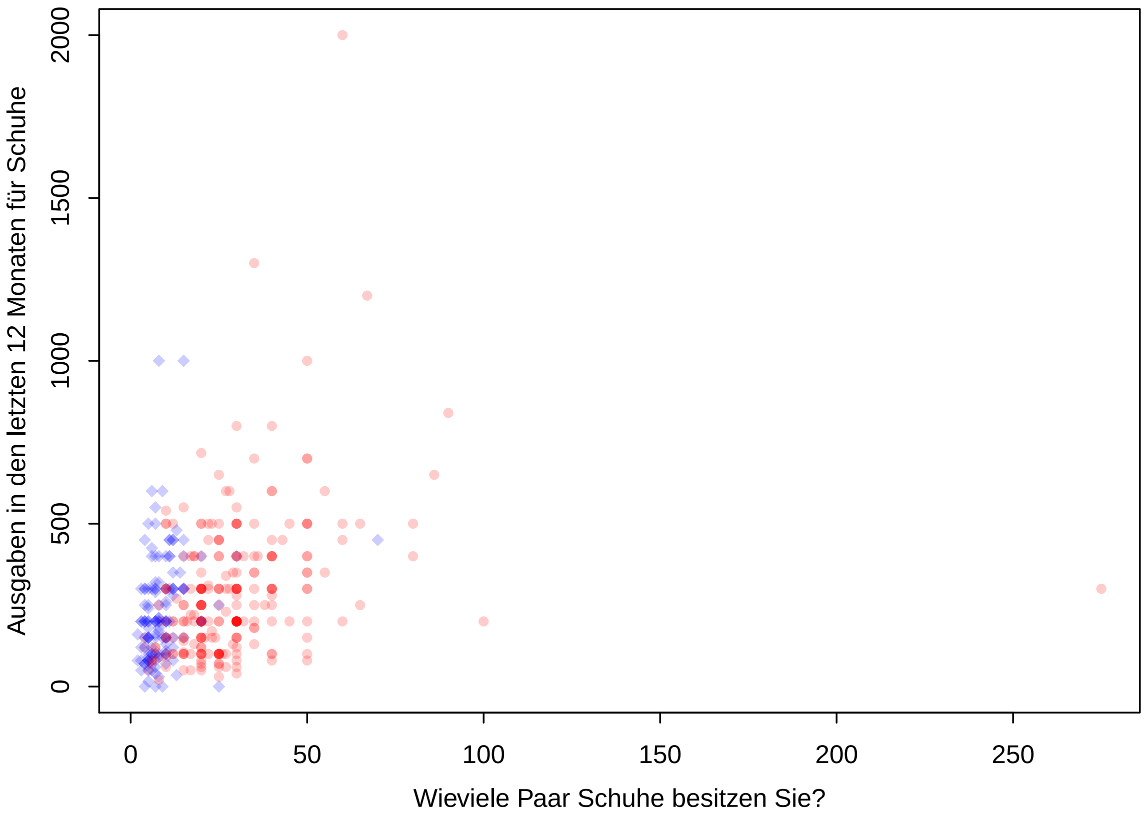
Lösungshinweis:

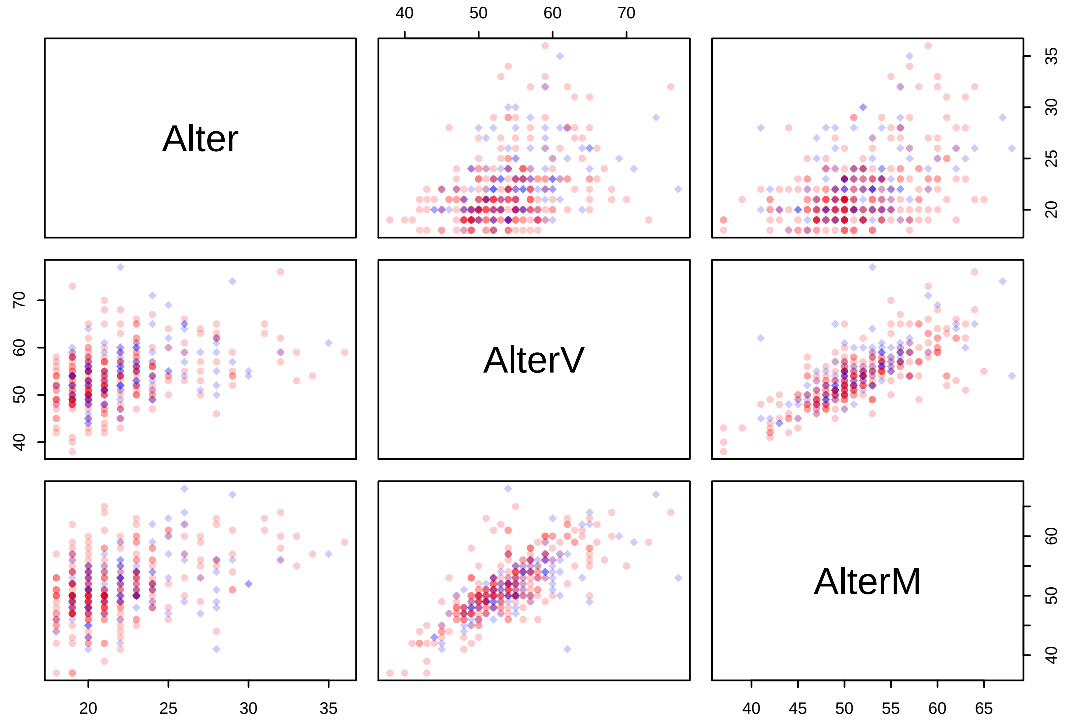
a) Folgende Grafiken ergeben sich:

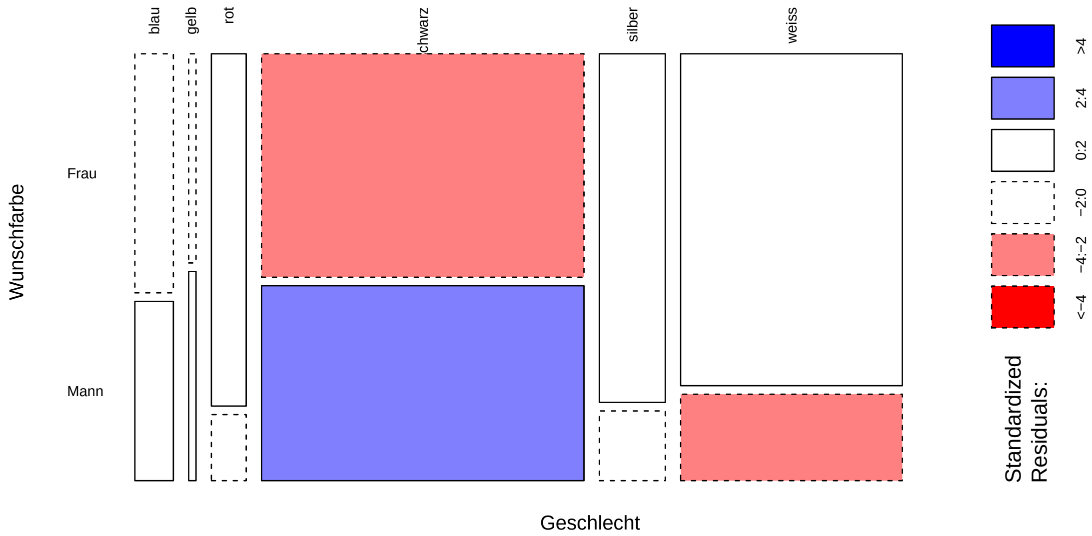


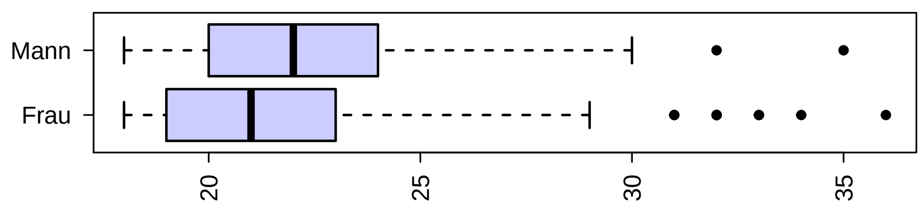

y
Mann
Frau
blau
schwarz
silber
weiss
x
Frau
Mann
Daten$AlterM
Daten$AlterV
Alter
AlterM
AlterV

b) Resultate:

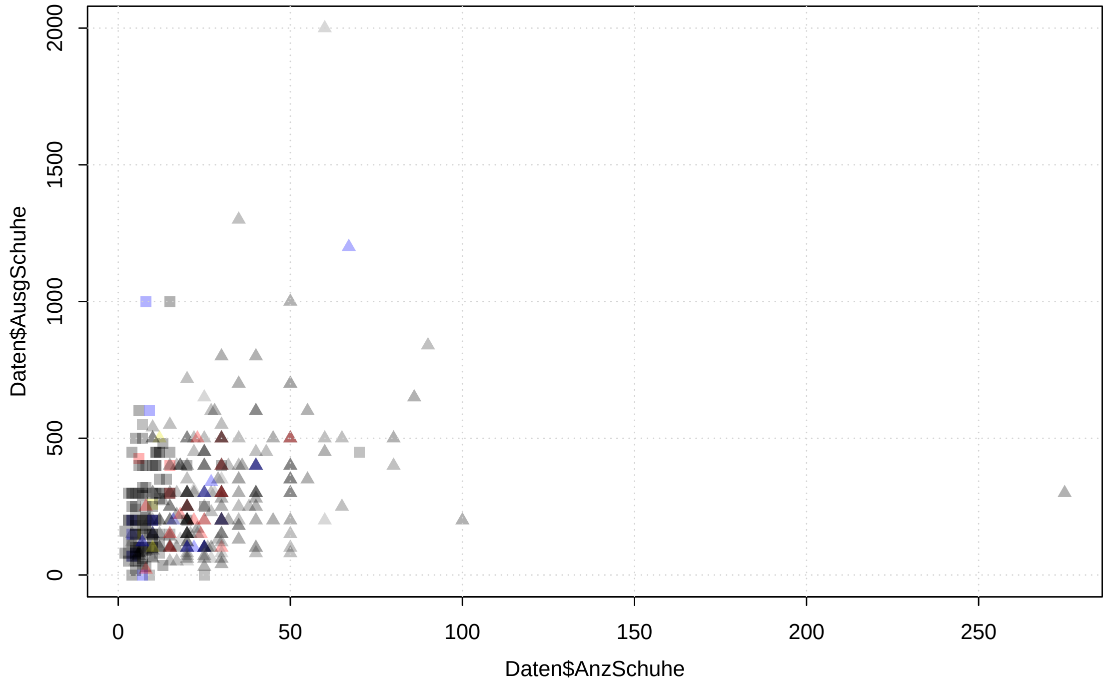

Ausgaben in den letzten 12 Monaten für Schuhe
Wieviele Paar Schuhe besitzen Sie?


Alter
AlterV
AlterM
40
50
60
70
20
25
30
35
45
55
65


blau
gelb
rot
schwarz
silber
weiss
Wunschfarbe
Frau
Mann
Geschlecht
>4
2:4
0:2
−2:0
−4:−2
<−4
Standardized Residuals:


Mann
Frau
20
25
30
35

c)
```r
levels(Daten$Farbe)

## [1] "blau"    "gelb"    "rot"     "schwarz" "silber"
## [6] "weiss"

Farben=c(rgb(0,0,1,0.3), # blau
         rgb(.9,.9,0,0.3), # gelb
         rgb(1,0,0,0.3), # rot
         rgb(0,0,0,0.3), # schwarz
         rgb(0.5,0.5,0.5,0.3), # grau (silber),
         rgb(0.2,0.2,0.2,0.3)) # (weiss) hier schwierig, also: helleres grau
Symbole=c(17,15)
plot(Daten$AnzSchuhe, Daten$AusgSchuhe, col=Farben[Daten$Farbe],
     pch=Symbole[Daten$Geschlecht])
grid()
```



Aufgabe 10

R

Empirische Verteilungsfunktion, Quantile, Boxplot

mit der Funktion `ecdf()` wird eine empirische Verteilungsfunktion erstellt. Wird das Ergebnis von `ecdf()` in die Funktion `plot()` eingesetzt erhält man eine grafische Darstellung dieser Verteilungsfunktion. Die Funktion `quantile()` liefert empirische Quantile einer Urliste. Dabei gibt es verschiedene Varianten, wie diese Quantile berechnet werden können; mit dem Parameter `type` kann man diese unterschiedlichen Definitionen ansprechen. In der Vorlesung wurde `type = 2` definiert.

Gegeben ist die folgende Urliste:

```
x <- c(1, 2, 1, 2, 10, 10, 20, 1, 2, 1)
```

Lösen Sie folgende Aufgaben jeweils zuerst auf Papier und dann mit R:

a) Geben Sie die empirische Verteilungsfunktion $F(x)$ an und zeichnen Sie den Graph von F.
b) Berechnen Sie die folgenden empirischen Quantile $\tilde{x}_{p_i}$ mit

```
p <- c(0.2, 0.25, 0.4, 0.5, 0.6, 0.75, 0.99)
```

c) Zeichnen Sie (von Hand) einen Boxplot zum Merkmal X und überprüfen Sie mit der Assgabe von R das Ergebnis.

Lösungshinweis:

a) Sortierte Urliste und relative kumulierte Häufigkeiten

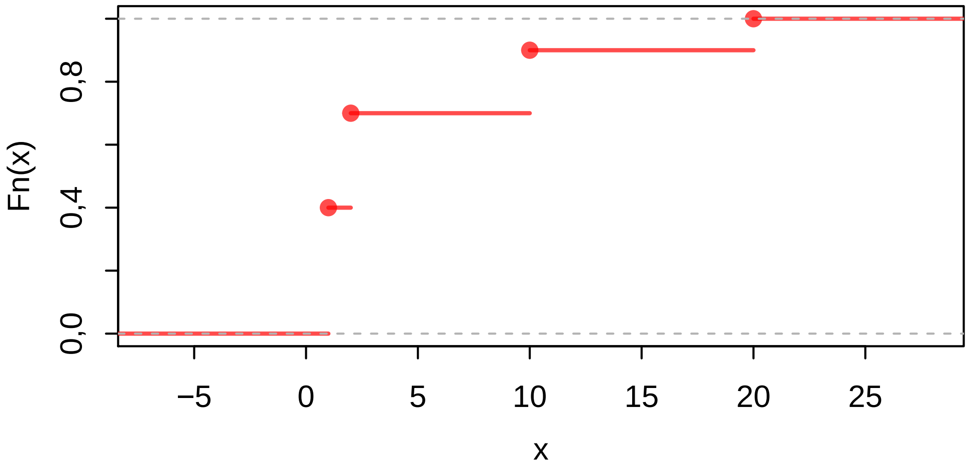
```
sort(x)

##  [1]  1  1  1  1  2  2  2 10 10 20

cumsum(table(x))/length(x)

##   1   2  10  20
## 0,4 0,7 0,9 1,0
```

damit ist

$$F(x) = \begin{cases} 0 & \text{für } \quad x < 1 \\ 0{,}4 & \text{für } \; 1 \leqq x < 2 \\ 0{,}7 & \text{für } \; 2 \leqq x < 10 \\ 0{,}9 & \text{für } 10 \leqq x < 20 \\ 1{,}0 & \text{für } 20 \leqq x \end{cases}$$

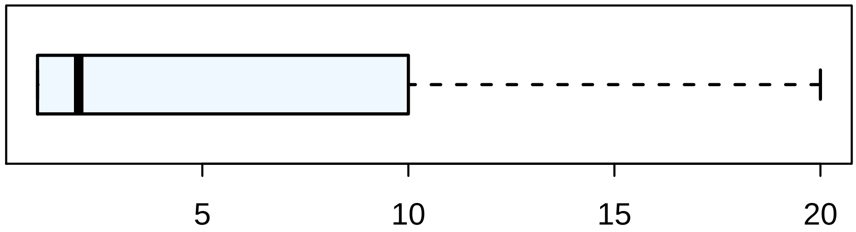
Graph von $F(x)$:

```
plot(ecdf(x), col = rgb(1, 0, 0, 0.7), lwd = 2, main = "")
```



b) Empirische Quantile und Boxplot

```
quantile(x, probs = p, type = 2)

## 20%  25%  40%  50%  60%  75%  99%
## 1,0  1,0  1,5  2,0  2,0 10,0 20,0

boxplot(x, col = "aliceblue", horizontal = TRUE, lwd = 1.5)
```



Aufgaben zur deskriptiven Statistik

Aufgabe 11

Deskriptiv: Häufigkeit1b (1)

Ein Einzelhändler registriert für einen Exklusivartikel im Verlauf von 30 Verkaufstagen folgende Verkaufszahlen:

Tag	1	2	3	4	5	6	7	8	9	10
Anzahl	5	2	3	0	0	1	3	6	0	2
Tag	11	12	13	14	15	16	17	18	19	20
Anzahl	1	0	1	0	2	3	5	1	0	0
Tag	21	22	23	24	25	26	27	28	29	30
Anzahl	3	5	3	1	0	0	0	6	3	1

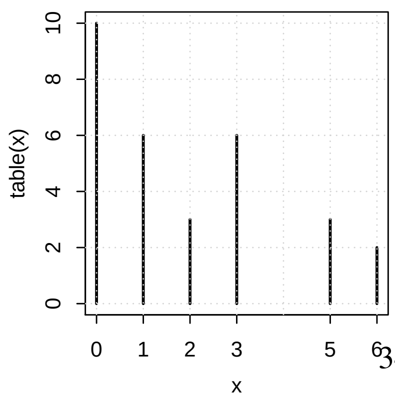
a) Berechnen Sie die absoluten und relativen Häufigkeiten der Ausprägungen sowie die absolute kumulierte Häufigkeit für $x = 4$.
b) Erstellen Sie das zugehörige Balkendiagramm und das Kreissektorendiagramm mithilfe der absoluten Häufigkeiten.

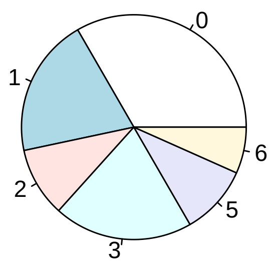
Lösungshinweis:

```r
x <- c(5, 2, 3, 0, 0, 1, 3, 6, 0, 2, 1, 0, 1, 0, 2, 3, 5,
    1, 0, 0, 3, 5, 3, 1, 0, 0, 0, 6, 3, 1)
T <- table(x)
T  # Ausgabe der Häufigkeiten
cumsum(T)  # Ausgabe der kumulierten Häufigkeiten
```

Ausprägung	0	1	2	3	5	6
Häufigkeit	10	6	3	6	3	2
kumuliert	10	16	19	25	28	30

```r
plot(table(x))
pie(table(x))
```





Aufgabe 12

Das Ergebnis der Untersuchung eines kardinalskalierten Merkmals X sei in folgender Tabelle wiedergegeben:

Ausprägung	1	2	3	4	7
Anzahl	4	4	6	4	2

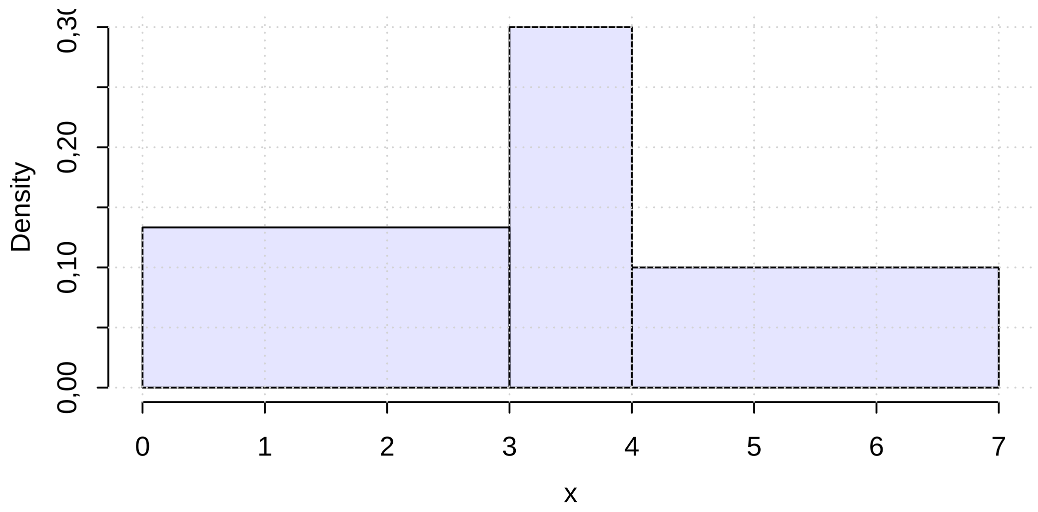
a) Bestimmen Sie das arithmetische Mittel, den Modus und den Median.
b) Berechnen Sie die mittlere quadratische Abweichung.
c) Obige Daten werden nun mittels der Intervalle $[0; 3)$, $[3; 4)$ und $[4; 7]$ klassiert. Bestimmen Sie die Rechteckhöhen des Histogramms.

Lösungshinweis:

```
x <- rep(c(1, 2, 3, 4, 7), times = c(4, 4, 6, 4, 2))
median(x)
mean(x)
SP <- max(x) - min(x)  # Spannweite
MQA <- mean((x - mean(x))^2)  # mittlere quadratische Abweichung
s <- sqrt(MQA)  # Standardabweichung
V <- s/mean(x)  # Variationskoeffizient
```

$$\begin{aligned} x_{\text{med}} &= 3, & \overline{x} &= 3, \\ \text{SP} &= 6, & s^2 &= 2{,}8, \\ s &\approx 1{,}67332, & V &\approx 0{,}55777. \end{aligned}$$

```
require(MASS)
hist(x, breaks = c(0, 3, 4, 7), right = FALSE, col = rgb(0,
    0, 1, 0.1), main = "")
grid()
```



Aufgabe 13

Eine Umfrage über den Bierkonsum Münchner Bürger ergibt bei 10 Personen folgende Zahlenreihe (Liter pro Woche):

3	10	1	2	3	0	2	1	0	3

Berechnen Sie den Modalwert, den Median, das arithmetische Mittel, die Spannweite, die mittlere quadratische Abweichung, die Standardabweichung und den Variationskoeffizienten.

Lösungshinweis:

```
x <- c(3, 10, 1, 2, 3, 0, 2, 1, 0, 3)
table(x)  # Kontingenztabelle
```

Ausprägung	0	1	2	3	10
Häufigkeit	2	2	2	3	1

also: $x_{\text{mod}} = 3$

```
median(x)
mean(x)
SP <- max(x) - min(x)  # Spannweite
MQA <- mean((x - mean(x))^2)  # mittlere quadratische Abweichung
s <- sqrt(MQA)  # Standardabweichung
V <- s/mean(x)  # Variationskoeffizient
```

$$\begin{aligned} x_{\text{med}} &= 2, \\ \overline{x} &= 2{,}5, \\ \text{SP} &= 10, \\ s^2 &= 7{,}45, \\ s &\approx 2{,}72947, \\ V &\approx 1{,}09179. \end{aligned}$$

Aufgabe 14

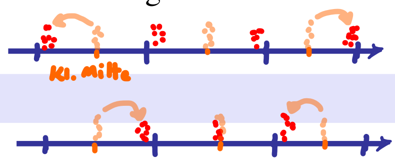
In der deutschen Fußballnationalmannschaft wird die Körpergröße (in cm) einiger (aktiver wie ehemaliger) Spieler erhoben. Es ergeben sich folgende Daten:

Körpergröße	[150, 160)	[160, 170)	[170, 175)	[175, 180)	[180, 190)	[190, 210)
Anzahl	5	35	47	38	101	27

a) Zeichnen Sie ein Histogramm der Daten.

b) Schätzen Sie das arithmetische Mittel und den Median der Körpergröße.

c) Schätzen Sie die Standardabweichung der Körpergröße, indem Sie davon ausgehen, dass alle Nennungen pro Klasse jeweils in der Klassenmitte liegen. Ist dieser Wert größer oder kleiner als die tatsächliche Standardabweichung in der Grundgesamtheit?



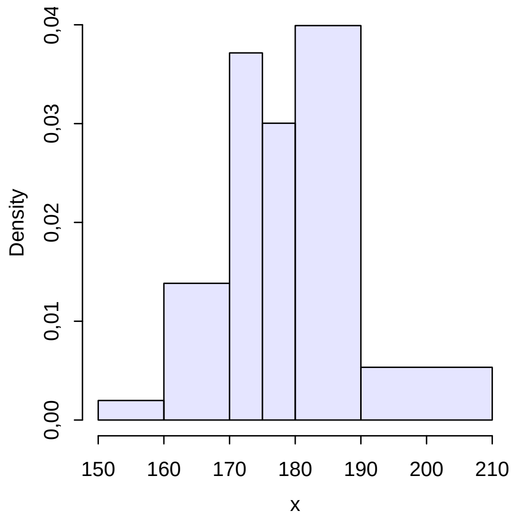
Lösungshinweis:

a)
```r
# definiere Werte jeweils in Mitte der Klasse
Klassenmitten <- c(155, 165, 172.5, 177.5, 185, 200)
Haeufigkeiten <- c(5, 35, 47, 38, 101, 27)
x <- rep(Klassenmitten, Haeufigkeiten)
table(x)

## x
##   155   165 172,5 177,5   185   200
##     5    35    47    38   101    27

# Anzahl der Objekte insgesamt
n <- length(x)
n

## [1] 253
```

Berechne Höhe$_i$ $= \frac{c \cdot h(a_i)}{\text{Breite}_i}$ mit $c = \frac{1}{n}$ (hier: $n = 253$) :

Körpergröße	[150, 160)	[160, 170)	[170, 175)	[175, 180)	[180, 190)	[190, 210)
Häufigkeit	5	35	47	38	101	27
Breite	10	10	5	5	10	20
Höhe	0,002	0,0138	0,0372	0,03	0,0399	0,0053

```r
hist(x, breaks = c(150, 160, 170, 175, 180, 190, 210),
     right = TRUE, col = rgb(0, 0, 1, 0.1),
     main = "")
```



b) Rechnet man jeweils (näherungsweise) mit den Klassenmitten, erhält man:

$$\overline{x} = 179{,}79249 \quad \text{und} \quad x_{\text{med}} = 185\,.$$

c) Für die Standardabweichung ergibt sich somit als Näherungswert:

$$s = 10{,}36781$$

Die Standardabweichung könnte so berechnet den wahren Wert über- oder unterschätzen: Angenommen alle tatsächlichen Werte würde immer möglichst nahe am arithmetischen Mittel aller Werte liegen, also jeweils am „dem arithmetischen Mittel zugewandten“ Rand der Intervalle, würde die Standardabweichung so überschätzt; im Fall der in Wirklichkeit maximal weit vom arithmetischen Mittel liegenden Werte würde die tatsächliche Standardabweichung so unterschätzt.

Aufgabe 15

Deskriptiv: Lage Streuung Vtgl.fkt. (Verteilung_2014_01)

10 Personen werden befragt, wieviel Sie in den Weihnachtsferien zugenommen haben. Für die empirische Verteilungsfunktion des abgefragten Merkmals $X \mathrel{\widehat{=}}$ „*Gewichtszunahme vom 23. Dezember bis zum darauf folgenden 7. Januar*“ ergibt sich:

$$F(x) = \begin{cases} 0{,}0 & \text{für} \quad x < -1{,}0 \\ 0{,}1 & \text{für} \quad -1{,}0 \leq x < 1{,}0 \\ 0{,}3 & \text{für} \quad 1{,}0 \leq x < 1{,}5 \\ 0{,}4 & \text{für} \quad 1{,}5 \leq x < 2{,}0 \\ 0{,}9 & \text{für} \quad 2{,}0 \leq x < 6{,}0 \\ 1{,}0 & \text{für} \quad x \geq 6{,}0 \end{cases}$$

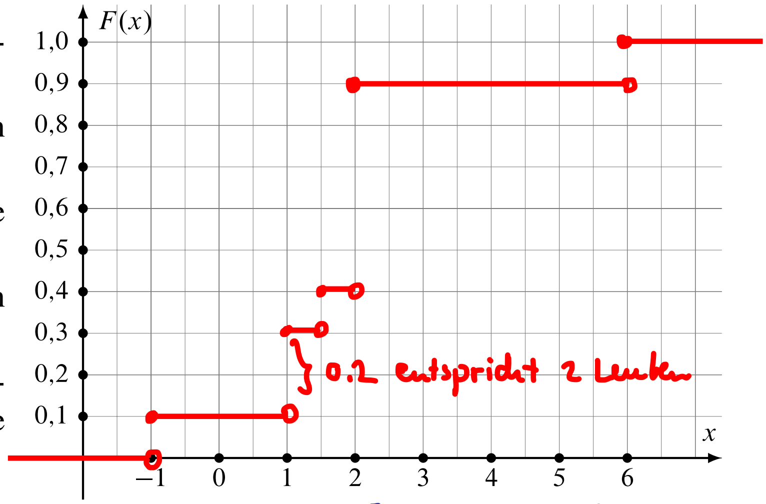
a) Zeichnen Sie $F(x)$ in nebenstehendes Koordinatensystem ein.

b) Schreiben Sie die ursprünglichen Daten als Urliste von X auf.

c) Berechnen Sie das arithmetische Mittel von X.

d) Bestimmen Sie den Median von X.

e) Berechnen Sie die mittlere quadratische Abweichung sowie die Standardabweichung von X.



-1 , 1,1,1.5,2,2,2,2,2,6

↑ $x_{med} = 2$

Lösungshinweis:

b) und a) Da die 10 %-Sprünge jeweils einer Person entsprechen (bei 10 Leuten), kann man die Urliste eindeutig aus F ablesen:

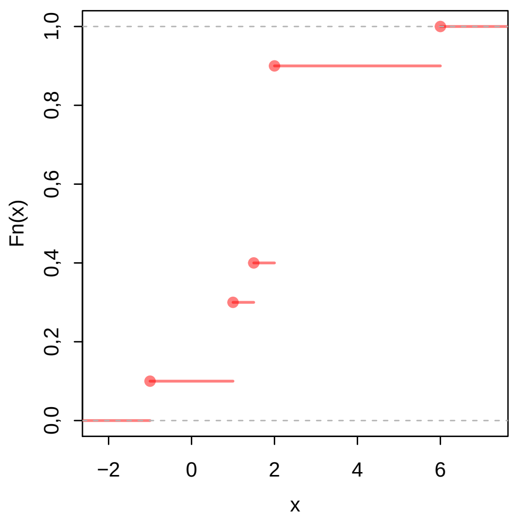
```
x <- c(-1, 1, 1, 1.5, 2, 2, 2, 2, 2, 6)
plot(ecdf(x), lwd = 2, col = rgb(1, 0, 0, 0.5), main = "")
```

empirical cumulativie distribution function

Setup → STAT → Freq. On.

TR: ▸ Mode → STAT → 1-Var
▸ Daten eingeben: -1, 1, 1, 1.5, ⋮, 6
▸ AC Shift → STAT → Var → $\bar{x}$, σ_x

Achtung: $s_x \neq \sigma_x$
induktive Stat.
deskr. Stat.



c) Für das arithmetische Mittel ergibt sich $\overline{x} = 1{,}85$.

d) Der Median ist $x_{\text{med}} = 2$.

e) $s^2 = 2{,}703$ und $s = 1{,}644$.

Aufgabe 16

Ein bestimmtes Gut wird von genau 7 Firmen produziert. Folgende Tabelle gibt an, wie viele tausend Stück jede Firma herstellt:

Firma:	A	B	C	D	E	F	G	
prod. Stückzahl:	3	2	3	5	6	15	6	(tausend Stück)

a) Bestimmen Sie den Medianwert der produzierten Stückzahlen.
b) Skizzieren Sie für $x-$Werte aus dem Intervall [0;20] den Verlauf der Funktion $F(x) =$ Anteil der Firmen, die höchstens $1000 \cdot x$ Stück produzieren.
c) Errechnen Sie die Knickpunkte der zugehörigen Lorenzkurve.
d) Errechnen Sie den normierten Gini-Koeffizienten.
e) Bestimmen Sie den Konzentrationskoeffizienten CR_2.

x_{med}

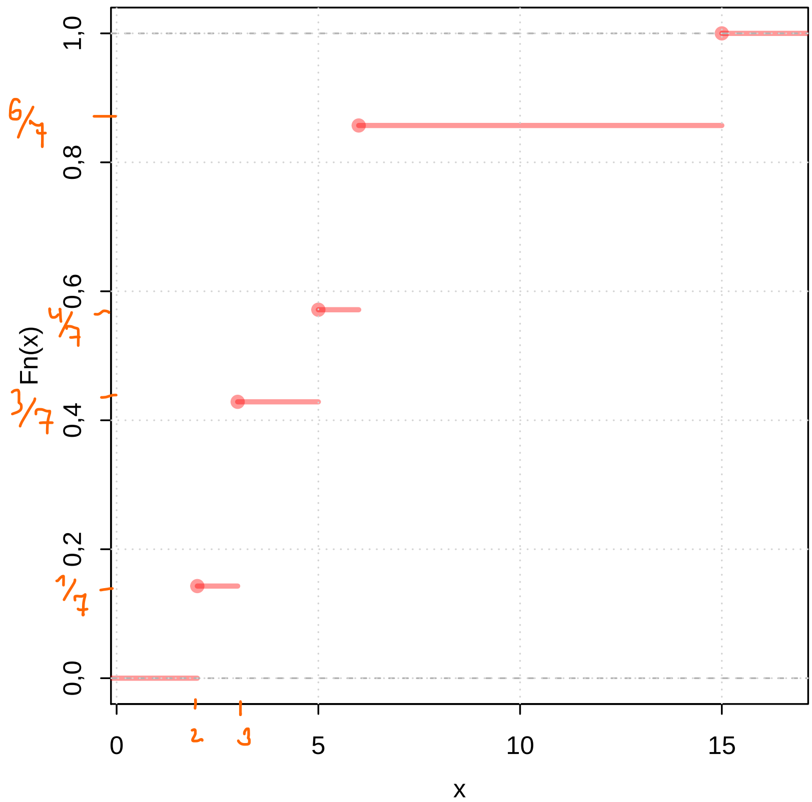
Lösungshinweis:

2 3 3 5 6 6 15

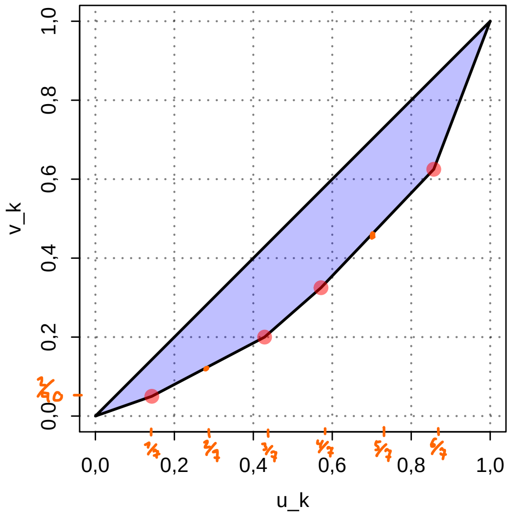
```
x <- c(3, 2, 3, 5, 6, 15, 6)
```

a) Median: $x_{\text{med}} = 5$

b)
```
plot(ecdf(x), lwd = 3, col = rgb(1, 0, 0, 0.4), main = "")
```



c)
```r
library(ineq)
plot(Lc(x), las = 2, col = rgb(1, 0, 0, 0.5))   #
```



A	B	C	D	E	F	G
3	2	3	5	6	15	6

$\Sigma = 40$

2 3 3 5 6 6 15

2/40

Für die Knicke ergibt sich:

$$\frac{2\sum_{i=1}^{n} i\,p_i - (n+1)}{n}$$

$= \frac{1}{7}\cdot\left[2\left(1\cdot\frac{2}{40}+2\cdot\frac{3}{40}+\dots+7\cdot\frac{15}{40}\right) - (7+1)\right]$

≈ 0.34286

$G^* = \frac{7}{6}\cdot G = 0.4$

k	u_k	v_k
1	0,143	0,050
3	0,429	0,200
4	0,571	0,325
6	0,857	0,625

d) $G = 0{,}34286$ und damit $G^* = \dfrac{n}{n-1}\cdot G = \dfrac{7}{6}\cdot 0{,}34286 = 0{,}4.$

e) $\mathrm{CR}_2 = 0{,}525.$

$= \frac{15+6}{40} = \frac{21}{40}$

Aufgabe 17

Bei einer Umfrage unter den Kindern einer Schulklasse nach der Anzahl ihrer Geschwister ergaben sich die folgenden Daten:

u_k: 6/20, 10/20, 15/20, 17/20, 19/20, 1
v_k: 0, 4/36, 14/36, 20/36, 30/36, 1

Anzahl Geschwister (a_i)	0	1	2	3	5	6
Häufigkeit (h_i)	6	4	5	2	2	1

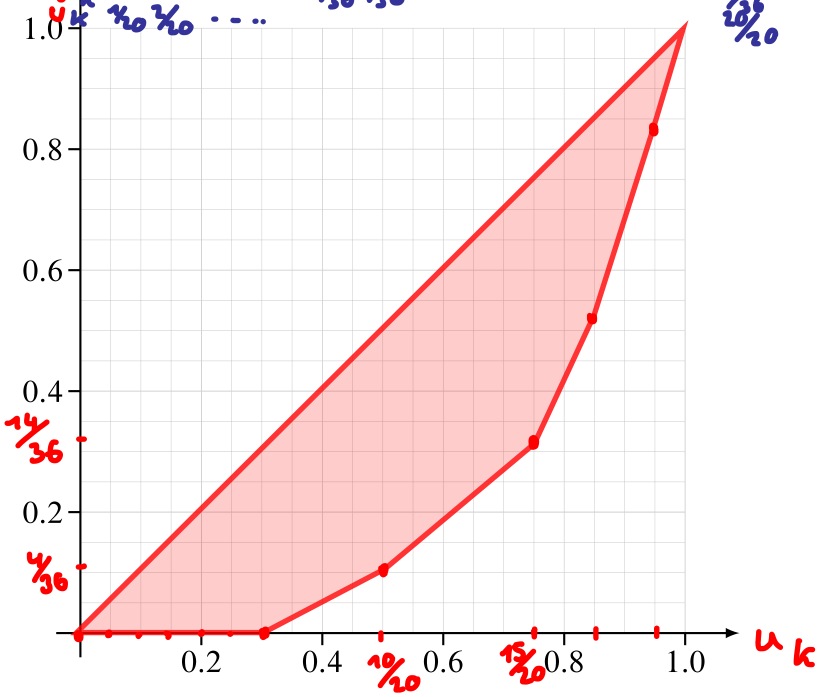
x_i: 0 0 0 0 0 0 1 1 1 1 2 2 2 2 2 3 3 5 5 6 Σ 36
p_i: 0 0 0 0 0 0 1/36 1/36 ...
v_k: 0 0 0 ... 0 1/36 2/36 ...
u_k: 1/20 2/20 ...

a) Erstellen Sie eine Tabelle, die die Koordinaten der Lorenzkurve enthält. Zeichnen Sie die Lorenzkurve in nebenstehendes Diagramm.

b) Bestimmen Sie den Mittelwert, den Median, die Standardabweichung, $f(4)$ und $F(4)$ der Verteilung.

c) Bestimmen Sie das 25 %- sowie das 75 %-Quantil der Daten.

d) Zeichnen Sie einen Boxplot der Daten.

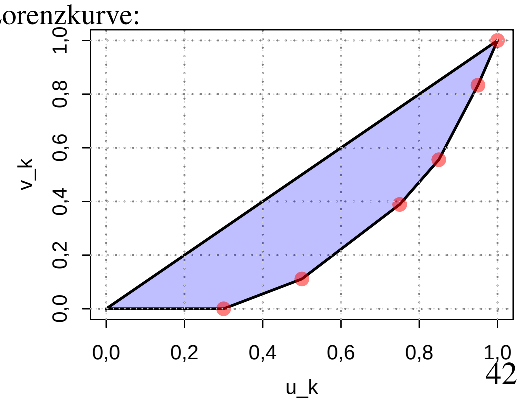


Lösungshinweis:

a) Tabelle:

i	1	2	3	4	5	6
u_k	0,30	0,50	0,75	0,85	0,95	1,00
v_k	0,00	0,11	0,39	0,56	0,83	1,00

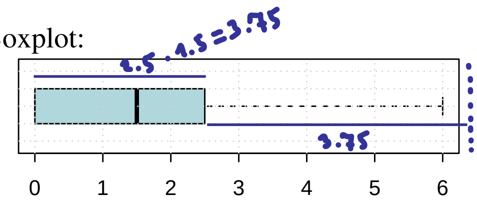
Lorenzkurve:



b) Für das arithmetische Mittel ergibt sich $\overline{x} = 1{,}8$. Der Median ist $x_{\text{med}} = 1{,}5$. Die Standardabweichung beträgt $s \approx 1{,}7776$, außerdem gilt $f(4) = 0$ und $F(4) = 0{,}85$.

c)
```
## 0% 25% 50% 75% 100%
## 0,0 0,0 1,5 2,5 6,0
```

d) Boxplot:



$\tilde{x}_{0,75} = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(2+3) = 2.5$

$n \cdot p = 20 \cdot 0.75 = 15 \in \mathbb{N}$

Setup → STAT → Freq. On/Off
Häufigk. Ein/Aus
Mode → STAT → 1-Var

x	Freq.
0	6
1	4
2	5
3	2
5	2
6	1

AC, Shift → STAT
→ MinMax
→ Var

Median
s, $\overline{x}$
Casio: σ_x

Aufgabe 18

Die Firma CelebWedCake liefert zu einem Festpreis von 200.000 € eine exklusive Premium-Hochzeitstorte an Prominente. In den letzten 5 Jahren wurden insgesamt 20 von diesen Torten verkauft. Pro Kunde ist die Anzahl der verkauften Torten in dieser Zeitspanne mittels der verschiedenen Ausprägungen a_i und den zugehörigen absoluten Häufigkeiten h_i erfasst:

i	1	2	3	4
a_i	1	2	4	5
h_i	5	3	1	h_4

$1\cdot 5 + 2\cdot 3 + 4\cdot 1 + 5\cdot h_4 = 20$

$\Leftrightarrow 5 h_4 = 5 \Leftrightarrow h_4 = 1$

a) Bestimmen Sie h_4.

b) Zeichnen Sie die Lorenzkurve,

c) berechnen Sie den normierten Gini-Koeffizienten sowie

d) den Herfindahl- und

e) den Exponentialindex der Anzahl der verkauften Torten pro Kunde.

Lösungshinweis:

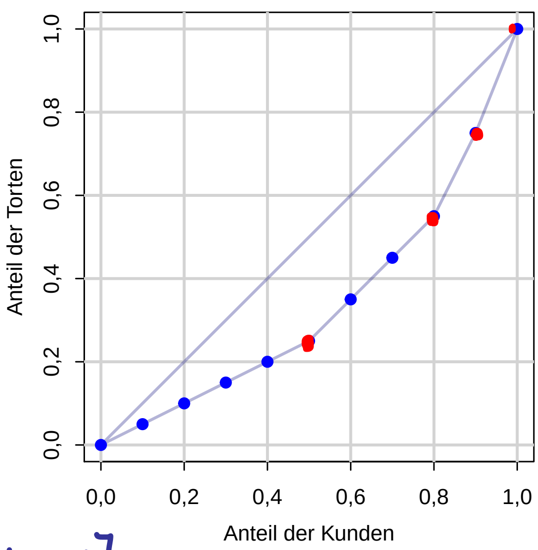
a) Insgesamt 20 Torten, laut Tabelle $1 \cdot 5 + 2 \cdot 3 + 4 \cdot 1 = 15$ Torten für bis zu 4 Torten. Bleibt ein Kunde mit 5 Torten, also $h_4 = 1$.

b) Lorenzkurve siehe rechts

c) Zehn Kunden. Gini: $G = 0{,}33$, normiert: $G_* = 0{,}36667$

d) Herfindahl: 0,145

e) Exponentialindex: 0,12146



$1,1,1,1,1,2,2,2,4,5 \qquad \Sigma = 20$

$\Rightarrow G = \frac{1}{10}\Big[2\Big(1\cdot\frac{1}{20} + 2\cdot\frac{1}{20} + 3\,\frac{1}{20} + 4\,\frac{1}{20} + 5\cdot\frac{1}{20} + 6\,\frac{2}{20} + \dots + 10\cdot\frac{5}{20}\Big) - (10+1)\Big]$

d) $H = \left(\frac{1}{20}\right)^2 + \dots + \left(\frac{5}{20}\right)^2$

e) $E = \left(\frac{1}{20}\right)^{1/20} \cdot \dots \left(\frac{5}{20}\right)^{5/20} = \frac{1}{20}\sqrt[20]{1^1 \cdot 1^1 \dots 2^2 \dots 5^5}$

Aufgabe 19

Die folgende Tabelle gibt jeweils den jährlichen Umsatz der weltweit 10 umsatzstärksten Softwareunternehmen an:

Nr. des Unternehmens:	1	2	3	4	5	6	7	8	9	10
Umsatz [in Mrd. US-Dollar]:	4	49	4	22	5	18	12	6	7	6

a) Zeichnen Sie die Lorenzkurve des Umsatzes.
b) Berechnen Sie den Gini-Koeffizienten und den normierten Gini-Koeffizienten.
c) Berechnen Sie den Herfindahl- sowie den Exponentialindex.

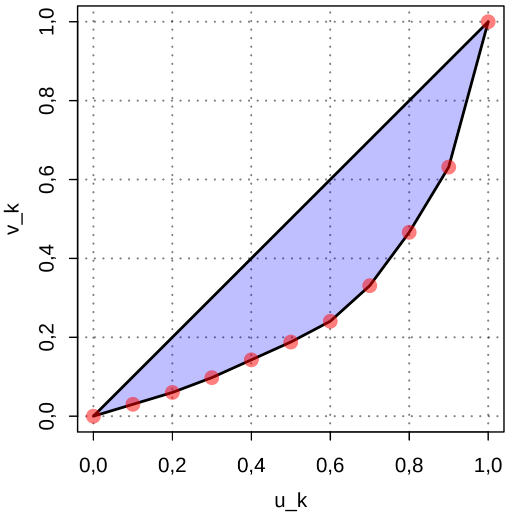
Lösungshinweis:

```
x <- c(4, 49, 4, 22, 5, 18, 12, 6, 7, 6)
## [1]  4  4  5  6  6  7 12 18 22 49
```

a) Kumulierte Anteile:

	0	1	2	3	4	5	6	7	8	9	10
u_k	0,000	0,100	0,200	0,300	0,400	0,500	0,600	0,700	0,800	0,900	1,000
v_k	0,000	0,030	0,060	0,098	0,143	0,188	0,241	0,331	0,466	0,632	1,000

b)
```
library(ineq)
plot(Lc(x), las = 2, col = rgb(1, 0, 0, 0.5))  #
```



c) $G = 0{,}46241$ und damit $G^* = \dfrac{n}{n-1} \cdot G = \dfrac{10}{9} \cdot 0{,}46241 = 0{,}51378.$

d) Herfindahl: 0,19962, Exponentialindex: 0,14633.

Aufgabe 20

Pia lädt 11 Freundinnen zu einem Damenabend ein. Es gibt 8 Flaschen Prosecco. Pro Flasche kann Sie 5 Gläser ausschenken.

- 3 Freundinnen müssen fahren und trinken nichts vom Prosecco,
- 3 Freundinnen trinken jeweils 3 Gläser,
- 4 Freundinnen trinken jeweils 5 Gläser,
- 1 Freundin trinkt 8 Gläser und
- Pia übernimmt den Rest.

a) Welchen Anteil am Prosecco muss Pia trinken?

b) Geben Sie die Häufigkeitsverteilung der Gläser pro Dame an.

c) Bestimmen Sie den Median, das arithmetische Mittel und die Standardabweichung der Gläser pro Dame.

d) Welche Werte nimmt die zur Anzahl der Gläser x pro Dame gebildete empirische Verteilungsfunktion $F(x)$ bei $x = 2$ und bei $x = 5$ an?

e) Zeichnen Sie die Lorenzkurve und

f) berechnen Sie den normierten Gini–Koeffizienten der Gläseranzahl pro Dame.

Lösungshinweis:

a) $(40 - (3 \cdot 0 + 3 \cdot 3 + 4 \cdot 5 + 1 \cdot 8) : 40 = \frac{3}{40} = 0{,}075$.

b)

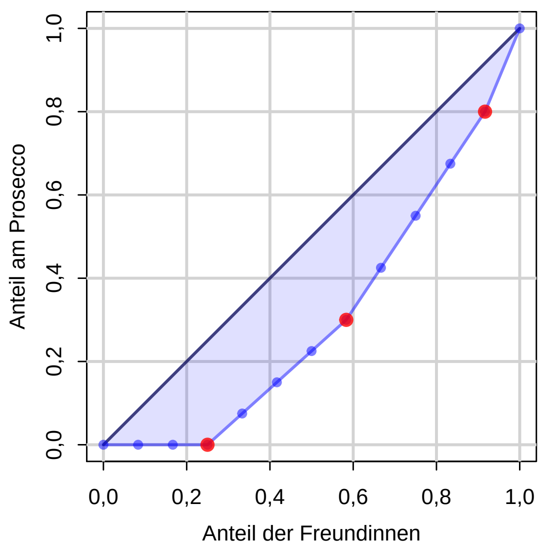
i	1	2	3	4
a_i	0	3	5	8
h_i	3	4	4	1

c) $x_{\text{med}} = 3$,
$\overline{x} = 3{,}33333$,
$s = 2{,}35702\,.$

d) $F(2) = 3/12 = 0{,}25$
$F(5) = 11/12 \approx 0{,}917$.

e) Lorenzkuve siehe rechts

f) 12 Leute: $G = 0{,}383$, norm.: $G_* = 0{,}418$



Aufgabe 21

An 5 aufeinanderfolgenden Zeitpunkten wurden Preise p und Mengen q zweier Güter G_1 und G_2 festgestellt:

Jahr	1		2		3		4		5	
Gut	p_1	q_1	p_2	q_2	p_3	q_3	p_4	q_4	p_5	q_5
G_1	2,5	5	3	4	2	6	3,5	3	4	2
G_2	3	6	2	10	4	5	4	4	3	10

Betrachtet werden im Folgenden die Berichtszeit 5 und die Basiszeit 1.

a) Berechnen Sie jeweils den Preisindex von Laspeyres, Paasche, Fisher sowie den von Marshall-Edgeworth.

b) Bestimmen Sie jeweils zu Gut 1 bzw. Gut 2 die durchschnittlichen Mengen $\overline{q}^{G_1}, \overline{q}^{G_2}$ der konsumierten Mengen über alle 5 Jahre. Berechnen Sie damit den Preisindex von Lowe.

c) Angenommen die Mengen zu den Zeitpunkten 1 bzw. 5 sind nicht bekannt. Wie hoch ist der maximale bzw. minimale Wert für den Preisindex von Laspeyres?

Lösungshinweis:

a) $P^{\mathrm{L}}_{1,5} = \dfrac{4 \cdot 5 + 3 \cdot 6}{2{,}5 \cdot 5 + 3 \cdot 6} \approx 1{,}246\,,\qquad P^{\mathrm{P}}_{1,5} = \dfrac{4 \cdot 2 + 3 \cdot 10}{2{,}5 \cdot 2 + 3 \cdot 10} \approx \overset{1{,}086}{\cancel{1{,}068}}$

$P^{\mathrm{F}}_{1,5} = \sqrt{P^{\mathrm{L}}_{1,5} \cdot P^{\mathrm{P}}_{1,5}} \approx 1{,}163\,,\qquad P^{\mathrm{ME}}_{1,5} = \dfrac{4 \cdot (5+2) + 3 \cdot (6+10)}{2{,}5 \cdot (5+2) + 3 \cdot (6+10)} \approx 1{,}16$

b) Durchschnittliche Mengen:

$$\overline{q}^{G_1} = \frac{1}{5}\sum_{i=1}^{5} q_i^{G_1} = \frac{1}{5}(5+4+6+3+2) = 4$$

$$\overline{q}^{G_2} = \frac{1}{5}\sum_{i=1}^{5} q_i^{G_2} = \frac{1}{5}(6+10+5+4+10) = 7$$

Damit ergibt sich für den Preisindex nach Lowe:

$$P^{\mathrm{LO}}_{1,5} = \frac{4 \cdot 4 + 3 \cdot 7}{2{,}5 \cdot 4 + 3 \cdot 7} \approx 1{,}194$$

c) „Extrem"-Warenkörbe:

- $q_1 = 1,\ q_2 = 0 \quad \Rightarrow \quad P^{\mathrm{L}}_{1,5} = \frac{4}{2{,}5} = 1{,}6$
- $q_1 = 0,\ q_2 = 1 \quad \Rightarrow \quad P^{\mathrm{L}}_{1,5} = \frac{3}{3} = 1{,}0$

Also: $1{,}0 \leqq P^{\mathrm{L}}_{1,5} \leqq 1{,}6$.

Aufgabe 22

Für zwei Güter G_1,G_2 sind in den Jahren $t = 1, \dots, 5$ folgende Preise x_t (zu G_1) bzw. y_t (zu G_2) pro Mengeneinheit (= ME) registriert worden.

t	1	2	3	4	5	
x_t	2	2	3	3	5	(€)
y_t	4	5	8	7	8	(€)

Im Jahr 1 wurden 9 ME von G_1 und 8 ME von G_2 verbraucht, im Jahr 4 waren dies 12 ME von G_1 und 9 ME von G_2.

- Bei den im folgenden betrachteten Indexzahlen besteht der Warenkorb nur aus G_1 und G_2; als Basisperiode ist stets das Jahr 1 angesetzt.
  a) Offensichtlich sind die Verbrauchsmengen nicht für jedes Jahr bekannt. Für welche $t \in \{2,3,4,5\}$ läßt sich dennoch ein Laspeyres–Preisindex P^L_{1t} ermitteln? (Kurze Begründung, aber keine Berechnung erforderlich!)
  b) Für welche $t \in \{2,3,4,5\}$ läßt sich ein Paasche-Preisindex P^P_{1t} ermitteln? Berechnen Sie *einen* solchen Indexwert P^P_{1t}.

- Nun beschreibt $y = \hat{a} + \hat{b}x$ die Regressionsgerade mit dem Preis von G_1 als unabhängiger und dem Preis von G_2 als abhängiger Variablen.
  c) Berechnen Sie $\hat{a}$ und $\hat{b}$.
  d) Prognostizieren Sie auf Basis dieser Regression den Preis pro ME des Gutes G_2, wenn G_1 pro ME 6 € kostet.

Lösungshinweis:

a) Berechnung von $P^{\mathrm{L}}_{1,t}$ für $t \in \{2, 3, 4, 5\}$ möglich, denn die Mengen im Basisjahr $t = 1$ sind bekannt.

b) Nur $P^{\mathrm{P}}_{1,4}$ kann man berechnen: $P^{\mathrm{P}}_{1,4} = \dfrac{3 \cdot 12 + 7 \cdot 9}{2 \cdot 12 + 4 \cdot 9} \approx 1{,}65$

c) $y = 2{,}9 + 1{,}16667 \cdot x$

d) $\hat{y} = 2{,}9 + 1{,}16667 \cdot 6 = 9{,}9$

Aufgabe 23

Zwei Personen sollen fünf verschiedene Produkte A bis E durch Angabe einer Reihenfolge beurteilen. Die Befragung ergab folgende Ergebnisse:

Produkt	Person I	Person II
A	5	3
B	2	1
C	3	4
D	4	2
E	1	5

Berechnen Sie den Korrelationskoeffizienten von Spearman.

Lösungshinweis:

Sowohl mit als auch mit Formel bzw. Taschenrechner ergibt sich $r_{\mathrm{SP}} = -0{,}3$.

```
Person1 <- c(5, 2, 3, 4, 1)
Person2 <- c(3, 1, 4, 2, 5)
```

```
# method='spearman' ist der Rangkorrelationskoeffizient
# (Schalter ist hier eigentlich überflüssig, da sowieso
# schon Ränge vorliegen)
cor(Person1, Person2, method = "spearman")

## [1] -0,3
```

Aufgabe 24

Ein Betrieb hat im Kalenderjahr 2004 zwölf neue Mitarbeiter eingestellt. Von diesen sind unter anderem folgende Daten bekannt:

Mitarbeiter Nr.	Geschlecht	Ausbildungsdauer (in Jahren)	Abschlussnote
1	männlich	9	4
2	weiblich	10	2
3	weiblich	10	4
4	männlich	11	4
5	weiblich	12	2
6	weiblich	13	2
7	weiblich	14	1
8	männlich	15	3
9	männlich	16	2
10	männlich	17	3
11	weiblich	19	3
12	männlich	22	2

a) Geben Sie die Skalierung der drei Merkmale Geschlecht, Ausbildungsdauer und Abschlussnote an.

b) Ermitteln Sie für jedes der drei Merkmale die folgenden Größen, soweit diese aufgrund des jeweiligen Skalenniveaus sinnvollerweise berechnet werden können:
   (1) Modus
   (2) Median
   (3) Arithmetisches Mittel
   (4) Mittlere quadratische Abweichung
   (5) Variationskoeffizient

c) Geben Sie für jedes der zwei Merkmalspaare
   (1) Geschlecht – Abschlussnote
   (2) Ausbildungsdauer – Abschlussnote
   einen statistisch sinnvollen Korrelationskoeffizienten an.
   (Die Korrelationskoeffizienten müssen nicht berechnet werden.)

Lösungshinweis:

a) Geschlecht: *nominal*,
   Ausbildungsdauer: *kardinal*,
   Abschlussnote: *ordinal*

b)

	Geschlecht	Ausbildungsdauer	Abschlussnote
Modus	(nicht eindeutig)	10	2
Median	–	13,5	2,5
Arithmetisches Mittel	–	14	–
Mittlere quadratische Abweichung	–	14,5	–
Variationskoeffizient	–	0,27	–

c) Geeignete Korrelationskoeffizienten:
   (1) Geschlecht – Abschlussnote: *Kontingenzkoeffizient*
   (2) Ausbildungsdauer – Abschlussnote: *Rangkorrelationskoeffizient*

	Skalierung von Y		
Skalierung von X	kardinal	ordinal	nominal
kardinal	Bravais-Pearson-Korrelationskoeffizient	Rangkorrelationskoeffizient von Spearman	Kontingenzkoeffizient
ordinal	Rangkorrelationskoeffizient von Spearman	Rangkorrelationskoeffizient von Spearman	Kontingenzkoeffizient
nominal	Kontingenzkoeffizient	Kontingenzkoeffizient	Kontingenzkoeffizient

Aufgabe 25

Bei einer Befragung von Passanten in einer Fußgängerzone bezüglich ihres Bierkonsums in Litern pro Woche und ihrer Selbsteinschätzung als Fußballfan ergaben sich folgende Daten:

20 Fußballfans und 120 Nichtfußballfans gaben einen Bierkonsum von höchstens 1 Liter pro Woche an. Zwischen 1 und 3 Liter pro Woche trinken 210 Fußballfans und 200 Nichtfußballfans. 150 Fußballfans und 90 Nichtfußballfans gaben einen Bierkonsum von mindestens 7 Liter an. 145 Fußballfans und 65 Nichtfußballfans lagen in der verbleibenden Zwischengruppe.

a) Stellen Sie die zugehörige Kontingenztabelle auf.
b) Errechnen Sie die Randhäufigkeiten.
c) Berechnen Sie die bedingte Verteilung des Bierkonsums für Fußballfans.
d) Sind Bierkonsum und die Fußballaffinität unabhängig?
e) Berechnen Sie ein geeignetes Zusammenhangsmaß für die beiden Merkmale.

Lösungshinweis:

(Lösung in R nicht prüfungsrelevant, mit Bleistift und Papier schon)

a)
```
Fussball.Bier

##           Fussball
## Bier        Fan kein Fan
##   [0; 1]     20      120
##   (1; 3)    210      200
##   [3; 7)    145       65
##   [7; inf)  150       90
```

b)
```
addmargins(Fussball.Bier)
```

$\frac{525 \cdot 140}{1000}$

## Bier	Fan		kein Fan		Sum
## [0; 1]	20	73,5	120	66,5	140
## (1; 3)	210	215,25	200	194,75	410
## [3; 7)	145	110,25	65	99,75	210
## [7; inf)	150	126	90	114	240
## Sum	525		475		1000

e) $\chi^2 = \frac{(20-73{,}5)^2}{73{,}5} + \frac{(120-66{,}5)^2}{66{,}5} + \ldots$

$\ldots + \frac{(90-114)^2}{114}$

$\approx 114{,}936$

$\Rightarrow k = \sqrt{\frac{\chi^2}{1000+\chi^2}} \approx 0.321$

c) Bedingte Verteilung der Spalte *Fan* bezogen auf alle Fans (das sind 525):

```
round(Fussball.Bier[, 1]/sum(Fussball.Bier[, 1]), 3)

##   [0; 1]    (1; 3)    [3; 7) [7; inf)
##    0,038     0,400     0,276    0,286
```

d) zum Beispiel Bierkonsum [0; 1], Fussball *Fan*: Erwartete Häufigkeit bei Unabhängigkeit:

$$\frac{140 \cdot 525}{1000} = 73{,}5 \neq 20,$$

also nicht unabhängig.

e) Kontingenzkoeffizient: $K = 0{,}32107$

Aufgabe 26

Deskriptiv: Kontingenzkoeffizient (Mensa_Kontingenzkoeffizient)

An einer Hochschule sollen die Studierenden ihre Mensa bezüglich der Qualität des Essens beurteilen. In einer Voruntersuchung haben 50 Studenten aus vier Studienjahren befragt ein bestimmtes Gericht bezüglich des Geschmacks als schlecht, mittel bzw. gut bewertet. Folgende Häufigkeitstabelle fasst die Ergebnisse zusammen:

	Studienjahr			
Bewertung	1	2	3	4
schlecht	0	6	3	1
mittel	5	9	1	0
gut	5	5	6	9

```
##           Studienjahr
## Bewertung  1  2  3  4 Sum
##   schlecht 2  4  2  2  10
##   mittel   3  6  3  3  15
##   gut      5  10 5  5  25
##   Sum     10 20 10 10  50
```

$\chi^2 = (0-2)^2/2 + (6-4)^2/4 + \dots$

$\dots + (9-5)^2/5$

$= 17{,}06$

Berechnen Sie den normierten Kontingenzkoeffizient zwischen der Zugehörigkeit zum Studienjahr und der vergebenen Bewertung.

$K = \sqrt{\frac{\chi^2}{\chi^2+50}} \approx 0{,}5044$

$K_{max} = \sqrt{\frac{3-1}{3}} \approx 0{,}8165$

$K^* \approx 0{,}618$

Lösungshinweis:

(Lösung in R nicht prüfungsrelevant, mit Bleistift und Papier schon)

```
Mensa = matrix(c(0, 6, 3, 1,
                 5, 9, 1, 0,
                 5, 5, 6, 9), nrow=3, byrow=T)
dimnames(Mensa) =
  list(Bewertung=c("schlecht", "mittel", "gut"),
       Studienjahr=c("1", "2", "3", "4"))

# wird noch nicht als Kontingenztabelle erkannt
Mensa = as.table(Mensa)

# Paket vcd: "visualizing categorical data"
library(vcd)
K = assocstats(Mensa)$cont
K_max = sqrt(2/3)
K_normiert = K/K_max
```

Mit den Randhäufigkeiten

```
addmargins(Mensa)

##            Studienjahr
## Bewertung   1  2  3  4 Sum
##   schlecht  0  6  3  1  10
##   mittel    5  9  1  0  15
##   gut       5  5  6  9  25
##   Sum      10 20 10 10  50
```

ergibt sich $K = 0{,}50445$ und die normierte Variante $K^* = 0{,}61783$.

Aufgabe 27

Bei 8 zufällig ausgewählten Arbeitnehmerinnen und Arbeitnehmern wird die Schuhgröße S und das jährliche Einkommen E erfasst. Es ergeben sich folgende Werte:

i	1	2	3	4	5	6	7	8
S_i	36	44	40	49	38	34	34	45
E_i	41910	53860	37360	73450	46720	39560	20470	69040

a) Berechnen Sie einen geeigneten Korrelationskoeffizienten zwischen Schuhgröße und Einkommen. Interpretieren Sie das Ergebnis.

b) Bestimmen Sie die Regressionsparameter eines linearen Modells, in dem die Höhe des Einkommens in Abhängigkeit von der Schuhgröße beschrieben wird.

c) Wieviel Einkommen würden Sie gemäß diesem Modell bei einer Person wie zum Beispiel Dirk Nowitzki mit Schuhgröße 54 und bei jemanden mit Größe 35 (z.B. Kylie Minogue) erwarten?

d) Zeichnen Sie die Werte zusammen mit der Regressionsgerade in ein geeignetes Koordinatensystem ein.

e) Wie hoch ist der Determinationskoeffizient dieses Modells? Was sagt diese Größe aus?

f) Bewerten Sie das Modell bezüglich Kausalität versus Korrelation und geben Sie eine potentielle latente Variable an.

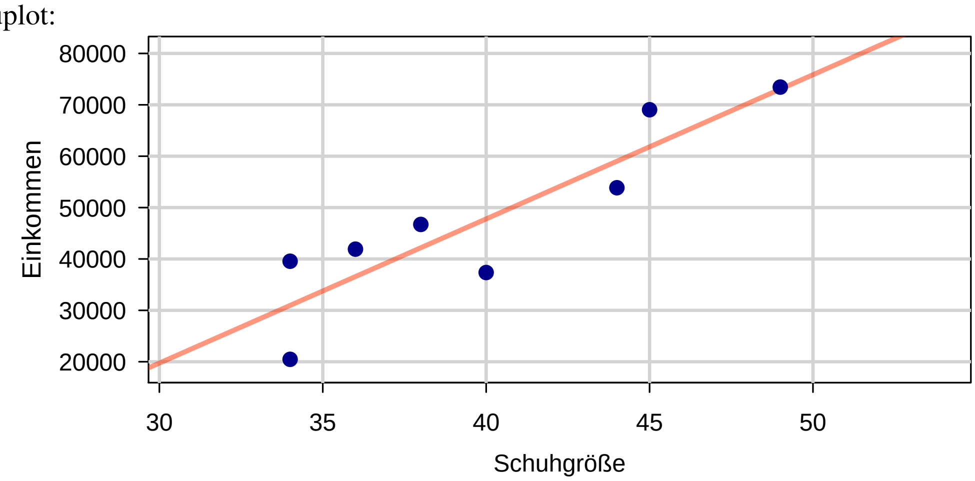
Lösungshinweis:

a) Bravais-Pearson: $r = 0{,}895$. -64434)

b) Modell: $E \approx 2805{,}75 \cdot S + (\cancel{-64400})$

c) Prognose: $E(54) = 87076{,}72$, $E(35) = 33767{,}51$

d) Streuplot:



e) Determinationskoeffizient: $R^2 \approx 0{,}8$, also stecken ca. 80 % der Information aus den Daten im Modell.

f) Vermutlich Keine Kausalität im Modell; Eventuelle latente Variable: Geschlecht

Aufgabe 28

In einem Unternehmen fragt man sich, ob zwischen Umsatz und Marketingkosten ein Zusammenhang besteht. Folgende betrieblichen Daten (in 1000 €) liegen vor:

Marketingkosten/Kunde	Umsatz/Kunde
1,4	210
1,8	220
1,9	240
2,4	241
2,8	320
3,2	400
3,6	410
4,0	480

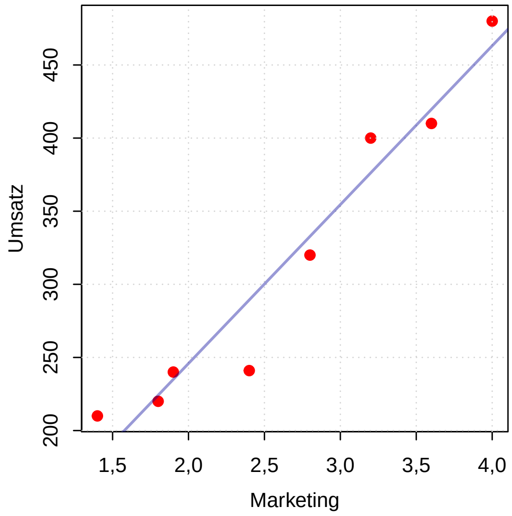
a) Erstellen Sie ein Streuungsdiagramm (y = Umsatz, x = Marketingkosten) und berechnen Sie den Bravais-Pearson- und den Rangkorrelationskoeffizienten.

b) Stellen Sie die Regressionsgerade $\hat{y} = \hat{a} + \hat{b}\,x$ auf und berechnen Sie den Determinationskoeffizienten.

Lösungshinweis:

„in Abh. von..."

a)
```
Marketing <- c(1.4, 1.8, 1.9, 2.4, 2.8, 3.2, 3.6, 4)
Umsatz <- c(210, 220, 240, 241, 320, 400, 410, 480)
Regression <- lm(Umsatz ~ Marketing)
plot(Marketing, Umsatz, pch = 20, col = "red", cex = 1.5)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
# Bravais-Pearson:
cor(Marketing, Umsatz)

## [1] 0,97048

# Rangkorrelation:
cor(Marketing, Umsatz, method = "spearman")

## [1] 1
```



b)
```
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```

Modell: $\hat{y} = 28{,}63 + 108{,}624 \cdot x$
Determinationskoeffizient: $R^2 \approx 0{,}94183$

Aufgabe 29

Von einer Firma sind über mehrere Jahre hinweg die Umsätze und die Beschäftigtenzahlen bekannt:

Jahr t:	1	2	3	4	5	6
Umsatz x_t (in Millionen €):	60	55	57	61	65	62
Anzahl y_t der Beschäftigten:	1000	1100	960	840	800	700

a) Berechnen Sie den Variationskoeffizienten des Umsatzes.
b) Berechnen Sie den Rangkorrelationskoeffizienten von Spearman zwischen den beiden Merkmalen Umsatz und Beschäftigtenzahl.
c) Berechnen Sie die Regressionsgerade $\hat{y} = \hat{a} + \hat{b}t$ der Beschäftigtenzahl in Abhängigkeit von der Zeit. Mit welcher Anzahl der Beschäftigten ist im Jahr 8 zu rechnen?

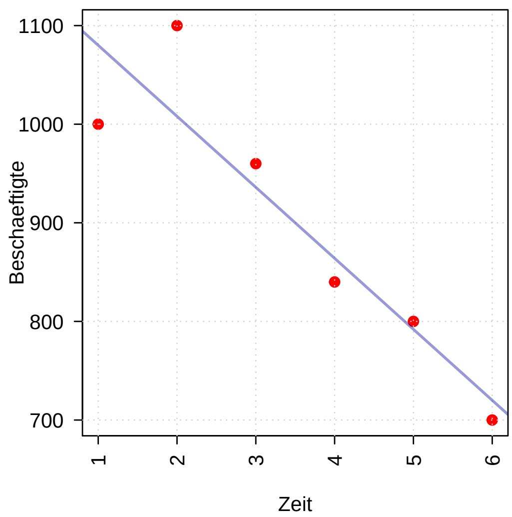
Lösungshinweis:

a)
```
Umsatz = c(60,55,57,61,65,62)
Beschaeftigte = 100 * c(10, 11, 9.6, 8.4, 8, 7)
m = mean(Umsatz)              # arithm. Mittel
s = sqrt(mean((Umsatz-m)^2))  # Standardabweichung
V = s/m                       # Variationskoeffizient
V

## [1] 0,054433
```

b) Rangkorrelationskoeffizient

```
# Rangkorrelation
cor(Umsatz, Beschaeftigte, method = "spearman")

## [1] -0,88571
```

c)
```
Zeit <- 1:6
Regression <- lm(Beschaeftigte ~ Zeit)
plot(Zeit, Beschaeftigte, pch = 20, col = "red", cex = 1.5,
    las = 2)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```



Modell: $\hat{y} = 1152 + -72 \cdot x$
Prognose: $y(8) = 1152 + -72 \cdot 8 = 576$

Aufgabe 30

An 5 aufeinander folgenden Zeitpunkten wurden Preise p und Mengen q eines Gutes festgestellt:

Zeitpunkt	1	2	3	4	5
p	2,5	3	2	3,5	4
q	5	4	6	3	2

a) Berechnen Sie den Rangkorrelationskoeffizienten von Spearman. Welche Vermutung wird durch das Ergebnis nahe gelegt?
b) Bestimmen Sie die Regressionsgerade $\hat{q} = \hat{a} + \hat{b}\, p$.
c) Wie groß ist der Korrelationskoeffizient von Bravais-Pearson? (Beachten Sie Ihr Ergebnis aus Teil b)!).

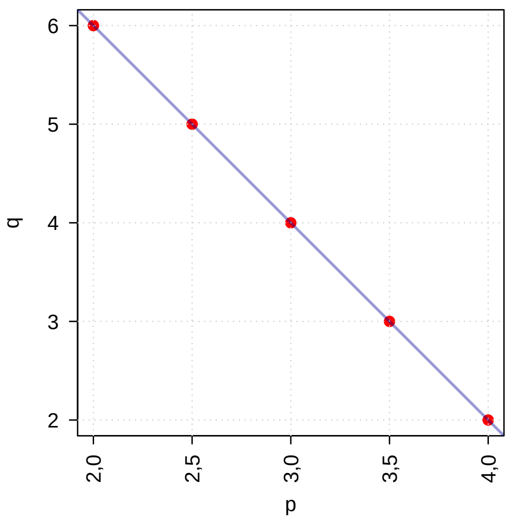
Lösungshinweis:

a) Rangkorrelationskoeffizient

```
p = c(2.5, 3, 2, 3.5, 4)
q = c(5, 4, 6, 3, 2)

# Rangkorrelation
cor(p, q, method="spearman")
## [1] -1
```

Perfekte Rangkorrelation, die Preise und die Mengen haben eindeutig gegenläufige Rangnummern. Vermutlich ist auch der Korrelationskoeffizient von Bravais-Pearson nahe bei -1.

b)
```
Regression <- lm(q ~ p)
plot(p, q, pch = 20, col = "red", cex = 1.5, las = 2)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```



Modell: $\hat{y} = 10 + -2 \cdot x$
Bravais-Pearson ist $r = -1$.

Aufgabe 31

Zwischen der Anzahl der Besucher eines Freibades und der Tageshöchsttemperatur wird ein Zusammenhang vermutet. Es wurden folgende Daten erhoben:

Tag	Besucheranzahl	Höchsttemperatur (in Grad Celsius)
1	340	35
2	150	25
3	250	28
4	300	32
5	240	26
6	220	28

a) Berechnen Sie ein geeignetes Zusammenhangsmaß zwischen der Besucheranzahl und der Höchsttemperatur.
b) Berechnen Sie die Regressionskoeffizienten der linearen Regression, wenn die Höchsttemperatur als einzige Einflussgröße für die Besucheranzahl erachtet wird.
c) Mit welcher Besucheranzahl ist bei einer Höchsttemperatur von 30° zu rechnen?

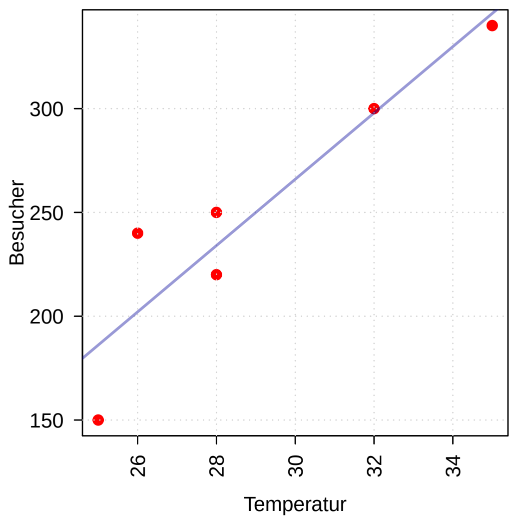
Lösungshinweis:

a) ~~Rangkorrelationskoeffizient~~ **Bravais-Pearson**

```
Besucher = c(340, 150, 250, 300, 240, 220)
Temperatur = c(35, 25, 28, 32, 26, 28)

# beide metrisch, also Bravais-Pearson
cor(Besucher, Temperatur)

## [1] 0,92216
```

b)

```
Regression <- lm(Besucher ~ Temperatur)
plot(Temperatur, Besucher, pch = 20, col = "red", cex = 1.5,
    las = 2)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```



Modell: $\hat{y} = -213{,}194 + 15{,}972 \cdot x$

c) Prognose: $\hat{y}(30) = -213{,}194 + 15{,}972 \cdot 30 = 265{,}97$

Aufgabe 32

Das Jahreseinkommen einiger Fußballnationalspieler ist zusammen mit der Anzahl der Tore, die sie in Länderspielen für Deutschland erzielen konnten in folgender Tabelle dargestellt:

Spieler Nummer	Jahreseinkommen [in Mio. €]	Anzahl Tore
1	1,1	0
2	0,8	1
3	2,3	3
4	4,2	3
5	1,7	1
6	0,9	0
7	3,7	4
8	0,7	1
9	2,8	2

a) Stellen Sie ein lineares Regressionmodell der Toranzahl in Abhängigkeit vom Spielereinkommen auf.
b) Geben Sie den Determinationskoeffizienten an und interpretieren Sie ihn.
c) Wieviel Tore würden Sie nach diesem Modell bei einem Einkommen von 10 Mio. erwarten?
d) Wieviel Tore müsste nach diesem Modell ein Spieler mehr schießen, wenn er 1 Mio. € mehr verdient?

Lösungshinweis:

a) ~~Rangkorrelationskoeffizient~~

```
Einkommen = c(1.1, 0.8, 2.3, 4.2, 1.7, 0.9, 3.7, 0.7, 2.8)
Tore = c(0, 1, 3, 3, 1, 0, 4, 1, 2)

Regression = lm(Tore ~ Einkommen)
a = Regression$coefficients[1]
b = Regression$coefficients[2]
```

Es ergibt sich: $\hat{y} = -0{,}218 + 0{,}932 \cdot x$

b)
```
R.Quadrat <- (cor(Tore, Einkommen))^2
```

Determinationskoeffizient: $R^2 = 0{,}7438$. Damit sind ca. 74 % der Streuung (Informationsgehalt) der gegebenen Daten durch das Modell erklärbar.

c) Für ein Einkommen von 10 ergibt sich: $\hat{y}(10) \approx -0{,}218 + 0{,}932 \cdot 10 \approx 9{,}103$.

d) Pro Million zusätzlichem Einkommen erhöht sich laut Modell die Toranzahl um $b \approx 0{,}932$, also fast um 1 Tor.

Aufgabe 33

Zu verschiedenen Zeitpunkten wird der Wasserstand x_t der Isar gemessen.

t	1	2	3	4	5	6
Wasserstand in cm	100	110	?	?	125	120

Die Messwerte zum Zeitpunkt 3 und 4 sind leider verloren gegangen. Es ist jedoch Folgendes bekannt:

$$\sum_{t=1}^{6} x_t = 705 \qquad \text{und} \qquad \sum_{t=1}^{6} x_t^2 = 83425$$

Außerdem ist bekannt, dass der Messwert zum Zeitpunkt 4 größer ist als der Messwert zum Zeitpunkt 3.

a) Ermitteln Sie die fehlenden Messwerte zum Zeitpunkt 3 und 4.
b) Prognostizieren Sie den Wasserstand zum Zeitpunkt $t = 7$ mittels einer linearen Regression.
c) Ermitteln Sie den Determinationskoeffizienten der Regression.
d) Begründen Sie kurz, ob Sie das Vorgehen aus Teilaufgabe b) für sinnvoll erachten.

Lösungshinweis:

a)
$$\sum_{i=1}^{6} x_t = 100 + 110 + x_3 + x_4 + 125 + 120 = 705 \quad \Leftrightarrow x_3 + x_4 = 250$$
$$\Leftrightarrow x_4 = 250 - x_3$$
$$\sum_{i=1}^{6} x_t^2 = 100^2 + 110^2 + x_3^2 + x_4^2 + 125^2 + 120^2 = 83425 \quad \Leftrightarrow x_3^2 + (250 - x_3)^2 = 31300$$

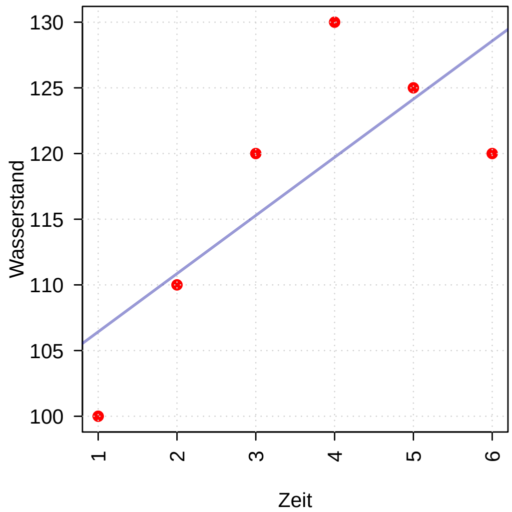
$$\Leftrightarrow \; 2x_3^2 - 500x_3 + 31200 = 0$$
$$\Leftrightarrow \; x_{3/4} = \frac{1}{4}\left(500 \pm \sqrt{25000 - 4 \cdot 2 \cdot 31200}\right) = \begin{cases} 120 \\ 130 \end{cases}$$
$$\Leftrightarrow \; x_3 = 120,\, x_4 = 130$$

b) $\hat{x}(t) \approx 102 + 4{,}429 \cdot x$. Damit ist $\hat{x}(7) \approx 102 + 4{,}429 \cdot 7 = 133$
c) $R^2 \approx 0{,}584$.
d) Ist vermutlich nicht sehr sinnvoll, nur den Trend als Grundlage der Prognose heranzuziehen; das Wasser könnte ja auch zyklisch steigen bzw. fallen.

(Lösung in R:)

```
Zeit = 1:6
Wasserstand = c(100,110,120,130,125,120)
Regression = lm(Wasserstand ~ Zeit)
a = Regression$coefficients[1]
b = Regression$coefficients[2]

# Prognose x(7):
x.7 = predict(Regression, data.frame(Zeit=7))

# Determinationskoeffizient:
R.Quadrat = (cor(Zeit, Wasserstand))^2
```

```
plot(Zeit, Wasserstand, pch=20, col="red", cex=1.5, las=2)
grid()
abline(Regression, col=rgb(0,0,0.6,0.4), lwd=2)
```



Aufgaben zur Kombinatorik

Aufgabe 34

Wie viele verschiedene Zusammenstellungen von genau 5 Buchstaben können aus den 26 Buchstaben des Alphabets gebildet werden, wenn Wiederholungen zulässig bzw. nicht zulässig sind?

Lösungshinweis:

- Mit Wiederholungen: $26^5 = 11881376$
- Ohne Wiederholungen: $\dfrac{26!}{(26-5)!} = 7893600$

Aufgabe 35

Bei der Beurteilung der Klangqualität von 10 Lautsprecher-Boxen ist in der Weise zu verfahren, dass die Tester jeweils zwei Boxen durch aufeinander folgendes Anhören miteinander vergleichen. Um die Objektivität der Tester zu überprüfen, soll auch jede Box mit sich selbst in der angegebenen Weise verglichen werden. Wie viele Hörvergleiche sind durchzuführen, wenn es auf die Reihenfolge, in der zwei Boxen angehört werden, nicht ankommt?

Lösungshinweis:

Mit Wiederholung, ohne Reihenfolge, $n = 10, k = 2$:

$$\binom{n+k-1}{k} = \binom{11}{2} = 55$$

Aufgabe 36

Ein Kartenspiel mit 32 verschiedenen Karten soll so unter 4 Spieler aufgeteilt werden, dass jeder genau 8 Karten erhält.

a) Wie viele unterschiedliche Möglichkeiten gibt es?
b) Wie viele Möglichkeiten gibt es, dass ein Spieler alle vier Asse erhält?
c) Bilden Sie den Quotienten des Ergebnisses von b) und a) und interpretieren Sie den erhaltenen Wert.

a) Annahme: Jeder Spieler bekommt 8 Karten auf einmal

Anzahl Mgl. für die 8 Karten Spieler 1 : $\binom{32}{8}$

Spieler 2 : $\binom{24}{8}$

Spieler 3 : $\binom{16}{8}$

Spieler 4 : $\binom{8}{8}$

Insgesamt $\binom{32}{8}\cdot\binom{24}{8}\cdot\binom{16}{8}\cdot\binom{8}{8} = 9{,}956 \cdot 10^{16}$

b) $\underbrace{\binom{4}{4}}_{\text{Anz. Mgl. dass Spieler 1 die 4 Asse bekommt}}\cdot\underbrace{\binom{28}{4}}_{\text{Anz. Mgl. für die restlichen 4 Nicht-As-Karten von Spieler 1}} \cdot \underbrace{\binom{24}{8}}_{\text{Sp. 2}} \cdot \underbrace{\binom{16}{8}}_{\text{Sp. 3}} \cdot \underbrace{\binom{8}{8}}_{\text{Sp. 4}} \cdot \underbrace{4}_{\text{Jeder könnte „Sp. 1" sein}} \approx 7{,}752 \cdot 10^{14}$

c) $\dfrac{4\cdot\binom{28}{4}}{\binom{32}{8}} = 0.00779 = 7{,}79\,‰$

Lösungshinweis:

a) $\binom{32}{8}\cdot\binom{24}{8}\cdot\binom{16}{8}\cdot\binom{8}{8} \approx 9{,}95611 \times 10^{16}$

b) $\underbrace{\binom{4}{4}\cdot\binom{28}{4}}_{\text{Anz.M. 1. Spieler 4 Asse}}\cdot\binom{24}{8}\cdot\binom{16}{8}\cdot\binom{8}{8}\cdot 4 \approx 7{,}75225 \times 10^{14}$

c) $\dfrac{\text{Antwort aus b)}}{\text{Antwort aus a)}} \approx 0{,}00779$

Aufgabe 37

Gegeben seien die Ziffern 1, 2, 3, 4, 5, 6, 7, 8, 9.

a) Wie viele dreistellige Zahlen können daraus gebildet werden, wenn jede Ziffer höchstens einmal vorkommen darf?
b) Wie viele der so gebildeten Zahlen sind gerade, wie viele ungerade?
c) Wie viele dieser Zahlen sind durch 5 teilbar?
d) Wie viele dieser Zahlen sind kleiner als 200 bzw. größer als 500?

a) $9 \cdot 8 \cdot 7 = 504$ ☐☐☐

b) gerade: $504 \cdot \frac{4}{9}$ $= 224$

4 → Endziffer „2", „4", „6", „8"

9 → alle Endziffern

ungerade: $504 \cdot \frac{5}{9}$ $= 280$

5 → „1", „3", „5", „7", „9"

(alternativ (von rechts nach links Ziffern auffüllen

gerade ☐☐☐ $7 \cdot 8 \cdot 4 = 224$

ungerade ☐☐☐ $7 \cdot 8 \cdot 5 = 280$)

c) letzte Ziffer muss „5" sein; damit $1 \cdot 8 \cdot 7 = 56$ Mgl.

d) kleiner 200: Erste Ziffer „1": $1 \cdot 8 \cdot 7 = 56$ Mgl.

größer 500 Erste Ziffer „5", „6", „7", „8", „9": $5 \cdot 8 \cdot 7 = 280$ Mgl.

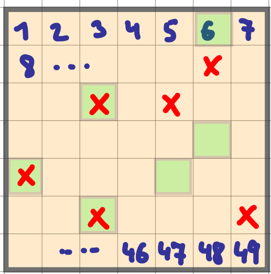
Lösungshinweis:

a) $9 \cdot 8 \cdot 7 = 504$
b) $7 \cdot 8 \cdot 4 = 224$
c) $7 \cdot 8 \cdot 1 = 56$
d) Kleiner als 200: $1 \cdot 8 \cdot 7 = 56$, größer als 500: $5 \cdot 8 \cdot 7 = 280$

Aufgabe 38

Wie viele Möglichkeiten gibt es, im Zahlenlotto „6 aus 49“ genau 3,4,5, beziehungsweise 6 richtige Zahlen anzukreuzen?

3 Richtige



Anzahl Mögl. 3 Kreuze auf 6 richtige Felder zusetzen (Auswahl von $k=3$ aus $n=6$ ohne Reihenfolge, ohne WH)

$$\binom{6}{3} = \frac{6\cdot 5\cdot 4}{1\cdot 2\cdot 3} = 20$$

Anzahl Mögl. die verbleibenden 3 Kreuze auf die $49-6=43$ falschen Felder zu verteilen:

$$\binom{43}{3} = \frac{43\cdot 42\cdot 41}{1\cdot 2\cdot 3} = 12341$$

insgesamt: $\binom{6}{3}\cdot\binom{43}{3} = 246820$

4 Richtige: $\binom{6}{4}\cdot\binom{43}{2} = 13545$

5 Richtige: $\binom{6}{5}\cdot\binom{43}{1} = 258$

6 Richtige: 1 Mgl.

richtig angekreuzt	Anzahl Möglichkeiten
3	$\binom{6}{3}\cdot\binom{43}{3} = 246820$
4	$\binom{6}{4}\cdot\binom{43}{2} = 13545$
5	$\binom{6}{5}\cdot\binom{43}{1} = 258$
6	$\binom{6}{6}\cdot\binom{43}{0} = 1$

Aufgabe 39

Eine Statistik-Klausur bestehe aus insgesamt 10 Aufgaben mit den (absteigend sortierten) Punktzahlen

$$22, 20, 16, \underbrace{12, 12}_{2!}, 10, \underbrace{8, 8}_{2!}, \underbrace{6, 6}_{2!} = 8 .$$

c)

Die Bearbeitung der einzelnen Aufgaben sei in beliebiger Reihenfolge zulässig.

a) Wie viele unterschiedliche Anordnungen (unterschiedliche Bearbeitungsreihenfolgen) gibt es, wenn alle Aufgaben bearbeitet werden?

b) Wie viele unterschiedliche Anordnungen (unterschiedliche Auswahlen der Aufgaben sowie unterschiedliche Bearbeitungsreihenfolgen) gibt es, wenn nur 5 Aufgaben bearbeitet werden?

c) Eine Studentin verfolgt die Strategie, die Aufgaben in absteigender Reihenfolge der erreichbaren Punktzahlen zu bearbeiten. Haben mehrere Aufgaben eine übereinstimmende Punktzahl, wählt Sie irgendeine Anordnung dieser Aufgaben. Wie viele unterschiedliche Bearbeitungsreihenfolgen zur Bearbeitung aller Aufgaben bleiben bei dieser Strategie möglich?

a) $10! = 3628800$ (alle Permutationen)

b) $\frac{10!}{(10-5)!} = 10 \text{ nPr } 5 = 30240$

ohne WH
mit RF

Lösungshinweis:

a) $10! = 3628800$ Möglichkeiten.

b) $\frac{10!}{(10-5)!} = 30240$ Möglichkeiten.

c) $2 \cdot 2 \cdot 2 = 8$ Möglichkeiten.

Aufgaben zur Wahrscheinlichkeitstheorie

Aufgabe 40

WTheorie: Laplace-Wahrscheinlichkeit (1)

Wie groß ist die Wahrscheinlichkeit, mit viermaligem Werfen eines Würfels

a) viermal 6
b) keine 6
c) mindestens eine 6
d) der Reihe nach 6,6,6,5
e) dreimal 6 und einmal 5
f) genau die Augensumme 7
g) mindestens zweimal die gleiche Zahl

zu erhalten?

$\Omega = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, \dots, 6\}\}$; $|\Omega| = 6^4 = 1296$

a) $A = \{(6,6,6,6)\} \Rightarrow P(A) = \frac{1}{1296} \approx 0{,}00077 \,\hat{=}\, 0{,}8$ ‰

b) $B = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, \dots, 5\}\} \Rightarrow |B| = 5^4 \Rightarrow P(B) = \frac{625}{1296} \approx 0.48$

c) $C = \overline{B} \Rightarrow P(C) = 1 - P(B) \approx 0.52$

d) $D = \{(6,6,6,5)\} \Rightarrow P(D) = \frac{1}{1296}$ (wie a)

e) $E = \{(6,6,6,5), (6,6,5,6), (6,5,6,6), (5,6,6,6)\}$

$\Rightarrow P(E) = \frac{4}{1296} \approx 0.0031$

f) $F = \{(1,2,2,2),$ ← 4 Perm. $\frac{4!}{3!} = 4$
$(1,1,2,3),$ ← 12 Perm. $\frac{4!}{2!} = 4 \cdot 3 = 12$
$(1,1,1,4)$ ← 4 Perm. $\frac{4!}{3!} = 4$

$\Rightarrow |F| = 4 + 12 + 4 = 20$

$\Rightarrow P(F) = \frac{20}{1296} \approx 0{,}0154$

g) $P(G) = 1 - P(\text{alle 4 sind verschieden})$
$= 1 - \frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4}$
$= 1 - \frac{5 \cdot 2}{36} = \frac{26}{36} = \frac{13}{18} \approx 0{,}72$

Lösungshinweis:

$\Omega = \{(x_1, x_2, x_3, x_4) \text{ mit } x_i = 1, \dots, 6\} \Rightarrow |\Omega| = 6^4 = 1296$

a) $A = \{(6,6,6,6)\} \Rightarrow |A| = 1 \Rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{1}{1296} \approx 0{,}00077$
b) $P(B) = \frac{5^4}{6^4} \approx 0{,}48225$
c) $P(C) = 1 - P(B) \approx 0{,}51775$
d) $P(D) = P(A)$
e) $P(E) = \frac{4}{1296} \approx 0{,}00309$
f)

1, 1, 1, 4	: 4 Permutationen
1, 1, 2, 3	: $\frac{4!}{2!} = 12$ Perm.
1, 2, 2, 2	: 4 Perm.
Summe	: 20 Möglichkeiten

$P(F) = \frac{20}{1296} = 0{,}01543$

g) $P(G) = 1 - P(\text{„Alle vier sind unterschiedlich“}) = 1 - \frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \frac{13}{18} \approx 0{,}72222.$

Aufgabe 41

In einem Raum befinden sich n Personen, von denen niemand am 29. Februar Geburtstag hat. Nehmen Sie weiterhin an, dass Sie selbst auch nicht am 29. Februar Geburtstag haben.

a) Sei $n = 3$. Wie groß ist die Wahrscheinlichkeit, dass mindestens zwei der drei Personen am gleichen Tag (Tag und Monat) Geburtstag haben?
b) Wie viele Leute müssen sich im Raum befinden, so dass die Wahrscheinlichkeit mindestens 50 % beträgt, dass mindestens zwei Personen am gleichen Tag Geburtstag haben?
c) Sei $n = 100$. Wie groß ist die Wahrscheinlichkeit, dass mindestens noch eine Person am selben Tag Geburtstag hat wie Sie selbst?
d) Wie viele Personen müssen sich im Raum befinden, so dass die Wahrscheinlichkeit mindestens 50 % beträgt, dass mindestens noch eine Personen am gleichen Tag Geburtstag hat wie Sie selbst?

Lösungshinweis:

```
P.1 = function(n){1- prod(365:(365-n+1)/365)}
P.2 = function(n){1- (364/365)^n}
```

TR: Table $f(x) = 1 - 365\,\boxed{nPr}\,x : 365^x$
from = 10 , to = 30 , step = 1

a) $P(3) = P(\text{„Mind. zwei von drei am gleichen Tag Geburtstag“})$
$= 1 - P(\text{„alle an verschiedenen Tagen Geburtstag“})$
$$= 1 - \frac{365 \cdot 364 \cdot 363}{365^3} = 0{,}0082$$

b) n Leute: $P(n) = 1 - \dfrac{365!}{(365-n)!} \cdot \dfrac{1}{365^n}$

```
options(digits=5)
n=1:20
df = data.frame(n,        P=sapply(n, P.1),
                n21=n+20, P=sapply(n+20, P.1),
                n41=n+40, P=sapply(n+40, P.1),
                n61=n+60, P=sapply(n+60, P.1))
print(df, row.names=FALSE)
```

```
##   n         P n21     P.1 n41     P.2 n61     P.3
##   1 0,0000000  21 0,44369  41 0,90315  61 0,99509
##   2 0,0027397  22 0,47570  42 0,91403  62 0,99591
##   3 0,0082042  23 0,50730  43 0,92392  63 0,99660
##   4 0,0163559  24 0,53834  44 0,93289  64 0,99719
##   5 0,0271356  25 0,56870  45 0,94098  65 0,99768
##   6 0,0404625  26 0,59824  46 0,94825  66 0,99810
##   7 0,0562357  27 0,62686  47 0,95477  67 0,99844
##   8 0,0743353  28 0,65446  48 0,96060  68 0,99873
##   9 0,0946238  29 0,68097  49 0,96578  69 0,99896
##  10 0,1169482  30 0,70632  50 0,97037  70 0,99916
##  11 0,1411414  31 0,73045  51 0,97443  71 0,99932
##  12 0,1670248  32 0,75335  52 0,97800  72 0,99945
##  13 0,1944103  33 0,77497  53 0,98114  73 0,99956
##  14 0,2231025  34 0,79532  54 0,98388  74 0,99965
##  15 0,2529013  35 0,81438  55 0,98626  75 0,99972
##  16 0,2836040  36 0,83218  56 0,98833  76 0,99978
##  17 0,3150077  37 0,84873  57 0,99012  77 0,99982
##  18 0,3469114  38 0,86407  58 0,99166  78 0,99986
##  19 0,3791185  39 0,87822  59 0,99299  79 0,99989
##  20 0,4114384  40 0,89123  60 0,99412  80 0,99991
```

Also: Ab 23 ist P größer als 50 %. im TR: $1-\left(\frac{364}{365}\right)^{100}$

c) $n = 100: \quad P(c)) = 1 - \dfrac{364^{100}}{365^{100}} = 0{,}23993$

d) $1 - \left(\frac{364}{365}\right)^n \geq 0{,}5 \quad \Leftrightarrow \quad n \geq \frac{\ln(0{,}5)}{\ln(364)-\ln(365)} \approx 252{,}65199$, also müssen mind. 253 Leute (außer Ihnen) noch im Raum sein.

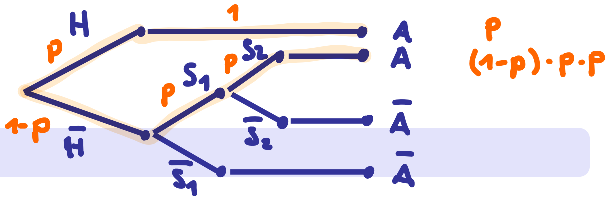
Aufgabe 42

Ein dreimotoriges Flugzeug stürzt ab, wenn der Hauptmotor in der Mitte ausfällt oder beide Seitenmotoren ausfallen. Es wird angenommen, dass jeder der Flugzeugmotoren mit der Wahrscheinlichkeit p auf einem bestimmten Flug ausfällt. Ferner wird angenommen, dass der Ausfall eines Motors unabhängig vom Verhalten der anderen Motoren erfolgt.

A bezeichne das Ereignis, dass ein Flugzeug dieses Typs infolge von Motorversagen abstürzt.

a) Ist die Wahrscheinlichkeit $P(A)$ größer oder kleiner als p? Bitte begründen Sie Ihre Antwort.

b) Bestimmen Sie $P(A)$ für $p = 0{,}01$.



Lösungshinweis:

$H \equiv$ Hauptmotor fällt aus, $S_{1/2} \equiv$ Seitenmotor 1 bzw. 2 fällt aus.

$P(A) = p + (1-p)pp$

a) $P(A) = P(H \cup (\overline{H} \cap S_1 \cap S_2)) = p + (1-p) \cdot p \cdot p > p$.

b) $p = 0{,}01$: $P(A) = 0{,}01 + 0{,}99 \cdot 0{,}01^2 = 0{,}0101$

Aufgabe 43

Ein Kraftfahrzeughändler weiß aus langjähriger Erfahrung, dass bei den in Zahlung genommenen Wagen 50% Mängel am Motor, 70% an der Karosserie und 30% an Motor und Karosserie aufweisen. Wie groß ist die Wahrscheinlichkeit dafür, dass ein in Zahlung genommener Wagen

a) ohne Mängel an Motor und Karosserie ist,
b) auch einen Mangel am Motor besitzt, wenn bekannt ist, dass die Karosserie schadhaft ist?

Lösungshinweis:

Abkürzungen: (M)otormangel; (K)arosserieschaden.

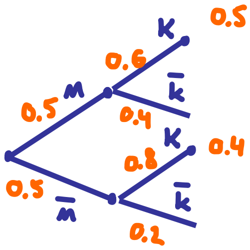
Gegeben: $P(M) = 0{,}5, \quad P(K) = 0{,}7, \quad P(M \cap K) = 0{,}3.$

0.5 · x = 0.3

Damit ergibt sich:

	K	$\overline{K}$	
M	0,30	0,20	0,50
$\overline{M}$	0,40	0,10	0,50
	0,70	0,30	

geg.



a) $P\left(\overline{M} \cap \overline{K}\right) = 0{,}10$

b) $P\left(M|K\right) = \dfrac{P(M \cap K)}{P(K)} = \dfrac{0{,}3}{0{,}7} = \dfrac{3}{7} \approx 0{,}42857$

Aufgabe 44

Der Bauer Bertram hat 3 Hühner (Erna, Lisa und Moni). Erna ist seine Lieblingshenne, denn sie liefert durchschnittlich 40 % aller pro Jahr gelegten Eier, während Lisa und Moni nur jeweils 30 % schaffen. Da die Eier ein Mindestgewicht haben müssen, gibt es einen gewissen Ausschuß (A). Bei Erna und Lisa beträgt er jeweils 3 % und bei Moni 5 %.

Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Ei

a) von Lisa stammt
b) zu klein ist
c) von Lisa stammt, wenn bekannt ist, dass es zu klein ist?

Lösungshinweis:

Aufgabe 45

Alexandra, Bernhard und Claudio sind als heilige drei Könige verkleidet von Haus zu Haus unterwegs. Bei jedem Haus lassen Sie den Zufall entscheiden, wer von den dreien ein Gedicht aufsagen darf. Dazu würfeln sie jeweils vorher einmal mit einem fairen Würfel. Alexandra sagt das Gedicht, wenn eine 1 fällt, Bernhard bei einer 2 oder 3 und Claudio darf bei 4, 5 oder 6 rezitieren. Alexandra sagt das Gedicht mit einer Wahrscheinlichkeit von 80 % perfekt (mit einer Wahrscheinlichkeit von 20 % ist mindestens ein kleiner Fehler dabei), Bernhard sagt das Gedicht mit einer Wahrscheinlichkeit von 90 % perfekt auf, Claudio mit einer Wahrscheinlichkeit von 95 %.

a) Mit welcher Wahrscheinlichkeit wird das Gedicht bei einem beliebigen Haus perfekt zum Vortrag gebracht?

b) Frau Maier erzählt am Tag nach dem Besuch der drei Ihrer Nachbarin, dass sich bei Ihr ein Sternsinger beim Gedicht ganz schön verhaspelt hätte. Mit welcher Wahrscheinlichkeit hat Bernhard das Gedicht bei Frau Maier aufgesagt?

Lösungshinweis:

Abkürzungen: $A \mathrel{\hat{=}}$ Alexandra sagt das Gedicht, analog (B)ernhard bzw. (C)laudio.
$F \mathrel{\hat{=}}$ Gedicht mit Fehler aufgesagt.

Gegeben: $P(A) = 1/6, \quad P(B) = 2/6, \quad P(C) = 3/6,$
$P(F|A) = 0{,}2, \quad P(F|B) = 0{,}1, \quad P(F|C) = 0{,}05$

a)
$$\begin{aligned} P(\overline{F}) &= 1 - P(F) \\ &= 1 - \left[P(F|A) \cdot P(A) + P(F|B) \cdot P(B) + P(F|C) \cdot P(C)\right] \\ &= 1 - \left[0{,}2 \cdot 1/6 + 0{,}1 \cdot 2/6 + 0{,}05 \cdot 3/6\right] \\ &= 1 - \tfrac{4+4+3}{120} \\ &= \tfrac{109}{120} \approx 0{,}90833 \end{aligned}$$

b)
$$P(B|F) = \frac{P(B \cap F)}{P(F)} = \frac{P(F|B) \cdot P(B)}{P(F)} = \frac{0{,}1 \cdot 2/6}{11/120} = \frac{4}{11} \approx 0{,}36364.$$

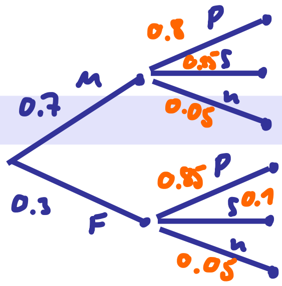
Aufgabe 46

In einer Großbank kommen 80 % der männlichen Kreditkunden ihren Kreditverpflichtungen pünktlich nach, 15 % schleppend nach, und bei 5 % muss die Bank den Kredit abschreiben. Bei den weiblichen Kreditkunden sind die entsprechenden Zahlen 85 %, 10 % und 5 %. Von den Kreditkunden der Bank sind 70 % männlich.

a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person ihren Kreditverpflichtungen pünktlich nachkommt?

b) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person weiblich ist, falls die Person ihren Kreditverpflichtungen nur schleppend nachkommt?

c) Sind die Ereignisse „Kunde ist männlich" und „Kunde zahlt pünktlich" stochastisch unabhängig?



b) $P(F|S)$

$= \frac{P(F \cap S)}{P(S)}$

$= \frac{0.03}{0.03 + 0.7 \cdot 0.15}$

Lösungshinweis:

Zahlungsmoral: (p)ünktlich, (s)chleppend, (n)ie.
Geschlecht: (M)ann, (F)rau.

Gegeben: $P(p|M) = 0{,}8, \quad P(s|M) = 0{,}15, \quad P(n|M) = 0{,}05,$

$P(p|F) = 0{,}85, \quad P(s|F) = 0{,}10, \quad P(n|F) = 0{,}05,$

$P(M) = 0{,}7, \quad P(F) = 0{,}3$

a) $P(p) = P(p|M) \cdot P(M) + P(p|F) \cdot P(F) = 0{,}8 \cdot 0{,}7 + 0{,}85 \cdot 0{,}3 = 0{,}815.$

b) $$P(F|s) = \frac{P(F \cap s)}{P(s)} = \frac{P(s|F) \cdot P(F)}{P(s|F) \cdot P(F) + P(s|M) \cdot P(M)}$$
$$= \frac{0{,}1 \cdot 0{,}3}{0{,}1 \cdot 0{,}3 + 0{,}15 \cdot 0{,}7} = \frac{2}{9} \approx 0{,}22222.$$

c) $P(M|p) = \frac{P(p|M) \cdot P(M)}{P(p)} = \frac{0{,}8 \cdot 0{,}7}{0{,}815} \approx 0{,}68712 \neq 0{,}7 = P(M),$
also sind die Ereignisse M und p nicht unabhängig.

Aufgabe 47

In der Stadt D wird im Mittel zu 10 % schwarz gefahren. 70 % der Schwarzfahrer haben keine Fahrkarte, während die anderen 30 % gefälschte oder illegal besorgte Karten besitzen. Von den ehrlichen Fahrgasten haben im Mittel 5 % ihre Fahrkarte vergessen. Mit welcher Wahrscheinlichkeit ist ein kontrollierter Fahrgast, der keine Karte vorzeigen kann, ein Schwarzfahrer?

Lösungshinweis:

Abkürzungen: (S)chwarzfahrer; zeigt (K)arte.

Gegeben: $P(S) = 0{,}1, \quad P\left(\overline{K}|S\right) = 0{,}7, \quad P(K|S) = 0{,}3,$
$P\left(\overline{K}|\overline{S}\right) = 0{,}05$

Lösung:

$$P\left(S|\overline{K}\right) = \frac{P\left(\overline{K}|S\right) \cdot P(S)}{P\left(\overline{K}\right)} = \frac{P\left(\overline{K}|S\right) \cdot P(S)}{P\left(\overline{K}|S\right) \cdot P(S) + P\left(\overline{K}|\overline{S}\right) \cdot P\left(\overline{S}\right)}$$
$$= \frac{0{,}7 \cdot 0{,}1}{0{,}7 \cdot 0{,}1 + 0{,}05 \cdot 0{,}9} = \frac{14}{23} \approx 0{,}6087$$

Aufgabe 48

10.000 Flugreisende, die aus einem südlichen Land nach Deutschland einreisen werden auf eine ansteckende tropische Krankheit getestet. Ein positiver Test deutet auf eine Erkrankung hin, allerdings nicht sicher. 9 Leute, bei denen der Test positiv ausgefallen ist sind tatsächlich krank. 9899 Leute mit negativem Testergebnissen sind nicht krank. Insgesamt war der Test bei 9900 Untersuchten negativ. Wie hoch ist die Wahrscheinlichkeit, dass bei einer beliebig unter diesen 10.000 Flugreisenden ausgewählten Person

a) Der Test positiv ausfällt,
b) die Person krank ist,
c) die Person gesund ist, obwohl der Test positiv ausgefallen ist,
d) der Test positiv ausfällt, wenn bekannt ist, dass die Person gesund ist.

Lösungshinweis:

Abkürzungen: Test ist (p)ositiv; Person hat (K)rankheit.

Gegeben: $P(K \cap p) = 0{,}0009, \quad P\left(\overline{p} \cap \overline{K}\right) = 0{,}9899, \quad P\left(\overline{p}\right) = 0{,}99.$

Damit ergibt sich:

	p	$\overline{p}$	
K	0,0009	0,0001	0,001
$\overline{K}$	0,0091	0,9899	0,999
	0,01	0,99	

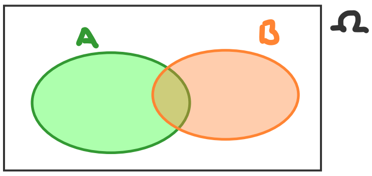
a) $P(p) = 1 - 0{,}99 = 0{,}01$
b) $P(K) = 0{,}001$
c) $P\left(\overline{K}|p\right) = \dfrac{0{,}0091}{0{,}01} = 0{,}91$
d) $P\left(p|\overline{K}\right) = \dfrac{0{,}0091}{0{,}999} \approx 0{,}00911$

Aufgabe 49

Geben Sie zu den Ereignissen A, B die bedingte Wahrscheinlichkeit $P(A|B)$ an, wenn

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

a) $A \subset B$,
b) $B \subset A$,
c) $A = \Omega$,
d) $B = \Omega$,
e) $A \cap B = \{\}$.



Lösungshinweis:

a) $\dfrac{P(A)}{P(B)}$,

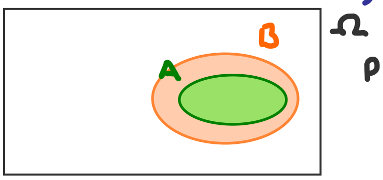
b) $\dfrac{P(B)}{P(B)} = 1$,

c) $\dfrac{P(B)}{P(B)} = 1$,

d) $\dfrac{P(A)}{1} = P(A)$,

e) $\dfrac{P(\{\})}{P(B)} = 0$.

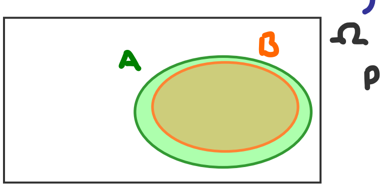
a)



hier: $A \cap B = A$

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

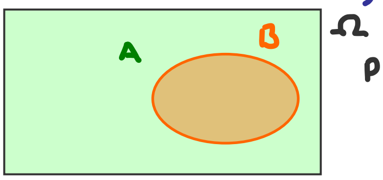
b)



hier: $A \cap B = B$

$$P(A|B) = \frac{P(B)}{P(B)} = 1$$

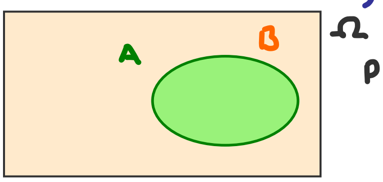
c)



hier: $A \cap B = B$

$$P(A|B) = \frac{P(B)}{P(B)} = 1$$

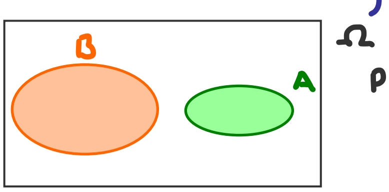
d)



hier: $A \cap B = A$

$$P(A|B) = \frac{P(A)}{\underbrace{P(B)}_{=1}} = P(A)$$

a)



hier: $A \cap B = \{\}$

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

Aufgabe 50

Ein Schießbudenbesitzer hat festgestellt, dass die Trefferwahrscheinlichkeit in den späten Abendstunden 0,1 pro Schuss beträgt.

a) Wie hoch ist die Wahrscheinlichkeit, bei 5 Schüssen mindestens 2 Treffer zu erzielen? R

b) Wie viele Schüsse sind notwendig, um mit einer Wahrscheinlichkeit von mindestens 0,9 mindestens einen Treffer zu erzielen?

$P(Y \geq 1) \geq 0.9$

$\Leftrightarrow \quad 1 - P(Y=0) \geq 0.9 \quad \Leftrightarrow \quad \ldots$

Lösungshinweis:

$X \mathrel{\hat{=}}$ Anzahl Treffer bei 5 Schüssen. Damit gilt: $X \sim \mathrm{B}(n = 5;\ p = 0{,}1)$

a) $P(X \geq 2) = 1 - P(X \leq 1) \approx 0{,}0815.$
mit R:

```
1 - pbinom(1, size = 5, prob = 0.1)
## [1] 0,08146
```

$n = 5$

↓x p →	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
0	0.9510	0.9039	0.8587	0.8154	0.7738	0.7339	0.6957	0.6591	0.6240	0.5905
1	0.9990	0.9962	0.9915	0.9852	0.9774	0.9681	0.9575	0.9456	0.9326	0.9185
2	1.0000	0.9999	0.9997	0.9994	0.9988	0.9980	0.9969	0.9955	0.9937	0.9914
3	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

b) $Y \mathrel{\hat{=}}$ Anzahl Treffer bei n Schüssen; damit

$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{n}{0} \cdot 0{,}1^0 \cdot 0{,}9^n = 1 - 0{,}9^n \geq 0{,}9$

$\Leftrightarrow\ 0{,}9^n \leq 0{,}1 \quad \Leftrightarrow \quad n \geq \dfrac{\ln 0{,}1}{\ln 0{,}9} \approx 21{,}85435$, also mindestens 22 Schuss sind nötig

Aufgabe 51

Eine binomialverteilte Zufallsvariable X habe einen Erwartungswert von 2 und eine Varianz von 4/3. Wie groß ist die Wahrscheinlichkeit für x=2?

Lösungshinweis:

```
E[X]    = np       = 2
Var[X] = np(1-p) = 4/3

=> p=1/3, n=6
```

$$P(X=2) = \binom{6}{2} \left(\tfrac{1}{3}\right)^2 \cdot \left(\tfrac{2}{3}\right)^4 = 0.3292$$

Aufgabe 52

In der Klausur zur Statistik werden 25 Multiple-Choice-Fragen gestellt mit jeweils vier Antwortmöglichkeiten, von denen genau eine richtige anzukreuzen ist. Wie wahrscheinlich ist es, mindestens 12 Punkte zu erhalten, wenn man nur rät? R

Lösungshinweis:

$X \widehat{=}$ Anzahl der richtig beantworteten Fragen

$X \sim B(n = 25, p = 0{,}25)$

$P(X \geq 12) = 1 - P(x \leq 11) \approx 0{,}01073$

$n = 25$

↓x p →	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25
0	0.7778	0.6035	0.4670	0.3604	0.2774	0.2129	0.1630	0.1244	0.0946	0.0718	0.0038	0.0008
1	0.9742	0.9114	0.8280	0.7358	0.6424	0.5527	0.4696	0.3947	0.3286	0.2712	0.0274	0.0070
2	0.9980	0.9868	0.9620	0.9235	0.8729	0.8129	0.7466	0.6768	0.6063	0.5371	0.0982	0.0321
3	0.9999	0.9986	0.9938	0.9835	0.9659	0.9402	0.9064	0.8649	0.8169	0.7636	0.2340	0.0962
4	1.0000	0.9999	0.9992	0.9972	0.9928	0.9850	0.9726	0.9549	0.9314	0.9020	0.4207	0.2137
5	1.0000	1.0000	0.9999	0.9996	0.9988	0.9969	0.9935	0.9877	0.9790	0.9666	0.6167	0.3783
6	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995	0.9987	0.9972	0.9946	0.9905	0.7800	0.5611
7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9989	0.9977	0.8909	0.7265
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9532	0.8506
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9827	0.9287
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9944	0.9703
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9985	0.9893
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Mit R:

```
1 - pbinom(11, size = 25, prob = 0.25)
## [1] 0,010734
```

Aufgabe 53

M N

Im Laufe eines Jahres werden von 52 aufeinanderfolgenden Ausgaben einer wöchentlich erscheinenden Zeitschrift 11 beliebige Ausgaben mit einer bestimmten Annonce versehen. Wie groß ist die Wahrscheinlichkeit, dass ein Leser von 20 beliebigen (aber verschiedenen) Ausgaben

n

a) zwei Ausgaben
b) keine Ausgabe
c) 20 Ausgaben
d) sämtliche 11 Ausgaben
e) mindestens eine Ausgabe

mit einer Annonce erhält?

Lösungshinweis:

$X \hat{=}$ Anzahl der Zeitschriften mit der Annonce, $X \sim \text{Hyp}(M = 11, N = 52, n = 20)$

a) $P(X = 2) = \frac{\binom{11}{2}\binom{41}{18}}{\binom{52}{20}} \approx 0{,}0882275$

b) $P(X = 0) = \frac{\binom{11}{0}\binom{41}{20}}{\binom{52}{20}} \approx 0{,}002136$

c) Das geht nicht, also $P(X = 20) = 0$.

d) $P(X = 11) = \frac{\binom{11}{11}\binom{41}{9}}{\binom{52}{20}} \approx 0{,}0000028$

e) $P(X \geq 1) = 1 - P(\text{„Teilaufgabe b)“}) = 0{,}997864$

Lösung in R:

```
a = dhyper(x=2, m=11, n=41, k=20)
b = dhyper(x=0, m=11, n=41, k=20)
c = dhyper(x=20, m=11, n=41, k=20)
d = dhyper(x=11, m=11, n=41, k=20)
e = 1 - dhyper(x=0, m=11, n=41, k=20)
print(data.frame(Aufgabe=c("a", "b", "c", "d", "e"),
                 Ergebnis=round(c(a,   b,   c,   d,   e), 7)),
      row.names=FALSE)

##  Aufgabe  Ergebnis
##        a 0,0882275
##        b 0,0021360
##        c 0,0000000
##        d 0,0000028
##        e 0,9978640
```

Aufgabe 54

Unter den 20 Passagieren eines Charterfluges befinden sich zwei Bewaffnete, die das Flugzeug entführen wollen. Zehn Passagiere werden zufällig ausgewählt und genau untersucht. Wie groß ist die Wahrscheinlichkeit dafür, dass die beiden Bewaffneten unentdeckt bleiben?

Lösungshinweis:

$X \mathrel{\hat{=}}$ Anzahl der entdeckten Bombenleger, $X \sim \text{Hyp}(M = 2, N = 20, n = 10)$

$$P(X = 0) = \frac{\binom{2}{0}\binom{18}{10}}{\binom{20}{10}} \approx 0{,}2368421$$

Lösung in R:

```
P <- dhyper(x = 0, m = 2, n = 18, k = 10)
P

## [1] 0,2368421
```

Aufgabe 55

Den drei Studentinnen Anna, Julia und Laura steht eine Statistikklausur bevor. Leider hatten die drei keine Zeit, die Vorlesung zu verfolgen. So verlassen Sie sich auf die Aussage des Dozenten, dass alle 5 Klausuraufgaben zufällig aus einer Liste von 50 veröffentlichten Aufgaben ausgewählt werden.

n N

Die Klausur ist bestanden, wenn mindestens vier Aufgaben richtig gelöst werden. Gehen Sie im Folgenden davon aus, dass eine nicht vorbereitete Aufgabe sicher falsch und eine vorbereitete Aufgabe sicher richtig gelöst wird.

Anna: $M = 25$ Laura: $M = 45$
Julia: $M = 40$

Alle drei wählen eine bestimmte Anzahl von Aufgaben zufällig und unabhängig voneinander aus und bereiten sich auf diese Fragen intensiv vor. Anna bereitet sich auf die Hälfte der Aufgaben vor. Julia geht davon aus, dass es reicht, sich auf vierzig der fünfzig Aufgaben vorzubereiten. Laura möchte nichts dem Zufall überlassen, schafft es aber wegen einer Krankheit nur sich auf 45 Aufgaben vorzubereiten.

a) Mit welcher Wahrscheinlichkeit bestehen die drei Kandidatinnen jeweils die Prüfung? R
b) Wie hoch ist die Wahrscheinlichkeit, dass alle drei bestehen? R
c) Wie hoch ist die Wahrscheinlichkeit, dass mindestens eine von den dreien besteht? R
d) Wie hoch ist die Wahrscheinlichkeit, dass Anna in drei Klausurversuchen (mit jeweils den gleichen Konditionen und den gleichen vorbereiteten Aufgaben wie beim ersten Versuch) durchfällt? Wie hoch ist diese Wahrscheinlichkeit bei Laura?

Lösungshinweis:

$X_{A,J,L} \mathrel{\hat{=}}$ Anzahl der richtig gelösten Aufgaben, entspricht der Anzahl der vom Prüfer ausgewählten Aufgaben, die (A)nna, (J)ulia, (L)aura jeweils vorbereitet haben.

$X_{A,J,L}$ ist jeweils hypergeometrisch verteilt.

a) $$P(X_A \geq 4) = \frac{\binom{25}{4}\cdot\binom{25}{1}}{\binom{50}{5}} + \frac{\binom{25}{5}\cdot\binom{25}{0}}{\binom{50}{5}} \approx 0{,}17434 = p_A$$
$$P(X_J \geq 4) = \frac{\binom{40}{4}\cdot\binom{10}{1}}{\binom{50}{5}} + \frac{\binom{40}{5}\cdot\binom{10}{0}}{\binom{50}{5}} \approx 0{,}7419 = p_J$$
$$P(X_L \geq 4) = \frac{\binom{45}{4}\cdot\binom{5}{1}}{\binom{50}{5}} + \frac{\binom{45}{5}\cdot\binom{5}{0}}{\binom{50}{5}} \approx 0{,}92825 = p_L$$

b) $p_A \cdot p_J \cdot p_L \approx 0{,}12006$

c) $1-(1-p_A)\cdot(1-p_J)\cdot(1-p_L) \approx 0{,}98471$

d) Anna: $(1-p_A)^3 \approx 0{,}56287$
Laura: $(1-p_L)^3 \approx 0{,}00037$

```r
# Lösung zu a)
pA <- sum(dhyper(4:5, m = 25, n = 25, 5))
pJ <- sum(dhyper(4:5, m = 40, n = 10, 5))
pL <- sum(dhyper(4:5, m = 45, n = 5, 5))
```

```r
c(pA, pJ, pL,                 # Lösung a)
pA * pJ * pL,                 # Lösung b)
1- (1-pA) * (1-pJ) * (1-pL)) # Lösung c)

## [1] 0,17434 0,74190 0,92825 0,12006 0,98471
```

Aufgabe 56

Das Rechenzentrum der Hochschule habe festgestellt, dass während einer Betriebszeit von einem Tag mit der Wahrscheinlichkeit 0,905 kein Ausfall des Systems zu verzeichnen ist. Die Anzahl der Systemausfälle sei Poisson-verteilt.

a) Bestimmen Sie den Parameter λ der Poisson-Verteilung.
b) Wie groß ist die Wahrscheinlichkeit, dass an einem Tag genau zwei Systemausfälle zu verzeichnen sind?
c) Berechnen Sie die Wahrscheinlichkeit, dass es bei 5 gleichartigen Systemen, die unabhängig voneinander laufen, zu mindestens einem Ausfall am Tage kommt.

R

Lösungshinweis:

a) $X \mathrel{\hat{=}}$ „Anzahl Ausfälle“ $\Rightarrow X \sim P(\lambda)$
$P(X=0) = \frac{\lambda^0}{0!} \cdot e^{-\lambda} = 0{,}905 \Leftrightarrow \lambda = -\ln 0{,}905 \approx 0{,}09982$

b) $P(X=2) = \frac{\lambda^2}{2!} \cdot e^{-\lambda} \approx 0{,}00451$

R

```
lambda = -log(0.905)
pb = dpois(2, lambda = lambda)
pb

## [1] 0,0045088
```

c) $Y \mathrel{\hat{=}}$ „Anzahl Systeme mit mind. einem Ausfall“
$\Rightarrow Y \sim B(n=5,\ p=P(X \geq 1) = 1-0{,}905 = 0{,}095)$
$P(Y \geq 1) = 1 - P(Y=0) = 1 - \binom{5}{0} \cdot p^0 \cdot (1-p)^5 = 1 - 0{,}905^5 \approx 0{,}39292$

Aufgabe 57

$\lambda = 3$

In einer Online-Redaktion weiß man, dass ein Webredakteur gemessen am output sehr wenige sprachliche Fehler produziert. Im Durchschnitt werden drei Fehler pro Monat festgestellt. Die Anzahl der Fehler pro Monat kann als Poisson-verteilt angenommen werden und ist jeweils unabhängig von den anderen Monaten.

Wie groß ist die Wahrscheinlichkeit,

a) dass der Redakteur mehr als 9 Fehler pro Monat begeht,
b) für mehr als 3 Fehler, wenn man weiß, dass er schon 2 Fehler gemacht hat,
c) dass er während eines Jahres in mindestens 3 Monaten keinen Fehler produziert?

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

Lösungshinweis:

$\downarrow x \quad \lambda \rightarrow$	2.5	2.75	3
0	0.0821	0.0639	0.0498
1	0.2873	0.2397	0.1991
2	0.5438	0.4815	0.4232
3	0.7576	0.7030	0.6472
4	0.8912	0.8554	0.8153
5	0.9580	0.9392	0.9161
6	0.9858	0.9776	0.9665
7	0.9958	0.9927	0.9881
8	0.9989	0.9978	0.9962
9	0.9997	0.9994	0.9989
10	0.9999	0.9999	0.9997

$X \hat{=}$ „Anzahl der Fehler pro Monat“, also $X \sim P(\lambda = 3)$

a) $P(X > 9) = 1 - P(X \leq 9) \approx 1 - 0{,}9989 \approx 0{,}0011$

b) $P(X > 3 | X \geq 2) = \dfrac{P(X > 3)}{P(X \geq 2)} = \dfrac{1 - P(X \leq 3)}{1 - P(X \leq 1)} = \dfrac{1 - 0{,}64723}{1 - 0{,}19915} \approx 0{,}44049$

c) $Y \hat{=}$ „Anzahl der Monate ohne Fehler in einem Jahr“, also $Y \sim B\big(n = 12; p = P(X = 0)\big)$, wobei $p = \frac{3^0}{0!} \cdot \mathrm{e}^{-3} \approx 0{,}04979$. Damit gilt:

$$\begin{aligned}
P(Y \geq 3) &= 1 - P(Y \leq 2) = 1 - \big[P(Y = 0) + P(Y = 1) + P(Y = 2)\big] \\
&= 1 - \Big[\tbinom{12}{0} p^0 (1-p)^{12} + \tbinom{12}{1} p^1 (1-p)^{11} + \tbinom{12}{2} p^2 (1-p)^{10}\Big] \\
&\approx 1 - \big[0{,}54182 + 0{,}34067 + 0{,}09817\big] \approx 0{,}01935
\end{aligned}$$

```r
Pa = 1 - ppois(9, lambda = 3)                       # Teilaufgabe a)
Pb = (1-ppois(3,lambda = 3)) / (1-ppois(1,lambda = 3)) # Teil b)
p = dpois(0, lambda = 3)                            # in c) benutzt
Pc = 1 - pbinom(2, size = 12, prob = p)             # Ergebnis c)
```

```r
c(Pa, Pb, p, Pc)
## [1] 0,0011025 0,4404912 0,0497871 0,0193476
```

Aufgabe 58

Die portugiesische Fußballnationalmannschaft schießt pro Spiel durchschnittlich 1 Tor. Gehen Sie davon aus, dass die Anzahl der Tore pro Spiel poissonverteilt ist. Wie groß ist die Wahrscheinlichkeit, dass die Mannschaft

a) in einem Spiel höchstens 1 Tor erzielt, R
b) in einem Spiel genau 7 Tore erzielt, R
c) während der Gruppenphase einer Fußball-WM (3 Spiele) mindestens einmal 7 Tore schießt? R

Lösungshinweis:

a) $X \hat{=}$ „Anzahl Tore bei einem Spiel", also $X \sim P(\lambda = 1)$. Damit:
$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1^0}{0!} \cdot \mathrm{e}^{-1} + \frac{1^1}{1!} \cdot \mathrm{e}^{-1} = 2 \cdot \mathrm{e}^{-1} \approx 0{,}73576$
(alternativ mit Tabelle)

R

```
ppois(1, lambda = 1)

## [1] 0,73576
```

b) $P(X = 7) = \frac{1^7}{7!} \cdot \mathrm{e}^{-1} \approx 0{,}00007$

R

```
dpois(7, lambda = 1)

## [1] 0,000072992
```

c) $Y \hat{=}$ „Anzahl der Spiele in Gruppenphase mit genau 7 Toren",
also $Y \sim B(n = 3, p)$ mit $p = P(X = 7) \approx 0{,}00007$. Damit ergibt sich:

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{3}{0} p^0 (1-p)^3 \approx 0{,}00022$$

R

```
p = dpois(7, lambda = 1)
1 - dbinom(0, size = 3, prob = p)

## [1] 0,00021896
```

Aufgabe 59

Johanna und Benedikt haben zu Beginn Ihres Studiums geheiratet, sich aber schon vor dem Abschluss wieder scheiden lassen. Im Verlauf der folgenden $n = 40$ Jahre begegnen sich die beiden jedoch häufiger wieder, wobei allerdings die Wahrscheinlichkeit evtl. wieder zu heiraten in jedem Jahr nur 1/200 beträgt. Wie groß ist dann (bei Unabhängigkeit) die Wahrscheinlichkeit, dass die beiden

a) nicht wieder heiraten (x=0)
b) noch einmal heiraten (x=1)?

R

Berechnen Sie die beiden Ergebnisse jeweils mit der Binomial- sowie der Poissonverteilung und beurteilen Sie die Abweichungen.

Lösungshinweis:

$X \sim B\left(n = 40; p = \frac{1}{200}\right), \qquad Y \sim P(\lambda = 0{,}2)$

$$\begin{aligned} P(X=0) &= \binom{40}{0}\left(\tfrac{1}{200}\right)^0 \cdot \left(1-\tfrac{1}{200}\right)^{40} \\ &\approx 0{,}81832 \\ P(X=1) &= \binom{40}{1}\left(\tfrac{1}{200}\right)^1 \cdot \left(1-\tfrac{1}{200}\right)^{39} \\ &\approx 0{,}16449 \end{aligned} \qquad \begin{aligned} P(Y=0) &= \frac{0{,}2^0}{0!} \cdot \mathrm{e}^{-0{,}2} \\ &\approx 0{,}81873 \\ P(Y=1) &= \frac{0{,}2^1}{1!} \cdot \mathrm{e}^{-0{,}2} \\ &\approx 0{,}16375 \end{aligned}$$

Abweichungen klein, Poisson-Approximation funktioniert hier prima.

```r
c(dbinom(0:1, size = 40, prob = 1/200),
  dpois(0:1, lambda = 40/200))

## [1] 0,81832 0,16449 0,81873 0,16375
```

Aufgabe 60

Ein Computerhersteller will eine neue Bestückungsmaschine für Platinen kaufen. Die Ausschußrate soll höchstens 5 % sein. Zur Kontrolle wird ein Probelauf mit 20 Platinen durchgeführt. Sind mehr als k Platinen fehlerhaft bestückt, so muss die Produktion gestoppt und kostenfrei nachgebessert werden.

a) Mit welcher Wahrscheinlichkeit erhält man bei einer tatsächlichen Ausschussrate von 5 % höchstens 3 fehlerhafte Platinen?

b) Wie muss die Zahl k gewählt werden, damit die Wahrscheinlichkeit für einen Produktionsstopp trotz ausreichender Ausschussrate kleiner als 10 % ist?

Lösungshinweis:

$X \mathrel{\hat{=}}$ „Anzahl fehlerhafter Platinen“, damit gilt: $X \sim B(n = 20; p = 0{,}05)$

a) $P(X \leq 3) = F(3) \approx 0{,}9841$

b) $P(X > k) < 0{,}1 \quad \Leftrightarrow 1 - P(X \leq k) < 0{,}1$
$\Leftrightarrow P(X \leq k) > 0{,}9 \quad \Leftrightarrow F(k) > 0{,}9$

k	$F(k)$	
0	0,35849	$< 0{,}9$
1	0,73584	$< 0{,}9$
2	0,92452	$> 0{,}9$

$\Rightarrow k$ muss mindestens 2 sein; d.h. Produktionsstop bei 3 oder mehr defekten Platinen.

Aufgabe 61

Die Gesamtdauer X eines Projektes wird als normalverteilt mit dem Parameter $\mu = 10$ (Wochen) angenommen. Ferner wird für die Wahrscheinlichkeit $P(8 \leq X \leq 12)$ der Wert 0,8 geschätzt. Man bestimme den Parameter σ.

$$P(8 \leq X \leq 12) = F(12) - F(8) = \Phi\left(\frac{12-10}{\sigma}\right) - \Phi\left(\frac{8-10}{\sigma}\right)$$

$$= \Phi\left(\tfrac{2}{\sigma}\right) - \Phi\left(-\tfrac{2}{\sigma}\right)$$

$$= \Phi\left(\tfrac{2}{\sigma}\right) - \left(1 - \Phi\left(\tfrac{2}{\sigma}\right)\right)$$

$$= 2\Phi\left(\tfrac{2}{\sigma}\right) - 1 = 0.8$$

$$\Leftrightarrow \quad \Phi\left(\tfrac{2}{\sigma}\right) = 0.9$$

$$\Rightarrow \quad \tfrac{2}{\sigma} \approx 1{,}28 \quad \Leftrightarrow \quad \sigma = \frac{2}{1{,}28} \approx 1{,}5625$$

$x_1 \backslash x_2$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147

Aufgabe 62

Das Körpergewicht X (in kg) zufällig ausgewählter Personen aus einer Grundgesamtheit sei normalverteilt mit den Parametern μ und σ. Es gilt:

$$P(X \leq 80) = \tfrac{1}{2} \qquad \text{und} \qquad P(X \leq 70) = \tfrac{1}{4}\,.$$

a) Geben Sie μ und σ an.

b) Berechnen Sie $P(X \geq 100)$.

c) Wieviel Prozent der Personen der Grundgesamtheit, die mindestens 100 kg wiegen, wiegen über 110 kg?

a) $P(X \leq 80) = 0.5 \Rightarrow \mu = 80$

$P(X \leq 70) = F(70) = \Phi\left(\frac{70-80}{\sigma}\right) = \Phi\left(-\frac{10}{\sigma}\right) = 0.25$

$\Leftrightarrow \quad 1 - \Phi\left(\frac{10}{\sigma}\right) = 0.25 \quad \Leftrightarrow \quad \Phi\left(\frac{10}{\sigma}\right) = 0.75$

$\Leftrightarrow \quad \frac{10}{\sigma} \approx 0.67 \quad \Leftrightarrow \quad \sigma \approx \frac{10}{0.67} \approx 14{,}93$

b) $P(X \geq 100) = 1 - P(X \leq 100) = 1 - \Phi\left(\frac{100-80}{14{,}93}\right) = 1 - \Phi(1{,}34)$

$= 1 - 0.90988 = 0.09012$

c) $P(X \geq 110 \mid X \geq 100) = \dfrac{P(X \geq 110)}{P(X \geq 100)} = \dfrac{1 - \Phi\left(\frac{110-80}{14{,}93}\right)}{0.09012} = \dfrac{1 - \Phi(2{,}01)}{0.09012}$

$= \dfrac{1 - 0.97778}{0.09012} \approx 0.2466$

$x_1 \setminus x_2$	0	0.01	0.02	0.03	0.04
0	0.50000	0.50399	0.50798	0.51197	0.51595
0.1	0.53983	0.54380	0.54776	0.55172	0.55567
0.2	0.57926	0.58317	0.58706	0.59095	0.59483
0.3	0.61791	0.62172	0.62552	0.62930	0.63307
0.4	0.65542	0.65910	0.66276	0.66640	0.67003
0.5	0.69146	0.69497	0.69847	0.70194	0.70540
0.6	0.72575	0.72907	0.73237	0.73565	0.73891
0.7	0.75804	0.76115	0.76424	0.76730	0.77035
0.8	0.78814	0.79103	0.79389	0.79673	0.79955
0.9	0.81594	0.81859	0.82121	0.82381	0.82639
1	0.84134	0.84375	0.84614	0.84850	0.85083
1.1	0.86433	0.86650	0.86864	0.87076	0.87286
1.2	0.88493	0.88686	0.88877	0.89065	0.89251
1.3	0.90320	0.90490	0.90658	0.90824	0.90988
1.4	0.91924	0.92073	0.92220	0.92364	0.92507

Aufgabe 63

Die Zufallsvariable X beschreibt den täglichen Umsatz in einer Eisdiele. Es wird angenommen, dass $X \sim N(\mu, \sigma)$ gilt. Außerdem sei bekannt, dass der Umsatz an 30,854 % der Tage mindestens 1500 € und an 30,854 % der Tage weniger als 900 € beträgt.

a) Bestimmen Sie μ und σ.
b) Wie groß ist die Wahrscheinlichkeit, dass der Umsatz an einem Tag mehr als 3000 € beträgt?
c) Wie hoch müsste der Umsatz mindestens gemäß der Verteilung an den 5 % besten Tagen sein?
d) Wie groß ist die Wahrscheinlichkeit, dass der Umsatz negativ ist? Was bedeutet das?

$$P(X \geq 1500) = 0.30854 = P(X < 900)$$

a)
$$\left.\begin{aligned} 1 - P(X \leq 1500) &= 0.30854 \\ P(X \leq 900) &= 0.30854 \end{aligned}\right\}$$

$$\Leftrightarrow \left\{\begin{aligned} 1 - F(1500) &= 0.30854 \\ F(900) &= 0.30854 \end{aligned}\right.$$

$$\Leftrightarrow \left\{\begin{aligned} 1 - \phi\left(\frac{1500-\mu}{\sigma}\right) &= 0.30854 \\ \phi\left(\frac{900-\mu}{\sigma}\right) &= 0.30854 \end{aligned}\right.$$

$$\Leftrightarrow \left\{\begin{aligned} \phi\left(\frac{1500-\mu}{\sigma}\right) &= 0.69146 \\ \phi\left(\frac{\mu-900}{\sigma}\right) &= \text{"} \end{aligned}\right.$$

$$\Leftrightarrow \left\{\begin{aligned} 1500 - \mu &= \sigma \cdot 0.5 \quad (1) \\ -900 + \mu &= \sigma \cdot 0.5 \quad (2) \end{aligned}\right.$$

$$\Leftrightarrow (1) + (2): \quad 600 = \sigma, \quad \mu = 1200$$

sehr ausführlich, nicht elegant

$$\phi(x) = 1 - \phi(-x)$$

b)
$$P(X > 3000) = 1 - F(3000) = 1 - \phi\left(\frac{3000-1200}{600}\right)$$
$$= 1 - \phi(3) = 0.0013$$

c)
$$P(X \geq x) = 0.05$$
$$\Leftrightarrow F(x) = 0.95$$
$$\Leftrightarrow \phi\left(\frac{x-1200}{600}\right) = 0.95$$
$$\Leftrightarrow \frac{x-1200}{600} = 1.64$$
$$\Leftrightarrow x = 2184$$

d)
$$P(X < 0) = F(0) = \phi(-2) = 1 - \phi(2) = 0.0227$$

(Modell funktioniert nicht für $x < 0$)

Aufgabe 64

Eine normalverteilte Zufallsvariable X soll untersucht werden. Zwei Tatsachen sind von X bekannt:

- $P(X > 20) = 20\%$
- $P(X < 1) = 1\%$

Bestimmen Sie damit:

a) $\mathrm{Sta}[X] = \sigma$ sowie $\mathrm{E}[X] = \mu$

b) $P(X = 20)$

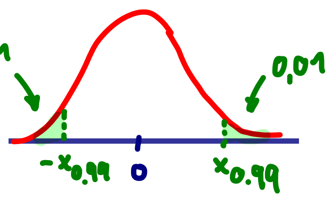
c) $P(X \leq 20)$

d) $P(X \geq 25)$

e) $P(X \geq 25 | X \geq 20)$

f) $P(X \geq 20 | X \geq 25)$

$$\phi(x) = 1 - \phi(-x)$$



a) $\left.\begin{array}{l} P(X>20) = 0.2 \\ P(X<1) = 0.01 \end{array}\right\} \left.\begin{array}{l} 1-F(20) = 0.2 \\ F(1) = 0.01 \end{array}\right\} \left.\begin{array}{l} 1-\phi\left(\frac{20-\mu}{\sigma}\right) = 0.2 \\ \phi\left(\frac{1-\mu}{\sigma}\right) = 0.01 \end{array}\right\} \begin{array}{l} \phi\left(\frac{20-\mu}{\sigma}\right) = 0.8 \\ \phi\left(\frac{\mu-1}{\sigma}\right) = 0.99 \end{array}$

$\left.\begin{array}{l} 20-\mu = 0.84\cdot\sigma \ ① \\ \mu-1 = 2{,}33\cdot\sigma \ ② \end{array}\right\}$ ① + ②: $19 = (0.84+2{,}33)\sigma$

$\Rightarrow \sigma = 5{,}9937$

$\Rightarrow \mu = 2{,}33\cdot\sigma + 1 = 14{,}965$

b) $P(X=20) = 0$

c) $P(X \leq 20) = 1-0.2 = 0.8$

d) $P(X \geq 25) = 1 - \phi\left(\frac{25-14{,}965}{5{,}9937}\right) = 1-\phi(1{,}67) = 1-0.95254 = 0{,}04746$

e) $P(X \geq 25 \mid X \geq 20) = 0{,}04746/0.20 = 0{,}2373$

f) $P(X \geq 20 \mid X \geq 25) = 1$

Aufgabe 65

Schokoladennikoläuse mit einem Sollgewicht von 200g sollen bzgl. ihres Gewichts kontrolliert werden. Es stellt sich heraus, dass

- das Gewicht X der Nikoläuse normalverteilt ist,
- die Wahrscheinlichkeit, dass ein Nikolaus mindestens 200g wiegt bei 30 % liegt und
- ein Nikolaus mit einer Wahrscheinlichkeit von 99 % höchstens 210g wiegt.

Berechnen Sie bzw. geben Sie ohne Rechnung aber mit Begründung an:

a) Wie groß ist die Standardabweichung σ sowie der Erwartungswert μ von X?

b) Wie groß ist die Wahrscheinlichkeit, zufällig einen Nikolaus mit einem Gewicht von exakt 200g (±0g) auszuwählen?

c) Wie groß ist die Wahrscheinlichkeit, dass ein Nikolaus weniger als 190g wiegt?

d) Nikoläuse mit weniger als 195g werden aussortiert. Wie groß ist die Wahrscheinlichkeit, dass ein Nikolaus aus diesem Ausschuss zwischen 190g und 195g wiegt?

a) $X \sim N(\mu; \sigma)$

$$\left.\begin{array}{l} P(X \geq 200) = 0.30 \\ P(X \leq 210) = 0.99 \end{array}\right\} \quad \left.\begin{array}{l} 1 - F(200) = 0.30 \\ F(210) = 0.99 \end{array}\right\} \quad \begin{array}{l} F(200) = 0.70 \\ F(210) = 0.99 \end{array}$$

$$\left.\begin{array}{l} \Phi\left(\frac{200-\mu}{\sigma}\right) = 0.70 \\ \Phi\left(\frac{210-\mu}{\sigma}\right) = 0.99 \end{array}\right\} \quad \begin{array}{l} \frac{200-\mu}{\sigma} \approx 0.52 \\ \frac{210-\mu}{\sigma} \approx 2{,}33 \end{array}$$

$x_1 \backslash x_2$	0	0.01	0.02	0.03
0	0.50000	0.50399	0.50798	0.51197
0.1	0.53983	0.54380	0.54776	0.55172
0.2	0.57926	0.58317	0.58706	0.59095
0.3	0.61791	0.62172	0.62552	0.62930
0.4	0.65542	0.65910	0.66276	0.66640
0.5	0.69146	0.69497	0.69847	0.70194
0.6	0.72575	0.72907	0.73237	0.73565
0.7	0.75804	0.76115	0.76424	0.76730
0.8	0.78814	0.79103	0.79389	0.79673
0.9	0.81594	0.81859	0.82121	0.82381
1	0.84134	0.84375	0.84614	0.84850
1.1	0.86433	0.86650	0.86864	0.87076
1.2	0.88493	0.88686	0.88877	0.89065
1.3	0.90320	0.90490	0.90658	0.90824
1.4	0.91924	0.92073	0.92220	0.92364
1.5	0.93319	0.93448	0.93574	0.93699
1.6	0.94520	0.94630	0.94738	0.94845
1.7	0.95543	0.95637	0.95728	0.95818
1.8	0.96407	0.96485	0.96562	0.96638
1.9	0.97128	0.97193	0.97257	0.97320
2	0.97725	0.97778	0.97831	0.97882
2.1	0.98214	0.98257	0.98300	0.98341
2.2	0.98610	0.98645	0.98679	0.98713
2.3	0.98928	0.98956	0.98983	0.99010

$$\left.\begin{array}{ll} 200 - \mu = 0.52 \cdot \sigma & (1) \\ 210 - \mu = 2.33 \cdot \sigma & (2) \end{array}\right.$$

(2) − (1): $10 = (2.33 - 0.52)\,\sigma$

$\Leftrightarrow \sigma = \frac{10}{1.81} \approx 5{,}525$

in (2): $\mu = 210 - 2.33 \cdot 5.525 = 197{,}13$

b) $P(X = 200) = 0$

c) $P(X < 190) = F(190) = \Phi\left(\frac{190 - 197{,}13}{5{,}525}\right)$

$\approx \Phi(-1{,}29) = 1 - \Phi(1{,}29) \approx 1 - 0.90147$
$= 0{,}09853$

d) $P(190 \leq X \leq 195 \mid X \leq 195) = \frac{P(190 \leq X \leq 195)}{P(X \leq 195)}$

$= \frac{F(195) - F(190)}{F(195)} = 1 - \frac{F(190)}{F(195)} \overset{c)}{=} 1 - \frac{0.09853}{\Phi\left(\frac{195-197{,}13}{5{,}525}\right)}$

$= 1 - \frac{0.09853}{\underbrace{1 - \Phi(+0.39)}_{0.65173}} = 1 - 0.09853 : 0.34827 \approx 0.717$

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

Aufgabe 66

Die zufallsabhängige Nachfrage X nach einem Gut in einer Zeitperiode ist gemäß der folgenden Wahrscheinlichkeitsfunktion verteilt:

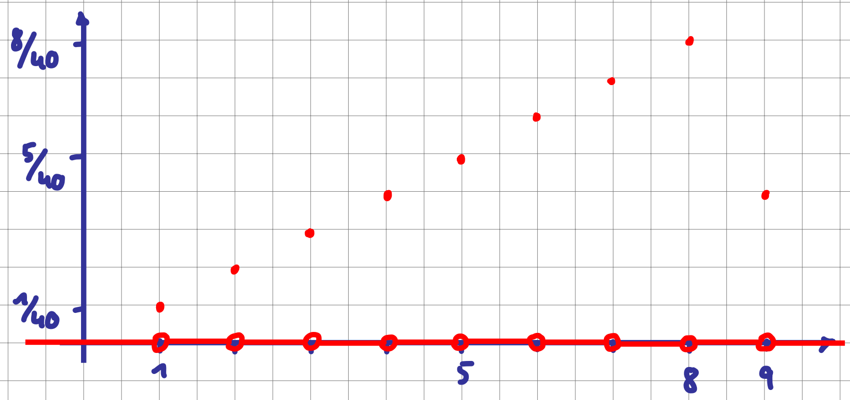
$$P(x = n) = \frac{n}{40} \quad \text{für } n = 1, \dots, 8 \quad \text{und} \quad P(x = 9) = \frac{1}{10}$$

a) Skizzieren Sie den Verlauf der Wahrscheinlichkeitsfunktion.
b) Mit welcher Wahrscheinlichkeit werden mindestens 6 Stück des Gutes nachgefragt?
c) Berechnen Sie den Erwartungswert und die Standardabweichung der Nachfrage.

Lösungshinweis:

x	1	2	3	4	5	6	7	8	9
$P(X=x)$	$1/40$	$2/40$	$3/40$	$\dots$			$\dots$	$8/40$	$1/10$

a)



b) $P(X \geq 6) = P(X \in \{6,7,8,9\}) = 6/40 + 7/40 + 8/40 + 1/10 = 25/40 = 5/8 = 0.625 = 62{,}5\%$

c) $E[X] = 1 \cdot 1/40 + 2 \cdot 2/40 + \dots + 8 \cdot 8/40 + 9 \cdot 1/10 = 6$

$\mathrm{Sta}[X] = \sqrt{\mathrm{Var}[X]} = \sqrt{E[X^2] - E^2[X]}$

$= \sqrt{1^2 \cdot 1/40 + 2^2 \cdot 2/40 + \dots + 8^2 \cdot 8/40 + 9^2 \cdot 1/10 - 6^2} = \sqrt{4.5} \approx 2.12$

Aufgabe 67

Die Lebensdauer einer Maschine sei eine über dem Zeitintervall [0,65] gleichverteilte Zufallsvariable. Berechnen Sie

a) den Erwartungswert der Lebensdauer
b) die Varianz der Lebensdauer
c) die Wahrscheinlichkeit, dass die Lebensdauer zwischen 13 und 39 liegt.

Lösungshinweis:

a) $E[X] = \frac{65+0}{2} = 32.5$

b) $Var[X] = \frac{(b-a)^2}{12} = \frac{65^2}{12} = 352.08$

c) $P(13 < X < 39) = \frac{39-13}{65} = 0.4$

Aufgabe 68

Eine Unternehmung sieht sich auf dem Absatzmarkt zufällig schwankender Nachfrage gegenübergestellt. Die Höhe der Nachfrage X sei folgendermaßen verteilt:

$$f(x) = \begin{cases} \frac{1}{12}, & \text{für } 0 \leq x \leq 12 \\ 0, & \text{sonst} \end{cases}$$

Die Produktion der Unternehmung wird unmittelbar abgesetzt, d.h. es existieren keine Absatzlager. Die Kostenfunktion der Unternehmung lautet $Y = 2X + 10$. Man gebe den Erwartungswert und die Varianz der Kosten an.

Lösungshinweis:

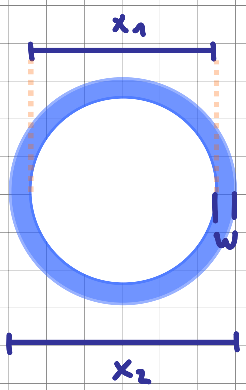
$$E[Y] = E[2X+10] = 2\,E[X] + 10 = 2 \cdot \frac{0+12}{2} + 10 = 22$$

$$Var[Y] = Var[2X+10] = 2^2\, Var[X] = 4 \cdot \frac{(12-0)^2}{12} = 48$$

Aufgabe 69

Ein Röhrenwerk produziert Stahlröhren, deren Durchmesser produktionsbedingten Schwankungen unterliegen. Für den Innendurchmesser X_1 hat man $\mathrm{E}(X_1) = 800$ und $\mathrm{Var}(X_1) = 0{,}01$ und für den Außendurchmesser X_2 hat man $\mathrm{E}(X_2) = 810$ und $\mathrm{Var}(X_2) = 0{,}02$. Bestimmen Sie Erwartungswert und Varianz für die Wandstärke der Röhren, wenn angenommen werden kann, dass Innen- und Außendurchmesser voneinander unabhängig schwanken.

Lösungshinweis:



Wandstärke $\hat{=}$ W

$$W = \tfrac{1}{2}(X_2 - X_1)$$

$$E[W] = E[\tfrac{1}{2}(X_2 - X_1)] = \tfrac{1}{2}(E[X_2] - E[X_1]) = \tfrac{1}{2}(810 - 800) = 5$$

da X_1, X_2 unabhängig

$$Var[W] = Var[\tfrac{1}{2}(X_2 - X_1)] = \tfrac{1}{4}(Var[X_2] + Var[-X_1]) = \tfrac{1}{4}(Var[X_2] + (-1)^2 Var[X_1]) = \tfrac{1}{4}(0.02 + 0.01) = 0.0075$$

$$Var[a + bX] = b^2 \cdot Var[X]$$

Aufgabe 70

Der Besitzer eines Zeitschriftenladens hat für einen längeren Zeitraum in der Vergangenheit folgende tägliche Nachfrageverteilung nach einer bestimmten Tageszeitung beobachtet:

pro Tag nachgefragte Exemplare	0	1	2	3	4	> 4
Nachfragewahrscheinlichkeit	0,20	0,30	0,20	0,20	0,10	0

Er rechnet für die Zukunft mit keiner Änderung der Nachfrageverteilung. Der Einkaufspreis eines Exemplars beträgt 0,50 €, der Verkaufspreis 1,50 €. Unverkaufte Exemplare können nicht zurückgegeben werden. Für einen längeren Zeitraum muss er eine feste Zahl von Zeitungen pro Tag bestellen. Wie viele Zeitungen pro Tag sollte er bestellen, um seinen erwarteten Gewinn zu maximieren?

Lösungshinweis:

$k \hat{=}$ Anzahl Zeitschriften, die pro Tag bestellt werden

$G \hat{=}$ Gewinn pro Tag

$k=0$: $G=0$

$k=1$: $G = \begin{cases} -0{,}5€ & \text{mit W. } 0.2 \\ 1{,}0€ & \text{mit W. } 0.8 \end{cases}$

$\Rightarrow E[G] = -0.5€ \cdot 0.2 + 1{,}0€ \cdot 0.8 = 0.7€$

$k=2$: $G = \begin{cases} -1€ & \text{mit W. } 0.2 \\ 0{,}5€ & \text{mit W. } 0.3 \\ 2{,}0€ & \text{mit W. } 0.5 \end{cases}$

optimal

$E[G] = -1€ \cdot 0.2 + 0{,}5€ \cdot 0.3 + 2{,}0€ \cdot 0.5 = 0.95€$

$k=3$: $E[G] = -1{,}5€ \cdot 0.2 + 0€ \cdot 0.3 + 1{,}5€ \cdot 0{,}2 + 3{,}0€ \cdot 0{,}3 = 0.90€$

$k=4$: $E[G] = -2€ \cdot 0.2 + \cdots + 4€ \cdot 0.1 = 0.55€$

Aufgabe 71

Die Versicherungsgesellschaft *Ollionz* verlangt als Prämie das 1,3-fache des Erwartungswertes ihrer Zahlungen an den Versicherungsnehmer. Es werden einjährige Lebensversicherungen des folgenden Typs betrachtet:

Der Betrag von 100.000 € ist zu zahlen, wenn der Versicherungsnehmer innerhalb eines Jahres nach Abschluss des Vertrages stirbt. Im Erlebensfall ist keine Zahlung zu leisten.

Die einjährige Sterbewahrscheinlichkeit betrage 0,006.

a) Wie hoch ist die Prämie für einen derartigen Vertrag?

b) Betrachten Sie den aus einem solchen Vertrag resultierenden Gewinn G. Wie hoch sind der Erwartungswert $E[G]$ und die Varianz $\mathrm{Var}[G]$ des Gewinns?

c) Berechnen Sie die Standardabweichung σ sowie das Intervall $[\mu - 3\sigma; \mu + 3\sigma]$ (den Drei-Sigma-Bereich). Interpretieren Sie dieses Intervall.

d) Die Versicherungsgesellschaft schließt 50.000 derartige Verträge ab. Bestimmen Sie den Drei-Sigma-Bereich des durchschnittlichen Gewinns pro Vertrag

$$\overline{G} = \frac{1}{50.000}\left[G_1 + G_2 + \cdots + G_{50.000}\right],$$

der aus den Verträgen resultiert und interpretieren Sie diese Zahl.

Hinweis: Setzen Sie die Unabhängigkeit der Einzelgewinne G_i voraus.

a) $P = 1{,}3 \cdot E[\text{Zahlung}] = 1{,}3 \cdot [100000 \cdot 0.006 + 0 \cdot 0.994]$

$= 780\,€$

b) Gewinn $G = \begin{cases} 780 & \text{mit W. } 0.994 \\ 780 - 100000 & \text{mit W. } 0.006 \end{cases}$

$E[G] = 780 \cdot 0.994 + (780 - 100') \cdot 0.006$
$= 180\,€$

Auszahlung

$\mathrm{Var}[G] = \mathrm{Var}[780 - Z] = \mathrm{Var}[Z] = E[Z^2] - E^2[Z]$

$\mathrm{Var}[a + bX] = b^2 \mathrm{Var}[X]$

$= 100000^2 \cdot 0.006 \; (+\; 0^2 \cdot 0.994) - 600^2$

$= 59.640.000$

c) $\sigma = \sqrt{\mathrm{Var}[G]} \approx 7722{,}69\,€$

3-Sigma-Bereich:

$[180 \pm 3 \cdot 7722{,}69] = [-22989{,}08 ; 23349{,}07]$

→ sehr riskant, Verluste wahrscheinlich

d) $\mathrm{Var}\left[\frac{1}{50000} \cdot (G_1 + \dots + G_{50000})\right]$

$= \left(\frac{1}{50000}\right)^2 \left[\underbrace{\mathrm{Var}[G_1]}_{=59640000} + \dots + \underbrace{\mathrm{Var}[G_{50000}]}_{= \dots}\right]$

$= \frac{50000 \cdot 59640000}{50000^2} = \frac{59640000}{50000} = 1192{,}80$

$\Rightarrow \sigma = 34{,}54$

$\Rightarrow$ 3-Sigma: $[180 \pm 3 \cdot 34{,}54] = [76{,}38 ; 283{,}62]$

Aufgabe 72

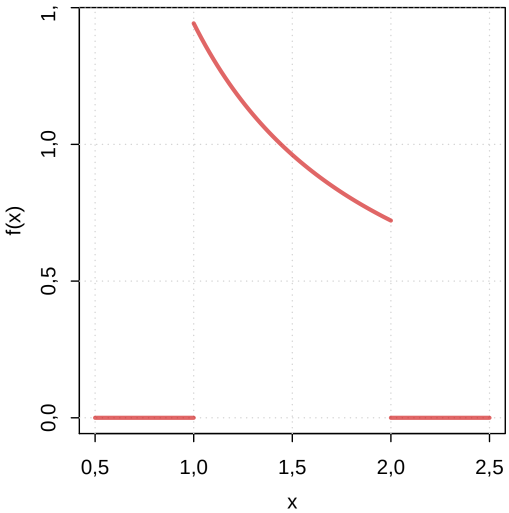
Gegeben ist zur Zufallsvariablen X die Dichtefunktion f gemäß

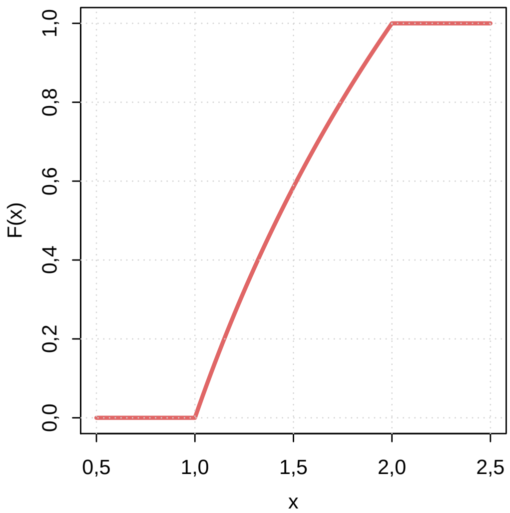
$$f(x) = \begin{cases} \dfrac{1}{x \cdot \ln 2} & \text{für} \quad 1 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

a) Skizzieren Sie den Graph von f.
b) Bestimmen Sie zu X die Verteilungsfunktion $F(x) = P(X \leq x)$ und skizzieren Sie auch deren Graph.
c) Berechnen Sie $P(1.2 \leq X \leq 1.8)$.
d) Berechnen Sie $\mathrm{E}[X]$ sowie
e) $\mathrm{Var}[X]$.

Lösungshinweis:

a) Graph von f:





$$E[x] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$
$$= \int_1^2 \cancel{x} \cdot \frac{1}{\cancel{x} \cdot \ln 2}\,dx$$
$$\approx \int_1^2 \frac{1}{0.69}\,dx = \left[\frac{1}{0.69} \cdot x\right]_1^2$$
$$= \frac{1}{0.69} \cdot (2-1) \approx 1.44$$

b) $$F(x) = \begin{cases} 0 & \text{für} \quad x < 1 \\ \dfrac{\ln(x)}{\ln(2)} & \text{für} \quad 1 \leq x \leq 2 \\ 1 & \text{für} \quad x > 2 \end{cases}$$

(Skizze siehe oben)

c) $P(1.2 \leq X \leq 1.8) = F(1.8) - F(1.2) = \dfrac{\ln(1.8) - \ln(1.2)}{\ln(2)} = \dfrac{\ln(3/2)}{\ln(2)} \approx 0{,}58496.$

d) $\mathrm{E}[X] = \dfrac{1}{\ln 2} \displaystyle\int_1^2 x \cdot \frac{1}{x}\,\mathrm{d}x = \frac{1}{\ln 2} \approx 1{,}4427$

e) $$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X] = \frac{1}{\ln 2} \int_1^2 x^2 \cdot \frac{1}{x}\,\mathrm{d}x - \left(\frac{1}{\ln 2}\right)^2$$
$$= \frac{1}{\ln 2}\left[\frac{2^2}{2} - \frac{1^2}{2}\right] - \left(\frac{1}{\ln 2}\right)^2 = \frac{3}{2\ln 2} - \frac{1}{\ln^2 2} \approx 0{,}08267.$$

$$\mathrm{Var}[x] = \frac{1}{\ln 2} \int_1^2 x^2 \cdot \frac{1}{x}\,dx - \left(\frac{1}{\ln 2}\right)^2$$
$$= \frac{1}{\ln 2} \cdot \left[\frac{1}{2}x^2\right]_1^2 - \left(\frac{1}{\ln 2}\right)^2 = \frac{1}{\ln 2} \cdot \frac{1}{2}(2^2 - 1^2) - \left(\frac{1}{\ln 2}\right)^2 = \frac{1 \cdot 3}{\ln(2) \cdot 2} - \frac{1}{\ln^2 2}$$

Aufgabe 72

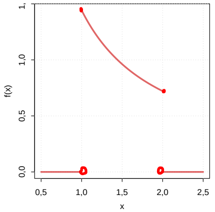
Gegeben ist zur Zufallsvariablen X die Dichtefunktion f gemäß

$$f(x) = \begin{cases} \dfrac{1}{x \cdot \ln 2} & \text{für} \quad 1 \le x \le 2 \\ 0 & \text{sonst} \end{cases}$$

a) Skizzieren Sie den Graph von f.

b) Bestimmen Sie zu X die Verteilungsfunktion $F(x) = P(X \le x)$ und skizzieren Sie auch deren Graph.

c) Berechnen Sie $P(1.2 \le X \le 1.8)$.



b) $1 \le x \le 2$

$$F(x) = \int_1^x \frac{1}{t \cdot \ln 2}\, dt$$

$$= \frac{1}{\ln 2} \cdot \int_1^x \frac{1}{t}\, dt$$

$$= \frac{1}{\ln 2} \left[\ln t \right]_1^x$$

$$= \frac{\ln x - \ln 1}{\ln 2} = \frac{\ln x}{\ln 2}$$

$$\Rightarrow F(x) = \begin{cases} 0 & \text{für} \quad x < 1 \\ \ln(x)/\ln 2 & \text{f.} \quad 1 \le x \le 2 \\ 1 & \text{f.} \quad x > 2 \end{cases}$$

c) $P(1.2 \le x \le 1.8) = F(1.8) - F(1.2)$

$$= \frac{\ln 1.8 - \ln 1.2}{\ln 2} \approx 0.585$$

Aufgabe 73

Gegeben ist eine stetige Zufallsvariable X mit der zugehörigen Dichtefunktion $f : \mathbb{R} \to \mathbb{R}$:

$$f(x) = \begin{cases} a \cdot \frac{1}{10}x & \text{für} \quad 0 \le x \le 10 \\ a\left(2 - \frac{1}{10}x\right) & \text{für} \quad 10 < x \le 20 \\ 0 & \text{sonst} \end{cases}$$

a) Welchen Wert muss a haben, so dass f eine Dichtefunktion ist.

b) Berechnen Sie $P(X \le 12)$.

c) Bestimmen Sie zu X die Verteilungsfunktion $F(x)$.

d) Benutzen Sie F, um $P(3 \le X \le 25)$ zu ermitteln.

Lösungshinweis:

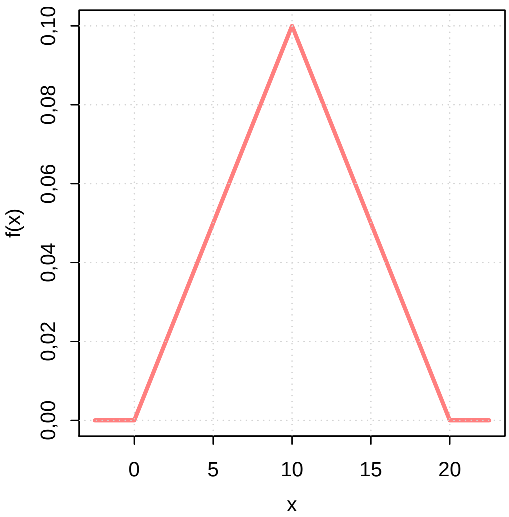
a) Graph von f ist ein Dreieck. Damit f Dichte ist, muss Fläche unter f gleich 1 sein. Also: $20 \cdot a \cdot \frac{1}{2} = 1 \iff a = \frac{1}{10}$. Also ist

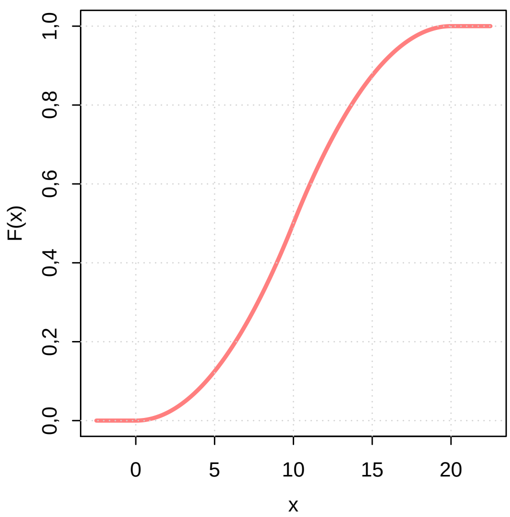
$$f(x) = \begin{cases} 0.01x & \text{für} \quad 0 \le x \le 10 \\ 0.2 - 0.01x & \text{für} \quad 10 < x \le 20 \\ 0 & \text{sonst} \end{cases}$$

b) $$P(X \le 12) = \int_{-\infty}^{12} f(x)\mathrm{d}x = \int_0^{10} 0.01x\mathrm{d}x + \int_{10}^{12} (0.2 - 0.01x)\mathrm{d}x$$
$$= 0{,}01 \cdot \frac{1}{2} \cdot 10^2 + \left[0.2x - 0.005 \cdot x^2\right]_{10}^{12} = 0.5 + (2.4 - 0.72) - (2 - 0.5) = 0.68$$

c) $$F(x) = \begin{cases} 0 & \text{für} \quad x < 0 \\ 0.005x^2 & \text{für} \quad 0 \le x \le 10 \\ -1 + 0.2x - 0.005x^2 & \text{für} \quad 10 < x \le 20 \\ 1 & \text{für} \quad x > 20 \end{cases}$$

d) $P(3 \le X \le 25) = F(25) - F(3) = 1 - 0.005 \cdot 3^2 = 0.955$





Aufgabe 73

Gegeben ist eine stetige Zufallsvariable X mit der zugehörigen Dichtefunktion $f:\mathbb{R}\to\mathbb{R}$:

$$f(x) = \begin{cases} a\cdot\frac{1}{10}x & \text{für} \quad 0\le x\le 10 \\ a\left(2-\frac{1}{10}x\right) & \text{für} \quad 10 < x \le 20 \\ 0 & \text{sonst} \end{cases}$$

a) Welchen Wert muss a haben, so dass f eine Dichtefunktion ist.
b) Berechnen Sie $P(X\le 12)$.
c) Bestimmen Sie zu X die Verteilungsfunktion $F(x)$.
d) Benutzen Sie F, um $P(3\le X\le 25)$ zu ermitteln.

a) $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{10} a\cdot\frac{1}{10}x\,dx + \int_{10}^{20} a\left(2-\frac{1}{10}x\right)dx$

$= \frac{a}{10}\cdot\left[\frac{1}{2}x^2\right]_0^{10} + a\left[2x-\frac{1}{20}x^2\right]_{10}^{20}$

$= a\cdot\left(\frac{1}{20}\cdot(10^2-0^2) + 2\cdot 20 - \frac{1}{20}\cdot 20^2 - \left(2\cdot 10 - \frac{1}{20}\cdot 10^2\right)\right)$

$= a\cdot(5+40-20-20+5) = 10a = 1$

$\Leftrightarrow \quad a = 1/10$

$\left(f(x) = \begin{cases} \frac{1}{100}x & \text{f. } 0\le x\le 10 \\ 0.2-\frac{1}{100}x & \text{f. } 10\le x\le 20 \end{cases}\right.$

b) $P(X\le 12) = F(12) = \int_{-\infty}^{12} f(x)\,dx$

$= \int_0^{10}\frac{1}{100}x\,dx + \int_{10}^{12} 0.2-0.01x\,dx$

$= \left[\frac{1}{200}x^2\right]_0^{10} + \left[0.2x-0.005x^2\right]_{10}^{12}$

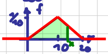
$= 0.5 + (2.4-0.72-(2-0.5))$

$= 0.68$

$F(x)$:

Für $0\le x\le 10$: $F(x) = \int_0^x \frac{1}{100}t\,dt$

$= \left[\frac{1}{200}t^2\right]_0^x = \frac{1}{200}x^2$



$10\le x\le 20$:

$F(x) = \underbrace{\int_0^{10} f(x)\,dx}_{=0.5} + \int_{10}^{x} 0.2-0.01t\,dt$

$= 0.5 + \left[0.2t-0.005t^2\right]_{10}^{x}$

$= 0.5 + 0.2x - 0.005x^2 - (2-0.5)$

$= -0.005x^2+0.2x-1$

$F(x) = \begin{cases} 0 & \text{für} \quad x<0 \\ 0.005x^2 & \text{für} \quad 0\le x<10 \\ -0.005x^2+0.2x-1 & \text{f.} \quad 10\le x<20 \\ 1 & \text{f.} \quad x\ge 20 \end{cases}$

d) $P(3\le X\le 25) = F(25)-F(3)$

$= 1 - 0.005\cdot 3^2 = 0.955$

Aufgabe 74

Für eine Zufallsvariable X sei die Wahrscheinlichkeit $P(|X-10| \geq 2{,}5)$ mit p^* abgekürzt.

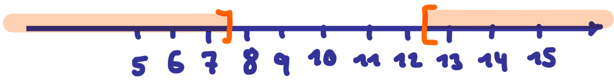
Berechnen Sie p^* für die Fälle, daß X

a) im Intervall [5; 15] gleichverteilt ist
b) einer $N(\mu;\sigma)$–Verteilung mit $\mu = 10$ und $\sigma^2 = 4$ genügt
c) die Werte 3, 6, 9, 12, 15 jeweils mit der Wahrscheinlichkeit 0,2 annimmt
d) Poisson–verteilt ist mit dem Parameter $\lambda = 10$.

Empfehlung: Veranschaulichen Sie den Bereich $\{x \in \mathbb{R} : |x-10| \geq 2{,}5\}$ auf einer Zahlengeraden.

Lösungshinweis:

$P(|X-10| \geq 2{,}5)$



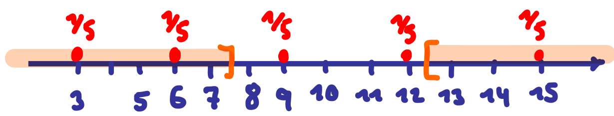
a)

$P(5 \leq X \leq 7.5) + P(12.5 \leq X \leq 15)$
$= 0.25 + 0.25$
$= 0.5$

b) $X \sim N(\mu = 10; \sigma = 2)$

$P(X \leq 7.5) + P(X \geq 12.5)$
$= 1 - P(7.5 \leq X \leq 12.5)$
$= 1 - (F(12.5) - F(7.5))$
$= 1 - \left(\Phi\left(\frac{12.5-10}{2}\right) - \Phi\left(\frac{7.5-10}{2}\right)\right)$
$= 1 - (\Phi(1.25) - (1 - \Phi(1.25)))$
$= 2 - 2\Phi(1.25) = 2 - 2 \cdot 0.89435$
$= 0.211$

c)



$P(\{3, 6, 15\}) = 3 \cdot \frac{1}{5} = 0.6$

d) $1 - P(\{8, 9, 10, 11, 12\})$
$= 1 - (F(12) - F(7))$
$= 1 - (0.7916 - 0.2202)$
$= 0.4286$

Poissonverteilung $X_\lambda \sim P(\lambda)$, Verteilungsfunktionen
$F_\lambda(x) = P(X_\lambda \leqslant x)$

$\downarrow x$ $\lambda \rightarrow$	6.25	6.5	6.75	7	7.25	7.5	7.75	8	8.25	8.5	8.75	9	9.25	9.5	10
0	0.0019	0.0015	0.0012	0.0009	0.0007	0.0006	0.0004	0.0003	0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0000
1	0.0140	0.0113	0.0091	0.0073	0.0059	0.0047	0.0038	0.0030	0.0024	0.0019	0.0015	0.0012	0.0010	0.0008	0.0005
2	0.0517	0.0430	0.0357	0.0296	0.0245	0.0203	0.0167	0.0138	0.0113	0.0093	0.0076	0.0062	0.0051	0.0042	0.0028
3	0.1303	0.1118	0.0958	0.0818	0.0696	0.0591	0.0501	0.0424	0.0358	0.0301	0.0253	0.0212	0.0178	0.0149	0.0103
4	0.2530	0.2237	0.1970	0.1730	0.1514	0.1321	0.1149	0.0996	0.0862	0.0744	0.0640	0.0550	0.0471	0.0403	0.0293
5	0.4064	0.3690	0.3338	0.3007	0.2699	0.2414	0.2152	0.1912	0.1694	0.1496	0.1317	0.1157	0.1013	0.0885	0.0671
6	0.5662	0.5265	0.4876	0.4497	0.4132	0.3782	0.3449	0.3134	0.2838	0.2562	0.2305	0.2068	0.1849	0.1649	0.1301
7	0.7089	0.6728	0.6359	0.5987	0.5615	0.5246	0.4884	0.4530	0.4186	0.3856	0.3540	0.3239	0.2954	0.2687	0.2202
8	0.8204	0.7916	0.7611	0.7291	0.6960	0.6620	0.6274	0.5925	0.5577	0.5231	0.4890	0.4557	0.4232	0.3918	0.3328
9	0.8978	0.8774	0.8549	0.8305	0.8043	0.7764	0.7471	0.7166	0.6852	0.6530	0.6203	0.5874	0.5545	0.5218	0.4579
10	0.9462	0.9332	0.9183	0.9015	0.8828	0.8622	0.8399	0.8159	0.7903	0.7634	0.7352	0.7060	0.6760	0.6453	0.5830
11	0.9737	0.9661	0.9571	0.9467	0.9345	0.9208	0.9053	0.8881	0.8692	0.8487	0.8266	0.8030	0.7781	0.7520	0.6968
12	0.9880	0.9840	0.9790	0.9730	0.9658	0.9573	0.9475	0.9362	0.9234	0.9091	0.8932	0.8758	0.8568	0.8364	0.7916
13	0.9949	0.9929	0.9904	0.9872	0.9832	0.9784	0.9727	0.9658	0.9578	0.9486	0.9380	0.9261	0.9129	0.8981	0.8645
14	0.9979	0.9970	0.9958	0.9943	0.9923	0.9897	0.9866	0.9827	0.9781	0.9726	0.9661	0.9585	0.9499	0.9400	0.9165

Aufgabe 75

Die Zufallsvariable X beschreibt die Anzahl der möglichen Tore der deutschen Nationalmannschaft bei Länderspielen. Es kommen nur die folgenden Ergebnisse mit den angegebenen Wahrscheinlichkeiten vor:

x	0	1	2	3	4
$P(X = x)$	0,4	0,5	0,07	0,025	0,005

Wie groß ist der

a) Erwartungswert von X, $E[X] = 0.735$ $Sta[X] = 0.738$

b) die Varianz und die Standardabweichung von X und

c) $P(|X - E[X]| \leq Sta[X])$?

$P(|X - 0.735| \leq 0.738) = P(\{0,1\}) = 0.4 + 0.5 = 0.9$

Lösungshinweis:

✓ 0 : $|0 - 0.735| = 0.735 \leq 0.738$

✓ 1 : $|1 - 0.735| = 0.265 \leq 0.738$

↯ 2 : $|2 - 0.735| = 1.265 \not\leq 0.738$

↯ 3 : ...

a) $E[X] = \sum x_i \cdot f(x_i) = 0 \cdot 0.4 + 1 \cdot 0.5 + 2 \cdot 0.07 + 3 \cdot 0.025 + 4 \cdot 0.005 = 0.735$

b) $Var[X] = E[X^2] - E^2[X] = 0^2 \cdot 0.4 + 1^2 \cdot 0.5 + \dots + 4^2 \cdot 0.005 - 0.735^2 = 0.544775$

$Sta[X] \approx 0.7381$

c) $P(|X - 0.735| \leq 0.7381) = P(X \in \{0,1\}) = 0.4 + 0.5 = 0.9$

Aufgabe 76

Die Zufallsvariablen X und Y haben die nebenstehende gemeinsame Wahrscheinlichkeitsfunktion. Man berechne

x \ y	0	1	2	
1	0,1	0,2	0,3	0.6
2	0,2	0	0,2	0.4
	0.3	0.2	0.5	

a) die Randverteilungen, Erwartungswerte und Varianzen für X und Y,
b) die Kovarianz und Korrelation zwischen X und Y.

a) $E[X] = 1 \cdot 0.6 + 2 \cdot 0.4 = 1.4$

$E[Y] = 0 \cdot 0.3 + 1 \cdot 0.2 + 2 \cdot 0.5 = 1.2$

$Var[X] = E[X^2] - E^2[X] = 1^2 \cdot 0.6 + 2^2 \cdot 0.4 - 1.4^2$
$= 0.24$

$Var[Y] = \cdots = 0^2 \cdot 0.3 + 1^2 \cdot 0.2 + 2^2 \cdot 0.5 - 1.2^2$
$= 0.76$

b) $Cov[X,Y] = E[X \cdot Y] - E[X] \cdot E[Y]$

$= 1 \cdot 0 \cdot 0.1 + 1 \cdot 1 \cdot 0.2 + 1 \cdot 2 \cdot 0.3$
$+ 2 \cdot 0 \cdot 0.2 + 2 \cdot 1 \cdot 0 + 2 \cdot 2 \cdot 0.2$
$- 1.4 \cdot 1.2$

$= -0.08$

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}} = \frac{-0.08}{\sqrt{0.24 \cdot 0.76}} \approx -0.187$$

Aufgabe 77

Man betrachte die Klausurergebnisse in Mathematik (Zufallsvariable X) und Statistik (Zufallsvariable Y). Die gemeinsame Wahrscheinlichkeitsfunktion der zweidimensionalen Zufallsvariablen (X, Y) sei durch die folgende Tabelle gegeben:

$x\downarrow$ $y\rightarrow$	1	2	3	4	5	
1	0,04	0,03	0,02	0,01	0,00	0.1
2	0,04	0,10	0,03	0,02	0,01	0.2
3	0,02	0,08	0,20	0,08	0,02	0.4
4	0,01	0,02	0,04	0,10	0,03	0.2
5	0,00	0,01	0,03	0,03	0,03	0.1
	0.11	0.24	0.32	0.24	0.09	

a) Bestimmen Sie die Wahrscheinlichkeiten
   (i) in Mathematik zu bestehen und in Statistik nicht zu bestehen
   (ii) in beiden Klausuren zu bestehen
   (iii) in beiden Klausuren nicht zu bestehen
   (iv) in beiden Klausuren besser als 3 zu erhalten
   (v) in beiden Klausuren zwischen 2 und 4 zu erreichen.

b) Geben Sie die Randwahrscheinlichkeits- und Randverteilungsfunktionen an.

c) Sind die beiden Zufallsvariablen unabhängig?

d) In Mathematik erzielen Beate, Peter, Helga und Bernd die Noten 1, 2, 3 und 4. Wie sehen für diese vier Kandidaten die Notenchancen bei der Statistikklausur aus?

d)

$P(Y \mid X=x)$	x	y: 1	2	3	4	5
Beate	1	0.4	0.3	0.2	0.1	0
Peter	2	0.2	0.5	0.15	0.1	0.05
Helga	3	0.05	0.2	0.5	0.2	0.05
Bernd	4	0.05	0.1	0.2	0.5	0.15

$$P(Y=1 \mid X=1) = \frac{P(Y=1 \text{ und } X=1)}{P(X=1)} = \frac{0.04}{0.10} = 0.4$$

(i) 0.06

$x\downarrow$ $y\rightarrow$	1	2	3	4	5
1	0,04	0,03	0,02	0,01	0,00
2	0,04	0,10	0,03	0,02	0,01
3	0,02	0,08	0,20	0,08	0,02
4	0,01	0,02	0,04	0,10	0,03
5	0,00	0,01	0,03	0,03	0,03

(ii) 0.84

$x\downarrow$ $y\rightarrow$	1	2	3	4	5
1	0,04	0,03	0,02	0,01	0,00
2	0,04	0,10	0,03	0,02	0,01
3	0,02	0,08	0,20	0,08	0,02
4	0,01	0,02	0,04	0,10	0,03
5	0,00	0,01	0,03	0,03	0,03

(iii) 0.03

$x\downarrow$ $y\rightarrow$	1	2	3	4	5
1	0,04	0,03	0,02	0,01	0,00
2	0,04	0,10	0,03	0,02	0,01
3	0,02	0,08	0,20	0,08	0,02
4	0,01	0,02	0,04	0,10	0,03
5	0,00	0,01	0,03	0,03	0,03

(iv) 0.21

$x\downarrow$ $y\rightarrow$	1	2	3	4	5
1	0,04	0,03	0,02	0,01	0,00
2	0,04	0,10	0,03	0,02	0,01
3	0,02	0,08	0,20	0,08	0,02
4	0,01	0,02	0,04	0,10	0,03
5	0,00	0,01	0,03	0,03	0,03

(v) 0.67

$x\downarrow$ $y\rightarrow$	1	2	3	4	5
1	0,04	0,03	0,02	0,01	0,00
2	0,04	0,10	0,03	0,02	0,01
3	0,02	0,08	0,20	0,08	0,02
4	0,01	0,02	0,04	0,10	0,03
5	0,00	0,01	0,03	0,03	0,03

b)

z	1	2	3	4	5
$P(X=z)$	0.1	0.2	0.4	0.2	0.1
$P(X\leq z)$	0.1	0.3	0.7	0.9	1.0
$P(Y=z)$	0.11	0.24	0.32	0.24	0.09
$P(Y\leq z)$	0.11	0.35	0.67	0.91	1.00

$x\downarrow$ $y\rightarrow$	1	2	3	4	5	
1	0,04	0,03	0,02	0,01	0,00	0.1
2	0,04	0,10	0,03	0,02	0,01	0.2
3	0,02	0,08	0,20	0,08	0,02	0.4
4	0,01	0,02	0,04	0,10	0,03	0.2
5	0,00	0,01	0,03	0,03	0,03	0.1
	0.11	0.24	0.32	0.24	0.09	

c) $\left.\begin{array}{l} P(X=1 \wedge Y=1) = 0.04 \\ P(X=1)\cdot P(Y=1) = 0.1 \cdot 0.11 = 0.011 \end{array}\right\}$ nicht unabhängig

Aufgabe 78

Der Erwartungswert μ in der Grundgesamtheit soll durch die Stichprobenfunktion

$$\hat{\Theta} = \frac{1}{n}(X_1 + \cdots + X_n)$$
$$\hat{\Theta}' = \frac{2}{n^2}\sum_{i=1}^{n} i\, X_i$$
$$\hat{\Theta}'' = \frac{2}{n(n+1)}(X_1 + 2X_2 + \cdots + nX_n)$$

geschätzt werden. Prüfen Sie die Erwartungstreue der Schätzfunktionen und ermitteln Sie die wirksamste unter diesen.

Hinweise: Es gilt $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$ und $\sum_{i=1}^{n} i^2 = \frac{1}{3}n(n+1)(n+\frac{1}{2})$

$$\sum_{i=1}^{n} i\cdot 5 = 1\cdot 5 + 2\cdot 5 + \dots + n\cdot 5$$

Erwartungstreue:

$E[\hat\Theta] = E[\bar X] = \mu \Rightarrow \hat\Theta$ ist erwartungstreu

$$E[\hat\Theta'] = E\left[\frac{2}{n^2}\sum_{i=1}^{n} i X_i\right] = \frac{2}{n^2}\sum_{i=1}^{n} i\, \underbrace{E[X_i]}_{=\mu}$$
$$= \frac{2}{n^2}\cdot\mu\cdot\underbrace{(1+2+\dots+n)}_{\text{Hinweis einsetzen}} = \frac{\cancel{2}}{n^2}\cdot\mu\cdot\frac{1}{\cancel{2}}n(n+1) = \mu\cdot\frac{n+1}{n} > \mu$$

$\Rightarrow \hat\Theta'$ ist nicht erwartungstreu

$$E[\hat\Theta''] = \dots = \frac{\cancel{2}}{\cancel{n(n+1)}}\cdot\mu\cdot\frac{1}{\cancel{2}}\cancel{n(n+1)} = \mu \Rightarrow \hat\Theta'' \text{ ist erwartungstreu}$$

Annahme: $\hat\Theta$ ist wirksamer als $\hat\Theta''$

$$\Leftrightarrow \operatorname{Var}[\hat\Theta] < \operatorname{Var}[\hat\Theta'']$$
$$\Leftrightarrow \operatorname{Var}[\bar X] < \operatorname{Var}\left[\frac{2}{n(n+1)}\cdot(1X_1+2X_2+\dots+nX_n)\right]$$
$$\Leftrightarrow \frac{\sigma^2}{n} < \left(\frac{2}{n(n+1)}\right)^2\cdot\left(1^2\cdot\underbrace{\operatorname{Var}[X_1]}_{\sigma^2}+2^2\cdot\underbrace{\operatorname{Var}[X_2]}_{\sigma^2}+\dots+n^2\underbrace{\operatorname{Var}[X_n]}_{\sigma^2}\right)$$
$$\Leftrightarrow \frac{\cancel{\sigma^2}}{\cancel{n}} < \frac{4\cdot\cancel{\sigma^2}}{n^{\cancel{2}}(n+1)^2}\cdot\underbrace{(1^2+2^2+\dots+n^2)}_{\text{Hinweis 2 einsetzen}}$$
$$\Leftrightarrow \cancel{n}(n+1)^{\cancel{2}} < 4\cdot\frac{1}{3}\cancel{n}\cancel{(n+1)}(n+\tfrac{1}{2})$$
$$\Leftrightarrow n+1 < \frac{4}{3}(n+\tfrac{1}{2})$$
$$\Leftrightarrow 3n+3 < 4n+2 \Leftrightarrow 1 < n$$

Aufgabe 78

Der Erwartungswert μ in der Grundgesamtheit soll durch die Stichprobenfunktion

$$\hat{\Theta} = \frac{1}{n}(X_1 + \cdots + X_n) = \bar{x}$$

$$\hat{\Theta}' = \frac{2}{n^2}\sum_{i=1}^{n} i\, X_i$$

$$\hat{\Theta}'' = \frac{2}{n(n+1)}(X_1 + 2X_2 + \cdots + nX_n)$$

geschätzt werden. Prüfen Sie die Erwartungstreue der Schätzfunktionen und ermitteln Sie die wirksamste unter diesen.

Hinweise: Es gilt $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$ und $\sum_{i=1}^{n} i^2 = \frac{1}{3}n(n+1)(n+\frac{1}{2})$

$$E[\hat{\Theta}] = E\left[\tfrac{1}{n}(X_1 + \cdots + X_n)\right]$$

$$= \tfrac{1}{n}\cdot\Big(\underbrace{E[X_1]}_{=\mu} + \underbrace{E[X_2]}_{=\mu} + \cdots + \underbrace{E[X_n]}_{=\mu}\Big)$$

$$= \tfrac{1}{n}\cdot\Big(\underbrace{\mu + \mu + \cdots + \mu}_{n\text{-mal}}\Big)$$

$$= \tfrac{1}{n}\cdot n\cdot\mu = \mu \quad \Rightarrow \quad \hat{\Theta} \text{ ist erwartungstreu}$$

$$E[\hat{\Theta}'] = E\left[\tfrac{2}{n^2}\left(1\cdot X_1 + 2\cdot X_2 + \cdots + n\cdot X_n\right)\right]$$

$$= \tfrac{2}{n^2}\cdot\Big(1\cdot\underbrace{E[X_1]}_{=\mu} + 2\cdot\underbrace{E[X_2]}_{=\mu} + \cdots + n\cdot\underbrace{E[X_n]}_{=\mu}\Big)$$

$$= \tfrac{2}{n^2}\cdot\left(1\cdot\mu + 2\cdot\mu + \cdots + n\cdot\mu\right)$$

siehe Hinweise

$$= \tfrac{2\mu}{n^2}\cdot(1+2+\cdots+n) = \tfrac{2\mu}{n^{\cancel{2}}}\cdot\tfrac{1}{2}\cdot\cancel{n}\cdot(n+1)$$

$$= \mu\cdot\frac{n+1}{n} > \mu \quad \Rightarrow \quad \hat{\Theta}' \text{ nicht erw.treu}$$

$$E[\hat{\Theta}''] = E\left[\tfrac{2}{n\cdot(n+1)}\cdot\left(1X_1 + \cdots + nX_n\right)\right]$$

$$= \tfrac{2\mu}{n(n+1)}\cdot(1+2+\cdots+n) = \tfrac{\cancel{2}\mu}{\cancel{n(n+1)}}\cdot\tfrac{1}{\cancel{2}}\cdot\cancel{n(n+1)} = \mu \quad \Rightarrow \quad \hat{\Theta}'' \text{ ist erw.treu}$$

Wirksamkeit

$$\mathrm{Var}[\hat{\Theta}] = \mathrm{Var}[\bar{x}] = \frac{\sigma^2}{n}$$

(aus VL, siehe Tabelle (S.163))

- Gegeben: Einfache Stichprobe $X_1,\ldots,X_n$, mit $E(X_i)=\mu$, $\mathrm{Var}(X_i)=\sigma^2$ — Beliebige Verteilung,

Stichprobenfunktion V	Bezeichnung	$E(V)$	$\mathrm{Var}(V)$
$\sum_{i=1}^{n} X_i$	Merkmalssumme	$n\mu$	$n\sigma^2$
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	Stichprobenmittel	μ	$\frac{\sigma^2}{n}$

$$\sum_{i=1}^{n} i^2 = \tfrac{1}{3}n(n+1)(n+\tfrac{1}{2})$$

$$\mathrm{Var}[\hat{\Theta}''] = \mathrm{Var}\left[\frac{2}{n(n+1)}\sum i\, X_i\right]$$

$$= \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} i^2\cdot\underbrace{\mathrm{Var}[X_i]}_{=\sigma^2}$$

$$= \frac{4\sigma^2}{n^2(n+1)^2}\cdot\sum_{i=1}^{n} i^2$$

$$= \frac{4\sigma^2}{n^{\cancel{2}}(n+1)^{\cancel{2}}}\cdot\tfrac{1}{3}\cdot\cancel{n}\cdot\cancel{(n+1)}\cdot(n+\tfrac{1}{2})$$

Hinweis

$$= \frac{4\sigma^2}{n(n+1)}\cdot\tfrac{1}{3}\cdot(n+\tfrac{1}{2})$$

Behauptung: $\mathrm{Var}[\hat{\Theta}] < \mathrm{Var}[\hat{\Theta}'']$

$$\Leftrightarrow \quad \frac{\cancel{\sigma^2}}{\cancel{n}} < \frac{4\cancel{\sigma^2}}{\cancel{n}(n+1)}\cdot\tfrac{1}{3}\cdot(n+\tfrac{1}{2})$$

$$\Leftrightarrow \quad n+1 < \tfrac{4}{3}\cdot(n+\tfrac{1}{2})$$

$$\Leftrightarrow \quad 3n+3 < 4n+2$$

$$\Leftrightarrow \quad 1 < n$$

$$\Rightarrow \quad \hat{\Theta} \text{ ist wirksamer als } \hat{\Theta}''$$

Aufgabe 79

Gegeben sei eine einfache Stichprobe $X_1, \ldots, X_n$ aus einer Grundgesamtheit mit unbekanntem Erwartungswert μ. Zur Schätzung von μ wird der Einsatz von linearen Stichprobenfunktionen der folgenden Gestalt erwogen:

$$\hat{\Theta}_1 = \frac{1}{n-1} \sum_{i=1}^{n-1} X_i + \alpha \cdot X_n \quad \text{mit} \quad -1 \leq \alpha \leq 1$$

a) Bestimmen Sie α so, dass $\hat{\Theta}_1$ erwartungstreu für μ ist.

b) Welche Schätzfunktion ist wirksamer: Die erwartungstreue Variante von $\hat{\Theta}_1$ von Teilaufgabe a) oder eine neue Funktion $\hat{\Theta}_2$ mit

$$\hat{\Theta}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i \;?$$

Lösungshinweis:

a)

$$E[\hat{\Theta}_1] = E\left[\frac{1}{n-1}\sum_{i=1}^{n-1} X_i + \alpha X_n\right]$$
$$= E\left[\frac{1}{n-1}\sum_{i=1}^{n-1} X_i\right] + E[\alpha X_n]$$
$$= \frac{1}{n-1}\sum_{i=1}^{n-1} \underbrace{E[X_i]}_{=\mu} + \alpha \underbrace{E[X_n]}_{=\mu}$$
$$= \frac{1}{\cancel{n-1}} \cdot (\cancel{n-1}) \cdot \mu + \alpha\mu$$
$$= \mu(1+\alpha) \quad \Rightarrow \quad \alpha = 0$$

b) Annahme: $\hat{\Theta}_2$ wirksamer als $\hat{\Theta}_1$

$$\Leftrightarrow \quad Var[\bar{X}] < Var\left[\frac{1}{n-1}\sum_{i=1}^{n-1} X_i\right]$$
$$\Leftrightarrow \quad \frac{\sigma^2}{n} < \frac{1}{(n-1)^2} \cdot \sum_{i=1}^{n-1} \underbrace{Var[X_i]}_{=\sigma^2}$$
$$\Leftrightarrow \quad \frac{\cancel{\sigma^2}}{n} < \frac{1}{(n-1)^{\cancel{2}}} \cdot \cancel{(n-1)}\,\cancel{\sigma^2}$$
$$\Leftrightarrow \quad n-1 < n \quad \checkmark$$

Aufgabe 80

Bei der Prüfung der Füllmenge von Fruchtsaftflaschen ergaben sich folgende Werte:

ccm	197	198	199	200	201	202	203	204	205	206	207
Anzahl	2	1	3	1	3	1	2	1	1	0	1

Nach Angaben des Abfüllers ist die Füllmenge normalverteilt mit einer Varianz von $\sigma^2 = 2{,}25$.

a) Man gebe ein Schätzintervall für den Erwartungswert μ zum Niveau $1 - \alpha = 0{,}94$.
b) Welcher Stichprobenumfang n garantiert eine Länge von 1 für das Schätzintervall?

Lösungshinweis:

1. Festlegen des Konfidenzniveaus $1 - \alpha$
2. Bestimmung des $\left(1 - \frac{\alpha}{2}\right)$-Fraktils c der $N(0,1)$-Verteilung
3. Berechnen des Stichprobenmittels $\bar{x}$
4. Berechnen des Wertes $\frac{\sigma c}{\sqrt{n}}$
5. Ergebnis der Intervall-Schätzung:

$$\left[\bar{x} - \frac{\sigma c}{\sqrt{n}} \; ; \; \bar{x} + \frac{\sigma c}{\sqrt{n}}\right]$$

a) ① $1-\alpha = 0{,}94$

② $x_{0,97} \approx 1{,}88 = c$

③ $\bar{x} = 201$

④ $\frac{\sigma \cdot c}{\sqrt{16}} = \frac{1{,}5 \cdot 1.88}{4} = 0.705$

⑤ $KI = [200.295 \; ; \; 201.705]$

b) $L = \frac{2\sigma c}{\sqrt{n}} \iff n \geq (2 \cdot \sigma \cdot c)^2 = 4 \cdot 2{,}25 \cdot 1{,}88^2 \approx 31{,}81$, also mind. 32

```
# -------------------
# Loesung in R:
# -------------------

a = c(197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207)
h = c(2, 1, 3, 1, 3, 1, 2, 1, 1, 0, 1)

x = rep(a, h)       # Aufstellen der Urliste
n = length(x)       # Stichprobenumfang
c = qnorm(0.97)     # x_1-alpha/2
x.m = mean(x)       # Stichprobenmittel
sigma = sqrt(2.25) # hier gegeben

KI = x.m + sigma*c/sqrt(n) * c(-1,1)
KI
```

Aufgabe 81

$X_1, \dots, X_{31}$ beschreibe eine einfache Stichprobe aus einer beliebig verteilten Grundgesamtheit. Aus den Ergebnissen wurden $\bar{x} = 9$ und $s^2 = 31/4$ errechnet. Zur Irrtumswahrscheinlichkeit $\alpha = 0{,}05$ bestimme man

a) ein Schätzintervall für den Erwartungswert μ,
b) unter der Annahme, dass die Grundgesamtheit normalverteilt ist, ein Schätzintervall für die Varianz σ^2.

Lösungshinweis:

a) σ unbekannt, 30 Freiheitsgrade → t-Vtlg
(auch Ok: Normalvtlg)

$x_{0.975} = 2.042 = c$

$$kI = \left[9 \mp \frac{\sqrt{31/4} \cdot 2.042}{\sqrt{31}} \right] = [9 \mp 1.021] = [7.979 \,;\, 10.021]$$

N.vtlg $[8.02 \,;\, 9.98]$

b) $\chi^2(30)$: $x_{0.025} = 16.79$; $x_{0.975} = 46{,}98$

$(n-1)\, s^2 = 30 \cdot 31/4 = 232{,}5$

$$kI = \left[\frac{232{,}5}{46{,}98} \,;\, \frac{232{,}5}{16.79} \right] = [4{,}949 \,;\, 13{,}848]$$

```
# Aufgabe 51 (Konfidenzintervall)

# a)
c = qt(1-(0.05)/2, 30)
delta = sqrt(31/4)*c/sqrt(31)
m = 9
KI = c(m-delta, m+delta)
KI

# b)
c1 = qchisq(0.025, 30)
c2 = qchisq(0.975, 30)
z  = 30*31/4
KI = c(z/c2, z/c1)
KI
```

Aufgabe 82

Ein Barista dosiert in einer Espressobar die Menge Kaffeepulver in Gramm bei 10 zufällig ausgewählten Espressobezügen folgendermaßen:

7.3	8.2	7.0	9.2	8.1	6.9	7.1	8.5	9.0	8.5

Berechnen Sie für diesen Barista ein Konfidenzintervall für die Varianz der Kaffeemenge pro Espresso zum Konfidenzniveau 0,95.

Lösungshinweis:

χ^2 Vtlg (9): $c_1 = x_{0.025} = 2.70$

$c_2 = x_{0.975} = 19.02$

$z = (n-1)\, s^2 = 9 \cdot 0.8495^2$

$KI = \left[\frac{z}{c_2} ; \frac{z}{c_1}\right] = [0.3415 ; 2.4059]$

```
# ------------------------------------
# Loesung in R (nicht klausurrelevant)
# ------------------------------------

library(asbio)
x = c(7.3, 8.2, 7.0, 9.2, 8.1, 6.9, 7.1, 8.5, 9.0, 8.5)
ci.sigma(x, conf=0.95)

## 95% Confidence interval for population variance
##  Estimate      2.5%     97.5%
## 0.7217778 0.3414855 2.4055789
```

Aufgabe 83

In einem Spielkasino werden Zweifel geäußert, dass ein bestimmter Würfel fair ist, d.h. alle Zahlen gleich häufig auftreten. Der Spielleiter fordert einen Zweifler auf, ein Signifikanzniveau α zwischen 0,01 und 0,40 anzugeben, zu dem die Hypothese H_0, dass der Würfel fair ist, getestet werden soll. Welches α wird der Zweifler wählen, wenn er möchte, dass der Würfel aus dem Spiel genommen wird?

Lösungshinweis:

Je größer das Signifikanzniveau,
desto größer der Ablehnungsbereich,
desto eher wird auch ein fairer Würfel
(als gezinkt) abgelehnt.

$\rightarrow \alpha = 0.40$

Aufgabe 84

Ein Arbeiter braucht für die Bearbeitung eines Werkstücks im Durchschnitt 7 Minuten (420 sec. $= \mu_0$). Ein Fachmann schlägt, um eine Zeitersparnis zu erreichen ($\mu < \mu_0$), eine andere Bearbeitungsart vor und will die Effektivität seines Vorschlags mithilfe einer Stichprobe vom Umfang $n = 16$ testen. Führen Sie diesen Test (Hypothese $H_0 : \ \mu = \mu_0$ gegen $H_1 : \ \mu < \mu_0$) zum Signifikanzniveau $\alpha = 0{,}05$ bzw. $0{,}01$ durch. Dabei sei ferner vorausgesetzt, dass die Grundgesamtheit normalverteilt ist. Die Stichprobe ergab folgende Werte: $\bar{x} = 408$ und $s = 25{,}7$.

Lösungshinweis:

Test parametrisch

↓ ja

Parameter?

↓ μ

Stichprobe klein? σ unbekannt?

↓ ja

einseitig? zweiseitig?

↓ einseitig (Unterschreitung)

- t-Vtlg mit 15 Freiheitsgraden
- $B = (-\infty\,;\,-x_{1-\alpha})$

1. $\alpha = 0.05$ $\qquad\qquad$ $\alpha = 0.01$
2. $v = \dfrac{\bar{x}-\mu_0}{s}\sqrt{n} = \dfrac{408-420}{25.7}\sqrt{16} = -1{,}8677$
3. $t(15):\ x_{0.95} = 1{,}753$ $\qquad\qquad$ $x_{0.99} = 2{,}603$
4. $B = (-\infty\,;\,-1{,}753) \Rightarrow v \in B \Rightarrow H_0$ wird verworfen

   $B = (-\infty\,;\,-2{,}603) \Rightarrow v \notin B \Rightarrow$ " nicht "

Aufgabe 85

Die Personalabteilung eines Großunternehmens hat den Verdacht, dass die Mitarbeiter die Mittagspause (maximal 1 Stunde) im Durchschnitt überziehen. Mit einer einfachen Stichprobe der Pausenlänge von 20 Mitarbeitern soll getestet werden, ob die Zeiten eingehalten werden oder ob im Mittel überzogen wird. Es ergibt sich für die Pausendauer ein Stichprobenmittel von 70 Minuten und eine Stichprobenstandardabweichung von 15 Minuten. Die Pausendauer eines Mitarbeiters kann als normalverteilte Zufallsvariable angenommen werden.

Formulieren Sie Nullhypothese und Gegenhypothese und testen Sie zum Signifikanzniveau von 5 %.

Lösungshinweis:

$H_0: \mu = 60 \quad H_1: \mu > 60$; also $v = \frac{70-60}{15}\sqrt{20} \approx 2.98$

t-Vtlg, 19 FG: $x_{0.95} = 1.729 \Rightarrow B = (1.729; \infty) \Rightarrow$ H_0 verwerfen

Aufgabe 86

Angeblich sollen Studierende sich in der Nacht vor einer Klausur kürzer in der Tiefschlafphase befinden also im Durchschnitt aller Nächte. Eine einfache Stichprobe von 61 Studenten wird diesbezüglich untersucht. Im Durchschnitt wurde in dieser Stichprobe 48 Minuten Tiefschlaf in den letzten 24h vor der Klausur gemessen, mit einer Stichprobenvarianz von 196.

mittlere s^2

Bestimmen Sie ein 95 %-Konfidenzintervall für die Tiefschlaflänge aller Studierenden am Tag vor der Prüfung.

Lösungshinweis:

Normal vteg: $c = x_{0.975} \approx 1.96$

$$KI = \left[48 \pm \frac{\sqrt{196} \cdot 1.96}{\sqrt{61}}\right] = [44.487\,;\,51.513]$$

Aufgabe 87

Die Hochschule X möchte wissen, wie gut Ihre Sudenten über internationale aktuelle Nachrichten aus der Politik informiert sind. 40 zufällig ausgewählten Studierenden werden Fragen zu 100 Nachrichten der letzten beiden Monate gestellt. Im Durchschnitt können die Befragten 58 Fragen richtig beantworten bei einer Stichprobenstandardabweichung von 3,2.

a) Berechnen Sie ein 99 %-Konfidenzintervall für die durchschnittlich von allen Studenten der Hochschule richtig beantwortete Anzahl der Fragen.

b) Im Landesdurchschnitt aller Studenten aller Hochschulen ergeben sich 65 Punkte. Testen Sie zum Signifikanzniveau von 5 %, ob der der Durchschnitt der Punktzahl an der Hochschule X niedriger ist als im Landesdurchschnitt.

Lösungshinweis:

$n > 30$: Normal-Vtlg

a) ~~29 Freiheitsgrade~~, ~~t-Vtlg~~

$x_{0.995} \approx 2.\cancel{756}\,58 = c$;

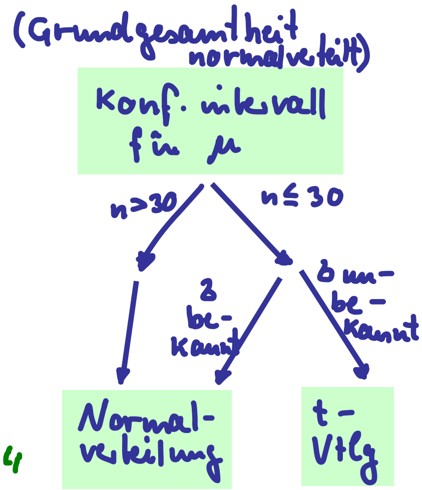
$\Rightarrow KI = [\bar{x} \mp \frac{sc}{\sqrt{n}}] = [\cancel{56.39}\ 56{,}69 ; \cancel{59.61}\ 59{,}30]$

b) $H_0 : \mu = 65 \qquad H_1 : \mu < 65$;

$v = \frac{\bar{x} - \mu}{s} \sqrt{n} = \frac{58 - 65}{3.2} \sqrt{\cancel{30}\,40} \approx \cancel{-11.98}\ -13.835$

$x_{0.05} = -x_{0.95} = \cancel{-1.699}\ -1.64 \Rightarrow B = (-\infty ; \cancel{-1.699}\ -1.64)$

$\Rightarrow H_0$ wird verworfen



σ	S
Streuung in Grundgesamtheit	Streuung in Stichprobe

Aufgabe 88

Studierende beschweren sich, dass die Klausuren in Statistik zu schwer seien. Der Dozent möchte das natürlich im Sinne der Studierenden verbessern und lässt durch das Prüfungsamt die Metallklammern der Klausuren durch eine leichtere Variante aus Kunststoff ersetzen. Die Studierenden glauben aber nicht, dass dadurch die Klausuren leichter werden. Der Dozent möchte das beweisen und zieht eine einfache Stichprobe von 10 Klausuren, bei der er folgende Gewichte misst:

65.28	65.84	64.47	63.82	66.64	62.55	65.74	64.90	65.87	65.29

Gehen Sie im folgenden davon aus, dass es sich beim Gewicht der Klausur um eine normaverteilte Zufallsvariable handelt.

a) Bestimmen Sie zum Konfidenzniveau 0,95 ein symmetrisches Schätzintervall für den Erwartungswert des Gewichts einer Klausur.
b) Mit Metallklammer n hatten die Klausuren früher ein durchschnittliches Gewicht von 65,86 g. Testen Sie zum Signifikanzniveau von 5 % auf Basis der Stichprobe, ob die Klausuren jetzt leichter sind.
c) Was bedeutet bei dem Test von b) der Fehler 2. Art?

Lösungshinweis:

a) t-Verteilung mit 9 Freiheitsgraden: $c = x_{0,975} \approx 2{,}262$; $\overline{x} = 65{,}04$, $s \approx 1{,}18$. Damit:

$$\mathrm{KI} = \left[\overline{x} \pm \frac{sc}{\sqrt{n}}\right] \approx [64{,}196;\ 65{,}884]$$

b) $\mu_0 = 65{,}\overset{86}{\cancel{68}}$ g. Also: $H_0 : \mu = \mu_0$ gegen $H_1 : \mu < \mu_0$. Damit:

$$v = \frac{(\overset{65.04}{\overline{x}} - \overset{65.86}{\mu_0})}{\underset{1.18}{s}} \underset{\sqrt{10}}{\sqrt{n}} \approx -2{,}1975$$

Verwerfungsbereich: $B = [-\infty; -x_{0,95}] \approx [-\infty; -1{,}833]$. Damit ist $v \in B$, die Klausuren sind also laut Test signifikant leichter geworden.

c) Einen Fehler 2. Art zu begehen würde hier bedeuten, nicht zweifelsfrei sagen zu können, dass die Klausuren leichter geworden sind, obwohl sie in Wirklichkeit im Durchschnitt leichter sind.

```
# --------------------
# Loesung in R
# --------------------

x = c(65.28, 65.84, 64.47, 63.82, 66.64, 62.55, 65.74, 64.90, 65.87, 65.29)

# Teilaufgabe a)
KI = t.test(x, conf.level=0.95)$conf.int
KI[1:2]

# Teilaufgabe b)
t.test(x, mu=65.86, alternative = "less")
# damit: p-value  < alpha, also: H0 verwerfen
```

Aufgabe 89

Die deutsche Nationalmannschaft hat in 50 zufällig ausgewählten Länderspielen die folgenden Toranzahlen geschossen:

Anzahl der Tore pro Spiel	0	1	2	3	4
Anzahl der Spiele mit diesem Ergebnis	18	22	5	3	2

Gehen Sie im folgenden davon aus, dass die erhobenen Toranzahlen aus einer einfachen Stichprobe der Grundgesamtheit aller Länderspiele der deutschen Nationalmannschaft stammen.

a) Bestimmen Sie das 95 %-Konfidenzintervall für den Erwartungswert μ der Toranzahl in allen Länderspielen.

b) Die Standardabweichung der Tore pro Spiel in der Grundgesamtheit aller Länderspiele sei jetzt mit $\sigma = 1{,}0$ gegeben. Wie groß muss der Umfang einer Stichprobe sein, damit das 95 %-Konfidenzintervall für μ nicht breiter als 0,5 Tore ist?

Lösungshinweis:

R

```
t.test(x, conf.level=0.95)
```

a) Rechne näherungsweise mit Normalverteilung: $x_{0,975} = c \approx 1{,}959964$ 1,96

TR : $\bar{x} = 0{,}98, \quad s = 1{,}039819$

KI = [0,692; 1,268] $\bar{x} - \frac{sc}{\sqrt{n}} \approx 0.98 - \frac{1.0398 \cdot 1.96}{\sqrt{50}}$

b) Mit $\sigma = 1$ folgt:

$$L = 2\frac{\sigma c}{\sqrt{n}} \leq 0{,}5 \quad \Rightarrow \quad \sqrt{n} \geq 4\sigma c \quad \Rightarrow \quad n \geq 16 \cdot 1^2 \cdot c^2 \approx 61{,}46$$

also: mind. 62

Aufgabe 90

Am Tag der Bundestagswahl werden kurz nach dem Schließen der Wahllokale 300 zufällig ausgewählte Wahlberechtigte gefragt, ob sie gewählt hatten. 250 der Befragten bejahten das.

Schätzen Sie mit der Approximation der Normalverteilung zum Konfidenzniveau 99 % ein Konfidenzintervall für die Wahlbeteiligung aller Wahlberechtigten.

Lösungshinweis:

$\bar{x} = \frac{250}{300}$, $\quad 5 \leq \sum_i x_i = 250 \leq 295$ ✓

$\hat{\sigma} = \sqrt{\frac{250}{300} \cdot \frac{50}{300}} \approx 0{,}3727 \qquad c = x_{1-\alpha/2} = x_{0.995} \approx 2{,}58$

$KI = \left[\bar{x} \pm \frac{\hat{\sigma} c}{\sqrt{n}}\right] \approx [0{,}778\ ;\ 0{,}889]$

Aufgabe 91

Herr Meyer betreibt einen Schnellimbiss für Vegetarier. Eines Tages wird er von einem Kunden wegen Betrugs und Etikettenschwindels verklagt. Der Kunde kann per Fotos nachweisen, dass 8 von 25 von ihm bestellten Gemüsesuppen gar nicht vegetarisch waren, da sich eine Fliege darin befand. Das Gericht verlangt auf Basis dieser als einfach akzeptierten Stichprobe einen Hypothesentest, mit dem der Anteil μ aller Gemüsesuppen mit Fliegen als über den gesetzlich zugelassenen 10 % nachgewiesen wird.

a) Ist die Approximation durch die Normalverteilung in diesem Fall gerechtfertigt?

b) Formulieren Sie H_0 sowie H_1.

c) Führen Sie den Hypothesentest zu einem Signifikanzniveau von $\alpha = 2{,}28\,\%$ durch.

Lösungshinweis:

a) $5 \leq 8 \leq 25-5$ ✓

b) $H_0: \mu = 0.1$, $H_1: \mu > 0.1$

c) $v = \dfrac{8/25 - 0.1}{\sqrt{0.1 \cdot 0.9}} \sqrt{25} \approx 3{,}67$

$B = (x_{1-0.0228}; \infty) = (x_{0.9772}; \infty) = (2; \infty)$

$\Rightarrow$ H_0 ablehnen, Fliegenanteil zu hoch

Aufgabe 92

Herr Untermann möchte auf der Karriereleiter in seiner Firma, einem Telekommunikationsunternehmen, etwas vorankommen und schlägt deshalb folgende Maßnahme vor: Der Datendurchsatz der bisher üblichen Flatrates beim Internetzugang von Kunden soll ab einem Volumen von 1GB downloads pro Monat extrem gedrosselt werden. Die alten Konditionen können die Kunden dann optional zukaufen. Bisher hat Herr Untermanns Firma einen Marktanteil von 45 % bei dieser Art flatrates. Eine Stichprobe unter allen potentiellen Kunden vom Umfang $n = 2500$ ergab, dass immerhin noch 43% der potentiellen Kunden diese Leistung mit den verschlechterten Konditionen nachfragen wollen. Herr Untermann schließt daraus, dass sich die Kunden von der Verschlechterung der Vertragsbedingungen nicht abschrecken lassen und freut sich auf die Mehreinnahmen durch seinen Plan.

a) Formulieren Sie in diesem Szenario die Nullhypothese H_0 und die Alternative H_1.
b) Worin besteht in diesem Beispiel das Risiko, den Fehler 1. Art zu begehen?
c) Was bedeutet hier der Fehler 2. Art?
d) Würden Sie an der Stelle von Herrn Untermann ein möglichst kleines Signifikanzniveau α zugrunde legen und dadurch einen größeren Fehler 2. Art (β) in Kauf nehmen oder umgekehrt β klein halten und dabei ein größeres α zulassen?

Lösungshinweis:

a) H_0: Marktanteil bleibt nach Einführung stabil bei 45%
H_1: " sinkt " auf unter 45%

b) Neuer Tarif wird nicht eingeführt, obwohl Marktanteil nicht sinken würde.

c) Neuer Tarif wird eingeführt obwohl Marktanteil dadurch sinken wird

d) Kommt auf Risikoneigung und Kostenstruktur an:

α klein, β groß: risikofreudig, eher Einführung
α groß β klein: risikoavers, eher keine Einf.

Aufgabe 93

100 zufällig ausgewählten Personen einer bestimmten Bevölkerungsschicht werden zwei Fragen vorgelegt, nämlich ob sie (mindestens) ein Smartphone besitzen bzw. ob sie soziale Netze im Internet an mehr als 3 Stunden am Tag nutzen. Es ergeben sich folgende Antworten:

	Smartphone	
>3h soziale Netze	ja	nein
ja	12	24
nein	38	26

Testen Sie zum Signifikanzniveau $\alpha = 5\,\%$, ob in dieser Bevölkerungsschicht die beiden Merkmale „Smartphonebesitz“ bzw. „starke Nutzung sozialer Netzwerke“ voneinander unabhängig sind.

Lösungshinweis:

Aufgabe 94

Wie hängen der Bierkonsum in Litern pro Woche und die Selbsteinschätzung als Fußballfan zusammen? Personen einer einfachen Stichprobe wurden diesbezüglich befragt:
20 Fußballfans und 120 Nichtfußballfans gaben einen Bierkonsum von höchstens 1 l pro Woche an. Zwischen 1 und 3 l pro Woche trinken 210 Fußballfans und 200 Nichtfußballfans. 150 Fußballfans und 90 Nichtfußballfans gaben einen Bierkonsum von mindestens 7 l an. 145 Fußballfans und 65 Nichtfußballfans lagen in der verbleibenden Zwischengruppe.

Es soll nun auf Basis der Stichprobe die Unabhängigkeit der Fußballaffinität vom Bierkonsum in der Gesamtbevölkerung getestet werden.

a) Formulieren Sie H_0 bzw. H_1.
b) Was bedeutet hier der Fehler 1. bzw. der Fehler 2. Art?
c) Führen Sie den Test zum Signifikanzniveau von 10 % durch.

Lösungshinweis:

Aufgabe 95

500 Vollzeitbeschäftigte eines Krankenhauses wurden für eine Studie zur Abhängigkeit von Überstunden und dem Geschlecht zufällig ausgewählt. Die Probanden wurden nach Ihrer durchschnittlichen Wochenarbeitszeit im letzten Jahr befragt. Dabei wurde die Wochenarbeitszeit in die drei Klassen „*unter 40*“, „*von 40 bis unter 45*“ und „*45 und mehr*“-Wochenarbeitsstunden eingeteilt; außerdem wurden die Mitarbeiter nach Geschlecht unterschieden. Die Ergebnisse der Stichprobe sind der folgenden Tabelle zu entnehmen:

	durchschn. Wochenarbeitszeit [h]		
	unter 40	40 bis unter 45	45 und mehr
Frauen	165	55	25
Männer	175	45	35

Testen Sie bitte zur Irrtumswahrscheinlichkeit von 5 %, ob in diesem Krankenhaus die wöchentliche Arbeitszeit unabhängig vom Geschlecht ist.

Lösungshinweis:

Aufgabe 96

Ist die Fahrtzeit mit dem PKW für eine vorgegebene Strecke vom Geschlecht des Fahrers bzw. der Fahrerin abhängig? Diese Frage soll mit einer einfachen Stichprobe untersucht werden. Es ergibt sich:

	Fahrtzeit		
	kurz	mittel	lang
Mann	524	455	221
Frau	413	263	124

a) Formulieren Sie H_0 bzw. H_1.

b) Was bedeutet der Fehler 1. Art hier?

c) Testen Sie zum Signifikanzniveau von 5 %, ob die Fahrtzeit unabhängig vom Geschlecht ist.

Lösungshinweis: