

# Kapitel 2

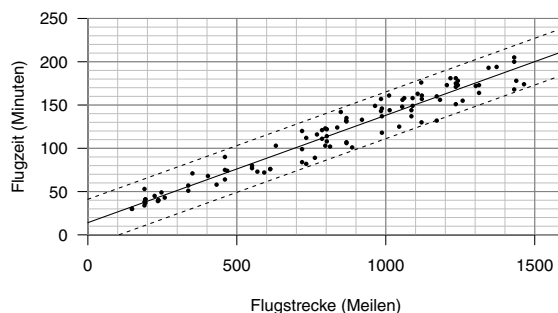
## Lösungen

Dieser zweite Teil des Buches besteht aus der Sammlung der zumeist ausführlich dargestellten Lösungen zur Aufgabensammlung aus dem Aufgabenteil dieses Buches. Folglich ist sie ebenso in dieselben 13 Themen untergliedert.

Wie bereits mehrfach erwähnt, werden zu den meisten der Lösungen auch die entsprechenden in der Statistiksoftware **R** geschriebenen Programme zur Verfügung gestellt. Eine Einführung zu dieser Software finden Sie im Anhang dieses Buches und natürlich auch in der online-Hilfe der Software selbst. Um sich den Zugang zu dieser Statistiksoftware zu vereinfachen, lohnt es sich, **R** auch schlicht als Taschenrechner zu benutzen noch bevor man beginnt, komplexere Probleme damit zu lösen. Sie können die bereitgestellten Programme auch jederzeit beliebig modifizieren und somit spielend den Gebrauch dieser Software erlernen.

### 2.1 Einführende Konzepte und Grundbegriffe - Lösungen

[ 1.1 ] a)  $\approx 50; 80; 140; 175$    b)  $\approx 700; 1\,100$    c) (100, 125); (115, 165)  
d)



Die durchgezogene Linie verläuft etwa durch die Mitte der Punktwolke. Man könnte die Flugzeit bei gegebener Flugstrecke durch den Punkt (d.h. durch die y-Koordinate des Punktes) auf dieser Geraden vorhersagen. Die zwei Strichlinien bilden so etwas wie ein Vorhersage- oder Konfidenzband. Wir würden intuitiv sagen, dass wir

Abweichungen von der mittleren erwarteten Zeit zumeist innerhalb dieses Bands erwarten würden. Dieses Band ist im Grunde ein über die Flugzeitachse gleitendes Konfidenzintervall; für eine gegebene Flugstrecke vertrauen wir darauf, dass sich die Flugzeit auf einem bestimmten Intervall, ablesbar an den Strichlinien, beschränkt. Allerdings liegt diesem Band oder Intervallen (noch) kein Wahrscheinlichkeitsmodell zugrunde, so dass wir auch keine Wahrscheinlichkeitsaussagen machen können.

[ 1.2 ] a) Das zweite Rechteck hat die Höhe 0.001. Multiplizieren wir mit der Klassenbreite 200 ergibt sich für diese Klasse der Anteil  $0.2 = 1/5$ . Die Höhe des ersten und letzten Rechtecks ist jeweils ein Drittel der Höhe des zweiten Rechtecks, also ist der Anteil jeweils  $1/15$ . Die Höhe des vierten Rechtecks übersteigt die Höhe des zweiten Rechtecks um die Hälfte der Höhe des ersten Rechtecks, d.h. der Anteil ist  $1/5 + 1/30 = 7/30$ . Somit bleibt für das dritte Rechteck ein Anteil von  $1 - 1/15 - 1/5 - 7/30 - 1/15 = (30 - 2 - 6 - 7 - 2)/30 = 13/30$ . Die Anteile in den fünf Klassen sind also:

1	2	3	4	5
$2/30$	$6/30$	$13/30$	$7/30$	$2/30$

Selbstverständlich kann man auch alle Höhen direkt an der Skala der y-Achse ablesen. i)  $7/30 + 2/30 = 9/30 = 3/10$  ii)  $2/30 + 6/30 = 8/30 = 4/15$   
 iii)  $13/30 + 7/30 = 20/30 = 2/3$ . b) Man multipliziert die an a) berechneten Anteile mit der Stichprobengröße  $n = 30$ : i) 9 ii) 8 iii) 20.

[ 1.3 ] a) W b) F c) W d) W e) F

[ 1.4 ] a) W b) F c) W d) W e) W

[ 1.5 ] a) W b) W c) F d) F e) W f) F

[ 1.6 ] a) F b) F c) W d) W e) W

[ 1.7 ] Der Begriff *Grundgesamtheit* ist in diesem Zusammenhang schwer einzugrenzen: Sind es

- alle Tageblattleser?
- die Tageblattleser vom 12.04.2008?

- die Tageblattleser vom 12.04.2008, die diese Frage gelesen haben?
- alle Bewohner im Einzugsbereich des Göttinger Tageblatts?
- zusätzlich alle, die Zugang zum Internet haben?
- alle, die Zugang zum Internet haben und dieses in der fraglichen Zeit von Stellung der Frage bis Auswertung der Antworten genutzt haben?

Stichprobe müssten dann alle sein, die die Frage beantwortet haben. Es ist keine zufällige Stichprobe, da nicht jeder Tageblattleser Zugang zum Internet hat. Nicht jeder Leser ist motiviert, die Frage zu beantworten. Vermutlich sind auch mehrfache Beantwortung dieser Fragen durch dieselbe Person möglich. Repräsentativität ist daher auch anzuzweifeln.

[ 1.8 ] a) W b) W c) W d) W e) F

[ 1.9 ] Zuerst bemerke man, dass  $P(X > 10) = P(X \leq 10) = 0.5$  und offenbar  $P(8 < X \leq 10) = P(10 < X \leq 12) = 0.25$ . a)  $P(6 < X \leq 8) = P(X \leq 10) - P(6 < X) - P(8 < X \leq 10) = 0.5 - 0.09 - 0.25 = 0.16$  und  $P(12 < X \leq 14) = P(6 < X \leq 8) = 0.16$ . Die weiteren Aufgaben löse man analog: b) 0.091 bzw. 0.25 c) jeweils 0.9 d) jeweils 0.75.

[ 1.10 ] a)  $1 - 0.01 = 0.99$  b)  $1 - 0.92 = 0.08$  c)  $1 - 0.7 = 0.3$   
d)  $1 - 0.65 = 0.35$  e)  $1 - 0.65 - 0.01 = 0.34$

[ 1.11 ] a) 0.5 b)  $0.5 + 0.477 = 0.977$  c)  $0.5 - 0.477 = 0.023$  d) 0.023 e) 0.954

## 2.2 Deskriptive Statistik - Lösungen

[ 2.1 ] a) nominal skaliert, b) ordinal skaliert, c) ordinal skaliert, d) nominal skaliert, e) nominal skaliert, f) metrisch, g) metrisch

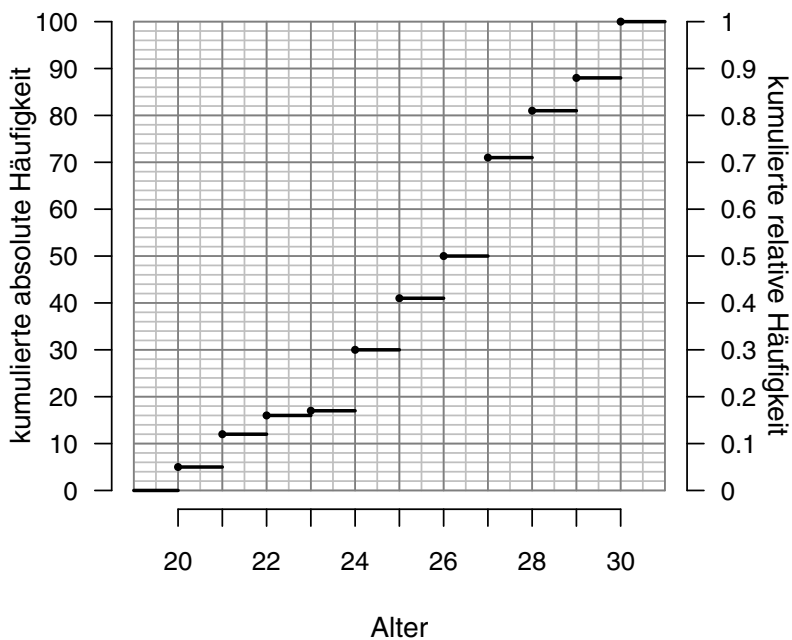
[ 2.2 ] a) F b) F c) W d) W e) W f) F g) W h) W

[ 2.3 ] a)

Alter	$N_i$	$N_i/N$	$K_i$	$K_i/N$
20	5	0.05	5	0.05
21	7	0.07	12	0.12
22	4	0.04	16	0.16
23	1	0.01	17	0.17
24	13	0.13	30	0.30
25	11	0.11	41	0.41
26	9	0.09	50	0.50
27	21	0.21	71	0.71
28	10	0.10	81	0.81
29	7	0.07	88	0.88
30	12	0.12	100	1.00

b)  $\mu = 25.89$ ;  $\sigma^2 = 8.0579$  c) Modalwert: 27; Median: 26.5;  $Q_1 = 24$ ;  $Q_3 = 28$   
 (siehe auch **R**-Programm)

d)

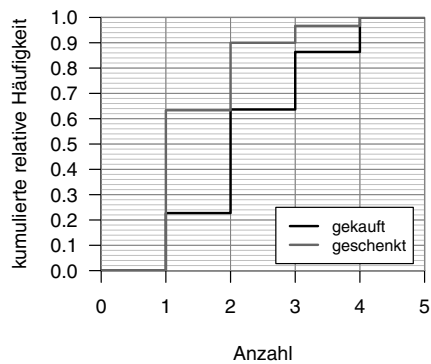
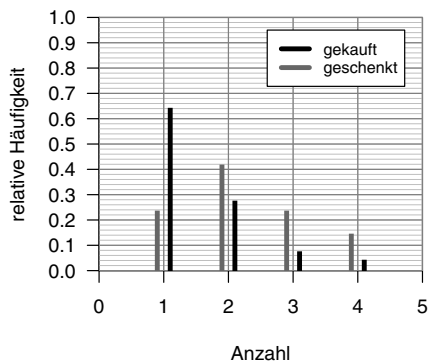


e) i) 0.17 ii) 0.29 iii) 0.2 iv) 0.33 v) 0.54

[ 2.4 ] a)

Anzahl Bücher		1	2	3	4	N
Gekauft	$N_i$	5	9	5	3	22
	$N_i/N$	5/22	9/22	5/22	3/22	
	$K_i$	5	14	19	22	
	$K_i/N$	5/22	14/22	19/22	1	
Geschenkt	$N_i$	19	8	2	1	30
	$N_i/N$	19/30	8/30	2/30	1/30	
	$K_i$	19	27	29	30	
	$K_i/N$	19/30	27/30	29/30	1	

b) linke Grafik c) rechte Grafik.



d) i) 0.773 bzw. 0.367, ii) 0.364 bzw. 0.1, iii) 0.636 bzw. 0.9, iv) 0.864 bzw. 0.967  
e)

Kennzahl	Gekauft	Geschenkt
Mittelwert	2.273	1.5
Median	2	1
Varianz	0.926	0.583
Standardabweichung	0.962	0.764

[ 2.5 ] a) ordinal skaliertes Merkmal b) siehe Tabelle:

	Mittelwert	Median	Modalwert	Spannweite	Varianz	Stand.ab.
TC Ludwigsburg	1.143	1	1	1	0.122	0.350
Braunschweiger TSC	1.857	2	2	1	0.122	0.350
Siemenstadt Berlin	4.429	5	5	2	0.531	0.728
Casino Nürnberg	3.286	3	3	1	0.204	0.452
Schwarz-Gold Göttingen	4.286	4	4 und 5	2	0.490	0.700

## [ 2.6 ]

Kennzahl	gesamt	weiblich	männlich
Mittelwert	12	12	12
Modalwert	12	12	8, 16
Median	12	12	12
Spannweite	8	0	8
Varianz	8	0	16
Standardabweichung	2.828	0	4

## [ 2.7 ] a)

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$\sum$	$\sum/N$	$\sigma^2$
Datensatz I:	5	9	7	8	6	35	7	—
$x_i^2$	25	81	49	64	36	255	51	2
Datensatz II:	15	19	17	18	16	85	17	—
$x_i^2$	225	361	289	324	256	1455	291	2

b) Die beiden Datensätze haben beide die Varianz 2. Datensatz II erhält man, indem zu den Werten aus Datensatz I jeweils die Zahl 10 addiert wird. Die Streuung verändert sich somit nicht und die Varianz ist gleich.

c)

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$\sum$	$\sum/N$	$\sigma^2$
Datensatz I:	5	9	7	8	6	35	7	—
$x_i - \mu$	-2	2	0	1	-1	0	—	—
$(x_i - \mu)^2$	4	4	0	1	1	10	—	2
Datensatz II:	15	19	17	18	16	85	17	—
$x_i - \mu$	-2	2	0	1	-1	0	—	—
$(x_i - \mu)^2$	4	4	0	1	1	10	—	2

d) Das Ergebnis ist natürlich das gleiche, da die Formeln äquivalent sind. Es wird deutlich, dass die Abstände der Ausprägungen beider Datensätze zum jeweiligen Mittelwert gleich sind.

e) Beide Datensätze haben die Varianz 2. Man erhält diese Datensätze, indem man zu den Daten aus I jeweils 5 subtrahiert bzw. 6 addiert. Dadurch ändert sich die Varianz nicht.

[ 2.8 ] a)

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$\sum$	$\sum/N$	$\sigma^2$
Datensatz I:	5	9	7	8	6	35	7	—
$x_i^2$	25	81	49	64	36	255	51	2
Datensatz II:	10	18	14	16	12	70	14	—
$x_i^2$	100	324	196	256	144	1020	204	8

b) Die Werte aus Datensatz II erhält man, indem man die Werte aus Datensatz I mit 2 multipliziert. Die Varianz von Datensatz II ist nun  $2^2 = 4$ -mal so groß wie die Varianz von Datensatz I.

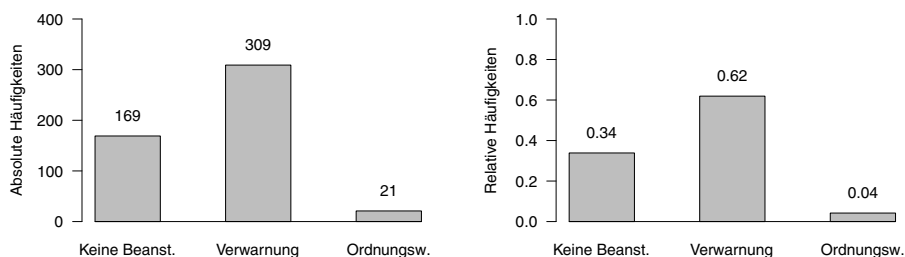
c)

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$\sum$	$\sum/N$	$\sigma^2$
Datensatz I:	5	9	7	8	6	35	7	—
$x_i - \mu$	-2	2	0	1	-1	0	—	—
$(x_i - \mu)^2$	4	4	0	1	1	10	—	2
Datensatz II:	10	18	14	16	12	70	14	—
$x_i - \mu$	-4	4	0	2	-2	0	—	—
$(x_i - \mu)^2$	16	16	0	4	4	40	—	8

d) Es wird für Datensatz II deutlich, dass sich durch die Multiplikation mit 2 die Abstände der Ausprägungen zum Mittelwert ebenfalls verdoppelt haben. Durch Quadrieren ergibt sich somit eine 4-fach größere Varianz für Datensatz II.

e) Die Varianzen sind 18 und 32, denn Datensatz III, erhält man, indem man die Daten aus I mit 3 multipliziert und IV erhält man, indem man die Daten aus I mit 4 multipliziert, d.h. die Varianzen sind dann  $3^2 = 9$ - bzw.  $4^2 = 16$ -mal so groß.

[ 2.9 ] a) Rangskaliert, diskret. b) Keine Beanstandung, Verwarnung, Ordnungswidrigkeit c) siehe unten d) i) 0.66 ii) 0.96

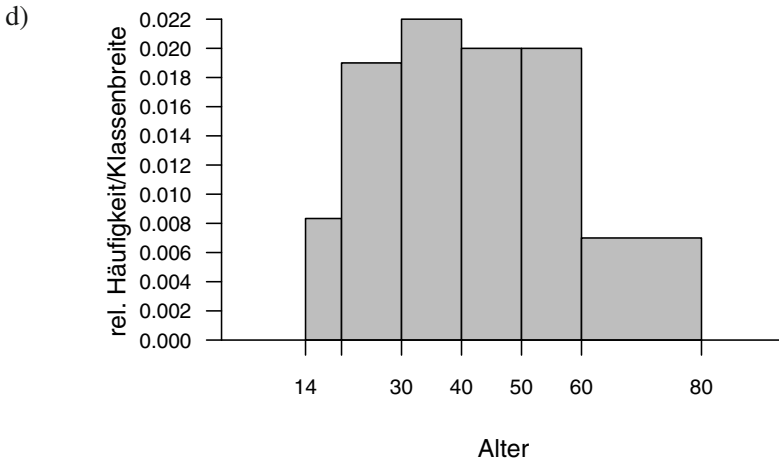


[ 2.10 ] Die Ergebnisse können z.T. nur ungefähr abgelesen werden:

a) 43; 33; 54. Die Box geht von 33 bis 54, die Mittellinie ist bei 43.

- b) i) Anteile: 0.1; 0.4; 0.7; 0.9      Anzahlen:  $0.1 \cdot 200 = 20$ ; 80; 140; 180  
 ii) Anteile: 0.9; 0.6; 0.3; 0.1      Anzahlen: 180; 120; 60; 20  
 iii) Anteile: 0.6; 0.5      Anzahlen: 120; 100  
 c) 30; 40; 50

[ 2.11 ] a)  $\mu = 17 \cdot 0.05 + 25 \cdot 0.19 + 35 \cdot 0.22 + 45 \cdot 0.20 + 55 \cdot 0.20 + 70 \cdot 0.14 = 43.1$  (siehe auch **R**-Programm) b)  $\sigma^2 = (17)^2 \cdot 0.05 + (25)^2 \cdot 0.19 + (35)^2 \cdot 0.22 + (45)^2 \cdot 0.20 + (55)^2 \cdot 0.20 + (70)^2 \cdot 0.14 - (43.1)^2 = 241.09$   $\sigma = \sqrt{241.09} \approx 15.527$  (siehe auch **R**-Programm) c)  $Q_1 \in (30, 40)$ ;  $Q_2 \in (40, 50)$ ;  $Q_3 \in (50, 60)$

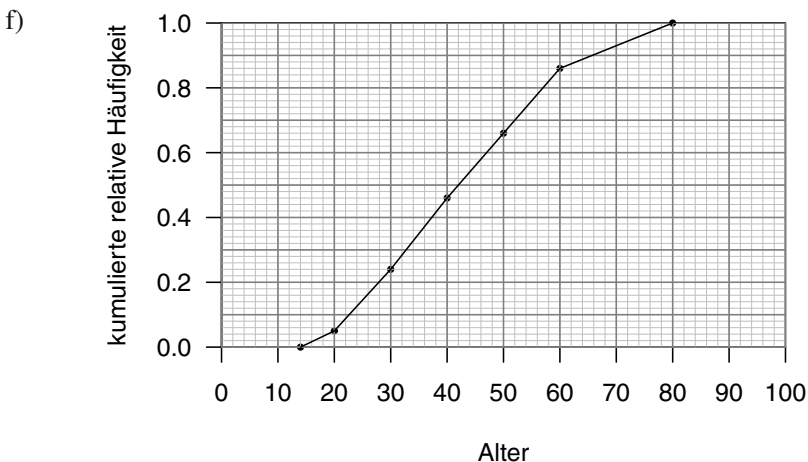


e) Wir benötigen die Höhen der Rechtecke

Alter	14 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 80
Anteil/Klassenbr.	$0.05/6 \approx 0.008$	0.019	0.022	0.020	0.020	0.007

Zu berechnen ist die Fläche des Histogramms von

- i) 14 bis 25:  $6 \cdot 0.05/6 + 5 \cdot 0.019 = 0.05 + 0.095 = 0.145$   
 ii) 55 bis 80:  $5 \cdot 0.020 + 20 \cdot 0.007 = 0.24$   
 iii) 25 bis 45:  $5 \cdot 0.019 + 10 \cdot 0.022 + 5 \cdot 0.020 = 0.095 + 0.220 + 0.100 = 0.415$



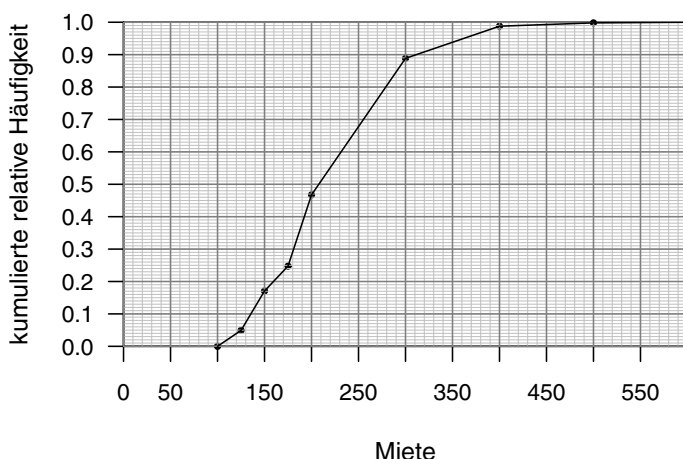


g) Aus der Grafik sind die Zahlen nur ungefähr abzulesen:  $Q_1 \approx 31$ ;  $Q_2 \approx 42$ ;  $Q_3 \approx 54.5$ . h) i)  $\approx 0.14$  ii)  $\approx 0.10$  iii)  $\approx 0.56 - 0.14 = 0.42$ .

[ 2.12 ] a) Qualitativ oder nominal bzw. metrisch in gruppierter Form

b) i) Stab- oder Säulendiagramm ii) Histogramm mit unterschiedlichen Klassenbreiten iii) Da wir die maximale Miete nicht kennen, kann man entweder diese Klasse bis zum rechten Ende der Grafik zeichnen oder sie künstlich auf die gleiche Intervallbreite des linken Nachbarn beschränken. Im Histogramm und in den folgenden Berechnungen haben wir die Klasse *über 500* durch  $500 - 600$  ersetzt.

c) Wohnung (allein);  $176 - 200$  d) Die Anteile sind (nach Division durch 97.8):  $0.049; 0.122; 0.077; 0.220; 0.421; 0.099; 0.010; 0.002$ . Rechnet man mit diesen gerundeten Zahlen weiter, ergibt sich: i)  $0.3215$ ; ii)  $0.9385$ ; iii)  $0.84$  (siehe auch **R**-Programm) e) Summenkurve: siehe Grafik. Quartile:  $Q_1 \approx 175$ ;  $Q_2 \approx 210$ ;  $Q_3 \approx 270$ .



[ 2.13 ] a) F b) W c) W d) W e) W f) F

[ 2.14 ] a) W b) F c) W d) W e) F f) F

[ 2.15 ] a) W b) W c) F d) W e) W f) F g) W h) W

[ 2.16 ] a) W b) W c) F d) W e) F f) W g) F h) F i) W j) F

[ 2.17 ] a) - d)

	Min	Max	Spannw.	$Q_1$	Med.	$Q_3$	$Q_3 - Q_1$	$\mu$	$\sigma^2$	$\sigma$
F	153	193	40	167	170	174.3	7.3	170.3	49.2	7.0
M	151	201	50	178	183	189	11	182.7	60.8	7.8

Siehe **R**-programm. Dort erhält man zum Beispiel:

```
>summary(Groessef)
Min.   1st Qu.  Median    Mean   3rd Qu.    Max.
153.0   167.0   170.0   170.3   174.3   193.0
```

Beachten Sie, dass in **R** bei der Berechnung der Varianz und Standardabweichung durch  $n - 1$  statt  $n$  geteilt wird.

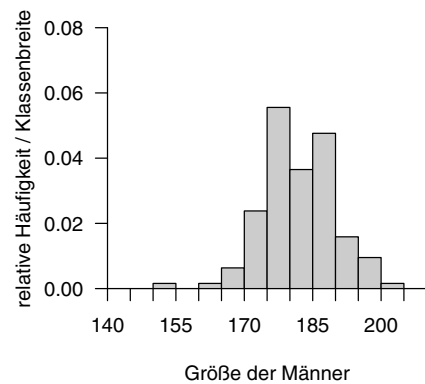
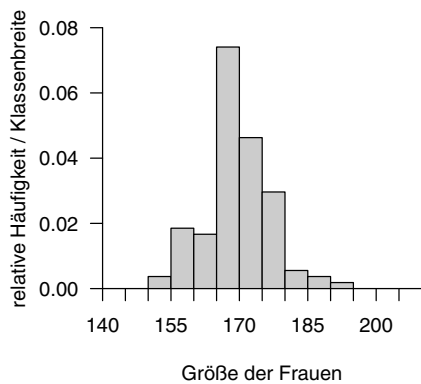
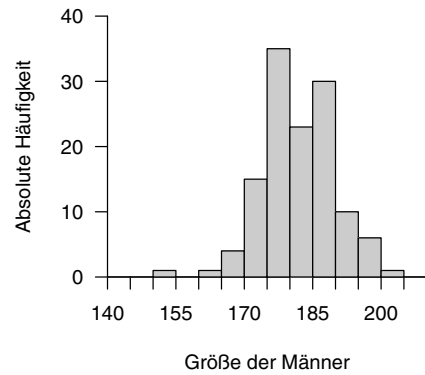
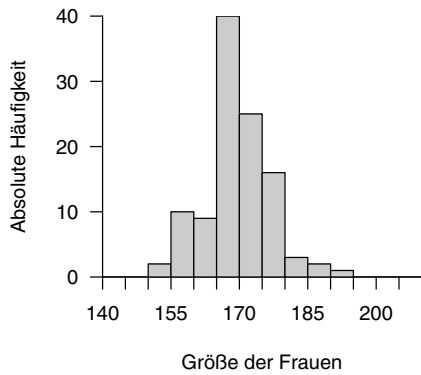
e) Obere Tabelle für die weiblichen, untere für die männlichen Studierenden<sup>1</sup>.

Intervall	$N_i$	$N_i/N$	$K_i$	$K_i/N$
$150 \leq x_i \leq 155$	2	0.019	2	0.019
$155 < x_i \leq 160$	10	0.093	12	0.111
$160 < x_i \leq 165$	9	0.083	21	0.194
$165 < x_i \leq 170$	40	0.370	61	0.565
$170 < x_i \leq 175$	25	0.231	86	0.796
$175 < x_i \leq 180$	16	0.148	102	0.944
$180 < x_i \leq 185$	3	0.028	105	0.972
$185 < x_i \leq 190$	2	0.019	107	0.991
$190 < x_i \leq 195$	1	0.009	108	1.000

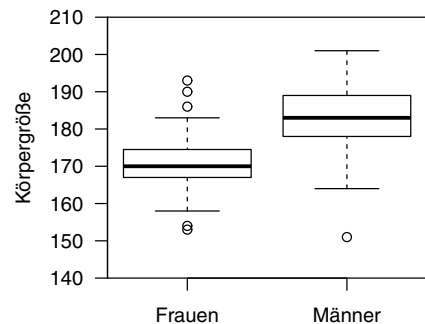
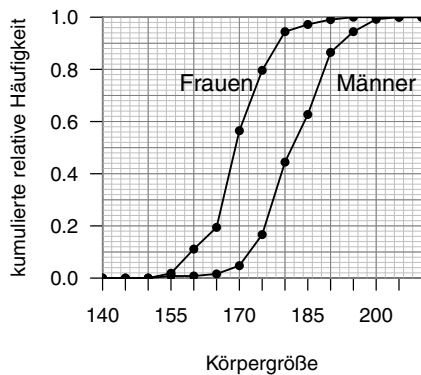
Intervall	$N_i$	$N_i/N$	$K_i$	$K_i/N$
$150 \leq x_i \leq 155$	1	0.008	1	0.008
$155 < x_i \leq 160$	0	0.000	1	0.008
$160 < x_i \leq 165$	1	0.008	2	0.017
$165 < x_i \leq 170$	4	0.033	6	0.050
$170 < x_i \leq 175$	15	0.124	21	0.174
$175 < x_i \leq 180$	35	0.289	56	0.463
$180 < x_i \leq 185$	18	0.149	74	0.612
$185 < x_i \leq 190$	30	0.248	104	0.860
$190 < x_i \leq 195$	10	0.083	114	0.942
$195 < x_i \leq 200$	6	0.050	120	0.992
$200 < x_i \leq 205$	1	0.008	121	1.000

<sup>1</sup> Beachten Sie, dass es zu Rundungsungenauigkeiten kommen kann: Die kumulativen relativen Häufigkeiten wurden nach der Formel  $K_i/N$  berechnet und dies stimmt nicht unbedingt mit der Summe der (gerundeten) relativen Häufigkeiten überein.

f) Histogramme mit absoluten Häufigkeiten (oben) und g) normierte Histogramme (unten: relative Häufigkeiten/Klassenbreite)



h) und i) Summenkurven und Boxplots



[ 2.18 ] Diese Aufgabe kann weitgehend im Kopf und mit Taschenrechner gelöst werden. In **R** würde man zuerst Vektoren mit der Größe und Anzahl sowie dem Alter und Anzahl erstellen, mit denen analog zu vorangegangenen Aufgaben leicht die Tabellenwerte berechnet und die Kennzahlen bestimmt werden könnten. Ein Beispiel:

```
gross<-c(189,200,201,207,188,190,195,193,183,208,201,188,183,204)
> table(gross)                # Berechnet Häufigkeitstabelle
gross
183 188 189 190 193 195 200 201 204 207 208
  2   2   1   1   1   1   1   2   1   1   1
> anzahl<-as.vector(table(gross))    # Vektor der Häufigkeiten
> anzahl
[1] 2 2 1 1 1 1 1 2 1 1 1
round(anzahl/sum(anzahl),3)          # relative Häufigkeiten
[1] 0.143 0.143 0.071 0.071 0.071 0.071 0.071 0.071 0.071 0.143
[9] 0.071 0.071 0.071
> cumsum(anzahl)                  # kumulierte Häufigkeiten
[1] 2 4 5 6 7 8 9 11 12 13 14
> round(cumsum(anzahl)/sum(anzahl),3) # kum.-relat. Häufigkeiten
[1] 0.143 0.286 0.357 0.429 0.500 0.571 0.643 0.786
[9] 0.857 0.929 1.000
> summary(gross)                  # Zusammenfassung
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  183.0  188.3   194.0   195.0   201.0   208.0
> var(gross)                      # Varianz in R (Nenner: n - 1!)
[1] 72.46154
```

a) i)–ii)

Größe in cm	$N_i$	$N_i/N$	$K_i$	$K_i/N$
183	2	0.143	2	0.14286
188	2	0.143	4	0.28572
189	1	0.071	5	0.35715
190	1	0.071	6	0.42858
193	1	0.071	7	0.50001
195	1	0.071	8	0.57144
200	1	0.071	9	0.64287
201	2	0.143	11	0.78573
204	1	0.071	12	0.85716
207	1	0.071	13	0.92859
208	1	0.071	14	1

Alter	$N_i$	$N_i/N$	$K_i$	$K_i/N$
18	1	0.071	1	0.0714
22	2	0.143	3	0.2143
23	3	0.214	6	0.4286
24	3	0.214	9	0.6429
25	2	0.143	11	0.7858
26	1	0.071	12	0.8572
27	1	0.071	13	0.9286
30	1	0.071	14	1.0000

Nationalität	$N_i$	$N_i/N$	$K_i$	$K_i/N$
<i>D</i>	5	0.357	–	–
<i>US</i>	9	0.643	–	–

iii)–vi)

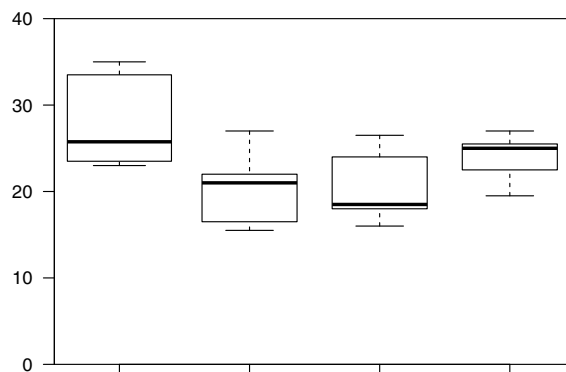
Kennzahl	Größe	Alter	Nationalität
Mittelwert	195	24	–
Median	194	24	–
Modalwert(e)	183, 188, 201	23, 24	US
Minimum	183	18	–
Maximum	208	30	–
Spannweite	25	12	–
Varianz	67.286	7	–
Standardabweichung	8.203	2.646	–

b) i) 0.571, ii) 0.429, iii) 0.5, iv) 0.429, v) 0.571, vi) 0.429

c) i) 0.929, ii) 0.071, iii) 0.214, iv) 0.786    d) i) 0.143, ii) 0.143

e) 0.4; f) 0.143; g) 0.111; h) 0.071

[ 2.19 ] Der erste Boxplot liegt wesentlich höher als die drei anderen. Diese wiederum steigen von links nach rechts an.

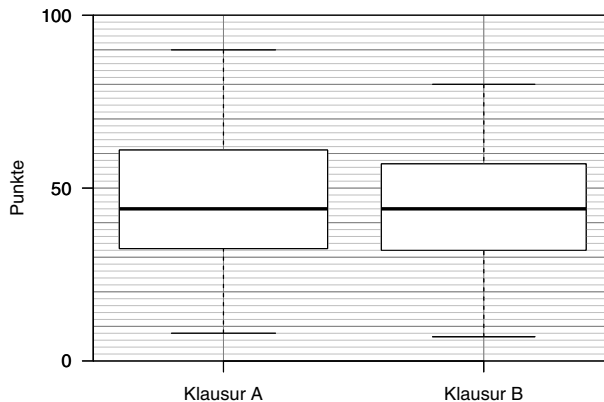


[ 2.20 ] a)

Note	Punkte	$N_i$	$N_i/N$	$x_i^M$	$N_i x_i^M$	$N_i (x_i^M)^2$
5.0	[0, 31]	83	0.234	15.5	1286.5	19940.75
4.0	[32, 35]	32	0.090	33.5	1072	35912
3.7	[36, 40]	34	0.096	38	1292	49096
3.3	[41, 44]	33	0.093	42.5	1402.5	59606.25
3.0	[45, 49]	28	0.079	47	1316	61852
2.7	[50, 53]	25	0.071	51.5	1287.5	66306.25
2.3	[54, 58]	27	0.076	56	1512	84672
2.0	[59, 62]	21	0.059	60.5	1270.5	76865.25
1.7	[63, 67]	27	0.076	65	1755	114075
1.3	[68, 72]	15	0.042	70	1050	73500
1.0	[73, 90]	29	0.082	76.5	2218.5	169715.25
Summe	—	354	1	—	15462.5	811540.8

Mittelwert ist  $15462.5/354 \approx 43.679$ , Varianz ist  $811540.8/354 - (15462.5/354)^2 = 384.6$ . b) Mittelwert Noten:  $(5.0 \cdot 83 + 4.0 \cdot 32 + \dots + 1.0 \cdot 29)/354 \approx 3.186$ .

[ 2.21 ] In der B-Klausur ist die Streuung etwas geringer, das dritte Quartil und das Maximum sind geringer.



[ 2.22 ] a) Die Tabelle der Häufigkeiten sieht wie folgt aus<sup>2</sup>

<sup>2</sup> Beachten Sie, dass es Ungenauigkeiten durch Rundung geben kann. Sie erhalten nicht immer dasselbe Ergebnis, wenn Sie anstelle  $K_i/N$  (exakt) die kumulierten Summen von  $N_i/N$  bilden.

Überschreitung	Strafe	$N_i$	$N_i/N$	$K_i$	$K_i/N$
11 – 15	20	3204	0.453	3204	0.453
16 – 20	30	1806	0.255	5010	0.708
21 – 25	40	963	0.136	5973	0.844
26 – 30	50	536	0.076	6509	0.920
31 – 40	75	436	0.062	6945	0.981
41 – 50	100	99	0.014	7044	0.995
51 – 60	150	24	0.003	7068	0.999
61 – 70	275	6	0.001	7074	1.000
> 70	375	3	0.000	7077	1.000
Summe	–	7077	1.000	–	–

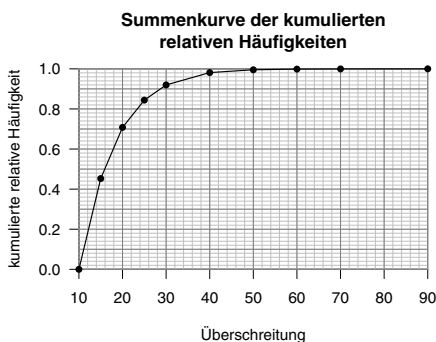
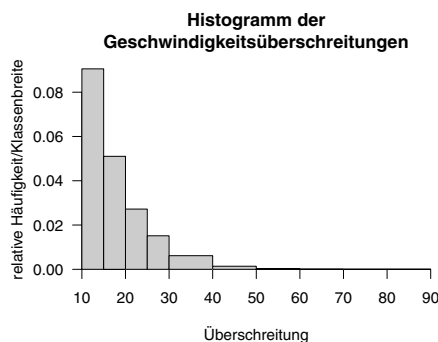
b) Median (für Strafe) ist 30 und die Modalklasse ist die Klasse 11 – 15 bzw. 20.

c) Überschreitung: Mittelwert:  $\mu = (13 \cdot 3204 + 18 \cdot 1806 + 23 \cdot 963 + 28 \cdot 536 + 35.5 \cdot 436 + 45.5 \cdot 99 + 55.5 \cdot 24 + 65.5 \cdot 6 + 85.5 \cdot 3) / 7077 = 133281 / 7077 = 18.83298 \approx 18.833$ .

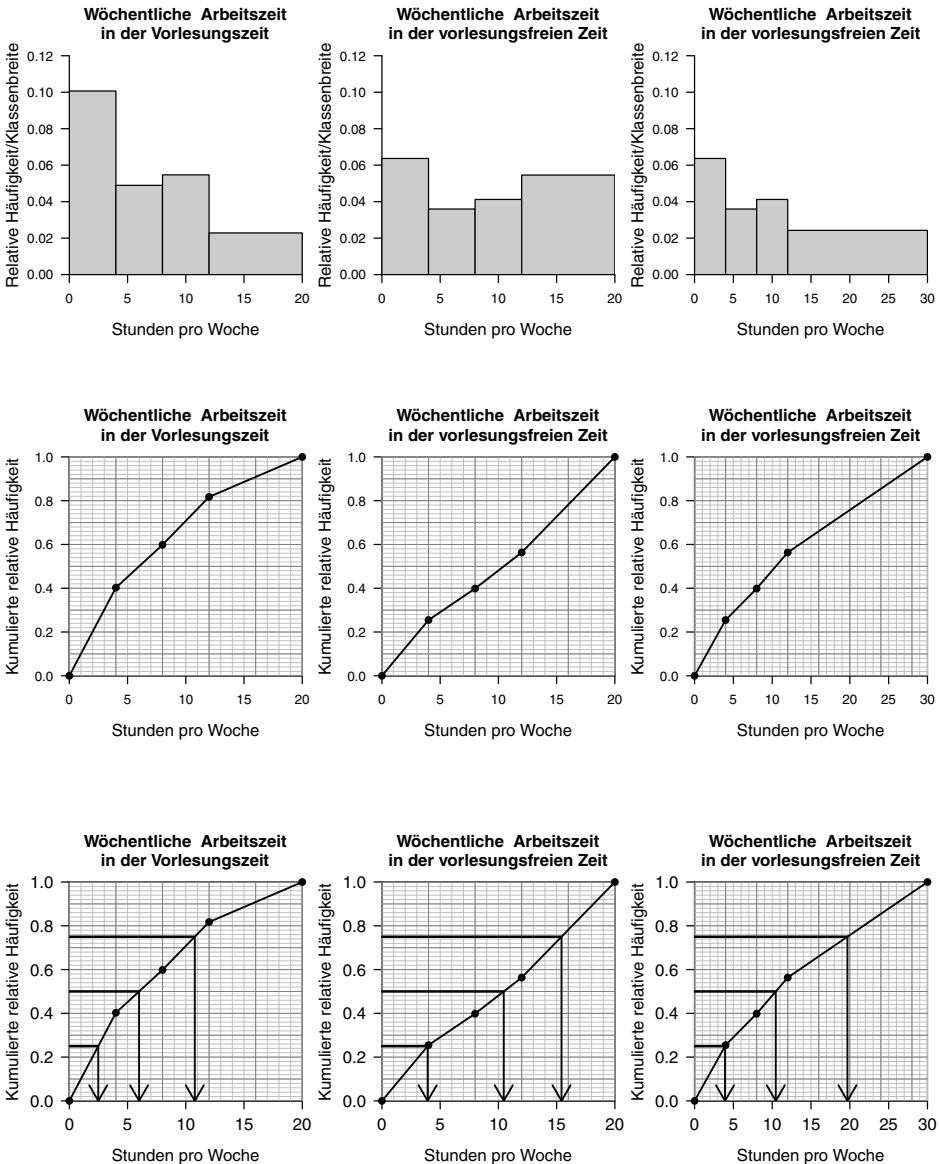
$\left( \sum_{i=1}^9 (x_i^M)^2 \cdot N_i \right) / N = (13^2 \cdot 3204 + 18^2 \cdot 1806 + 23^2 \cdot 963 + 28^2 \cdot 536 + 35.5^2 \cdot 436 + 45.5^2 \cdot 99 + 55.5^2 \cdot 24 + 65.5^2 \cdot 6 + 85.5^2 \cdot 3) / 7077 = 2932293 / 7077 = 414.3412 \approx 414.341 \Rightarrow$  Varianz:  $\sigma^2 = \left( \sum_{i=1}^9 (x_i^M)^2 \cdot N_i \right) / N - \mu^2 = 414.3412 - 18.83298^2 = 59.66011 \approx 59.660$

Strafe: Mittelwert:  $\mu = (20 \cdot 3204 + 30 \cdot 1806 + 40 \cdot 963 + 50 \cdot 536 + 75 \cdot 436 + 100 \cdot 99 + 150 \cdot 24 + 275 \cdot 6 + 375 \cdot 3) / 7077 = 232555 / 7077 = 32.86068 \approx 32.861$

$\left( \sum_{i=1}^9 (x_i)^2 \cdot N_i \right) / N = (20^2 \cdot 3204 + 30^2 \cdot 1806 + 40^2 \cdot 963 + 50^2 \cdot 536 + 75^2 \cdot 436 + 100^2 \cdot 99 + 150^2 \cdot 24 + 275^2 \cdot 6 + 375^2 \cdot 3) / 7077 = 10645925 / 7077 = 1504.299 \Rightarrow$  Varianz:  $\sigma^2 = \left( \sum_{i=1}^9 (x_i^M)^2 \cdot N_i \right) / N - \mu^2 = 1504.299 - 32.86068^2 = 424.4751 \approx 424.475$



[ 2.23 ] Der Umgang mit der Klasse  $\geq 12$  kann sehr subjektiv erfolgen. Für die Lösung wählen wir für die Vorlesungszeit eine Klasse von 12 bis 20, für die vorlesungsfreie Zeit auch eine Klasse von 12 bis 20 (Alternative 1) aber zusätzlich eine Klasse von 12 bis 30 (Alternative 2).



Vorlesungszeit:  $Q_1 \approx 2.5$ , Median  $\approx 6$ ,  $Q_3 \approx 10.8$ , Modalklasse:  $[0; 4]$

Vorlesungsfreie Zeit A1:  $Q_1 \approx 3.8$ , Median  $\approx 10.4$ ,  $Q_3 \approx 15.5$ , Modalklasse:  $[0; 4]$

Vorlesungsfreie Zeit A2:  $Q_1 \approx 3.8$ , Median  $\approx 10.4$ ,  $Q_3 \approx 19.7$ , Modalklasse:  $[0; 4]$

Vorlesungszeit:  $\mu = 7.1$ ,  $\sigma^2 = 27.1$ ; Vorlesungsfreie Zeit A1:  $\mu = 10.0$ ,  $\sigma^2 = 34.4$

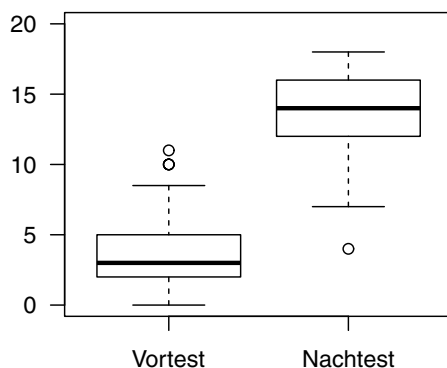
Vorlesungsfreie Zeit A2:  $\mu = 12.2$ ,  $\sigma^2 = 66.7$



[ 2.24 ] a) Vortest:  $\mu = 261.5/75 \approx 3.487$ ;  $\sigma^2 = 1410.75/75 - 3.487^2 \approx 6.651$ ;  $\sigma \approx 2.579$ . Nachtest:  $\mu = 13.62$ ;  $\sigma^2 \approx 8.239$ ;  $\sigma \approx 2.870$

b) **R** verwendet bei der Berechnung der Varianz den Nenner  $n - 1 = 74$  statt  $n = 75$ .

c) Die Box für den Nachtest liegt wesentlich höher (Mitte der Box bei 14) als für den Vortest (Mitte bei 3). Das Maximum im Vortest war 11. Im Nachtest liegt die Box oberhalb von 11, so dass 75% mehr als 11 Punkte erreicht haben.



d) i) Im Vortest hatten 53 von 75 weniger als 5 Punkte, das ist ein Anteil von  $53/75 = 0.707$  oder 70.7%. Weniger als 8:  $70/75 = 0.933$  oder 93.3%. Weniger als 10:  $72/75 = 0.96$  oder 96%. ii) Nachtest: Mehr als 12:  $55/75 = 0.733$  oder 73.3%. Mehr als 14:  $29/75 = 0.36$  oder 36%. Mehr als 16:  $13/75 = 0.173$  oder 17.3%.

[ 2.25 ] a)  $1517/10 = 151.7$  b)  $188 - 117.8 = 70.2$  c) 83 d)  $241299/10 - 151.7^2 = 1117.01$  e) 1241.122 (in **R** wird der Nenner  $n - 1$  verwendet!)

## 2.3 Wahrscheinlichkeiten - Lösungen

[ 3.1 ] a) W b) F c) F d) W e) W f) F

[ 3.2 ] a) W b) W c) F d) W e) W f) F g) F h) F

[ 3.3 ] a) F b) F c) W d) W e) F f) W

[ 3.4 ] a) Ja b) 10 c)  $1/10$  d) i)  $6/10 = 3/5$  bzw.  $4/10 = 2/5$

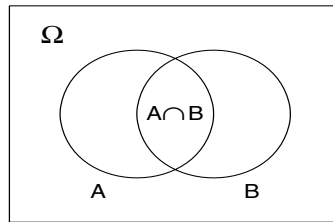
ii)  $4/10 = 2/5$ ;  $3/10$ ;  $2/20 = 1/5$ ;  $1/10$  iii)  $1/10 \cdot 1/10 = 1/100 = 0.01$

iv)  $\frac{3}{10} \cdot \frac{3}{10} \cdot \frac{3}{10} = \frac{27}{100} = 0.027$

e) i)  $\frac{1/10}{9/10} = \frac{1}{9}$ ; es bleiben nur 9 Zahlen ii)  $\frac{1/9}{5/9} = \frac{1}{5}$ ; es bleiben nur 1,3,5,7,9

iii)  $\frac{1/5}{2/5} = \frac{1}{2}$ ; es bleiben nur 3 und 9 iv) 1 denn nur 3 ist kleiner als 6

[ 3.5 ]



b) i) 0.4 ; ii)  $P(A \cap B) = 0.1 \Rightarrow P(A) = 0.1 + 0.2 = 0.3$ ; iii) 0.6;

iv)  $P(A \setminus B) = 0.3 \Rightarrow P(A) = 0.2 + 0.3 = 0.5$ ;

v)  $P(A \cap B) = 0.4 \Rightarrow P(A) = 0.4 + 0.2 = 0.6$

[ 3.6 ] a) F b) W c) W d) F e) F

[ 3.7 ] a) W b) W c) W d) W e) F f) F

[ 3.8 ] a) W b) W c) W d) F e) F

[ 3.9 ] a) i)  $P(A \cap B) = P(A|B) \cdot P(B)$  ii)  $P(A \cap B) = P(B|A) \cdot P(A)$

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

b)  $\frac{P(A|B)}{P(B|A)} = \frac{P(A \cap B)/P(B)}{P(A \cap B)/P(A)} = \frac{P(A)}{P(B)} \Rightarrow P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$  und  $P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)}$

- [ 3.10 ] a) marginale oder unbedingte Wahrscheinlichkeit  $P(A) = P(\text{Ja}) = 0.68$   
 b) bedingte Wahrscheinlichkeit  $P(A|B) = P(\text{Ja}|\text{Alter} \leq 34) = 0.82$   
 c) bedingte Wahrscheinlichkeit  $P(A|B) = P(\text{Ja}|\text{Alter} > 35) = 0.58$   
 d) bedingte Wahrscheinlichkeit  $P(A|B) = P(\text{Ja}|\text{Ostdeutschland}) = 0.75$   
 e) bedingte Wahrscheinlichkeit  $P(A|B) = P(\text{Ja}|\text{Westdeutschland}) = 0.66$   
 f) bedingte Wahrscheinlichkeit  $P(A|B) = P(\text{Ja}|\text{Mutter}) = 0.39$

- [ 3.11 ] a) Tabelle der gemeinsamen Wahrscheinlichkeiten in Prozent:

	CDU/CSU	SPD	FDP	Linke	Grüne
Dalai Lama	11.84	14.43	5.82	5.50	8.34
Papst Benedikt	21.46	10.19	2.63	4.88	0.78

Die Werte werden analog zum obigen Beispiel berechnet, siehe auch **R**-Programm.

- b) Tabelle der bedingten Wahrscheinlichkeiten in Prozent, dass eine Person, die den Dalai Lama bzw. den Papst als Vorbild betrachtet, die gegebenen Parteien wählt:

gegeben Vorbild	CDU/CSU	SPD	FDP	Linke	Grüne
Dalai Lama	26.91	32.80	13.25	12.50	18.96
Papst Benedikt	51.10	24.26	6.27	11.61	1.85

Die Werte werden analog zum obigen Beispiel berechnet, siehe auch **R**-Programm.

- c) Gemeinsame Wahrscheinlichkeiten  $P(\text{Nein und Partei})$  in Prozent:

	CDU/CSU	SPD	FDP	Linke	Grüne
Nein	15.91	20.94	4.23	7.13	8.92

Tabelle der bedingten Wahrscheinlichkeiten in Prozent, dass eine Person, die *Nein* gesagt hat, die gegebenen Parteien wählt:

gegeben	CDU/CSU	SPD	FDP	Linke	Grüne
Nein	25.66	33.78	6.82	11.49	14.39

- [ 3.12 ] a)  $\frac{1}{365} \approx 0.002739726$  b)  $\frac{1}{365^2} = \frac{1}{133225} \approx 0.0000075$

- c)  $\frac{1}{365^3} = \frac{1}{48627125} \approx 0.000000021$  d) Die Wahrscheinlichkeit, dass zwei zufällig

ausgewählte Personen am 7. Januar Geburtstag haben ist nach b)  $\left(\frac{1}{365}\right)^2$ . Entsprechend ist die Wahrscheinlichkeit, dass 9 zufällig ausgewählte Personen nicht am 7.

Januar Geburtstag haben  $\left(\frac{364}{365}\right)^9$ . Bei 11 Personen gibt es  $\frac{11 \cdot 10}{2}$  Möglichkeiten

ein Paar von zwei Personen auszuwählen, denn die erste Person ist frei wählbar, d.h. es gibt 11 Möglichkeiten. Für die zweite Person gibt es dann noch 10 Möglichkeiten, d.h. man hat  $11 \cdot 10$  Möglichkeiten. Nun ist es bei einem Paar jedoch egal,

ob die Person als erste oder zweite ausgewählt wurde, d.h. es gibt bei 11 Personen  $\frac{11 \cdot 10}{2} = 11 \cdot 5 = 55$  verschiedene Paare von zwei Personen. Diese Anzahl der Möglichkeiten 2 aus 11 auszuwählen wird mit  $\binom{11}{2}$  bezeichnet, was wie '2 aus 11' oder im Englischen '11 choose 2' gelesen wird<sup>3</sup>. Die Wahrscheinlichkeit ist also

$$\binom{11}{2} \left(\frac{1}{365}\right)^2 \left(\frac{364}{365}\right)^9 = 55 \cdot 0.002739726^2 (1 - 0.002739726)^9 = 0.0004027668$$

In **R**: `choose(11, 2) / 365^2 * (364/365)^9`

$$\begin{aligned} \text{e) } \binom{18}{3} \left(\frac{1}{365}\right)^3 \left(\frac{364}{365}\right)^{15} &= \frac{18 \cdot 17 \cdot 16}{1 \cdot 2 \cdot 3} \cdot 0.002739726^3 (1 - 0.002739726)^{15} \\ &= 0.00001610421 \quad \text{In **R**: } \text{choose}(18, 3) / 365^3 * (364/365)^{15} \end{aligned}$$

[ 3.13 ] a)  $P(S = 4 | X_1 = 2) = P(X_2 = 2) = 1/6$ ;

b)  $P(S \leq 4 | X_1 = 2) = P(X_2 \leq 2) = P(X_2 = 1) + P(X_2 = 2) = 1/6 + 1/6 = 1/3$ ;

c)  $P(S > 4 | X_1 = 2) = P(X_2 > 2) = \sum_{i=3}^6 P(X_2 = i) = 4/6 = 2/3$ ;

$$\begin{aligned} \text{d) } P(S = 4 | X_1 \leq 2) &= \frac{P(S = 4, X_1 \leq 2)}{P(X_1 \leq 2)} = \frac{P(X_2 = 3, X_1 = 1) + P(X_2 = 2, X_1 = 2)}{1/3} = \\ &= \frac{(1/6) \cdot (1/6) + (1/6) \cdot (1/6)}{1/3} = 1/6 \end{aligned}$$

$$\begin{aligned} \text{e) } P(S \leq 4 | X_1 \leq 2) &= \frac{P(S = 2, X_1 \leq 2) + P(S = 3, X_1 \leq 2) + P(S = 4, X_1 \leq 2)}{P(X_1 \leq 2)} = \\ &= \frac{1/36 + 2/36 + 2/36}{1/3} = \frac{5}{12} \end{aligned}$$

$$\begin{aligned} \text{f) } P(S > 4 | X_1 \leq 2) &= \frac{P(X_1 \leq 2, S > 4)}{P(X_1 \leq 2)} = \frac{P(X_1 = 1, X_2 \geq 4) + P(X_1 = 2, X_2 \geq 3)}{1/3} = \\ &= \frac{(1/6) \cdot (3/6) + (1/6) \cdot (4/6)}{1/3} = \frac{7}{12} \end{aligned}$$

$$\begin{aligned} \text{g) } P(S = 6 | X_1 > 2) &= \frac{P(X_1 > 2, S = 6)}{P(X_1 > 2)} = \frac{\sum_{i=3}^5 P(X_1 = i, X_2 = 6 - i)}{2/3} = \frac{3 \cdot \frac{1}{6} \cdot \frac{1}{6}}{2/3} = \\ &= \frac{9}{72} = \frac{1}{8} \end{aligned}$$

$$\text{h) } P(S \leq 6 | X_1 > 2) = \frac{\sum_{i=3}^5 P(X_1 = i, S \leq 6)}{P(X_1 > 2)} = \frac{\frac{1}{6} \cdot \frac{3}{6} + \frac{1}{6} \cdot \frac{2}{6} + \frac{1}{6} \cdot \frac{1}{6}}{2/3} = \frac{18}{72} = \frac{1}{4}$$

<sup>3</sup> Mathematisch handelt es sich um einen Binomialkoeffizienten. Diese und die folgende Teilaufgabe ist auch mit Hilfe der Binomialverteilung (siehe Kap. 5) zu lösen.

$$\text{i) } P(S > 6 | X_1 > 2) = \frac{\sum_{i=3}^6 P(X_1 = i, S > 6)}{2/3} = \frac{\frac{1}{6} \cdot \left( \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \right)}{2/3} = \frac{3}{4}$$

[ 3.14 ] Es handelt sich hier um bedingte Wahrscheinlichkeiten. Wir interessieren uns für die Wahrscheinlichkeit des Ereignisses A (Pfeil landet in der Gewinnzone) unter der Bedingung, dass das Ereignis B bereits eingetreten ist.

a) Der Anteil der Gewinnzone beträgt unter der Bedingung, dass der Pfeil nicht in der linken oberen Kreishälfte stehen geblieben ist  $30^\circ/270^\circ$ , also  $P(A|B) = 1/9$ .

Genauer nach der Formel:  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  geht es so: Sei  $\alpha$  der Winkel zwischen der positiven y-Achse und dem Pfeil, so ist  $A = \{\alpha \leq 30^\circ\}$ , während  $B = \{\alpha \leq 270^\circ\}$ . Dann ist  $A \cap B = \{\alpha \leq 30^\circ, \alpha \leq 270^\circ\} = \{\alpha \leq 30^\circ\}$  und  $\frac{P(A \cap B)}{P(B)} = \frac{30/360}{270/360} = 30/270 = 1/9$ .

b)  $P(A|B) = 30/180 = 1/6$  c)  $P(A|B) = 30/90 = 1/3$  d)  $P(A|B) = 30/60 = 1/2$

e)  $P(A|B) = 30/45 = 2/3$  f)  $P(A|B) = 1$

[ 3.15 ] a)  $(1 - 0.92)^3 = (0.08)^3 = 0.000512$  b)  $(0.92)^3 = 0.778688 \approx 0.779$

c) Mit der Notation  $\{\circ \bullet \bullet\}$  deuten wir an, dass die erste Glühbirne nach 800 Stunden noch brennt, während die zweite und dritte schon dunkel sind, d.h. nicht mehr brennen. Die gesuchte Wahrscheinlichkeit<sup>4</sup> ist

$$P(\{\circ \bullet \bullet\} \cup \{\bullet \circ \bullet\} \cup \{\bullet \bullet \circ\}) = P(\{\circ \bullet \bullet\}) + P(\{\bullet \circ \bullet\}) + P(\{\bullet \bullet \circ\}).$$

Wegen der Unabhängigkeit ist dies  $3 \cdot (0.92 \cdot 0.08 \cdot 0.08) = 0.017664 \approx 0.018$ .

$$\text{d) } P(\{\circ \circ \bullet\} \cup \{\circ \bullet \circ\} \cup \{\bullet \circ \circ\}) = 3 \cdot (0.92 \cdot 0.92 \cdot 0.08) = 0.203136 \approx 0.203$$

[ 3.16 ] a)  $P(\text{dreimal } 6) = (1/6)^3 \approx 0.00463$

b)  $P(\text{keinmal } 6) = (5/6)^3 \approx 0.57870$

c)  $P(\text{alle gerade}) = 0.5^3 = 0.125$  d) Mit  $\{6, \times, \times\}$  deuten wir an, dass der erste Würfel eine 6 zeigt, während die anderen beiden keine 6 zeigen.

$$P(\text{einmal } 6) = P(\{6, \times, \times\} \cup \{\times, 6, \times\} \cup \{\times, \times, 6\}) = P(\{6, \times, \times\}) + P(\{\times, 6, \times\}) + P(\{\times, \times, 6\}) = 3 \cdot 1/6 \cdot (5/6)^2 \approx 0.34722$$

e) Die Summe der drei Augenzahlen ist nur dann 17, wenn zweimal die 6 und einmal die 5 gewürfelt wird, d.h. die gesuchte Wahrscheinlichkeit ist  $P(\{6, 6, 5\} \cup \{6, 5, 6\} \cup \{5, 6, 6\}) = P(\{6, 6, 5\}) + P(\{6, 5, 6\}) + P(\{5, 6, 6\}) = 3(1/6)^3 \approx 0.01389$

f) Es gibt die 6 Möglichkeiten  $\{1, 1, 3\}; \{1, 3, 1\}; \{3, 1, 1\}; \{1, 2, 2\}; \{2, 1, 2\}; \{2, 2, 1\}$  für die Augensumme 5 ist. Jede hat die Wahrscheinlichkeit  $(1/6)^3$ , d.h. die gesuchte Wahrscheinlichkeit ist  $6 \cdot (1/6)^3 \approx 0.027778$ .

<sup>4</sup> Diese und die folgende Teilaufgabe ist auch mit Hilfe der Binomialverteilung (siehe Kap. 5) zu lösen.

[ 3.17 ] Wir benutzen für 'nicht A' das Zeichen  $A^c$  (also 'A Komplement'). Man bedenke, dass die drei Ereignisse unabhängig voneinander auftreten; daher kann man einfach wie folgt rechnen:

$$\text{a) } P(A_1^c \cap A_2^c \cap A_3^c) = (1 - 0.01) \cdot (1 - 0.02) \cdot (1 - 0.03) = 0.99 \cdot 0.98 \cdot 0.97 = 0.941094$$

$$\text{b) } P(A_1 \cap A_2 \cap A_3) = 0.01 \cdot 0.02 \cdot 0.03 = 0.000006$$

$$\begin{aligned} \text{c) } P(A_1 \cup A_2 \cup A_3) &= P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3^c) + P(A_1 \cap A_2^c \cap A_3) + P(A_1^c \cap A_2 \cap A_3) \\ &+ P(A_1 \cap A_2^c \cap A_3^c) + P(A_1^c \cap A_2 \cap A_3^c) + P(A_1^c \cap A_2^c \cap A_3) = 0.01 \cdot 0.02 \cdot 0.03 + 0.01 \cdot 0.02 \cdot 0.97 + 0.01 \cdot 0.98 \cdot 0.03 + 0.99 \cdot 0.02 \cdot 0.03 + 0.01 \cdot 0.98 \cdot 0.97 + 0.99 \cdot 0.02 \cdot 0.97 + 0.99 \cdot 0.98 \cdot 0.03 \approx 0.0589 \end{aligned}$$

d) Gesucht ist die Wahrscheinlichkeit, dass höchstens zwei Fehlerarten auftreten, d.h. dass nicht alle drei Fehlerarten auftreten, d.h.  $1 - P(A_1 \cap A_2 \cap A_3) = 1 - 0.000006 = 0.999994$  (siehe b).

$$\text{e) } P(A_1 \cap A_2^c \cap A_3^c) + P(A_1^c \cap A_2 \cap A_3^c) + P(A_1^c \cap A_2^c \cap A_3) = 0.01 \cdot 0.98 \cdot 0.97 + 0.99 \cdot 0.02 \cdot 0.97 + 0.99 \cdot 0.98 \cdot 0.03 = 0.057818$$

$$\text{f) } P(A_1 \cap A_2 \cap A_3^c) + P(A_1 \cap A_2^c \cap A_3) + P(A_1^c \cap A_2 \cap A_3) = 0.01 \cdot 0.02 \cdot 0.97 + 0.01 \cdot 0.98 \cdot 0.03 + 0.99 \cdot 0.02 \cdot 0.03 = 0.001082$$

[ 3.18 ] a)  $0.16^3 = 0.004096$ ; b)  $(1 - 2 \cdot 0.16)^3 = 0.68^3 = 0.314432$ ; c)  $0.004096$ ; d)  $(1 - 0.16)^3 = 0.592704$ ; e) ebenso, also  $0.592704$ ; f)  $0.16 \cdot 0.68 \cdot 0.16 = 0.017408$

## 2.4 Verteilungen und ihre Eigenschaften - Lösungen

[ 4.1 ] a) F b) W c) F d) W e) F f) W g) F h) W i) W

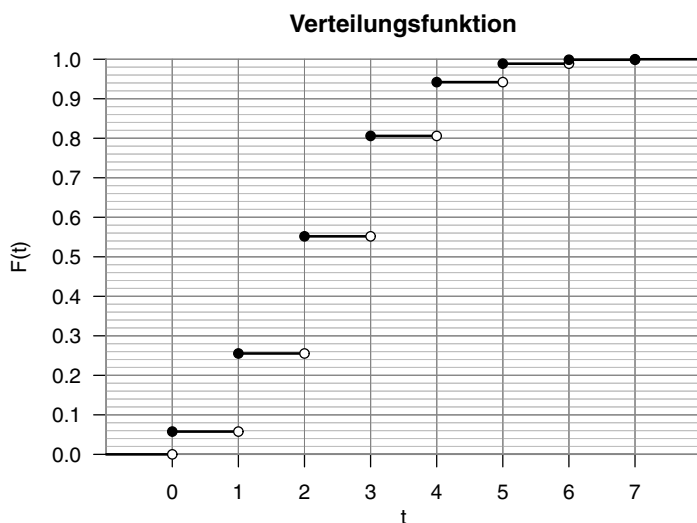
[ 4.2 ] a) F b) F c) W d) W e) F f) W g) F

[ 4.3 ] a) W b) W c) F d) W e) F f) W g) F h) F

[ 4.4 ] a) F b) W c) W d) F e) W f) W g) F

[ 4.5 ] In etwa sollten Sie die folgenden Werte erhalten.

$$\text{a) } P(X=x) = \begin{cases} 0.06 & x=0 \\ 0.20 & x=1 \\ 0.29 & x=2 \\ 0.25 & x=3 \\ 0.14 & x=4 \\ 0.05 & x=5 \\ 0.01 & x=6 \\ 0.00 & x=7 \end{cases} \quad \text{b) } F_X(t) = \begin{cases} 0.00 & t < 0 \\ 0.06 & 0 \leq t < 1 \\ 0.26 & 1 \leq t < 2 \\ 0.55 & 2 \leq t < 3 \\ 0.80 & 3 \leq t < 4 \\ 0.94 & 4 \leq t < 5 \\ 0.99 & 5 \leq t < 6 \\ 1.00 & t \geq 6 \end{cases}$$



c)  $E(X) = 2.4$ ;  $E(X^2) = 7.46$ ;  $\text{Var}(X) = 1.7$

d) i)  $P(X \geq a) = 1 - P(X < a) = 1 - P(X \leq a-1) = 1 - F_X(a-1) = 0.74$ ; 0.20; 0.01

ii)  $P(X > b) = 1 - P(X \leq b) = 1 - F_X(b) = 0.45$ ; 0.06; 0.00

iii)  $P(2 < X < 6) = P(2 < X \leq 5) = F_X(5) - F_X(2) = 0.44$

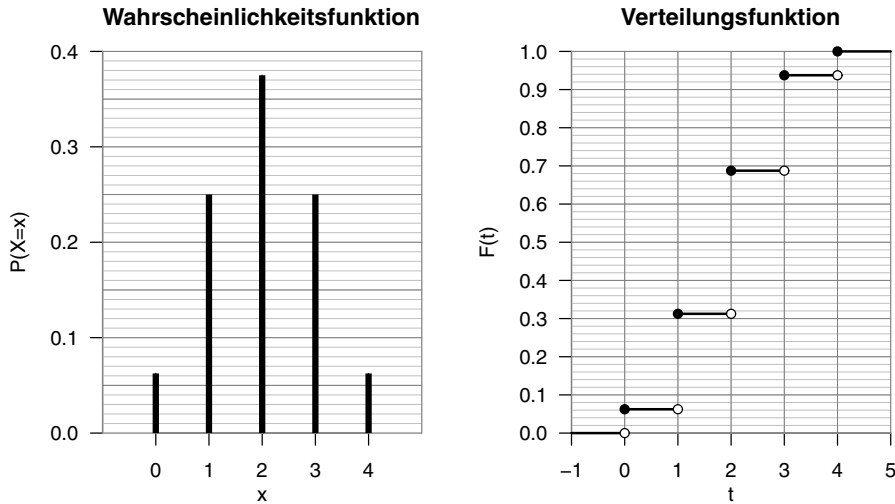
iv)  $P(2 \leq X \leq 6) = P(1 < X \leq 6) = F_X(6) - F_X(1) = 0.74$

v)  $P(2 < X \leq 6) = F_X(6) - F_X(2) = 0.45$

vi)  $P(2 \leq X < 6) = P(1 < X \leq 5) = F_X(5) - F_X(1) = 0.73$

[ 4.6 ] a)  $P_X(x) \geq 0$  für alle  $x$  und  $\sum_{x=0}^4 P_X(x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$ .

b) und h)

c) Schwerpunkt bei 2. Damit ist  $E(X) = 2$ .

$$d) \mu = E(X) = 0 \cdot \left(\frac{1}{16}\right) + 1 \cdot \left(\frac{4}{16}\right) + 2 \cdot \left(\frac{6}{16}\right) + 3 \cdot \left(\frac{4}{16}\right) + 4 \cdot \left(\frac{1}{16}\right) = 2$$

$$E(X^2) = 0^2 \cdot \left(\frac{1}{16}\right) + 1^2 \cdot \left(\frac{4}{16}\right) + 2^2 \cdot \left(\frac{6}{16}\right) + 3^2 \cdot \left(\frac{4}{16}\right) + 4^2 \cdot \left(\frac{1}{16}\right) = 5 \Rightarrow$$

$$\sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2 = 5 - 2^2 = 1$$

 $\alpha_3 = 0$  (Die Wahrscheinlichkeitsfunktion  $P_X(x)$  ist symmetrisch).

$$\alpha_4 = \frac{E(X-\mu)^4}{\sigma^4} = \frac{E(X-2)^4}{1^4} =$$

$$(0-2)^4 \left(\frac{1}{16}\right) + (1-2)^4 \left(\frac{4}{16}\right) + (2-2)^4 \left(\frac{6}{16}\right) + (3-2)^4 \left(\frac{4}{16}\right) + (4-2)^4 \left(\frac{1}{16}\right) = 2.5$$

$$e) P(X > 0) = \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = \frac{15}{16} \quad P(X < 4) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16}$$

$$P(1 < X \leq 3) = \frac{6}{16} + \frac{4}{16} = \frac{5}{8} \quad P(2 \leq X < 4) = \frac{6}{16} + \frac{4}{16} = \frac{5}{8}$$

$$P(X < 2) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16} \quad P(X \geq 2) = \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = \frac{11}{16}$$

$$P(X > 2) = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}$$

f)

t	$-\infty < t < 0$	$0 \leq t < 1$	$1 \leq t < 2$	$2 \leq t < 3$	$3 \leq t < 4$	$4 \leq t < \infty$
$F_X(t)$	$\frac{0}{16}$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	$\frac{16}{16} = 1$

$$g) P(X > 0) = 1 - P(X \leq 0) = 1 - F_X(0) = 1 - \frac{1}{16} = \frac{15}{16}$$

$$P(X < 4) = P(X \leq 3) = F_X(3) = \frac{15}{16}$$

$$P(1 < X \leq 3) = F_X(3) - F_X(1) = \frac{15}{16} - \frac{5}{16} = \frac{5}{8}$$

$$P(2 \leq X < 4) = P(1 < X \leq 3) = \frac{5}{8} \text{ (siehe oben)}$$

$$P(X < 2) = P(X \leq 1) = F_X(1) = \frac{5}{16}$$

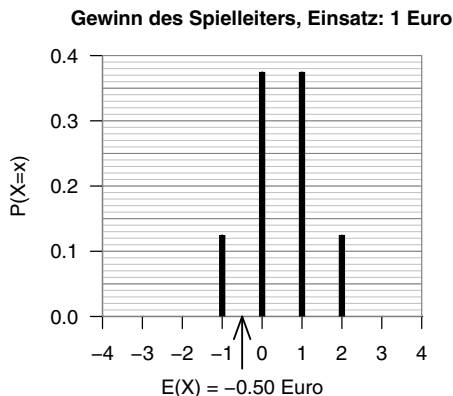
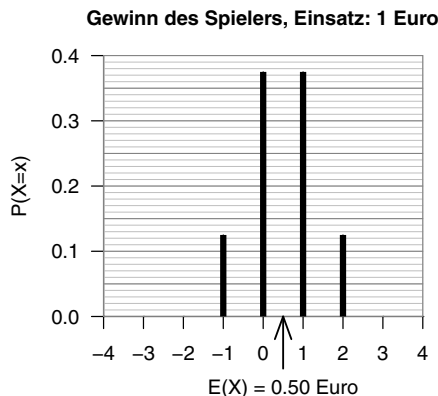
$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - F_X(1) = 1 - \frac{5}{16} = \frac{11}{16}$$

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F_X(2) = 1 - \frac{11}{16} = \frac{5}{16}$$



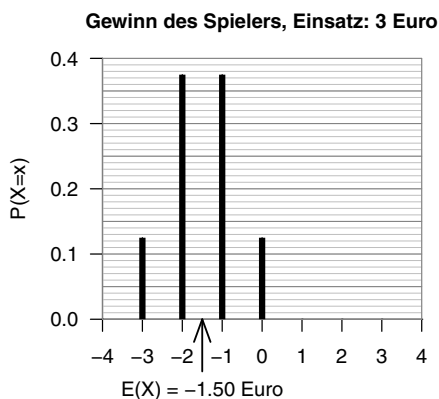
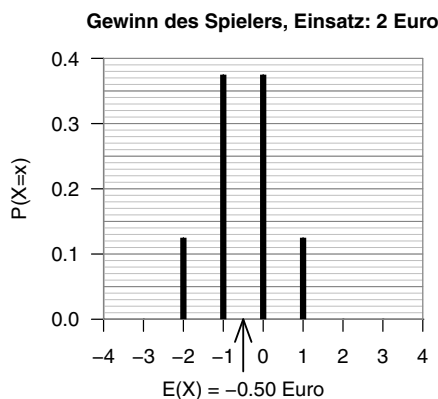
[ 4.7 ] a)

$$P_X(x) = \begin{cases} 1/8 & x = -1 \\ 3/8 & x = 0 \\ 3/8 & x = 1 \\ 1/8 & x = 2 \\ 0 & \text{sonst} \end{cases} \quad F_X(t) = \begin{cases} 0 & t < -1 \\ 1/8 & -1 \leq t < 0 \\ 4/8 & 0 \leq t < 1 \\ 7/8 & 1 \leq t < 2 \\ 1 & 2 \leq t \end{cases}$$

b)  $E(X) = 1/2$ ;  $\text{Var}(X) = 3/4$     c)  $\approx 800 \cdot E(X) = 400 \text{ Euro}$ 

$$d) P_Y(y) = \begin{cases} 1/8 & y = -2 \\ 3/8 & y = -1 \\ 3/8 & y = 0 \\ 1/8 & y = 1 \\ 0 & \text{sonst} \end{cases}$$

Die Grafik ist oben rechts.

e)  $E(Y) = -1/2$ ;  $\text{Var}(Y) = 3/4$ . Dies ist plausibel, da der erwartete Verlust dem erwarteten Gewinn des Spielers entsprechen muss, denn es gilt  $X = -Y$ .f) Sei  $X_2$  bzw.  $X_3$  der Gewinn des Spielers bei 2 bzw. 3 Euro Einsatz.

$$P_{X_2}(x) = \begin{cases} 1/8 & x = -2 \\ 3/8 & x = -1 \\ 3/8 & x = 0 \\ 1/8 & x = 1 \\ 0 & \text{sonst} \end{cases} \quad P_{X_3}(x) = \begin{cases} 1/8 & x = -3 \\ 3/8 & x = -2 \\ 3/8 & x = -1 \\ 1/8 & x = 0 \\ 0 & \text{sonst} \end{cases}$$

Beachten Sie: Die Wahrscheinlichkeitsfunktion wird um eine bzw. zwei Einheiten nach links verschoben. Der Erwartungswert (Schwerpunkt der Wahrscheinlichkeitsfunktion) verschiebt sich auch, die Varianz verändert sich nicht.

$$E(X_2) = -1/2; E(X_3) = -3/2; \text{Var}(X_2) = \text{Var}(X_3) = 3/4.$$

g) Für den Spielleiter muss  $E(Y) = 0$  sein, d.h. der Einsatz muss 1.5 Euro sein.

[ 4.8 ] a)  $P_X(3) = 0.2$ . b)

$t$	$-\infty < t < 1$	$1 \leq t < 2$	$2 \leq t < 3$	$3 \leq t < 4$	$4 \leq t < 5$	$5 \leq t < \infty$
$F_X(t)$	0	0.2	0.5	0.7	0.9	1

c) 0.5; 0.7; 0.5; 0.7 d) 0.8; 0.3

e)  $E(X) = 2.7; E(X^2) = 8.9 \Rightarrow \sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2 = 8.9 - 2.7^2 = 1.61$   
Standardabweichung:  $\sigma = \sqrt{1.61} \approx 1.27$ .

[ 4.9 ] a)  $E(X) = \int_1^\infty x \cdot 3x^{-4} dx = \int_1^\infty 3x^{-3} dx = \left. \frac{3}{-2} x^{-2} \right|_1^\infty = -\frac{3}{2x^2} \Big|_1^\infty = 0 - (-\frac{3}{2}) = \frac{3}{2}; E(X^2) = \int_1^\infty x^2 \cdot 3x^{-4} dx = \int_1^\infty 3x^{-2} dx = \left. \frac{3}{-1} x^{-1} \right|_1^\infty = -\frac{3}{x} \Big|_1^\infty = 0 - (-3) = 3 \Rightarrow \sigma^2 = \text{Var}(X) = E(X^2) - (EX)^2 = 3 - \frac{9}{4} = \frac{3}{4}; \sigma = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}.$

b) Für  $t \geq 1$  ist  $F_X(t) = \int_{-\infty}^t f_X(x) dx = \int_1^t 3x^{-4} dx = \left. -\frac{1}{x^3} \right|_1^t = 1 - t^{-3} = 1 - \frac{1}{t^3}$ , d.h.

$$F_X(t) = \begin{cases} 0 & t < 1 \\ 1 - t^{-3} & t \geq 1 \end{cases} \quad \text{c) } P(X \leq 3/2) = F(3/2) = 1 - (3/2)^{-3} =$$

$$1 - (2/3)^3 = 19/27; \quad P(X \geq 3/2) = 1 - F(3/2) = 1 - 19/27 = 8/27;$$

$$P\left(\frac{3}{2} \leq X \leq \frac{3+\sqrt{3}}{2}\right) = F\left(\frac{3+\sqrt{3}}{2}\right) - F\left(\frac{3}{2}\right) \approx 0.9245 - 0.7037 = 0.2208$$

[ 4.10 ] a) Damit  $f_X(x)$  eine Dichte ist, muss  $f_X(x) \geq 0$  sein für alle  $x \in \mathbb{R}$  und  $\int_{-\infty}^\infty f_X(x) dx = 1$ . Da  $f_X(x) = 0$ , für  $x \leq 0$  und  $x \geq 1 \Rightarrow \int_{-\infty}^\infty f_X(x) dx = \int_0^1 f_X(x) dx = \int_0^1 ax^4 dx = \left. \frac{ax^5}{5} \right|_0^1 = \frac{a}{5} = 1$  und damit  $a = 5$ .

$$\begin{aligned} \text{b) } P(X < 0.5) &= \int_0^{0.5} 5x^4 dx = x^5 \Big|_0^{0.5} = 0.5^5 = 0.03125; \quad P(X > 0.7) = \int_{0.7}^1 5x^4 dx \\ &= x^5 \Big|_{0.7}^1 = 1 - 0.7^5 \approx 0.83193; \quad P(0.6 < X < 0.8) = \int_{0.6}^{0.8} 5x^4 dx = x^5 \Big|_{0.6}^{0.8} = 0.8^5 - \\ &0.6^5 \approx 0.24992; \end{aligned}$$

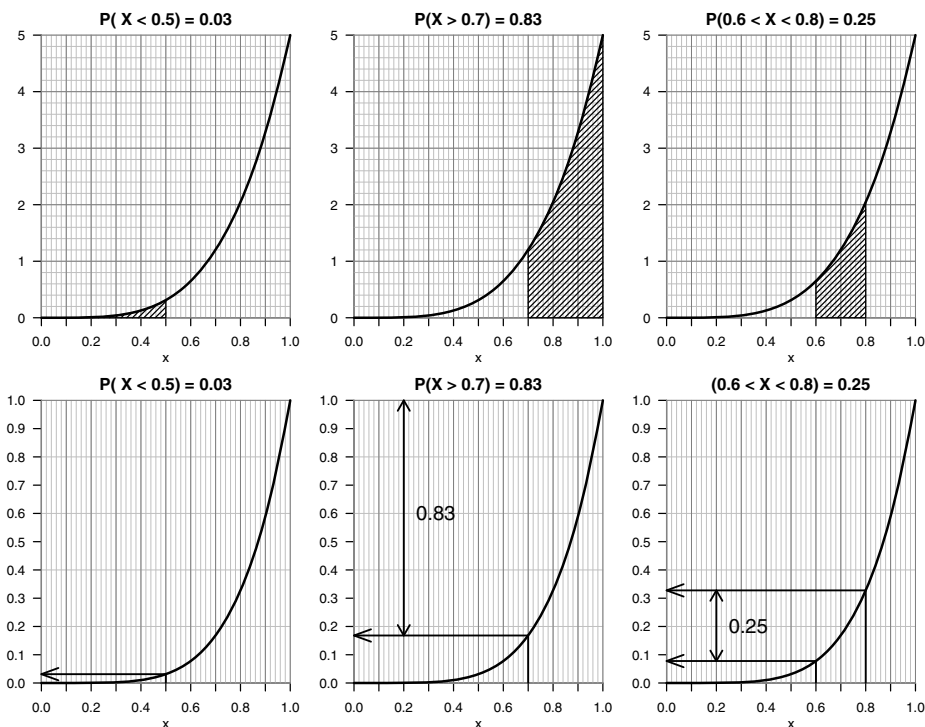
$$\text{c) } F_X(t) = \begin{cases} 0 & t < 0 \\ t^5 & 0 \leq t \leq 1 \\ 1 & 1 < t \end{cases} \quad \text{nach unseren obigen Rechnungen.}$$

$$\begin{aligned} \text{d) } P(X < 0.5) &= F_X(0.5) = 0.5^5; \quad P(X > 0.7) = 1 - F_X(0.7) = 1 - 0.7^5; \\ P(0.6 < X < 0.8) &= F_X(0.8) - F_X(0.6) = 0.8^5 - 0.6^5. \end{aligned}$$

e)

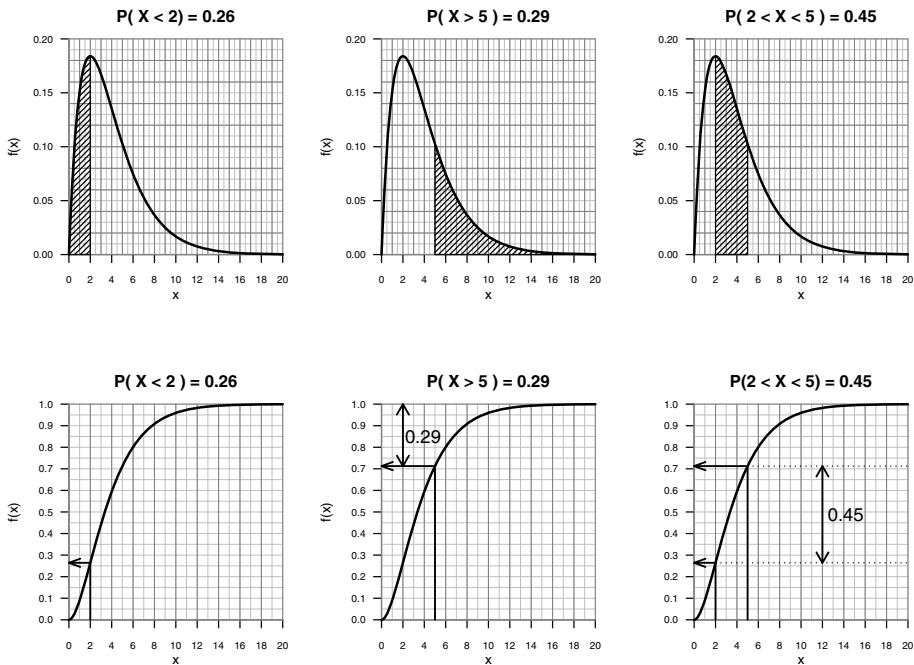
```
> c(pbeta(0.5, 5, 1), 1 - pbeta(0.7, 5, 1), pbeta(0.8, 5, 1) - pbeta(0.6, 5, 1))
[1] 0.03125 0.83193 0.24992
```

f)

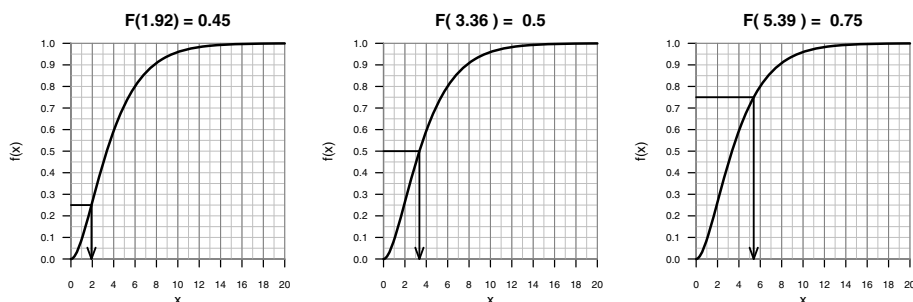


$$\begin{aligned} \text{g) } E(X) &= \int_0^1 x \cdot 5x^4 dx = \int_0^1 5x^5 dx = \frac{5}{6}x^6 \Big|_0^1 = \frac{5}{6}; \quad E(X^2) = \int_0^1 x^2 \cdot 5x^4 dx = \\ &\int_0^1 5x^6 dx = \frac{5}{7}x^7 \Big|_0^1 = \frac{5}{7}; \quad \text{Var}(X) = \frac{5}{7} - \left(\frac{5}{6}\right)^2 \approx 0.0198; \quad \sigma = \sqrt{\text{Var}(X)} \approx 0.1409. \end{aligned}$$

- [ 4.11 ] a)  $F_X(t) = 0$  für  $t \leq 0$ . Für  $t > 0$  ist  $F_X(t) = \frac{1}{4} \int_0^t x e^{-x/2} dx$ . Mit Hinweis für  $k = 1$  ist dies  $\frac{1}{4} \left( -2xe^{-x/2} \Big|_0^t + 2 \int_0^t e^{-x/2} dx \right) = \frac{1}{4} \left( -2te^{-t/2} - 4 \left[ e^{-x/2} \right]_0^t \right) = -\frac{t}{2} e^{-t/2} - e^{-t/2} + 1$ , d.h.  $F_X(t) = 1 - e^{-t/2} (1 + t/2)$  für  $t > 0$ .
- b)  $P(X \leq 2) = F_X(2) = 1 - e^{-1} (1 + 2/1) \approx 0.2642$ ;  $P(X > 5) = 1 - F_X(5) = e^{-5/2} (1 + 5/2) \approx 0.2873$ ;  $P(2 \leq X \leq 5) = F_X(5) - F_X(2) \approx (1 - 0.2873) - 0.2642 = 0.4485$
- c)



- i) Für jeden Punkt  $t$  der  $x$ -Achse entspricht die Fläche unterhalb der Dichtefunktion von 0 bis zu diesem Punkt  $t$  dem Wert der Verteilungsfunktion  $F_X(t) = P(X \leq t) = \int_0^t f_X(x) dx$ . Damit ist  $P(X \leq 2)$  die Fläche unterhalb der Dichtefunktion von 0 bis 2, während  $P(X > 5)$  die Fläche unterhalb der Dichtefunktion rechts von 5, d.h. von 5 bis  $\infty$  ist. Schließlich ist  $P(2 \leq X \leq 5)$  die Fläche unterhalb der Dichtefunktion von 2 bis 5.
- ii) Wir bestimmen die Stelle  $t$ , so dass die Verteilungsfunktion den Wert  $p$  annimmt, d.h. wir bestimmen die Umkehrfunktion der Verteilungsfunktion  $F_X^{-1}(p)$ . Für die Dichtefunktion bedeutet das: Wir bestimmen den Punkt  $t$  auf der  $x$ -Achse, so dass die Fläche links von  $t$  den Wert  $p$  hat.



Die Ablesungen sind nur annähernd möglich: i) 0.25; 0.29; 0.44; ii) 1.9; 3.3; 5.5.

d) Mit dem Hinweis für  $k = 2$  ist  $E(X) = \frac{1}{4} \int_0^\infty x^2 e^{-x/2} dx = \left[ -\frac{1}{2} x^2 e^{-x/2} \right]_0^\infty + \int_0^\infty x e^{-x/2} dx = 0 + \int_0^\infty x e^{-x/2} dx$ . Mit dem Hinweis für  $k = 1$  ist  $\int_0^\infty x e^{-x/2} dx = -2x e^{-x/2} \Big|_0^\infty + 2 \int_0^\infty e^{-x/2} dx = 0 - 4e^{-x/2} \Big|_0^\infty = 0 + 4 = 4$ , d.h.  $E(X) = 4$ . Analog

ergibt sich  $E(X^2) = \frac{1}{4} \int_0^\infty x^3 e^{-x/2} dx = 24$ . Damit ist die Varianz  $\sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2 = 24 - 4^2 = 8$ . Die Standardabweichung ist  $\sigma = \sqrt{8} \approx 2.828$ .

e) i) 0.4422; 0.7127; 0.8641 ii) 0.4060; 0.1991; 0.0916 iii) 0.2705; 0.3298; 0.3449.

[ 4.12 ] a) Der Schwerpunkt der Dichtefunktion liegt ungefähr bei 7.

b) i)  $P(X > 7) = 0.43$ ; ii)  $P(X < 7) = 0.57$  c) i)  $P(X \leq a) = 0.11$ ; 0.34; 0.75; 0.81; 0.96 ii)  $P(X \geq b) = 0.96$ ; 0.78; 0.33; 0.19; 0.10 iii)  $P(X \in (a, b]) = 0.47$ ; 0.62; 0.31 d)  $t_1 = 2.2$ ; 2.8; 3.4; 3.8 e)  $t_2 = 14.1$ ; 12.0; 10.7; 9.8

f)  $\alpha = 0.05$ :  $t_1 = 1.7$ ;  $t_2 = 16.0$ .  $\alpha = 0.10$ :  $t_1 = 2.2$ ;  $t_2 = 14.1$ .  $\alpha = 0.20$ :  $t_1 = 2.8$ ;  $t_2 = 12.0$ .  $P(t_1 \leq X \leq t_2) = 1 - \alpha = 0.95$ ; 0.9; 0.8.

[ 4.13 ] a) Aus  $F_X(x) = \int_{-\infty}^x f_X(t) dt$  folgt:  $F'_X(x) = f_X(x)$  und  $F'_X(x) = ce^{cx}$ .

b) i)  $\approx 0.018$ ; 0.050; 0.135; 0.368 ii)  $\approx 0.982$ ; 0.950; 0.865; 0.632 iii) 0.350; 0.233.

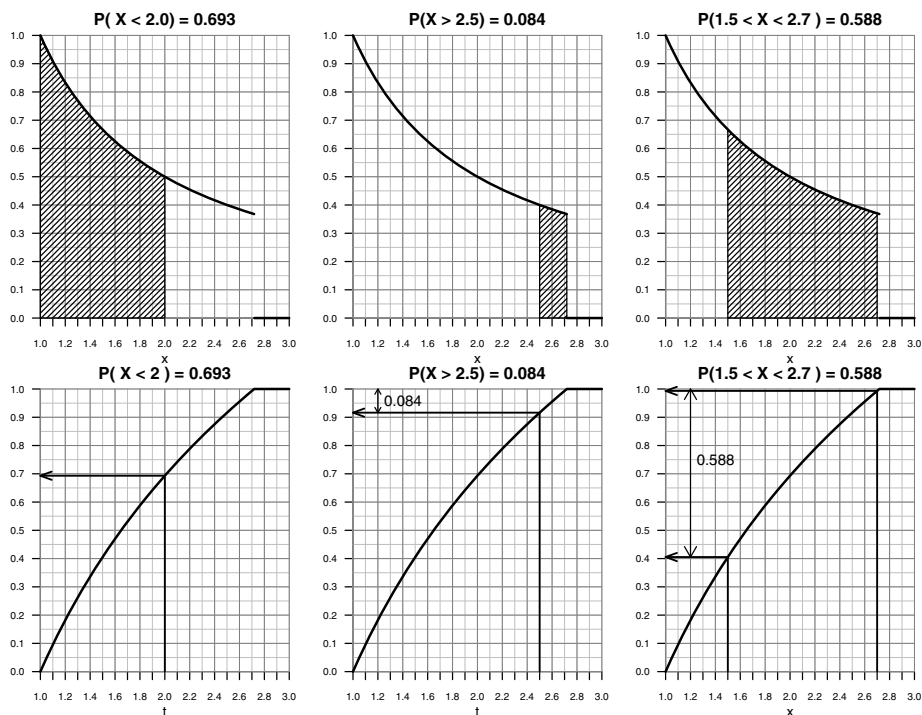
[ 4.14 ] a) Da  $f_X(x) = 1/x \geq 0$  für  $x \in [1, a]$  und  $f_X(x) = 0$  für  $x \notin [1, a]$ , ist  $f_X(x) \geq 0$  für alle  $x \in \mathbb{R}$ . Zweitens, muss die Bedingung

$\int_{-\infty}^\infty f_X(x) dx = \int_1^a \frac{1}{x} dx = [\log x]_1^a = \log(a) - \log(1) = \log(a) = 1$  erfüllt<sup>5</sup> sein. Dies gilt genau dann, wenn  $a = e \approx 2.718282$ .

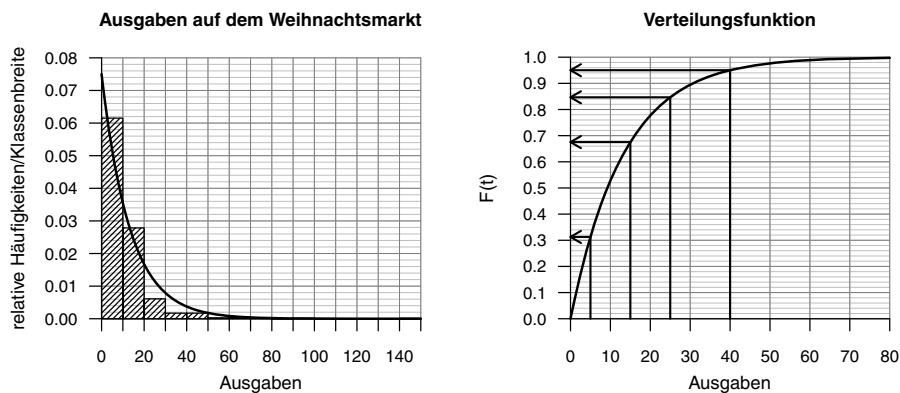
$$\text{b) } F_X(t) = \begin{cases} 0 & -\infty < t \leq 1 \\ \log(t) & 1 \leq t \leq e \\ 1 & e \leq t < \infty \end{cases}$$

<sup>5</sup> Mit log wird hier der natürliche Logarithmus bezeichnet.

c)  $P(X < 2) = F_X(2) = \log(2) \approx 0.693$ ;  $P(X > 2.5) = 1 - P(X \leq 2.5) = 1 - \log(2.5) \approx 0.084$ ; und  $P(1.5 < X \leq 2.7) = \log(2.7) - \log(1.5) \approx 0.588$ .



[ 4.15 ] a) und b) linke Grafik, e) rechte Grafik



c) i)  $\approx 0.777$ ; ii)  $\approx 0.367$ ; iii)  $\approx 0.024$ . d) Anteile:

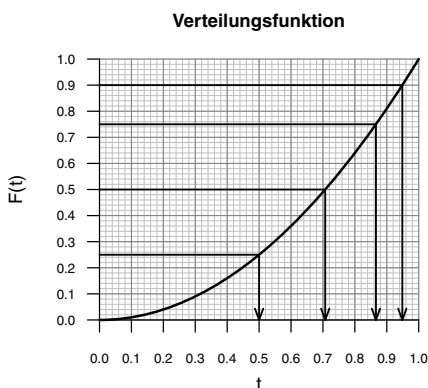
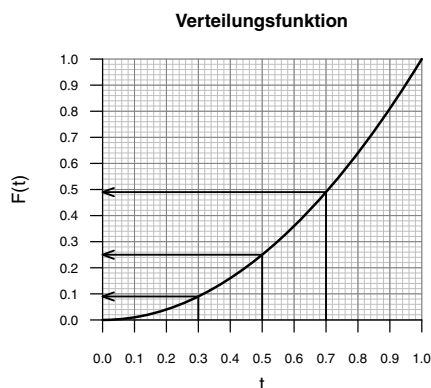
i)  $(433 + 196)/704 \approx 0.893$ ; ii)  $(196 + 43)/704 \approx 0.339$ ; iii)  $8/704 \approx 0.011$

e) siehe rechte Grafik: i)  $\approx 0.31$ ; ii)  $\approx 0.85 - 0.68 = 0.17$ ; iii)  $\approx 1 - 0.95 = 0.05$ .

f) i) 0.37; ii)  $0.50 - 0.21 = 0.29$ ; iii)  $1 - 0.68 = 0.32$ .

[ 4.16 ] a)  $P(X < 0.3) = 0.09$ ;  $P(X > 0.5) = 1 - 0.25 = 0.75$ ;  $P(0.5 < X < 0.7) = 0.49 - 0.25 = 0.24$

b) und c)



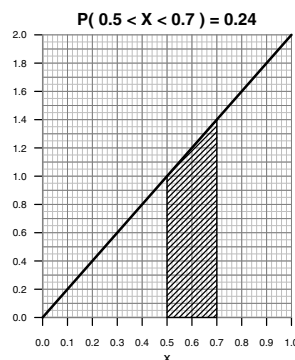
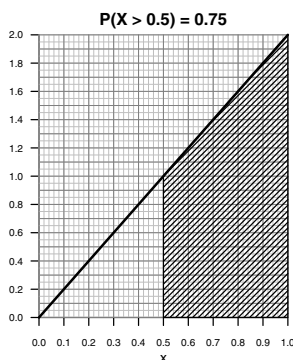
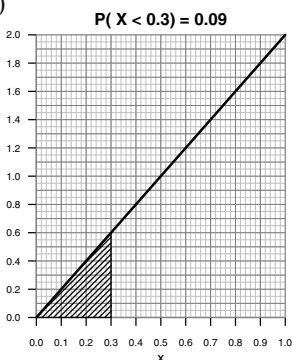
Die in c) gesuchten Werte sind auf der  $x$ -Achse abzulesen:  $0.5$ ,  $\approx 0.71$ ,  $\approx 0.87$ ,  $\approx 0.95$ . d) Für  $t \in [0, 1]$  und  $p \in [0, 1]$  gilt  $F_X(t) = p \iff p = t^2 \iff t = \sqrt{p}$ , d.h. für  $p \in [0, 1]$  gilt  $F_X^{-1}(p) = \sqrt{p}$ . e) Fragen wie in c): Für welches  $t$  hat die Verteilungsfunktion den Wert  $p$ ? f)  $\sqrt{0.05} \approx 0.2236$ ;  $\sqrt{0.1} \approx 0.3162$ ,  $\sqrt{0.25} = 0.5$  und  $\sqrt{0.5} \approx 0.7071$ .

g) `> qbeta(c(0.05, 0.1, 0.25, 0.5), 2, 1)`

[1] 0.2236068 0.3162278 0.5000000 0.7071068

$$h) f_X(x) = \frac{dF_X(x)}{dx} = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

i)



[ 4.17 ] a)  $f_X(x) = 3(1-x)^2 \geq 0$  (und 0 sonst)  $\Rightarrow f_X(x) \geq 0$  für alle  $x \in \mathbb{R}$ .

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^1 (3 - 6x + 3x^2) dx = 3x - 3x^2 + x^3 \Big|_0^1 = 1.$$

$$b) P(X \leq 0.3) = \int_0^{0.3} (3 - 6x + 3x^2) dx = 3x - 3x^2 + x^3 \Big|_0^{0.3} = 0.657$$

$$P(X > 0.5) = \int_{0.5}^1 (3 - 6x + 3x^2) dx = 3x - 3x^2 + x^3 \Big|_{0.5}^1 = 0.125$$

$$P(0.25 < X \leq 0.75) = \int_{0.25}^{0.75} (3 - 6x + 3x^2) dx = 3x - 3x^2 + x^3 \Big|_{0.25}^{0.75} = 0.40625$$

c)  $F_X(t) = \int_{-\infty}^t f_X(x) dx$ . Es gibt drei Fälle: (A)  $t < 0$ , (B)  $0 \leq t \leq 1$ , (C)  $t > 1$ :

(A)  $F_X(t) = \int_{-\infty}^t 0 dx = 0$

(B)  $F_X(t) = \int_0^t (3 - 6x + 3x^2) dx = 3x - 3x^2 + x^3 \Big|_0^t = (3t - 3t^2 + t^3) = 1 - (1 - t)^3$

(C)  $F_X(t) = \int_{-\infty}^t f_X(x) dx = \int_{-\infty}^0 0 dx + \int_0^1 (3 - 6x + 3x^2) dx + \int_1^t 0 dx = 0 + 1 + 0 = 1$

Alternativ: Die Ableitung der Verteilungsfunktion muss die Dichtefunktion ergeben.

d)  $\text{pbeta}(0.3, 1, 3); 1 - \text{pbeta}(0.5, 1, 3); \text{pbeta}(0.75, 1, 3) - \text{pbeta}(0.25, 1, 3)$   
 $\Rightarrow 0.65700 \quad 0.12500 \quad 0.40625$

e)  $\text{qbeta}(c(0.25, 0.5, 0.75), 1, 3) \Rightarrow 0.0914397 \quad 0.2062995 \quad 0.3700395$

[ 4.18 ] a)  $f_X(x) \geq 0$  und  $\int_{-\infty}^{\infty} f_X(x) dx = \int_{-1}^1 \frac{1}{2} dx = \frac{1}{2}x \Big|_{-1}^1 = \frac{1}{2} - (-\frac{1}{2}) = 1$ .

b) Für  $-1 \leq t \leq 1$  ist  $F_X(t) = \int_{-1}^t \frac{1}{2} dx = \frac{1}{2}x \Big|_{-1}^t = \frac{t}{2} + \frac{1}{2} = \frac{t+1}{2}$ , d.h.

$$F_X(t) = \begin{cases} 0 & t < -1 \\ \frac{t+1}{2} & -1 \leq t \leq 1 \\ 1 & t > 1 \end{cases}$$

c)  $EX = EX^3 = 0$  und  $EX^2 = \frac{1}{3}$ .

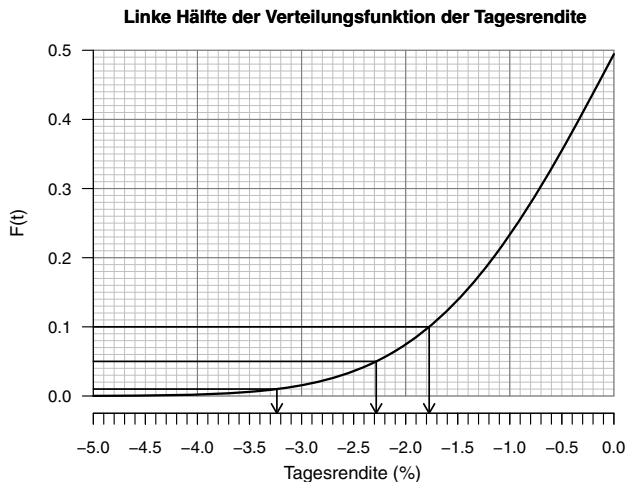
d)  $\text{Var}(X) = \frac{1}{3} - 0^2 = \frac{1}{3}$ .

e) Für  $0 \leq y \leq 1$ :  $P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y}+1}{2} - \frac{-\sqrt{y}+1}{2} = \sqrt{y}$ , d.h.

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \sqrt{y} & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

f)  $f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$

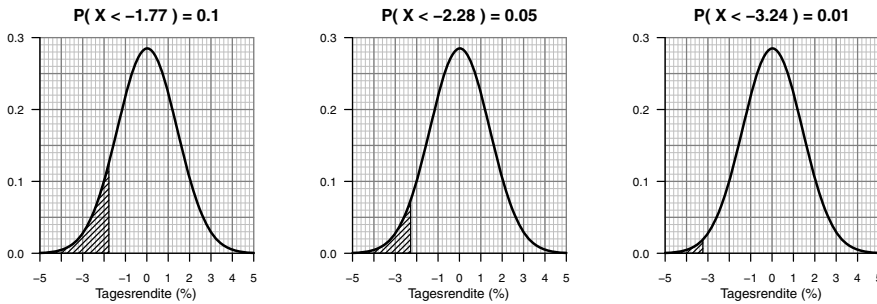
[ 4.19 ] a)



Quantile:  $\approx -1.77; -2.28 \quad -3.24$     VaR:  $\approx 1.8\%; 2.3\%; 3.2\%$



b)



c) Mit Wahrscheinlichkeit 0.9; 0.95 bzw. 0.99 wird der Verlust nicht größer als  $5000 \cdot 0.0177 \approx 89$  Euro;  $5000 \cdot 0.0228 = 114$  Euro bzw.  $5000 \cdot 0.0324 = 162$  Euro.

[ 4.20 ] a)  $E(a) = \int_{-\infty}^{\infty} a f_X(x) dx = a \int_{-\infty}^{\infty} f_X(x) dx = a \cdot 1 = a.$

b) Es gilt  $E(aX^k) = \int_{-\infty}^{\infty} ax^k f_X(x) dx = a \cdot \int_{-\infty}^{\infty} x^k f_X(x) dx = a \cdot E(X^k)$  und damit  $E(aX) = aE(X).$

Allgemein gilt:  $\text{Var}(\heartsuit) = E(\heartsuit^2) - (E(\heartsuit))^2.$

Für  $\heartsuit = aX$ :

$$\text{Var}(aX) = E((aX)^2) - (E(aX))^2 = E(a^2 X^2) - E(aX) \cdot E(aX) = (\text{siehe oben})$$

$$a^2 E(X^2) - aE(X) \cdot aE(X) = a^2 (E(X^2) - (E(X))^2) = a^2 \text{Var}(X).$$

c)  $E(X+b) = \int_{-\infty}^{\infty} (x+b) f_X(x) dx = \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} f_X(x) dx = E(X) + b.$

Für  $\heartsuit = X+b$ :  $\text{Var}(X+b) = E((X+b)^2) - (E(X+b))^2$

$$= E(X^2 + 2bX + b^2) - (E(X)^2 + 2bE(X) + b^2)$$

$$= E(X^2 - (E(X))^2) + (E(2bX) - 2bE(X)) + (E(b^2) - b^2)$$

$$= \text{Var}(X) + (2bE(X) - 2bE(X)) + (b^2 - b^2) = \text{Var}(X).$$

d) Kombinieren wir a) und b), erhalten wir die wichtigen allgemeinen Resultate:

$$\boxed{E(aX+b) = aE(X) + b} \quad \text{und} \quad \boxed{\text{Var}(aX+b) = a^2 \text{Var}(X)}.$$

e) Da unsere Berechnungen für die Varianz ohne Verwendung der Dichtefunktion, sondern direkt für die allgemeine Definition der Varianz durchgeführt wurden, genügt es, die Beweise für die Erwartungswerte zu wiederholen. Offenbar gilt

$$E(aX^k) = \sum_i ax_i^k P(X = x_i) = a \cdot \sum_i x_i^k P(X = x_i) = a \cdot E(X^k) \text{ und}$$

$$E(X+b) = \sum_i (x_i+b) P(X = x_i) = \sum_i x_i P(X = x_i) + b \cdot \sum_i P(X = x_i) = E(X) + b.$$

Also lassen sich diese allgemeinen Resultate auf diskrete Verteilungen übertragen.

[ 4.21 ] a)  $\int_0^{\infty} xe^{-x} dx = 1! = 1.$

b)  $\mu = E(X); \quad \sigma^2 = \text{Var}(X) = E(X^2) - (E(X))^2;$

Schiefe:  $\alpha_3 = \frac{E((X-\mu)^3)}{\sigma^3};$

Kurtosis:  $\alpha_4 = \frac{E((X-\mu)^4)}{\sigma^4}.$

Wir benötigen:

$$E(X) = \int_0^\infty x^2 e^{-x} dx = 2! = 2. \quad E(X^2) = \int_0^\infty x^3 e^{-x} dx = 3! = 6.$$

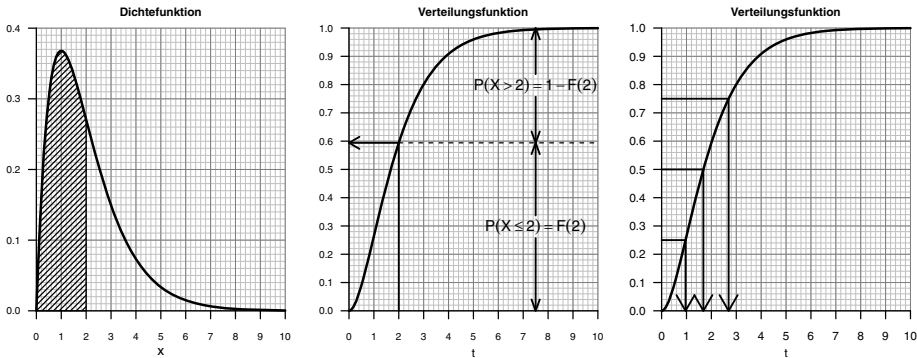
$$E((X - \mu)^3) = \int_0^\infty (x - 2)^3 x e^{-x} dx = \int_0^\infty x^4 e^{-x} - 6x^3 e^{-x} + 12x^2 e^{-x} - 8x e^{-x} dx \\ = 4! - 6 \cdot 3! + 12 \cdot 2! - 8 \cdot 1! = 4.$$

$$E((X - \mu)^4) = \int_0^\infty (x - 2)^4 x e^{-x} dx \\ = \int_0^\infty x^5 e^{-x} - 8x^4 e^{-x} + 24x^3 e^{-x} - 32x^2 e^{-x} + 16x e^{-x} dx \\ = 5! - 8 \cdot 4! + 24 \cdot 3! - 32 \cdot 2! + 16 \cdot 1! = 24.$$

$$\text{Also } \mu = 2; \quad \sigma^2 = 6 - 2^2 = 2; \quad \alpha_3 = \frac{4}{2^{3/2}} = \sqrt{2}; \quad \alpha_4 = \frac{24}{2^2} = 6.$$

c)  $P(X \leq E(X)) = P(X \leq 2) = 0.59$ ;  $P(X \geq E(X)) = P(X \geq 2) = 0.41$ . (Siehe unten linkes und mittleres Bild.)

d)  $t \approx 0.95, 1.7, 2.7$  (Siehe rechtes Bild.)



[ 4.22 ] a) Es gilt  $f_X(x) \geq 0$  für alle  $x \in \mathbb{R}$ , da  $\lambda > 0$  und  $e^{-\lambda x} > 0$  für  $x \geq 0$ . Ferner gilt  $\int_{-\infty}^\infty f_X(x) dx = \int_0^\infty \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^\infty = 0 - (-1) = 1$ .

b) Mit dem Hinweis aus der Fußnote folgt:  $\mu = E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1!}{\lambda} = \frac{1}{\lambda}$  und  $E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2!}{\lambda^2} = \frac{2}{\lambda^2}$ .  $\text{Var}(X) = E(X^2) - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ . Für die Schiefe und Kurtosis benötigen wir ebenfalls das dritte und vierte Moment. Mit dem Hinweis aus der Aufgabenstellung erhalten wir  $E(X^3) = 3!/ \lambda^3 = 6/ \lambda^3$  und  $E(X^4) = 4!/ \lambda^4 = 24/ \lambda^4$ . Mit dem Hinweis aus Aufg. 21 erhalten wir

$$\begin{aligned} E((X - \mu)^3) &= E(X^3 - 3X^2 \cdot \mu + 3X \cdot \mu^2 - \mu^3) \\ &= E(X^3) - 3E(X^2) \cdot \mu + 3\mu \cdot \mu^2 - \mu^3 \\ &= 3!/ \lambda^3 - 3 \cdot (2/ \lambda^2) \cdot (1/ \lambda) + 3/ \lambda^3 - 1/ \lambda^3 = 2/ \lambda^3 \\ E((X - \mu)^4) &= E(X^4 - 4X^3 \cdot \mu + 6X^2 \cdot \mu^2 - 4X \cdot \mu^3 + \mu^4) \\ &= E(X^4) - 4E(X^3) \cdot \mu + 6E(X^2) \cdot \mu^2 - 4\mu \cdot \mu^3 + \mu^4 = \\ &= 24/ \lambda^4 - 24/ \lambda^4 + 12/ \lambda^4 - 4/ \lambda^4 + 1/ \lambda^4 = 9/ \lambda^4 \end{aligned}$$

$$\text{Schiefe: } \alpha_3 = \frac{E((X - \mu)^3)}{\sigma^3} = \frac{2}{\lambda^3} \cdot \frac{\lambda^3}{1} = 2 \quad \text{Kurtosis: } \alpha_4 = \frac{E((X - \mu)^4)}{\sigma^4} = \frac{9}{\lambda^4} \cdot \frac{\lambda^4}{1} = 9$$

## 2.5 Diskrete Verteilungen - Lösungen

[ 5.1 ] a) W b) W c) F d) W e) W

[ 5.2 ] a) W b) W c) F d) W e) W f) F g) F

[ 5.3 ] a) F b) W c) W d) W e) W f) W g) W h) F

[ 5.4 ] a) Die verschiedenen Ausgaben geben die Wahrscheinlichkeiten  $P(X = k)$  für  $k = 45, 46, \dots, n$  einer  $b(n, 0.9)$ -Verteilung an mit  $n = 51$  (erste Zeile) bis  $n = 54$  (letzte Zeile). Daraus sehen wir  $P(X > 50) = P(X \geq 51)$  und

$$\begin{array}{llll} P(X = 51) & = & 0.005 & = 0.005 \\ P(X = 51) + P(X = 52) & = & 0.024 + 0.004 & = 0.028 \\ P(X = 51) + P(X = 52) + P(X = 53) & = & 0.064 + 0.022 + 0.004 & = 0.090 \\ P(X = 51) + P(X = 52) + P(X = 53) + P(X = 54) & = & 0.115 + 0.06 + 0.02 + 0.003 & = 0.198 \end{array}$$

Hier bezieht sich wieder die erste Zeile auf  $n = 51$ , die letzte auf  $n = 54$ .

Somit ist die Antwort auf a) 51 für  $\alpha = 0.01$ ; 52 für  $\alpha = 0.005$ ; und 53 für  $\alpha = 0.1$ .

b) Ebenso ergibt sich daraus unmittelbar für b) 0.05; 0.028; 0.090

c)  $P(X > 50) = 0.028$ ; 0.090; 0.198 für  $n = 52, 53, 54$

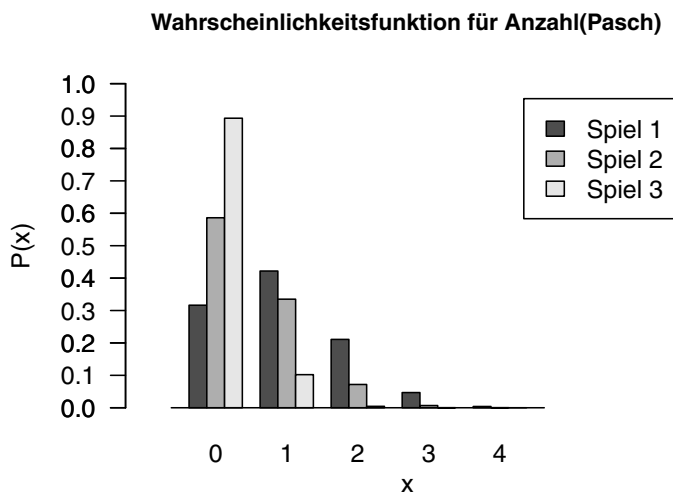
[ 5.5 ] Die **R**-Ausgabe berechnet die Verteilungsfunktion  $F_X(t) = P(X \leq t)$  einer  $b(n, 0.9)$ -verteilten Zufallsvariablen mit  $n = 55, 56, \dots, 65$  und  $\pi = 0.9$  an den Stellen  $t = 49$  und  $t = 50$ . Man kann daraus ablesen, dass die Unterbesetzungswahrscheinlichkeit  $P(X \leq 49) = 0.476$  für  $n = 55$  ist. Für  $n = 59$  ist  $P(X \leq 49) = 0.067$  erstmals  $\leq 0.10$ . Für  $\alpha = 0.05$ ; 0.01; 0.005 muss  $n = 60$ ; 62; 63 sein.

[ 5.6 ] a) Binomial  $b(4, \pi_i)$  mit  $\pi_i = 1/4$ ;  $1/8$ ;  $1/36$ .  $E(X_i) = n\pi_i = 1$ ;  $1/2$ ;  $1/9$ .  $\text{Var}(X_i) = n\pi_i(1 - \pi_i) = 3/4$ ;  $7/16$ ;  $(1/9) \cdot (35/36) \approx 0.1080$

b)

		$x$	0	1	2	3	4
$\pi$	$n$	$\binom{n}{x}$	1	4	6	4	1
1/4	4	$P(X_1 = x)$	0.3164	0.4219	0.2109	0.0469	0.0039
1/8	4	$P(X_2 = x)$	0.5862	0.3350	0.0718	0.0068	0.0002
1/36	4	$P(X_3 = x)$	0.8934	0.1021	0.0044	0.0001	0.0000

c)



d)

$i$	$P(X_i \leq 0)$	$P(X_i \leq 1)$	$P(X_i \leq 2)$	$P(X_i \leq 3)$	$P(X_i \leq 4)$
1	0.3164	0.7383	0.9492	0.9961	1.0000
2	0.5862	0.9212	0.9930	0.9998	1.0000
3	0.8934	0.9955	0.9999	1.0000	1.0000

$i$	$P(X_i > 0)$	$P(X_i \geq 2)$	$P(1 \leq X_i \leq 3)$	$P(1 < X_i \leq 3)$	$P(0 < X_i < 3)$
1	0.6836	0.2617	0.6797	0.2578	0.6328
2	0.4138	0.0788	0.4136	0.0786	0.4068
3	0.1066	0.0045	0.1066	0.0045	0.1065

[ 5.7 ] a) Es handelt sich um ein symmetrisches Zufallsexperiment, d.h. jede der 10 Ziffern wird mit Wahrscheinlichkeit  $1/10 = 0.1$  als Superzahl gezogen.

b) In allen Aufgaben gibt die relevante Zufallsvariable  $X$  die Anzahl der Erfolge in  $n = 10$  unabhängigen Ziehungen mit Erfolgswahrscheinlichkeit  $\pi = 0.1$  an, d.h.

$$X \sim b(10, 0.1), \text{ d.h. } P(X = k) = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k} = \binom{n}{k} \pi^k (1-\pi)^{n-k}.$$

$$\text{i) } P(X = 4) = \binom{10}{4} 0.1^4 0.9^6 \approx 0.0112; \quad \text{dbinom}(4, 10, 0.1)$$

$$\text{ii) } P(X = 0) = \binom{10}{0} 0.1^0 0.9^{10} \approx 0.3487; \quad \text{dbinom}(0, 10, 0.1)$$

$$\text{iii) } P(X = 2) = \binom{10}{2} 0.1^2 0.9^8 \approx 0.1937; \quad \text{dbinom}(2, 10, 0.1)$$

$$\text{iv) } P(X = 1) = \binom{10}{1} 0.1^1 0.9^9 \approx 0.3874; \quad \text{dbinom}(1, 10, 0.1)$$

- c) i)  $X \sim b(3, 0.1) \Rightarrow P(X=3) = \binom{3}{3} 0.1^3 0.9^0 = 0.001$ ; `dbinom(3, 3, 0.1)`  
 ii)  $X \sim b(6, 0.1) \Rightarrow P(X=4) = \binom{6}{4} 0.1^4 0.9^2 \approx 0.0012$ ; `dbinom(4, 6, 0.1)`  
 iii)  $0.1 \cdot 0.1 = 0.01$ ; iv)  $0.1 \cdot 0.1 \cdot 0.1 = 0.001$ ; d)  $0.1 \cdot 0.1 \cdot 0.1 = 0.001$ ;  
 e)  $X \sim b(10, 0.1) \Rightarrow E(X) = n\pi = 10 \cdot 0.1 = 1$ ;  $\text{Var}(X) = n\pi(1-\pi) = 10 \cdot 0.1 \cdot 0.9 = 0.9$ ; Standardabweichung:  $\sqrt{\text{Var}(X)} = \sqrt{0.9} \approx 0.9487$

[ 5.8 ] Die Anzahl  $X$  der Linkshänder ist binomialverteilt mit  $n = 24$  und  $\pi = 0.1$  bzw.  $0.15$ . a)  $\pi = 0.1 \Rightarrow P(X=2) = \binom{24}{2} \cdot 0.1^2 \cdot 0.9^{22} \approx 0.272$ ;  $\pi = 0.15 \Rightarrow P(X=2) = \binom{24}{2} \cdot 0.15^2 \cdot 0.85^{22} \approx 0.174$ . `dbinom(2, 24, c(0.1, 0.15))`  
 b)  $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) \approx 0.564$  für  $\pi = 0.1$  bzw.  $\approx 0.280$  für  $\pi = 0.15$ . `pbinom(2, 24, c(0.1, 0.15))`  
 c)  $P(X \geq 2) = 1 - (P(X=0) + P(X=1)) \approx 0.708$  für  $\pi = 0.1$  bzw.  $\approx 0.894$  für  $\pi = 0.15$ . `1 - pbinom(1, 24, c(0.1, 0.15))`  
 d)  $P(X > 2) = 1 - P(X \leq 2) = 1 - (P(X=0) + P(X=1) + P(X=2)) \approx 0.436$  für  $\pi = 0.1$  bzw.  $\approx 0.720$  für  $\pi = 0.15$ . `1 - pbinom(2, 24, c(0.1, 0.15))`.

[ 5.9 ] a)  $b(100, 0.1)$ ; b)  $E(X) = 100 \cdot 0.1 = 10$ ;  $\text{Var}(X) = 100 \cdot 0.1 \cdot 0.9 = 9$ ;  $\sigma = \sqrt{9} = 3$ ; c) Achtung: verwendet man durchgehend **R**, erhält man leicht andere Ergebnisse. Die Abweichungen erklären sich durch Rundungsfehler.  $P(X=6) = 0.060$ ;  $P(X \leq 6) = 0.118$ ;  $P(X < 10) = P(X \leq 9) = 0.451$ ;  $P(X > 10) = 1 - P(X \leq 10) = 0.416$ . `dbinom(6, 100, 0.1)`;  
`pbinom(6, 100, 0.1)`; `pbinom(9, 100, 0.1)`; `1 - pbinom(10, 100, 0.1)`

[ 5.10 ] a) W b) W c) F d) F e) W f) F

[ 5.11 ] a) W b) W c) W d) F e) W

[ 5.12 ] a) Hypergeometrisch, zuerst mit (A)  $N_e = 20, N_m = 30, n = 8$  und möglichen Werten  $0, 1, \dots, 8$ ; dann mit (B)  $N_e = 6, N_m = 14, n = 8$  und möglichen Werten  $0, 1, \dots, 6$ .

b) Im **R**-Ausdruck waren die Werte der Verteilungsfunktionen angegeben, demnach gilt für die Wahrscheinlichkeitsfunktionen:

$x$	0	1	2	3	4	5	6	7	8
$P(X_A = x)$	0.0109	0.0758	0.2102	0.3026	0.2473	0.1172	0.0314	0.0044	0.0002
$P(X_B = x)$	0.0238	0.1635	0.3576	0.3178	0.1192	0.0174	0.0007	0.0000	0.0000

c) Die Ergebnisse lassen sich aus der Tabelle unter b) ablesen, da nur die entsprechenden Wahrscheinlichkeiten summiert werden müssen, siehe auch **R**-Programm. Alternativ kann man natürlich die Verteilungsfunktion (siehe **R**-Ausdruck in der Aufgabenstellung) benutzen, d.h.

$$\text{i) } P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$$

$$\text{ii) } P(X \geq a) = 1 - P(X < a) = 1 - P(X \leq a - 1) = 1 - F_X(a - 1)$$

$$\text{iii) } P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - F_X(a)$$

$$\text{iv) } P(a \leq X < b) = P(a - 1 < X \leq b - 1) = F_X(b - 1) - F_X(a - 1)$$

$$\text{v) } P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$$

$$\text{vi) } P(a < X < b) = P(a < X \leq b - 1) = F_X(b - 1) - F_X(a)$$

i) A: 0.7031; 0.0046    B: 0.4551; 0.0000    ii) A: 0.9133; 0.0360    B: 0.8127; 0.0007  
 iii) A: 0.8773; 0.4003    B: 0.8120; 0.1373    iv) A: 0.7601; 0.3959    B: 0.7946; 0.1373  
 v) A: 0.6671; 0.1530    B: 0.4544; 0.0181    vi) A: 0.5499; 0.1486    B: 0.4370; 0.0181

[ 5.13 ] a) i) Mit der hypergeometrischen Verteilung mit den Parametern  $N_e = 1$ ,  $N_m = 48$ ,  $n = 6$ . ii) Gesucht ist  $P(X = 1) = \frac{\binom{1}{1} \cdot \binom{48}{5}}{\binom{49}{6}} = \frac{48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \cdot \frac{6}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{6}{49}$ .

In **R**: `dhyper(1, 1, 48, 6)`    Ausgabe: 0.1224490

b) Mit Hilfe der Unabhängigkeit folgt: i)  $\frac{1}{49}$ ; ii)  $\frac{48}{49} \cdot \frac{1}{48} = \frac{1}{49}$ ; iii)  $\frac{48}{49} \cdot \frac{47}{48} \cdot \frac{1}{47} = \frac{1}{49}$ . Analog ergibt sich auch in den Fällen iv) - vi) jeweils  $\frac{1}{49}$  und somit ist die Summe:  $6 \cdot \frac{1}{49} = \frac{6}{49}$ . c) Wenn  $49 \cdot 6$  Zahlen gezogen wird, erwartet man die Lieblingszahl  $\frac{49 \cdot 6}{49}$  mal, d.h.  $49 \cdot \pi = \frac{49 \cdot 6}{49}$ . Auflösen nach  $\pi$  ergibt  $\pi = \frac{6}{49}$ . d) Es geht jeweils um die Anzahl  $X$  der Erfolge in  $n = 10$  unabhängigen Versuchen mit Erfolgswahrscheinlichkeit  $\pi = 6/49$ , d.h.  $X \sim b(10, 6/49)$ .

$$\text{i) } P(X = 1) = \binom{10}{1} \left(\frac{6}{49}\right)^1 \left(1 - \frac{6}{49}\right)^9 \approx 0.3779; \text{ dbinom}(1, 10, 6/49)$$

$$\text{ii) } P(X = 0) = \left(1 - \frac{6}{49}\right)^{10} \approx 0.2708; \text{ dbinom}(0, 10, 6/49)$$

$$\text{iii) } P(X = 4) = \binom{10}{4} \left(\frac{6}{49}\right)^4 \left(1 - \frac{6}{49}\right)^6 \approx 0.0216; \text{ dbinom}(4, 10, 6/49)$$

$$\text{e) } b(10, 6/49) \text{ mit } E(X) = n\pi = 10 \cdot 6/49 \approx 1.224; \text{ Var}(X) = n\pi(1 - \pi) \approx 1.075.$$

[ 5.14 ] a) Binomialverteilung mit  $n = 8$  und  $\pi = 250/1000 = 1/4 = 0.25$

b)  $P(X = x) = \binom{8}{x} 0.25^x 0.75^{8-x}$ . Für  $x = 0, 1, 2$  ergibt sich 0.1001; 0.2670; 0.3115 oder in **R**: `round(dbinom(0:2, 8, 0.25), 4)`

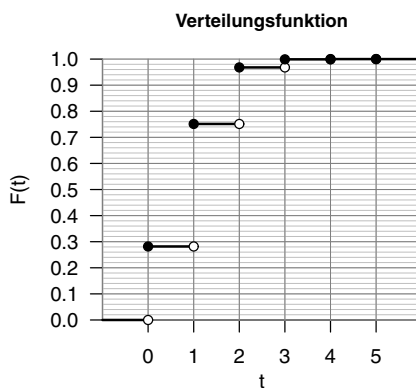
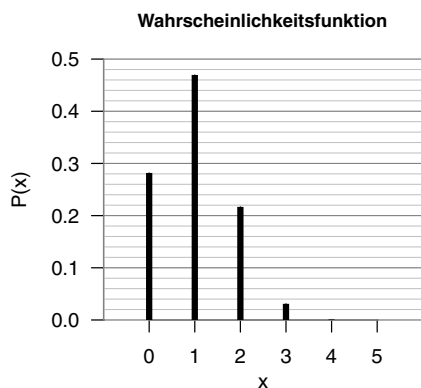
c) Hypergeom.:  $F_X(t) = 0.0992; 0.3662; 0.6789$     `phyper(0:2, 250, 750, 8)`  
 Binomial:  $F_X(t) = 0.1001; 0.3671; 0.6786$     `pbinom(0:2, 8, 0.25)`

[ 5.15 ] a) Hypergeometrisch mit  $N_e = 4, N_m = 16$  und  $n = 5$ , d.h.  $h(4, 16, 5)$ .

b)

$x$	0	1	2	3	4	sonst	<b>R</b> -Befehl
$P_X(x)$	0.2817	0.4696	0.2167	0.0310	0.0010	0	<code>dhyper(0:4, 4, 16, 5)</code>

$t$	$t < 0$	$0 \leq t < 1$	$1 \leq t < 2$	$2 \leq t < 3$	$3 \leq t < 4$	$4 \leq t$	<b>R-Befehl</b>
$F_X(t)$	0	0.2817	0.7513	0.9680	0.9990	1	phyper(0:4,4,16,5)



c) Schwerpunkt  $\approx 1$  d)  $E(X) = 1 \cdot 0.4696 + 2 \cdot 0.2167 + 3 \cdot 0.0310 + 4 \cdot 0.0010 = 1$   
`sum(0:4*dhyper(0:4,4,16,5))`

$$E(X^2) = 1^2 \cdot 0.4696 + 2^2 \cdot 0.2167 + 3^2 \cdot 0.0310 + 4^2 \cdot 0.0010 = 1.6314$$

$$\text{sum}((0:4)^2 * \text{dhyper}(0:4,4,16,5))$$

$$\text{Var}(X) = 1.6314 - 1^2 = 0.6314; \text{ Standardabweichung: } \sqrt{0.6314} \approx 0.7946$$

e)  $h(16,4,5)$

[ 5.16 ] a) F b) W c) W d) F e) W f) W g) F h) F i) F

[ 5.17 ] a) W b) W c) W d) F e) W f) W

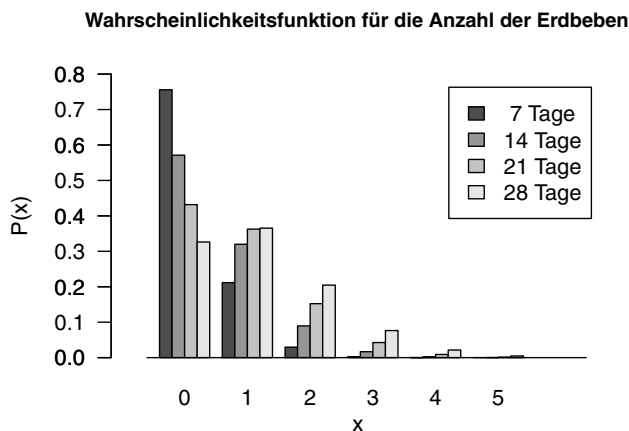
[ 5.18 ] a)  $E(X) = \text{Var}(X) = \lambda \cdot t = 0.28; 0.56; 0.84; 1.12$

Für b) - f) siehe auch **R**-Programm.

b)

$t$	$\lambda \cdot t$	$P(N_t = 0)$	$P(N_t = 1)$	$P(N_t = 2)$	$P(N_t = 3)$
7	0.28	0.7558	0.2116	0.0296	0.0028
14	0.56	0.5712	0.3199	0.0896	0.0167
21	0.84	0.4317	0.3626	0.1523	0.0426
28	1.12	0.3263	0.3654	0.2046	0.0764

c)



d)

$t$	$\lambda \cdot t$	$P(N_t \leq 0)$	$P(N_t \leq 1)$	$P(N_t \leq 2)$	$P(N_t \leq 3)$
7	0.28	0.7558	0.9674	0.9970	0.9998
14	0.56	0.5712	0.8911	0.9807	0.9974
21	0.84	0.4317	0.7943	0.9466	0.9892
28	1.12	0.3263	0.6917	0.8963	0.9727

e)

$t$	$\lambda \cdot t$	$P(N_t > 1)$	$P(N_t \geq 1)$	$P(N_t > 2)$	$P(N_t \geq 3)$
7	0.28	0.0326	0.2442	0.0030	0.0030
14	0.56	0.1089	0.4288	0.0193	0.0193
21	0.84	0.2057	0.5683	0.0534	0.0534
28	1.16	0.3083	0.6737	0.1037	0.1307

f)

$t$	$\lambda \cdot t$	$P(1 \leq N_t \leq 3)$	$P(1 \leq N_t < 3)$	$P(1 < N_t \leq 3)$	$P(1 < N_t < 3)$
7	0.28	0.2440	0.2412	0.0324	0.0296
14	0.56	0.4262	0.4095	0.1063	0.0896
21	0.84	0.5575	0.5149	0.1949	0.1523
28	1.16	0.6464	0.5700	0.2810	0.2046

[ 5.19 ] Die Poissonverteilung  $Po(\lambda)$  hat die Wahrscheinlichkeitsfunktion

$$P_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \text{ für } x = 0, 1, 2, \dots, \text{ und Verteilungsfunktion } F_X(x) = e^{-\lambda} \sum_{k=0}^x \frac{\lambda^k}{k!}.$$

Damit ergibt sich  $P(Y = y) = P(X - 1 = y) = P((x = y - 1))$ , da  $Y = X + 1$ . Für die Berechnungen vergleiche gegebenenfalls auch **R**-Befehle:



```
dpois(0:5,0.5)      # für a) und b)
ppois(0:5,0.5)      # für c)
1-ppois(0:5,0.5)    # für d)
1.0 ; 1-ppois(0:4,0.5) # für e), da Y immer mindestens 1 ist
```

a) 0.6065; b) 0.3033, 0.0758, 0.0126, 0.0016, 0.0002; c) 0.6065, 0.9098, 0.9856, 0.9982, 0.9998, 1.0000; d) 0.3935, 0.0902, 0.0144, 0.0018, 0.0002, 0.0000; e) 1.000, 0.3935, 0.0902, 0.0144, 0.0018, 0.0002; f)  $E(Y) = E(X+1) = E(X) + 1 = \lambda + 1 = 1.5$ ;  $\text{Var}(Y) = \text{Var}(X+1) = \text{Var}(X) = \lambda = 0.5$ .

[ 5.20 ] a) Wir wissen, dass  $P(X_0 > 0) = 1 - P(X_0 = 0) = 1 - \frac{\lambda^0}{0!} e^{-\lambda} = 1 - e^{-\lambda}$ , also  $P_{X_1}(x) = \frac{P_{X_0}(x)}{1 - e^{-\lambda}} = \frac{\lambda^x}{x!} e^{-\lambda}$  für  $x = 1, 2, 3, \dots$ . Um zu zeigen, dass dies tatsächlich eine Wahrscheinlichkeitsfunktion definiert, ist zu prüfen, dass  $P_{X_1}(x) \geq 0$  für alle  $x = 1, 2, \dots$  und  $\sum_{x=1}^{\infty} P_{X_1}(x) = 1$ . Da  $P_{X_0}(x) \geq 0$  und  $1 - e^{-\lambda} > 0$ , ist  $P_{X_1}(x) \geq 0$  für

alle  $x = 1, 2, \dots$ . Aus dem Hinweis folgt, dass  $\sum_{x=1}^{\infty} P_{X_0}(x) = 1 - P_{X_0}(0) = 1 - e^{-\lambda}$ .

Damit gilt  $\sum_{x=1}^{\infty} P_{X_1}(x) = \frac{1}{1 - e^{-\lambda}} \sum_{x=1}^{\infty} P_{X_0}(x) = \frac{1 - e^{-\lambda}}{1 - e^{-\lambda}} = 1$ . b)  $P_{X_1}(1) = \frac{\lambda^1 e^{-\lambda}}{1 - e^{-\lambda}} = \frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}} = \frac{0.5 \cdot e^{-0.5}}{1 - e^{-0.5}} \approx 0.7707$ .  $\text{dpois}(1, 0.5) / (1 - \text{dpois}(0, 0.5))$   $P_{X_1}(2) = \frac{\lambda^2 e^{-\lambda}}{1 - e^{-\lambda}} = \frac{\lambda^2 e^{-\lambda}}{1 - e^{-\lambda}} = \frac{0.125 \cdot e^{-0.5}}{1 - e^{-0.5}} = 0.1927$ .  $\text{dpois}(2, 0.5) / (1 - \text{dpois}(0, 0.5))$

c)  $E(X_1) = \sum_{x=1}^{\infty} x P_{X_1}(x) = \sum_{x=1}^{\infty} x \frac{P_{X_0}(x)}{1 - e^{-\lambda}} = \frac{1}{1 - e^{-\lambda}} \sum_{x=0}^{\infty} x P_{X_0}(x) = \frac{EX_0}{1 - e^{-\lambda}} = \frac{\lambda}{1 - e^{-\lambda}}$ .

Für  $\lambda = 0.5$  ist  $E(X_0) = \lambda = 0.5$  und  $E(X_1) = \frac{0.5}{1 - e^{-0.5}} \approx 1.2707$ .

Da die Anzahl der geschenkten Bücher mindestens 1 ist, muss auch der Erwartungswert mindestens 1 sein.

[ 5.21 ] a) Es wird jeweils für  $x = 0, 1, 2, 3$  die Wahrscheinlichkeitsfunktion der Binomialverteilung mit den Parametern  $n = 40$ ;  $\pi = 0.1$  bzw.  $n = 80$ ;  $\pi = 0.05$  bzw.  $n = 160$ ;  $\pi = 0.025$  berechnet.

b) Wenn  $n$  groß und  $\pi$  klein ist, kann die Binomialverteilung durch eine Poissonverteilung mit  $\lambda = n \cdot \pi$  approximiert werden. In allen drei Fällen ist  $n \cdot \pi = 4$ .

c) Mithilfe der Formel  $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$  wobei  $\lambda = 4$  und  $k = 0, 1, 2, 3$ , erhalten wir 0.0183; 0.0733; 0.1465; 0.1954, oder mit **R**: `dpois(0:3, 4)`.

d) Siehe Tabelle:

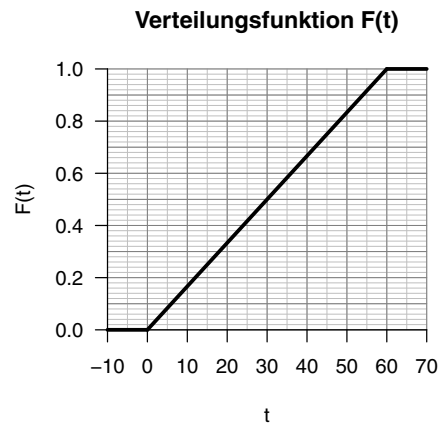
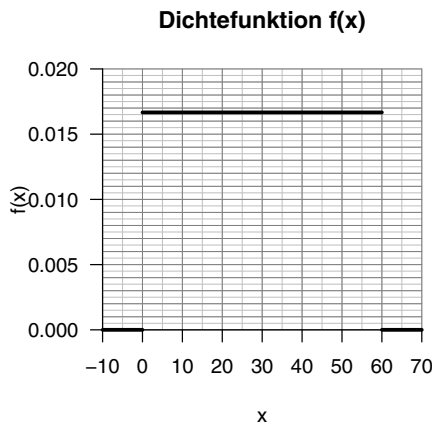
Verteilung	$F_X(t)$	$t=0$	$t=1$	$t=2$	$t=3$	R-Befehl
$b(40, 0.1)$	$\sum_{x=0}^t \binom{40}{x} 0.1^x 0.9^{40-x}$	0.015	0.080	0.223	0.423	<code>pbinom(0:3, 40, 0.1)</code>
$b(80, 0.05)$	$\sum_{x=0}^t \binom{80}{x} 0.05^x 0.95^{80-x}$	0.017	0.086	0.231	0.428	<code>pbinom(0:3, 80, 0.05)</code>
$b(160, 0.025)$	$\sum_{x=0}^t \binom{160}{x} 0.025^x 0.975^{160-x}$	0.017	0.089	0.234	0.431	<code>pbinom(0:3, 160, 0.025)</code>
$Po(4)$	$\sum_{x=0}^t \frac{4^x}{x!} e^{-4}$	0.018	0.092	0.238	0.434	<code>ppois(0:3, 4)</code>

[ 5.22 ] a)  $1 - \alpha = 0.95 \Rightarrow x_{un} = 1; x_{ob} = 10$ , d.h.  $[1; 10]$  und  $1 - \alpha = 0.99 \Rightarrow x_{un} = 0; x_{ob} = 11$ , d.h.  $[0; 11]$ . b)  $1 - \alpha = 0.95 \Rightarrow x_{un} = 2; x_{ob} = 9$ , d.h.  $[2; 14]$  und  $[0; 9]$  und  $1 - \alpha = 0.99 \Rightarrow x_{un} = 1; x_{ob} = 11$ , d.h.  $[1; 14]$  und  $[0; 11]$ .

## 2.6 Stetige Verteilungen - Lösungen

[ 6.1 ] a) Eine Zufallsvariable  $X$  ist rechteckverteilt, wenn angenommen wird, dass die möglichen Ausprägungen von  $X$  im Intervall von  $a$  bis  $b$  gleichmäßig verteilt sind. Hier bietet sich, da das Intervall  $[a, b]$  als  $[0, 60]$  gegeben ist, eine Zeitmessung an, wobei nur die Minute betrachtet wird, in der ein zufälliges Ereignis eintritt, z.B. ein Anruf in einem Call-Center.

b)



$$c) f(x) = \begin{cases} 1/60 & 0 \leq x \leq 60 \\ 0 & \text{sonst} \end{cases}$$

$$F(t) = \begin{cases} 0 & t < 0 \\ t/60 & 0 \leq t \leq 60 \\ 1 & t > 60 \end{cases}$$

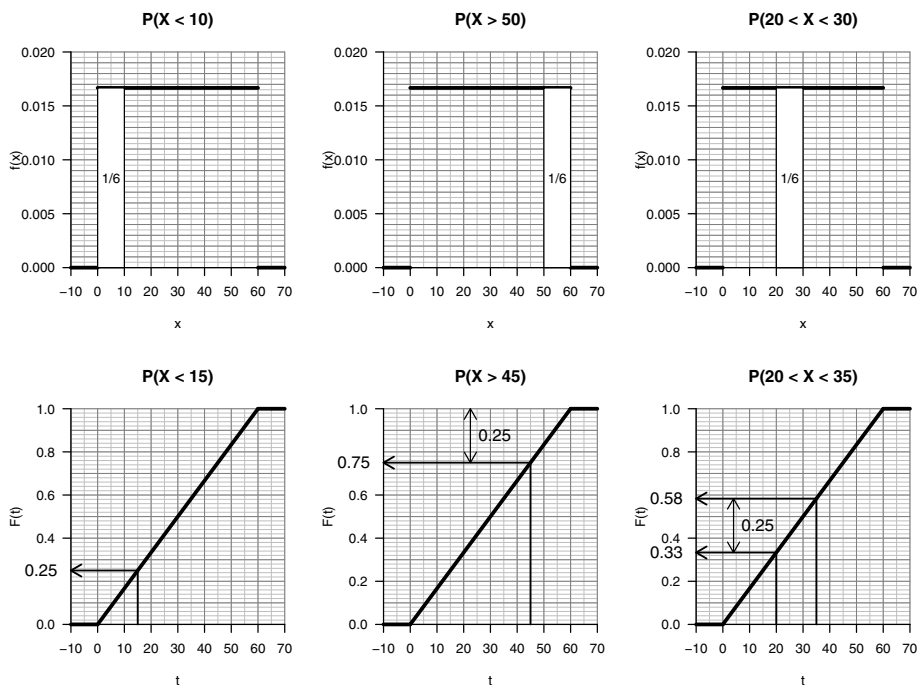
d)  $E(X) = 30$   $\text{Var}(X) = (60 - 0)^2 / 12 = 300$ . e)  $P(X < 10) = P(X > 50) = P(20 < X \leq 30) = 1/6$ ;  $P(X < 15) = P(X > 45) = P(20 < X \leq 35) = 1/4$ .

Wir sehen, Rechteck- oder Gleichverteilung meint, dass alle Intervalle gleicher

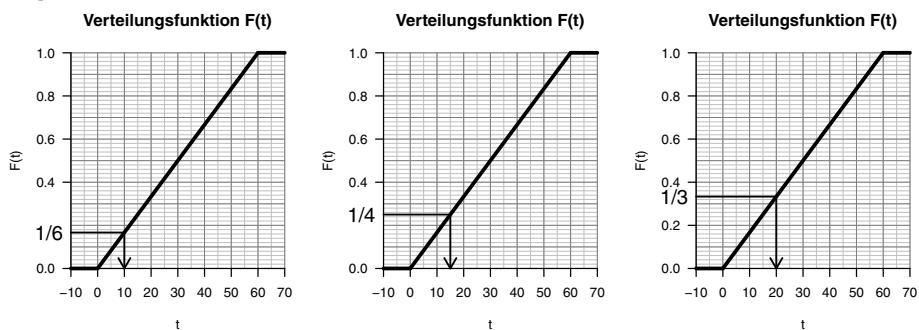
Länge im Wertebereich von  $X$  stets dieselbe Wahrscheinlichkeit haben. Mit **R**:

`punif(10, 0, 60); 1-punif(50, 0, 60); punif(30, 0, 60)-punif(20, 0, 60)` bzw.

`punif(15, 0, 60); 1-punif(45, 0, 60); punif(35, 0, 60)-punif(20, 0, 60)`



f)  $t_p = 10; 15; 20$ .



g)  $F(t_p) = p \Leftrightarrow t_p/60 = p \Leftrightarrow t_p = 60 \cdot p$  für  $0 \leq p \leq 1$ .

[ 6.2 ] a) W b) F c) W d) F e) F f) F g) W

[ 6.3 ] a)  $X \sim \text{Exp}(\lambda) \Rightarrow E(X) = 1/\lambda = 1/0.04 = 25$ ;  $\text{Var}(X) = 1/\lambda^2 = 625$

b) i)  $P(X \leq 25) = \int_0^{25} f_X(x) dx = \int_0^{25} 0.04e^{-0.04x} dx = -e^{-0.04x} \Big|_0^{25} = 1 - e^{-0.04 \cdot 25} = 1 - e^{-1} \approx 0.632$ . Entsprechend:  $P(X < 50) \approx 0.865$  und  $P(X < 75) \approx 0.950$ .

ii)  $P(X > 60) = \int_{60}^{\infty} f(x) dx = -e^{-0.04x} \Big|_{60}^{\infty} = 0 + e^{-0.04 \cdot 60} = e^{-2.4} \approx 0.091$ . Entsprechend:  $P(X > 90) \approx 0.027$ ;  $P(X > 120) \approx 0.008$ . iii)  $P(25 < X < 50) = \int_{25}^{50} f(x) dx = -e^{-0.04x} \Big|_{25}^{50} = -e^{-0.04 \cdot 50} + e^{-0.04 \cdot 25} = -e^{-2} + e^{-1} \approx 0.233$ .

c)  $F_X(t) = 0$  für  $t < 0$ . Für  $t > 0$  ist  $F_X(t) = \int_0^t f_X(x) dx = \int_0^t e^{-0.04x} dx = -e^{-0.04x} \Big|_0^t = -e^{-0.04t} - (-1) = 1 - e^{-0.04t}$ . i)  $P(X \leq 25) = F_X(25) = 1 - e^{-0.04 \cdot 25} = 1 - e^{-1} \approx 0.632$ . Entsprechend:  $P(X \leq 50) = F_X(50)$ ;  $P(X \leq 75) = F_X(75)$ . In **R**: `pexp(c(25, 50, 75), 0.04)` ii)  $P(X > 60) = 1 - P(X \leq 60) = 1 - F_X(60) = 1 - (1 - e^{-0.04 \cdot 60}) = e^{-2.4} \approx 0.091$ . Entsprechend:  $P(X > 90) = 1 - F_X(90)$ ;  $P(X > 120) = 1 - F_X(120)$ . In **R**: `1-pexp(c(60, 90, 120), 0.04)` iii)  $P(25 < X < 50) = F_X(50) - F_X(25) = (1 - e^{-0.04 \cdot 50}) - (1 - e^{-0.04 \cdot 25}) = -e^{-2} + e^{-1} \approx 0.233$ . In **R**: `pexp(50, 0.04) - pexp(25, 0.04)` d) Für  $t > 0$  gilt  $P(X > t) = 1 - F_X(t) = 1 - (1 - e^{-0.04t}) = e^{-0.04t}$ . Damit ergeben sich für  $t = 30, 60, 100, 200$  und  $365$  folgende Wahrscheinlichkeiten  $0.3012; 0.0907; 0.0183; 0.0003; 0.0000$ . In **R**: `1-pexp(c(30, 60, 100, 200, 365), 0.04)`

[ 6.4 ] a) Die Überlebensfunktion ist  $S(t) = P(X > t) = 1 - F(t) = e^{-\lambda t}$ .

b)  $F_X(t_p) = p \iff 1 - e^{-\lambda t_p} = p \iff 1 - p = e^{-\lambda t_p} \iff \log(1 - p) = -\lambda t_p \iff t_p = -\log(1 - p)/\lambda \geq 0$  da  $0 < 1 - p \leq 1$ .

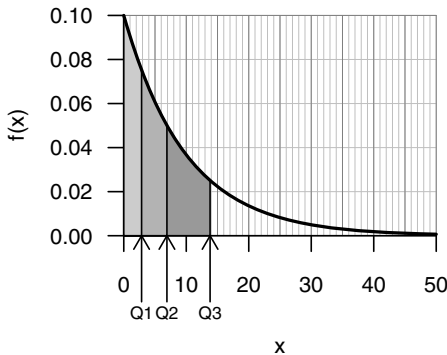
c) Mit b) folgt  $t_{0.25} = t_{1/4} = -\log(3/4)/\lambda = (\log 4 - \log 3)/\lambda$

$t_{0.5} = t_{1/2} = -\log(1/2)/\lambda = (\log 2 - \log 1)/\lambda = \log 2/\lambda$

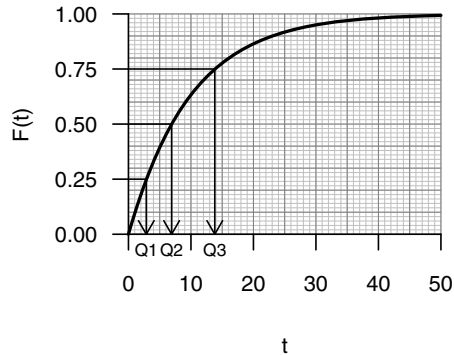
$t_{0.75} = t_{3/4} = -\log(1/4)/\lambda = (\log 4 - \log 1)/\lambda = \log 4/\lambda$ .

d) Mit c) folgt  $t_{1/4} = (\log 4 - \log 3)/0.1 = 10 \cdot (\log 4 - \log 3) \approx 2.877$ ;  $t_{1/2} = 10 \cdot \log 2 \approx 6.931$ ;  $t_{3/4} = 10 \cdot \log 4 \approx 13.863$ .

Dichtefunktion



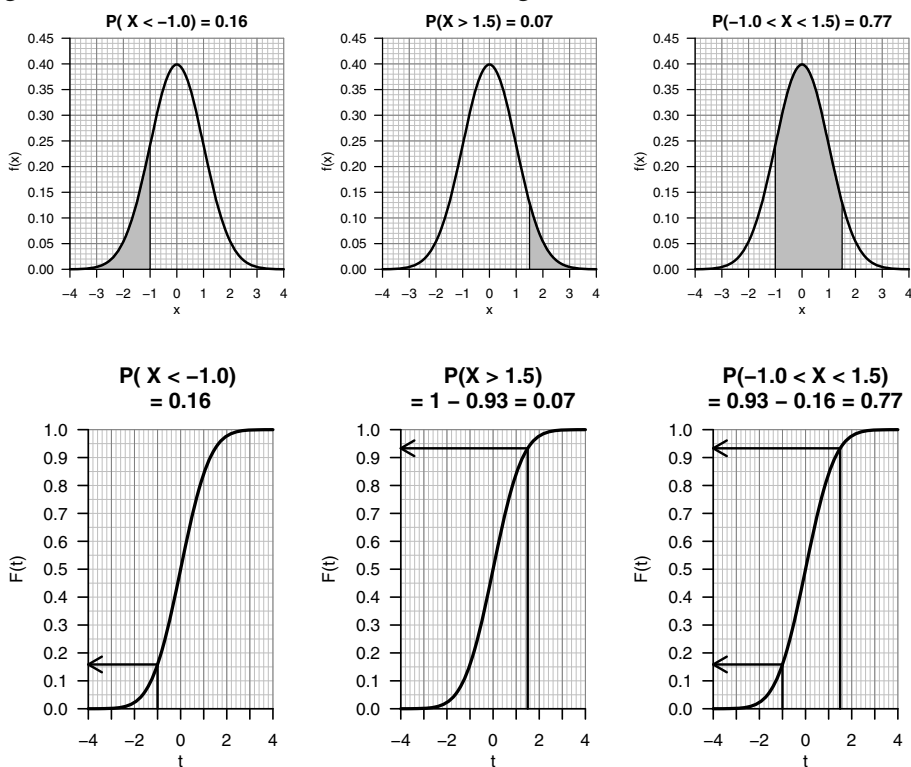
Verteilungsfunktion



e) Nach c) gilt  $t_{0,5} = \log 2 / \lambda$ . Dies gilt genau dann, wenn  $\lambda = \log 2 / t_{0,5}$ . Aus den Grafiken liest man ab:  $t_{0,5} = 10$  bzw. 25, d.h.  $\lambda = \log 2 / 10 \approx 0.069$  bzw.  $\lambda = \log 2 / 25 \approx 0.028$ .

[ 6.5 ] a) W b) F c) W d) F e) W f) W g) F

[ 6.6 ] a) Wir geben erst ein Beispiel für die grafische Lösung für Problemstellungen i)-iii) und danach die numerischen Lösungen.



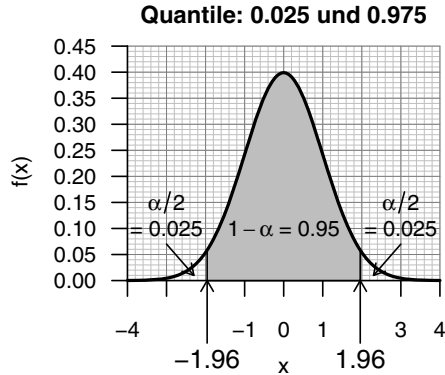
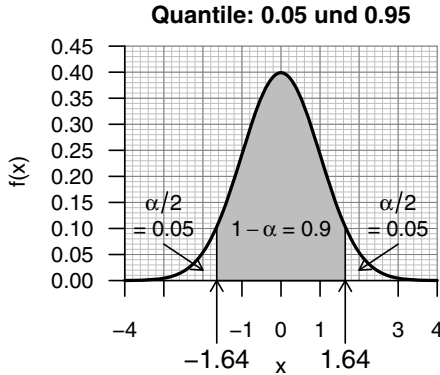
i) 0.02; 0.07; 0.16; 0.31; 0.50; 0.69; 0.84; 0.93; 0.98; 0.99; 1.00

ii) 0.98; 0.84; 0.50; 0.16; 0.02    iii) 0.82; 0.82; 0.98; 0.98; 0.95; 0.68

b) -1.64; -1.28; -1.04, -0.84    c) 1.64; 1.28; 1.04, 0.84

d)  $(-1.96, 1.96)$ ;  $(-1.64, 1.64)$ ;  $(-1.28, 1.28)$ ;

$P(t_1 \leq X \leq t_2) = 1 - \alpha = 0.95; 0.90; 0.80$



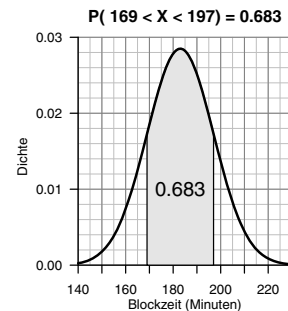
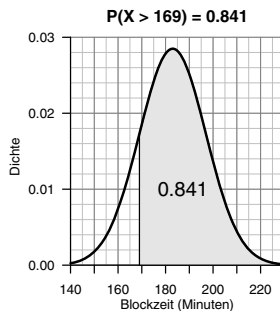
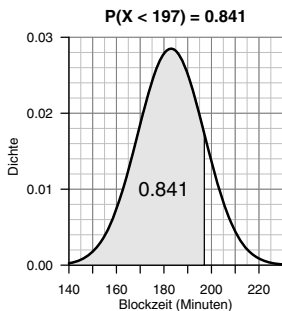
e) **R-Programm:**

```
alpha<-c(0.05,0.1,0.2)
qnorm(alpha/2);qnorm(1-alpha/2)
[1] -1.959964 -1.644854 -1.281552
[1] 1.959964 1.644854 1.281552
```

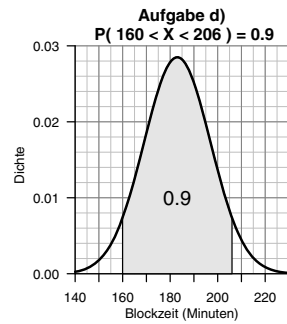
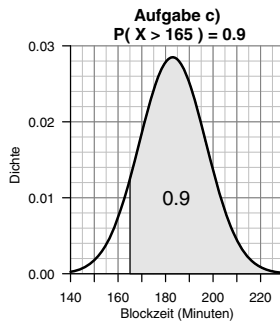
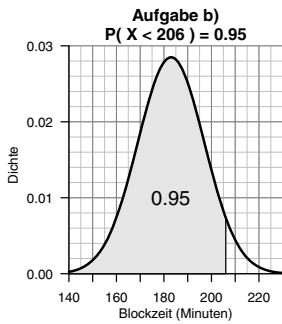
```
# alpha Werte
# t1 und t2 bestimmen
(t1-Werte)
(t2-Werte)
```

[ 6.7 ]

- a) i)  $P(X \leq 197) = \Phi\left(\frac{197-183}{14}\right) = \Phi(1) = 0.841$  ;  
 $P(X \leq 211) = \Phi\left(\frac{211-183}{14}\right) = \Phi(2) = 0.977$   
 ii)  $P(X > 169) = 1 - \Phi\left(\frac{169-183}{14}\right) = 1 - \Phi(-1) = \Phi(1) = 0.841$ ;  
 $P(X > 155) = 1 - \Phi\left(\frac{155-183}{14}\right) = 1 - \Phi(-2) = \Phi(2) = 0.977$   
 iii)  $P(169 < X \leq 197) = P(X \leq 197) - P(X \leq 169) = \Phi(1) - \Phi(-1)$   
 $= 0.841 - 0.159 = 0.682$   
 iv)  $P(155 < X \leq 211) = P(X \leq 211) - P(X \leq 155) = \Phi(2) - \Phi(-2)$   
 $= 0.977 - 0.023 = 0.954$



- b)  $P(X \leq t) = 0.95 \Leftrightarrow \Phi\left(\frac{t-183}{14}\right) = 0.95 \Leftrightarrow \frac{t-183}{14} = 1.64 \Leftrightarrow t = 205.96$ ,  
 wobei Tabelle A.2 benutzt wurde.



c)  $P(X > t) = 0.90 \Leftrightarrow P(X \leq t) = 0.1 \Leftrightarrow \Phi\left(\frac{t-183}{14}\right) = 0.1 \Leftrightarrow \frac{k-183}{14} = -1.28 \Leftrightarrow k = 165.08$ , wobei wieder Tabelle A.2 benutzt wurde.

d)  $P(X < t_1) = 0.05 \Leftrightarrow \Phi\left(\frac{t_1-183}{14}\right) = 0.05 \Leftrightarrow \frac{t_1-183}{14} = -1.64 \Leftrightarrow t_1 = 160.04$ ;  $P(X > t_2) = 0.05 \Leftrightarrow \Phi\left(\frac{t_2-183}{14}\right) = 0.95 \Leftrightarrow t_2 = 205.96$ , siehe b)

e)  $P(t_1 \leq X \leq t_2) = 0.9$

[ 6.8 ] a) `qnorm(c(0.005, 0.025))` ergibt  $z_{0.005} = -2.575829$  und  $z_{0.025} = -1.959964$ .

b) Warngrenzen:

$$WG_U = 10 - 1.959964 \cdot 0.3/3 \approx 9.80; WG_O = 10 + 1.959964 \cdot 0.3/3 \approx 10.20$$

Eingriffsgrenzen:

$$EG_U = 10 - 2.575829 \cdot 0.3/3 \approx 9.74; EG_O = 10 + 2.575829 \cdot 0.3/3 \approx 10.26$$

[ 6.9 ] Die Ausgabe ist mit **R** erzeugt und gibt die Verteilungsfunktion der Standardnormalverteilung  $\Phi(t)$  für  $t = -2, -1.5, -1, -0.5, \dots, 1.5, 2$  an.

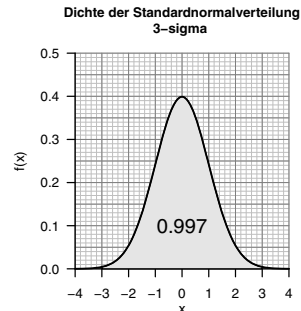
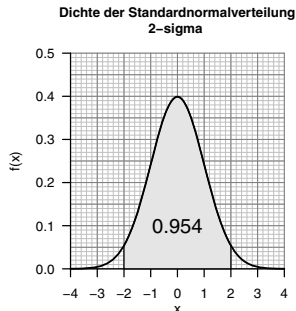
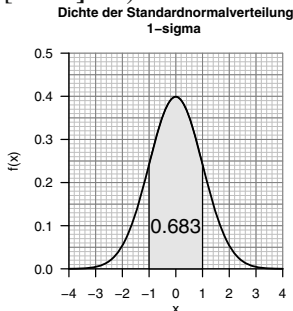
$$P(X \leq 20) = F(20) = \Phi\left(\frac{20-15}{5}\right) = \Phi(1) = 0.8413; \quad P(12.5 < X \leq 22.5) =$$

$$F(22.5) - F(12.5) = \Phi\left(\frac{22.5-15}{5}\right) - \Phi\left(\frac{12.5-15}{5}\right) = \Phi(1.5) - \Phi(-0.5) =$$

$$0.9332 - 0.3085 = 0.6247; \quad P(X > 10) = 1 - F(10) = 1 - \Phi\left(\frac{10-15}{5}\right) = 1 -$$

$$\Phi(-1) = 1 - 0.1587 = 0.8413.$$

[ 6.10 ] a)



b) i) 0.6827; 0.9545; 0.9973, da  $\Phi(q)$  mit  $\text{pnorm}(q)$  berechnet wird, d.h.  $\Phi(1) - \Phi(-1)$  wird durch  $\text{pnorm}(1) - \text{pnorm}(-1)$  berechnet.

ii) Mit Flächen sind im folgenden Flächen unterhalb der Dichtefunktion der  $N(0, 1)$ -Verteilung gemeint.  $\Phi(-3)$  ist die Fläche links von  $-3$ . Diese ist wegen der Symmetrie um 0 genau so groß wie die Fläche rechts von 3, d.h.  $2 \cdot \Phi(-3) = 1 - (\Phi(3) - \Phi(-3))$ , da  $\Phi(3) - \Phi(-3)$  die Fläche zwischen  $-3$  und 3 ist und die Gesamtfläche unterhalb der Dichte 1 ist. Daraus folgt  $\Phi(-3) = (1 - 0.9973)/2 = 0.00135$ . Entsprechend folgt  $\Phi(-2) = (1 - 0.9545)/2 = 0.02275$  und  $\Phi(-1) = (1 - 0.6827)/2 = 0.15865$ . Aus Symmetriegründen ist  $\Phi(0) = 0.5$ . Ferner ist  $\Phi(1) + (1 - \Phi(1)) = 1$ , wobei  $1 - \Phi(1)$  die Fläche rechts von 1 ist, die genau so groß ist wie die Fläche links von  $-1$  und diese ist  $\Phi(-1)$ , d.h.  $\Phi(1) + \Phi(-1) = 1 \iff \Phi(1) = 1 - \Phi(-1) = 1 - 0.15865 = 0.84135$ . Entsprechend ist  $\Phi(2) = 1 - \Phi(-2) = 0.97725$  und  $\Phi(3) = 1 - \Phi(-3) = 0.99865$ .

c) Mit den Ergebnissen aus b) folgt

i)  $P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 0.6827$ ;  $P(0 < Z < 1) = \Phi(1) - \Phi(0) = 0.84135 - 0.5 = 0.34135$ ;  $P(Z > 1) = 1 - \Phi(1) = \Phi(-1) = 0.15865$ ;

$P(Z > -1) = 1 - \Phi(-1) = \Phi(1) = 0.84135$ ;  $P(Z \leq -1) = \Phi(-1) = 0.15865$ ;

$P(|Z| > 1) = \Phi(-1) + (1 - \Phi(1)) = 2 \cdot \Phi(-1) = 0.3173$

ii)  $P(-2 < Z < 2) = \Phi(2) - \Phi(-2) = 0.9545$ ;  $P(0 < Z < 2) = \Phi(2) - \Phi(0) = 0.47725$ ;  $P(Z > 2) = 1 - \Phi(2) = 0.02275$ ;  $P(Z > -2) = 1 - \Phi(-2) = \Phi(2) = 0.97725$ ;  $P(Z \leq -2) = \Phi(-2) = 0.02275$ ;  $P(|Z| > 2) = \Phi(-2) + (1 - \Phi(2)) = 2 \cdot \Phi(-2) = 0.0455$

iii)  $P(-3 < Z < 3) = \Phi(3) - \Phi(-3) = 0.9973$ ;  $P(0 < Z < 3) = \Phi(3) - \Phi(0) = 0.49865$ ;  $P(Z > 3) = 1 - \Phi(3) = 0.00135$ ;  $P(Z > -3) = 1 - \Phi(-3) = \Phi(3) = 0.99865$ ;  $P(Z \leq -3) = \Phi(-3) = 0.00135$ ;  $P(|Z| > 3) = \Phi(-3) + (1 - \Phi(3)) = 2 \cdot \Phi(-3) = 0.0027$

iv)  $P(1 < Z < 2) = \Phi(2) - \Phi(1) = 0.97725 - 0.84135 = 0.1359$

$P(1 < |Z| < 3) = P(-3 < Z < -1) + P(1 < Z < 3) = 2 \cdot P(1 < Z < 3) = 2 \cdot (\Phi(3) - \Phi(1)) = 0.3146$  (dabei wurde die Symmetrie der Dichtefunktion benutzt).

$P(-1 < Z < 2) = \Phi(2) - \Phi(-1) = 0.97725 - 0.15865 = 0.8186$ ;

$P(-3 < Z < -2) = \Phi(-2) - \Phi(-3) = 0.02275 - 0.00135 = 0.0214$ .

d) Wir verwenden stets: Wenn  $X \sim N(\mu, \sigma^2)$ , dann  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ .

i)  $\mu = 20; \sigma = 3 \Rightarrow P(17 < X < 23) = P\left(\frac{17-20}{3} < \frac{X-20}{3} < \frac{23-20}{3}\right) =$

$P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 0.6827$ . Entsprechend ist  $P(14 < X < 26) = P(-2 < Z < 2) = \Phi(2) - \Phi(-2) = 0.9545$ ;  $P(20 < X < 29) = P(0 < Z < 3) = \Phi(3) - \Phi(0) = 0.49865$ ;  $P(X > 17) = P(Z > -1) = 1 - \Phi(-1) = 0.84135$

ii)  $\mu = 10; \sigma = 2 \Rightarrow P(8 \leq X \leq 12) = P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = 0.6827$ ;  $P(6 \leq X \leq 14) = P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = 0.9545$

$P(X \leq 8) = P(Z \leq -1) = \Phi(-1) = 0.15865$ ;  $P(X > 10) = 1 - \Phi(0) = 0.5$

iii)  $\mu = 5; \sigma = 5 \Rightarrow P(-5 \leq X \leq 5) = P(-2 \leq Z \leq 0) = \Phi(0) - \Phi(-2) = 0.47725$ ;  $P(-5 \leq X \leq 10) = P(-2 \leq Z \leq 1) = \Phi(1) - \Phi(-2) = 0.8186$

$P(-5 \leq X \leq 15) = P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = 0.9545$



[ 6.11 ] a)  $X \sim U(a, b) \Rightarrow \mu = \frac{b-a}{2}$  und  $\sigma = \frac{b-a}{\sqrt{12}} = \frac{b-a}{2\sqrt{3}}$ . Das Intervall  $(\mu - \sigma, \mu + \sigma)$  hat die Länge  $2\sigma = \frac{b-a}{\sqrt{3}}$  und damit die Wahrscheinlichkeit  $\frac{1}{b-a} \cdot \frac{b-a}{\sqrt{3}} = \frac{1}{\sqrt{3}} \approx 0.577$ . b) Das Intervall  $(\mu - 2\sigma, \mu + 2\sigma)$  hat die Länge  $4\sigma = \frac{2}{\sqrt{3}}(b-a) > b-a$ . Da es den gleichen Mittelpunkt wie  $(a, b)$  hat, überragt es  $(a, b)$ , so dass die gesuchte Wahrscheinlichkeit 1 ist.

[ 6.12 ] a) F b) W c) W d) W e) F f) W g) W h) F

[ 6.13 ] Die  $b(150, 0.4)$  Verteilung wird approximiert durch die Normalverteilung mit  $\mu = 150 \cdot 0.4 = 60$  und  $\sigma^2 = 150 \cdot 0.4 \cdot 0.6 = 36$ .

a) Ohne Stetigkeitskorrektur:

$P(X < a) = P(X \leq a-1) = F_b(a-1) \approx F_N(a-1) = \Phi\left(\frac{a-1-60}{6}\right)$ , wobei  $(a-61)/6 \approx -2.17, -1.17, 0.83, 1.83$ . Mit Tabelle A.1 folgt  $P(X < a) \approx 0.015, 0.121, 0.797, 0.966$ .

```
a<-c(48, 54, 66, 72)           # Werte von a
round(pbinom(a-1, 150, 0.4), 4) # Binomial, exakt
0.0176 0.1391 0.8206 0.9716
round(pnorm((a-1-60)/6), 4)    # Normalapproximation
round(pnorm(a-1, 60, 6), 4)    # andere Möglichkeit
0.0151 0.1217 0.7977 0.9666
```

Mit Stetigkeitskorrektur:

$P(X < a) = F_b(a-1) \approx F_N(a-1+0.5) = F_N(a-0.5) = \Phi\left(\frac{a-0.5-60}{6}\right)$ , wobei  $(a-60.5)/6 \approx -2.08, -1.08, 0.92, 1.92$ . Mit Tabelle A.1 folgt  $P(X < a) \approx 0.019, 0.140, 0.821, 0.973$ .

```
round(pnorm(a-0.5, 60, 6), 4)  # Appr. mit Korrektur
round(pnorm((a-0.5-60)/6), 4) # standardisiert
0.0186 0.1393 0.8203 0.9724
```

b) Ohne Stetigkeitskorrektur:

$P(X \leq a) = F_b(a) \approx F_N(a) = \Phi\left(\frac{a-60}{6}\right)$ , wobei  $(a-60)/6 = -2, -1, 1, 2$ . Mit Tabelle A.1 folgt  $P(X \leq a) \approx 0.023, 0.159, 0.841, 0.977$ .

```
round(pbinom(a, 150, 0.4), 4)  # Binomial, exakt
0.0265 0.1799 0.8605 0.9807
round(pnorm(a, 60, 6), 4)     # Normalapproximation
```

```
round(pnorm((a-60)/6),4)      # andere Möglichkeit
0.0228 0.1587 0.8413 0.9772
```

#### Mit Stetigkeitskorrektur:

$P(X \leq a) = F_b(a) \approx F_N(a + 0.5) = \Phi\left(\frac{a + 0.5 - 60}{6}\right)$ , wobei  $(a - 59.5)/6 \approx -1.92, -0.92, 1.08, 2.08$ . Mit Tabelle A.1 folgt  $P(X \leq a) \approx 0.027, 0.179, 0.860, 0.981$ .

```
round(pnorm(a+0.5, 60, 6),4)    # Appr. mit Korrektur
round(pnorm((a+0.5-60)/6),4)    # standardisiert
0.0276 0.1797 0.8607 0.9814
```

#### c) Ohne Stetigkeitskorrektur:

$P(X > a) = 1 - P(X \leq a) = 1 - F_b(a) \approx 1 - F_N(a) = 1 - \Phi\left(\frac{a - 60}{6}\right)$ , wobei  $(a - 60)/6 = -2, -1, 1, 2$ . Mit Tabelle A.1 folgt  $P(X > a) \approx 0.977, 0.841, 0.159, 0.023$ .

```
round(1-pbinom(a, 150, 0.4),4)  # Binomial, exakt
0.9735 0.8201 0.1395 0.0193
round(1-pnorm(a, 60, 6),4)      # Normalapproximation
round(1-pnorm((a-60)/6),4)      # standardisiert
0.9772 0.8413 0.1587 0.0228
```

#### Mit Stetigkeitskorrektur:

$P(X > a) = 1 - P(X \leq a) = 1 - F_b(a) \approx 1 - F_N(a + 0.5) = 1 - \Phi\left(\frac{a + 0.5 - 60}{6}\right)$ , wobei  $(a - 59.5)/6 \approx -1.92, -0.92, 1.08, 2.08$ . Mit Tabelle A.1 folgt  $P(X > a) \approx 0.973, 0.821, 0.140, 0.019$ .

```
round(1-pnorm(a+0.5, 60, 6),4)  # Appr. mit Korrektur
round(1-pnorm((a+0.5-60)/6),4)  # standardisiert
0.9724 0.8203 0.1393 0.0186
```

#### d) Ohne Stetigkeitskorrektur:

$P(\bar{X} \geq a) = 1 - P(\bar{X} < a) = 1 - P(X \leq a - 1) = 1 - F_b(a - 1) \approx 1 - F_N(a - 1) = 1 - \Phi\left(\frac{a - 1 - 60}{6}\right)$ , wobei  $(a - 61)/6 \approx -2.17, -1.17, 0.83, 1.83$ . Mit Tabelle A.1 folgt  $P(\bar{X} \geq a) \approx 0.985, 0.879, 0.203, 0.034$ .

```
round(1-pbinom(a-1, 150, 0.4),4) # Binomial exakt
0.9824 0.87609 0.1794 0.0284
round(1-pnorm(a-1, 60, 6),4)     # Normalapproximation
round(1-pnorm((a-1-60)/6),4)     # standardisiert
0.9849 0.8783 0.2023 0.0334
```

#### Mit Stetigkeitskorrektur:

$P(\bar{X} \geq a) = 1 - F_b(a - 1) \approx 1 - F_N(a + 0.5 - 1) = 1 - \Phi\left(\frac{a - 0.5 - 60}{6}\right)$ , wobei  $(a - 60.5)/6 \approx -2.08, -1.08, 0.92, 1.92$ . Mit Tabelle A.1 folgt  $P(\bar{X} \geq a) \approx 0.981, 0.860, 0.179, 0.027$ .

```
round(1-pnorm(a-0.5,60,6),4)      # Appr. mit Korrektur
round(1-pnorm((a-0.5-60)/6),4)    # standardisiert
0.9814 0.8607 0.1797 0.0276
```

e) Ohne Stetigkeitskorrektur:

$$P(48 \leq X \leq 72) = P(47 < X \leq 72) = F_b(72) - F_b(47) \approx F_N(72) - F_N(47) = \Phi\left(\frac{72-60}{6}\right) - \Phi\left(\frac{47-60}{6}\right) \approx \Phi(2) - \Phi(-2.17) = 0.977 - 0.015 = 0.962 \text{ (mit Tabelle A.1).}$$

```
round(pbinom(72,150,0.4)-pbinom(47,150,0.4),4) # exakt
0.9632
round(pnorm(72,60,6)-pnorm(47,60,6),4)      # Appr.
round(pnorm((72-60)/6)-pnorm((47-60)/6),4) # stand.
0.9621
```

Mit Stetigkeitskorrektur:

$$P(48 \leq X \leq 72) = F_b(72) - F_b(47) \approx F_N(72+0.5) - F_N(47+0.5) = \Phi\left(\frac{72.5-60}{6}\right) - \Phi\left(\frac{47.5-60}{6}\right) \approx \Phi(2.08) - \Phi(-2.08) = 0.981 - 0.019 = 0.962 \text{ (Tabelle A.1)}$$

```
round(pnorm(72.5,60,6)-pnorm(47.5,60,6),4) # mit Korr.
round(pnorm((72.5-60)/6)-pnorm((47.5-60)/6),4) # stand.
0.9628
```

f) Ohne Stetigkeitskorrektur:

$$P(48 < X \leq 66) = F_b(66) - F_b(48) \approx F_N(66) - F_N(48) = \Phi\left(\frac{66-60}{6}\right) - \Phi\left(\frac{48-60}{6}\right) = \Phi(1) - \Phi(-2) = 0.841 - 0.023 = 0.818$$

```
round(pbinom(66,150,0.4)-pbinom(48,150,0.4),4) # exakt
0.834
round(pnorm(66,60,6)-pnorm(48,60,6),4)      # Normalappr.
round(pnorm((66-60)/6)-pnorm((48-60)/6),4) # stand.
0.8186
```

Mit Stetigkeitskorrektur:

$$P(48 < X \leq 66) = F_b(66) - F_b(48) \approx F_N(66.5) - F_N(48.5) = \Phi\left(\frac{66.5-60}{6}\right) - \Phi\left(\frac{48.5-60}{6}\right) \approx \Phi(1.08) - \Phi(-1.92) = 0.860 - 0.027 = 0.833$$

```
round(pnorm(66.5,60,6)-pnorm(48.5,60,6),4)
round(pnorm((66.5-60)/6)-pnorm((48.5-60)/6),4)
0.833
```

g) Ohne Stetigkeitskorrektur:

$$P(54 \leq X < 72) = P(53 < X \leq 71) = F_b(71) - F_b(53) \approx F_N(71) - F_N(53) = \Phi\left(\frac{71-60}{6}\right) - \Phi\left(\frac{53-60}{6}\right) \approx \Phi(1.83) - \Phi(-1.17) = 0.845$$

```
round(pbinom(71,150,0.4)-pbinom(53,150,0.4),4) # exakt
0.8325
round(pnorm(71,60,6)-pnorm(53,60,6),4) # Normalappr.
round(pnorm((71-60)/6)-pnorm((53-60)/6),4)
0.845
```

Mit Stetigkeitskorrektur:

$$P(54 \leq X < 72) = F_b(71) - F_b(53) \approx F_N(71.5) - F_N(53.5) = \Phi\left(\frac{71.5-60}{6}\right) - \Phi\left(\frac{53.5-60}{6}\right) = \Phi(1.92) - \Phi(-1.08) = 0.973 - 0.140 = 0.833 \text{ (Tab. A.1)}$$

```
round(pnorm(71.5,60,6)-pnorm(53.5,60,6),4) # mit Korrr.
round(pnorm((71.5-60)/6)-pnorm((53.5-60)/6),4)
0.833
```

h) Ohne Stetigkeitskorrektur:

$$P(54 < X < 66) = P(54 < X \leq 65) = F_b(65) - F_b(54) \approx F_N(65) - F_N(54) = \Phi\left(\frac{65-60}{6}\right) - \Phi\left(\frac{54-60}{6}\right) \approx \Phi(0.83) - \Phi(-1) = 0.797 - 0.159 = 0.638$$

```
round(pbinom(65,150,0.4)-pbinom(54,150,0.4),4) # exakt
0.6407
round(pnorm(65,60,6)-pnorm(54,60,6),4) # Appr.
round(pnorm((65-60)/6)-pnorm((54-60)/6),4) # stand.
0.639
```

Mit Stetigkeitskorrektur:

$$P(54 < X < 66) = F_b(65) - F_b(54) \approx F_N(65.5) - F_N(54.5) = \Phi\left(\frac{65.5-60}{6}\right) - \Phi\left(\frac{54.5-60}{6}\right) \approx \Phi(0.92) - \Phi(-0.92) = 0.821 - 0.179 = 0.642$$

```
round(pnorm(65.5,60,6)-pnorm(54.5,60,6),4) # mit Korrr.
round(pnorm((65.5-60)/6)-pnorm((54.5-60)/6),4)
0.6407
```

**[ 6.14 ]** Die approximierende Normalverteilung hat die Parameter  $\mu = 100 \cdot 0.5 = 50$  und  $\sigma^2 = 100 \cdot 0.5 \cdot 0.5 = 25$ , d.h.  $\sigma = 5$ . Sei  $F$  die Verteilungsfunktion der Binomialverteilung und  $\Phi$  die Verteilungsfunktion der  $N(0,1)$ -Verteilung. Dann gilt mit Tabelle A.1:

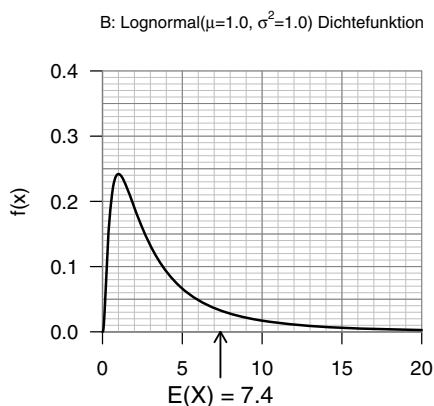
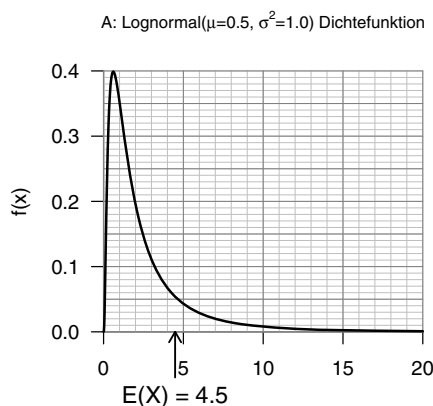
$$\begin{aligned} \text{a) } P(X \leq 40) &= F(40) \approx \Phi\left(\frac{40.5-50}{5}\right) = \Phi\left(-\frac{9.5}{5}\right) = \Phi(-1.9) = 0.029 \\ P(X < 45) &= P(X \leq 44) = F(44) \approx \Phi\left(\frac{44.5-50}{5}\right) = \Phi(-1.1) = 0.136 \end{aligned}$$

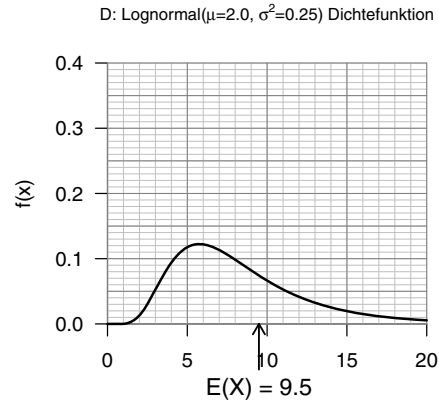
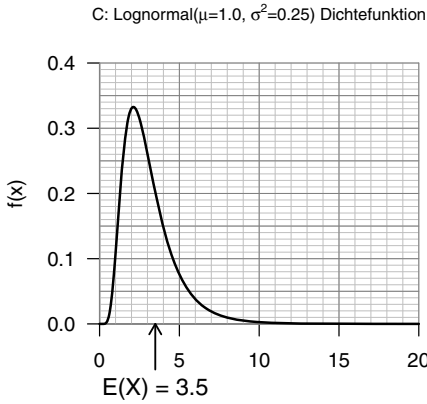
- b)  $P(X > 52) = 1 - P(X \leq 52) = 1 - F(52) \approx 1 - \Phi\left(\frac{52.5 - 50}{5}\right) = 1 - \Phi(0.5) = 1 - 0.691 = 0.309$ ;  $P(X \geq 55) = P(X > 54) = 1 - P(X \leq 54) = 1 - F(54) \approx 1 - \Phi\left(\frac{54.5 - 50}{5}\right) = 1 - \Phi(0.9) = 1 - 0.816 = 0.184$
- c)  $P(45 < X \leq 52) = F(52) - F(45) \approx \Phi\left(\frac{52.5 - 50}{5}\right) - \Phi\left(\frac{45.5 - 50}{5}\right) = \Phi(0.5) - \Phi(-0.9) = 0.691 - 0.184 = 0.507$ ;  $P(47 < X < 53) = P(47 < X \leq 52) = F(52) - F(47) \approx \Phi\left(\frac{52.5 - 50}{5}\right) - \Phi\left(\frac{47.5 - 50}{5}\right) = \Phi(0.5) - \Phi(-0.5) = 0.691 - 0.309 = 0.382$
- d)  $P(49 \leq X \leq 58) = P(48 < X \leq 58) = F(58) - F(48) \approx \Phi\left(\frac{58.5 - 50}{5}\right) - \Phi\left(\frac{48.5 - 50}{5}\right) = \Phi(1.7) - \Phi(-0.3) = 0.955 - 0.382 = 0.573$ ;  $P(43 \leq X < 54) = P(42 < X \leq 53) = F(53) - F(42) \approx \Phi\left(\frac{53.5 - 50}{5}\right) - \Phi\left(\frac{42.5 - 50}{5}\right) = \Phi(0.7) - \Phi(-1.5) = 0.758 - 0.067 = 0.691$

[ 6.15 ] a) F b) W c) F d) F e) W f) W

[ 6.16 ] a)

$\mu$	$\sigma^2$	$E(X)$	$\text{Var}(X)$	$\alpha_3$	$\alpha_4$
0.5	1	4.482	12.696	6.185	113.936
1	1	7.389	34.513	6.185	113.936
1	0.25	3.490	2.695	1.750	8.898
2	0.25	9.488	19.912	1.750	8.898





b) Beachtet man, dass  $\log(X) \sim N(\mu, \sigma^2)$  und somit  $(\log(X) - \mu)/\sigma \sim N(0, 1)$ , so folgt:  $F_X(t) = P(X \leq t) = P(\log(X) \leq \log(t)) = P\left(\frac{\log(X) - \mu}{\sigma} \leq \frac{\log(t) - \mu}{\sigma}\right) = \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$ .

c) In allen Teilaufgaben verwenden wir die in b) bewiesene Formel.

i)  $P(X \leq 5) = \Phi\left(\frac{\log(5) - \mu}{\sigma}\right)$ , wobei  $\frac{\log(5) - \mu}{\sigma} = 1.11(A); 0.61(B); 1.22(C); -0.78(D)$ . Mit Tabelle A.1 folgt  $P(X \leq 5) = 0.867(A); 0.729(B); 0.889(C); 0.218(D)$ . (Mit **R** ergeben sich Abweichungen in der dritten Stelle nach dem Dezimalpunkt.) ii)  $P(X > 10) = 1 - P(X \leq 10) = 1 - \Phi\left(\frac{\log(10) - \mu}{\sigma}\right)$ , wobei  $\frac{\log(10) - \mu}{\sigma} = 1.80(A); 1.30(B); 2.61(C); 0.61(D)$ . Mit Tabelle A.1 folgt  $P(X > 10) = 0.036(A); 0.097(B); 0.005(C); 0.271(D)$ .

iii)  $P(X \leq e^{0.5}) = \Phi\left(\frac{\log(e^{0.5}) - 0.5}{1}\right) = \Phi(0.5 - 0.5) = \Phi(0) = 0.5$ .

iv) Für B, C gilt  $P(X > e) = 1 - P(X \leq e) = 1 - \Phi\left(\frac{\log(e) - 1}{\sigma}\right) = 1 - \Phi(0) = 0.5$ .

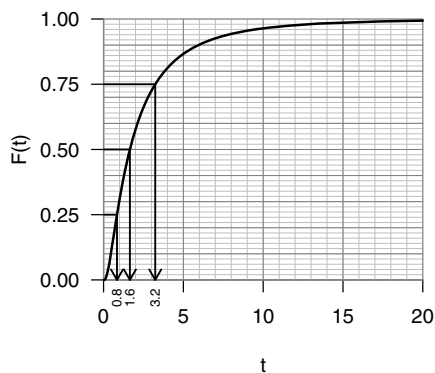
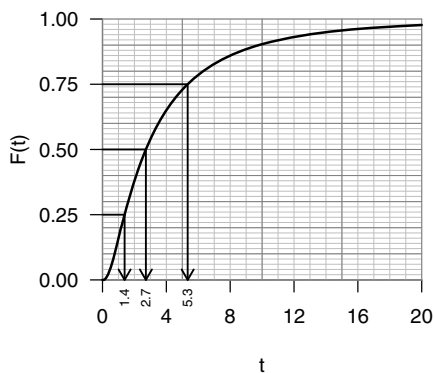
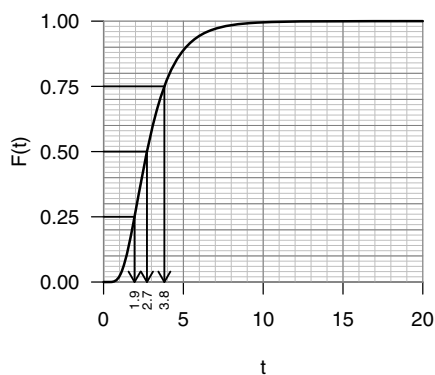
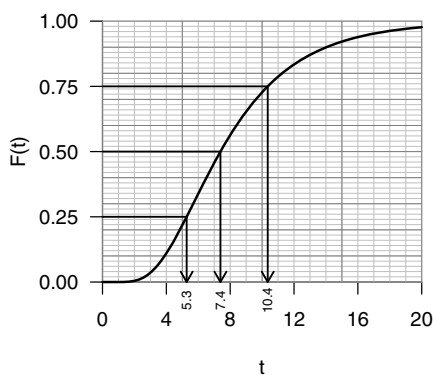
v)  $P(X \leq e^2) = \Phi\left(\frac{\log(e^2) - 2}{0.5}\right) = \Phi(0) = 0.5$ . vi)  $P(X > e^{-0.5}) = 1 - P(X \leq e^{-0.5}) = 1 - \Phi\left(\frac{\log(e^{-0.5}) - 0.5}{1}\right) = 1 - \Phi(-1) = 1 - 0.159 = 0.841$ . vii)

$P(X \leq 1) = \Phi\left(\frac{\log(1) - 1}{1}\right) = \Phi(-1) = 0.159$ .

viii)  $P(X \leq e^{0.75}) = \Phi\left(\frac{\log(e^{0.75}) - 0.5}{0.5}\right) = \Phi(-0.5) = 0.309$ .

ix)  $P(X > e^{1.75}) = 1 - P(X \leq e^{1.75}) = 1 - \Phi\left(\frac{\log(e^{1.75}) - 2}{0.5}\right) = 1 - \Phi(-0.5) = 0.691$ .

d) ii) Ungefähr: A : 0.8; 1.6; 3.2    B : 1.4; 2.7; 5.3    C : 1.9; 2.7; 3.8    D : 5.3; 7.4; 10.4

A: Lognormal( $\mu=0.5$ ,  $\sigma^2=1.0$ ) VerteilungsfunktionB: Lognormal( $\mu=1.0$ ,  $\sigma^2=1.0$ ) VerteilungsfunktionA: Lognormal( $\mu=1.0$ ,  $\sigma^2=0.25$ ) VerteilungsfunktionB: Lognormal( $\mu=2.0$ ,  $\sigma^2=0.25$ ) Verteilungsfunktion

[ 6.17 ] a) W b) W c) F d) W e) F

## 2.7 Modellangepassung und Parameterschätzung - Lösungen

[ 7.1 ] a) W b) W c) F d) W e) F f) F

[ 7.2 ] a) F b) W c) W d) W e) F f) F g) W h) W

[ 7.3 ] a)  $E(X) = \int_1^{\infty} x \alpha x^{-\alpha-1} dx = \int_1^{\infty} \alpha x^{-\alpha} dx = \left[ \frac{\alpha}{-\alpha+1} x^{-\alpha+1} \right]_1^{\infty} = 0 - \frac{\alpha}{-\alpha+1} = \frac{\alpha}{\alpha-1}$

b) Methode der Momente:  $E(X) = \frac{\alpha}{\alpha-1} \Rightarrow \alpha = E(X)(\alpha-1) \Rightarrow \alpha(1-E(X)) = -E(X) \Rightarrow \alpha = \frac{E(X)}{E(X)-1} \Rightarrow \hat{\alpha} = \frac{\bar{x}}{\bar{x}-1}$

c) ML-Methode:  $L(\alpha) = \alpha^n \prod_{i=1}^n x_i^{-\alpha-1}$

$$\log L(\alpha) = n \log(\alpha) - (\alpha + 1) \sum_{i=1}^n \log(x_i) \Rightarrow \frac{d \log L(\alpha)}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^n \log(x_i) = 0 \Rightarrow$$

$$\frac{n}{\hat{\alpha}} = \sum_{i=1}^n \log(x_i) \Rightarrow \hat{\alpha} = \frac{n}{\sum_{i=1}^n \log(x_i)}; \text{ d) Meth. Momente: } \hat{\alpha} = \bar{x}/(\bar{x} - 1) \approx 2.083$$

ML-Meth.:  $\hat{\alpha} = n / \left( \sum_{i=1}^n \log x_i \right) \approx 2.159$ ; e) Der Fehler des Schätzers ( $\hat{\alpha} - \alpha$ ) beträgt für den Schätzer nach der Methode der Momente  $2.083 - 2 = 0.083$  und für den Schätzer nach der ML-Methode  $2.159 - 2 = 0.159$ .

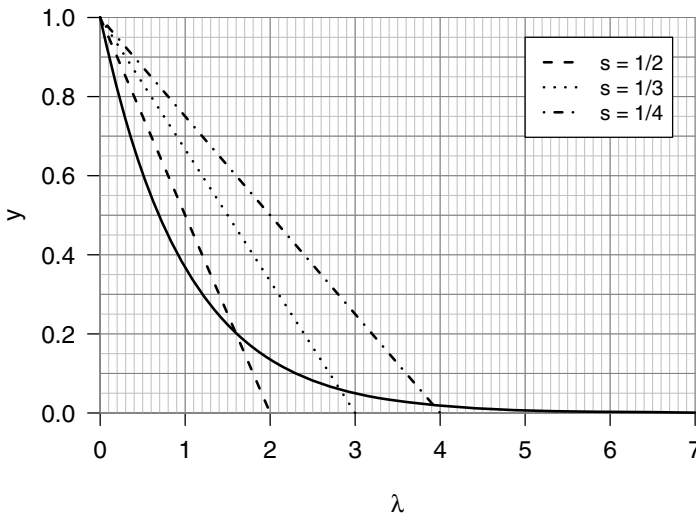
[7.4] a)  $\lambda = 2/E(X)$  und somit  $\hat{\lambda} = 2/\bar{x}$ . b) Die Likelihoodfunktion ist:  $L(\lambda) = \prod_{i=1}^n f(x_i) = \lambda^{2n} \cdot \left( \prod_{i=1}^n x_i \right) \cdot e^{-\lambda \sum_{i=1}^n x_i} \Rightarrow \log(L(\lambda)) = 2n \log(\lambda) + \sum_{i=1}^n \log(x_i) - \lambda \sum_{i=1}^n x_i$

$$\Rightarrow \frac{d \log(L(\lambda))}{d\lambda} = \frac{2n}{\lambda} - \sum_{i=1}^n x_i = 0 \iff \frac{\hat{\lambda}}{2n} = 1 / \sum_{i=1}^n x_i \iff \hat{\lambda} = 2n / \sum_{i=1}^n x_i = 2/\bar{x}.$$

c)  $\hat{\lambda} \approx 0.380$

[7.5] a)  $\mu = \frac{\lambda}{1 - e^{-\lambda}} \iff e^{-\lambda} = \frac{\mu - \lambda}{\mu} = 1 - \frac{\lambda}{\mu}$ , d.h. für den Schätzer nach der Methode der Momente gilt  $e^{-\hat{\lambda}} = 1 - \frac{\hat{\lambda}}{\bar{x}}$ . b) Die Gerade  $y = f(\lambda) = 1 - \frac{\lambda}{\bar{x}}$  schneidet nach a) den Graphen der Funktion  $y = e^{-\lambda}$  an der Stelle  $\hat{\lambda}$ . c) Zu zeichnen sind die Geraden mit dem y-Achsenabschnitt 1 und den Steigungen  $1/2$ ;  $1/3$  bzw.  $1/4$ . Man liest als  $\lambda$ -Koordinate die Schnittpunkte der Geraden mit dem Graphen der Funktion  $e^{-\lambda}$ , d.h. die Punkte  $\hat{\lambda} = 1.6$ ;  $2.8$  bzw.  $3.9$  ab.

Die Funktionen  $y = e^{-\lambda}$  und  $y = 1 - s\lambda$





d) i)

$x$	$N_i$	$N_i/N$	$x \cdot N_i/N$	$\hat{P}_X(x)$
1	274	0.377	0.377	0.402
2	222	0.306	0.612	0.324
3	171	0.236	0.707	0.174
4	59	0.081	0.325	0.070
$\Sigma$	726	1.000	2.021	0.970

Der Mittelwert ist  $\bar{x} = 2.021$ . ii) Die Steigung der Geraden ist  $\frac{1}{\bar{x}} = \frac{1}{2.021} = 0.4948046 \approx 0.495$ . iii)  $\hat{\lambda} \approx 1.6$ .

iv) Siehe auch **R**-Programm.  $\hat{P}(X > 4) = 1 - \sum_{j=1}^4 P_X(j) \approx 0.02967319 \approx 0.0297$ .

$x$	1	2	3	4	$> 4$
$\hat{P}_X(x)$	0.4048	0.3238	0.1727	0.0691	0.0297

[ 7.6 ] a) W b) F c) W d) W e) F f) F g) F

[ 7.7 ] a)  $\hat{\pi} = 2/20 = 0.1$ ; b)  $\widehat{SE}(\hat{\pi}) = \sqrt{0.1 \cdot (1 - 0.1)/20} = 0.06708204 \approx 0.067$ ; c) Da  $\hat{\pi}$  ein erwartungstreuer Schätzer von  $\pi$  ist, ist der geschätzte mittlere quadratische Fehler gleich der geschätzten Varianz von  $\hat{\pi}$ , d.h.  $0.1 \cdot 0.9/20 = 0.0045$ .

[ 7.8 ]  $\hat{\pi} = 4/39 = 0.1025641 \approx 0.1026$ ;  $\widehat{Var}(\hat{\pi}) = \hat{\pi} \cdot (1 - \hat{\pi})/n = (4/39) \cdot (35/39)/39 = 0.002360121 \approx 0.0024$ ;  $\widehat{SE}(\hat{\pi}) = \sqrt{\hat{\pi} \cdot (1 - \hat{\pi})/n} = 0.04858107 \approx 0.0486$ ;  $\widehat{MQF}(\hat{\pi}) = \widehat{Var}(\hat{\pi}) = 0.0024$  da der Bias Null ist.

[ 7.9 ] Im ersten Fall ist die Zustimmung 73%, d.h.  $\hat{\pi} = 0.73$ . Dann ist  $\widehat{SE}(\hat{\pi}) = \sqrt{\hat{\pi}(1 - \hat{\pi})/n} = \sqrt{0.73(1 - 0.73)/10} = 0.1403923 \approx 0.1404 = 14.04\%$ , falls  $n = 10$  ist. Für  $n = 100$  erhalten wir 4.44%, für  $n = 1000$  noch 1.4% und für  $n = 10000$  schließlich 0.44%. Analog erhält man für  $\hat{\pi} = 59\%$ : 15.55%, 4.92%, 1.56%, 0.49%; für  $\hat{\pi} = 56\%$ : 15.70%, 4.96%, 1.57%, 0.50%; für  $\hat{\pi} = 52\%$ : 15.80%, 5.00%, 1.58%, 0.50% und für  $\hat{\pi} = 49\%$ : 15.81%, 5%, 1.58%, 0.50% .

[ 7.10 ] a) Beachten Sie, dass  $n = 775$  und  $\widehat{SE}(\hat{\pi}) = \sqrt{\hat{\pi}(1 - \hat{\pi})/n}$  ist.

Eigenschaft	Zustimmung in %	$\hat{\pi}$	$\widehat{SE}(\hat{\pi})$
heimatlich	91	0.91	0.0103
humorvoll	85	0.85	0.0128
gemütlich	79	0.79	0.0146
typisch norddeutsch	75	0.75	0.0156
gehört zum Ohnsorg-Theater	67	0.67	0.0169
stirbt aus	62	0.62	0.0174
sprechen auch junge Leute wieder	55	0.55	0.0179
kenne ich aus Talk op Platt	41	0.41	0.0177
nicht zeitgemäß	30	0.30	0.0165
grob	27	0.27	0.0159
nur für alte Menschen	21	0.21	0.0146

b) Für  $\hat{\pi} = 0.55$ ;

c)  $f'(\hat{\pi}) = 1 - 2\hat{\pi} = 0 \iff \hat{\pi} = 1/2 = 0.5$ . Da  $f''(\hat{\pi}) = -2 < 0$ , handelt es sich in der Tat um ein Maximum. d) Mit der Lösung in c) benutze man  $\hat{\pi} = 0.5$ . Damit erhält man für den maximalen geschätzten Standardfehler  $\sqrt{0.5 \cdot (1 - 0.5)/n} = 0.5/\sqrt{n} = 0.5/\sqrt{775} = 0.01796053 \approx 0.018$ .

e) Nach d) ist der maximale Wert  $0.5/\sqrt{n}$ . Wir rechnen mit **R**:

```
> n<-c(30,50,100,500,1000); round(0.5/sqrt(n),3)
[1] 0.091 0.071 0.050 0.022 0.016
```

f) Da  $\widehat{\text{var}}(\hat{\pi}) = \hat{\pi} \cdot (1 - \hat{\pi})/n = (\widehat{SE}(\hat{\pi}))^2$  nach d) maximal ist für  $\hat{\pi} = 0.5$ , muss gelten  $(0.5 \cdot 0.5)/n \leq (\text{SE}_{\max})^2 \iff n \geq \frac{0.25}{(\text{SE}_{\max})^2}$ .

g) Mit **R** berechnen wir

```
> semax<-c(0.05,0.04,0.03,0.02,0.01) # Maximalwert
> round(0.25/semax^2,1)
100.0 156.2 277.8 625.0 2500.0 # n nach Formel gerundet
> ceiling(0.25/semax^2)
100 157 278 625 2500 # nach oben auf ganze Zahl gerundet
```

[ 7.11 ] a)-c)

Quelle	Z1	$\hat{\pi}$	$\widehat{SE}(\hat{\pi})$	Z1 + Z2	$\hat{\pi}$	$\widehat{SE}(\hat{\pi})$	Z1 + Z2 + Z3	$\hat{\pi}$	$\widehat{SE}(\hat{\pi})$
Radio	0	0.000	0.000	0	0.000	0.000	0	0.000	0.000
Zeit.	42	0.532	0.056	68	0.613	0.046	79	0.637	0.043
Blick	6	0.076	0.030	6	0.054	0.021	7	0.056	0.021
Eule	4	0.051	0.025	4	0.036	0.018	4	0.032	0.016
Hallo	1	0.013	0.013	1	0.009	0.009	1	0.008	0.008
Plakat.	8	0.101	0.034	10	0.090	0.027	11	0.089	0.026
Hand.	6	0.076	0.030	7	0.063	0.023	7	0.056	0.021
Bek.	4	0.051	0.025	6	0.054	0.021	6	0.048	0.019
Inter.	2	0.025	0.018	2	0.018	0.013	2	0.016	0.011
Sonst.	6	0.076	0.030	7	0.063	0.023	7	0.056	0.021
Gesamt	79			111			124		

Sowohl der geschätzte Anteil  $\hat{\pi}$  als auch der geschätzte Standardfehler von  $\hat{\pi}$  sind in den meisten Kategorien unwesentlich kleiner geworden, da  $n$  gestiegen ist und keine weiteren Beobachtungen in diesen Kategorien gemacht worden sind. Lediglich im Bereich Zeitung stieg  $\hat{\pi}$  an. Aufgrund der relativen kleinen Stichprobengröße von Z3 hätte man zumindest diese Umfrage weglassen können.

d)  $SE_{max} = \sqrt{\frac{0.5 \cdot (1-0.5)}{n}} = \frac{0.5}{\sqrt{n}} = 0.0563$  (für  $n = 79$ ),  $0.0475$  (für  $n = 111$ ),  $0.0449$  (für  $n = 124$ ); i) von 79 auf 111: Verringerung: 0.0088, entspricht  $\approx 15.64\%$ ; von 111 auf 124: Verringerung: 0.0026, entspricht  $\approx 5.39\%$ . ii) nein, der maximale Standardfehler ist bereits für  $n = 79$  kleiner als 0.1.

[ 7.12 ]  $\hat{\lambda} = \bar{x} = (1 \cdot 9 + 2 \cdot 5 + 3 \cdot 3)/22 = 28/22 \approx 1.273$ ;  $\text{Var}(\hat{\lambda}) = \lambda/n$  und  $\widehat{\text{Var}}(\hat{\lambda}) = \hat{\lambda}/n = (28/22)/22 = 0.05785124 \approx 0.058$ ;  $\widehat{SE}(\hat{\lambda}) = \sqrt{\widehat{\text{Var}}(\hat{\lambda})} = \sqrt{(28/22)/22} = 0.2405228 \approx 0.241$ . Da  $\hat{\lambda}$  erwartungstreu, ist der geschätzte mittlere quadratische Fehler gleich der geschätzten Varianz, also  $\approx 0.058$ .

[ 7.13 ] Alle Rechenwege sind analog zur letzten Aufgabe:  $a\hat{\lambda} = \bar{x} = 15/30 = 0.5$ ;  $\widehat{\text{Var}}(\hat{\lambda}) = \widehat{\text{MQF}}[\hat{\lambda} = \hat{\lambda}/n = 0.5/30 \approx 0.017$  und  $\widehat{SE}(\hat{\lambda}) = \sqrt{0.5/30} \approx 0.129$ .

[ 7.14 ] a) F b) W c) W d) W e) W f) F

[ 7.15 ] a) W b) W c) F d) W e) F f) W g) W

[ 7.16 ] a) W b) F c) W d) W e) W

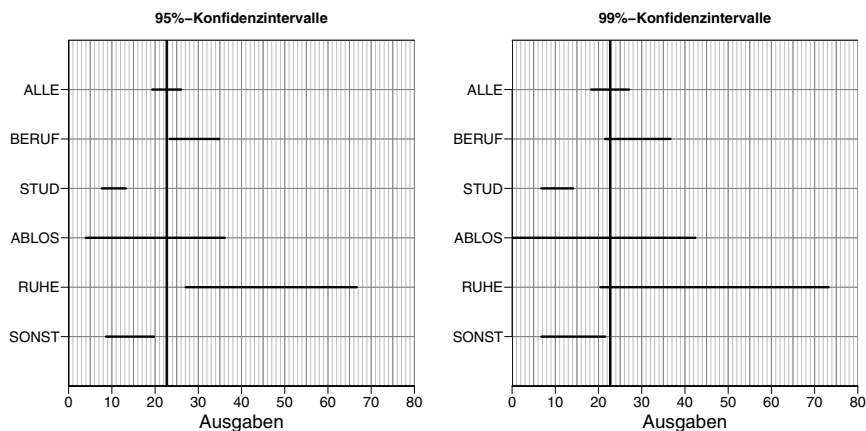
[ 7.17 ] a) Die Grenzen des Konfidenzintervalls sind  $\bar{x} \mp t_{n-1, \alpha/2} S_*/\sqrt{n}$ .

Gruppe	ALLE	BERUF	STUD	ABLOS	RUHE	SONST
$\bar{x}$	22.69	29.09	10.46	20.07	46.86	14.18
$S_*$	43.78	51.05	20.66	28.04	68.95	22.78
$n$	659	306	213	14	49	67
$t_{n-1, 0.025}$	1.964	1.968	1.971	2.160	2.011	1.997
$C_-$	19.34	23.35	7.67	3.88	27.06	8.62
$C_+$	26.04	34.83	13.25	36.26	66.66	19.74
$t_{n-1, 0.005}$	2.583	2.592	2.599	3.012	2.682	2.652
$C_-$	18.28	21.53	6.78	-2.50	20.44	6.80
$C_+$	27.10	36.65	14.14	42.64	73.28	21.56

Die Quantile der  $t$ -Verteilung wurden mit **R** berechnet:

```
> n<-c(659, 306, 213, 14, 49, 67)           # Stichprobengrößen
> round(qt(0.975, n-1), 3)                   # t-Quantile für 95%
[1] 1.964 1.968 1.971 2.160 2.011 1.997
> round(qt(0.995, n-1), 3)                   # t-Quantile für 99%
[1] 2.583 2.592 2.599 3.012 2.682 2.652
```

b)



c) beim 95% Konfidenzintervall: ALLE, ABLOS; beim 99% Konfidenzintervall: ALLE, BERUF, ABLOS, RUHE

d) Das Histogramm hat Ähnlichkeit mit der Dichte einer Normalverteilung, d.h. die Mittelwerte sind annähernd normalverteilt.

[ 7.18 ] a) F b) F c) W d) F e) F f) W

[ 7.19 ]  $\hat{\pi} = 4/16 = 0.25$ ;  $1 - \alpha = 0.95 \Rightarrow \alpha/2 = 0.025 \Rightarrow z_{\alpha/2} = 1.96$  (Tabelle A.2)  $\Rightarrow C^- = 0.25 - 1.96 \cdot \sqrt{0.25 \cdot 0.75/16} \approx 0.0378$  und  $C^+ = 0.25 + 1.96 \cdot \sqrt{0.25 \cdot 0.75/16} \approx 0.4622$ , d.h.  $(C^-, C^+) = (0.0378, 0.4622)$ .

[ 7.20 ] a) Wir rechnen mit **R**:

```
pidach<-c(0.91,0.85,0.79,0.75,0.67,0.62,0.55,0.41,0.30,0.27,0.21)
se<-round(sqrt(pidach*(1-pidach)/775),4) # schätzt SE
qua <- c(0.950,0.975) # qua=1-alpha/2
round(qnorm(qua,4) # gibt 1.6449 und 1.9600
Cmin90<-round(pidach-1.6449*se,3) # Untere Grenze 90%
Cplus90<-round(pidach+1.6449*se,3) # Obere Grenze 90%
Cmin95<-round(pidach-1.9600*se,3) # Untere Grenze 95%
Cplus95<-round(pidach+1.9600*se,3) # Obere Grenze 95%
```

Eigenschaft	$\hat{\pi}$	$\widehat{SE}(\hat{\pi})$	90%	95%
heimatlich	0.91	0.0103	(0.893,0.927)	(0.890,0.930)
humorvoll	0.85	0.0128	(0.829,0.871)	(0.825,0.875)
gemütlich	0.79	0.0146	(0.766,0.814)	(0.761,0.819)
typisch norddeutsch	0.75	0.0156	(0.724,0.776)	(0.719,0.781)
gehört zum Ohnsorg-Theater	0.67	0.0169	(0.642,0.698)	(0.637,0.703)
stirbt aus	0.62	0.0174	(0.591,0.649)	(0.586,0.654)
sprechen auch junge Leute wieder	0.55	0.0179	(0.521,0.579)	(0.515,0.585)
kenne ich aus Talk op Platt	0.41	0.0177	(0.381,0.439)	(0.375,0.445)
nicht zeitgemäß	0.30	0.0165	(0.273,0.327)	(0.268,0.332)
grob	0.27	0.0159	(0.244,0.296)	(0.239,0.301)
nur für alte Menschen	0.21	0.0146	(0.186,0.234)	(0.181,0.239)

b) 0.55; c) Für 0.5, da dies den größten Standardfehler und damit das breiteste Intervall gibt. d) Die Breite der Intervalle ist  $B = 2 \cdot z_{\alpha/2} \sqrt{\hat{\pi}(1-\hat{\pi})/n}$ , d.h. die maximale Breite ist  $B_{max} = 2 \cdot z_{\alpha/2} \sqrt{0.5^2/n} = z_{\alpha/2}/\sqrt{n} = 1.6449/\sqrt{775} \approx 0.059$  für  $1 - \alpha = 0.9$  und  $B_{max} = 1.96/\sqrt{775} \approx 0.070$ .

e) Wir rechnen mit **R**:

```
> n<-c(30,50,100,500,1000) # Stichprobengrößen
> round(qnorm(0.95)/sqrt(n),3) # Max. Breite 90%
[1] 0.300 0.233 0.164 0.074 0.052
> round(qnorm(0.975)/sqrt(n),3) # Max. Breite 95%
[1] 0.358 0.277 0.196 0.088 0.062
```

f) Nach d) gilt  $B_{\max} = z_{\alpha/2}/\sqrt{n} \iff \sqrt{n} = z_{\alpha/2}/B_{\max} \iff n = (z_{\alpha/2}/B_{\max})^2$ . Wir rechnen mit **R**:

```
> Bmax<-c(0.1,0.05,0.02)          # Maximale Breiten
> ceiling((qnorm(0.95)/Bmax)^2)    # Nötiges n bei 90%
[1] 271 1083 6764
> ceiling((qnorm(0.975)/Bmax)^2)  # Nötiges n bei 95%
[1] 385 1537 9604
```

[ 7.21 ] a) Das ist gleichbedeutend mit dem Konfidenzintervall, dh. mit einer bestimmten Wahrscheinlichkeit werden die jeweiligen wahren Anteilswerte von einem Konfidenzintervall (hier der Länge 6 *Prozentpunkte*) um den jeweiligen geschätzten Anteilswert überdeckt. b) Da das Maximum für  $\pi = 0.5$  angenommen wird, ist der maximale Standardfehler  $\sqrt{0.5^2/1002} = 0.0157956 \approx 0.0158$ . c)  $0.03/0.0157956 = 1.899263$ . d) 1.899 entspricht in etwa dem 0.97 Quantil der Normalverteilung. Zu berechnen mit `pnorm(1.899)`. Das ergibt in etwa ein 94% Konfidenzintervall.

[ 7.22 ] a) Die allgemeinen Formeln lauten  $C^- = \hat{\pi} - z_{\alpha/2} \cdot \sqrt{\hat{\pi} \cdot (1 - \hat{\pi})/n}$ ;  $C^+ = \hat{\pi} + z_{\alpha/2} \cdot \sqrt{\hat{\pi} \cdot (1 - \hat{\pi})/n}$ . Wir bestimmen  $z_{\alpha/2}$  aus Tabelle A.2 zu 1.64; 1.96; 2.58.

$$C_{0.90}^- = 5/27 - 1.64 \sqrt{(5/27 \cdot 22/27)/27} = 0.06258404 \approx 0.063$$

$$C_{0.90}^+ = 5/27 + 1.64 \sqrt{(5/27 \cdot 22/27)/27} = 0.3077863 \approx 0.308$$

$$C_{0.95}^- = 5/27 - 1.96 \sqrt{(5/27 \cdot 22/27)/27} = 0.03866186 \approx 0.039$$

$$C_{0.95}^+ = 5/27 + 1.96 \sqrt{(5/27 \cdot 22/27)/27} = 0.3317085 \approx 0.332$$

$$C_{0.99}^- = 5/27 - 2.58 \sqrt{(5/27 \cdot 22/27)/27} = -0.007687356, \text{ d.h. } C^- = 0$$

$$C_{0.99}^+ = 5/27 + 2.58 \sqrt{(5/27 \cdot 22/27)/27} = 0.3780577 \approx 0.378$$

b) Ein Konfidenzintervall zum Niveau  $1 - \alpha$  enthält den wahren Parameter mit einer Wahrscheinlichkeit von  $1 - \alpha$ . Wird die Wahrscheinlichkeit  $1 - \alpha$  größer, muss also auch das Konfidenzintervall länger werden.

c) Bei Erhöhung des Stichprobenumfangs unter sonst gleichen Bedingungen wird das Konfidenzintervall für den Anteilswert  $\pi$  bei einem vorgegebenen Signifikanzniveau kleiner aufgrund der gewachsenen Information. Man bedenke, dass wir Wahrscheinlichkeiten berechnen, also Werte unter Unsicherheit; mehr Information vermindert die Unsicherheit. In der Formel drückt sich das durch den kleiner werden Standardfehler aus. Beispielhaft für  $\alpha = 0.01$ :

$$C_{0.99}^- = 5/27 - 2.58 \sqrt{(5/27 \cdot 22/27)/270} = 0.1241935 \approx 0.124$$

$$C_{0.99}^+ = 5/27 + 2.58 \sqrt{(5/27 \cdot 22/27)/270} = 0.2461768 \approx 0.246$$

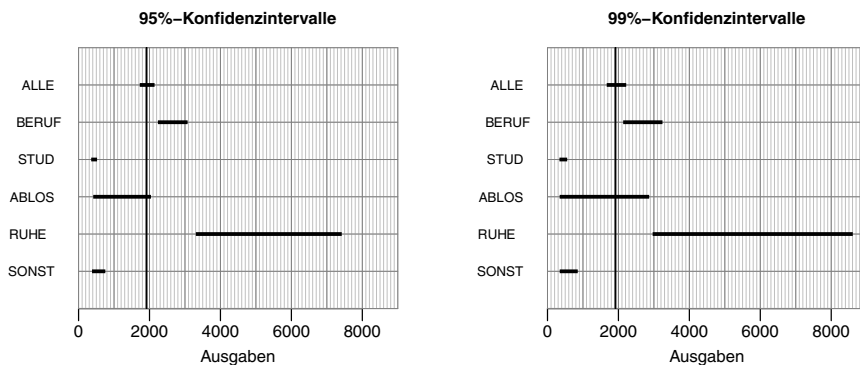
Damit haben wir  $(C_{0.99}^-, C_{0.99}^+) = (0.124; 0.246)$  im Vergleich zu  $(0; 0.378)$ .

[ 7.23 ] a) und b), zuerst für 95, dann für 99% Konfidenzintervall:

Gruppe	ALLE	BERUF	STUD	ABLOS	RUHE	SONST
$n$	659	306	213	14	49	67
$S_*^2$	1916.79	2605.65	426.98	786.23	4754.21	519.09
$nS_*^2 = (n-1)S_*^2$	1261248	794723	90520	10221	228202	34260
$\chi_{n-1, \alpha/2}^2$	730.98	355.27	254.22	24.74	69.02	90.35
$C_-$	1725.43	2236.94	356.07	413.21	3306.19	379.20
$\chi_{n-1, 1-\alpha/2}^2$	588.81	258.51	173.57	5.01	30.76	45.43
$C_+$	2142.02	3074.20	521.52	2040.63	7420.12	754.10

Gruppe	ALLE	BERUF	STUD	ABLOS	RUHE	SONST
$nS_*^2 = (n-1)S_*^2$	1261248	794723	90520	10221	228202	34260
$\chi_{n-1, \alpha/2}^2$	755.20	372.37	268.79	29.82	76.97	99.33
$C_-$	1670.10	2134.24	336.77	342.76	2964.87	344.91
$\chi_{n-1, 1-\alpha/2}^2$	568.32	245.14	162.72	3.57	26.51	40.16
$C_+$	2219.27	3241.92	556.30	2867.01	8607.96	853.12

c)



d) In beiden Fällen nur ALLE und ABLOS.

[ 7.24 ] **R** berechnet mit dem Befehl `var` den erwartungstreuen Schätzer  $S_*^2$ . Dies kann mit den Befehlen `?var` oder `help(var)` überprüft werden. Die hier angegebenen Ergebnisse wurden mit Taschenrechner berechnet. Das ebenso bereitgestellte **R**-Programm zu dieser Aufgabe hingegen rundet erst ganz am Ende, wodurch es zu numerischen Abweichungen kommen kann.

$$a) C^- = \bar{x} - t_{n-1; \alpha/2} \cdot \frac{S_*}{\sqrt{n}} \quad C^+ = \bar{x} + t_{n-1; \alpha/2} \cdot \frac{S_*}{\sqrt{n}}$$

$$C_{0.90}^- = 5444.444 - 1.86 \cdot \frac{\sqrt{211893}}{\sqrt{9}} = 5159.047 \quad C_{0.90}^+ = 5444.444 + 1.86 \cdot \frac{\sqrt{211893}}{\sqrt{9}} = 5729.841 \Rightarrow (C_{0.90}^-, C_{0.90}^+) = (5159.047, 5729.841)$$

$$C_{0.95}^- = 5444.444 - 2.31 \cdot \frac{\sqrt{211893}}{\sqrt{9}} = 5089.999 \quad C_{0.95}^+ = 5444.444 + 2.31 \cdot \frac{\sqrt{211893}}{\sqrt{9}} = 5798.889 \Rightarrow (C_{0.95}^-, C_{0.95}^+) = (5089.999, 5798.889)$$

$$\text{b) } n = 25: \quad C_{0.95}^- = 5444.444 - 2.06 \cdot \frac{\sqrt{211893}}{\sqrt{25}} = 5254.793 \quad C_{0.95}^+ = 5444.444 + 2.06 \cdot \frac{\sqrt{211893}}{\sqrt{25}} = 5634.095 \Rightarrow (C_{0.95}^-, C_{0.95}^+) = (5254.793, 5634.095)$$

$$n = 60: \quad C_{0.95}^- = 5444.444 - 2.00 \cdot \frac{\sqrt{211893}}{\sqrt{60}} = 5325.590 \quad C_{0.95}^+ = 5444.444 + 2.00 \cdot \frac{\sqrt{211893}}{\sqrt{60}} = 5563.298 \Rightarrow (C_{0.95}^-, C_{0.95}^+) = (5325.590, 5563.298): \quad \text{i) Länge des Konfidenzintervalls} = C^+ - C^- = 708.890 \text{ für } n = 9; \quad 379.302 \text{ für } n = 25; \quad 237.708 \text{ für } n = 60. \quad \text{ii) Die Länge des Konfidenzintervalls wird mit zunehmendem Stichprobenumfang kleiner.}$$

$$\text{c) } C^- = \frac{n \cdot S^2}{\chi_{n-1; \alpha/2}^2} \quad C^+ = \frac{n \cdot S^2}{\chi_{n-1; 1-\alpha/2}^2} \quad nS^2 = (n-1)S_*^2 = 8 \cdot 211893 = 1695144$$

$$C_{0.90}^- = 1695144/15.51 = 109293.617 \quad C_{0.90}^+ = 1695144/2.73 = 620931.868 \Rightarrow (C_{0.90}^-, C_{0.90}^+) = (109293.617, 620931.868)$$

$$C_{0.98}^- = 1695144/20.09 = 84377.501 \quad C_{0.98}^+ = 1695144/1.65 = 1027360.000 \Rightarrow (C_{0.98}^-, C_{0.98}^+) = (84377.501, 1027360.000). \text{ Das Konfidenzintervall zum Konfidenzniveau } 1 - \alpha = 0.98 \text{ ist breiter.}$$

$$\text{d) } nS^2 = (n-1)S_*^2 = 24 \cdot 211893 = 5085432$$

$$C_{0.90}^- = 5085432/36.42 = 139632.949 \quad C_{0.90}^+ = 5085432/13.85 = 367179.206 \Rightarrow (C_{0.90}^-, C_{0.90}^+) = (139632.949, 367179.206). \text{ Die Breite des Konfidenzintervalls ist für } n = 25 \text{ geringer als bei } n = 9.$$

## 2.8 Hypothesentests - Lösungen

[ 8.1 ] a) W b) F c) W d) F e) F f) W g) W

[ 8.2 ] a) F b) W c) W d) F e) W f) F g) W

[ 8.3 ] a) F b) W c) F d) W e) F f) W g) F

[ 8.4 ] a) F b) W c) W d) F

[ 8.5 ] a) W b) F c) W d) W e) W



[ 8.6 ] a) F b) F c) W d) W e) W f) W g) F h) W

[ 8.7 ] Die abzusichernde Aussage ist jeweils die Alternativhypothese.

a) Nullhypothese und Alternativhypothese:  $H_0: \pi \geq \pi_0 = 0.55$  und  $H_1: \pi < 0.55$ .

Berechnung der Prüfgröße: Mit  $X = 35; n = 104$  und  $\pi_0 = 0.55$  ist

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{35 - 104 \cdot 0.55}{104 \cdot 0.55 \cdot 0.45} \approx -4.376.$$

Bestimmung des Ablehnungsbereiches: Wenn die Alternativhypothese wahr ist, ist die Prüfgröße klein, d.h. der Ablehnungsbereich liegt links, d.h. er geht von  $-\infty$  bis zum kritischen Wert. Die kritischen Werte entnimmt man der Tabelle A.2:  $-1.28; -1.64; -2.33$  für  $\alpha = 0.1; 0.05, 0.01$ . Alternativ kann man den **R**-Befehl `qnorm(c(0.1, 0.05, 0.01))` verwenden, was `[1] -1.281552 -1.644854 -2.326348` ergibt. Damit sind die Ablehnungsbereiche für  $\alpha = 0.1: (-\infty, -1.28]$ , für  $\alpha = 0.05: (-\infty, -1.64]$ , für  $\alpha = 0.01: (-\infty, -2.33]$ . Damit liegt  $Z = -4.376$  für alle drei Werte von  $\alpha$  im Ablehnungsbereich, d.h. die gegebene Aussage lässt sich auf allen drei Niveaus statistisch absichern.

Der P-Wert ist annähernd 0, wie man Tabelle A1 entnehmen kann oder mit dem **R**-Befehl `pnorm(-4.376)` berechnen kann.

b) Nullhypothese und Alternativhypothese:  $H_0: \pi \leq \pi_0 = 0.47$  und  $H_1: \pi > 0.47$ .

Berechnung der Prüfgröße: Mit  $X = 64; n = 106$  und  $\pi_0 = 0.47$  ist

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{64 - 106 \cdot 0.47}{106 \cdot 0.47 \cdot 0.53} \approx 2.760.$$

Bestimmung des Ablehnungsbereiches: Wenn die Alternativhypothese wahr ist, ist die Prüfgröße groß, d.h. der Ablehnungsbereich liegt rechts, d.h. er geht vom kritischen Wert bis  $\infty$ . Die kritischen Werte nach Tabelle A.2 sind:  $1.28; 1.64; 2.33$  für  $\alpha = 0.1; 0.05, 0.01$ . Alternativ kann man den **R**-Befehl

`qnorm(c(0.9, 0.95, 0.99))` verwenden, was

`[1] 1.281552 1.644854 2.326348` ergibt.

Damit sind die Ablehnungsbereiche für  $\alpha = 0.1: [1.28, \infty)$ , für  $\alpha = 0.05: [1.64, \infty)$ , für  $\alpha = 0.01: [2.33, \infty)$ . Damit liegt  $Z = 2.760$  für alle drei Werte von  $\alpha$  im Ablehnungsbereich, d.h. die gegebene Aussage lässt sich auf allen drei Niveaus statistisch absichern.

Der P-Wert ist annähernd 0, wie man Tabelle A1 entnehmen kann oder mit dem **R**-Befehl `1-pnorm(2.760)` berechnen kann.

c) Nullhypothese und Alternativhypothese:  $H_0: \pi = \pi_0 = 0.5$  und  $H_1: \pi \neq 0.5$ .

Berechnung der Prüfgröße: Mit  $X = 68; n = 106$  und  $\pi_0 = 0.5$  ist

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{68 - 106 \cdot 0.5}{106 \cdot 0.5 \cdot 0.5} \approx 2.914.$$

Bestimmung des Ablehnungsbereiches: Die Alternativhypothese ist  $\pi \neq 0.5$ , d.h. die Nullhypothese wird abgelehnt, wenn die Prüfgröße sehr klein oder sehr groß ist. Die kritischen Werten sind  $\pm z_{\alpha/2}$ , welche man Tabelle A2 entnimmt oder mit dem **R**-Befehl `qnorm(c(0.95, 0.975, 0.995))` berechnet, was `1.644854 1.959964 2.575829` ergibt. Die Ablehnungsbereiche sind für  $\alpha = 0.1: (-\infty, -1.64] \cup [1.64, \infty)$ , für  $\alpha = 0.05: (-\infty, -1.96] \cup [1.96, \infty)$ , für  $\alpha = 0.01: (-\infty, -2.58] \cup [2.58, \infty)$ . Damit

liegt  $Z = 2.914$  für alle drei Werte von  $\alpha$  im Ablehnungsbereich, d.h. die gegebene Aussage lässt sich auf allen drei Niveaus statistisch absichern.

Der P-Wert ist annähernd 0, wie man Tabelle A1 entnehmen kann oder mit

$2 * (1 - \text{pnorm}(2.914))$  berechnen kann. Beachten Sie dabei den Faktor 2!

d) Die Prüfgröße ist jetzt  $Z = \frac{52 - 104 \cdot 0.55}{\sqrt{104 \cdot 0.55 \cdot 0.45}} \approx -1.02$ . Der P-Wert ist

$P(Z < -1.02) = 0.154$  (nach Tabelle A1) oder  $\text{pnorm}(-1.02)$

[8.8] a) Test über den Mittelwert:  $H_0: \mu \geq \mu_0 = 77$  und  $H_1: \mu < 77$ .

Prüfgröße:  $T = \frac{\bar{x} - \mu_0}{\sqrt{S_*^2/n}} = \frac{59.57143 - 77}{\sqrt{87.46392/98}} \approx -18.448$ . Unter  $H_0$  gilt  $T \sim t_{97}$  und unter  $H_1$  ist die PG klein, d.h. der Ablehnungsbereich ist auf der linken Seite. Die kritischen Werte<sup>6</sup> erhält man mit  $\text{qt}(c(0.1, 0.05, 0.01), 97)$ , was  $-1.290340$   $-1.660715$   $-2.365407$  ergibt. Damit folgen die Ablehnungsbereiche für  $\alpha = 0.1$ :  $(-\infty; -1.290]$ , für  $\alpha = 0.05$ :  $(-\infty; -1.661]$ , für  $\alpha = 0.01$ :  $(-\infty; -2.365]$ . Der Wert der Prüfgröße  $T = -18.448$  liegt für alle drei  $\alpha$  im Ablehnungsbereich, d.h. die gegebene Aussage lässt sich auf allen drei Niveaus statistisch absichern. Der P-Wert ist annähernd 0, siehe  $\text{pt}(-18.448, 97)$ .

Test über die Varianz:  $H_0: \sigma^2 \geq \sigma_0^2 = 157$  und  $H_1: \sigma^2 < 157$ . Prüfgröße:  $PG = nS^2/\sigma_0^2 = (n-1) \cdot S_*^2/\sigma_0^2 = 97 \cdot 87.46392/157 \approx 54.038$ . Unter  $H_0$  gilt  $PG \sim \chi_{n-1}^2 = \chi_{97}^2$ . Die Ablehnungsbereiche sind linksseitig mit den kritischen Werten  $\approx 79.633; 75.282; 67.562$ , die man mit  $\text{qchisq}(c(0.1, 0.05, 0.01), n-1)$  erhält. Die Ablehnungsbereiche sind für  $\alpha = 0.1$ :  $[0, 79.633]$ , für  $\alpha = 0.05$ :  $[0, 75.282]$ , für  $\alpha = 0.01$ :  $[0, 67.562]$ ; 5%:  $[0, 75.282]$ ; 10%:  $[0, 79.633]$ . Da  $PG = 54.038$  für alle drei  $\alpha$  im Ablehnungsbereich liegt, lässt sich die gegebene Aussage auf allen drei Niveaus statistisch absichern. Der P-Wert ist fast Null, siehe  $\text{pchisq}(54.038, 97)$ , was  $0.0001246892$  ergibt.

b) Test über den Mittelwert:  $H_0: \mu = \mu_0 = 69$  und  $H_1: \mu \neq 69$ .

Prüfgröße:  $T = \frac{\bar{x} - \mu_0}{\sqrt{S_*^2/n}} = \frac{59.57143 - 69}{\sqrt{87.46392/98}} \approx -9.980$ . Unter  $H_0$  gilt  $T \sim t_{n-1} = t_{97}$ .

Der Ablehnungsbereich ist beidseitig. Die kritischen Werte sind gegeben durch  $\pm t_{n-1, \alpha/2}$ , was man z.B. mit  $\text{qt}(c(0.05, 0.025, 0.005), 97)$  berechnen kann. Die gerundete Ausgabe ist  $-1.661$   $-1.985$   $-2.627$  und somit sind die Ablehnungsbereiche für  $\alpha = 0.1$ :  $(-\infty, -1.661] \cup [1.661, \infty)$ ; für  $\alpha = 0.05$ :  $(-\infty, -1.985] \cup [1.985, \infty)$ ; für  $\alpha = 0.01$ :  $(-\infty, -2.627] \cup [2.627, \infty)$ . Die Prüfgröße  $T = -9.980$  liegt für alle drei  $\alpha$  im Ablehnungsbereich, d.h. die gegebene Aussage lässt sich auf allen Niveaus statistisch absichern. Der P-Wert kann mit  $2 * \text{pt}(-9.980, 97)$  berechnet werden, was  $1.480840\text{e-}16$  ergibt, was  $1.480840 \cdot 10^{-16}$  bedeutet, d.h. der P-Wert ist nahezu Null.

Test über die Varianz  $H_0: \sigma^2 = \sigma_0^2 = 195$  gegen  $H_1: \sigma^2 \neq 195$ .

Prüfgröße:  $PG = 97 \cdot 87.46392/195 \approx 43.507$ . Unter  $H_0$  gilt  $PG \sim \chi_{97}^2$ . Die kritischen Werte sind  $\chi_{97; (1-\alpha/2)}^2$  und  $\chi_{97; \alpha/2}^2$ , die wir mit

$\text{round}(\text{qchisq}(c(0.05, 0.95, 0.025, 0.975, 0.005, 0.995), 97), 3)$  berechnen:  
75.282 120.990 71.642 126.141 64.878 136.619

<sup>6</sup> Mit Tabelle A3 lassen sich die kritischen Werte nur ungefähr bestimmen.

Die Ablehnungsbereiche sind für  $\alpha = 0.1$ :  $[0, 75.282] \cup [120.990; \infty)$ ; für  $\alpha = 0.05$ :  $[0, 71.642] \cup [126.141; \infty)$ ; für  $\alpha = 0.01$ :  $[0, 64.878] \cup [136.619, \infty)$ . Da  $PG = 43.507$  in allen drei Ablehnungsbereichen liegt, lässt sich die gegebene Aussage auf allen drei Niveaus statistisch absichern.

Den P-Wert erhält man mit  $2 \cdot \text{pchisq}(43.507, 97)$ , was  $\approx 0$  ergibt.

c) Test über den Mittelwert:  $H_0: \mu \leq \mu_0 = 170$  gegen  $H_1: \mu > 170$ .

Prüfgröße:  $T = \frac{182.6905 - 170}{\sqrt{61.30343/126}} \approx 18.194$ . Unter  $H_0$  gilt  $T \sim t_{n-1} = t_{125}$ .

Die Ablehnungsbereiche sind rechtsseitig mit den durch **R** berechneten kritischen Werten  $\text{round}(\text{qt}(c(0.9, 0.95, 0.99), 125), 2)$ , was 1.29 1.66 2.36 ergibt (vergleiche Tabelle A3). Ablehnungsbereiche: Für  $\alpha = 0.1$ :  $[1.29, \infty)$ ; für  $\alpha = 0.05$ :  $[1.66, \infty)$ ; für  $\alpha = 0.01$ :  $[2.36, \infty)$ . Da  $T = 18.194$  in allen drei Ablehnungsbereichen liegt, lässt sich die gegebene Aussage auf allen drei Niveaus statistisch absichern. Der P-Wert ist  $\approx 0$ , siehe  $1 - \text{pt}(18.194, 125)$ .

Test über die Varianz:  $H_0: \sigma^2 \leq \sigma_0^2 = 50$  gegen  $H_1: \sigma^2 > 50$ . Prüfgröße:  $PG = 125 \cdot 61.30343/50 \approx 153.259$ . Unter  $H_0$  gilt  $PG \sim \chi_{125}^2$ . Der Ablehnungsbereich ist rechts. Mit  $\text{round}(\text{qchisq}(c(0.9, 0.95, 0.99), 125), 2)$  erhalten wir die kritischen Werte 145.64 152.09 164.69. Die Ablehnungsbereiche sind für  $\alpha = 0.1$ :  $[145.64, \infty)$ ; für  $\alpha = 0.05$ :  $[152.09, \infty)$ ; für  $\alpha = 0.01$ :  $[164.69, \infty)$ . Die Prüfgröße  $PG = 153.259$  liegt für  $\alpha = 0.1$  und  $\alpha = 0.05$  im Ablehnungsbereich, d.h. für  $\alpha = 0.1$  und  $\alpha = 0.05$  lässt sich die Nullhypothese widerlegen und damit die Alternative statistisch absichern, jedoch nicht für  $\alpha = 0.01$ . Den P-Wert erhält man mit  $\text{round}(1 - \text{pchisq}(153.259, 125), 4)$ , was 0.0437 ergibt.

[ 8.9 ] a) Hier ist ein zweiseitiger Test mit der Nullhypothese  $H_0: \pi = \pi_0 = 0.5$  gegen  $H_1: \pi \neq 0.5$  zu verwenden. Die Ablehnungsbereiche für  $\alpha = 0.1, 0.05, 0.01$  sind aus der Ausgabe der Verteilungsfunktion der Binomialverteilung abzulesen:

$A = \{0, 1, \dots, 105\} \cup \{132, 133, \dots, 237\}$  für  $\alpha = 0.1$ ;

$A = \{0, 1, \dots, 102\} \cup \{135, 136, \dots, 237\}$  für  $\alpha = 0.05$ ;

$A = \{0, 1, \dots, 98\} \cup \{139, 140, \dots, 237\}$  für  $\alpha = 0.01$ .

Der Wert 74 für die Prüfgröße liegt jeweils im Ablehnungsbereich, d.h. die Nullhypothese wird verworfen. Der P-Wert für 74 ist  $2 \cdot \text{pbinom}(74, 237, 0.5)$ , d.h.  $\approx 0$  nach der **R**-Ausgabe in der Aufgabe.

b)  $H_0: \pi > 0.5$  gegen  $H_1: \pi \leq 0.5$  mit den Ablehnungsbereichen:

$A = \{0, 1, \dots, 108\}$  für  $\alpha = 0.1$ ;  $A = \{0, 1, \dots, 105\}$  für  $\alpha = 0.05$ ;

$A = \{0, 1, \dots, 100\}$  für  $\alpha = 0.01$ . P-Wert:  $\text{pbinom}(74, 237, 0.5)$ , d.h.  $\approx 0$ .

c)  $H_0: \pi < 0.5$  gegen  $H_1: \pi \geq 0.5$  mit den Ablehnungsbereichen

$A = \{128, 129, \dots, 237\}$  für  $\alpha = 0.1$ ;  $A = \{132, 133, \dots, 237\}$  für  $\alpha = 0.05$ ;

$A = \{136, 137, \dots, 237\}$  für  $\alpha = 0.01$ . P-Wert:  $1 - \text{pbinom}(73, 237, 0.5)$ , d.h.  $\approx 1$ .

[ 8.10 ]  $H_0: \pi \leq \pi_0 = 0.23$  gegen  $H_1: \pi > 0.23$ . Prüfgröße:  $Z = \frac{433 - 1044 \cdot 0.23}{\sqrt{1044 \cdot 0.23 \cdot (1 - 0.23)}} \approx 14.185$ . Der P-Wert ist fast Null, siehe Tabelle A1 oder  $1 - \text{pnorm}(14.185)$ .  $H_0$  wird

verworfen. Es ist statistisch abgesichert, dass die Anzahl der Drogenfahrten gestiegen ist.

[ 8.11 ]  $H_0 : \pi \geq \pi_0 = 0.49$  gegen  $H_1 : \pi < 0.49$ . Prüfgröße:  $Z = \frac{92 - 500 \cdot 0.49}{\sqrt{500 \cdot 0.49 \cdot (1 - 0.49)}} \approx -13.687$ . Der P-Wert ist laut Tabelle A1  $\approx 0$ , siehe auch in **R**: `pnorm(-13.687)`.  $H_0$  wird verworfen. Man hat damit statistisch abgesichert, dass der Anteil der Ökostrombezieher in Wiesbaden kleiner als 49% ist.

[ 8.12 ]  $H_0 : \pi = \pi_0 = 0.167$  gegen  $H_1 : \pi \neq 0.167$ . Prüfgröße:  $Z = \frac{0.22 - 0.167}{\sqrt{0.167 \cdot (1 - 0.167) / 500}} \approx 3.177$ . Der P-Wert ist laut Tabelle A1  $\approx 0$ , siehe auch in **R**: `2 * (1 - pnorm(3.177))`.  $H_0$  wird verworfen. Man hat statistisch abgesichert, dass die Anteil der Ja-Antworten in Wiesbaden und Mainz von 16.7% verschieden ist.

[ 8.13 ] a) Anzahl ‘+’ (und Anzahl ‘-’)  $\sim b(12, 0.5)$ .

b) Sei  $\pi$  die Wahrscheinlichkeit, dass „*unterschätzt*“ wird. Dann:  $H_0: \pi \leq \pi_0 = 0.5$  gegen  $H_1: \pi > 0.5$ .

c) Die Ablehnungsbereiche sind rechtsseitig. Mit **R** berechnen wir die Wahrscheinlichkeitsfunktion für  $x \geq 6$ :

```
> round(dbinom(6:12, 12, 0.5), 3)
[1] 0.226 0.193 0.121 0.054 0.016 0.003 0.000
```

Ablehnungsbereich:  $A = \{11, 12\}$  für  $\alpha = 0.01$ ;  $A = \{10, 11, 12\}$  für  $\alpha = 0.05$ ;  $A = \{9, 10, 11, 12\}$  für  $\alpha = 0.10$ . Es wurde 9-mal unterschätzt, d.h. die PG liegt im Ablehnungsbereich für  $\alpha = 0.1$ . Für  $\alpha = 0.05$  und  $0.01$  kann  $H_0$  nicht verworfen werden. Der P-Wert ist  $P(PG \geq 9) = 1 - P(PG \leq 8) \approx 0.073$ , was mit `1 - pbinom(8, 12, 0.5)` berechnet werden kann.

[ 8.14 ] a)  $H_0: \pi \geq \pi_0 = 0.5$  gegen  $H_1: \pi < 0.5$ . Ablehnungsbereiche<sup>7</sup> für  $\alpha = 0.01$ :  $\{0, 1, 2\}$ ; für  $\alpha = 0.05$ :  $\{0, 1, 2, 3\}$ ; für  $\alpha = 0.1$ :  $\{0, 1, 2, 3, 4\}$ .

b)  $H_0: \pi \leq \pi_0 = 0.5$  gegen  $H_1: \pi > 0.5$ . Ablehnungsbereiche:  $\alpha = 0.01$ :  $\{13, 14, 15\}$ ; für  $\alpha = 0.05$ :  $\{12, 13, 14, 15\}$ ; für  $\alpha = 0.1$ :  $\{11, 12, 13, 14, 15\}$ .

c)  $H_0: \pi = \pi_0 = 0.5$  gegen  $H_1: \pi \neq 0.5$ .  $A = \{0, 1, 2, 3\} \cup \{12, 13, 14, 15\}$ .

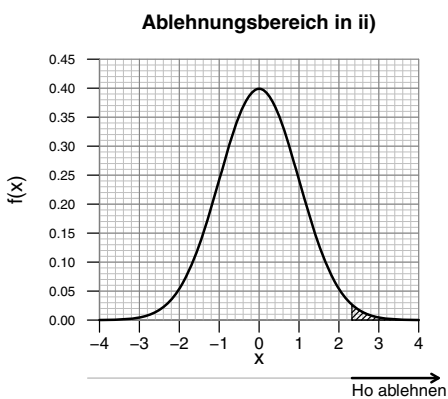
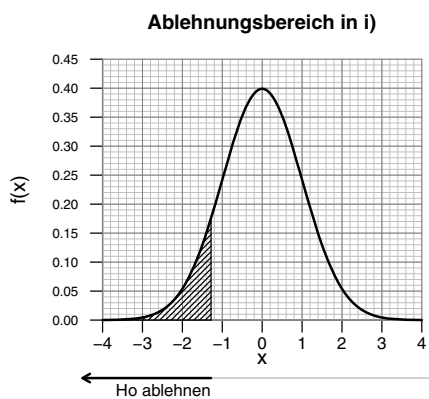
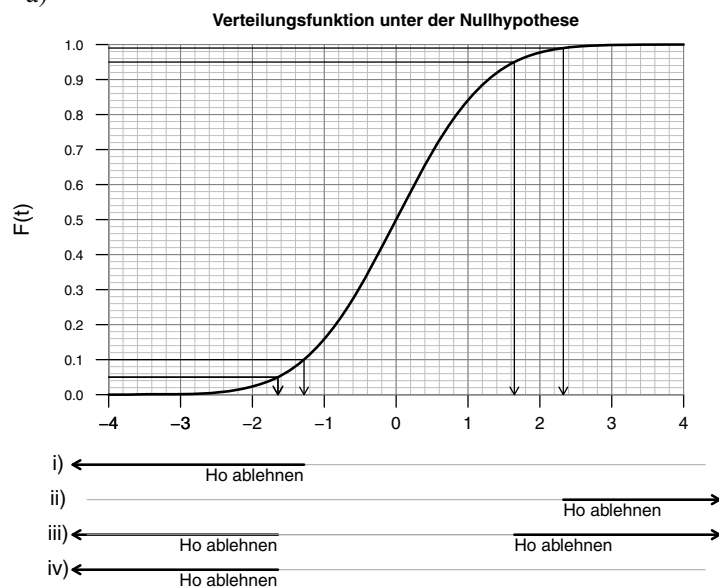
d) (i)  $P(PG \leq 5) \approx 0.15$ , (ii)  $P(PG \geq 10) = 1 - P(PG \leq 9) \approx 0.15$ ,

(iii)  $2 \cdot P(PG \leq 5) \approx 0.3$ .

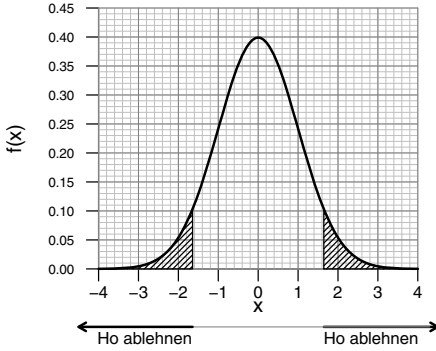
<sup>7</sup> Überprüfen Sie die Ergebnisse mit `sum(dbinom(A, 15, 0.5))`, wobei für A die Werte aus dem Ablehnungsbereich einzusetzen sind.

- [ 8.15 ] a)  $\alpha = 0.05 : A = \{0, 1\}$ ;  $\alpha = 0.1 : A = \{0, 1, 2\}$ .  
 b)  $\alpha = 0.05 : A = \{9, 10\}$ ;  $\alpha = 0.1 : A = \{8, 9, 10\}$ .  
 c)  $\alpha = 0.01 : A = \{0\} \cup \{10\}$ ;  $\alpha = 0.05 : A = \{0, 1\} \cup \{9, 10\}$ .  
 d) (i)  $P(PG \leq 2) \approx 0.055$ ; (ii)  $P(PG \geq 7) \approx 0.17$ ; (iii)  $2 \cdot P(PG \geq 9) \approx 0.02$ .

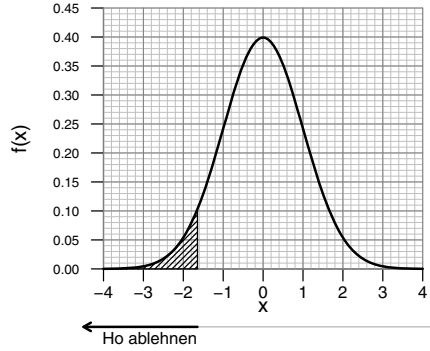
[ 8.16 ] a)



**Ablehnungsbereich in iii)**

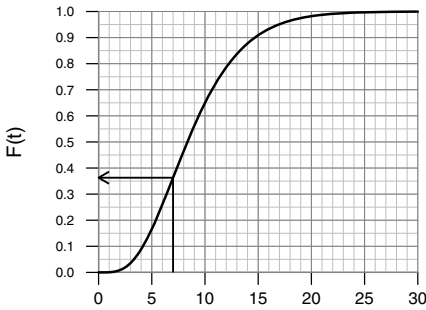


**Ablehnungsbereich in iv)**

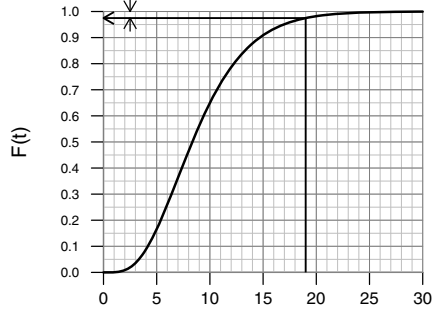


- i) Links:  $\alpha = 0.1 : A = (-\infty, -1.28]$  kann mit  $q_{\text{norm}}(0.1)$  geprüft werden.  
 ii) Rechts:  $\alpha = 0.01 : A = [2.32, \infty)$  kann mit  $q_{\text{norm}}(0.99)$  geprüft werden.  
 iii) Beidseitig:  $\alpha = 0.1 : A = (-\infty, -1.64] \cup [1.64, \infty)$  kann mit  $q_{\text{norm}}(0.05)$  und  $q_{\text{norm}}(0.95)$  geprüft werden.  
 iv) Links:  $\alpha = 0.05 : A = (-\infty, -1.64]$  kann mit  $q_{\text{norm}}(0.05)$  geprüft werden.
- b)

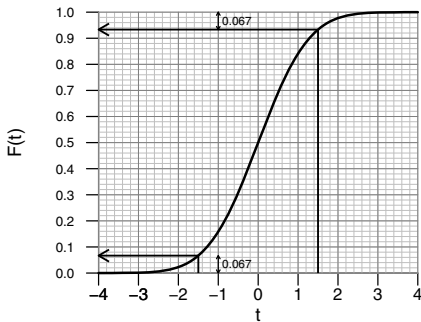
**i) P-Wert = 0.363**



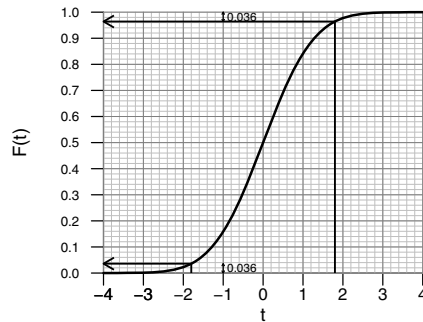
**ii) P-Wert = 1 - 0.975 = 0.025**

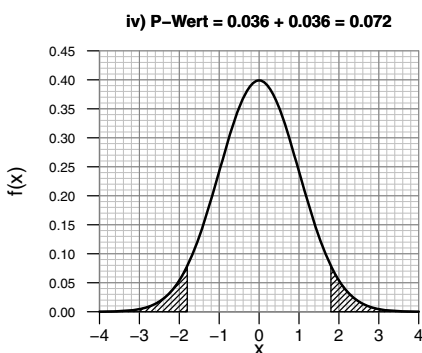
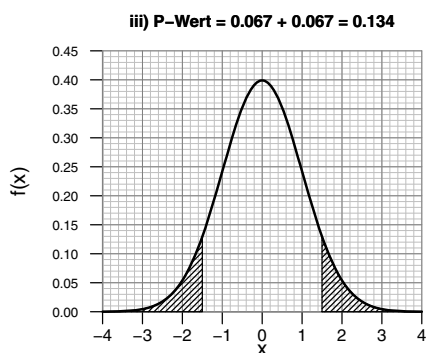
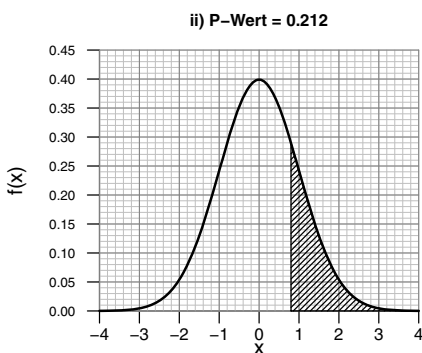
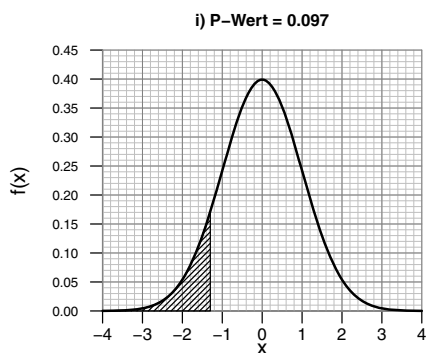


**iii) P-Wert = 0.067 + 0.067 = 0.134**



**iv) P-Wert = 0.036 + 0.036 = 0.072**





i)  $P(PG \leq -1.3) = 0.097$  kann mit `pnorm(-1.3)` geprüft werden.

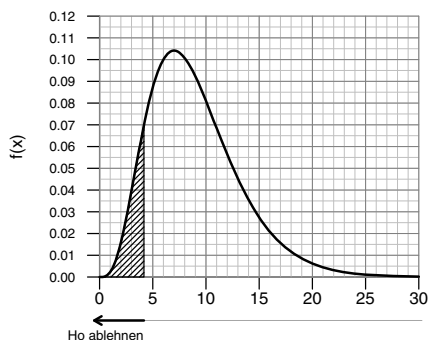
ii)  $P(PG \geq 0.8) = 1 - P(PG \leq 0.8) = 0.212$ . In **R**: `1-pnorm(0.8)`

iii)  $P(PG \leq -1.5) + P(PG \geq 1.5) = P(PG \leq -1.5) + 1 - P(PG \leq 1.5) = 0.134$  kann mit `2*pnorm(-1.5)` oder `pnorm(-1.5)+1-pnorm(1.5)` geprüft werden.

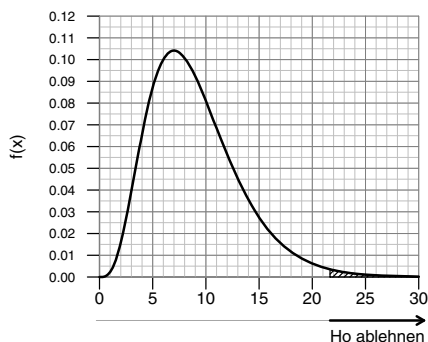
iv)  $P(PG \geq 1.8) + P(PG \leq -1.8) = 1 - P(PG \leq 1.8) + P(PG \leq -1.8) = 0.072$  kann mit `2*pnorm(-1.8)` oder `pnorm(-1.8)+1-pnorm(1.8)` geprüft werden.

[ 8.17 ] a)

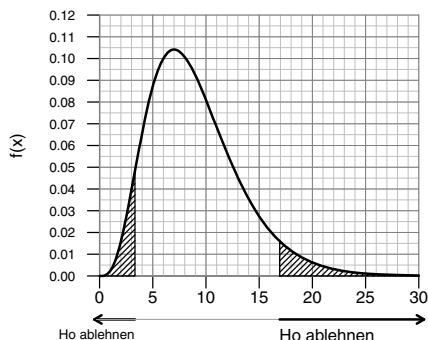
**Ablehnungsbereich in i)**



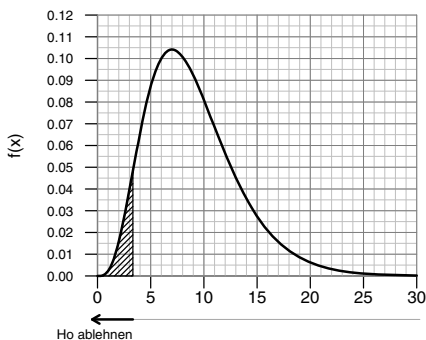
**Ablehnungsbereich in ii)**



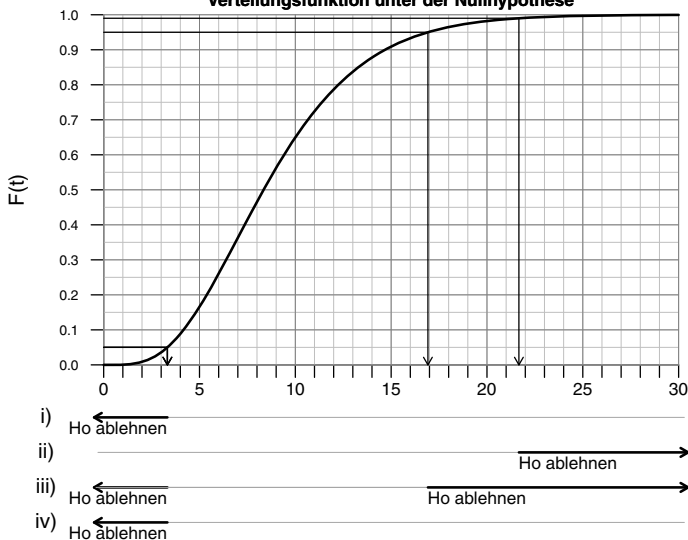
Ablehnungsbereich in iii)



Ablehnungsbereich in iv)



Verteilungsfunktion unter der Nullhypothese



i) Links:  $\alpha = 0.05$ ;  $A = [0, 3.33]$  kann mit  $qchisq(0.05, 9)$  geprüft werden.

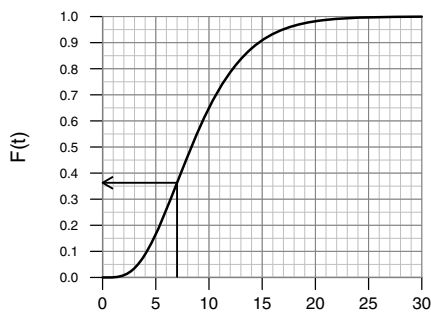
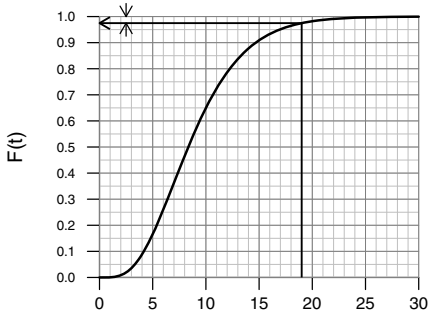
ii) Rechts:  $\alpha = 0.01$ ;  $A = [21.67, \infty)$  kann mit  $qchisq(0.99, 9)$  geprüft werden.

iii) Beidseitig:  $\alpha = 0.1$ ;  $A = [0, 3.33] \cup [16.92, \infty)$ ;  $\mathbf{R}$ :  $qchisq(c(0.05, 0.95), 9)$

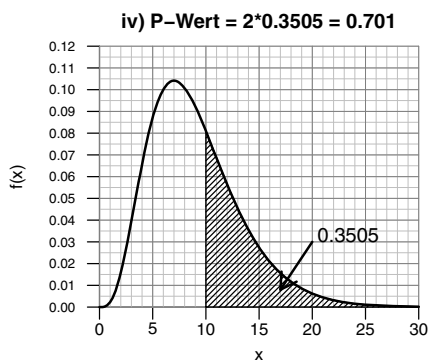
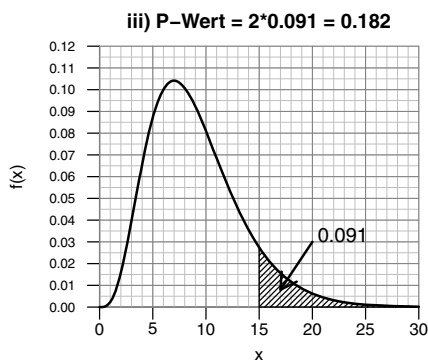
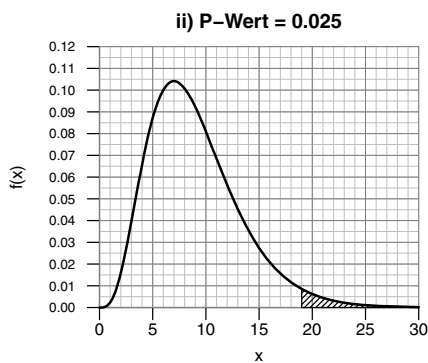
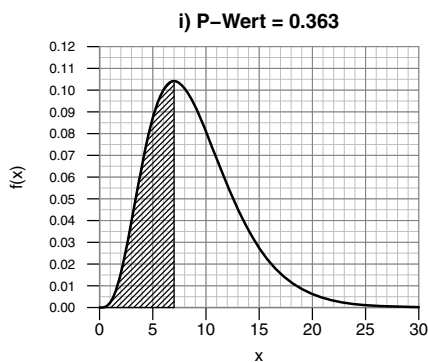
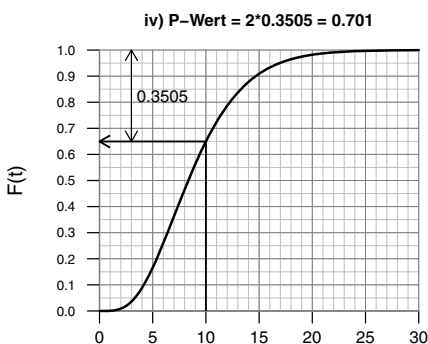
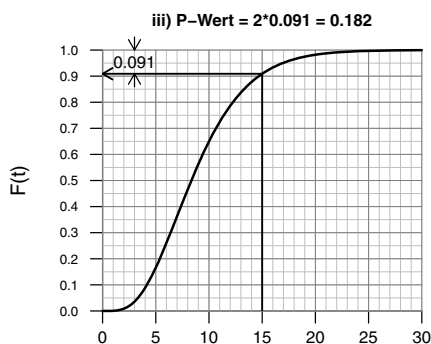
iv) Links:  $\alpha = 0.1$ ;  $A = [0, 4.17]$  kann mit  $qchisq(0.1, 9)$  geprüft werden.

b)

i) P-Wert = 0.363

ii) P-Wert =  $1 - 0.975 = 0.025$ 





- i)  $P(PG \leq 7) = 0.363$ , kann mit `pchisq(7, 9)` geprüft werden.  
 ii)  $P(PG \geq 19) = 0.025$ , kann mit `1-pchisq(19, 9)` geprüft werden.  
 iii)  $2 \cdot P(PG \geq 15) = 0.182$ , kann mit `2 * (1-pchisq(15, 9))` geprüft werden.  
 iv)  $2 \cdot P(PG \geq 10) = 0.701$ , kann mit `2 * (1-pchisq(10, 9))` geprüft werden.

[ 8.18 ] a) Wie üblich ist die abzusichernde Hypothese die Alternativhypothese, d.h.  $H_0: \pi \leq \pi_0 = 0.23$  gegen  $H_1: \pi > 0.23$ . Damit ist der Ablehnungsbereich rechts.

b) Die kritischen Werte erhalten wir aus dem Ausdruck.  $\alpha = 0.1 : A = \{7, 8, \dots, 17\}$ . Beachten Sie  $P(X \geq 7) = 1 - P(X < 7) = 1 - P(X \leq 6) = 1 - 0.926 = 0.074$ . Für  $\alpha = 0.05$  ist  $A = \{8, 9, \dots, 17\}$ . Machen Sie die Probe mit

`sum(dbinom(7:12, 17, 0.23))` und `sum(dbinom(8:12, 17, 0.23))`.

c) Der P-Wert ist  $P(PG \geq 6) = 1 - P(PG \leq 5) = 1 - 0.823 = 0.177$  nach dem **R**-Ausdruck oder nach `1-pbinom(5, 17, 0.23)`.

d) Die Nullhypothese kann bei  $\alpha = 0.1$  nicht verworfen werden. Wird die Nullhypothese verworfen, so ist die Irrtumswahrscheinlichkeit gleich dem P-Wert, d.h. 0.177.

### [ 8.19 ]

a)

Gruppe	BERUF	STUD	ABLOS	RUHE	SONST
$\bar{x}$	29.09	10.46	20.07	46.86	14.18
$S_*$	$\sqrt{2605.65}$	$\sqrt{426.98}$	$\sqrt{786.23}$	$\sqrt{4754.21}$	$\sqrt{519.09}$
$n$	306	213	14	49	67
$PG = T$	2.193	-8.638	-0.350	2.454	-3.057
$FG$	305	212	13	48	66
$A_{0.05}$	$(-\infty; -1.96] \cup [1.96; \infty)$	$(-\infty; -1.96] \cup [1.96; \infty)$	$(-\infty; -2.16] \cup [2.16; \infty)$	$(-\infty; -2.01] \cup [2.01; \infty)$	$(-\infty; -2] \cup [2; \infty)$
P-Wert	0.028	0.000	0.732	0.018	0.000

Die P-Werte können mit `2*pt(-2.193, 305)`; `2*pt(-8.638, 212)`; `2*pt(-0.350, 13)`; `2*pt(-2.454, 48)`; `2*pt(-0.3057, 66)` in **R** geprüft werden.

b) Nur das Konfidenzintervall für die Arbeitslosen enthält  $\mu_0$ . Dies ist die einzige Gruppe, für die die Nullhypothese nicht verworfen wird bei  $\alpha = 0.05$ .

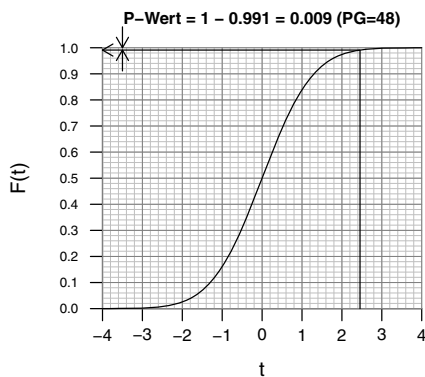
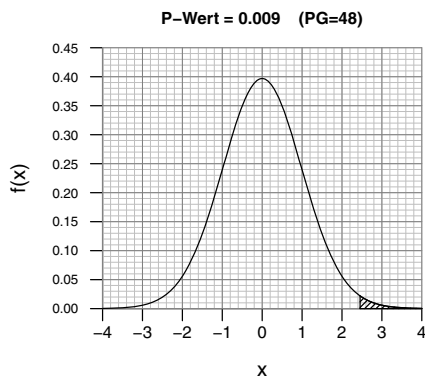
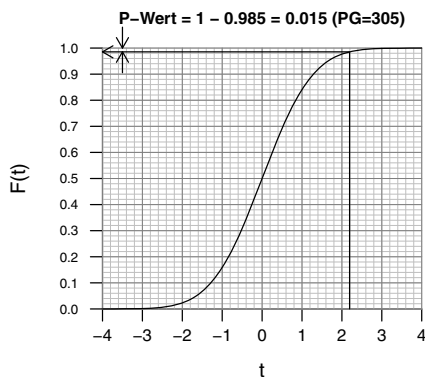
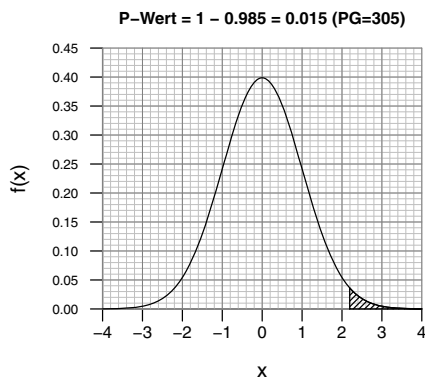
c)  $H_0$  wird für Studenten und Sonstige verworfen, nicht aber für die der Berufstätigen, der Arbeitslosen und der Ruheständler, da die Konfidenzintervalle für diese Gruppen  $\mu_0$  enthalten.

d) Rechtsseitiger Test auf Mittelwert.

$$i) T_{\text{Beruf}} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S_*} = \frac{(29.09 - 22.69)\sqrt{306}}{\sqrt{2605.65}} = 2.193$$

$$T_{\text{Ruhe}} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S_*} = \frac{(46.86 - 22.69)\sqrt{49}}{\sqrt{4754.21}} = 2.454$$

ii)  $P(PG \geq 2.193) = 0.015$   $P(PG \geq 2.454) = 0.009$ , mit `1-pt(2.193, 305)` und `1-pt(2.454, 48)`. Mit steigender Stichprobengröße nähert sich die  $t$ -Verteilung der Standardnormalverteilung an. Die untere Dichtefunktion ist folglich etwas flacher.



e) Linkssseitiger Test auf Mittelwert.

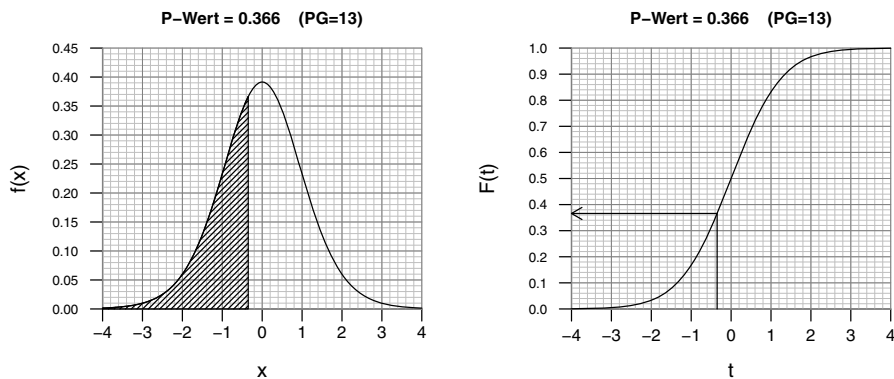
$$i) T_{\text{Stud}} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S_*} = \frac{(10.46 - 22.69)\sqrt{213}}{\sqrt{426.98}} = -8.638$$

$$T_{\text{Ablos}} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S_*} = \frac{(20.07 - 22.69)\sqrt{14}}{\sqrt{768.23}} = -0.350$$

$$ii) \text{STUD: } \alpha = 0.05 : A = (-\infty, -1.65] \quad \alpha = 0.05 : (-\infty, -1.77]$$

$$\alpha = 0.01 : A = (-\infty, -2.34] \quad \alpha = 0.01 : (-\infty, -2.65]$$

iii)  $P(PG \leq -8.638) \approx 0$ ,  $P(PG \leq -0.350) = 0.366$ . (Mit **R**: `pt(-8.638, 212)` und `pt(-0.350, 13)`). Mit steigender Stichprobengröße nähert sich die  $t$ -Verteilung der Standardnormalverteilung an. Die untere Dichtefunktion ist folglich ein wenig flacher. Wir zeigen die Bilder nur für den 2. Fall, da der P-Wert im 1. Fall  $\approx 0$  ist.

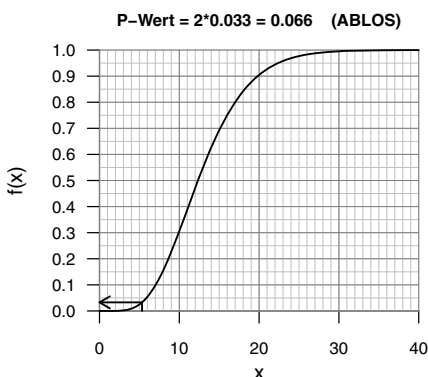
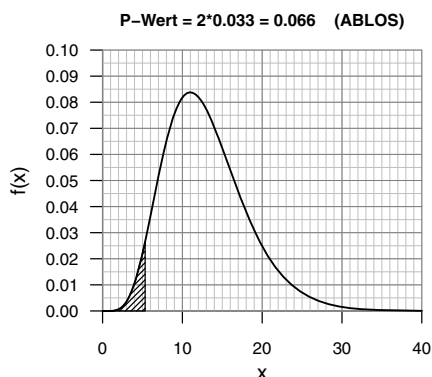


[ 8.20 ] a) Bedenken Sie, dass  $nS^2 = (n-1)S_*^2$ .

Gruppe	BERUF	STUD	ABLOS	RUHE	SONST
$S_*^2$	2605.65	426.98	786.23	4754.21	519.09
$n$	306	213	14	49	67
$nS^2$	794723.2	90519.76	10220.99	228202.1	34259.94
$PG$	414.61	47.225	5.332	199.054	17.874
$FG$	305	212	13	48	66
$H_0$	$\sigma^2 = \sigma_0^2$	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 = \sigma_0^2$	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 \geq \sigma_0^2$
$A_{0.05}$	$[0; 258.51] \cup [355.27; \infty)$	$[0; 179.30]$	$[0; 5.01] \cup [24.74; \infty)$	$[65.17; \infty)$	$[0; 48.31]$
$A_{0.01}$	$[0; 245.14] \cup [372.37; \infty)$	$[0; 167.06]$	$[0; 3.57] \cup [29.82; \infty)$	$[73.68; \infty)$	$[0; 42.24]$
P-Wert	0.00	0.00	0.07	0.00	0.00

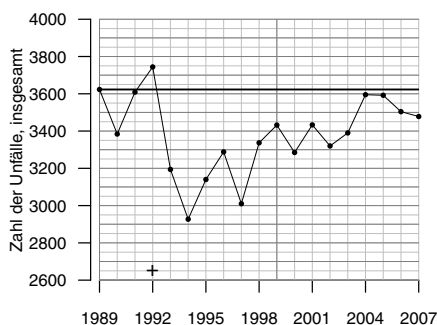
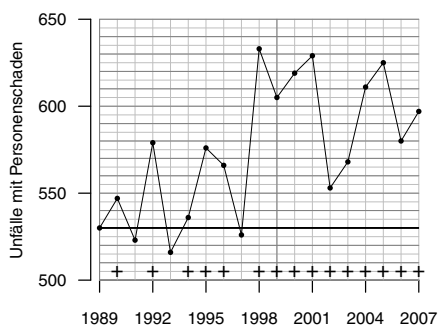
Die P-Werte können mit **R** wie folgt berechnet werden:

```
2*(1-pchisq(414.612, 305)); pchisq(47.225, 212); 2*pchisq(5, 332, 13);
1-pchisq(119.054, 48); pchisq(17.874, 66)
```



b) Die Ergebnisse der Tests sind anhand der Konfidenzintervalle zu erwarten. Das 95%- und 99%-KI der Arbeitslosen enthält jeweils  $\sigma_0^2$ , weshalb  $H_0$  auf 5%- und 1%-Signifikanzniveau nicht verworfen werden kann. Bei den Berufstätigen ist  $\sigma_0^2$  nicht in den Konfidenzintervallen enthalten, weshalb  $H_0$  auf allen üblichen Niveaus verworfen werden kann. Die Nullhypothese wird für alle weiteren Gruppen, außer natürlich die alle umfassende, verworfen.

[ 8.21 ] a) und c) [siehe auch R-Programm]



b)  $H_0: \pi \leq \pi_0 = 0.5$  und  $H_1: \pi > 0.5$ . Wir haben 15 '+'-Zeichen gezählt, d.h.  $PG = 15$ . Ablehnungsbereich für  $\alpha = 0.05$  ist  $A = \{13, 14, \dots, 18\}$  und für  $\alpha = 0.01$  ist  $A = \{15, 16, 17, 18\}$ . Der P-Wert ist  $P(PG \geq 15) = 0.004$ . Die Nullhypothese ist bei  $\alpha = 0.01$  abzulehnen.

c)  $H_0: \pi \geq \pi_0 = 0.5$  und  $H_1: \pi < 0.5$ . Wir zählen nur ein '+'-Zeichen, d.h.  $PG = 1$ . Ablehnungsbereich für  $\alpha = 0.05$  ist  $A = \{0, 1, \dots, 5\}$  und für  $\alpha = 0.01$  ist  $A = \{0, 1, \dots, 4\}$ . Der P-Wert ist  $P(PG \geq 1) \approx 0$ , d.h. die Nullhypothese ist abzulehnen.

[ 8.22 ] a) Nach Tabelle A2 ist  $z_{0.005} = 2.58$  und  $z_{0.025} = 1.96$ . Oder  $qnorm(0.995)$  und  $qnorm(0.975)$ .

- b)  $EG_U = 10 - 2.58 \cdot 0.3/\sqrt{9} \approx 9.742$        $EG_O = 10 + 2.58 \cdot 0.3/\sqrt{9} \approx 10.258$   
 $WG_U = 10 - 1.96 \cdot 0.3/\sqrt{9} \approx 9.804$        $WG_O = 10 + 1.96 \cdot 0.3/\sqrt{9} \approx 10.196$
- c) Das Verfahren kann als zweiseitiger Hypothesentest zur Nullhypothese  $H_0: \mu = \mu_0$  gegen  $H_1: \mu \neq \mu_0$  bei bekannter Varianz  $\sigma^2$  aufgefasst werden. Es wird gewarnt bzw. eingegriffen, wenn die Prüfgröße in den Ablehnungsbereich fällt.
- d) Irrtümlich gewarnt wird mit einer Wahrscheinlichkeit von 0.05. Die Wahrscheinlichkeit eines irrtümlichen Eingriffs beträgt 0.01.

[ 8.23 ] a) Hypothesentest, über  $\mu$  bei unbekannter Varianz  $\sigma^2$ , d.h. ein t-Test, wobei  $\mu = E[X_A - 3 \cdot X_E]$ . Die Prüfgröße ist  $t$ -verteilt mit 74 Freiheitsgraden (*degrees of freedom*, siehe df).

b) Es handelt sich um einen zweiseitigen Test mit  $H_0: \mu = 0$  und  $H_1: \mu \neq 0$ .

c)  $H_0$  kann bei allen drei Signifikanzniveaus verworfen werden, da der P-Wert jeweils geringer ist.

d) Das Intervall  $[1.545436, 4.774654]$  ist ein Konfidenzintervall für  $\mu$ , dass den wahren Mittelwert  $\mu$  mit einer Wahrscheinlichkeit von  $1 - \alpha = 0.95$  enthält.

e) Eigentlich interessiert uns  $H_1: \mu > 0$ . Das kann in R mit

`t.test(Ausgangstest-3*Eingangstest, alternative="greater", mu=0)` erreicht werden.

f) Irrtumswahrscheinlichkeit = P-Wert =  $P(PG \geq 3.8998) = 0.00021/2 = 0.000105$ .

[ 8.24 ] a)  $H_0: \sigma^2 \geq 1284955$  und  $H_1: \sigma^2 < 1284955$ .

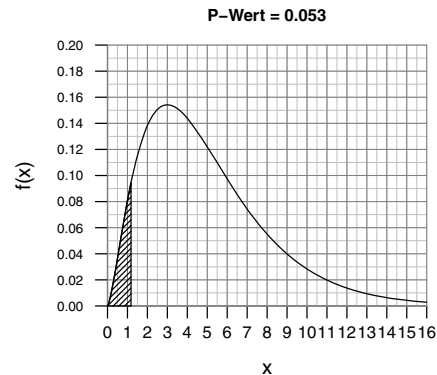
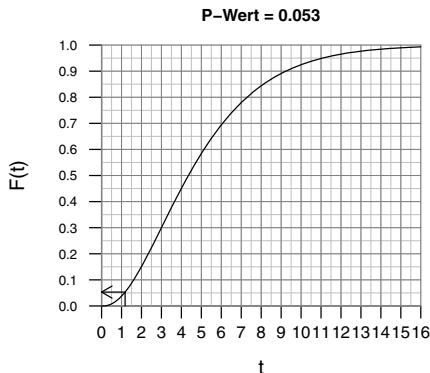
b)  $H_0$  wird verworfen, wenn Prüfgröße  $PG = \frac{n \cdot S^2}{\sigma_0^2}$  in den Ablehnungsbereich  $[0, 1.15]$  fällt, der mit der  $\chi_5^2$ -Verteilung (Tabelle A4) bestimmt wurde.

c) Aus den Daten erhalten wir  $S^2 = 252568.9 \Rightarrow PG = \frac{n \cdot S^2}{\sigma_0^2} = \frac{6 \cdot 252568.9}{1284955} \approx 1.179$ .

Unter  $H_0$  gilt  $PG \sim \chi_5^2$ .

d) Der P-Wert ist  $P(PG \leq 1.179)$ , was zwischen 1.1 und 1.2 liegt. Damit liegt der P-Wert zwischen 0.046 und 0.055. Der genaue Wert beträgt 0.0531239 und kann mit `pchisq(1.179, 5)` bestimmt werden.

e) Siehe auch R-Programm.



f) Da der P-Wert ungefähr 0.053 beträgt, wird die Nullhypothese bei  $\alpha = 0.1$  abgelehnt und die Alternative gilt als statistisch abgesichert. Bei  $\alpha = 0.05$  kann  $H_0$  nicht verworfen werden.

[ 8.25 ] a)  $H_0: \sigma^2 \leq 1284955$  und  $H_1: \sigma^2 > 1284955$ .

b) Die Prüfgröße ist  $PG = \frac{nS^2}{\sigma_0^2} = \frac{(n-1)S_*^2}{\sigma_0^2} = \frac{59 \cdot 1690277}{1284955} = 77.611$ . (Beachten Sie:  $nS^2 = (n-1)S_*^2$ )

c) Unter  $H_0$  gilt  $PG \sim \chi_{59}^2$ -Verteilung. Den kritischen Wert kann man einer Tabelle, der Grafik oder aus `qchisq(0.95, 59)` ablesen: 77.931 und somit wird  $H_0$  verworfen, wenn  $PG \in A_{0.05} = [77.931, \infty)$ . Hier ist  $PG = 77.611 < 77.931$ , d.h.  $H_0$  kann bei  $\alpha = 0.05$  nicht verworfen werden.

d) Der P-Wert ist  $P(PG \geq 77.611) = 1 - P(PG \leq 77.611)$ . Da  $77.6 < PG < 77.7$ , liegt der P-Wert nach der **R**-Ausgabe zwischen  $1 - 0.948$  und  $1 - 0.947$ , d.h. zwischen 0.052 und 0.053. Genau bestimmen kann man ihn mit `1-pchisq(77.611, 59)`.

[ 8.26 ] a)  $H_0: \mu \leq 6160.05$  und  $H_1: \mu > 6160.05$ .

b)  $T = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S_*} = \frac{(7314.85 - 6160.05) \cdot \sqrt{60}}{\sqrt{1690277}} = 6.880$ . Unter  $H_0$  gilt  $PG \sim t_{59}$ .

c) Für  $\alpha = 0.05$  wird die Nullhypothese verworfen, wenn  $T \in A_{0.05} = [1.67, \infty)$ . Da  $6.880 > 1.67$  wird  $H_0$  verworfen.

d) Der P-Wert ist fast Null, siehe `1-pt(6.880, 59)`.

## 2.9 Paare von Zufallsvariablen - Lösungen

[ 9.1 ] a) F b) W c) W d) W e) F f) W g) W h) W i) W j) W

[ 9.2 ] a) F b) W c) W d) W e) F f) W g) W h) W i) F j) W k) W

[ 9.3 ] a) W b) F c) F d) W e) W

[ 9.4 ] a) W b) W c) F d) W e) W f) F g) W h) F i) W

[ 9.5 ] a) W b) W c) W d) W e) W f) W g) F h) W

[ 9.6 ] a) F b) W c) W d) W e) W f) F g) W h) W i) F j) W k) F

[ 9.7 ] a) i)  $\frac{5}{32}; \frac{4}{32} = \frac{1}{8}; \frac{3}{32}$  ii)  $\frac{2}{32} = \frac{1}{16}; \frac{3}{32}; \frac{6}{32} = \frac{3}{16}$  iii)  $\frac{12}{32} = \frac{3}{8}; \frac{4}{32} = \frac{1}{8}; \frac{5}{32}; \frac{17}{32}$  iv)  $\frac{13}{32}; \frac{7}{32}; \frac{7}{32}$

b)

$x \backslash y$	0	1	2	3	$P_X(x)$
0	$3/32$	$2/32$	$2/32$	$1/32$	$8/32 = 1/4$
1	$2/32$	$2/32$	$1/32$	$3/32$	$8/32 = 1/4$
2	$2/32$	$1/32$	$3/32$	$2/32$	$8/32 = 1/4$
3	$1/32$	$3/32$	$2/32$	$2/32$	$8/32 = 1/4$
$P_Y(y)$	$8/32 = 1/4$	$8/32 = 1/4$	$8/32 = 1/4$	$8/32 = 1/4$	$32/32 = 1$

$$c) P_{Y|X}(y|X=x) = \frac{P(X=x, Y=y)}{P_X(x)}; P_{X|Y}(x|Y=y) = \frac{P(X=x, Y=y)}{P_Y(y)}.$$

$$P_{Y|X}(y|X=0) = \begin{cases} 3/8 & y=0 \\ 1/4 & y=1 \\ 1/4 & y=2 \\ 1/8 & y=3 \\ 0 & \text{sonst} \end{cases} \quad P_{Y|X}(y|X=1) = \begin{cases} 1/4 & y=0 \\ 1/4 & y=1 \\ 1/8 & y=2 \\ 3/8 & y=3 \\ 0 & \text{sonst} \end{cases}$$

$$P_{Y|X}(y|X=2) = \begin{cases} 1/4 & y=0 \\ 1/8 & y=1 \\ 3/8 & y=2 \\ 1/4 & y=3 \\ 0 & \text{sonst} \end{cases} \quad P_{Y|X}(y|X=3) = \begin{cases} 1/8 & y=0 \\ 3/8 & y=1 \\ 1/4 & y=2 \\ 1/4 & y=3 \\ 0 & \text{sonst} \end{cases}$$

$$P_{X|Y}(x|Y=0) = \begin{cases} 3/8 & x=0 \\ 1/4 & x=1 \\ 1/4 & x=2 \\ 1/8 & x=3 \\ 0 & \text{sonst} \end{cases} \quad P_{X|Y}(x|Y=1) = \begin{cases} 1/4 & x=0 \\ 1/4 & x=1 \\ 1/8 & x=2 \\ 3/8 & x=3 \\ 0 & \text{sonst} \end{cases}$$

$$P_{X|Y}(x|Y=2) = \begin{cases} 1/4 & x=0 \\ 1/8 & x=1 \\ 3/8 & x=2 \\ 1/4 & x=3 \\ 0 & \text{sonst} \end{cases} \quad P_{X|Y}(x|Y=3) = \begin{cases} 1/8 & x=0 \\ 3/8 & x=1 \\ 1/4 & x=2 \\ 1/4 & x=3 \\ 0 & \text{sonst} \end{cases}$$



d)

$s \setminus t$	0	1	2	3
0	3/32	5/32	7/32	8/32
1	5/32	9/32	12/32	16/32
2	7/32	12/32	18/32	24/32
3	8/32	16/32	24/32	32/32

e) Aus der Tabelle in Lösung d) lesen wir ab: 12/32 ; 18/32 ; 12/32 ; 9/32

f)  $\text{Kov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$ . Dabei ist

$$E(X \cdot Y) = \sum_{i=0}^3 \sum_{j=0}^3 i \cdot j \cdot P(X=i, Y=j) = (1 \cdot 1 \cdot 2 + 1 \cdot 2 \cdot 1 + 1 \cdot 3 \cdot 3 + 2 \cdot 1 \cdot 1 + 2 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 2 + 3 \cdot 1 \cdot 3 + 3 \cdot 2 \cdot 2 + 3 \cdot 3 \cdot 2) / 32 = 78/32.$$

$$E(X) = E(Y) = \frac{0+1+2+3}{4} = \frac{6}{4} = \frac{3}{2} \Rightarrow \text{Kov}(X, Y) = 78/32 - (3/2) \cdot (3/2) = 3/16;$$

$$E(X^2) = E(Y^2) = \frac{0^2 + 1^2 + 2^2 + 3^2}{4} = \frac{14}{4} = \frac{7}{2} \Rightarrow \text{Var}(X) = \text{Var}(Y) = \frac{7}{2} - \left(\frac{3}{2}\right)^2 = \frac{14-9}{4} = \frac{5}{4} \Rightarrow \rho = \frac{\text{Kov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{3/16}{5/4} = \frac{3}{20} = 0.15.$$

g) Nein, da die Kovarianz  $\neq 0$  ist. Andere Möglichkeit:  $X$  und  $Y$  sind nicht unabhängig, da z.B.  $P_X(0) \cdot P_Y(0) = 1/16 \neq P(0, 0) = 3/32$ .[ 9.8 ] a) i)  $\frac{1}{3}; \frac{1}{4}; \frac{1}{4}$ ; ii)  $\frac{1}{4}; \frac{5}{12}; \frac{1}{3}$ 

$$\text{b) } P_X(x) = \begin{cases} 7/12 & x=0 \\ 5/12 & x=1 \\ 0 & \text{sonst} \end{cases} \quad P_Y(y) = \begin{cases} 5/12 & y=0 \\ 3/12 & y=1 \\ 4/12 & y=2 \\ 0 & \text{sonst} \end{cases}$$

$$\text{c) } P_{Y|X}(y|X=0) = \begin{cases} 3/7 & y=0 \\ 2/7 & y=1 \\ 2/7 & y=2 \\ 0 & \text{sonst} \end{cases} \quad P_{Y|X}(y|X=1) = \begin{cases} 2/5 & y=0 \\ 1/5 & y=1 \\ 2/5 & y=2 \\ 0 & \text{sonst} \end{cases}$$

$$P_{X|Y}(x|Y=0) = \begin{cases} 3/5 & x=0 \\ 2/5 & x=1 \\ 0 & \text{sonst} \end{cases} \quad P_{X|Y}(x|Y=1) = \begin{cases} 2/3 & x=0 \\ 1/3 & x=1 \\ 0 & \text{sonst} \end{cases}$$

$$P_{X|Y}(x|Y=2) = \begin{cases} 1/2 & x=0 \\ 1/2 & x=1 \\ 0 & \text{sonst} \end{cases}$$

d)

$s \setminus t$	0	1	2
0	3/12	5/12	7/12
1	5/12	8/12	12/12

e)  $\text{Kov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 5/12 - 5/12 \cdot 11/12 = 5/144$ .  $\text{Var}(X) = E(X^2) - E^2(X) = \frac{5}{12} - \frac{25}{144} = \frac{35}{144}$ ,  $\text{Var}(Y) = E(Y^2) - E^2(Y) = \frac{19}{12} - \frac{121}{144} = \frac{107}{144}$ ,  $\Rightarrow \rho = \text{Kov}(X, Y) / \sqrt{\text{Var}(X) \cdot \text{Var}(Y)} = 0.0817$

f) Nein, da  $\text{Kov}(X, Y) \neq 0$ .

[ 9.9 ] Wegen der Unabhängigkeit ist die gemeinsame Wahrscheinlichkeitsfunktion das Produkt der beiden Randwahrscheinlichkeitsfunktionen:

$x \setminus y$	0	1	2
0	0.30	0.18	0.12
1	0.20	0.12	0.08

[ 9.10 ] a)

$x \setminus y$	0	1	2	$P_X(x)$
0	2/12	1/12	1/12	1/3
1	1/12	2/12	1/12	1/3
2	1/12	1/12	2/12	1/3
$P_Y(y)$	1/3	1/3	1/3	1

$X$  und  $Y$  sind nicht unabhängig, da z.B.  $P_X(0) \cdot P_Y(0) = (1/3) \cdot (1/3) = 1/9 \neq 2/12 = P_{XY}(0,0)$ .

b)

$s \setminus t$	0	1	2
0	2/12	3/12	4/12
1	3/12	6/12	8/12
2	4/12	8/12	12/12

c)  $\text{Kov}(X, Y) = E(XY) - E(X)E(Y) = 7/6 - 1 \cdot 1 = 1/6$ ,  $\text{Var}(X) = \text{Var}(Y) = 2/3$ ,  $\rho = (1/6) / \sqrt{(2/3)^2} = 1/4$ .

d)  $P_{X|Y}(x|1) = 1/4$  für  $x = 0$ ;  $1/2$  für  $x = 1$ ;  $1/4$  für  $x = 2$  und sonst Null.

e)  $E(X|Y = 1) = 1$ ,  $E(Y|X = 2) = 1.25$ .

$\text{Var}(X|Y = 1) = E(X^2|Y = 1) - E^2(X|Y = 1) = 1.5 - 1 = 0.5$ .

$\text{Var}(Y|X = 2) = E(Y^2|X = 2) - E^2(Y|X = 2) = 2.25 - 1.5625 = 0.6875$ .

[ 9.11 ] a)  $P(X > 0, Y \leq 1) = 0.2 + 0.1 = 0.3$ ;  $P(X > 0, Y > 0) = 0.1 + 0.1 = 0.2$ ;  $P(X < 1, Y < 2) = 0.3 + 0.2 = 0.5$

b)

$x \setminus y$	0	1	2	$P_X(x)$
0	0.3	0.2	0.1	0.6
1	0.2	0.1	0.1	0.4
$P_Y(y)$	0.5	0.3	0.2	1

$$\text{c) } P_{Y|X}(y|x=1) = \begin{cases} 1/2 & y=0 \\ 1/4 & y=1 \\ 1/4 & y=2 \\ 0 & \text{sonst} \end{cases} \quad \text{d) } P_{X|Y}(x|y=1) = \begin{cases} 2/3 & x=0 \\ 1/3 & x=1 \\ 0 & \text{sonst} \end{cases}$$

$$\text{e) } E(Y|X=1) = 3/4 \text{ und } E(X|Y=1) = 1/3.$$

f)

$s \setminus t$	0	1	2
0	0.3	0.5	0.6
1	0.5	0.8	1.0

$$\text{g) } \text{Kov}(X, Y) = 0.3 - 0.4 \cdot 0.7 = 0.02$$

$$\rho = \frac{0.02}{\sqrt{(0.4 - 0.16) \cdot (1.1 - 0.49)}} \approx 0.052.$$

h)  $X$  und  $Y$  sind nicht unabhängig verteilt, da  $\text{Kov}(X, Y) \neq 0$ . oder  $P(X=1) \cdot P(Y=1) = 0.4 \cdot 0.3 = 0.12 \neq 0.1 = P(X=1, Y=1)$ .

[ 9.12 ] a) Mithilfe des Hinweises berechnen wir für  $x > 0$ :

$$f_X(x) = \int_0^1 y e^{-yx} dy = \left[ -\frac{y}{x} e^{-yx} \right]_0^1 + \frac{1}{x} \int_0^1 e^{-yx} dy = \frac{-1}{x} e^{-x} - \left[ \frac{1}{x^2} e^{-yx} \right]_0^1 = \frac{1}{x^2} - \frac{e^{-x}}{x} - \frac{e^{-x}}{x^2}.$$

$$f_Y(y) = \int_0^\infty x y e^{-yx} dx = [1 - e^{-yx}]_0^\infty = 1 \text{ für } 0 < y < 1, \text{ d.h. } Y \sim U(0, 1);$$

Rechteckverteilung im Intervall  $[0, 1]$ .

$$\text{b) } f_{X|Y}(x|y) = f_{X,Y}(x, y) / f_Y(y) = y e^{-yx} \text{ für gegebenes } y \text{ und } f_{Y|X}(y|x) = f_{X,Y}(x, y) / f_X(x) = y e^{-yx} / \left( \frac{1}{x^2} - \frac{e^{-x}}{x} - \frac{e^{-x}}{x^2} \right) \text{ für gegebenes } x.$$

c) Exponentialverteilt mit  $\lambda = y = 0.5$

d) Die bedingte Erwartung ist  $1/\lambda = 1/y = 3/2$  mit Varianz  $9/4$ , bzw. 3 mit der Varianz 9.

e) Für  $x=1$  ist die bedingte Dichte  $f_{Y|X}(y|x) = y e^{-y} / \left( \frac{1}{x^2} - \frac{e^{-x}}{x} - \frac{e^{-x}}{x^2} \right) = y e^{-y} / (1 - 2 \cdot e^{-1})$  und somit keine der uns geläufigen Verteilungen. Es ist aber leicht zu prüfen, dass dies in der Tat eine Dichtefunktion ist. Bedingter Erwartungswert und Varianz berechnen sich mithilfe von  $\int y f_{Y|1}(y|1) dy$  und  $\int y^2 f_{Y|1}(y|1) dy$  woraus sich  $E(Y|X=1) = 1.4530$  und  $\text{Var}(Y|X=1) = 0.431$  ergeben.

f) Zu i) Hier könnte man entweder direkt die Dichte benutzen oder die gemeinsame Verteilungsfunktion bestimmen durch ähnliche Integrationen wie oben durchgeführt. Zu ii) Durch Einsetzen in die Randverteilung von  $X$ . Zu iii) Durch

Einsetzen in die Randverteilung von  $Y$ . Da dies die Rechteckverteilung auf  $[0, 1]$  ist, folgt  $P(Y \geq 0.5) = 0.5$  und  $P(Y \leq 0.3) = 0.3$ . Zu iv) Hier verwenden wir die Exponentialverteilung mit  $\lambda = 0.5$  bzw.  $= 0.3$ .

g) Offenbar nein, da das Produkt der Randdichten nicht die gemeinsame Dichte ergibt.

h) Wenn wir c) betrachten, dann wissen wir, dass der Erwartungswert von  $X$  gegeben  $Y$  invers proportional zu  $Y$  ist. Somit erwarten wir eine negative Korrelation.

i)  $E(Y) = 0.5$  und  $\text{Var}(Y) = 1/12$ .

[ 9.13 ] a) Exponentiell mit  $\lambda = y$ , z.B.  $\lambda = y = 2$

b)  $E(X|Y = 2) = 1/2 = 0.5$ ,  $\text{Var}(X|Y = 2) = 1/2^2 = 0.25$ .

c)  $f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y) = ye^{-yx} \cdot e^{-y} = ye^{-y(x+1)} = ye^{-y(x+1)}$  für  $0 < x < \infty$  und  $0 < y < \infty$

d)  $F_{XY}(s,t) = \int_0^s \int_0^t ye^{-y(x+1)} dx dy = \int_0^s ye^{-y} \int_0^t e^{-xy} dx dy$   
 $= \int_0^s e^{-y} [-e^{-yx}]_0^t dy = \int_0^s e^{-y}(1 - e^{-yt}) dy = \int_0^s e^{-y} dy - \int_0^s e^{-y(t+1)} dy = (1 - e^{-s}) - \frac{1}{t+1}(1 - e^{-s(t+1)})$

e)  $P(X \leq 1, Y \leq 1)$ ; mit **R**:  $1 - \exp(-1) - 0.5 * (1 - \exp(-2))$  : 0.1997882

$P(X \leq 2, Y \leq 3)$ ; mit **R**:  $1 - \exp(-2) - 0.25 * (1 - \exp(-8))$  : 0.6147486

f)  $f_X(x) = \int_0^\infty ye^{-y(x+1)} dy = \frac{1}{(x+1)^2}$  für  $x > 0$  (sonst Null).

g)  $E(X) = \int_0^\infty \frac{x}{(x+1)^2} dx = 0.5$  und  $E(X^2) = \int_0^\infty \frac{x^2}{(x+1)^2} dx = 1/3 \Rightarrow \text{Var}(X) = E(X^2) - (E(X))^2 = 1/3 - 1/4 = 1/12$ .

h)  $f_{Y|X}(y|x) = f_{XY}(x,y)/f_X(x) = ye^{-y(x+1)}/(x+1)^2$

i) Hierzu können Sie entweder die Kovarianz berechnen mit Hilfe der Formel  $E(XY) - E(X)E(Y)$  oder argumentativ über die bedingten Verteilungen vorgehen: An der bedingten Verteilung von  $X$  gegeben  $Y$  sehen wir, dass für wachsende  $Y$  der Erwartungswert für  $X$  kleiner wird. Daraus leiten wir eine negative Korrelation ab.

[ 9.14 ] a) Damit  $f_{XY}$  eine gemeinsame Dichte ist, muss gelten

$$1 = a \int_0^1 \int_0^1 x^3(1-y)^2 dx dy = a \int_0^1 (1-y)^2 \left( \int_0^1 x^3 dx \right) dy =$$

$$a \int_0^1 (1-y)^2 \left[ \frac{x^4}{4} \right]_0^1 dy = \frac{a}{4} \int_0^1 (1-y)^2 dy = \frac{a}{4} \int_0^1 (1-2y+y^2) dy =$$

$$\frac{a}{4} \left[ y - y^2 + \frac{1}{3}y^3 \right]_0^1 = \frac{a}{4} \left( 1 - 1 + \frac{1}{3} - 0 \right) = \frac{a}{12}, \text{ d.h. es muss gelten } \frac{a}{12} = 1 \iff a = 12$$

$$f_{XY}(x,y) = \begin{cases} 12x^3(1-y)^2 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

$$\text{b) } f_X(x) = 12x^3 \int_0^1 (1-y)^2 dy = 12x^3 \int_0^1 (1-2y+y^2) dy = 12x^3 \left[ y - y^2 + \frac{y^3}{3} \right]_0^1 =$$

$$= 12x^3(1 - 1 + 1/3 - 0) = 4x^3, \text{ d.h. } f_X(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

$$f_Y(y) = 12(1-y)^2 \int_0^1 x^3 dx = 12(1-y)^2 \left[ \frac{x^4}{4} \right]_0^1 = 12(1-y)^2 \cdot \left( \frac{1}{4} - 0 \right) = 3(1-y)^2,$$

$$\text{d.h. } f_Y(y) = \begin{cases} 3(1-y)^2 & 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

$X$  und  $Y$  sind unabhängig, da  $f_{XY}(x, y) = 12x^3(1-y)^2 = [4x^3] \cdot [3(1-y)^2] = f_X(x) \cdot f_Y(y)$ .

c) Für  $0 \leq s \leq 1$  gilt  $F_X(s) = 4 \int_0^s x^3 dx = [x^4]_0^s = s^4$  und für  $0 \leq t \leq 1$  gilt:  $F_Y(t) = 3 \int_0^t (1-y)^2 dy = 3 \int_0^t (1-2y+y^2) dy = 3 [y - y^2 + y^3/3]_0^t = 3t - 3t^2 + t^3 - 0 = t^3 - 3t^2 + 3t$ , d.h.

$$F_X(s) = \begin{cases} 0 & s < 0 \\ s^4 & 0 \leq s \leq 1 \\ 1 & s > 1 \end{cases} \quad F_Y(t) = \begin{cases} 0 & t < 0 \\ t^3 - 3t^2 + 3t & 0 \leq t \leq 1 \\ 1 & t > 1 \end{cases}$$

Wegen der Unabhängigkeit ist die gemeinsame Verteilungsfunktion das Produkt der beiden Randverteilungsfunktionen:

$$F_{XY}(s, t) = F_X(s) \cdot F_Y(t) = \begin{cases} 0 & s < 0 \text{ oder } t < 0 \\ s^4(t^3 - 3t^2 + 3t) & 0 \leq s \leq 1; 0 \leq t \leq 1 \\ s^4 & 0 \leq s \leq 1; t > 1 \\ t^3 - 3t^2 + 3t & s > 1; 0 \leq t \leq 1 \\ 1 & s > 1 \text{ und } t > 1 \end{cases}$$

$$\begin{aligned} \text{d) } P\left(X < \frac{1}{2}, Y < \frac{1}{4}\right) &= P\left(X < \frac{1}{2}\right) \cdot P\left(Y < \frac{1}{4}\right) = F_X\left(\frac{1}{2}\right) \cdot F_Y\left(\frac{1}{4}\right) \\ &= \frac{1}{2^4} \cdot \left(\frac{1}{4^3} - \frac{3}{4^2} + \frac{3}{4}\right) = \frac{1}{16} \cdot \left(\frac{1}{64} - \frac{3}{16} + \frac{3}{4}\right) = \frac{1}{16} \cdot \frac{1-12+48}{64} = \frac{37}{16 \cdot 64} \approx 0.036; \end{aligned}$$

$$\begin{aligned} P\left(X > \frac{1}{3}, Y < \frac{1}{2}\right) &= P\left(X > \frac{1}{3}\right) \cdot P\left(Y < \frac{1}{2}\right) = \left(1 - F_X\left(\frac{1}{3}\right)\right) \cdot F_Y\left(\frac{1}{2}\right) \\ &= \left(1 - \frac{1}{3^4}\right) \cdot \left(\frac{1}{2^3} - \frac{3}{2^2} + \frac{3}{2}\right) = \frac{80}{81} \cdot \frac{1-6+12}{8} = \frac{80 \cdot 7}{81 \cdot 8} \approx 0.864. \end{aligned}$$

e) Wegen der Unabhängigkeit  $\text{Kov}(X; Y) = 0$ . f) Wegen der Unabhängigkeit stimmen die bedingten Dichten mit den Randdichten überein.

[ 9.15 ] Wegen der Unabhängigkeit gilt  $P_{XY}(x, y) = P_X(x)P_Y(y)$ . Daher folgt:

- a)  $P(Y > 0) = 0.28/0.7 = 0.4$ ;  $P(X \leq 0) = 1 - 0.7 = 0.3$ ;  $P(Y \leq 0) = 1 - 0.4 = 0.6$ .  
 b)  $P(X > 0, Y \leq 0) = 0.7 \cdot 0.6 = 0.42$ ;  $P(X \leq 0, Y > 0) = 0.3 \cdot 0.4 = 0.12$ ;  $P(X \leq 0, Y \leq 0) = 0.3 \cdot 0.6 = 0.18$

[ 9.16 ]

- a) Wenn  $X = 0.5$ , dann ist  $Y = 0.25$   
 b) Wenn  $Y = 0.09$ , dann  $X \in \{-0.3; 0.3\}$   
 c) Es gilt

$$F_X(x) = \begin{cases} 0 & x < -1 \\ (1+x)/2 & -1 \leq x \leq 1 \\ 1 & x > 1 \end{cases} \quad F_Y(y) = \begin{cases} 0 & y < 0 \\ \sqrt{y} & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

und somit  $P(X > 0.8) = 1 - P(X \leq 0.8) = 1 - 0.9 = 0.1$  und  $P(Y < 0.25) = 0.5$

d) Wenn  $X > 0.8$ , dann ist  $Y = X^2 > 0.64$ . Somit ist die gesuchte Wahrscheinlichkeit gleich Null.

e)  $Kov(X, Y) = Kov(X, X^2) = E(X^3) - E(X)E(X^2) = 0 - 0 \cdot \frac{1}{3} = 0$ , denn für  $k \in \mathbb{N}$  gilt

$$E(X^k) = \int_{-1}^1 \frac{1}{2} x^k dx = \left[ \frac{1}{2(k+1)} x^{k+1} \right]_{-1}^1 = \frac{1}{k+1}$$

wenn  $k$  gerade und sonst Null. Dennoch sind  $X$  und  $Y$  offensichtlich abhängig, da  $Y$  als Funktion von  $X$  definiert ist. Man kann z.B. prüfen, dass das Produkt der Randdichten nicht die gemeinsame Dichte ergibt und ebenso das Produkt der Randwahrscheinlichkeiten nicht die gemeinsame etc., vergleiche etwa d).

[ 9.17 ] Die Dichte der Exponentialverteilung ist  $f(x) = \lambda e^{-\lambda x}$  für  $x \geq 0$  und sonst Null mit Erwartungswert  $1/\lambda$  und Varianz  $1/\lambda^2$ . Die Verteilungsfunktion ist  $F(t) = (1 - e^{-\lambda t})$  für  $t > 0$  und sonst Null.

a) Wenn  $E(B_1) = 1/\lambda = 5$  ist, folgt  $\lambda = 1/5 = 0.2$ ;  $f_{B_1}(b) = 0.2e^{-0.2b}$  für  $b \geq 0$

$$\text{und Null sonst. } F_{B_1}(t) = \begin{cases} 0 & \text{für } t < 0 \\ 1 - e^{-0.2t} & \text{für } t \geq 0 \end{cases}$$

$\text{Var}(B_1) = 1/\lambda^2 = 1/(1/5)^2 = 25$ . Die Wahrscheinlichkeiten berechnet man durch Integration der Dichtefunktion, einfacher jedoch mit der Verteilungsfunktion oder in **R** mit `pexp(t, 0.2)` bzw. `1-pexp(t, 0.2)`: 0.451, 0.865, 0.819, 0.368

b) Dieselben Ergebnisse wie in a), da es die gleiche Verteilung ist.

c) Wegen der Unabhängigkeit ist die gemeinsame Dichte das Produkt der Randdichten:

$$f_{B_1, B_2}(b_1, b_2) = \begin{cases} 0 & b_1 < 0 \text{ oder } b_2 < 0 \\ 0.04e^{-0.2b_1-0.2b_2} & b_1 \geq 0 \text{ und } b_2 \geq 0 \end{cases}$$

- d) Wegen der Unabhängigkeit ist die gemeinsame Verteilungsfunktion das Produkt der Randverteilungsfunktionen:  $F_{B_1, B_2}(s, t) = F_{B_1}(s)F_{B_2}(t) =$

$$\begin{cases} 0 & s < 0 \text{ oder } t < 0 \\ 1 + e^{-0.2s-0.2t} - e^{-0.2s} - e^{-0.2t} & s \geq 0 \text{ und } t \geq 0 \end{cases}$$

- e) Man erhält in beiden Fällen die gleichen Ergebnisse, da wegen der Unabhängigkeit die (oben berechneten) Wahrscheinlichkeiten der univariaten Verteilungen multipliziert die gemeinsame ergeben:<sup>8</sup> 0.204, 0.748, 0.390, 0.390, 0.369, 0.318.
- f) Man verwende die Formel für die Verteilungsfunktion oder berechne gemeinsame Verteilungsfunktion  $F_{X_1, X_2}(x_1, x_2)$  mit **R**.

$$2 * \text{pexp}(x1, 0.2) * \text{pexp}(x2, 0.2) - \text{pexp}(x1, 0.2) ^ 2$$

- i) 0.204, 0.748; ii) 0.577, 0.367; iii) 0.204, 0.748.

Bei iii) ist zu beachten, dass  $P(X_1 \leq 10, X_2 \leq 3) = P(X_1 \leq 3, X_2 \leq 3)$  und  $P(X_1 \leq 15, X_2 \leq 10) = P(X_1 \leq 10, X_2 \leq 10)$ , denn das Minimum kann nicht größer als das Maximum sein.

- g) Integration mit den in der Aufgabenstellung gegebenen Hinweisen ergibt:

$$f_{X_1}(x_1) = \begin{cases} 2\lambda e^{-2\lambda x_1} & 0 \leq x_1 < \infty \\ 0 & \text{sonst} \end{cases}$$

$$f_{X_2}(x_2) = \begin{cases} 2\lambda e^{-\lambda x_2}(1 - e^{-\lambda x_2}) & 0 \leq x_2 < \infty \\ 0 & \text{sonst} \end{cases}$$

- i) Exponential; Parameter  $(2\lambda) = 0.4$ ;  $E(X_1) = 2.5$ ;  $\text{Var}(X_1) = 6.25$ ;

$$F_{X_1}(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-0.4t} & t \geq 0 \end{cases} \quad \text{und somit}$$

- ii)  $P(X_1 \leq 3) = 1 - e^{-0.4 \cdot 3} = 0.699$ ;  $P(X_1 > 5) = 1 - P(X_1 \leq 5) = 0.135$ ;  $P(3 < X_1 \leq 10) = P(X_1 \leq 10) - P(X_1 \leq 3) = 0.283$ .

- iii)  $F_{X_2}(t) = \int_0^t f_{X_2}(x) dx = \int_0^t 2\lambda e^{-\lambda x} - 2\lambda e^{-2\lambda x} dx = 2 - 2e^{-\lambda t} - (1 - e^{-2\lambda t}) = 1 - 2e^{-\lambda t} + e^{-2\lambda t}$

- iv) Mit Hilfe von  $F_{X_2}(t)$  ergibt sich leicht: 0.204; 0.600; 0.544

$$\text{h) } f_{X_2|X_1}(x_2|x_1) = \frac{2\lambda^2 e^{-\lambda x_1} e^{-\lambda x_2}}{2\lambda e^{-2\lambda x_1}} = \lambda e^{-\lambda(x_2 - x_1)} \quad \text{für } x_2 \geq x_1.$$

- i)  $\int_{x_1}^t \lambda e^{-\lambda(x_2 - x_1)} dx_2 = 1 - e^{-\lambda(t - x_1)}$  für  $t \geq x_1$ .

- j) 0.181; 0.368; 0.148.

[ 9.18 ] In **R** mit `cor(X1, X3)` bzw. `cor(X2, X6)` mit den Ergebnissen<sup>9</sup> 0.8689758 bzw. 0.2978948.

Oder nach der Formel:  $\widehat{\text{Kov}}(X_i, X_j) / \sqrt{\widehat{\text{Var}}(X_i) \widehat{\text{Var}}(X_j)}$

<sup>8</sup> Es kann hier Ungenauigkeiten durch Rundungen geben: Obige Ergebnisse wurden mit **R** berechnet und erst am Ende gerundet.

<sup>9</sup> Es spielt hier keine Rolle, dass **R** bei der Berechnung der Kovarianz und der Varianzen den Nenner  $n - 1$  verwendet, denn bei der Berechnung der Korrelation kürzen sich diese Nenner weg.

Generell erwartet man in beiden Kombinationen eher positive Korrelationen, da die Kombination X1 und X3 beide eng mit der Preis-Leistungsfrage im Moment der Anschaffung (positiv) verbunden sind, während X2 und X6 für Vielfahrer von besonderer Bedeutung sind. Da allerdings X2 für alle preisbewussten Kunden wichtig ist, ist der Fahrkomfort auch für Kunden wichtig, die beim Auto weniger aufs Geld schauen. Daher sollte hier die Korrelation niedriger sein, was sie auch ist.

## 2.10 Anpassungs- und Unabhängigkeitstest - Lösungen

[ 10.1 ] a) F b) F c) W d) W e) W f) W g) F h) W i) F j) F k) F

[ 10.2 ] a) F b) W c) F d) W e) F f) W

[ 10.3 ] a) Die Prüfgröße ist:  $\sum_{i=1}^K (f_{i,o} - f_{i,e})^2 / f_{i,e}$  wobei  $K$  die Anzahl der gewählten Klassen ist,  $f_{i,o}$  die Anzahl der Beobachtungen (also die Häufigkeit) in Klasse  $i$  und  $f_{i,e}$  die laut Nullhypothese erwartete Anzahl. So wie die Aufgabenstellung formuliert ist, würde man als Nullhypothese formulieren, dass die jeweilige Prognose annähernd der tatsächlichen Stimmenverteilung folgt (insgesamt 4 getrennte Hypothesen). Somit setzt man stets  $f_{i,o} = \text{beobachtete Stimmenzahl in der Wahl}$  und  $f_{i,e} = \text{prognostizierte Prozentzahl} \cdot \frac{1}{100} \cdot \text{insgesamt beobachtete Stimmenzahl}$ . Die insgesamt beobachtete Stimmenzahl ist 3426593. Wir erhalten die PGs: 69 930.5 (ARD), 129 920.6 (ZDF), 210 877.9 (NDR), 165 439.5 (BILD).

Mit den kritischen Werten aus Tabelle A4 erhalten wir für die  $\chi^2_5$ -Verteilung (alternativ mit **R**: `qchisq(c(0.9, 0.95, 0.99), 5)`) die Ablehnungsbereiche: für  $\alpha = 0.1$  :  $[9.24, \infty)$ ; für  $\alpha = 0.05$  :  $[11.07, \infty)$ ; für  $\alpha = 0.01$  :  $[15.09, \infty)$ .

b) Sämtliche Prüfgrößen fallen deutlich in den Ablehnungsbereich. So dass alle Modelle abzulehnen sind.

c)  $\chi^2$ -Kriterium =  $PG/FG$ . Dann ist die Reihenfolge:

ARD 13986.09 , ZDF 25984.12, Bild 33087.91, NDR 42175.58.

[ 10.4 ] a) Mit einem Chi-Quadrat-Anpassungstest.  $H_0$ : Bei den Daten der zweiten Zeile handelt es sich um eine echte Stichprobe aus der in Zeile 3 angegebenen Verteilung.

b)  $PG = \sum_{i=1}^K (f_{i,o} - f_{i,e})^2 / f_{i,e}$  mit  $K$  Anzahl der Klassen, wobei  $f_{i,o}$  die beobachteten Häufigkeiten und  $f_{i,e}$  die unter  $H_0$  erwarteten Häufigkeiten sind, d.h.  $f_{i,e} = 3.7, 5.3, 7.4, 9.0, 10.2, 11.5, 11.8, 11.3, 10.9, 9.8, 9.2$ . Es folgt  $PG = 21.183$ . Unter  $H_0$  gilt  $PG \sim \chi^2_{10}$ . Die kritischen Werte erhalten wir aus Tabelle A4 oder mit `qchisq(c(0.9, 0.95, 0.99), K-1)` und damit die Ablehnungsbereiche für  $\alpha =$



0.1:  $[15.99; \infty)$ ; für  $\alpha = 0.05$ :  $[18.31; \infty)$ ; für  $\alpha = 0.01$ :  $[23.21; \infty)$ . Die Prüfgröße fällt bei  $\alpha = 0.1$  und  $0.05$  in den Ablehnungsbereich, d.h. die Nullhypothese ist zu verwerfen. Bei  $\alpha = 0.01$  kann die Nullhypothese nicht verworfen werden. Laut Tabelle liegt der P-Wert zwischen  $0.01$  und  $0.05$  und mit  $1-\text{pchisq}(21.183, 10)$  erhalten wir  $\approx 0.0199$ .

[ 10.5 ] Nach der Methode der Momente ist  $\hat{\lambda} = 2/\bar{x} = 0.38$ .

- a) Bei 20 Beobachtungen sind die erwarteten Häufigkeiten in Klassen mit der Wahrscheinlichkeit  $0.25$  jeweils  $f_{i,e} = 20 \cdot 0.25 = 5$ . Die vier Klassen sind:  $(0; 2.530)$ ,  $[2.530; 4.417)$ ,  $[4.417; 7.086)$ , und  $[7.086; \infty)$ . Wir zählen die beobachteten Häufigkeiten ab:  $f_{i,o} = 5, 7, 4, 4$ . Damit erhalten wir  $PG = 1.2$ . Unter  $H_0$  gilt  $PG \sim \chi^2_{K-1-1=2}$  (man bedenke, dass ein Parameter geschätzt wurde). Tabelle A4 entnimmt man, dass die PG für die üblichen Signikanzniveaus nicht in den Ablehnungsbereich fällt. Den P-Wert kann man mit  $1-\text{pchisq}(1.2, 2)$  bestimmen, was  $\approx 0.549$  ergibt.
- b) Für  $\lambda = 0.5$  sind die vier Klassen:  $(0; 1.923)$ ,  $[1.923; 3.357)$ ,  $[3.357; 5.385)$ , und  $[5.385; \infty)$ , so dass wir  $f_{i,o} = 4, 2, 7, 7$  und damit die  $PG = 3.6$  erhalten. Unter  $H_0$  gilt  $PG \sim \chi^2_{K-1=3}$ , da wir keinen Parameter geschätzt haben. Der P-Wert ist  $\approx 0.308$  ( $1-\text{pchisq}(3.6, 3)$ ), so dass wir auch diese Nullhypothese nicht verwerfen können.
- c) Nein, denn die erwarteten Häufigkeiten sollten in jeder Klasse mindestens 5 sein.

[ 10.6 ] a)  $n = 93$ ;  $f_{ie} = 93/10 = 9.3$

$$PG = \sum_{i=1}^K \frac{(f_{io} - f_{ie})^2}{f_{ie}} = \frac{(1 - 9.3)^2}{9.3} + \frac{(2 - 9.3)^2}{9.3} + \dots + \frac{(10 - 9.3)^2}{9.3} \approx 30.118.$$

Die Prüfgröße ist unter der Nullhypothese  $\chi^2$ -verteilt mit 9 Freiheitsgraden.

b) Mit Tabelle A4 ergeben sich die Ablehnungsbereiche:  $\alpha = 0.10$ :  $A = [14.68, \infty)$ ;  $\alpha = 0.05$ :  $A = [16.92, \infty)$ ;  $\alpha = 0.01$ :  $A = [21.67, \infty)$ . Die Nullhypothese wird demnach für alle drei  $\alpha$  verworfen.

c) Außer für  $\alpha = 0.0001$  wird die Prüfgröße für alle  $\alpha$  verworfen.

d) Der P-Wert liegt zwischen  $0.0005$  und  $0.0001$ . Mit  $1-\text{pchisq}(30.118, 9)$  kann der P-Wert berechnet werden, was  $\approx 0.0004$  ergibt. Die Nullhypothese wird verworfen und es gilt als statistisch abgesichert, dass die zehn Ziffern mit unterschiedlicher Wahrscheinlichkeit genannt werden.

[ 10.7 ] a) (Siehe auch **R**-Programm.)  $\hat{\lambda} = (19 \cdot 0 + 8 \cdot 1 + 2 \cdot 2 + 1 \cdot 3)/30 = 0.5$ .

Anzahl Bücher – 1	0	1	2	3	Summe
Beobachtete Häufigkeiten	19	8	2	1	30
Wahrscheinlichkeit nach Modell	0.607	0.303	0.076	0.014	1

Anzahl Bücher – 1	0	1	$\geq 2$	Summe
Beobachtete Häufigkeiten	19	8	3	30
Erwartete Häufigkeiten	18.21	9.09	2.7	30
Beobachtet - Erwartet	0.79	-1.09	0.30	--
$(\text{Beobachtet} - \text{Erwartet})^2$	0.6241	1.1881	0.0900	--
$(\text{Beobachtet} - \text{Erwartet})^2 / \text{Erwartet}$	0.0343	0.1307	0.0333	0.1983

- b) Da wir nur mit  $K = 3$  Klassen arbeiten und zudem einen Parameter der Verteilung geschätzt haben, ist  $FG = K - 2 = 1$  und die Verteilung der Teststatistik (PG) ist unter  $H_0$  demnach  $\chi^2_1$ . Mit Tabelle A4 erhalten wir die kritischen Werte nach der  $\chi^2_1$ -Verteilung und damit  $A_{0.1} = [2.71, \infty)$ ,  $A_{0.05} = [3.84, \infty)$  und  $A_{0.01} = [6.63, \infty)$ .
- c)  $H_0$  kann bei keinem der üblichen Signifikanzniveaus verworfen werden.
- d) Aus der **R**-Ausgabe folgt mit  $PG \approx 0.2$ ;  $FG = 1$  der P-Wert  $\approx 1 - 0.345 = 0.655$ .

[ 10.8 ] a) Laut **R**-Ausgabe ist  $\hat{\mu} = 16.93$  und  $\hat{\sigma}^2 = 18.12$ .

b) Vergleiche **R**-Programm und beachten Sie dabei, dass wir in der folgenden Tabelle mit gerundeten Werten arbeiten. Die exakten Werte, die man mit dem **R**-Programm erhält, weichen von diesen ab.

Zeit	$t \leq 10$	$10 < t \leq 14$	$14 < t \leq 18$	$18 < t \leq 22$	$t > 22$
Beob. Häuf.	2	19	22	23	12
Erw. Häuf.	3.90	15.60	27.30	21.84	9.36
Beob. - Erw.	-1.90	3.40	-5.30	1.16	2.64
$(\text{Beob.} - \text{Erw.})^2$	3.6100	11.5600	28.0900	1.3456	6.9696
$(\text{Beob.} - \text{Erw.})^2 / \text{Erw.}$	0.926	0.741	1.029	0.062	0.745

Die Summe der Werte der unteren Zeile ergibt  $PG = 3.503$ . Unter  $H_0$  gilt  $PG \sim \chi^2$  mit  $K - r - 1 = 5 - 2 - 1 = 2$  Freiheitsgraden. Mit Tabelle A4 erhalten wir  $A_{0.10} = [4.61, \infty)$ ;  $A_{0.05} = [5.99, \infty)$  und  $A_{0.01} = [9.21, \infty)$ . Der P-Wert kann mit `1-pchisq(3.503, 2)` berechnet werden und ist  $\approx 0.1735$ .

c) Vergleiche **R**-Programm: Die unter  $H_0$  erwartete Häufigkeit für jede der 10 Klassen ist  $= 78/10 = 7.8$ .

Intervall	Beob.	Erw.	$(\text{Beob.} - \text{Erw.})^2 / \text{Erw.}$
$t \leq 11.47$	11	7.8	1.313
$11.47 < t \leq 13.34$	9	7.8	0.185
$13.34 < t \leq 14.70$	4	7.8	1.851
$14.70 < t \leq 15.85$	7	7.8	0.082
$15.85 < t \leq 16.93$	8	7.8	0.005
$16.93 < t \leq 18.01$	4	7.8	1.851
$18.01 < t \leq 19.16$	8	7.8	0.005
$19.16 < t \leq 20.52$	8	7.8	0.005
$20.52 < t \leq 22.39$	14	7.8	4.928
$t > 22.39$	5	7.8	1.005

Die Summe der Werte der letzten Spalte ergibt  $PG = 11.23$ . Unter  $H_0$  gilt  $PG \sim \chi^2$  mit  $K - r - 1 = 10 - 2 - 1 = 7$  Freiheitsgraden. Mit Tabelle A4 erhalten wir  $A_{0.10} = [12.02, \infty)$ ;  $A_{0.05} = [14.07, \infty)$  und  $A_{0.01} = [18.48, \infty)$ . Der P-Wert kann mit `1-pchisq(11.23, 7)` berechnet werden und ist  $\approx 0.1289$ .

[ 10.9 ] a) In der Tabelle sind die zu  $\Phi(z) = 0.1, 0.2, 0.3, \dots, 0.9$  gehörigen  $z$ -Werte zu bestimmen, die jedoch nicht immer eindeutig sind. Genauer geht es mit **R**:

```
> round(qnorm((1:9)/10), digits=3)
[1] -1.282 -0.842 -0.524 -0.253  0.000  0.253  0.524  0.842  1.282
```

Dazu kommen  $-\infty$  und  $\infty$  als erste bzw. letzte Klassengrenze. Die erwartete Häufigkeit pro Klasse ist  $0.1 \cdot 100 = 10$ . Unter  $H_0$  gilt  $PG \sim \chi_9^2$ . Ablehnungsbereiche:  $A_{0.10} = [14.68; \infty)$ ;  $A_{0.05} = [16.92; \infty)$ ;  $A_{0.01} = [21.67; \infty)$ .

b)  $PG = ((13-10)^2 + (10-10)^2 + (15-10)^2 + (6-10)^2 + (9-10)^2 + (6-10)^2 + (7-10)^2 + (16-10)^2 + (10-10)^2 + (8-10)^2) / 10 = 11.6$ .

P-Wert =  $1 - P(PG \leq 11.6) \approx 0.237$ . (`1 - pchisq(11.6, 9)`) Die Nullhypothese  $H_0$ : Die Beobachtungen sind  $N(0, 1)$ -verteilt kann auf keinem der drei üblichen Signifikanzniveaus verworfen werden.

c) Nun ist  $PG = 8.6$  und  $H_0$  kann nicht verworfen werden.

d) Die  $PG$  bleibt 11.6. Die Freiheitsgrade ändern sich, da zwei Parameter geschätzt wurden, d.h. die Prüfgröße ist unter  $H_0$  jetzt  $\chi_{9-2}^2 = \chi_7^2$ -verteilt. Die Ablehnungsbereiche sind jetzt:  $A_{0.10} = [12.02 ; \infty)$ ;  $A_{0.05} = [14.07 ; \infty)$ ;  $A_{0.01} = [18.48 ; \infty)$ . Mit (`1 - pchisq(11.6, 7)`) erhält man als P-Wert  $\approx 0.115$ .  $H_0$  kann auf keinem der drei üblichen Signifikanzniveaus verworfen werden.

In **R** lassen sich die Rechnungen und Lösungen wie folgt überprüfen

```
> beob1 <- c(13,10,15,6,9,6,7,16,10,8) # Beobachtete Häuf. b)
> Erw <- rep(10,10) # Erwartete Häuf.
> diff <- beob1-Erw # Beobachtet - Erwartet
> diff; diff^2; diff^2/Erw;
> PG1 <- sum(diff^2/Erw) # Prüfgröße
> 1-pchisq(PG1,10-1) # P-Wert
> qchisq(c(0.9,0.95,0.99),9) # kritische Werte

> beob2 <- c(15,13,9,8,9,9,4,12,9,12) # Beobachtete Häuf. c)
> diff <- beob2-Erw # Beobachtet - Erwartet
> diff; diff^2; diff^2/Erw;
> PG <- sum(diff^2/Erw) # Prüfgröße
> 1-pchisq(PG,10-1) # P-Wert

> 1-pchisq(PG1,10-3) # P-Wert für d)
> qchisq(c(0.9,0.95,0.99),7) # kritische Werte
```

[ 10.10 ] a) Die folgenden Ergebnisse hängen davon ab, an welcher Stelle gerundet wird. Hier wurden zunächst die Wahrscheinlichkeiten mit **R** oder einem Taschenrechner berechnet und dann auf drei Dezimalstellen gerundet. Die gerundeten Wahrscheinlichkeiten wurden mit 96 multipliziert, um die erwarteten Häufigkeiten zu berechnen, die auf eine Dezimalstelle gerundet wurde. Es wurde mit diesen gerundeten Werten weiter gerechnet. Auch die Beiträge zur Prüfgröße (letzte Zeile) wurden gerundet und mit diesen Zahlen wurde weiter gerechnet. Wenn man ein exaktes Ergebnis benötigt, sollte man keine Zwischenrundungen vornehmen.

Erfolge	0	1	2	3
Beobachtet	37	30	19	10
Wahrscheinlichkeit	0.287	0.444	0.229	0.039
Erwartet	27.6	42.6	22.0	3.7
Beobachtet - Erwartet	9.4	-12.6	-3.0	6.3
(Beobachtet - Erwartet) <sup>2</sup>	88.36	158.76	9.00	39.69
(Beobachtet - Erwartet) <sup>2</sup> /Erwartet	3.201	3.727	0.409	10.727

Die Prüfgröße ist die Summe der Werte in der letzten Zeile:  $PG = 18.064$ . Unter  $H_0$  gilt  $PG \sim \chi^2_{4-1-1} = \chi^2_2$ . Mit Tabelle A4 ergibt sich  $A_{0.05} = [5.99 ; \infty)$  und  $A_{0.01} = [9.21 ; \infty)$ . Der P-Wert ist  $P(PG \geq 18.064) \approx 0.0001$ , was mit `1 - pchisq(18.064, 2)` berechnet werden kann. Die Nullhypothese  $H_0$ : „Die Daten sind  $b(3, 0.3)$  verteilt“ kann bei allen üblichen Signifikanzniveaus verworfen werden.

b) Die Daten sind nicht unabhängig identisch binomialverteilt. Jede Versuchsperson hat ein anderes  $\pi$ , d.h. wir haben hier mehrere Verteilungen bzw. eine Mischverteilung.

[ 10.11 ] Die 5 Klassen haben jeweils die Wahrscheinlichkeit 0.2, so dass die erwarteten Häufigkeiten  $0.2 \cdot (13 + 9 + 5 + 11 + 2) = 0.2 \cdot 40 = 8$  sind. Damit folgt  $PG = ((13-8)^2 + (9-8)^2 + (5-8)^2 + (11-8)^2 + (2-8)^2)/8 = (25 + 1 + 9 + 9 + 36)/8 = 10$ . Unter  $H_0$  gilt  $PG \sim \chi^2_3$ , da  $FG = K - 1 - 1 = 5 - 2 = 3$ . Die kritischen Werte sind nach Tabelle A4 7.81 und 11.34. Für  $\alpha = 0.05$  wird die Nullhypothese also verworfen, für  $\alpha = 0.01$  kann die Nullhypothese nicht verworfen werden. Der P-Wert ist nach der **R**-Ausgabe  $1 - 0.981 = 0.019$ .

[ 10.12 ] a) W b) W c) F d) W e) F f) W g) F

[ 10.13 ] a) F b) F c) W d) W e) F f) W g) W h) F

[ 10.14 ] Es wurden jeweils die Klassen von unten nach oben bzw. von rechts nach links soweit zusammengefasst, bis die erwarteten Häufigkeiten mindestens 5 pro Klasse waren. Die Teststatistik ist  $PG = \sum_i \sum_j \frac{n_{ij} - m_{ij}}{m_{ij}}$ , wobei  $n_{ij}$  die beobachteten Häufigkeiten und  $m_{ij} = n_{i.} n_{.j} / n$  die unter Unabhängigkeit erwarteten Häufigkeiten sind. Dabei ist  $n$  die Stichprobengröße und  $n_{i.}$  bzw.  $n_{.j}$  sind die Randhäufigkeiten.

table(Mathe, Deutsch):

In R können wir die unter Unabhängigkeit erwarteten Häufigkeiten mit folgendem Befehl berechnen:

```
> round(chisq.test(table(Mathe, Deutsch))$expected, 1)
```

```
Deutsch
Mathe  1    2    3    4    5
  1  4.5 18.0 14.9  2.4  0.2
  2  9.3 37.4 31.0  5.0  0.4
  3  7.6 30.6 25.4  4.1  0.3
  4  4.2 17.1 14.2  2.3  0.2
  5  0.4  1.8  1.5  0.2  0.0
```

Wir fassen jeweils die Klassen 3, 4, 5 zusammen und erhalten:

Mathe	Deutsch			$\sum$	Erwartete Häufigkeiten		
	1	2	$\geq 3$		1	2	$\geq 3$
1	9	17	14	40	4.464	18.026	17.511
2	10	46	27	83	9.262	37.403	36.335
$\geq 3$	7	42	61	110	12.275	49.571	48.155
$\sum$	26	105	102	233			

Es folgt  $PG = 16.654$ . Unter  $H_0$  gilt  $PG \sim \chi_4^2$  mit Ablehnungsbereichen  $A_{0.10} = [7.78, \infty)$ ;  $A_{0.05} = [9.49, \infty)$ ;  $A_{0.01} = [13.28, \infty)$ .

$H_0$  kann auf allen drei Signifikanzniveaus verworfen werden. Es ist statistisch abgesichert, dass die Merkmale abhängig sind. Der P-Wert ist  $P(PG \geq 16.654) < 0.01$ .

Mit  $1 - \text{pchisq}(16.654, 4)$  erhält man  $\approx 0.0023$ .

table(Kunst, Sport):

Kunst	Sport		$\sum$	Erw. Häufigkeiten	
	1	$\geq 2$		1	$\geq 2$
1	19	21	40	17.674	22.326
2	38	64	102	45.070	56.930
$\geq 3$	38	35	73	32.256	40.744
$\sum$	95	120	215		

$PG = 3.998$ ;  $PG \sim \chi_2^2$  mit  $A_{0.10} = [4.61, \infty)$ ;  $A_{0.05} = [5.99, \infty)$ ;  $A_{0.01} = [9.21, \infty)$ .

Die Nullhypothese der Unabhängigkeit kann auf keinem der drei üblichen Signifikanzniveaus verworfen werden. Für den P-Wert gilt  $0.1 < P(PG \geq 3.998) < 0.5$ . Mit `1-pchisq(3.998, 2)` ergibt sich  $\approx 0.1355$

`table(Geschlecht, Mathe):`

	Mathe					Erwartete Häufigkeiten			
Geschlecht	1	2	3	$\geq 4$	$\sum$				
0	18	40	40	26	124	21.026	44.748	36.122	22.104
1	21	43	27	15	106	17.974	38.252	30.878	18.896
$\sum$	39	83	67	41	230				

$PG = 4.431; PG \sim \chi_3^2$  mit  $A_{0.10} = [6.25, \infty)$ ;  $A_{0.05} = [7.81, \infty)$ ;  $A_{0.01} = [11.34, \infty)$ .  $H_0$  kann bei keinem der üblichen Signifikanzniveaus verworfen werden. Für den P-Wert gilt  $0.1 < P(PG \geq 4.431) < 0.5$ . Mit `1-pchisq(4.431, 3)` ergibt sich  $\approx 0.2185$ .

`table(Geschlecht, Sport):`

	Sport				Erwartete Häufigkeiten		
Geschlecht	1	2	$\geq 3$	$\sum$			
0	67	40	14	121	54.853	51.627	14.52
1	35	56	13	104	47.174	44.373	12.48
$\sum$	102	96	27	225			

$PG = 11.525; PG \sim \chi_2^2$  mit  $A_{0.10} = [4.61, \infty)$ ;  $A_{0.05} = [5.99, \infty)$ ;  $A_{0.01} = [9.21, \infty)$ .  $H_0$  kann auf allen drei Signifikanzniveaus verworfen werden. Es ist statistisch abgesichert, dass die beiden Merkmale abhängig sind. Für den P-Wert gilt  $P(PG \geq 11.525) < 0.01$ . Mit `1-pchisq(11.525, 2)` ergibt sich  $\approx 0.0031$ .

# [ 10.15 ]

	Erwartet		Beob. - Erw.		$(\text{Beob.} - \text{Erw.})^2 / \text{Erw.}$	
Qualität \ Geschlecht	Männer	Frauen	Männer	Frauen	Männer	Frauen
sehr gut	58.79	63.21	-3.79	3.79	0.24	0.23
gut	111.80	120.20	8.20	-8.20	0.60	0.56
mäßig	124.82	134.18	5.18	-5.18	0.21	0.20
nur einige Wörter	53.97	58.03	-6.97	6.97	0.90	0.84
gar nicht	23.61	25.39	-2.61	2.61	0.29	0.27

$PG = 4.34$ ;  $PG \sim \chi_4^2$ . Die kritischen Werte sind 7.78, 9.49 bzw. 13.28, d.h.  $H_0$  kann auf keinem der drei Signifikanzniveaus verworfen werden. Der P-Wert ist  $P(PG \geq 4.34) \approx 0.36$ .

[ 10.16 ] a)

	Bild F	Bild E	Bild D	Bild C	Bild B	Bild A
Frau	0.520	3.121	23.409	43.177	26.530	6.242
Mann	0.480	2.879	21.591	39.823	24.470	5.758

b) Nein, sie sollten  $\geq 5$  sein. c) Die erwarteten Häufigkeiten für die neuen zusammengelegten Klassen sind bei Frau: 3.641 bzw. bei Mann: 3.359.

d)  $PG = 1.719$  und  $PG \sim \chi^2$  mit  $(r-1) \cdot (s-1) = (5-1) \cdot (2-1) = 4$  Freiheitsgraden. Damit ist nach Tabelle A4:  $A_{0.1} = [7.78; \infty)$ . Die  $PG$  fällt nicht in den Ablehnungsbereich. Die Nullhypothese der Unabhängigkeit kann nicht verworfen werden. Der P-Wert beträgt  $\approx 0.787$ .

[ 10.17 ]

Den nach der Formel des Buches zu berechnenden Test erhält man mit **R**

```
chisq.test(as.table(matrix(c(6,1,5,17),2,byrow=T)),correct=F)
      Pearson's Chi-squared test
data:  as.table(matrix(c(6, 1, 5, 17), 2, byrow = T))
X-squared = 8.9486, df = 1, p-value = 0.002777
```

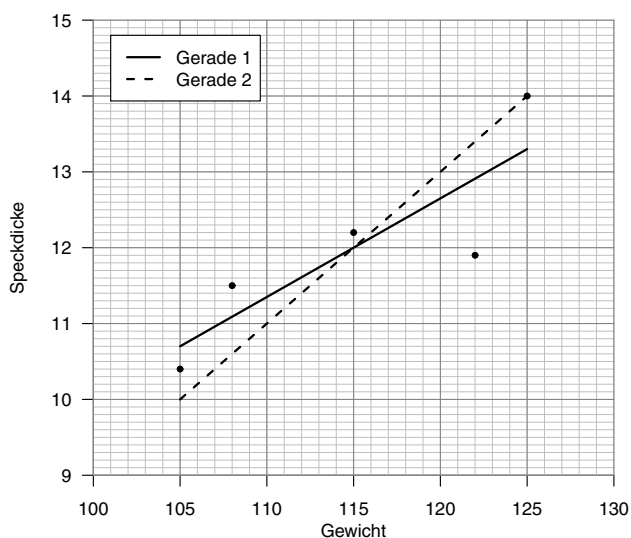
Nun ist  $PG = 8.9486$ ;  $PG \sim \chi_1^2$  und  $A_{0.05} = [3.84; \infty)$ ;  $A_{0.01} = [6.63; \infty)$ . In beiden Fällen wird somit die Nullhypothese verworfen. Der P-Wert ist 0.002777.

Im Fall einer  $2 \times 2$ -Kontingenztafel führt **R** einen Pearson's Chi-squared test with Yates' continuity correction durch, wenn das Argument `correct = F` nicht verwendet wird. Dann ergibt sich ein anderer Wert für  $PG$ .

## 2.11 Einfache Regressionsanalyse - Lösungen

[ 11.1 ]  $\hat{\theta}_1 \approx 17.945$ ,  $\hat{\theta}_2 \approx 0.03$ , d.h.  $y = 17.945 + 0.03x$ .

[ 11.2 ] a) und b)



c) Gerade 1 liegt näher an den Beobachtungen.

d) Gerade 1:

Schwein	1	2	3	4	5	Summe
Gewicht $x$	105	108	115	122	125	575
Speckdicke $y$	10.4	11.5	12.2	11.9	14.0	60.0
$\hat{y}$	10.70	11.09	12.00	12.91	13.30	—
$e_i$	-0.30	0.41	0.20	-1.01	0.70	—
$e_i^2$	0.0900	0.1681	0.0400	1.0201	0.4900	1.8082

Gerade 2:

Schwein	1	2	3	4	5	Summe
Gewicht $x$	105	108	115	122	125	575
Speckdicke $y$	10.4	11.5	12.2	11.9	14.0	60.0
$\hat{y}$	10.0	10.6	12.0	13.4	14.0	—
$e_i$	0.4	0.9	0.2	-1.5	0.0	—
$e_i^2$	0.16	0.81	0.04	2.25	0.00	3.26

Summe der Quadrate der Residuen für Gerade 1: 1.8082, für Gerade 2: 3.26. Somit ist SQ für Gerade 1 kleiner und Gerade 1 die bessere.

e) Beste Gerade nach der Methode der kleinsten Quadrate:  $\hat{y} = -2.973 + 0.130x$ .



[ 11.3 ] a) Die Lösungsgrafik ist in Aufg. 23 verwendet worden.

b)  $\hat{\theta}_2 = \frac{\text{sum}(\text{Kuend} * \text{Ablos}) - \text{sum}(\text{Kuend}) * \text{sum}(\text{Ablos}) / n}{\text{sum}(\text{Kuend} * \text{Kuend}) - \text{sum}(\text{Kuend}) * \text{sum}(\text{Kuend}) / n} \approx 2.132$  und

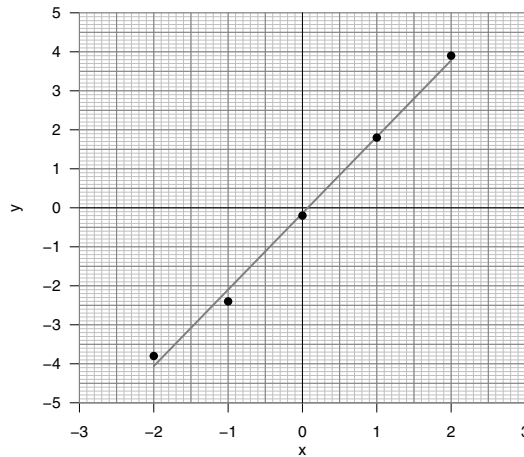
$\hat{\theta}_1 = \text{sum}(\text{Ablos}) / n - \hat{\theta}_2 \text{sum}(\text{Kuend}) / n \approx 1.864.$

c) Den Verlauf der Beobachtungen (oder der Residuen) betrachtend, scheint eine Parabel das bessere Modell zu sein.

[ 11.4 ] a) Gerade durch den Ursprung (0,0). b) Zu minimieren ist die Summe der quadrierten Residuen:  $\sum_{i=1}^n (y_i - \theta x_i)^2$ . Nullsetzen der Ableitung nach  $\theta$  ergibt:  $0 = -2 \sum_{i=1}^n x_i (y_i - \theta x_i) = -2 \sum_{i=1}^n x_i y_i + 2\theta \sum_{i=1}^n x_i^2$ . Damit folgt  $\hat{\theta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$ .

c)  $\hat{\theta} = \frac{19.6}{10} = 1.96$

d)



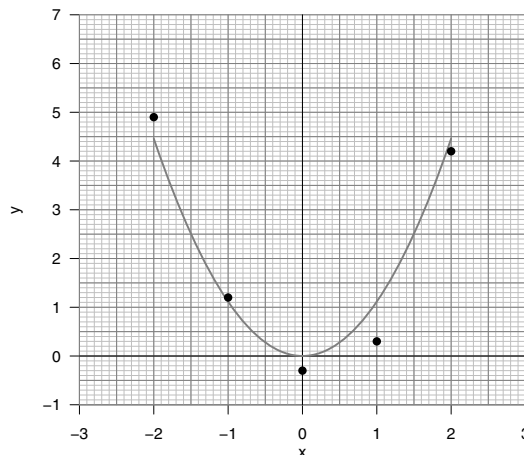
[ 11.5 ] a) Parabel ohne konstanten und linearen Term durch den Ursprung (0,0).

b) Zu minimieren ist die Summe der quadrierten Residuen:  $\sum_{i=1}^n (y_i - \theta x_i^2)^2$ . Nullsetzen der Ableitung nach  $\theta$  ergibt:

$$0 = -2 \sum_{i=1}^n x_i^2 (y_i - \theta x_i^2) = -2 \left( \sum_{i=1}^n x_i^2 y_i - \theta \sum_{i=1}^n x_i^4 \right) \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}.$$

c)  $\hat{\theta} = 37.9/34 \approx 1.115$

d)



[ 11.6 ] a) Parabel ohne konstanten aber mit linearem Term, d.h. Parabel durch (0,0).

b) i) Zu minimieren ist die Summe der quadrierten Residuen. Nullsetzen der partiellen Ableitungen nach  $\theta_1$  und  $\theta_2$  führt zu

$$\sum_{i=1}^n x_i \cdot \theta_1 + \sum_{i=1}^n x_i^3 \cdot \theta_2 = \sum_{i=1}^n x_i y_i$$

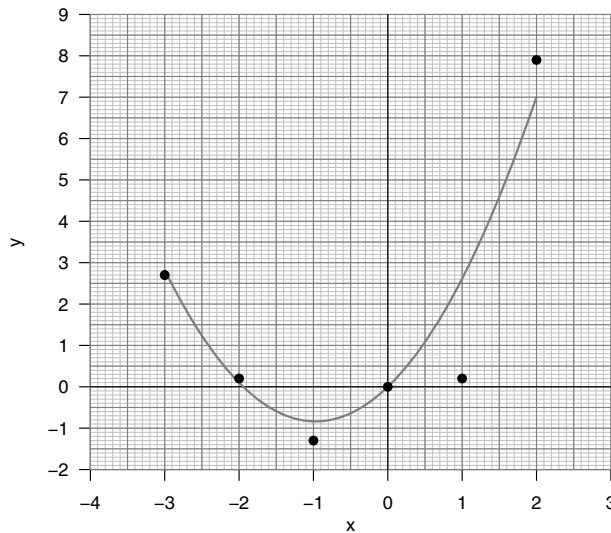
$$\sum_{i=1}^n x_i^3 \cdot \theta_1 + \sum_{i=1}^n x_i^4 \cdot \theta_2 = \sum_{i=1}^n x_i^2 y_i$$

$$\text{ii) } \hat{\theta}_1 = \frac{\sum_{i=1}^n x_i y_i \cdot \sum_{i=1}^n x_i^4 - \sum_{i=1}^n x_i^3 \cdot \sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n x_i^4 - \left( \sum_{i=1}^n x_i^3 \right)^2}$$

$$\hat{\theta}_2 = \frac{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n x_i^2 y_i - \sum_{i=1}^n x_i y_i \cdot \sum_{i=1}^n x_i^3}{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n x_i^4 - \left( \sum_{i=1}^n x_i^3 \right)^2}$$

$$\text{c) } \hat{\theta}_1 = \frac{2513.2}{1456} \approx 1.726 \quad \hat{\theta}_2 = \frac{1294}{1456} \approx 0.889$$

d)



[ 11.7 ] a) Logarithmische Funktion, auch semi-log-Modell genannt, da nur eine Seite der Modellgleichung logarithmiert ist.

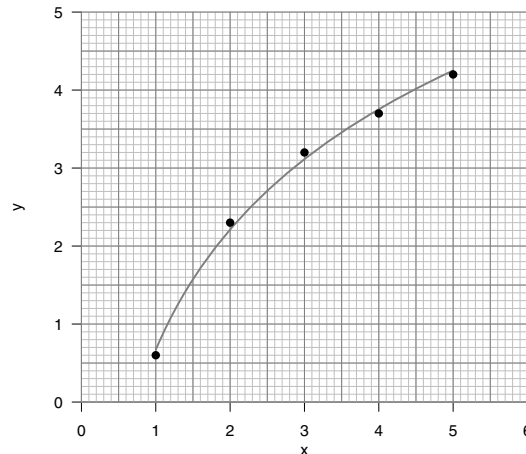
b) Wir ersetzen in den Formeln für die einfache lineare Regression  $x$  durch  $\log(x)$ :

$$\text{Steigung: } \hat{\theta}_2 = \frac{\sum_{i=1}^n (\log(x_i) y_i) - \frac{1}{n} \sum_{i=1}^n \log(x_i) \sum_{i=1}^n y_i}{\sum_{i=1}^n (\log(x_i))^2 - \frac{1}{n} \left( \sum_{i=1}^n \log(x_i) \right)^2} = \frac{\sum_{i=1}^n (\log(x_i) y_i) - n \overline{\log(x)} \bar{y}}{\sum_{i=1}^n (\log(x_i))^2 - n (\overline{\log(x)})^2}$$

$$\text{Achsenabschnitt: } \hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n y_i - \hat{\theta}_2 \cdot \left( \frac{1}{n} \sum_{i=1}^n \log(x_i) \right) = \bar{y} - \hat{\theta}_2 \overline{\log(x)}$$

Beachten Sie die Notation:  $\overline{\log(x)}$  bedeutet  $\frac{\sum_{i=1}^n \log(x_i)}{n}$  analog zu  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ .

c)  $\hat{\theta}_2 \approx 2.225$ ;  $\hat{\theta}_1 \approx 0.670$ .



[ 11.8 ] a) Exponentialfunktion ohne konstanten Term. Es wird ebenfalls semi-log Modell genannt, da man auch  $\log(y) = \log(\theta_1) + \theta_2 x$  schreiben kann, also ein lineares Modell in  $x$  für  $\log(y)$ . b) Wir erhalten die Schätzer für  $\theta_2$  und  $\log(\theta_1)$ , indem wir in den Formeln für die einfache lineare Regression  $y$  durch  $\log(y)$  ersetzen.

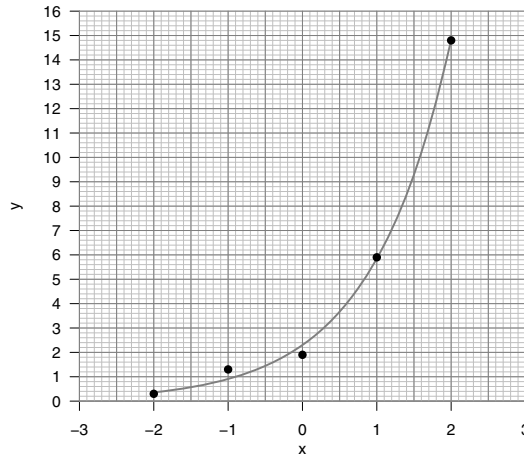
$$\text{Steigung: } \hat{\theta}_2 = \frac{\sum_{i=1}^n (x_i \log(y_i)) - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n \log(y_i)}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i \log(y_i)) - n \bar{x} \overline{\log(y)}}{\sum_{i=1}^n (x_i)^2 - n \bar{x}^2}$$

$$\text{Achsenabschnitt: } \widehat{\log(\theta_1)} = \frac{1}{n} \sum_{i=1}^n \log(y_i) - \hat{\theta}_2 \cdot \left( \frac{1}{n} \sum_{i=1}^n x_i \right) = \overline{\log y} - \hat{\theta}_2 \bar{x}$$

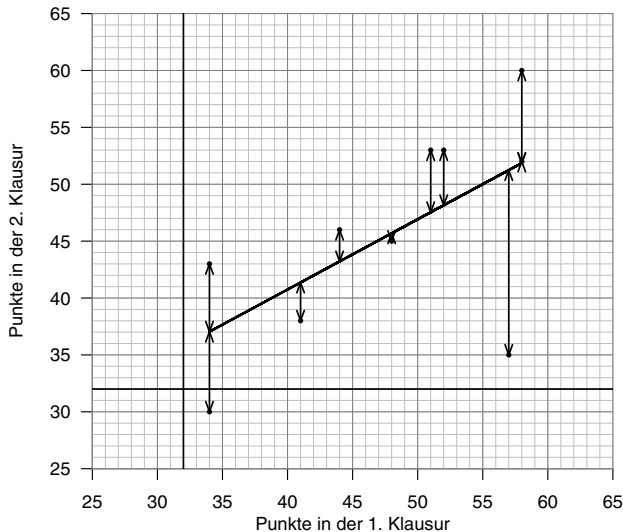
Daraus folgt:<sup>10</sup>  $\hat{\theta}_1 = \exp(\overline{\log y} - \hat{\theta}_2 \bar{x})$

c) Steigung:  $\hat{\theta}_2 \approx 0.931$ ; Achsenabschnitt:  $\widehat{\log(\theta_1)} \approx 0.834$ ;  $\hat{\theta}_1 \approx 2.302$

<sup>10</sup> Für die Notation  $\overline{\log(y)}$  siehe Lösung zu Aufg. 7.



[ 11.9 ] a) Konstante:  $\hat{\theta}_1 \approx 16.051$ ; Steigung  $\hat{\theta}_2 \approx 0.617$  (siehe auch **R**-Programm)



c) Vorhersage:  $16.051 + 0.617 \cdot 40 = 40.731$

d) Der geschätzte Standardfehler der Vorhersage ist  $\sqrt{\hat{\sigma}^2 \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right)} \approx$

8.476. Beachten Sie: Die Punkte  $\hat{y}_i$  auf der Geraden (zur Berechnung von  $\hat{\sigma}^2$ ) wurden mit den oben angegeben gerundeten Werten für  $\hat{\theta}_1$  und  $\hat{\theta}_2$  berechnet. Danach wurde erst ganz zum Schluss gerundet.

e) Mit dem Quantil  $t_{8,0.05} = 1.86$  und obigen Resultaten erhalten wir das Prognoseintervall  $[40.731 - 8.476 \cdot 1.86; 40.731 + 8.476 \cdot 1.86] \approx [24.97; 56.50]$ .

[ 11.10 ] a) Die Gerade wurde nach der Methode der kleinsten Quadrate angepasst, d.h. die Summe der Quadrate der Residuen (das sind die senkrechten Abstände der

Punkte von der Geraden) wird minimiert. Optisch allerdings hat man den Eindruck, als müsste die Gerade steiler sein, d.h. die Steigung müsste größer sein.

b) Ein Problem der Regression, das den Widerspruch zwischen bester Anpassung nach der Methode der kleinsten Quadrate einerseits und dem optischen Eindruck andererseits etwas erklären kann, ist die Tatsache, dass die Beobachtungen sowohl nach unten als auch nach oben zensiert (beschränkt) sind, d.h. die  $x$ - und  $y$ -Werte sind nach unten durch 0 und nach oben durch 18 beschränkt. Man sieht an der linken Abbildung, dass die Streuung um die Gerade mit wachsendem  $x$  abnimmt.

c) Nein. Mit zunehmendem  $x$  nimmt die Anzahl der Punkte oberhalb der waagerechten Gerade ab, d.h. die Residuen sind nicht zufällig um 0 herum verteilt.

d) Das Histogramm deutet auf eine leichte Linksschiefe. Die rechte Grafik zeigt eine klare Veränderung der Streuung mit wachsenden  $x$ -Werten (Werten im Eingangstest). Man spricht von Heteroskedastizität. Hier nimmt mit zunehmender Punktzahl im Eingangstest die Residuenstreuung ab. Auch dies ist typisch für von unten zensierte Daten (keine Werte kleiner Null in unserem Fall).

e) i) Steigung:  $\approx 0.49$ ; Achsenabschnitt:  $\approx 11.91$ . Dies stimmt mit der Grafik überein.

ii) Residualvarianz ist das Quadrat des Residual standard error, d.h.  $2.613^2 \approx 6.83$ .

iii) Den Residual standard error kann man am ehesten vom Histogramm ablesen. Da die Residuenverteilung uns leicht an eine zentrierte Normalverteilung erinnert, sollten etwa 68% der Fläche des Histogramms im Intervall  $[-\sigma; \sigma]$  bzw. 95% im Intervall  $[-2\sigma; 2\sigma]$  liegen.

iv) Der Prozentsatz der Variation, der durch die Regression erklärt wird, ist  $R^2 = 0.1933$  (siehe Multiple R-Squared: 0.1933).

v) Bei einer einfachen linearen Regression entspricht  $R^2$  dem Quadrat des Korrelationskoeffizienten zwischen den beiden Variablen, d.h.  $\hat{\rho} = +\sqrt{0.1933} \approx 0.4397$ . Beachten Sie, dass die Korrelation wegen der steigenden Geraden positiv ist.

vi) Die  $F$ -statistic prüft im Falle der einfachen linearen Regression die Nullhypothese, dass der Koeffizient der erklärenden Variablen Eingangstest Null ist. Er ist äquivalent zum zweiseitigen  $t$ -Test für Eingangstest.

vii)  $H_0$  wird für große Werte verworfen, d.h. der  $P$ -Wert ist  $P(PG \geq 17.49)$ . Somit ist  $P(PG \leq 17.49) = 1 - 0.00007935 \approx 1$ .

[ 11.11 ] Siehe R-Programm, Abweichungen sind wegen Rundungen möglich.

a) Vorhersage für Lesen:  $\hat{y} = 76.8329 + 0.8282 \cdot 510 \approx 499$

für Naturwissenschaften:  $\hat{y} = -31.42409 + 1.08915 \cdot 510 \approx 524$

b)  $\widehat{SF}(\text{Vorhersage}) = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n s_x^2 - n(\bar{x})^2}} =$

$$5.178 \sqrt{1 + \frac{1}{16} + \frac{(510 - 8007/16)^2}{4009187 - 16 \cdot (8007/16)^2}} \approx 5.442 \text{ für Lesen}$$

$$4.433 \sqrt{1 + \frac{1}{16} + \frac{(510 - 8007/16)^2}{4009187 - 16 \cdot (8007/16)^2}} \approx 4.659 \text{ für Naturwissenschaften}$$

c) Prognoseintervalle:  $1 - \alpha = 0.95$ :

Lesen:  $[499.2149 - 5.441509 \cdot 2.14; 499.2149 + 5.441509 \cdot 2.14] \approx [488; 511]$

Naturwiss.:  $[524.0424 - 4.658596 \cdot 2.14; 524.0424 + 4.658596 \cdot 2.14] \approx [514; 534]$

$1 - \alpha = 0.99$ :

Lesen:  $[499.2149 - 5.441509 \cdot 2.98; 499.2149 + 5.441509 \cdot 2.98] \approx [483; 515]$

Naturwiss.:  $[524.0424 - 4.658596 \cdot 2.98; 524.0424 + 4.658596 \cdot 2.98] \approx [510; 538]$ .

[ 11.12 ] Für die meisten Aufgabenteile und Grafiken siehe auch **R**-Programm.

a)  $\hat{y} \approx 122.043 - 10.218x$  b) Residuen:  $y_i - \hat{y}_i = 1.775; -2.307; -0.989; 0.229; 2.447; -0.735; -0.417$ . Das ergibt  $SQ(Res) = \sum_{i=1}^7 (y_i - \hat{y}_i)^2 \approx 16.205$ .

c)  $SQ(Total) = \sum_{i=1}^7 (y_i - \bar{y})^2 = 2939.534$ .

d)  $SQ(Regression) = \sum_{i=1}^7 (\hat{y}_i - \bar{y})^2 = SQ(Total) - SQ(Res) = 2939.534 - 16.205 = 2923.329$ .

e)  $R^2 = SQ(Regression)/SQ(Total) = 2923.329/2939.534 \approx 0.994$ . Damit werden 99.4% der Variation durch die Regression erklärt. 0.6% bleiben unerklärt.

f)  $|\hat{\rho}| = \sqrt{R^2} \approx 0.997$ . Da die Korrelation hier eindeutig negativ ist (siehe Abbildung), beträgt  $\hat{\rho} \approx -0.997$ .

g) Nein, da wir keine kausal erklärende Variable haben, sondern nur die Zeit, obwohl man manchmal Voraussagen in die Zukunft unter der Annahme macht, dass sich der gegenwärtige Trend in die Zukunft fortsetzt.

h)  $\widehat{\text{Var}}(\hat{\theta}_1) = \frac{\hat{\sigma}^2}{n} \left( 1 + \frac{(\bar{x})^2}{\hat{\sigma}_x^2} \right)$   $\widehat{\text{Var}}(\hat{\theta}_2) = \frac{\hat{\sigma}^2}{n\hat{\sigma}_x^2}$  wobei  $\hat{\sigma}^2 = SQ(Res)/(n-2) = 16.205/5 = 3.241$ ,  $\bar{x} = \frac{1}{7} \sum_{i=1}^7 i = 4$  und  $\hat{\sigma}_x^2 = \frac{1}{7} \sum_{i=1}^7 i^2 - (\bar{x})^2 = 20 - 16 = 4$ .

$\widehat{\text{Var}}(\hat{\theta}_1) = \frac{3.241}{7} \left( 1 + \frac{16}{4} \right) = 2.315$  und  $\widehat{\text{Var}}(\hat{\theta}_2) = \frac{3.241}{7 \cdot 4} \approx 0.116$ .

i) Die Konfidenzintervalle sind  $[\hat{\theta}_i - t_{n-2, \alpha/2} \widehat{SE}(\hat{\theta}_i), \hat{\theta}_i + t_{n-2, \alpha/2} \widehat{SE}(\hat{\theta}_i)]$ , wobei die hier vorkommenden geschätzten Standardfehler die Quadratwurzeln aus den in h) berechneten geschätzten Varianzen sind. Es ist  $t_{5, 0.05} = 2.02$ . Damit sind die Konfidenzintervalle für

$\theta_1$ :  $[122.043 - 2.02 \cdot \sqrt{2.315}, 122.043 + 2.02 \cdot \sqrt{2.315}] \approx [119.0, 125.1]$

$\theta_2$ :  $[-10.218 - 2.02 \cdot \sqrt{0.116}, -10.218 + 2.02 \cdot \sqrt{0.116}] \approx [-10.9, -9.5]$

[ 11.13 ] a)

Modell	FG	SQ	DQ	F
Differenz	1	263.4	263.400	7.7515
$M_1$	90	3058.8	33.987	
$M_2$	91	3322.2		

b) Die Prüfgröße ist F-verteilt mit 1 und 90 Freiheitsgraden.

c) Der P-Wert ist kleiner als  $0.006541 < 0.01$ , so dass die Nullhypothese (Modell der Geraden) verworfen wird. Damit ist die Parabel besser zur Beschreibung der Daten geeignet.

[ 11.14 ] a)  $\hat{y} = -131 + 127.2x$ . Die Grafik mit der eingezeichneten Geraden wird in Aufg. 24 verwendet. Die Residuen sind dann:  $y_i - \hat{y}_i = 19.8; -14.4; -30.6; 25.2$ . Die Summe der Quadrate ist  $SQ(Res) = 2170.8$ .

b) Für die Parabel ist  $SQ(Res) = 145.8$ .

c)

Modell	FG	SQ	DQ	F
Differenz	1	2025.0	2025.0	13.889
$M_1$ : Parabel	1	145.8	145.8	
$M_2$ Gerade	2	2170.8		

d) Mit Tabelle A5 oder  $qf(0.9, 1, 1)$  ergibt sich als Ablehnungsbereich  $A = [39.86, \infty)$ . Die Nullhypothese kann nicht verworfen werden, d.h. man würde das Modell der Geraden verwenden. Der P-Wert lässt sich mit  $1 - pf(13.889, 1, 1)$  berechnen und ist  $\approx 0.167$ .

[ 11.15 ] a)

Modell	FG	FG-DQ=SQ	DQ	F
Differenz	2	49.14	24.57	5.571
$M_1$	12	52.92	$2.1^2 = 4.41$	
$M_2$	14	102.06	$2.7^2 = 7.29$	

b)  $\alpha = 0.05 \Rightarrow A = [3.89, \infty)$ , d.h. die Nullhypothese wird verworfen und man würde die kubische Funktion verwenden.  $\alpha = 0.01 \Rightarrow A = (6.93, \infty)$ , d.h. die Nullhypothese kann nicht verworfen werden und man würde die Gerade verwenden.

c) Der P-Wert liegt zwischen 0.01 und 0.05 Er kann mit  $1 - pf(5.571, 2, 12)$  in **R** berechnet werden.

[ 11.16 ] a)  $\hat{y} = -110.919 + 1.026 \cdot x$ , wobei  $x$  die Körpergröße und  $y$  das Gewicht ist.

b) Einsetzen der Werte für  $x$  in die Formel unter a) ergibt 73.761 bzw. 84.021.

c) Die  $1 - \alpha$  Vorhersageintervalle sind  $[\hat{y}_0 - \widehat{SF} \cdot t_{n-p;\alpha/2}; \hat{y}_0 + \widehat{SF} \cdot t_{n-p;\alpha/2}]$ , wobei  $p$  die Anzahl der geschätzten Parameter und  $\widehat{SF}^2 = \hat{\sigma}^2 \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right)$ .

Mit  $\alpha = 0.05$ ,  $\widehat{SF} \approx 9.70$  und  $t_{123;0.025} = 1.98$  ergibt sich etwa  $[54.50, 92.92]$  bzw.  $[64.75, 103.18]$  (bis auf Rundungsfehler).

d) Aus der **R**-Ausgabe liest man (gerundet) ab:  $\hat{y} = 486.334 - 5.575x + 0.018x^2$

e)

Modell	FG	SQ	DQ	F
Differenz	1	454.47	454.47	5.025
$M_1$	122	11033.69	90.44	
$M_2$	123	11488.16		

wobei  $SQ(M_2)$  in der Aufgabe gegeben war,  $SQ(M_1) = DF \cdot (Res. stand. error)^2 = 122 \cdot 9.51^2 = 11033.69$  und der Rest sich daraus errechnen lässt.

i) Der Ablehnungsbereich ist rechtsseitig. Die untere Grenze kann in **R** mit dem Befehl `qf(0.95, 1, 122)` berechnet werden. In der **R**-Ausgabe findet man dafür den Wert  $F_{1,122}(0.95) = 3.92$ . Damit ist  $A_{0.05} = [3.92; \infty)$ .

ii) Da die Prüfgröße im Ablehnungsbereich liegt, wird die Nullhypothese bei  $\alpha = 0.05$  verworfen und es gilt als statistisch abgesichert, dass die Parabel besser zur Beschreibung der Daten geeignet ist. Der P-Wert liegt zwischen 0.02 und 0.03.

[ 11.17 ]

Modell	FG	SQ	DQ	F
Differenz	1	513.1	513.1	5.7228
$M_1$	121	10 848.4	10 848.4/121	
$M_2$	122	11 361.5		

[ 11.18 ] a) Die steigende Gerade.

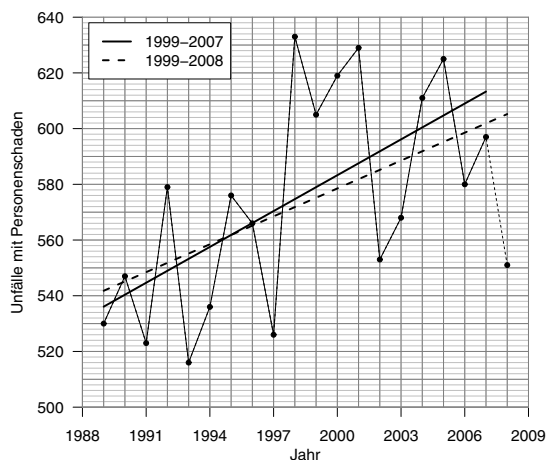
b) i)  $\hat{y} = -7995 + 4.289x$ 

ii) Es wird die Nullhypothese geprüft, dass der Achsenabschnitt bzw. die Steigung der Geraden Null ist. Die PG ist  $t$ -verteilt mit  $n - 2 = 17$  Freiheitsgraden.

iii) Es wird die Nullhypothese geprüft, dass die Steigung der Geraden Null ist, d.h. dass das Modell der waagerechten Geraden gilt. Alternative: Das Modell der nicht waagerechten Geraden gilt. Die F-Prüfgröße ist das Quadrat der zweiten  $t$ -Prüfgröße. Beachten Sie die Übereinstimmung der P-Werte. Die Behauptung der Grünen ist bei allen gängigen Signifikanzniveaus abzusichern. Der P-Wert ist 0.004329, so dass die Nullhypothese sogar bei  $\alpha = 0.005$  verworfen wird. Die Irrtumswahrscheinlichkeit ist gleich dem P-Wert 0.004329.

iv) 38.89% ergibt sich aus Multiple R-Squared: 0.3889.

c) i)





ii) Die gestrichelte Gerade in der obigen Abbildung ist flacher als die bisherige Gerade.

iii) 0.019 oder 1.9%; 26.95%. Die  $t$ -Prüfgrößen haben jetzt 18 Freiheitsgrade, die  $F$ -Prüfgröße hat 1 und 18 Freiheitsgrade.

[ 11.19 ] a) W b) W c) F d) W e) W f) W g) W h) F

[ 11.20 ] a) W b) W c) W d) F e) W f) W g) W h) F

Modell	FG	SQ	DQ	F
Differenz	2	40	20	2
$M_1$	20	200	10	
$M_2$	22	240		

[ 11.21 ] a) F b) F c) W d) W e) F

[ 11.22 ] a) W b) W c) F d) W e) F

[ 11.23 ] a) W b) W c) W d) W e) F

[ 11.24 ] a) F b) W c) F d) W e) F f) W g) W

## 2.12 Varianzanalyse - Lösungen

[ 12.1 ] a) W b) W c) W d) F e) W f) W g) W

[ 12.2 ]

Modell	FG	SQ	DQ	F
Differenz	4	40	10	1
$M_1$	20	200	10	
$M_2$	24	240		

a) W b) F c) W d) W e) F f) W g) W

[ 12.3 ] a) W b) W c) F d) F e) W

[ 12.4 ] a) W b) F c) F d) W e) F f) W g) W h) F i) F

[ 12.5 ] a)  $\hat{\mu}_1 = \bar{y}_1 = 12$ ;  $\hat{\mu}_2 = \bar{y}_2 = 11$ ;  $\hat{\mu}_1 = \bar{y}_1 = 13$  und  $\hat{\mu} = \bar{y}_{..} = 12$ .

b)

Modell	$FG$	$SQ$	$DQ$	$F$
Differenz	2	8	4	3
$M_1$	9	12	4/3	
$M_2$	11	20		

c)  $A = (4.26, \infty)$ . d) Man würde  $M_2$  verwenden, da die Nullhypothese (keine signifikanten Gruppenunterschiede) nicht verworfen wird. e) Er liegt leicht über 0.1, denn laut Tabelle ist der kritische Wert für  $\alpha = 0.1$  gleich 3.01.

[ 12.6 ] a) Mean Sq:  $126/3 = 42$  und  $2800/100 = 28$ . F value:  $42/28 = 1.5$ .

b) Der Wert der Verteilungsfunktion für 1.5 ist 0.78. Der P-Wert ist  $1 - 0.78 = 0.22$ .

c) Die Nullhypothese kann nicht verworfen werden, wenn  $\alpha = 0.1$ , d.h. man würde sich für das einfachere Modell  $M_2 : y_{ij} = \mu + e_{ij}$  mit einem gemeinsamen Erwartungswert  $\mu$  entscheiden.

d)

Modell	$FG$	$SQ$	$DQ$	$F$
Differenz	3	126	42	1.5
$M_1$	100	2800	28	
$M_2$	103	2926		

[ 12.7 ] Mit den folgenden Berechnungen in R

```
>be<-rep(c(1:5), 6)
>MM<-c(5849, 4916, 5518, 5535, 5200, 6830, 4305, 5031, 5659, 5545,
        5233, 5018, 5253, 4853, 6114, 5752, 4599, 4656, 5020, 5442,
        6332, 5935, 6059, 5691, 5940, 5653, 4936, 4939, 5564, 6356)
>anova(aov(MM ~ factor(be)))
```

Response: MM

```
      Df Sum Sq Mean Sq F value Pr(>F)
factor(be)  4 3820142   955036   4.0306 0.01173 *
Residuals 25 5923647   236946
```

ergibt sich für die Tabelle

Modell	FG	SQ	DQ	F
Differenz	4	3820142	955036	4.0306
$M_1$	25	5923647	236946	
$M_2$	29	9743789		

Wir erhalten als Wert der Prüfgröße  $F \approx 4.031$  und einen P-Wert von 0.01173 (F-Verteilung mit 4 und 25 Freiheitsgraden;  $1 - \text{pf}(4.0306, 4, 25)$ ), d.h. für jedes Signifikanzniveau  $\alpha > 0.01173$  (entsprechend 1.173%) schließen wir, dass der Betrieb einen Einfluss auf die Milchmenge hat.

[ 12.8 ]

Modell	FG	SQ	DQ	F
Differenz	3	115 533 032	38 511 011	37.93842
$M_1$	206	209 109 102	1 015 093	
$M_2$	209	324 642 134		

Unter der Nullhypothese ist die Teststatistik F-verteilt mit 3 und 206 Freiheitsgraden. Der **R**-Befehl  $1 - \text{pf}(37.93842, 3, 206)$  ergibt 0, d.h. der P-Wert ist 0. Die Nullhypothese wird verworfen.

[ 12.9 ] a) i)

Behandlung	1	2	3	4	5	$y_i$	$\bar{y}_i$
Unbehandelt	34	36	31	31	18	150	30
Präparat 1	33	38	45	35	22	173	34.6
Präparat 2	37	38	42	32	23	172	34.4
Präparat 3	38	36	38	32	30	174	34.8
Präparat 4	35	31	41	42	26	175	35

$$\bar{y}_{..} = 844/25 = 33.76$$

ii) Residuen für Modell 1:

Behandl.	1	2	3	4	5
Unbeh.	4	6	1	1	-12
Präparat 1	-1.6	3.4	10.4	0.4	-12.6
Präparat 2	2.6	3.6	7.6	-2.4	-11.4
Präparat 3	3.2	1.2	3.2	-2.8	-4.8
Präparat 4	0	-4	6	7	-9

iii) Residuen für Modell 2:

1	2	3	4	5
0.24	2.24	-2.76	-2.76	-15.76
-0.76	4.24	11.24	1.24	-11.76
3.24	4.24	8.24	-1.76	-10.76
4.24	2.24	4.24	-1.76	-3.76
1.24	-2.76	7.24	8.24	-7.76

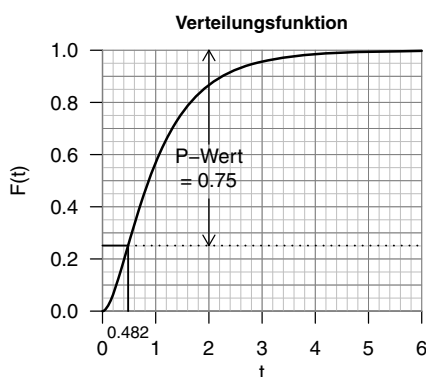
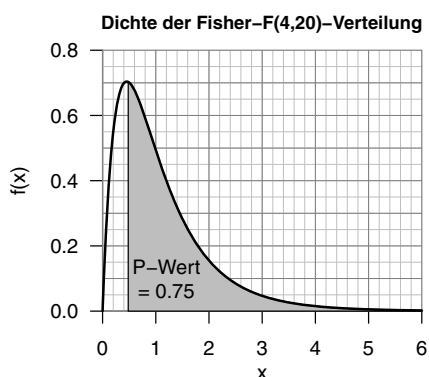
- iv)  $SQ(Res; M_1) = 927.2$ ;  $SQ(Res; M_2) = 1016.56$ . FG: 20 für  $M_1$  und 24 für  $M_2$   
 v)

	FG	$SQ(Res)$	$DQ(Res)$	$F(\text{oder } PG)$
Differenz	4	89.36	22.34	0.481881
Modell 1	20	927.2	46.36	
Modell 2	24	1016.56		

- b)  $F$ -verteilt mit 4 und 20 Freiheitsgraden.  $\alpha = 0.05 \Rightarrow A = [2.87, \infty)$  und  $\alpha = 0.01 \Rightarrow A = [4.43, \infty)$ . Die Nullhypothese kann nicht verworfen werden, d.h. es kann kein Unterschied in den Behandlungen nachgewiesen werden.

$$c) F = PG = \frac{\frac{1}{4} \sum_{i=1}^I 5(\bar{y}_i - \bar{y}_{..})^2}{\frac{1}{20} 927.2} =$$

$$\frac{1.2 \cdot ((30 - 33.76)^2 + (34.6 - 33.76)^2 + (34.4 - 33.76)^2 + (34.8 - 33.76)^2 + (35 - 33.76)^2)}{46.36} = 0.482$$



- [ 12.10 ] a) i)  $\bar{y}_1 = 27.75$   $\bar{y}_2 = 20.4$   $\bar{y}_3 = 20.6$   $\bar{y}_4 = 23.9$   $\bar{y}_i = 23.381$   
 ii) Residuen Modell 1:  $y_{ij} - \bar{y}_{..}$

Sem./Kand.	1	2	3	4	5	6
3.	-0.75	5.75	7.25	-3.25	-4.25	-4.75
4.	6.6	-4.9	-3.9	0.6	1.6	—
5.	-4.6	3.4	-2.1	-2.6	5.9	—
6.;7.	-4.4	1.6	3.1	1.1	-1.4	—

Residuen Modell 2:  $y_{ij} - \bar{y}_{..} = y_{ij} - 23.381$

Sem./Kand.	1	2	3	4	5	6
3.	3.619	10.119	11.619	1.119	0.119	-0.381
4.	3.619	-7.881	-6.881	-2.381	-1.381	-
5.	-7.381	0.619	-4.881	-5.381	3.119	-
6. oder 7.	-3.881	2.119	3.619	1.619	-0.881	-

$$\text{iii) } SQ(Res; M_1) = 336.475 \quad SQ(Res; M_2) = 535.4524 \approx 535.452$$

$$FG(M_1) = 21 - 4 = 17 \quad FG(M_2) = 21 - 1 = 20$$

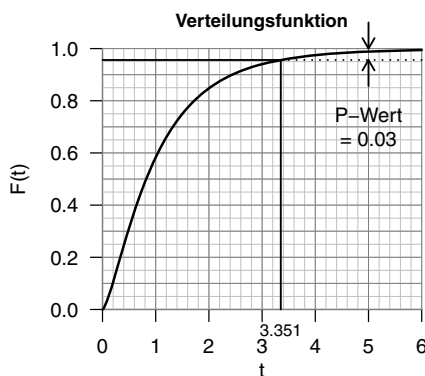
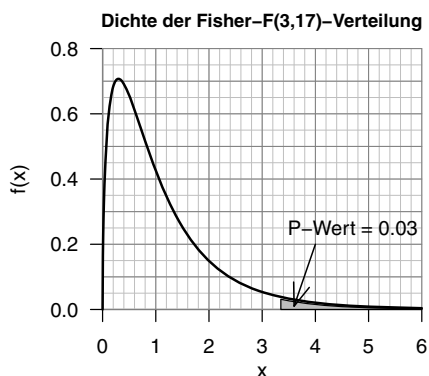
iv)

	FG	$SQ(Res)$	$DQ(Res)$	$F(\text{oder } PG)$
Differenz	3	198.977	66.32567	3.351
Modell 1	17	336.475	19.79265	
Modell 2	20	535.452		

b)  $PG \sim F_{3;17}$  Ablehnungsbereich:  $[3.20; \infty)$  bzw.  $[5.18; \infty)$

Bei einem Signifikanzniveau von 5% fällt die Prüfgröße in den Ablehnungsbereich und es gilt als statistisch abgesichert, dass ein Unterschied zwischen den vier Gruppen besteht. Bei einem Signifikanzniveau von 1% kann die Nullhypothese ( $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$ ) nicht verworfen werden, da die Prüfgröße außerhalb des Ablehnungsbereichs liegt.

$$\text{c) } PG = \frac{\frac{1}{3} (6(27.75 - 23.381)^2 + 5(20.4 - 23.381)^2 + 5(20.6 - 23.381)^2 + 5(23.9 - 23.381)^2)}{\frac{1}{17} 336.475} = 3.351$$



[ 12.11 ]

- a) Es gibt deutliche Unterschiede, die Gruppen 2, 3, 7 und 9 liegen deutlich tiefer.  
 b) Es wird die Nullhypothese geprüft, dass die Mittelwerte in den Gruppen gleich sind:  $\mu_1 = \mu_2 = \dots = \mu_{10}$

- c) Bei  $\alpha = 0.0005$  wird die Nullhypothese abgelehnt, da der P-Wert kleiner als 0.0005 ist. Bei  $\alpha = 0.0001$  kann die Nullhypothese nicht verworfen werden, weil der P-Wert größer als 0.0001 ist.
- d) Dies entspricht dem P-Wert, also 0.0003661.
- e) Der P-Wert ist die Fläche rechts von 4.0574. Es ist nichts mehr zu erkennen, deshalb zeigen wir hier keine Abbildung.
- f)

	<i>FG</i>	<i>SQ(Res)</i>	<i>DQ(Res)</i>	<i>F(oder PG)</i>
Differenz	9	226.14	25.13	4.0574
Modell 1	65	402.53	6.19	
Modell 2	74	628.67		

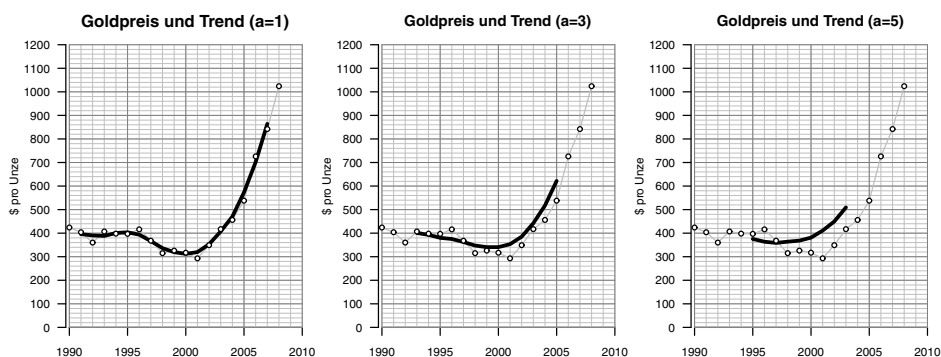
## 2.13 Zeitreihen und Indizes - Lösungen

[ 13.1 ] a) W b) F c) W d) W e) W f) F

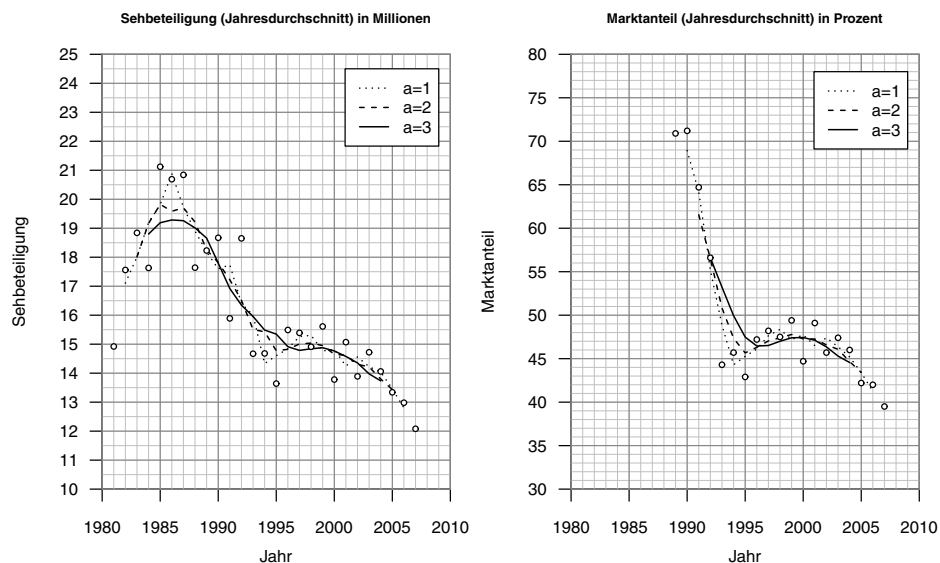
[ 13.2 ] a) W b) F c) W d) W e) W

[ 13.3 ]  
a) und b)

Jahr	Höchstpreis	a = 1	a = 3	a = 5
1990	424			
1991	404	396.000		
1992	360	390.333		
1993	407	388.333	400.857	
1994	398	400.667	392.857	
1995	397	403.667	380.143	375.636
1996	416	393.667	375.286	363.727
1997	368	366.333	362.429	358.727
1998	315	336.333	347.429	363.909
1999	326	319.333	340.571	368.364
2000	317	312.000	340.714	381.091
2001	293	319.667	353.286	411.000
2002	349	353.000	385.143	449.727
2003	417	407.333	442.286	509.364
2004	456	470.333	517.286	
2005	538	573.333	621.714	
2006	726	702.000		
2007	842	864.000		
2008	1024			



[ 13.4 ] Siehe auch **R**-Programm und beachten Sie, dass es Ungenauigkeiten wegen Rundungen geben kann.

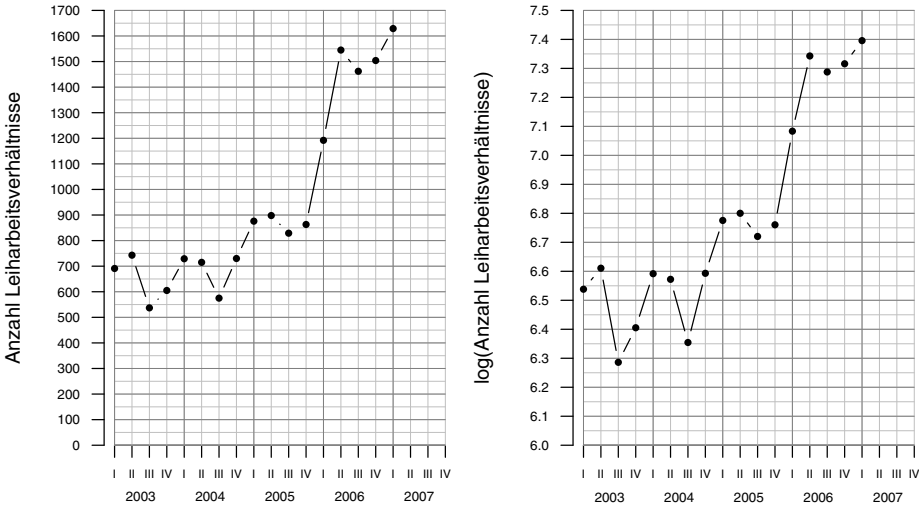


Jahr	Sehb.	$a = 1$	$a = 2$	$a = 3$	Markta.	$a = 1$	$a = 2$	$a = 3$
1981	14.92							
1982	17.56	17.11						
1983	18.84	18.01	18.01					
1984	17.63	19.20	19.17	18.80				
1985	21.12	19.81	19.82	19.19				
1986	20.69	20.88	19.58	19.28				
1987	20.84	19.72	19.70	19.26				
1988	17.64	18.90	19.21	19.01				
1989	18.23	18.18	18.25	18.66	70.9			
1990	18.67	17.60	17.82	17.80	71.2	68.93		
1991	15.89	17.74	17.22	16.92	64.7	64.17	61.54	
1992	18.65	16.40	16.51	16.35	56.6	55.20	56.50	56.61

1993	14.67	16.00	15.51	15.96	44.3	48.87	50.84	53.23
1994	14.68	14.33	15.43	15.49	45.7	44.30	47.34	49.94
1995	13.64	14.60	14.77	15.35	42.9	45.27	45.66	47.49
1996	15.49	14.84	14.82	14.91	47.2	46.10	46.30	46.46
1997	15.39	15.26	15.01	14.79	48.2	47.63	47.04	46.51
1998	14.91	15.30	15.04	14.84	47.5	48.37	47.40	47.00
1999	15.61	14.77	14.95	14.88	49.4	47.20	47.78	47.40
2000	13.78	14.82	14.65	14.77	44.7	47.73	47.28	47.43
2001	15.07	14.25	14.61	14.58	49.1	46.50	47.26	47.11
2002	13.89	14.56	14.30	14.35	45.7	47.40	46.58	46.36
2003	14.72	14.22	14.22	13.98	47.4	46.37	46.08	45.30
2004	14.06	14.04	13.80	13.73	46.0	45.20	44.66	44.56
2005	13.34	13.46	13.44		42.2	43.40	43.42	
2006	12.98	12.80			42.0	41.23		
2007	12.08				39.5			

[ 13.5 ]  $T_t = \frac{1}{24} \left( y_{t-6} + 2 \sum_{j=-5}^5 y_{t+j} + y_{t+6} \right)$  ergibt für die neuen Werte 90.7 und 89.1.

[ 13.6 ] a) Siehe auch **R**-Programm und beachten Sie dabei, dass es Ungenauigkeiten wegen Rundungen geben kann.





b) additives Modell:

Quartal	Daten	Trend	Daten-Trend	Saisonkomponente	Residuum
<i>II</i> 03	691				
<i>III</i>	743				
<i>IV</i>	537	648.750	-111.750	-84.438	-27.312
<i>I</i> 04	605	650.000	-45.000	-75.083	30.083
<i>II</i>	729	651.250	77.750	52.625	25.125
<i>III</i>	715	671.625	43.375	96.958	-53.583
<i>IV</i>	575	705.625	-130.625	-84.438	-46.187
<i>I</i> 05	730	746.875	-16.875	-75.083	58.208
<i>II</i>	876	801.500	74.500	52.625	21.875
<i>III</i>	898	849.875	48.125	96.958	-48.833
<i>IV</i>	829	906.000	-77.000	-84.438	7.438
<i>I</i> 06	863	1026.375	-163.375	-75.083	-88.292
<i>II</i>	1192	1186.375	5.625	52.625	-47.000
<i>III</i>	1545	1345.625	199.375	96.958	102.417
<i>IV</i>	1462	1480.375	-18.375	-84.438	66.063
<i>I</i> 07	1504				
<i>II</i>	1629				

multiplikatives Modell:

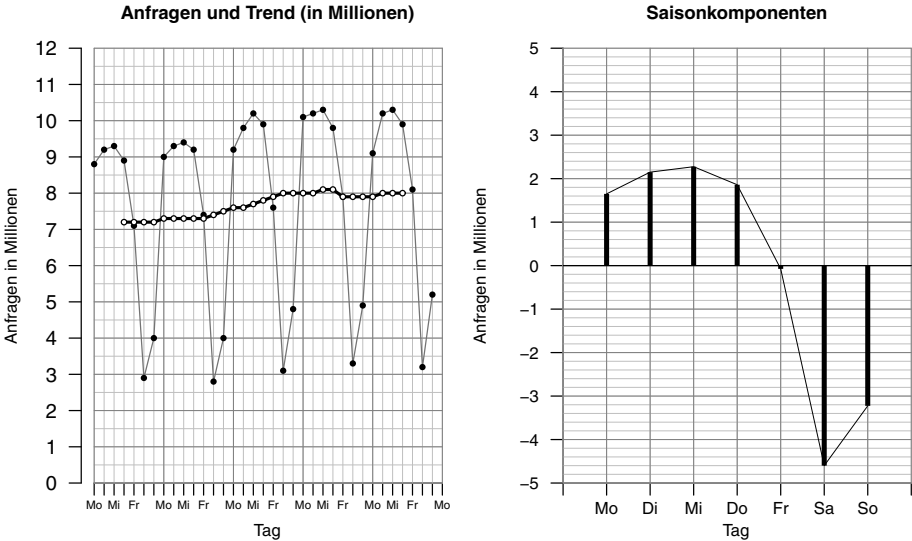
Quartal	log(Daten)	Trend	log(Daten)-Trend	Saisonkomponente	Residuum
<i>II</i> 03	6.54				
<i>III</i>	6.61				
<i>IV</i>	6.29	6.469	-0.179	-0.117	-0.062
<i>I</i> 04	6.41	6.470	-0.060	-0.073	0.013
<i>II</i>	6.59	6.473	0.117	0.085	0.032
<i>III</i>	6.57	6.502	0.068	0.092	-0.024
<i>IV</i>	6.35	6.549	-0.199	-0.117	-0.082
<i>I</i> 05	6.59	6.601	-0.011	-0.073	0.062
<i>II</i>	6.78	6.676	0.104	0.085	0.019
<i>III</i>	6.80	6.744	0.056	0.092	-0.036
<i>IV</i>	6.72	6.803	-0.083	-0.117	0.034
<i>I</i> 06	6.76	6.907	-0.147	-0.073	-0.074
<i>II</i>	7.08	7.046	0.034	0.085	-0.051
<i>III</i>	7.34	7.187	0.153	0.092	0.061
<i>IV</i>	7.29	7.297	-0.007	-0.117	0.110
<i>I</i> 07	7.32				
<i>II</i>	7.40				

c) Die Saisonfaktoren sind beginnend mit dem I. Quartal  $\approx 0.930, 1.089, 1.096, 0.890$ , d.h. 93.0%, 108.9%, 109.6%, 89.0%. Bedeutung: Der Wert für das erste Quartal beträgt 93.0% des Trendwertes, für das zweite Quartal 108.9% des Trends usw.

[ 13.7 ] a) Einfacher gleitender Durchschnitt mit  $a = 3$ .  
b)

Tag	Monat		Jahr	Anzahl	in Mio	$T_t$	$x_t - T_t$
Do	Aug	21	2008	8 855 458	8.9	7.2	1.7
Fr	Aug	22	2008	7 091 353	7.1	7.2	-0.1
Di	Aug	26	2008	9 322 135	9.3	7.3	2
So	Aug	31	2008	3 985 565	4.0	7.5	-3.5
Mi	Sep	3	2008	10 245 704	10.2	7.7	2.5
Mo	Sep	8	2008	10 093 619	10.1	8.0	2.1
Sa	Sep	13	2008	3 338 575	3.3	7.9	-4.6
Do	Sep	18	2008	9 873 709	9.9	8.0	1.9

c) und e), siehe **R**-Programm, Abweichungen zur Tabelle wegen Rundungen.

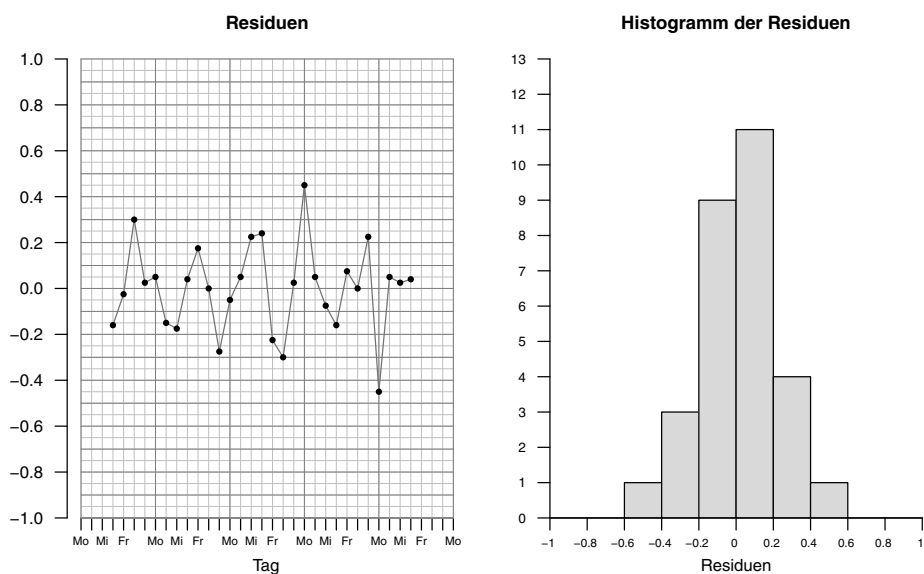


d)

Tag						Summe	Durchschnitt
Mo	--	1.7	1.6	2.1	1.2	6.6	1.650
Di	--	2.0	2.2	2.2	2.2	8.6	2.150
Mi	--	2.1	2.5	2.2	2.3	9.1	2.375
Do	1.7	1.9	2.1	1.7	1.9	9.3	1.860
Fr	-0.1	0.1	-0.3	0.0	--	-0.3	-0.075
Sa	-4.3	-4.6	-4.9	-4.6	--	-18.4	-4.600
So	-3.2	-3.5	-3.2	-3.0	--	-12.9	-3.225

f)

Tag	Monat		Jahr	Anzahl	in Mio	$T_t$	$e_t$
Do	Aug	21	2008	8855458	8.9	7.2	0.160
Fr	Aug	22	2008	7091353	7.1	7.2	-0.025
Di	Aug	26	2008	9322135	9.3	7.3	-0.150
So	Aug	31	2008	3985565	4.0	7.5	-0.275
Mi	Sep	3	2008	10245704	10.2	7.7	0.225
Mo	Sep	8	2008	10093619	10.1	8.0	0.450
Sa	Sep	13	2008	3338575	3.3	7.9	0.000
Do	Sep	18	2008	9873709	9.9	8.0	0.040

g) und h), siehe **R**-Programm

[ 13.8 ] a) Wir verwenden ein multiplikatives Modell, da die Saisonschwankungen mit steigenden Niveau ansteigen.

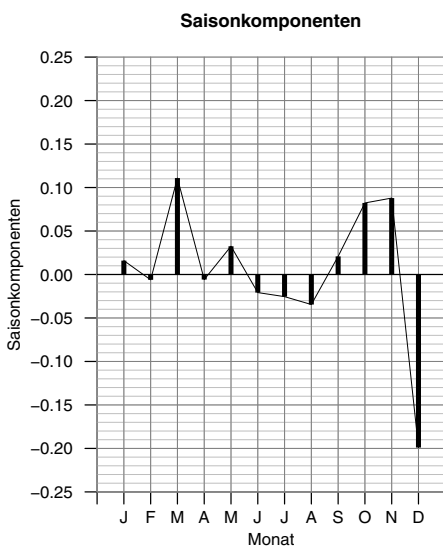
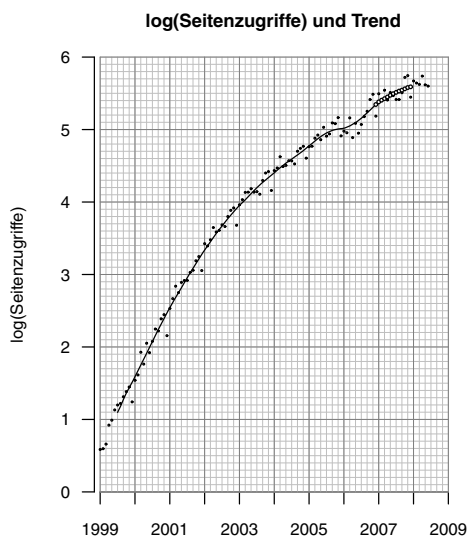
b) Der zentrierte gleitende Durchschnitt für monatliche Daten.

c)

Jahr	Monat	Zugriffe	$x_t = \log(\text{Zugriffe})$	$T_t$	$x_t - T_t$	$e_t$
2006	07	159190302	18.89	18.97	-0.08	-0.05
2006	08	177383253	18.99	19.01	-0.02	0.02
2006	09	190838123	19.07	19.04	0.03	0.01
2006	10	224604790	19.23	19.08	0.15	0.07
2006	11	241338215	19.30	19.12	0.18	0.09
2006	12	178558351	19.00	19.16	-0.16	0.04

2007	01	243125683	19.31	19.20	0.11	0.09
2007	02	221025505	19.21	19.22	-0.01	0.00
2007	03	255252638	19.36	19.24	0.12	0.01
2007	04	223004233	19.22	19.26	-0.04	-0.03
2007	05	247498040	19.33	19.29	0.04	0.01
2007	06	237110177	19.28	19.31	-0.03	-0.01
2007	07	224613993	19.23	19.33	-0.10	-0.07
2007	08	224603845	19.23	19.35	-0.12	-0.08
2007	09	246714374	19.32	19.36	-0.04	-0.06
2007	10	304682822	19.53	19.38	0.15	0.07
2007	11	312511429	19.56	19.39	0.17	0.08
2007	12	231952638	19.26	19.40	-0.14	0.06
2008	01	290163556	19.49	NA	NA	NA
2008	02	281938341	19.46	NA	NA	NA
2008	03	276733846	19.44	NA	NA	NA
2008	04	310243151	19.55	NA	NA	NA
2008	05	275019300	19.43	NA	NA	NA
2008	06	270322923	19.42	NA	NA	NA

d) und g)



e)

Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez
NA	NA	NA	NA	NA	NA	0.10	0.05	0.04	0.02	0.00	-0.28
-0.04	-0.05	0.18	-0.06	0.14	-0.07	0.00	0.09	-0.01	0.07	0.05	-0.31
-0.01	0.05	0.15	0.01	0.08	0.03	-0.05	0.00	-0.03	0.04	0.03	-0.22
0.09	0.00	0.03	0.13	0.02	-0.02	0.02	-0.05	0.03	0.06	0.05	-0.24
0.01	0.04	0.10	0.06	0.07	-0.02	-0.05	-0.13	0.02	0.10	0.09	-0.21
0.03	0.04	0.15	-0.01	-0.03	0.01	-0.03	-0.10	0.06	0.07	0.06	-0.14
-0.01	-0.04	0.04	0.05	-0.04	0.10	-0.05	-0.04	0.11	0.09	0.16	-0.10
-0.04	-0.08	0.11	-0.19	-0.01	-0.17	-0.08	-0.02	0.03	0.15	0.18	-0.16
0.11	-0.01	0.12	-0.04	0.04	-0.03	-0.10	-0.12	-0.04	0.15	0.17	-0.14
NA	NA	NA	NA	NA	NA						

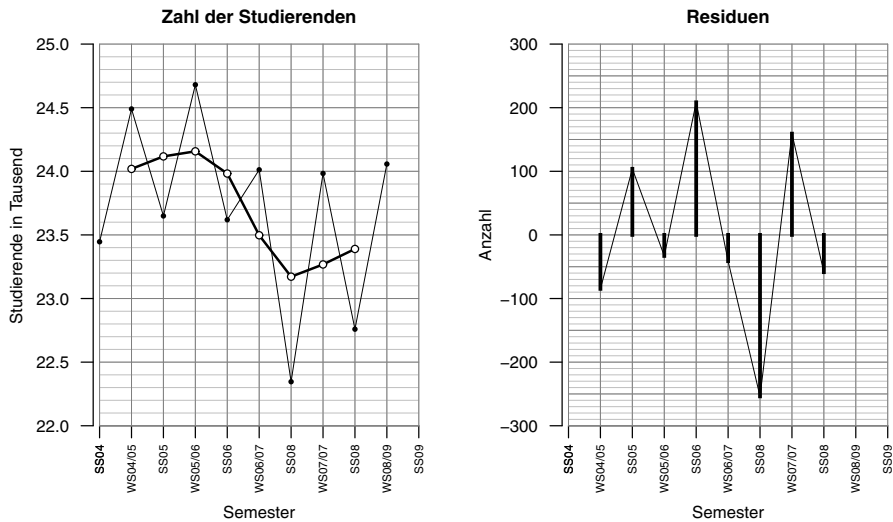
f)

	Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez
$S_t$	0.0175	-0.0063	0.1100	-0.0063	0.0337	-0.0212	-0.0267	-0.0356	0.0233	0.0833	0.0878	-0.2000
$e^{S_t}$	1.018	0.994	1.116	0.994	1.034	0.979	0.974	0.965	1.024	1.087	1.092	0.819
%	101.8	99.4	111.6	99.4	103.4	97.9	97.4	96.5	102.4	108.7	109.2	81.9

[ 13.9 ] a) - b)<sup>11</sup>

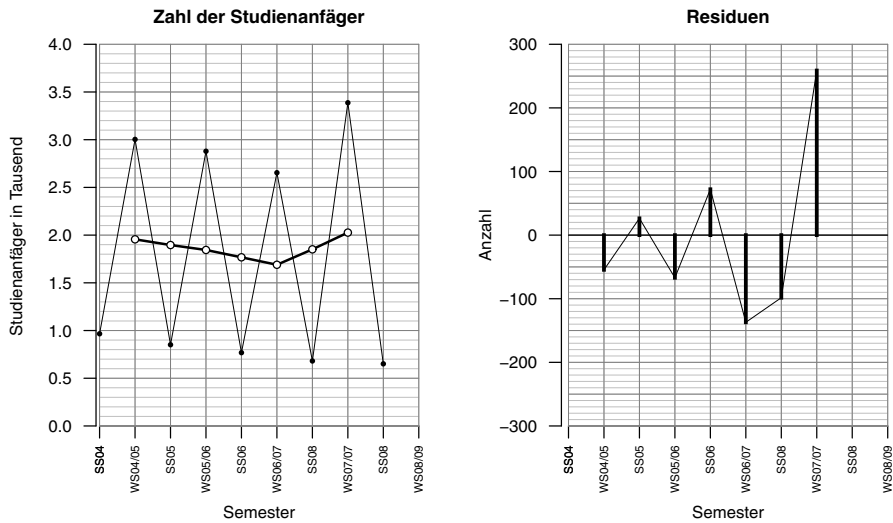
Semester	Anzahl	Trend	Anzahl - Trend	Saisonkomponente	Residuum
SS2004	23446				
WS2004/2005	24490	24018.75	471.25	556.0625	-84.8125
SS2005	23649	24117.00	-468.00	-571.9375	103.9375
WS2005/2006	24680	24157.00	523.00	556.0625	-33.0625
SS2006	23619	23982.50	-363.50	-571.9375	208.4375
WS2006/2007	24012	23497.25	514.75	556.0625	-41.3125
SS2007	22346	23171.75	-825.75	-571.9375	-253.8125
WS2007/2008	23983	23267.75	715.25	556.0625	159.1875
SS2008	22759	23389.50	-630.50	-571.9375	-58.5625
WS2008/2009	24057				

<sup>11</sup> Normalerweise würde man stärker runden. Wir haben dies hier nicht getan, um Ungenauigkeiten zu vermeiden.



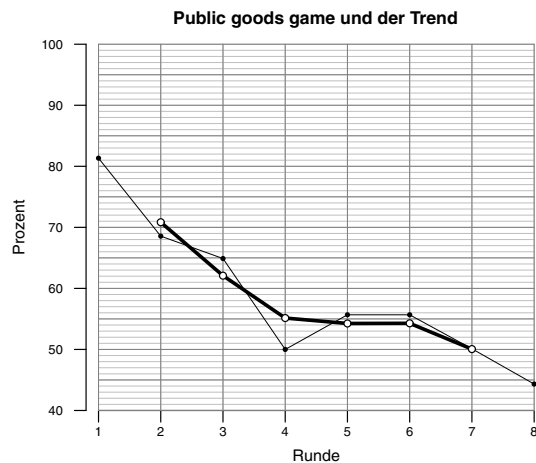
c) Siehe folgende Tabelle:

Semester	Anzahl	Trend	Anzahl - Trend	Saisonkomponente	Residuum
SS2004	966				
WS2004/2005	3003	1955.75	1047.25	1101.813	-54.563
SS2005	851	1896.00	-1045.00	-1071.333	26.333
WS2005/2006	2879	1844.25	1034.75	1101.813	-67.063
SS2006	768	1767.25	-999.25	-1071.333	72.083
WS2006/2007	2654	1689.25	964.75	1101.813	-137.063
SS2007	681	1850.75	-1169.75	-1071.333	-98.417
WS2007/2008	3387	2026.50	1360.50	1101.813	258.687
SS2008	651				



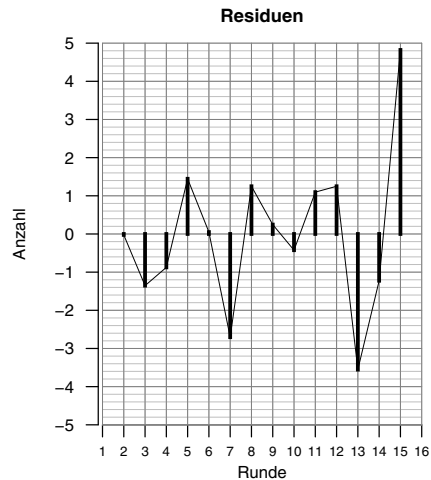
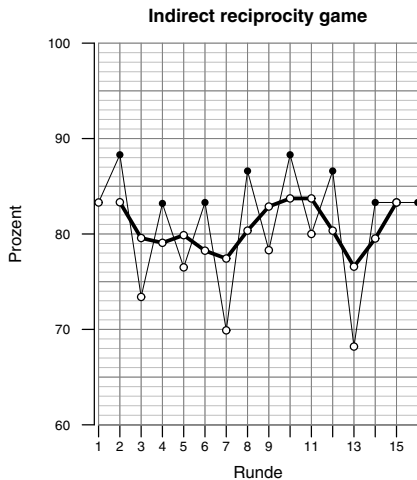
[ 13.10 ] a) Der Trend wird mit einem einfachen gleitenden Durchschnitt mit  $a = 1$  berechnet: (Anmerkung: Bei derart kleinem Stichprobenumfang, d.h. wenigen Daten, macht ein größeres  $a$  keinen Sinn.)

Runde	1	2	3	4	5	6	7	8
Kooperation in %	81.33	68.56	64.89	50.00	55.67	55.67	50.11	44.33
Trend	—	71.59	61.15	56.85	53.78	53.82	50.04	—
Residuen	—	-3.03	3.74	-6.85	1.89	1.85	0.07	—



b) Der Trend wird jetzt mit einem zentrierten Filter berechnet, analog zu dem für halbjährliche Daten:  $1/4, 1/2, 1/4$ :

Runde	$x_t$ = Kooperation in %	$T_t$	$x_t - T_t$	$S_t$	$e_t$
1 (IRG)	83.3	—	—	—	—
2 (PGG)	88.3	83.3	5.0	5.029	-0.029
3 (IRG)	73.4	79.6	-6.2	-4.829	-1.371
4 (PGG)	83.2	79.1	4.1	5.029	-0.929
5 (IRG)	76.5	79.9	-3.4	-4.829	1.429
6 (PGG)	83.3	78.2	5.1	5.029	0.071
7 (IRG)	69.9	77.4	-7.5	-4.829	-2.671
8 (PGG)	86.6	80.3	6.3	5.029	1.271
9 (IRG)	78.3	82.9	-4.6	-4.829	0.229
10 (PGG)	88.3	83.7	4.6	5.029	-0.429
11 (IRG)	80.0	83.7	-3.7	-4.829	1.129
12 (PGG)	86.6	80.3	6.3	5.029	1.271
13 (IRG)	68.2	76.6	-8.4	-4.829	-3.571
14 (PGG)	83.3	79.5	3.8	5.029	-1.129
15 (IRG)	83.3	83.3	0.0	-4.829	4.829
16 (PGG)	83.3	—	—	—	—



[ 13.11 ] a) W b) F c) W d) W

[ 13.12 ] a) F b) W c) W d) F e) W

[ 13.13 ] a) F b) W c) W d) W e) F f) W c) F d) F

[ 13.14 ] Den Umsatz erhält man als Produkt aus Preis und Verbrauch. Für die Messziffer zum Basisjahr müssen Preis, Verbrauch und Umsatz durch ihren Wert im Basisjahr geteilt werden.

a) 0.392, 0.723, 0.643, 0.608, 0.605, 0.707, 1, 1.207, 1.242

b) 1.082, 1.069, 1.084, 1.045, 1.024, 1.011, 1, 1.002, 0.913

c) 0.424, 0.773, 0.697, 0.635, 0.620, 0.715, 1, 1.210, 1.134

[ 13.15 ] Für die Messziffer zum Basisjahr muss der Goldpreis durch den Wert der jeweiligen Basisperiode geteilt werden.

Datum	2009	PMZ <sub>09</sub>	PMZ <sub>08</sub>	2008	PMZ <sub>09</sub>	PMZ <sub>08</sub>
20.02.	771	100.00	134.09	627	81.32	109.04
13.02.	727	94.29	126.43	617	80.03	107.30
06.02.	707	91.70	122.96	617	80.03	107.30
30.01.	718	93.13	124.87	622	80.67	108.17
23.01.	675	87.55	117.39	609	78.99	105.91
16.01.	627	81.32	109.04	608	78.86	105.74
09.01.	629	81.58	109.39	598	77.56	104.00
02.01.	628	81.45	109.22	575	74.58	100.00



c) i) um 9.22%, ii) um 18.55%; iii) um 25.42%; iv) um 34.09%

[ **13.16** ] a) Laspeyres: B: 1.051, 1.047; A: 1.008, 1.004. Paasche: B: 1.052, 1.053; A: 1.008, 1.005 b) Laspeyres: B: 1.000, 0.981; A: 0.998, 0.995. Paasche: B: 1.004, 0.986; A: 0.999, 0.996. c) B: 0.998, 0.962; A: 0.997, 0.961

$$[ \text{13.17} ] P_t^L = 100 \cdot \frac{20 \cdot 10 + 15 \cdot 12 + 8 \cdot 20 + 10 \cdot 12}{15 \cdot 10 + 12 \cdot 12 + 6 \cdot 20 + 10 \cdot 12} \approx 124; P_t^P = 100 \cdot \frac{20 \cdot 8 + 15 \cdot 13 + 8 \cdot 25 + 10 \cdot 12}{15 \cdot 8 + 12 \cdot 13 + 6 \cdot 25 + 10 \cdot 12} \approx 124$$

$$Q_t^L = 100 \cdot \frac{15 \cdot 8 + 12 \cdot 13 + 6 \cdot 25 + 10 \cdot 12}{15 \cdot 10 + 12 \cdot 12 + 6 \cdot 20 + 10 \cdot 12} \approx 102; Q_t^P = 100 \cdot \frac{20 \cdot 8 + 15 \cdot 13 + 8 \cdot 25 + 10 \cdot 12}{20 \cdot 10 + 15 \cdot 12 + 8 \cdot 20 + 10 \cdot 12} \approx 102$$

$$U_t = 100 \cdot \frac{675}{534} \approx 126$$

Statistikübungen für Bachelor- und Masterstudenten

Ein Arbeitsbuch mit einer Einführung in R

Böker, F.; Sperlich, S.; Zucchini, W.

2013, VIII, 378 S. 265 Abb., Softcover

ISBN: 978-3-642-34787-0