# Aufgabe 1

Im Skigebiet *Pelit 3000* wird jährlich vom örtlichen Liftbetreiber-Verband eine Unfallstatistik herausgegeben. In ihr aufgeschlüsselt finden sich folgende Merkmale:

- $A$ : "Schwierigkeitsgrad der Piste" mit den Ausprägungen
  - $a_1$ : "blau (geringer Schwierigkeitsgrad)"
  - $a_2$ : "rot (mittlerer Schwierigkeitsgrad)"
  - $a_3$ : "schwarz (hoher Schwierigkeitsgrad)"
- $B$ : "Schwere des Unfalls" mit den Ausprägungen
  - $b_1$ : "leicht"
  - $b_2$ : "schwer"

In der Saison 2011/2012 wurden 246 Unfälle registriert und die absoluten Häufigkeiten in folgender Tabelle veröffentlicht:

| $\downarrow B \backslash A \rightarrow$ | $a_1$ | $a_2$ | $a_3$ | |
|---|---|---|---|---|
| $b_1$ | ?? | ?? | ?? | 156 |
| $b_2$ | ?? | ?? | ?? | 90 |
| | 84 | 91 | 71 | 246 |

1. Welchen Merkmalstypen besitzt das Merkmal $A$? Nennen Sie eine Rechenoperation, die **nicht** gestattet ist.

2. Ergänzen Sie in der Tabelle die fehlenden Werte unter Angabe des vollständigen Rechenweges. Dazu sei Ihnen bekannt, dass rund..
   - .. 86.9048% der Unfälle, die auf blauen Pisten passiert sind, leicht waren und
   - .. 11.3821% aller Unfälle leichte waren und auf schwarzen Pisten geschehen sind.

*Hinweis*: Wenn Sie die Teilaufgabe 2. nicht bearbeiten können, verwenden Sie

$$n(a_1, b_1) = 73, \qquad n(a_1, b_2) = 11, \qquad n(a_2, b_1) = 55,$$
$$n(a_2, b_2) = 36, \qquad n(a_3, b_1) = 28, \qquad n(a_3, b_2) = 43.$$

3. Berechnen und interpretieren Sie $f(a_3|b_2)$.

4. Sind die Merkmale $A$ und $B$ stochastisch unabhängig?

5. Berechnen Sie die normierte Entropie $H_B^*$ des Merkmals $B$.

Um der Unfallrate auf der gefährlichsten schwarzen Piste nachzugehen, wurde diese in Abschnitte unterteilt und die Fahrt eines Testskifahrers aufgezeichnet. Dabei wurden folgende Daten erhoben:

| Streckenabschnitt | 1 | 2 | 3 | 4 | 5 | 6 |
|---:|---|---|---|---|---|---|
| Länge des Streckenabschnitts in km | 1.1 | 1.3 | 0.2 | 0.5 | 1.1 | 0.9 |
| Durchschnittsgeschwindigkeit des Skifahrers in km/h | 93 | 78 | 102 | 97 | 44 | 69 |

6. Welchen Mittelwert würden Sie zur Berechnung der Durchschnittsgeschwindigkeit über alle Abschnitte verwenden? Berechnen Sie diesen Mittelwert.

In **R** sei Ihnen ein Workspace mit folgender Tabelle gemeinsamer absoluter Häufigkeiten gegeben:

```
> M
       blau rot schwarz
leicht   83  58      18
schwer   13  38      41
```

7. Geben Sie den **R**-Befehl an, um die relativen Randhäufigkeiten der Matrix zu erhalten, und speichern Sie diese unter dem Namen `M1` ab.

8. Berechnen Sie die bedingte relative Häufigkeit dafür, dass ein Unfall schwer ist, wenn er auf einer blauen Piste geschieht.

**Schmierpapier zur Aufgabe 1**

# Aufgabe 2

Eine Stichprobe vom Umfang $n$ wird aus einer normalverteilten Grundgesamtheit $X$ gezogen, wobei $\mu = 0$ bekannt sei. Die Dichtefunktion von $X$ hat folgende Gestalt:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x \in \mathbb{R}, \sigma^2 > 0.$$

1. Zeigen Sie, dass die logarithmierte Likelihoodfunktion von $\sigma^2$ gegeben ist durch:

$$\ln L(\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2}.$$

2. Zeigen Sie, dass der ML-Schätzer $\hat{\sigma}^2_{ML}$ für den unbekannten Parameter $\sigma^2$ lautet (keine Berechnung der zweiten Ableitung nötig):

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^n X_i^2.$$

3. Zeigen Sie, dass für den Fall $\mu = 0$, $\hat{\sigma}^2_{ML}$ ein erwartungstreuer Schätzer für $\sigma^2$ ist, d.h. es gilt
$$E[\hat{\sigma}^2_{ML}] = \sigma^2.$$

Sie beobachten die Jahresrendite $x_i$ eines Aktienindex.

| Jahr $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | -0.03 | 0.12 | -0.24 | 0.06 | -0.01 | -0.01 | 0.10 | -0.15 | 0.10 | 0.03 |

Gehen Sie im Folgenden davon aus, dass die beobachteten Renditen Realisationen einer Stichprobe aus $X$ sind, wobei $X \sim N(\mu_X, \sigma^2_X)$ gilt.

4. Nach der Berechnung der Stichprobenvarianz $S^2_{X,10} = 0.0133$ vermuten Sie, dass die Varianz der Jahresrenditen größer als 0.01 ist. Sie konstruieren einen Hypothesentest der Form:
$$H_0 : \sigma^2_X \leq 0.01 \qquad H_1 : \sigma^2_X > 0.01.$$
Bestimmen Sie $k$ so, dass die Prüfgröße $T_n = (n-1)\frac{S^2_{X,10}}{\sigma^2}$ des Hypothesentest mit einer Wahrscheinlichkeit von 95% im Intervall $[2.09, k]$ liegt.

Zur Durchführung des obigen Hypothesentests steht Ihnen folgender **R**-Output zur Verfügung.

```
> X=c(-0.03,0.12,-0.24,0.06,-0.01,-0.01,0.10,-0.15,0.10,0.03)
> T=(length(X)-1)*var(X)/0.01
> T=12.001
> round(pchisq((length(X)-1)*var(X)/0.01, length(X)-1),4)
[1] 0.7867
> round(dchisq((length(X)-1)*var(X)/0.01, length(X)-1),4)
[1] 0.0564
> 1-round(pchisq((length(X)-1)*var(X)/0.01, length(X)-1),4)
[1] 0.2133
```

5. Vervollständigen Sie die Aussagen (a)-(d):

   (a) Der Wert der Teststatistik des obigen Hypothesentests lautet ______________.

   (b) Der p-Wert des obigen Hypothesentests lautet ______________.

   (c) Die Anzahl der Freiheitsgrade der $\chi^2$-Verteilung im obigen Output beträgt __________.

   (d) Der Befehl zur Berechnung der Stichprobenvarianz des Vektors `X` im obigen Output lautet ________________________.

Sie beobachten zusätzlich die Jahresrendite $y_i$ eines zweiten Aktienindix.

| Jahr $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 0.07 | -0.08 | 0.31 | 0.31 | -0.03 | -0.09 | -0.03 | 0.11 | -0.26 | 0.03 |

Gehen Sie wiederum davon aus, dass es sich bei den beobachteten Jahresrenditen um Realisationen einer Stichprobe aus $Y$ handelt, mit $Y \sim N(\mu_Y, \sigma_Y^2)$.

6. Sie vermuten, dass die Varianz der Jahresrenditen des zweiten Aktienindex $Y$ größer ist als die Varianz des ersten Aktienindex $X$. Testen Sie diese Vermutung bei einer Irrtumswahrscheinlichkeit von 5%, wenn Sie wissen, dass $S^2_{Y,10} = 0.0314$. Vervollständigen Sie das zugehörige Testverfahren:

   (a) Hypothesenpaar:

   (b) Kritischer Bereich:

   (c) Kritische Schranke:

   (d) Berechnung der Teststatistik:

   (e) Testentscheidung:

**Schmierpapier zur Aufgabe 2**

# Aufgabe 3

Im Jahr 2011 waren in Deutschland 158 Betriebe von Arbeitsniederlegungen betroffen. In einer Metropole streikt das Luftfahrtpersonal durchschnittlich 6 Mal pro Jahr. Sei die Zufallsvariable $X$: "Anzahl der Streiks des Luftfahrtspersonals in einem Jahr" poissonverteilt für $x = 0, 1, 2, \ldots$.

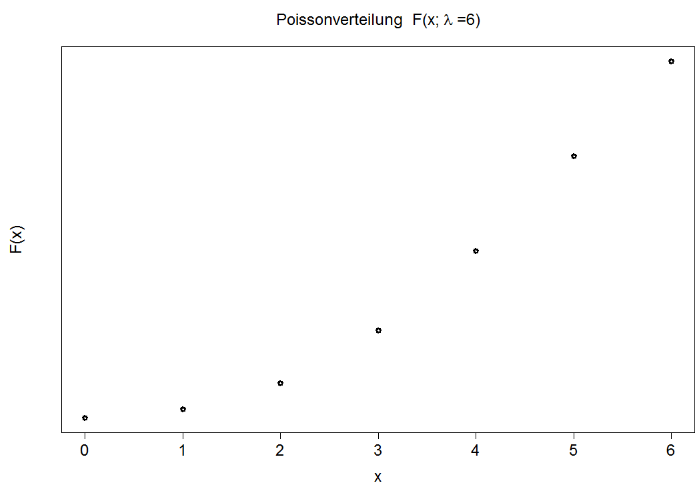
1. Bestimmen Sie den Parameter der Poissonverteilung für $X$. (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 3.)

2. Berechnen Sie die Wahrscheinlichkeit, dass im nächsten Jahr mehr als 6 Streiks vom Luftfahrtpersonal durchgeführt werden.

3. Geben Sie den Modus ggf. die Modi für die Zufallsvariable $X$ an. (Ohne Begründung)

4. Geben Sie die Varianz der Zufallsvariable $X$ an.

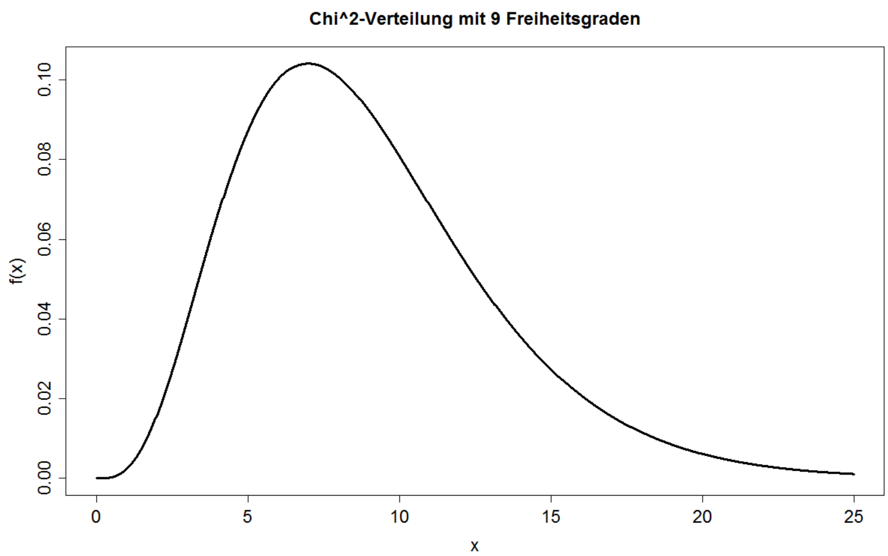
5. Markieren Sie in folgendem Plot das 30%-Quantil der Poissonverteilung mit Parameter $\lambda = 6$.



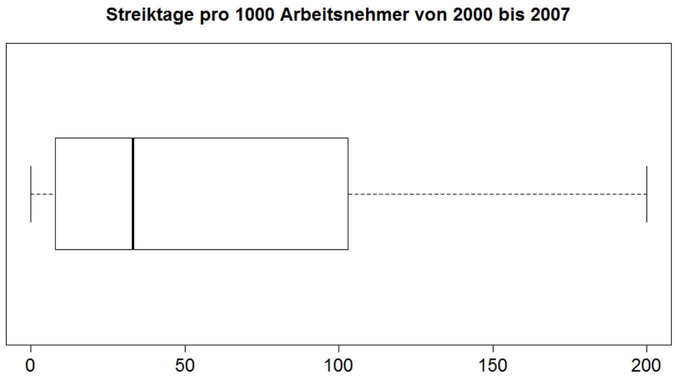
6. Sei die Zufallsvariable $Y$: "Dauer zwischen dem Beginnen zweier aufeinanderfolgender Streiks (in Tagen)" exponentialverteilt mit der Verteilungsfunktion $F_{Exp}(y;\lambda) = 1-e^{-\lambda y}$, für $y \geq 0$. Bestimmen Sie den Parameter $\lambda$ der Exponentialverteilung für $Y$. (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 0.1.)

7. Berechnen Sie die Wahrscheinlichkeit $P(Y > a + 30 | Y > a)$.

8. Die Dauer eines Streiks $Z$ (in Tagen) sei annähernd normalverteilt mit unbekannten $\mu$ und $\sigma^2$. Eine Zufallsstichprobe von 10 Streiks liegt Ihnen zur Analyse vor. Sie möchten auf 10%-Niveau testen, ob die Standardabweichung der Streikdauer 1 Tag beträgt. Markieren Sie den kritischen Bereich des Tests in folgendem Plot.



9. Sie haben die Anzahl der Streiktage pro 1000 Arbeitsnehmer in 12 Ländern analysiert. Folgender Box-Plot für die Daten wird erzeugt. Welche Streuungsmaße sind im Plot unmittelbar zu erkennen?




10. Mit dem R-Befehl ________________________ können 100 poissonverteilte Zufallszahlen mit $\lambda = 6$ erzeugt werden.

11. Der folgende R-Befehl speichert in `y` ____________________ von `x`.

```
> x = c(6,6,7,4,5,6,10,4,6,3,8,8,6,4,4,8,2)
> y = 1/(length(x)-1)*sum((x-mean(x))^2)
```

12. Der R-Output von folgendem Befehl ist ______________.

```
> (qnorm(0.22,2,0.4)-qnorm(0.17,2,0.4))>0
```

**Schmierpapier zur Aufgabe 3**

# Aufgabe 4

Eine Produktion von Stickpullovern soll auf Webfehler überprüft werden. Dazu sollen zufällig $n$ Pullover ausgewählt und auf Webfehler durchsucht werden. Die Anzahl der Webfehler pro Pullover sei verteilt mit einem unbekannten Mittelwert $\mu$ und Standardabweichung $\sigma = 0.25$. Sie können davon ausgehen, dass das standardisierte Stichprobenmittel $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ approximativ normalverteilt ist.

1. Bestimmen Sie den Stichprobenumfang $n$ so, dass die mittlere Anzahl von Webfehlern in der Stichprobe vom unbekannten Mittelwert $\mu$

   (a) mit einer Wahrscheinlichkeit von mindestens 95% um weniger als 0.04 abweicht,

   (b) mit einer Wahrscheinlichkeit von mindestens 99% um weniger als 0.06 abweicht.

2. Sei der Stichprobenumfang $n = 200$. Berechnen Sie die Wahrscheinlichkeit, dass die mittlere Anzahl von Webfehlern in der Stichprobe um weniger als 0.02 vom unbekannten Mittelwert $\mu$ abweicht.

3. In einer einfachen Stichprobe des Umfangs $n = 200$ werden im Mittel 0.5 Webfehler pro Pullover gefunden. Geben Sie das realisierte Konfidenzintervall für $\mu$ bei einem Konfidenzniveau von $1 - \alpha = 0.95$ an.

4. Welche der folgenden Aussagen sind richtig? (Begründung)

   (a) Der benötigte Stichprobenumfang wird größer, wenn die Sicherheitswahrscheinlichkeit steigt, mit der der Stichprobenmittelwert vom wahren Mittelwert um weniger als 0.1 abweicht.

   (b) Der benötigte Stichprobenumfang wird größer, wenn die Differenz größer gewählt wird, um die der Stichprobenmittelwert vom wahren Mittelwert maximal abweichen darf.

   (c) Der benötigte Stichprobenumfang wird kleiner, wenn die Varianz der Grundgesamtheit größer gewählt wird.

   (d) Der Mittelwert der Grundgesamtheit $\mu$ hat keinen Einfluss auf den benötigten Stichprobenumfang.

In **R** sei Ihnen nun ein Datenvektor `x` gegeben.

5. Was wird im folgenden **R** Code in der Variablen `ki` abgespeichert? (Exakte Bezeichnung)

```
> n=length(x)
> fg=n-1
> alpha=1-0.05
> xm=mean(x)
> sn=sd(x)
> diff=qt(1-alpha/2,fg)*sn/sqrt(n))
> xl=xm-diff
> xu=xm+diff
> ki=c(xl,xu)
```

**Schmierpapier zur Aufgabe 4**

# Aufgabe 1

1. Komparatives Merkmal, **nicht** erlaubte Rechenoperation: Addition uvm.

2. $f(b_1|a_1) = 0.869048$ und $f(b_1 \cap a_3) = 0.113821$ gegeben, Tabelle unter Angabe des Rechenwegs ausfüllen wie im Hinweis.

3. $$f(a_3|b_2) = \frac{f(b_2 \cap a_3)}{f(b_2)} = 0.4778$$

   47.47% der schweren Unfälle ereigneten sich auf einer schwarzen Piste .

4. Stochastische Unabhängigkeit, wenn z. B.

   $$f(a_3|b_2) \stackrel{!}{=} f(a_3)$$

   Nach 3. ist $f(a_3|b_2) = 0.4778 \neq \frac{71}{246} = f(a_3)$, damit sind die Merkmale $A$ und $B$ nicht stochastisch unabhängig.

5. $H_B^* = 0.9476$.

6. Harmonisches Mittel (bei ungleichen Abschnittslängen): $\bar{x}_H = 69.24$ km/h.

7. `> M1=c(colSums(M),rowSums(M))/sum(M)`

8. `> M[2,1]/M1[1]` oder 0.131 oder 0.1354

# Aufgabe 2

1. Likelihoodfunktion aufstellen

   $$\begin{aligned} L(\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{2\sigma^2}\right) \end{aligned}$$

   und anschließend logartihmieren, führt schließlich zu :

   $$\begin{aligned} \ln L(\sigma^2) &= -n \ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma^2} \\ &= -n/2 \ln(2\pi) - n/2 \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2. \end{aligned}$$

2. Erste Ableitung bilden und gleich 0 setzen :

$$\frac{\partial \ln L(\sigma^2)}{\partial \sigma^2} \stackrel{!}{=} 0.$$

Anschließend nach $\sigma^2$ auflösen :

$$\begin{aligned}
& -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_i^2 = 0 \\
& \Leftrightarrow \sum_{i=1}^{n} x_i^2 = n\sigma^2 \\
& \Rightarrow \hat{\sigma}^2_{ML} = n^{-1}\sum_{i=1}^{n} x_i^2.
\end{aligned}$$

3. Einsetzen und ausrechnen . Beachte, dass hier aufgrund der u.i.v. Annahme an $X_i$ $E[X_i^2] = Var(X_i) = \sigma^2$ gilt.

$$E\left[n^{-1}\sum_{i=1}^{n} X_i^2\right] = n^{-1}\sum_{i=1}^{n} E[X_i^2] = n^{-1}\sum_{i=1}^{n}\sigma^2 = \sigma^2.$$

4.

$$\begin{aligned}
P(2.09 < T_n < k) &= 0.95 \\
F_{\chi^2(9)}(k) - F_{\chi^2(9)}(2.09) &= 0.95 \\
F_{\chi^2(9)}(k) &= 0.96
\end{aligned}$$

In Tabelle nachschauen, dann ergibt sich k=17.61.

5. (a) Der Wert der Teststatistik des obigen Hypothesentests lautet 12.001

   (b) Der p-Wert des obigen Hypothesentests lautet 0.2133.

   (c) Die Anzahl der Freiheitsgrade der $\chi^2$-Verteilung im obigen Output beträgt 9.

   (d) Der Befehl zur Berechnung der Stichprobenvarianz des Vektors `X` im obigen Output lautet `var(X)`.

6. (a) $H_0 : \sigma_X \geq \sigma_Y$

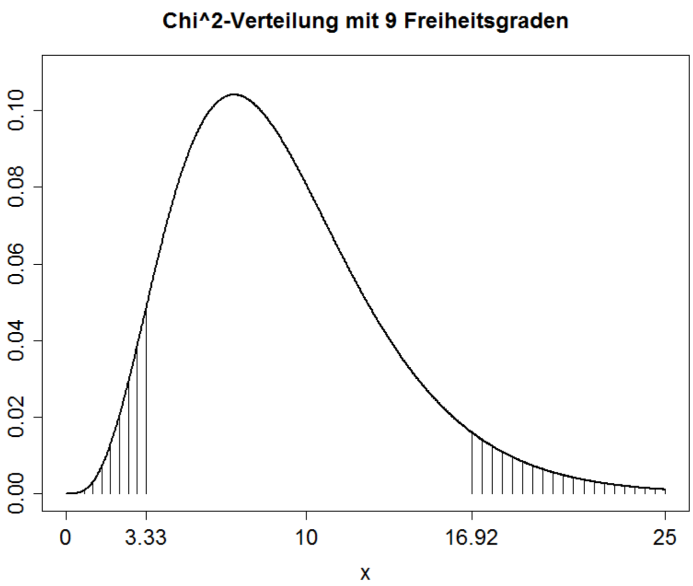
   (b) $t_n < f_{0.05;9;9}$ oder $t_n > f_{0.95;9;9}$

   (c) $f_{0.05;9;9} = 1/3.1789 = 0.3146$ oder $f_{0.95;9;9} = 3.1789$.

   (d) $t_n = \frac{S_Y^2}{S_X^2} = 0.04236$ oder $t_n = \frac{S_X^2}{S_Y^2} = 0.0314/0.0133 = 2.3609$.

   (e) $0.4236 \not< 0.3146$ oder $2.3607 \not> 3.1789$, d.h. $H_0$ so oder so nicht ablehnen.

# Aufgabe 3

1. 6
2. $P(X > 6) = 1 - F_{Pois}(6; 6) = 0.3937$
3. 5 und 6
4. 6
5. $X = 5$
6. $\lambda = \frac{6}{365}$
7. $P(Y > 30) = 1 - F_{Exp}(30, \frac{6}{365}) = 0.6114$
8. 

9. Spannweite und Quartilabstand.
10. `rpois(100,6)`
11. die Varianz
12. `TRUE`

# Aufgabe 4

1. (a)

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| < 0.04\right) \geq 0.95$$

umformen zu

$$n \geq \left(\frac{\lambda_{1-\frac{\alpha}{2}}}{\frac{\epsilon}{\sigma}}\right)^2 = \left(\frac{1.9600}{\frac{0.04}{0.25}}\right)^2 = 150.0625$$

$\rightarrow n = 151$

(b) mit anderen Werten: $n = 116$ .

2. $n = 200$,

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| < 0.02\right) = 0.7421$$

3.

$$\left[\bar{X}_n - \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{X}_n + \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = [0.4654; 0.5346]$$

4. Anhand von Formel aus 1

(a) Richtig, $1 - \alpha \uparrow$, $\lambda_{1-\frac{\alpha}{2}} \uparrow$ $\Rightarrow n \uparrow$

(b) Falsch, $\epsilon \uparrow$ $\Rightarrow n \downarrow$

(c) Falsch, $\sigma^2 \uparrow$, $\sigma \uparrow$ $\Rightarrow n \uparrow$

(d) Richtig, $\mu$ kommt in der Formel aus 1. nicht vor.

5. 1−`alpha`%-Konfidenzintervall eines Mittelwertes bei unbekannter Varianz.