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 Merkmalstyp 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 und interpretieren Sie diesen Wert im Vergleich zu $H_A^* = 0.9953$.

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:

- W : "Anzahl der Bäume am Rand des Streckenabschnitts"
- X : "Länge des Streckenabschnitts in km"
- Y : "Durchschnittsgeschwindigkeit des Skifahrers in km/h"
- Z : "Neigung der Piste in %"

Abschnitt	1	2	3	4	5	6
w_i	34	37	29	45	54	39
x_i	1.1	1.3	0.2	0.5	1.1	0.9
y_i	93	78	102	97	44	69
z_i	20.2	17.2	20.4	16.1	15.6	17.2

6. Berechnen Sie die Korrelation zwischen W und Z. Verwenden Sie dazu

$$\sum_{i=1}^{n}(w_i - \bar{y})(w_i - \bar{z}) = -79.9333, \qquad s_W^2 = 77.4667, \qquad s_Z^2 = 4.1937$$

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

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 σ^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 σ^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 σ^2 ist, d.h.
$$E[\hat{\sigma}^2_{ML}] = \sigma^2.$$

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.

5. Der Wert der Prüfgröße T_n des Hypothesentests beträgt 12.001. Berechnen Sie für obigen Hypothesentest den zugehörigen p-Wert. Hierfür steht Ihnen folgender Wert zur Verfügung $F_{\chi^2_{(9)}}(12.001) \approx 0.7867$.

6. Treffen Sie anhand des p-Wertes eine Testentscheidung für obigen Test bei einer Irrtumswahrscheinlichkeit von 5%.
   (**Hinweis**: Wenn Sie Teilaufgabe 5. nicht lösen konnten, dann und nur dann verwenden Sie als p-Wert 3%)

Sie beobachten zusätzlich die Jahresrendite y_i eines zweiten Aktienindex.

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 $Y \sim N(\mu_Y, \sigma_Y^2)$.

7. 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.)

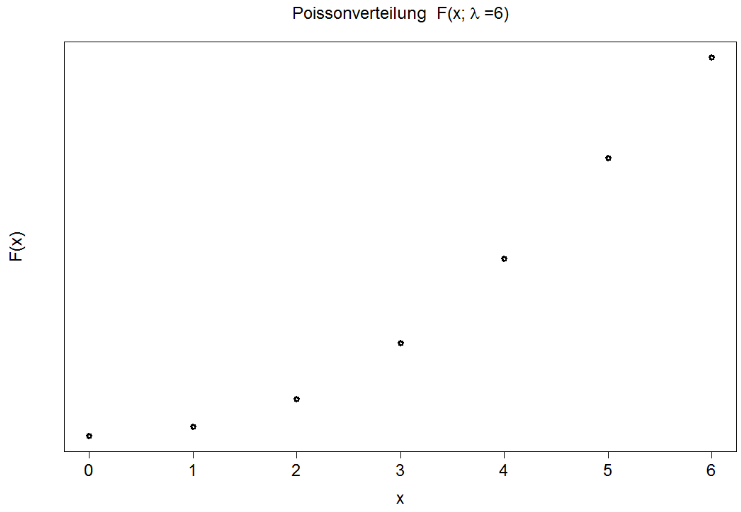
2. Berechnen Sie die Wahrscheinlichkeit, dass im nächsten Jahr mehr als 6 Streiks vom Luftfahrtpersonal durchgeführt werden.

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. Berechnen Sie die Wahrscheinlichkeit, dass in den kommenden 6 Monaten genau 3 Streiks stattfinden.

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



7. 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 λ der Exponentialverteilung für Y. (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 0.1.)

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

9. Die Dauer eines Streiks Z sei annähernd normalverteilt. Eine Zufallsstichprobe von 10 Streiks liegt Ihnen zur Analyse vor, wobei $\overline{z}_{10} = 1.7$ und $s_{10} = 0.6$ ist. Führen Sie einen Hypothesentest für

$$H_0 : \mu \leq 1.5 \qquad \text{gegen} \qquad H_1 : \mu > 1.5$$

zum Signifikanzniveau $\alpha = 10\%$ durch.

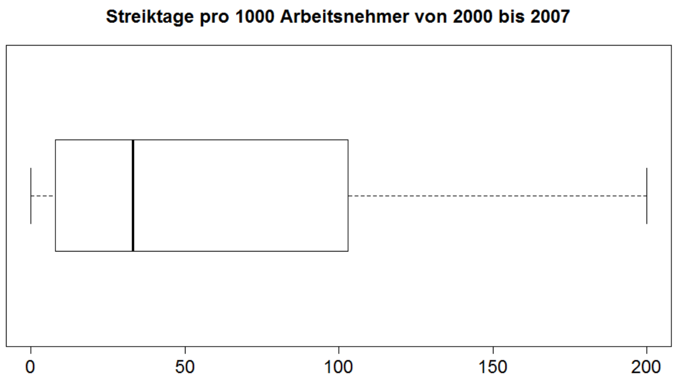
(a) Kritischer Bereich:

(b) Berechnung der Teststatistik T_n:

(c) Kritische Schranke:

(d) Testentscheidung:

10. 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?



Schmierpapier zur Aufgabe 3

Aufgabe 4

Von 100 zufällig ausgewählten Staaten liegen im `R`-Data Frame `staaten` die folgenden Merkmale vor:

Merkmal *BIP*:	Bruttoinlandsprodukt 2012 (in Bio.€)	Spalte `BIP`
Merkmal *Kontinent*:	Geographischer Kontinent des Staates	Spalte `Kontinent`
Merkmal *Status*:	Wirtschaftl. und soz. Entwicklungsstatus	Spalte `Status`

Ihnen liegt folgender `R`-Output vor:

```
'data.frame':   100 obs. of  3 variables:
 $ BIP      : num  3.72 3.93 3.42 3.51 3.16 3.19 7.38 3.74 5.47 3.15 ...
 $ Kontinent: Factor w/ 3 levels "Amerika","Asien",..: 2 3 2 2 2 3 3 2 3 3 ...
 $ Status   : Factor w/ 2 levels "Entwicklungsland",..: 1 2 2 1 2 1 1 1 2 2 ...
```

1. Geben Sie den `R`-Befehl an, mit dem der Output erzeugt wurde.

2. Wie viele Ausprägungen des Merkmals *Status* kommen im Datensatz vor?

3. Auf welchem Kontinent liegt der erste Staat im Datensatz?

4. Was berechnet die Funktion `cor`? Stellt der Befehl

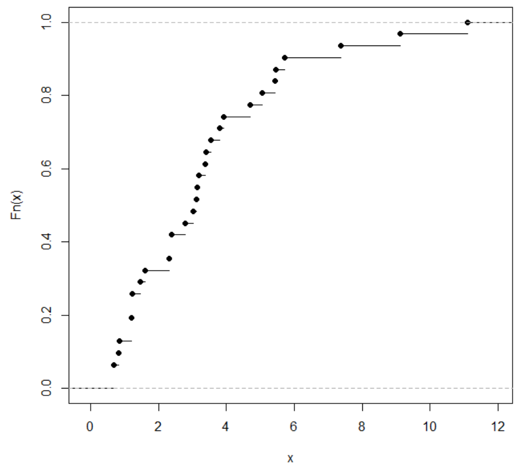
   ```
   > cor(staaten$BIP,staaten$Status)
   ```

   eine sinnvolle Operation dar? (mit <u>kurzer</u> Bergründung)

Betrachten Sie zusätzlich den folgenden Output

```
> fkt1=function(a){
+   quantile(a,0.5,type=1) }
> table(staaten$Kontinent)/nrow(staaten)
Amerika   Asien  Europa
   0.29    0.40    0.31
> As=staaten[staaten$Kontinent=="Asien",]
> Eu=staaten[staaten$Kontinent=="Europa",]
> Am=staaten[staaten$Kontinent=="Amerika",]
> (mz1=c(min(As$BIP),min(Eu$BIP),min(Am$BIP)))
[1] 0.36 0.69 0.42
> mz2=c(max(As$BIP),max(Eu$BIP),max(Am$BIP))
> mz3=c(fkt1(As$BIP),fkt1(Eu$BIP),fkt1(Am$BIP))
> mz4=c(mean(As$BIP),mean(Eu$BIP),mean(Am$BIP))
> (mz=cbind(mz1,mz2,mz3,mz4))
  mz1   mz2  mz3      mz4
 0.36  4.72 2.67 2.513250
 0.69 11.11 3.13 3.402581
 0.42  4.52 2.06 1.906897
```

sowie die dazugehörige Abbildung



5. Vervollständigen Sie die folgenden Aussagen:

(a) Im Durchschnitt erzielten die amerikanischen Staaten ein BIP von ____________ Bio.€.

(b) Das geringste BIP unter allen Staaten beträgt ____________ Bio.€.

(c) Die Hälfte der europäischen Staaten erzielte ein BIP von maximal ________ Bio.€.

(d) Die Abbildung zeigt die ____________________________________-funktion des Merkmals BIP der Staaten des Kontinents ____________.

6. Ausgehend von den vorliegenden Outputs, geben Sie R-Befehl(e) an, um das durchschnittliche BIP der asiatischen Entwicklungsländer zu berechnen.

7. Mit Hilfe eines χ^2-Unabhängigkeitstests soll bei einer Irrtumswahrscheinlichkeit von 10% überprüft werden, ob die Merkmale *Kontinent* und *Status* unabhängig sind. Es ist bekannt, dass für die Prüfgröße T des Testes gilt: $T \overset{appr}{\sim} \chi^2(2)$
Die beobachteten gemeinsamen absoluten Häufigkeiten liegen Ihnen bereits im Vektor `beo` vor und die entsprechenden gemeinsamen absoluten Häufigkeiten, die sich bei Unabhängigkeit ergeben würden, liegen im Vektor `erw` vor.

   Führen Sie den Test durch, indem Sie zunächst die R-Befehle angeben, um

   (a) ein Objekt `diff2` mit den quadrierten Differenzen zwischen beobachten und erwarteten Häufigkeiten, d.h. zwischen den einzelnen Elementen in `beo` und `erw`, zu erstellen,

   (b) ausgehend von `diff2` die realisierte Prüfgröße als Objekt `t` zu berechnen.

   (c) Treffen und begründen Sie nun die Testentscheidung anhand des Outputs:

```
> (k=qchisq(0.9,df=2))
[1] 4.60517
> t>k
[1] TRUE
```

   (d) Geben Sie den R-Befehl an, um den p-Wert des Tests zu berechnen.

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. $H_B^* = 0.9476$ ist kleiner als H_A^*, damit streut das Merkmal B geringer als A.

6. $$\frac{\frac{-79.9333}{6}}{\sqrt{77.4667}\sqrt{4.1937}} = -0.7391$$

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

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 σ^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. P-Wert richtig hinschreiben, richtig Umformen

$$P(T_n > 12.001 | H_0 \text{ gilt}) = 1 - P(T_n \leq 12.001 | H_0 \text{ gilt}) = 1 - 0.78647 = 0.2133.$$

6. Da $0.2133 > 0.05$ H_0 nicht ablehnen.

7. (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. $f_{Pois}(3; 3) = 0.2240$
6. $X = 5$
7. $\lambda = \frac{6}{365}$
8. $P(Y > 30) = 1 - F_{Exp}(30, \frac{6}{365}) = 0.6114$
9. $t_n > t_{1-\alpha;n-1}$
   $T_n = \sqrt{10}\frac{1.7-1.5}{0.6} = 1.0541$
   $t_{0.9;9} = 1.383$
   H_0 nicht ablehnen.
10. Spannweite und Quartilabstand.

Aufgabe 4

1. `str(staaten)`
2. 2
3. Asien
4. Korrelationskoeffizient nein, da Status qualitatives Merkmal (Korrelationskoeffizient nur für quantitative Daten)
5. (a) 1,91 Bio.€.
   (b) 0.36 Bio.€.
   (c) 3.13 Bio.€.
   (d) (empirische) Verteilungsfunktion Europa
6. z.B.

```
mean(As[As$Status=="Entwicklungsland","BIP"])
mean(As$BIP[As$Status=="Entwicklungsland"])
```

7. (a) `diff2=(beo-erw)^2`
   (b) `t=sum(diff2/erw)`
   (c) Nullhypothese wird abgelehnt, da realisierte Prüfgröße `t` kritische Schranke `k` überschreitet
   (d) `1-pchisq(t,2)`