Aufgabe 1

Sie wurden von der Verbraucherschutzzentrale Bayern beauftragt, die Kunden der Ulmen Investmentbank AG zu befragen, ob diese absichtlich falsch beraten wurden. Bei der Umfrage waren 60% der Kunden über 65 Jahre. 80% der falsch beratenen Kunden waren über 65 Jahre alt. Außerdem wissen Sie, dass 30% der Kunden richtig beraten wurden.

1. Stellen Sie eine Vierfeldertafel für die Zufallsvariablen R "Person ist älter als 65" und F "Person wurde falsch beraten" auf.

2. Geben Sie folgende Wahrscheinlichkeiten an:

   (a) Ein zufällig ausgewählter Kunde ist älter als 65 und wurde richtig beraten.

   (b) Ein zufällig ausgewählter Kunde, der falsch beraten wurde, ist jünger als 65.

   (c) Ein zufällig ausgewählter Kunde, der jünger als 65 ist, wurde falsch beraten.

3. Sind die beiden Zufallsvariablen R und F stochastisch unabhängig?

4. Gegeben sei die normalverteilte Zufallsvariable R_i "Jährliche logarithmierte Rendite der Schon-ewig-da-bank im $i-$ten Jahr", d.h. $R_i = \ln(S_i/S_{i-1})$, wobei S_i dem Wert des Aktienkurses am Ende des $i-$ten Jahres entspricht. Aufgrund der Zahlen der letzten 180 Jahre, ermitteln Sie als durchschnittliche jährliche logarithmierte Rendite $\bar{R} = 1/180 \sum_{i=1}^{180} R_i = -0.15$. Die Varianz ist bekannt und beträgt $\sigma_{\bar{R}}^2 = 1.44$.

   (a) Ein Bekannter behauptet, dass die mittlere logarithmierte Rendite der Schon-ewig-da-bank jährlich kleiner als 0 sei. Testen Sie diese Behauptung auf dem 5% Signifikanzniveau.

   (b) Sie kaufen trotz der Aussagen ihres Bekannten heute 10 Aktien der Schon-ewig-da-bank zum Tageskurs von 50 EUR. Mit welchem Wert ihres Investments rechnen Sie in genau einem Jahr, wenn Sie wissen, dass die tatsächliche durchschnittliche jährliche logarithmierte Rendite -0.14 beträgt? Wieviel Prozent des heutigen Wertes sind das?

5. Interpretieren Sie folgenden **R**-Code und geben Sie an, welcher Output auf dem Monitor erscheint:

```
>t=sqrt(180)*(-0.15-0)/(sqrt(1.44))
>ks=qnorm(0.05,0,1)
>if(t<ks){
>print("H0 wird abgelehnt")
>else
>print("H0 kann nicht verworfen werden") }
```

Aufgabe 2

In 100 ausgewählten Landkreisen in Deutschland wurde der Zusammenhang zwischen der durchschnittlichen Arbeitslosenqoute und der durchschnittlichen Kinderzahl je Frau im Jahr 2008 untersucht. Das Merkmal X stellt "Durchschnittliche Kinderzahl je Frau *von... bis...* in einem Landkreis im Jahr 2008" dar, und das Merkmal Y bezeichnet "Durchscnittliche Arbeitslosenquote *von... bis...* in einem Landkreis im Jahr 2008 in %".
Die relativen Häufigkeiten für die beiden Merkmale sind in nachfolgender Tabelle zusammengefasst:

$(\widetilde{x}_{i-1}; \widetilde{x}_i] \downarrow \backslash (\widetilde{y}_{j-1}; \widetilde{y}_j] \rightarrow$	0 - 6	6 - 10	10 - 16	16 - 25
0.9 - 1.3	0.05	0.03	0.01	0.00
1.3 - 1.5	0.45	0.07	0.18	0.05
1.5 - 1.8	0.10	0.05	0.01	0.00

1. Bestimmen Sie die Randverteilung des Merkmals Y und stellen Sie dessen relative Häufigkeiten sowie die approximierende Verteilungsfunktion geeignet graphisch dar. Welche Annahme müssen Sie hierbei treffen?

2. Bestimmen Sie den Median der Häufigkeitsverteilung des Merkmals Y und interpretieren Sie diesen.

3. Wieviel Prozent der betrachteten Landkreise wiesen im Jahre 2008 eine Arbeitslosenqoute unter 15% auf?

4. Berechnen Sie die Werte der zweidimensionalen Verteilungsfunktion $F_{X,Y}(x,y)$ an den Klassengrenzen. Interpretieren Sie den Wert $F_{X,Y}(\widetilde{x}_2, \widetilde{y}_3)$.

5. Gegeben seien $\overline{y} = 6.625$, $\overline{x} = 1.413$, $s_X^2 = 0.0179$ und $\sum_{i=1}^{3} \sum_{j=1}^{4} \frac{\widetilde{x}_{i-1}+\widetilde{x}_i}{2} \frac{\widetilde{y}_{i-1}+\widetilde{y}_i}{2} f_{ij} = 9.3265$. Berechnen Sie den Korrelationskoeffizienten für die Merkmale X und Y.

6. Die Klassengrenzen sowie die entsprechenden relativen Häufigkeiten der Klassen des Merkmals X aus der Angabe wurden in **R** bereits angegeben:

```
> x=c(0.9,1.3,1.5,1.8)
> fi=c(0.09, 0.75, 0.16)
```

   Geben Sie die Befehle zur Bestimmung von $\overline{x}$ in **R** an?

Aufgabe 3

Eine Zufallsvariable X ist doppelt exponentialverteilt, d.h. die Dichte besitzt folgende Gestalt:

$$f_X(x) = \theta/2 \cdot e^{-\theta|x|} = \begin{cases} (\theta/2)e^{\theta x} & \text{für} \quad x \leq 0, \\ (\theta/2)e^{-\theta x} & \text{für} \quad x > 0. \end{cases}$$

1. Zeigen Sie, dass $f_X(x)$ tatsächlich eine Dichtefunktion darstellt.
2. Zeigen Sie, dass die Verteilungsfunktion $F_X(x)$ folgende Gestalt hat:

   $$F_X(x) = \begin{cases} 1/2e^{\theta x} & \text{für} \quad x \leq 0, \\ 1 - 1/2e^{-\theta x} & \text{für} \quad x > 0. \end{cases}$$

3. Skizzieren Sie für $\theta = 1$ die Dichtefunktion $f_X(x)$ für $-3 \leq x \leq 3$.
4. Zeigen Sie, dass die Log-Likelihoodfunktion folgende Gestalt besitzt:

   $$\ln L(\theta; x_1, x_2, \ldots, x_n) = n\ln(\theta) - n\ln(2) - \theta \sum_{i=1}^{n} |x_i|.$$

   Leiten Sie anschließend den ML-Schätzer für θ her, ohne Berechung der 2. Ableitung.
   Hinweis: Falls Sie kein Ergebnis erhalten haben, verwenden Sie im Folgenden:
   $\theta = \frac{n}{\sum_{i=1}^{n} |x_i|}$.
5. Aufgrund einer Stichprobe wissen Sie, dass $\sum_{i=1}^{10} |x_i| = 12$. Zeichnen Sie für $\theta = 10/13, 10/12, 10/11$ die Funktionswerte der Log-Likelihoodfunktion in ein geeignetes Koordinatensystem ein.
6. Erzeugen Sie eine Zahlenfolge `y` in **R** der Länge 1000 für $-3 \leq y \leq 3$.
7. Schreiben Sie in die Funktion

   ```
   > my.fun=function(x){...}
   ```

   die Dichtefunktion einer doppelten Exponentialverteilung zum Parameter $\theta = 1$.
   Hinweis: Sie können die **R**-Funktion `abs(x)` verwenden, die den Absolutbetrag des Arguments `x` zurück gibt.
8. Plotten Sie die Funktion `my.fun` für den Vektor `y`.

Aufgabe 4

Es erfolgt eine Zeitmessung beim 1000-Meter-Lauf bei acht untrainierten Jugendlichen. Nach einem sechsmonatigen Training laufen die Probanden wieder eine 1000-Meter-Strecke. Die erfasste Zeit in Minuten für beide 1000-Meter-Läufe ist in nachfolgender Tabelle präsentiert:

Proband i	1	2	3	4	5	6	7	8
Laufzeit vor dem Trainig x_i	4.37	2.55	4.88	3.35	3.85	3.12	2.97	4.25
Laufzeit nach dem Trainig y_i	3.20	2.50	4.10	3.02	2.86	2.90	2.80	3.50

Die Zufallsvariablen V: "**logarithmierte** benötigte Zeit beim 1000-Meter-Lauf vor dem Training" und W: "**logarithmierte** benötigte Zeit beim 1000-Meter-Lauf nach dem Training " sind in der Grundgesamtheit mit (μ_V,σ_V^2) und (μ_W,σ_W^2) verteilt.

1. Wie ist das Stichprobenmittel $\overline{W}_n$ für großen Stichprobenumfang n verteilt? Begründen Sie kurz Ihre Antwort.

Nun seien V und W normalverteilt: $V \sim N(\mu_V, \sigma_V^2)$ und $W \sim N(\mu_W, \sigma_W^2)$. Anhand erhobener Stichprobe wurde bereits errechnet, dass $\overline{v}_8 = 1.2783$, $\overline{w}_8 = 1.1242$ und $s_W^2 = 0.0231$.

2. Berechnen Sie die Stichprobenvarianz von V.
   Hinweis: Beachten Sie dabei, dass die Zufallsvariablen V und W logarithmierte Zeiten darstellen.

3. (a) Testen Sie die Vermutung, dass das sechsmonatige Training einen signifikanten Einfluss auf die durchschnittliche Leistung der Probanden hat, mit einer Irrtumswahrscheinlichkeit von 5%.
   Zusätzlich wurden die gepoolte Stichprobenvarianz $\overline{s}^2 = 0.0361$ sowie die Stichprobenvarianz der Differenzen $s_D^2 = 0.0391$ berechnet.
   (b) Bei welcher **tabellierten** Irrtumswahrscheinlichkeit würde sich die Testentscheidung umkehren? (Keine Berechnung erforderlich)

4. (a) Wie groß muss der Stichprobenumfang gewählt werden, damit das Stichprobenmittel $\overline{W}_n$ mit einer Vertrauenswürdigkeit von 95% maximal um 0.05 von dem wahren Mittelwert μ_W abweicht. Die Varianz von W sei bekannt und betrage $\sigma_W^2 = 0.0231$.
   (b) Interpretieren Sie das realisierte Konfidenzintervall $[1.0742; 1.1742]$ für den Mittelwert μ_W.

5. Es soll eine Funktion in **R** zur Bestimmung des Konfidenzintervalls für die Varianz bei unbekanntem Mittelwert erstellt werden.
   (a) Geben Sie den Befehl zur Erstellung einer Funktion *KI.Var* an, der das Command-Window aufruft.
   (b) Nachdem sich das Command-Window geöffnet hat, sehen Sie nachfolgendes Schema. Vervollständigen Sie dieses:

```
function(S^2, n, niveau)
{
alpha=1-niveau
...
return(KI)
}
```

Lösung zu Aufgabe 1

1. Gegeben ist $P(R) = 0.6,\ P(\overline{F}) = 0.3,\ P(R|F) = 0.8$

	R	$\bar{R}$	
F	0.56	0.14	0.7
$\bar{F}$	0.04	0.26	0.3
	0.6	0.4	1

   Hilfsrechnung: $P(R \cap F) = P(R|F) \cdot P(F) = 0.8 \cdot 0.7 = 0.56$

2. Berechnen Sie folgende W'keiten:
   (a) $P(R \cap \bar{F}) = 0.04$ (aus Tabelle ablesen)
   (b) $P(\bar{R}|F) = \frac{P(\bar{R} \cap F)}{P(F)} = \frac{0.14}{0.7} = 0.2$
   (c) $P(F|\bar{R}) = \frac{P(\bar{R} \cap F)}{P(R)} = \frac{0.14}{0.4} = 0.35$

3. $P(R \cap F) = 0.56 \neq 0.7 \cdot 0.6 = P(R) \cdot P(F)$ A: Nein, Sie sind stocahstisch abhängig.

4. (a) Einseitiger Hypothesestest für den Mittelwert mit bekannter Varianz :
   $H_0 : \mu \geq 0$ gegen $H_0 : \mu < 0$
   Kritischer Bereich:

$$\sqrt{n}\frac{\overline{X}_n - \mu_0}{\sqrt{\sigma^2}} < \lambda(0.05) = -\lambda(1 - 0.05)$$

   Aufstellen der Teststatistik:
   $t = \sqrt{180} \cdot (-0.15 - 0)/\sqrt{1.44} = -1.6770$
   Berechnung der kritischen Schranke:
   $KS = \lambda(0.05) = -1.645$
   Testentscheidung:
   Da $t < KS$ wird H_0 auf dem 5%-Signifikanzniveau abgelehnt.

   (b) Wert der PF heute: $S_0 = 50 \cdot 10 = 500$
   Erwarteter Wert in einem Jahr: $S_1 = S_0 \cdot e^{-0.14} = 434.6791$. Prozentualer Wert: $S_1/S_0 = 434.6791/500 = 0.8693582 \approx 86.94\%$

5. 1. Zeile: Die Teststatistik aus 5.a wird ausgerechnet.
   2. Zeile: Die kritische Schranke aus 5.a wird berechnet, also das 95%− Quantil einer Standardnormalverteilung.
   3.-6. Zeile: Es wird mit Hilfe einer `if` Schleife eine Testentscheidung durchgeführt: Falls die Statistik kleiner als die kritische Schranke ist, erscheint die Ausgabe: H0 wird abgelehnt, falls nicht, dann: H0 wird nicht abgelehnt. Es erscheint als Output: `H0 wird abgelehnt`.

Lösung zu Aufgabe 2

1. Tabelle:

i	von ... bis unter ... %	Δy_i	f_i	F_i	f_i^*	y_i
1	0 - 6	6	0.60	0.60	0.1000	3.00
2	6 - 10	4	0.15	0.75	0.0375	8.00
3	10 - 16	6	0.20	0.95	0.0333	13.00
4	16 - 25	9	0.05	1	0.0056	20.50

   Zeichnung: Histogramm für relative Häufigkeiten . Polygonzug für die Verteilungsfunktion. Annahme: Gleichverteilung in den Klassen.

2. (a) Bestimmung des Medians für klassierte Daten:

$$y_{(0.5)} = \widetilde{y}_0 + \frac{0.5 - F(\widetilde{y}_0)}{f_1} \cdot \Delta y_1 = 0 + \frac{0.5 - 0.00}{0.60} \cdot 6 = 5.$$

   (b) Interpretation: Die Arbeitslosenquote von 5 % wird in mindestens 50 % aller betrachteten Landkreisen nicht überschritten .

3. Approximierende Verteilungsfunktion:

$$F^*(15) = (0.75 + (15 - 10) \cdot 0.0333) = 0.9165,$$

   $\Rightarrow$ Bei 91.65 % aller betrachteten Landkreise wurden maximal eine 15% Arbeitslosenquote ermittelt.

4. Die zweidimensionale Verteilungsfunktion $F_{X,Y}(x, y) = f(X \leq x, Y \leq y)$ lautet:

$X \downarrow \backslash Y \rightarrow$	0 - 6	6 - 10	10 - 16	16 - 25
0.9 - 1.3	0.05	0.08	0.09	0.09
1.3 - 1.5	0.50	0.60	0.79	0.84
1.5 - 1.8	0.60	0.75	0.95	1.00

   Der Wert $F_{X,Y}(\widetilde{x}_2 = 1.5, \widetilde{y}_3 = 16) = 0.79$ bedeutet, dass 79% der betrachteten Landkreise eine Arbeitslosenquote höchstens 16 % und gleichzeitig eine durchschnittliche Kinderzahl je Frau maximal 1.5 aufweisen.

5. Korrelationskoeffizient für Merkmale X und Y: $r_{XY} = \frac{s_{XY}}{s_X s_Y}$

   - Berechnung der Varianz von Y :

$$s_Y^2 = \sum_{i=1}^{k} y_i^2 f_i - \overline{y}^2 = 9 * 0.6 + 64 * 0.15 + 169 * 0.2 + 420.25 * 0.05 - 6.625^2 = 25.9219$$

   - Kovarianz von X und Y :

$$s_{XY} = s_{XY} = \overline{x \cdot y} - \overline{x} \cdot \overline{y} = 9.3265 - 1.4130 * 6.625 = -0.0346$$

- Korrelationskoeffizient :

$$r_{XY} = \frac{-0.0346}{\sqrt{0.0179 \cdot 25.9219}} = -0.0508$$

Es besteht ein sehr geringer negativer Zusammenhang zw. den Merkmalen X und Y.

6. Lösung zu **R**-Aufgaben:

```
> xl=x[1:(length(x)-1)]      # Untergrenzen der Klasssen
> xu=x[2:(length(x))]        # Obergrenzen der Klassen
> xm=(xu+xl)/2               # Klassenmitten
> sum(xm*fi)                 # Mittelwert von X
```

Lösung zu Aufgabe 3

1. Prüfe

   (a) $f(x) \geq 0$ für alle $x \in \mathbb{R}$: Erfüllt, da $\theta > 0$ und die Exponentialfunktion immer größer als 0.

   (b)

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} \frac{\theta}{2} e^{-\theta|x|} dx \\
&= 1/2 \left(\int_{-\infty}^{0} \theta e^{\theta x} dx + \int_{0}^{\infty} \theta e^{-\theta x} dx \right) \\
&= 1/2 \left(e^{\theta x}\Big|_{-\infty}^{0} + (-e^{-\theta x}\Big|_{0}^{\infty}) \right) \\
&= 1/2(1 - 0 + (0 - (-1)) = 1
\end{aligned}$$

2. Lösung z.B. durch Integration über f_X:

   (a) für $x \leq 0$

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} \frac{\theta}{2} e^{\theta t} dt = 1/2 e^{\theta t}\Big|_{-\infty}^{x} = 1/2 e^{\theta x}$$

   (b) für $x > 0$

$$\begin{aligned}
P(X \leq x) = F(x) &= F(0) + \int_{0}^{x} \frac{\theta}{2} e^{-\theta t} dt \\
&= 1/2 + 1/2(1 - e^{-\theta t})\Big|_{0}^{x} = 1 - 1/2 e^{-\theta x}
\end{aligned}$$

3. Graph der Dichtefunktion $f(x)$ für $-3 \leq x \leq 3$

4. Aufstellen der Log-Likelihoodfunktion

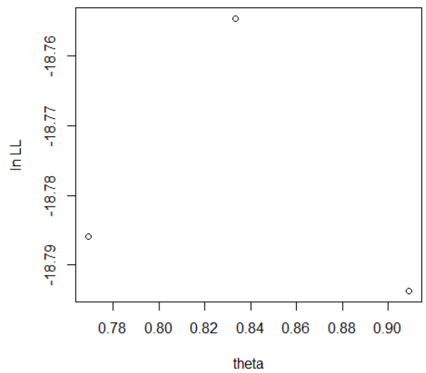
$$\begin{aligned}
L(\theta; x_i) &= \prod_{i=1}^{n} \frac{\theta}{2} e^{-\theta|x_i|} \\
&= \left(\frac{\theta}{2}\right)^n e^{-\sum_{i=1}^{n} \theta|x_i|} \\
&\Rightarrow lnL(\theta; x_i) = n(ln(\theta) - ln(2)) - \theta \sum_{i=1}^{n} |x_i|
\end{aligned}$$

   Bestimmung des Maximums mit Hilfe der ersten Ableitung:

$$\begin{aligned}
\frac{\partial lnL(\theta; x_i)}{\partial \theta} &= n/\theta - \sum_{i=1}^{n} |x_i| \stackrel{!}{=} 0 \\
&\Leftrightarrow \theta = \frac{n}{(\sum_{i=1}^{n} |x_i|)}
\end{aligned}$$

5. Graph von LL: Wertetabelle:

θ	10/13	10/12	10/11
$LL(\theta)$	-18.790	-18.755	-18.793



6. `> y=seq(-3,3,length=1000)`
7. `> my.fun=function(x){1/2*exp(-abs(x))}`
8. `> plot(y, my.fun(y))`

Lösung zu Aufgabe 4

1. Approximative Verteilung des Stichprobenmittels:

   $$\overline{W}_n \sim N\left(\mu_W; \frac{\sigma_W^2}{n}\right)$$

   Begründung: zentraler Grenzwertsatz.

2. Berechnung der Stichprobenvarianz von W:

   (a) Logarithmieren der angegebenen Zeiten:

Proband i	1	2	3	4	5	6	7	8
Laufzeit x_i	4.37	2.55	4.88	3.35	3.85	3.12	2.97	4.25
Log. Zeit v_i	1.4748	0.9361	1.5851	1.209	1.3481	1.1378	1.0886	1.4469

   (b) $s_V^2 = \frac{1}{8-1}\left((1.4748-1.2783)^2 + (0.9361-1.2783)^2 + \ldots + (1.4469-1.2783)^2\right) = 0.0491$

3. (a) Mittelwertdifferenztest bei verbundenen Stichproben :

   $$H_0 : \mu_V - \mu_W \leq 0 \text{ gegen } H_A : \mu_V - \mu_W > 0$$

   Kritischer Bereich:

   $$\sqrt{n}\frac{\overline{V}_n - \overline{W}_n - \delta_0}{S_D} > t_{1-\alpha;n-1}$$

   Teststatistik (aus der Angabe $S_D^2 = 0.0391$):

   $$t = \sqrt{8}\frac{1.2783 - 1.1242 - 0}{\sqrt{0.0391}} = 2.2042$$

   Kritische Schranke $t_{1-\alpha;n-1} = t_{0,95;7} = 1.895$
   Testentscheidung: Da $t = 2.2042 > 1.895 = t_{0.95;7}$, wir H_0 abgelehnt.

   (b) Die Testentscheidung kehrt sich erst bei $\alpha = 0.03$ um.

4. (a) KI für den Mittelwert bei bekannter Varianz, hierbei muss $\lambda_{1-\alpha/2} \cdot \frac{\sigma_W}{\sqrt{n}} \leq 0.05$ sein, d.h., $\frac{\lambda_{0.975}\cdot\sqrt{0.0231}}{\sqrt{n}} \leq 0.05$, wobei $\lambda_{0.975} = 1,96, \Rightarrow \quad n \geq (\frac{1.96}{0,05})^2 \cdot 0.0231 = 35.5, \Rightarrow n \geq 36$.

   (b) Mit einer Vertrauenswürdigkeit von 95% liegt der wahre Mittelwert im realisierten KI. **ODER:** Wenn man viele Stichproben zieht und die KI für den Mittelwert berechnet, dann fällt in 95 von 100 gezogenen Stichproben der wahre Parameter in das realisierte KI.

5. a)
```
> KI.Var=fix(KI.Var)
```
   b)
```
function(S^2, n, neveau))
{ alpha=1-niveau
KI_l=(n-1)S^2/qchisq(1-alpha/2,n-1)
KI_u=(n-1)S^2/qchisq(alpha/2,n-1)
KI=c(KI_l, KI_u)
return(KI)}
```