Aufgabe 1

Die Befragung von 100 mittelständischen Dienstleistungsunternehmen einer Metropolregion nach ihren *Umsätzen* sowie *Werbeausgaben* im Jahre 2008 erbrachte folgende Häufigkeitsfunktion:

$(\widetilde{x}_{i-1};\widetilde{x}_i]\ \downarrow \backslash (\widetilde{y}_{j-1};\widetilde{y}_j] \rightarrow$	20 - 40	40 - 55	55 - 80	80 - 100
1 - 2	0.12	0.16	0.10	0.02
2 - 4	0.03	0.07	0.14	0.09
4 - 5	0.00	0.03	0.10	0.14

Wobei das Merkmal X : "Werbeausgaben im Jahre 2008 *von... bis...* in Mio. EUR" und das Merkmal Y :"Umsatz im Jahre 2008 *von... bis...* in Mio. EUR" darstellen.

1. Berechnen Sie die zweidimensionale Verteilungsfunktion $F_{X,Y}(x,y)$ und interpretieren Sie den Wert $F_{X,Y}(X=x_3, Y=y_1)$.

2. Bestimmen Sie die Randverteilung des Merkmals *Umsatz* und stellen Sie dessen relative Häufigkeiten geeignet graphisch dar. Welche Annahme müssen Sie hierbei treffen?

Gehen Sie im Folgenden von der Gültigkeit der von Ihnen getroffenen Annahme aus.

3. Bestimmen Sie den Modus und den Median der Häufigkeitsverteilung des Merkmals *Umsatz*.

4. Wieviel Prozent der befragten Unternehmen erwirtschafteten einen Umsatz über 50 Mio. EUR?

5. Gegeben seien $\overline{x} = 2.805$ und $\sum_{i=1}^{3}\sum_{j=1}^{4} \frac{\widetilde{x}_{i-1}+\widetilde{x}_i}{2}\frac{\widetilde{y}_{i-1}+\widetilde{y}_i}{2} f_{ij} = 188.4375$. Berechnen Sie die Kovarianz von X und Y.

6. Die Werte der Merkmale U und V seien in **R** als Vektoren u und v definiert. Was wird in der folgenden Befehlreihenfolge in **R** berechnet?

```
> n=length(u)
> A=(sum(u*v)/n-mean(u)*mean(v))
> B=sum(u^2)/n-mean(u)^2
> C=sum(v^2)/n-mean(v)^2
> ruv=A/sqrt(B*C)
```

Aufgabe 2

Seien Z_i, für $i = 1, ..., n$, n stochastisch unabhängige, identisch Poisson-verteilte Zufallsvariablen, also

$$Z_i \overset{iid}{\sim} Poi(\lambda), \quad i = 1, ..., n$$

und sei $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

1. Geben Sie die exakte Verteilung von $\sum_{i=1}^{n} Z_i$ an und zeigen Sie damit, dass $P(\overline{Z}_n \leq z) = F_{Poi}(z \cdot n; n \cdot \lambda)$.

   Hinweis:
   Aus $X \sim Poi(\lambda_X)$ und $Y \sim Poi(\lambda_Y)$ und X,Y stochastisch unabhängig folgt $X + Y \sim Poi(\lambda_X + \lambda_Y)$.

2. Prüfen Sie die Bedingungen des zentralen Grenzwertsatzes. Wie ist $\overline{Z}_n$ demnach approximativ verteilt? Geben Sie die Parameter dieser Verteilung in Abhängigkeit von λ an.

Sei nun n=10 und $\lambda = 1$.

3. Bestimmen Sie die fehlenden Werte für $P(\overline{Z}_{10} \leq z)$ in folgender Tabelle einmal exakt und einmal mit der approximativen Verteilung.

z	0.00	0.10	0.15	0.50	1.00	2.00
$P(\overline{Z}_{10} \leq z)$ (exakte Verteilung)	0.0000	-	-	-	0.5830	0.9984
$P(\overline{Z}_{10} \leq z)$ (approx. Verteilung)	0.0008	0.0022	0.0036	-	-	-

4. Skizzieren Sie die Verteilungsfunktion der exakten und der approximativen Verteilung im Intervall von $[0; 2]$.

5. Geben Sie die Befehle an, um die entsprechende Zeichnung in **R** zu erstellen. Verwenden Sie folgendes Objekt x für die x-Achse.

```
> x = seq(0,2,,10^4)
```

Sei eine Stichprobe der Länge 10 aus einer Poisson-verteilten Grundgesamtheit mit unbekanntem λ gegeben.

6. Bestimmen Sie die exakte Wahrscheinlichkeit, dass das Stichprobenmittel um mehr als 0.4 vom wahren Mittelwert abweicht, unter der Annahme, dass die Grundgesamtheit Poisson-verteilt mit $\lambda = 1$ ist.

Das Stichprobenmittel $\bar{z}_{10}$ betrug 1.4.

7. Für welches Hypothesenpaar wäre die eben errechnete Wahrscheinlichkeit der p-Wert?

8. Treffen Sie eine entsprechende Testentscheidung nur anhand dieser Wahrscheinlichkeit bei einer Irrtumswahrscheinlichkeit von 5 %.

   Hinweis:
   Wenn Sie diese Teilaufgabe sonst nicht lösen können, dann geben Sie alternativ allgemein die Entscheidungsregel auf Basis von p-Werten an.

Aufgabe 3

Bei einem Zigarettenautomaten wurden folgende 56 Wartezeiten, also die Zeit zwischen einem Verkauf und dem nächsten, beobachtet:

Wartezeit	bis 5 Min.	von 5 bis 10 Min.	von 10 bis 15 Min.	ab 15 Min.
Anzahl	12	23	4	17

1. Kann mit Hilfe dieser Angabe die mittlere beobachtete Wartezeit exakt bestimmt werden? Begründen Sie Ihre Antwort kurz.

Nehmen Sie an, die Wartezeit W [in Minuten] bei diesem Automaten folge einer Exponentialverteilung mit Parameter $\lambda = \frac{1}{9}$.

2. Wie und mit welchen Parametern ist dann die Zufallsvariable A "Anzahl der Verkäufe in einer Minute" verteilt?

3. Mit wie vielen Verkäufen könnte demnach im Laufe eines Tages gerechnet werden?

4. Welche sachlichen Annahmen muss man treffen, damit die Wartezeit exponentialverteilt ist?

5. Bestimmen Sie die Wahrscheinlichkeiten dafür, dass eine Wartezeit in obigen Klassen (also: "bis 5 Min.","von 5 bis 10 Min.","von 10 bis 15 Min.","ab 15 Min.") auftritt.

6. Welche Häufigkeiten wären für diese Klassen zu erwarten gewesen?

7. Testen Sie mit einer Irrtumswahrscheinlichkeit von 4%, ob die getroffene Verteilungs- und Parameterannahme für die Wartezeit zu den Beobachtungen passt.

8. Welche Annahme könnte im gegebenen Sachverhalt verletzt sein? Geben Sie ein sachliches Beispiel.

Gegeben sei ein **R**-Workspace mit den Klassenhäufigkeiten als Objekt *ni*:

```
> ni
[1] 12 23  4 17
```

9. Geben Sie für jede der folgenden Befehlszeilen an, was berechnet wird.

```
> og=c(0,5,10,15,Inf)
> p=diff(pexp(og,1/9))
> n=sum(ni)
> sum((ni-n*p)^2/(n*p))
```

Aufgabe 4

Eine Warenteststiftung möchte den Wasserverbrauch pro Waschgang von Waschmaschinen im Standard- sowie im Sparmodus untersuchen. Dazu werden acht Waschmaschinen eines vergleichbaren Modells unterschiedlicher Hersteller getestet. Die Ergebnisse sind in nachfolgender Tabelle präsentiert:

Waschmaschine i	1	2	3	4	5	6	7	8
Wasserverbr. im Standardmodus in Liter x_i	45.5	48.9	50.6	53.7	49.4	55.2	60.0	52.7
Wasserverbr. im Sparmodus in Liter y_i	43.5	42.7	45.8	45.4	50.5	45.1	55.4	47.6

Es sei bekannt, dass die Zufallsvariablen X:"Wasserverbrauch pro Waschgang im Standardmodus in Liter" und Y: "Wasserverbrauch pro Waschgang im Sparmodus in Liter" normalverteilt sind mit (μ_X, σ_X^2) und (μ_Y, σ_Y^2).
Es wurde bereits errechnet, dass $\overline{x}_8 = 52$, $\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = 138.2$, $\overline{y}_8 = 47$ und $s_Y^2 = 19.84$ sind.

1. Geben Sie die Verteilung des Stichprobenmittels $\overline{X}_n$ an. Begründen Sie Ihre Antwort.
2. Bestimmen Sie die Stichprobenvarianz des Wasserverbrauchs im Standardmodus.
3. Waschmaschinenhersteller behaupten, dass mann im Sparmodus durchschnittlich mehr als 7 Liter Wasser pro Waschgang sparen kann. Testen Sie die Hypothese der Warenteststiftung, dass die tatsächliche durchschnittliche Ersparnis unter 7 Litern liegt, mit einer Irrtumswahrschenilichkeit von 5 %.
4. Es sei bekannt, dass $\sigma_X^2 = 20.00$ Liter^2 beträgt. Berechnen Sie das 96 % Konfidenzintervall für den Mittelwert von X. Wie lässt sich das realisierte Konfidenzintervall interpretieren?
5. Geben Sie den Befehl in **R** an, um den kritischen Wert aus der Teilaufgabe 3 zu bestimmen?
6. Die Daten für x und y aus obiger Tabelle wurden in **R** als Vektoren x und y eingelesen. Bestimmen Sie explizit (ohne Verwendung des Befehls *cov*) die Kovarianz von x und y in **R**.

Lösung zu Aufgabe 1

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

$X \downarrow \backslash Y \rightarrow$	20 - 40	40 - 55	55 - 80	80 - 100
1 - 2	0.12	0.28	0.38	0.4
2 - 4	0.15	0.38	0.62	0.73
4 - 5	0.15	0.41	0.75	1.00

   Der Wert $F_{X,Y}(X = x_3, Y = y_1) = 0.15$ bedeutet, dass 15% der befragten Unternehmen höchstens 40 bzw. von 20 bis 40 Mio. EUR erwirtschafteten.

2. Tabelle:

i	von ... bis unter ... Mio. EUR	Δy_i	f_i	F_i	f_i^*	y_i
1	20 - 40	20	0.15	0.15	0.0075	30.00
2	40 - 55	15	0.26	0.41	0.0173	47.50
3	55 - 80	25	0.34	0.75	0.0136	67.50
4	80 - 100	20	0.25	1	0.0125	90.00

   Zeichnung: Histogramm. Annahme: Gleichverteilung in den Klassen.

3. (a) Modus: $f_2^* = 0.0173 \geq f_i^*, i = 1, \ldots, 4 \quad \Rightarrow$ Klasse 2 ist die modale Klasse.

   (b) Bestimmung des Medians für klassierte Daten:

$$x_{(0.5)} = \widetilde{x}_2 + \frac{0.5 - F(\widetilde{x}_2)}{f_3} \cdot \Delta x_3 = 55 + \frac{0.5 - 0.41}{0.34} \cdot 25 = 61.6176.$$

4. Approximierende Verteilungsfunktion:

$$F^*(100) - F^*(50) = 1 - (0,15 + (50 - 40) \cdot 0,0173) = 1 - 0.323 = 0.677,$$

   $\Rightarrow$ 67.7% aller befragten Unternehmen erwitschafteten über 50 Mio. EUR.

5.
   - Berechnung des Mittelwerts von Y:

$$\bar{y} = \sum_{i=1}^{k} y_i f_i = 30.00 \cdot 0.15 + 47.50 \cdot 0.26 + 67.50 \cdot 0.34 + 90.00 \cdot 0.25 = 62.30$$

   - Kovarianz von X und Y: $s_{XY} = \overline{x \cdot y} - \overline{x} \cdot \overline{y} = 188.4375 - 2.805 \cdot 62.30 = 13.686$

6. Lösung zu **R**-Aufgaben:

```
> n=length(u)                       # L\"{a}nge des Vektors u
> A=(sum(u*v)/n-mean(u)*mean(v)) # Kovarianz von u und v
> B=sum(u^2)/n-mean(u)^2            # Varianz von u bzw.
> C=sum(v^2)/n-mean(v)^2            # Varianz von v
> ruv=A/sqrt(B*C)                   # Korrelationskoeffizient f\"{u}r u und v
```

Lösung zu Aufgabe 2

1. $\sum_{i=1}^{n} Z_i \sim Poi(\lambda_1 + \lambda_2 + ... + \lambda_n = n\lambda)$ und damit

$$P(\overline{Z}_n \leq z) = P(n\overline{Z}_n \leq z \cdot n) = P(\sum_{i=1}^{n} Z_i \leq z \cdot n) = F_{Poi}(z \cdot n; n \cdot \lambda).$$

2. Zentraler Grenzwertsatz

   - stochastische Unabhängigkeit ✓
   - identische Verteiltheit ✓
   - endlicher Mittelwert μ_Z und endliche Varianz σ_Z^2 ✓

   Für große n gilt dann

$$\overline{Z_n} \overset{approx}{\sim} N(\lambda; \lambda/n).$$

3. als Tabelle:

z	0.00	0.10	0.15	0.50	1.00	2.00
$F_{Poi}(10 \cdot z, 10)$	0.0000	**0.0005**	**0.0005**	**0.0671**	0.5830	0.9984
$F_N(z, 1, 0.1)$	0.0008	0.0022	0.0036	**0.0571**	**0.5000**	**0.9992**

   mit $F_N(0, 1, 0.1) = \Phi(\frac{0-1}{\sqrt{(0.1)}}) \approx \Phi(-3.16) = 1 - \Phi(3.16) = 1 - 0.9992 = 0.0008$, etc.

4. Zeichnung (Verwenden Sie den R-Code unten, um die Zeichnung zu erzeugen.)

5.
```
> plot(x,ppois(x*10,10))
> lines(x,pnorm(x,1,sqrt(.1)))
```

6. $P(|\overline{Z}_n - \lambda| > 0.4) = P(\overline{Z}_n \leq 1 - 0.4) + 1 - P(\overline{Z}_n \leq 1 + 0.4) =$
   $F_{Poi}(0.6 \cdot 10, 1 \cdot 10) + 1 - F_{Poi}(1.4 \cdot 10, 1 \cdot 10) = 0.2136$

7. Wahrscheinlichkeit aus TA 6 entspricht p-Wert für $H_0 : \lambda = 1$ vs. $H_A : \lambda \neq 1$.

8. Dieser ist bei gegebener Stichprobe $21.36\% > 5\%$, lehne daher H_0 nicht ab.

Lösung zu Aufgabe 3

1. Nein, denn die Verteilung innerhalb der Klassen ist nicht bekannt.

2. $A \sim Poi(\lambda)$, mit $\lambda = \frac{1}{9}$

3. $24 \cdot 60 \cdot E(A) = 160$

4. Gleiche Annahmen wie bei Poisson-Verteilung:

   - Wahrscheinlichkeit für das Auftreten eines Ereignis ist proportional zur Länge des betrachteten Intervals und unabhängig von dessen Lage.
   - Die Wahrscheinlichkeit von mehr als einem Ereignis in einem sehr kleinen Intervall ist vernachlässigbar klein.
   - Die Anzahlen der Ereignisse in zwei disjunkten Intervallen sind voneinander unabhängig.

5. $P(w_u < W < w_o) = F_{exp}(w_o; \frac{1}{9}) - F_{exp}(w_u; \frac{1}{9}) = e^{-\frac{1}{9}w_u} - e^{-\frac{1}{9}w_o}$, also

$w_u; w_o$	0; 5	5; 10	10; 15	15; ∞
$P(w_u < W < w_o)$	0.4262	0.2446	0.1403	0.1889

6. $E(n_i) = p_i n = n \cdot P(w_u < W < w_o)$, also

$w_u; w_o$	0; 5	5; 10	10; 15	15; ∞
$E(n_i)$	23.87	13.70	7.86	10.58

7. χ^2-Anpassungstest

   - Hypothese: $H_0 : W \sim Exp(\frac{1}{9})$
   - Prüfgröße: $\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2(k-1)$
   - Ablehnungsbereich: $\chi^2 > \chi^2_{k-1;0.96} = 8.31$
   - Approximationsregeln erfüllt, $\Rightarrow$ Befund:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = \frac{(12 - 23.87)^2}{23.87} + \frac{(23 - 13.70)^2}{13.70} + ... \approx 18.01$$

   - H_0 bei Irrtumswahrscheinlichkeit von 4% ablehnen.

8. Angenommen bei Regenwetter kauft niemand Zigaretten.$\Rightarrow$ Keine Unabhängigkeit von der Lage des Intervalls.

9. Pro Zeile:

   - Erstellt ein Objekt (vector) *og* mit den Obergrenzen der Klassen.
   - Berechnet einen Vektor mit den Klassenwahrscheinlichkeiten.
   - Erstellt ein Object (numeric) n mit der Stichprobenlänge.
   - Berechnet die Prüfgröße zu Test aus Teilaufgabe 6.

Lösung zu Aufgabe 4

1. Das Stichprobenmittel ist aufgrund der Reproduktivität der Normalverteilung ebenfalls normalverteilt mit

$$\text{Mittelwert:} \quad E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i)\right) = \frac{1}{n}\sum_{i=1}^{n}E(X_i) = \frac{1}{n}n\mu_X = \mu_X,$$

$$\text{Varianz:} \quad VAR(\bar{X}_n) = VAR\left(\frac{1}{n}\sum_{i=1}^{n}(X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}VAR(X_i) = \frac{1}{n^2}n\sigma_X^2 = \sigma_X^2/n$$

2. $\bar{s}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{8-1}138.2 = 19.7429$

3. Mittelwertdifferenztest bei verbundenen Stichproben.

   $H_0: \ \mu_X - \mu_Y \leq 7 \quad \text{gegen} \quad H_A: \mu_X - \mu_Y > 7$

   Kritischer Bereich: $\sqrt{n}\dfrac{\overline{X}_n - \overline{Y}_n - \delta_0}{S_D} > t_{1-\alpha;n-1}.$

   Teststatistik: $\sqrt{n}\dfrac{\overline{X}_n - \overline{Y}_n - \delta_0}{S_D} = \sqrt{8}\dfrac{(-2)}{3.4797} = -1.6257,$ wobei

$$\begin{aligned} S_D^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2 \text{ und } \overline{D}_n = \overline{X}_n - \overline{Y}_n = 5. \\ S_D^2 &= \frac{1}{7}((45.5 - 43.5 - 5)^2 + \cdots + (52.7 - 47.6 - 5)^2) = \\ &= 1/7((2-5)^2 + (6.2-5)^2 + (4.8-5)^2 + (8.3-5)^2 + (-1.1-5)^2 + (10.1-5)^2 \\ &+ (4.6-5)^2 + (5.1-5)^2) = 84.76/7 = 12.1086. \end{aligned}$$

   Kritische Schranke: $t_{1-\alpha;n-1} = t_{0.95;7} = 1.895$. **Testentscheidung**: Da $t = -1.6257 \not> 1.895 = -t_{0.95;7}$, kann H_0 nicht abgelehnt werden.

4. Das theortische Konfidenzintervall für den Mittelwert bei bekannter Varianz:

$$P\left(\overline{X}_n - \lambda_{1_\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_X \leq \overline{X}_n + \lambda_{1_\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

   $\alpha = 0.04, \ \lambda_{1-0.02} = 2.0537.$
   Das realisierte Konfidenzintervall:

$$KI = \left[52 - 2.0537\frac{\sqrt{20}}{\sqrt{8}}; 52 + 2.0537\frac{\sqrt{20}}{\sqrt{8}}\right] = [48.7528; 55.2472]$$

   Interpretation: Der wahre Mittelwert liegt mit einer **Vertrauenswürdigkeit** von 96% im realisierten KI. /**Alternativ:** Zieht man sehr viele einfache Stichptoben des Umfangs n aus der Grundgesamtheit, dann gibt $\gamma = 1 - \alpha$ den Anteil der Stichproben an, in denen der wahre Parameter in das realisierte KI fällt.

5. (a) `> qt(0.95, 7)`
   (b) 1/(length(x)-1)*sum((x-mean(x))*(y-mean(y)))