Aufgabe 1

150 Personen gaben bei einer Befragung an, wie viel Geld sie in diesem Jahr für Weihnachtsgeschenke ausgegeben haben. Die Ergebnisse der Befragung sind in nachfolgender Tabelle zusammengefasst, wobei das Merkmal X : "Für die Geschenke ausgegebener Betrag in EUR" ist:

i	Ausgegebener Betrag (X_i) in EUR (von / bis unter)	Anzahl der Befragten $n(X_i)$
1	50 - 55	10
2	55 - 70	50
3	70 - 80	50
4	80 - 85	33
5	85 - 100	7

1. Nennen Sie den für das Merkmal X geeigneten Skalentyp und begründen Sie Ihre Antwort.

2. Bestimmen Sie die relativen Häufigkeiten und die Verteilungsfunktion von X. Stellen Sie die approximierende Verteilungsfunktion von X geeignet graphisch dar und gehen Sie kurz auf die Annahme ein, die Sie hierbei treffen müssen.

3. Bestimmen Sie den Modus und den Median der Häufigkeitsverteilung von X und interpretieren Sie den Median.

4. Durchschnittlich gaben die Befragten für die Weihnachtsgeschenke $71,80$ EUR aus. Berechnen Sie die Varianz des Merkmals X.

5. Geben Sie die Befehle in **R** an, um die Zahlenreihe 52.5, 62.5, 75, 82.5, 92.5 als Vektor b zu definieren und den Mittelwert von b zu bestimmen.

6. Geben Sie eine Formel zur Berechnung der Varianz von b in **R** an, ohne die Befehle var() oder sd() zu verwenden.

Aufgabe 2

Bei der Produktion von LCD-Bildschirmen treten geringe Anzahlen von Pixelfehlern unvermeidlich auf. Es werden 100 zufällig ausgewählte LCD-Bildschirme eines Herstellers auf Pixelfehler untersucht. Die Ergebnisse der Untersuchung sind in nachfolgender Tabelle erfasst:

i	1	2	3	4	5	6
Anzahl der Pixelfehler pro Bildschirm x_i	0	1	2	3	4	≥ 5
Anzahl der Bildschirme n_i	14	32	22	11	14	7

Sei X: "Anzahl der Pixelfehler pro Bildschirm" Poisson-verteilt mit dem unbekannten Parameter λ.

1. Geben Sie an, wie und mit welchen Parametern hier das Stichprobenmittel $\overline{X}$ approximativ verteilt ist?

2. Schätzen Sie den Parameter der Poissonverteilung λ nach der Methode der Momente. Nehmen Sie dabei an, dass bei jedem der sieben Bildschirme aus der Klasse 6 genau 5 Pixelfehler festgestellt wurden.
   Hinweis: Verwenden Sie im Folgenden $\widehat{\lambda} = 2.0$, wenn Sie oben kein Ergebnis erhalten haben.

3. Überprüfen Sie die Hypothese $X \sim Poi(\lambda)$ anhand eines χ^2-Anpassungstests mit einer Irrtumswahrscheinlichkeit von 1%.

4. Nehmen Sie Stellung zu folgenden Aussagen. Schreiben Sie eine kurze Begründung, wenn Sie die Aussagen als falsch beurteilen:

   (a) Der Fehler 1. Art heißt auch α-Fehler und bedeutet, die Nullhypothese nicht abzulehnen, wenn sie falsch ist.

   (b) Der Zentrale Grenzwertsatz besagt, dass wenn $X_1, \cdots X_n$ stochastisch unanhängige, in der Grundgesamtheit mit μ_X und σ_X^2 verteilte Zufallsvariablen sind, dann ist das Stichprobenmittel für großen Stichprobenumfang n annähernd genau so verteilt wie die Grundgesamtheit von X.

   (c) Bei der Maximum-Likelihood-Schätzung wird der Parameter θ_{ML} so gewählt, dass dem Stichprobenbefund die höchste " Glaubwürdigkeit" zugeordnet wird.

5. Die Zufallsvariable Y sei normalverteilt mit $\mu_Y = 8$ und $\sigma_Y^2 = 25$. Geben Sie einen Befehl in **R** zur Berechnung der Wahrscheinlichkeit $P(Y \leq 15)$ an.

6. Die Zufallsvariable Z sei Poisson-verteilt mit dem Parameter $\lambda = 2.0$. Geben Sie die Befehle in **R** an, um die Wahrscheinlichkeit $P(Z = 4)$ und das 95% Quantil $z_{0.95}$ zu bestimmen.

Aufgabe 3

Die Befragung von 100 Teilnehmern einer Statistik-Klausur erbrachte folgende Häufigkeitsfunktion:

$X \downarrow \backslash Y \rightarrow$	0	1	2	3
1	0.04	0.06	0.1	0.16
2	0.06	0.22	0.1	0.02
3	0.14	0.04	0.04	0.02

Wobei das Merkmal X : "Note in der Statistik-Klausur" und das Merkmal Y : "Anzahl der besuchten R-Tutorien in Statistik" darstellen.

1. Bestimmen Sie die Randverteilung von der Klausurnote und geben Sie deren Median an.

2. Berechnen Sie die zweidimensionale Verteilungsfunktion $F_{X,Y}(x, y)$ und interpretieren Sie den Wert $F_{X,Y}(1, 3)$.

3. Bestimmen Sie die Häufigkeitsfunktion und den Mittelwert der Klausurnote unter der Bedingung, dass höchstens zwei R-Tutorien besucht wurden. Interpretieren Sie den Wert der bedingten Häufigkeitsfunktion an der Stelle $X = 1$.

4. Sind die beiden Merkmale X und Y statistisch unabhängig? Erklären Sie die statistische Unabhängigkeit am Sachverhalt.

5. Berechnen Sie den Mittelwert und die Varianz von X.

6. Gegeben sei $\bar{Y} = 1.4$ und $\overline{X \cdot Y} = 2.24$. Berechnen Sie die Kovarianz von X, Y.
   Hinweis: Wenn Sie in der Teilaufgabe 5 kein Ergebnis erhalten haben, verwenden Sie für $\bar{X} = 1.9$.

7. Sie haben bereits die Ausprägungen eines Merkmals Z und deren Häufigkeiten in **R** eingegeben:

   ```
   > Z=c(5, 7, 18, 28, 35)
   > n=c(15, 16, 12, 24, 4)
   ```

   Geben Sie die Befehle in **R** an, um den Modus und den Mittelwert von Z zu bestimmen.

Aufgabe 4

Ein Golfspieler möchte den Durchmesser von Golfbällen untersuchen. Dazu hat er von zwei unterschiedlichen Herstellern "Mygolf" und "Akin-Golf" eine Stichprobe gezogen. Die Ergebnisse sind in nachfolgender Tabelle erfasst:

Beobachtung i	1	2	3	4	5	6	7	8	9	10
Mygolf x_i	42.50	42.82	42.67	42.53	42.78	42.69	42.99	42.86	42.79	42.84
Akin-Golf y_i	42.75	42.50	42.49	42.69	42.75	42.55	42.60			

Es sei bekannt, dass die Zufallsvariablen X ("Durchmesser des Golfballs von Mygolf in mm") und Y ("Durchmesser des Golfballs von Akin-Golf in mm") normalverteilt sind mit (μ_X,σ_X^2) und (μ_Y,σ_Y^2).
Es wurde bereits errechnet, dass $\bar{x}_{10} = 42.747$, $s_X^2 = 0.0229$, $\bar{y}_7 = 42.6186$ und $s_Y^2 = 0.0126$.

1. Geben Sie die Verteilung des Stichprobenmittels $\bar{Y}_7$ an.

2. Es sei bekannt, dass der Durchmesser eines Golfballs bei mindestens 42.67 mm liegen soll.

   (a) Testen Sie die Vermutung, dass der Durchmesser des Golfballs von Akin-Golf im Durchschnitt diesen Wert unterschreitet, mit einer Irrtumswahrscheinlichkeit von 5%.

   (b) Bei welcher **tabellierten** Irrtumswahrscheinlichkeit würde sich die Testentscheidung umkehren?
   Hinweis: Keine Berechnung erforderlich. Beantworten Sie die Frage mit Hilfe der geeigneten Tabelle.

3. Testen Sie die Vermutung, dass die Varianz von X größer als die Varianz von Y ist, mit einer Irrtumswahrscheinlichkeit von 10%.

4. Nun sei bekannt, dass die Varianzen der beiden Zufallsvariablen gleich sind und das realisierte 96% Konfidenzintervall für die Differenz der Mittelwerte [-0.0235; 0.2803] beträgt. Interpretieren Sie dieses Ergebnis. Geben Sie das geeignete theoretische Konfidenzintervall an. (Keine Berechnung erforderlich.)

5. Nehmen Sie Stellung zu folgender Aussage. Begründen Sie kurz ihre Antwort, wenn Sie die Aussage als falsch beurteilen.

   (a) Der p-Wert gibt an, wie groß das Testniveau gewählt werden müsste, damit H_0 gerade noch akzeptiert würde. Wenn der p-Wert des Tests größer ist als das Niveau α, so kann H_0 nicht verworfen werden.

6. Sie definieren zuerst Vektoren $x1$ und $y1$ in **R**. Anschließend führen Sie einen Test mit folgendem Output durch: (Siehe folgende Seite)

```
> x1=c( 42.5,42.82,42.67,42.53,42.78,42.69,42.99,42.86,
  42.79, 42.84)
> y1=c(42.75,42.50,42.49,42.69,42.75,42.55,42.60)
> t.test(x1, y1, alternative="greater",mu=0, paired=FALSE, var.equal=TRUE,
  conf.level=0.95)
    Two Sample t-test
data:  x1 and y1
t = 1.9031, df = 15, p-value = 0.0382
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
0.01012285        Inf
sample estimates:
mean of x mean of y
42.74700  42.61857
```

Beschreiben Sie und interpretieren Sie den Output aus **R**:

- Welcher Test wurde angewendet?
- Nennen Sie die Null- und Alternativhypothese.
- Wie lautet die Testentscheidung? Begründen Sie Ihre Antwort.

.

Lösung zu Aufgabe 1

1. Geeignet für das Merkmal X ist eine Verhältnisskala, da das Merkmal einen natürlichen Nullpunkt jedoch keine natürliche Einheit besitzt.

2. Tabelle:

i	von ... bis unter ... EUR	$n(X_i)$	Δx_i	f_i	F_i	f_i^*	x_i
1	50 - 55	10	5	0.067	0.067	0.0134	52.50
2	55 - 70	50	15	0.333	0.4	0.0222	62.50
3	70 - 80	50	10	0.333	0.733	0.0333	75.00
4	80 - 85	33	5	0.22	0.953	0.044	82.50
5	85 - 100	7	15	0.047	1.000	0.003	92.50

   Zeichnung: Polygonzug für die empirische Verteilungsfunktion. Annahme: Gleichverteilung in den Klassen.

3. (a) Modale Klasse: $f_4^* = 0.044 \geq f_i^*, i = 1, \ldots, 4, \quad \Rightarrow$ Klasse 4 ist die modale Klasse.

   (b) Bestimmung Median für klassierte Daten:

   $$x_{(0.5)} = \widetilde{x}_2 + \frac{0.5 - F(\widetilde{x}_2)}{f_3} \cdot \Delta x_3 = 70 + \frac{0.5 - 0.4}{0.333} \cdot 10 = 73.00,$$

   d.h. mindestens 50% der Befragten haben höchstens 73.00 EUR ausgegeben.

4. Berechnung der Varianz von X:

   $$s_x^2 = \overline{x^2} - \overline{x}^2 = 52.50^2 * 0.067 + 62.50^2 * 0.333 + 75.00^2 * 0.333 +$$

   $$+ 82.50^2 * 0.22 + 92.50^2 * 0.047 - 71.80^2 = 5258.094 - 5155.24 = 102.854$$

5. Lösung zu **R**-Aufgaben:

```
> b=c(52.5, 62.5, 75, 82.5, 92.5)      # Definition des Vektors
> mean(b)                              # Berechnung des Mittelwertes
> 1/(length(b)-1)*sum((b-mean(b))^2)   # Berechnung der Varinaz
```

Lösung zu Aufgabe 2

1. Die Verteilung des Stichprobenmittels:

$$\bar{X}_n \overset{appr}{\sim} N(\mu_X = \lambda, \sigma_X^2/n = \frac{\lambda}{n}).$$

2. Schätzung von λ: (**Hinweis:** $E(X) = \mu_x = \lambda \Rightarrow \overline{X}_n = \frac{1}{n}\sum_{i=1}^n x_i = \widehat{\lambda}$.)

$$\widehat{\lambda} = \overline{X_n} = \frac{1}{100}(0 \cdot 14 + 1 \cdot 32 + 2 \cdot 22 + 3 \cdot 11 + 4 \cdot 14 + 5 \cdot 7) = 2.0.$$

3. χ^2-Anpassungstest für eine **zusammengesetzte Nullhypothese**: $H_0 : X$ folgt der Poissonverteilung $Poi(2.0)$.

- **Verteilung der Prüfgröße:**

$$\chi^2 = \sum_{i=1^{n^*}} \frac{(N_i - np_i^0)^2}{np_i^0} \overset{appr}{\sim} \chi^2(k-r-1) \text{ hier } \chi^2(6-1-1).$$

- Berechnung der unter der H_0 zu erwarteten Häufigkeiten:

$p_1^0 = P(X=0|H_0) = 0.1353$ $np_1^0 = 100 \cdot 0.1353 = 13.53\ldots,$

$p_6^0 = P(X \geq 5|H_0) = 1 - P(X \leq 4|H_0) = 1 - 0.9473 = 0.0527$ $np_6^0 = 5.27.$

- Die Regel für eine gute Approximation ($np_i^0 \geq 5$ für $k \leq 8$) ist erfüllt.
- Berechnung der Summanden der χ^2- Statistik:

i	X_i	n_i	np_i^0	$n_i - np_i^0$	$\frac{(n_i-np_i^0)^2}{np_i^0}$
1	$x=0$	14	13.53	0.47	0.0163
2	$x=1$	32	27.07	4.93	0.8979
3	$x=2$	22	27.07	-5.07	0.9500
4	$x=3$	11	18.04	-7.04	2.7473
5	$x=4$	14	9.02	4.98	2.7495
6	$x\geq 5$	7	5.27	1.73	0.5679
	$\sum$	100	100	0	7.9289

- **Prüfgröße:**

$$\chi^2 = \sum_{i=1^{n^*}} \frac{(N_i - np_i^0)^2}{np_i^0} = 7.9289.$$

- Kritischer Bereich: $\chi^2 > \chi^2_{0.99,6-1-1}$. Kritischer Wert: $\chi^2_{0.99,4} = 13.28$.
- Testentscheidung: Der empirische Wert der Teststatistik ist kleiner als der kritische Wert. Daher wird die H_0 bei $\alpha = 0.01$ nicht abgelehnt.

4. Stellungnahme:

   (a) Falsch. α-Fehler bedeutet die Nullhypothese abzulehnen, wenn sie richtig ist.

   (b) Falsch. Das Stichprobenmittel ist bei genannten Bedingugen annährend normalverteilt $\bar{X}_n \overset{appr}{\sim} N(\mu_X, \sigma_X^2/n)$.

   (c) Richtig.

5. Aufgaben zu **R**:

```
> pnorm(15, 8, 5)
> dpois(4,2)
> qpois(0.95,2)
```

Lösung zu Aufgabe 3

1. Die Randverteilung von X:

x_i	1	2	3
$f_X(x)$	0.36	0.40	0.24
F_X	0.36	0.76	1.00

   Der Median von X ist $x_{(0.5)} = 2$

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

$X \downarrow \backslash Y \rightarrow$	0	1	2	3
1	0.04	0.10	0.20	0.36
2	0.10	0.38	0.58	0.76
3	0.24	0.56	0.80	1.00

   Der Wert $F_{X,Y}(1, 3) = 0.36$ bedeutet, dass 36% der Befragten höchstens R-Tutorien besucht haben und eine Note 1 bekommen haben.

3.
   - Die bedingte Häufigkeitsfunktion von X $f_{(X|Y)}(x|Y \leq 2) = P(X = x|Y \leq 2) = \frac{P(X=x,Y\leq 2)}{P(Y\leq 2)}$:
     Bedingte Häufigkeitsfunktion an der Stelle $X = 1$

$$P(X = 1|Y \leq 2) = \frac{P(X = 1, Y = 0) + P(X = 1, Y = 1) + P(X = 1, Y = 2)}{P(Y = 0) + P(X = 1) + P(Y = 2)} =$$
$$= \frac{0.4 + 0.06 + 0.1}{0.24 + 0.32 + 0.24} = 0.2/0.8 = 0.25$$

$$P(X = 2|Y \leq 2) = 0.38/0.8 = 0.475$$

$$P(X = 3|Y \leq 2) = 0.22/0.8 = 0.275$$

x_i	1	2	3
$f_{X\|Y}(x\|Y \leq 2)$	0.25	0.475	0.275

   - Der bedingte Mittelwert von X:

$$E(X|Y \leq 2) = 1 * 0.25 + 2 * 0.475 + 3 * 0.275 = 2.025.$$

   - Interpretation der Häufigkeitsfunktion an der Stelle $X = 1$: D.h., dass 25% der Studenten, die höchstens zwei R-Tutorien besuchten, in der Klausur Note 1 erhielten.

4. Statistische Unabhängigkeit von X und Y liegt vor, wenn für jedes Paar gilt:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

   Z.B. für

$$0.14 = P(X = 3, Y = 0) \neq P(X = 3) \cdot P(Y = 0) = 0.24 \cdot 0.24 = 0.0576$$

   $\Rightarrow$ X und Y sind statistisch abhängig.

Interpretation von Unabhängigkeit in diesem Kontext: Die Anzahl der Besuche des R-Tutoriums eines Studenten hat keinen Einfluss auf seine Klausurnote.

5. Gesucht ist der Mittelwert der Randverteilung von X:

$$\bar{x} = \sum_{i=1}^{n} x_i f(x_i) = 1 * 0.36 + 2 * 0.4 + 3 * 0.24 = 1.88$$

Für die Varianz von X setze $g(X, Y) = (X - \mu_X)^2$:

$$s_X^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_{x_i} = \bar{x^2} - \bar{x}^2 = 1^2 \cdot 0.36 + 2^2 \cdot 0.4 + 3^2 \cdot 0.24 - 1.88^2 = 0.5856$$

6. Berechnung der Kovarianz von X, Y:

$$s_{X,Y} = \overline{x \cdot y} - \bar{x} \cdot \bar{y} = 2.24 - 1.88(1.9) \cdot 1.4 = -0.392(0.42)$$

7.
```
> Z[which(n==max(n))]    # Modus von Z
> sum(Z*n)/sum(n)        # Mittelwert von Z
```

Lösung zu Aufgabe 4

1. Verteilung des Stichprobenmittels:

$$\bar{Y}_7 \sim \mathcal{N}\left(\mu_Y; \frac{\sigma_Y^2}{7}\right)$$

2. (a) Mittelwerttest bei unbekannter Varianz:

$$H_0 : \mu \geq \mu_0 \text{ gegen } H_A : \mu < \mu_0$$

Teststatistik:

$$t = \sqrt{n}\frac{\bar{X}_n - \mu_0}{S_n} = \sqrt{7}\frac{42.6186 - 42.67}{\sqrt{0.0126}} = -1.2115$$

Kritische Schranke: $-t_{1-\alpha;n-1} = -t_{0,95;6} = -1.943$.
Testentscheidung: Da $t = -1.2115 > -1.943 = -t_{0.95;6}$, kann H_0 nicht angelehnt werden.

(b) Die Testentscheidung kehrt sich bei $\alpha = 15\%$ um.

3. Varianzquotiententest bei unverbundenen Stichpropen:

$$H_0 : \sigma_X \leq \sigma_Y \quad \text{gegen} \quad H_A : \sigma_X > \sigma_Y$$

Teststatistik: $f = \frac{S_X^2}{S_Y^2} = 0.0229/0.0126 = 1.8175$.
Kritische Schranke: $f_{1-0.1;9,6} = 2.958$.
Testentscheidung: Da $f = 1.8175 < 2.958 = f_{0.90;9,6}$, kann H_0 nicht abgelehnt werden.

4. KI für die Differenz von Mittelwerten bei unbekannten aber identischen Varianzen

$$P\left(\bar{X}_n - \bar{Y}_m - t_{1-\alpha/2;m+n-2}\sqrt{\bar{S}^2}\sqrt{\frac{m+n}{mn}} \leq \mu_X - \mu_Y \leq \bar{X}_n - \bar{Y}_m + t_{1-\alpha/2;m+n-2}\sqrt{\bar{S}^2}\sqrt{\frac{m+n}{mn}}\right)$$

$$\overline{S}^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

5. Interpretation: In 96 % von n gezogenen Stichproben liegt die wahre Differenz der Mittelwerte von X und Y im realisierten KIntervall.

(a) Richtig.

6.
- Mittelwertdifferenztest bei unverbundenen Stichproben (da paired=FALSE) und gleichen Varianzen bei IW 0.05.
- $H_0 : \mu_x - \mu_Y \leq 0$; $H_A : \mu_x - \mu_Y > 0$
- H_0 muss abgelehnt werden, da p-Wert (0.0382) kleiner als Irrtumswahrscheinlichkeit $\alpha = 0,05$ ist.