Aufgabe 1

Der Pharmakonzern Sachse hat ein neues Medikament gegen Bluthochdruck entwickelt, dessen Wirkung überprüft werden sollte.

Zu diesem Zweck wurde 120 Probanden das neue Medikament verabreicht, 80 weitere Probanden erhielten ein Placebo (ein wirkungsloses Scheinmedikament). Bei insgesamt 88 Probanden senkte sich daraufhin der Blutdruck. Bei 60 *Prozent* der Probanden, die ein Placebo erhielten, blieb der Blutdruck unverändert, bei 20 *Prozent* stieg er sogar. Bei keinem der Probanden, die das neue Medikament erhielten, stieg der Blutdruck.

1. Erstellen Sie die Tabelle der gemeinsamen absoluten Häufigkeiten und Randhäufigkeiten der Merkmale M "Art des eingenommenen Präparates" und B: "Art der Blutdruckänderung".

2. Bei wie viel Prozent der Probanden senkte sich der Blutdruck nicht?

3. Wie groß ist der Anteil der Probanden, die das neue Medikament erhielten und keine Änderung des Blutdrucks verzeichneten?

4. Um welchen Faktor unterscheidet sich der Anteil der Patienten mit gesenktem Blutdruck unter den Medikamentenempfängern von dem unter den Placeboempfängern?

Für eine genauere Beurteilung soll nun das Ausmaß der Blutdrucksenkung näher untersucht werden. Folgende Werte des (systolischen) Blutdrucks ergaben sich für 8 Probanden jeweils vor und kurze Zeit nach der Einnahme des neuen Medikaments.

Proband i	1	2	3	4	5	6	7	8
Blutdruck vor Einnahme (in mm Hg) x_i	160	153	161	182	171	169	146	162
Blutdruck nach Einnahme (in mm Hg) y_i	158	145	154	160	170	163	146	160

Die Zufallsvariablen X: "Blutdruck vor Einnahme des Medikaments" und Y: "Blutdruck nach Einnahme des Medikaments" sind in der Grundgesamtheit normalverteilt mit (μ_X, σ_X^2) und (μ_Y, σ_Y^2).

5. Schätzen Sie μ_Y nach der Methode der Momente.

6. Testen Sie bei einer Irrtumswahrscheinlichkeit $\alpha = 0.05$ die Vermutung, dass das neue Medikament den Blutdruck um mehr als 5 mm Hg senkt.

7. Führen Sie obigen Test in **R** durch.

8. Es sei nun bekannt, dass $\mu_X = 162$, $\mu_Y = 157$, $\sigma_x = \sigma_y = 12$ und $Cov(\bar{X}_8, \bar{Y}_8) = 17.5$. Berechnen Sie die Wahrscheinlichkeit, dass die Stichprobendifferenz $\bar{D}_8 = \bar{X}_8 - \bar{Y}_8$ kleiner ist als 5.5.

Aufgabe 2

Vor dem Finale des Eurovision Song Contest (ESC) wurde spekuliert, ob einige Euro-Länder auch fiskalpolitische Erwägungen bei der Punktevergabe berücksichtigen.
Nachfolgend sind für die 7 Länder, die am ESC-Finale teilnahmen und an der Finanzhilfe für Griechenland beteiligt waren, die Höhe der gewährten Hilfe sowie die von Griechenland vergebenen Punkte wiedergegeben:

Land e_ν	1	2	3	4	5	6	7
$X(e_\nu)$	2,9	22,3	16,8	1,3	2,1	9,8	0,2
$Y(e_\nu)$	6	2	8	0	0	0	12

Definiert seien das Merkmal X: "Finanzhilfe für Griechenland (in Mrd. Euro)" und das Merkmal Y : "von Griechenland erhaltene Punkte im ESC-Finale".

1. Bestimmen Sie den Modus von Merkmal Y.
2. (a) Wie hoch war der durchschnittliche Beitrag zur Finanzhilfe für Griechenland pro Land?

   (b) Geben Sie den Code an, mit der die Berechnung aus a) in **R** durchgeführt werden kann.
3. (a) Geben Sie den Median des Merkmals X an und interpretieren Sie diesen.

   (b) Bestimmen Sie das 80%-Quantil des Merkmals X.
4. Bekannt sei $s_X^2 = 64.5241$ und $\bar{y} = 4$.

   (a) Berechnen Sie den Korrelationskoeffizienten von X und Y und interpretieren Sie diesen.

   (b) Berechnen Sie das Korrelationsverhältnis $H^2_{X|Y}$ von X und Y und interpretieren Sie das Ergebnis.
   Hinweis: $H^2_{X|Y} = \sum_{j=1}^{l} (\bar{x}(y_j) - \bar{x})^2 \cdot f_{\cdot j}/s_X^2$
5. (a) Teilen Sie das Merkmal X in die Klassen $[0\text{-}10)$, $[10\text{-}20)$, $[20\text{-}25)$ und das Merkmal Y in die Klassen $[0\text{-}6)$ und $[6\text{-}13)$ ein und erstellen Sie dafür eine Tabelle der gemeinsamen relativen Klassenhäufigkeiten und der entsprechenden Randhäufigkeiten.

   (b) Wieviel Prozent der Länder mit einer Finanzhilfe von unter 10 Mrd. erhielten von Griechenland 6 oder mehr Punkte beim ESC?

   (c) Die Merkmalsausprägungen von X wurden in **R** bereits als Vektor **x** eingelesen. Erstellen Sie in **R** ein Histogramm für das Merkmal X mit den in 5(a) vorgegebenen Klassengrenzen.

Aufgabe 3

A geht für ein Fußballspiel ins Stadion, weil er Tore sehen möchte. Er weiß aus Erfahrung, dass im Mittel pro Spiel ($\hat{=}$ 90 Min.) 2.7 Tore fallen.

Betrachten Sie zunächst die Zufallsvariable T_k: "Anzahl der Tore die innerhalb von k Minuten fallen".

1. Könnte T_k einer Poisson-Verteilung folgen? Prüfen Sie die entsprechenden Annahmen.

Gehen Sie nun davon aus, die Annahmen seien erfüllt.

2. Bestimmen Sie den Parameter λ_{10} für die Zufallsvariable T_{10}.
3. Wie hoch ist die Wahrscheinlichkeit dafür, dass A

   (a) kein Tor verpasst,

   (b) mehr als ein Tor verpasst,

   wenn er für 10 Min. das Stadion verlässt.
4. Wie hoch ist die Wahrscheinlichkeit, dass bei einem Spiel kein Tor fällt?
5. Wie hoch ist die Wahrscheinlichkeit, dass in zwölf Spielen (je 90 Min.) weniger als 20 Tore fallen?

A will wissen, wann endlich das nächste Tor fällt. Betrachten Sie die Zufallsvariable M: "Minuten bis zum nächsten Tor".

6. Wie und mit welchen Parametern ist M unter den oben getroffenen Annahmen verteilt?
7. Wie viele Minuten wird es im Mittel dauern, bis das nächste Tor fällt?

A wartet geduldig 15 Minuten ab.

8. Wie viele Minuten wird es nun im Mittel dauern, bis das nächste Tor fällt?

Gegeben sei ein Vektor x mit einer Stichprobe.

9. Was berechnet die folgende Zeile?

   ```
   sum(log(dpois(x,lambda)))
   ```

   Welches statistische Konzept wird hier berechnet?
10. Wie können Sie die Wahrscheinlichkeit zu Teilaufgabe 5 in R exakt berechnen?

Aufgabe 4

Statistiker K hat eine neue Lieferung mit grünen und roten Bällen aus China bekommen. K interessiert der wahre Anteil roter Bälle p. Der Lieferant hat ihm p nicht genannt, jedoch kann K der Verpackung folgendes entnehmen:

40%的合成橡膠（丁腈橡膠）60%天然橡膠

Leider weiß K nicht, welches der Zeichen für grün und welches für rot steht. Daher will K der Lieferung 10 zufällig gewählte Bälle mit Zurücklegen entnehmen.

1. Welcher Verteilung folgt die Zufallsvariable R: "Anzahl der gezogenen roten Bälle"?

K hat zwei Hypothesen für p entwickelt:

$H_1 : p = 0.4 \qquad H_2 : p = 0.6$

2. Skizzieren Sie die Wahrscheinlichkeitsfunktion der Verteilung von R unter H_1.

K testet nun anhand von R die Hypothese H_1 gegen die Alternative H_2. Er wählt als Irrtumswahrscheinlichkeit $\alpha = 5.48\%$.

3. Geben Sie die kritische Grenze k^* an, für die H_1 gerade noch beibehalten wird.
4. Markieren Sie die Wahrscheinlichkeit für den Fehler erster Art in der Skizze aus Teilaufgabe 2.
5. Bestimmen Sie die Wahrscheinlichkeit für den Fehler zweiter Art zu k^*.
   Hinweis: Nutzen Sie ggf., dass $F_{Binom}(x, n, p) = 1 - F_{Binom}(n - x - 1, n, 1 - p)$.
6. Welche Möglichkeiten hat K, die Güte des Tests zu erhöhen?

K zieht nun 200 Bälle mit Zurücklegen. 95 Bälle sind rot. K rechnet aus, dass

$P(\mathrm{R} \leq 95 | n = 200; p = 0.6) \approx 0.0002.$

7. Geben Sie den p-Wert für den Test von H_1 gegen die Alternative H_2 an.
8. Geben Sie den p-Wert für den Test von H_2 gegen die Alternative H_1 an.
9. Welche der Nullhypothesen lehnen Sie bei einer Irrtumswahrscheinlichkeit von 5 % ab?

Aufgaben zu **R**:

10. Geben Sie Befehle in R an, um die Graphik zu Teilaufgabe 2 zu erzeugen.
11. Geben Sie die Befehle an, um K's Rechnung für $P(R \leq 95 | n = 200; p = 0.6)$ in R durchzuführen.

Aufgabe 1

1. Merkmal M mit m_1: Medikament; m_2: Placebo
   Merkmal B mit b_1: geringerer Bltudruck; b_2: unveränderter Blutdruck; b_3: höherer Blutdruck
   Gegeben:
   $n = 200$, $n(B_1) = 88$, $n(M_1) = 120$, $n(M_2) = 80$, $f(b_2|m_2) = 0.6$, $f(b_3|m_2) = 0.2$

	b_1	b_2	b_3	$\sum$
m_1	72	48	$\underline{0}$	$\underline{120}$
m_2	16	$\underline{48}$	$\underline{16}$	$\underline{80}$
$\sum$	$\underline{88}$	96	16	200

2. $f(B_2) + f(B_3) = 1 - \frac{n(B_1)}{n} = 0.56$

3. $f(M_1 \cap B_2) = \frac{48}{200} = 0.24$

4.
$$\frac{f(B_1|M_1)}{f(B_1|M_2)} = \frac{72/120}{16/80} = 3$$

5. $\bar{Y}_n = \frac{1}{8}(158 + 145 + \cdots + 160) = 157 = \hat{\mu}_Y$

6. Mittelwertdifferenztest bei verbundenen Stichproben.

$$H_0: \ \mu_X - \mu_Y \leq 5 \quad \text{gegen} \quad H_A: \mu_X - \mu_Y > 5$$

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

Teststatistik: $\sqrt{n}\dfrac{\overline{X}_n - \overline{Y}_n - \delta_0}{S_D} = \sqrt{8}\dfrac{(163 - 157 - 5)}{7.1114} = 0.3977$, wobei

$$
\begin{aligned}
S_D^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D_n})^2 \text{ und } \overline{D}_n = \overline{X}_n - \overline{Y}_n = 6. \\
S_D^2 &= \frac{1}{7}((160-158-6)^2 + \cdots + (162-160-6)^2) = \\
&= 1/7((2-6)^2 + (8-6)^2 + \cdots + (2-6)^2 \\
&+ (4.6-5)^2 + (5.1-5)^2) = 354/7 = 50.5714
\end{aligned}
$$

Kritische Schranke: $t_{1-\alpha;n-1} = t_{0.95;7} = 1.895$

Testentscheidung: Da $t = 0.3977 \not> 1.895 = t_{0.95;7}$, kann H_0 nicht abgelehnt werden

7.
```
> x=c(160,153,161,182,171,169,146,162)
> y=c(158,145,154,160,170,163,146,160)
> t.test(x,y,conf.level=0.95,mu=5,alternative='greater',paired=TRUE)
```

8. $\bar{X}_n \sim N(\mu_X; \sigma_X^2/n)$ und $\bar{Y}_n \sim N(\mu_Y; \sigma_Y^2/n) \to \bar{D}_n \sim N(\mu_{\bar{D}_n}, \sigma^2_{\bar{D}_n})$ mit

$$
\begin{aligned}
E(\bar{D}_n) &= E(\bar{X}_8) - E(\bar{Y}_8) = 162 - 157 = 5 \\
Var(\bar{D}_n) &= Var(\bar{X}_8) + Var(\bar{Y}_8) + 2(-1)Cov(\bar{X}_8, \bar{Y}_8) = \frac{12^2}{8} + \frac{12^2}{8} - 2 \cdot 17.5 = 1
\end{aligned}
$$

$\Rightarrow \quad P(\bar{D}_8 \leq 5.5) = \Phi(\frac{5.5-5}{1}) = \Phi(0.5) = 0.6915$

Aufgabe 2

1. $y_{mod} = y_1 = 0$

2. (a) $\bar{x} = \frac{1}{n}\sum_{\nu=1}^{n} X(e_\nu) = 7.9143$

   (b)
```
> x=c(2.9,22.3,16.8,1.3,2.1,9.8,0.2)
> mean(x)
```

3. (a) $x_{(0,5)} = \min\{x|F(x) \geq 0,5\} = x_4$

i	x_i	F_i
1	0.2	0.1429
2	1.3	0.2857
3	2.1	0.4286
4	2.9	0.5714
5	9.8	0.7143
6	16.8	0.8571
7	22.3	1.0000

   Mindestens die Hälfte der 7 Länder haben eine Finanzhilfe von maximal 2.9 Mrd. Euro geleistet.

   (b) $x_{(0,8)} = \min\{x|F(x) \geq 0,8\} = x_6$

4. (a)

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{28.4 - 7.9143 \cdot 4}{\sqrt{64.5241} \cdot \sqrt{19.4286}} = -0.0920$$

   sehr schwacher negativer linearer Zusammenhang zwischen X und Y

(b) $H^2_{X|Y} = \sum_{j=1}^{l} (\bar{x}(y_j) - \bar{x})^2 \cdot f_j / s^2_X$

j	1	2	3	4	5
y_j	0	2	6	8	12
f_j	0.4286	0.1429	0.1429	0.1429	0.1429
$\bar{x}(y_j)$	4.4	22.3	2.9	16.8	0.2
$(\bar{x}(y_j) - \bar{x})^2$	12.3503	206.9484	25.1432	78.9557	59.5104

$H^2_{X|Y} = 58.2460/64.5241 = 0.9027$

Y hat starken Einfluss auf X (X wird sehr gut durch Y erklärt)

5. (a) Tabelle:

$[\widetilde{x}_{i-1}; \widetilde{x}_i) \rightarrow$ $[\widetilde{y}_{j-1}; \widetilde{y}_j) \downarrow$	0 - 10	10 - 20	20 - 25	$f_{\cdot j}$
0 - 6	0.4286	0	0.1429	0.5714
6 - 13	0.2857	0.1429	0	0.4286
$f_{i\cdot}$	0.7143	0.1429	0.1429	1

(b)

$$\frac{f_{12}}{f_{1\cdot}} = \frac{0.2857}{0.7143} = 0.4$$

(c)
```
> grenzen=c(0,10,20,25)
> hist(x,grenzen)
```

Aufgabe 3

1. $T_k \sim Pois(\lambda_k)$, wenn

   - die Wahrscheinlichkeit für ein Tor proportional zur Länge des betrachteten Zeitraums, aber unabhängig von dessen Lage ist;
     Erfüllt, es sei denn, Spieler kommen motiviert aus Halbzeit oder fühlen sich gegen Ende des Spiels unter Druck, etc.
   - die Wahrscheinlichkeit, dass zwei Tore in einem sehr kleinen Zeitraum fallen unendlich klein ist;
     Ja, denn es muss nach jedem Tor neu angespielt werden.
   - die Wahrscheinlichkeit für Tore in disjunkten Intervallen unabhängig ist.
     Es sei denn, es gibt so etwas wie Torserien oder die Spieler hören nach 5 Toren auf weitere zu schießen, etc.

2. $E(\mathrm{T}_{90}) = 2.7 = \lambda_{90} \Rightarrow E(\mathrm{T}_{10}) = \frac{2.7}{90} \cdot 10 = 0.3 = \lambda_{10}$

3. (a) $P(\mathrm{T}_{10} = 0) = f_{Pois}(0, 0.3) = \frac{0.3^0 \exp(-0.3)}{0!} = 0.7408...$

   (b) $P(\mathrm{T}_{10} > 1) = 1 - P(\mathrm{T}_{10} = 0) - P(\mathrm{T}_{10} = 1) = 1 - \frac{0.3^0}{0!}e^{-0.3} - \frac{0.3^1}{1!}e^{-0.3} = 0.036...$

4. $\mathrm{T}_{90} \sim \mathrm{Pois}(\lambda_{90})$, $\lambda_{90} = 2.7 \Rightarrow P(\mathrm{T}_{90} = 0) = f_{Pois}(0, 2.7) = 0.0672...$

5. $\mathrm{T}_{12 \cdot 90} \sim \mathrm{Pois}(\lambda_{12 \cdot 90})$, $\lambda_{1080} = 2.7 \cdot 12 = 32.4$
   $P(\mathrm{T}_{1080} < 20) = P(\mathrm{T}_{1080} \leq 19) = F_{Pois}(19; 32.4) =$ *nicht tabelliert*
   wegen $\lambda_{1080} > 10$ ist Approximation durch Normalverteilung möglich
   $F_{Pois}(19; 32.4) \approx F_{Norm}(19+\mathbf{0.5}; 32.4, 32.4) = \Phi(\frac{19.5-32.4}{\sqrt{32.4}}) = \Phi(-2.27)$
   $= 1 - \Phi(2.27) = 0.0116...$

6. 2.7 Tore in 90 Min., also $\mathrm{E(M)}= \frac{90}{2.7} = \frac{1}{\lambda}$, damit gilt $\mathrm{M} \sim \mathrm{Exp}(\frac{2,7}{90})$

7. $E(\mathrm{M}) = \frac{90}{2.7}$

8. $E(\mathrm{M}-15|\mathrm{M}> 15) = E(\mathrm{M}) = \frac{2,7}{90}$, kein Gedächtnis

9. $\sum_{i=1}^{n} ln\,(f_{pois}(x_i; \lambda))$ berechnet die Likelihoodfunktion an der Stelle *lambda*.

10. ppois(19,32.4)

Aufgabe 4

1. R$\sim Binom(10, p)$
2. Unter H_1 gilt $P(R = x) = f_{Binom}(x, 10, 0.4)$. Die Werte von $f_{binom}(x, 10, 0.4)$ stehen in einer Spalte der Tabellensammlung. *Beachte: diskrete Zufallsvariable!*
3. K lehnt für zu große R H_1 ab und akzeptiert dabei sich in 5.48 % der Fälle zu täuschen, so dass für die kritische Schranke k^* gelten muss

   $$P(R > k^*|H_1) \doteq \alpha \Rightarrow 1 - P(R \leq k^*|p = 0.4) \doteq \alpha$$

   $$\Rightarrow P(R \leq k^*|p = 0.4) \doteq 1 - \alpha \quad \text{mit } 1 - \alpha = 1 - 0.0548 = 0.9452$$

   $$\Rightarrow F_{Binom}(k^*; 10, 0.4) \doteq 0.9452 \Rightarrow k^* = 6.$$

4. Grafik in `R`:

   ```
   x=0:10
   plot(x,dbinom(x,10,.4),col="blue",ylab='P(R=x)',frame.plot=F,pch=19)
   for(i in 7:10){lines(c(i,i),c(0,dbinom(i,10,.4)),col='red',lwd=4)}
   ```

5. β-Fehler zu H_1 vs. H_2: $P(H_1 \text{ annehmen}|H_2) = P(R \leq k^*|H_2) = P(R \leq k^*|p = 0.6)$
   $= F_{Binom}(k^*; 10, 0.6) = 1 - F_{Binom}(\text{9-k}^*; \text{10,0.4}) = 1 - F_{Binom}(3; 10, 0.4) = 0.6177$
6. K könnte Irrtumswahrscheinlichkeit α oder Stichprobenumfang n erhöhen.
7. p-Wert(H_1): $\quad P(\text{dieser oder extremerer Beobachtung}|H_1)$
   $= P(R \geq 95|p = 0.4) = 1 - F_{Binom}(94; 200, 0.4) =$ *nicht tabelliert*
   wegen $np(1 - p) > 9$ ist Approximation durch Normalverteilung möglich
   $\approx 1 - F_{Norm}(94 + \mathbf{0.5}, np, np(1 - p)) = 1 - \Phi(\frac{94.5-80}{\sqrt{200 \cdot 0.4(1-0.4)}}) = 1 - \Phi(2.09) = 0.0183$
8. p-Wert(H_2): $\quad P(\text{dieser oder extremerer Beobachtung}|H_2)$
   $= P(R \leq 95|p = 0.6) \approx 0.0002$ (aus Angabe)
9. Verwerfe Hypothese für die p-Wert $< \alpha$ gilt, hier also beide.
10. `x=0:10`
    `plot(x,dbinom(x,10,.4))`
11. `pbinom(95,200,.6)`