Aufgabe 1

In einer Umfrage wird der Besitz eines Smartphones (Merkmal X) und die Nutzungsdauer des Internets pro Monat (Merkmal Y) untersucht. Merkmal X hat zwei Ausprägungen: X_1: "Besitz" und X_2: "Nichtbesitz". Die Nutzungsdauer (in Stunden) wird in drei Kategorien eingeteilt: Y_1: weniger als 20, Y_2: von 20 bis 60, Y_3: mehr als 60. Bekannt ist:

- Insgesamt wurden 200 Handynutzer befragt, von denen 21% ein Smartphone besitzen, d.h. $f(X_1) = 0.21$.
- Kein Smartphone-Besitzer surft weniger als 20 Std. pro Monat, d.h. $f(X_1 \cap Y_1) = 0$.
- 14 Befragte haben ein Smartphone und surfen monatlich von 20 bis 60 Std., d.h. $n(X_1 \cap Y_2) = 14$.
- 39% der Befragten sind mehr als 60 Std. pro Monat online, d.h. $f(Y_3) = 0.39$.

1. Wieviele Befragte haben kein Smartphone?

2. Berechnen Sie den Anteil der Smartphone-Besitzer unter den Nutzern, die monatlich mehr als 60 Stunden online sind.

Sie wissen noch, dass 18.99% der Surfer, die kein Smartphone haben, pro Monat weniger als 20 Stunden das Internet nutzen, d.h. $f(Y_1|X_2) = 0.1899$.

3. Wieviele der Befragten surfen monatlich weniger als 20 Stunden?

Die genaue durchschnittliche monatliche Nutzungsdauer des Internets (Y) und die monatlichen Ausgaben für Handy (Z) von 8 Befragten werden zusammengefasst:

Befragte i	1	2	3	4	5	6	7	8
Y_i (in Stunden)	67	45	36	22	60	40	35	18
Z_i (in Euro)	80	50	23	20	95	48	15	10

4. Berechnen Sie die gemeinsame Verteilungsfunktion $F_{Y,Z}(36, 16)$. Interpretieren Sie das Ergebnis.

5. Bekannt seien $\overline{y}_8 = 40.375, \overline{z}_8 = 42.625, s^2_{Y;8} = 250.23, s^2_{Z;8} = 868.48$ und $\sum_{i=1}^{8} y_i z_i = 17203$. Berechnen Sie den Stichprobenkorrelationskoeffizienten.

6. Gegeben ist folgender `R`-Output:

```
        One Sample t-test
data:  dauer
t = 3.4078, df = 7, p-value = 0.005661
alternative hypothesis: true mean is greater than 20
99 percent confidence interval:
 22.45042      Inf
sample estimates:
mean of x
   40.375
```

(a) Durch welchen `R`-Befehl können Sie diesen `R`-Output erhalten?

(b) Treffen Sie die Testentscheidung.

7. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

   Sei X in der Grundgesamtheit mit (μ, σ^2) normalverteilt. Zwei Hypothesen $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$ werden bei einer Irrtumswahrscheinlichkeit von α getestet.

   (a) Die Wahrscheinlichkeit, eine richtige H_0 nicht zu verwerfen, beträgt $1 - \alpha$.

   (b) Die Wahrscheinlichkeit, eine falsche H_0 zu verwerfen, beträgt $1 - \beta$.

   (c) Die Wahrscheinlichkeit für den Fehler 2. Art sinkt bei festem Stichprobenumfang mit zunehmendem Signifikanzniveau.

   (d) Die Wahrscheinlichkeit für den Fehler 2. Art sinkt mit Vergrößerung der Stichprobe.

Aufgabe 2

Der Student Paul fährt jeden Tag mit dem Fahrrad zur Uni. Die Strecke kreuzt drei Fußgängerüberwege mit Bedarfsampeln, die durch Fußgänger aktiviert werden. Wird Paul an einer Bedarfsampel gestoppt, verzögert dies seine Weiterfahrt um 90 Sekunden. Wird Paul nicht aufgehalten, benötigt er 17 Minuten, bis er den Hörsaal erreicht.

Nehmen Sie an, die Zufallsvariable X: "Anzahl der Bedarfsampeln, die auf Pauls Weg zur Uni ein rotes Signal zeigen" ist binomialverteilt mit den Parametern $n = 3$ und $p = 0.4$, d.h. $X \sim \text{Bin}(3, 0.4)$.

1. Geben Sie das 90%-Quantil der Binomialverteilung mit Parametern $n = 3$ und $p = 0.4$ an. Interpretieren Sie dieses Quantil inhaltlich.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden, dass das gesuchte Quantil den Wert 3 hat)

2. Geben Sie Modus, Erwartungswert und Varianz der Zufallsvariablen X an.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden, den Wert 2 für den Modus, den Wert $\frac{5}{4}$ für den Erwartungswert und den Wert 3.1 für die Varianz)

3. Geben Sie einen `R`-Code an, mit dem Sie die Wahrscheinlichkeit dafür berechnen, dass mindestens eine und höchstens zwei Bedarfsampeln auf Pauls Weg zur Uni das Haltesignal zeigen.

Betrachten Sie zusätzlich die Zufallsvariable Y: "Fahrtzeit in Minuten, die Paul für seinen Weg zur Uni benötigt". Paul erkennt, dass er die Zufallsvariablen X und Y linear ineinander überführen kann:

$$Y = 17 + 1.5 \cdot X$$

4. Geben Sie Modus, Erwartungswert und Varianz der Zufallsvariablen Y an.

5. Pauls Lieblingsübung beginnt pünktlich um 8:00 Uhr. Wann muss Paul spätestens losfahren, wenn er mit Wahrscheinlichkeit von 0.9 rechtzeitig ankommen möchte?

6. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

(a) Die Poissonverteilung ist eine stetige Verteilung.

(b) Der Parameter λ der Poissonverteilung bestimmt den zugehörigen Erwartungswert.

(c) Mit dem `R`-Befehl `dpois(1, lambda=4)` wird eine Zufallszahl erzeugt, die aus einer Poissonverteilung mit Parameter $\lambda = 4$ stammt.

(d) Die Wahrscheinlichkeitsfunktion der Poissonverteilung nimmt negative Werte an, wenn der Parameter $\lambda < 1$ ist.

Aufgabe 3

In der Tiefgarage eines Einkaufszentrums müssen die Kunden einen Aufzug zu den Geschäften benutzen. Die Zufallsvariable X: "Wartezeit vor dem Aufzug" sei exponentialverteilt mit der Verteilungsfunktion $F_{Exp}(x; \lambda) = 1 - e^{-\lambda x}$, für $x \geq 0$. Durchschnittlich warten die Kunden 1.25 Minuten vor dem Aufzug.

1. Bestimmen Sie den Erwartungswert $E(X)$ und dadurch den Parameter λ.
   (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden λ =0.9)

2. Berechnen Sie die Wahrscheinlichkeit, dass ein Kunde mehr als 3 Minuten auf den Aufzug warten muss.

3. Geben Sie den `R`-Code an, um die Wahrscheinlichkeit, dass ein Kunde weniger als 2 Minuten auf den Aufzug warten muss, zu berechnen.

4. Mit einer Wahrscheinlichkeit von 60% liegt die Wartezeit im Intervall $[a, 4[$. Berechnen Sie den Wert von a.

5. Die Zufallsvariable Y: "Stromverbrauch (in kW) des Aufzugs pro Stunde" sei normalverteilt mit Varianz $\sigma_Y^2 = 2$. Eine Stichprobe von 32 solchen Aufzügen wurde geprüft. Das Stichprobenmittel ist $\overline{y}_{32} = 17.5$. Berechnen Sie ein realisiertes 90%-Konfidenzintervall für den mittleren stündlichen Stromverbrauch μ_Y.

6. Das Einkaufszentrum hat 3 baugleiche Aufzüge, die voneinander unabhängig sind. Der stündliche Stromverbrauch eines Aufzugs sei $Y_i \sim N(17, 2), i = 1, 2, 3$.

   (a) Gegeben ist Zufallsvariable S: "gesamter stündlicher Stromverbrauch der 3 Aufzüge". Wie ist S verteilt?

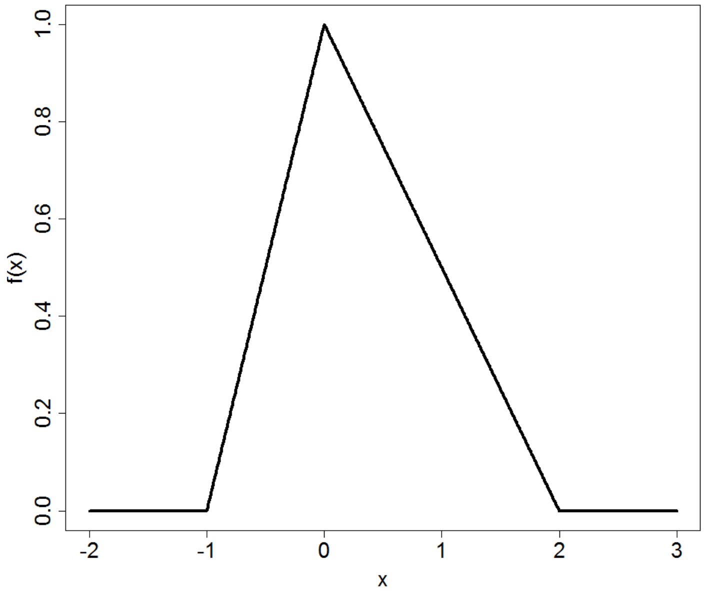
   (b) Geben Sie den `R`-Code an, um die Wahrscheinlichkeit $P(S \leq 60)$ zu berechnen.

7. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

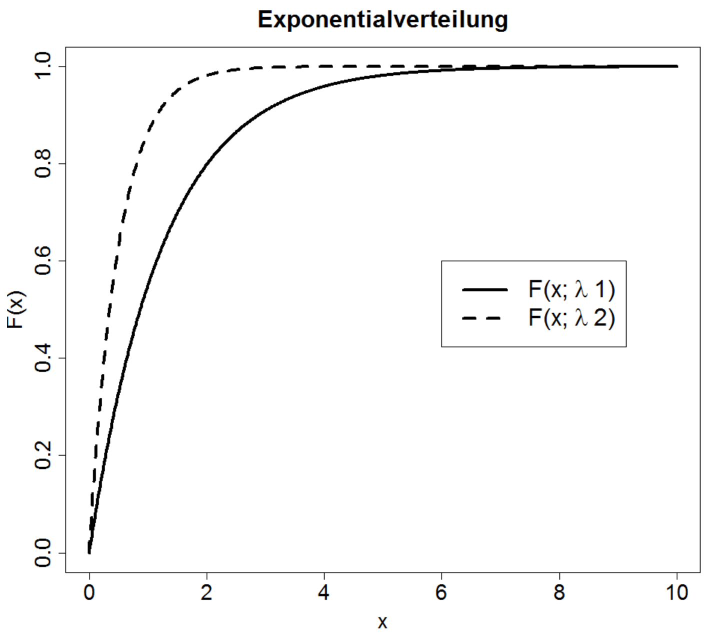
   (a) Die Dichtefunktion $f_X(x)$ einer stetigen Zufallsvariablen kann Werte größer als 1 annehmen.

(b) Für die Verteilungsfunktion einer stetigen Zufallsvariablen gilt $F_X(x) = \int_{-\infty}^{x} f(s)ds$.

(c) Die im Folgenden skizzierte Funktion $f(x)$ kann als Dichtefunktion interpretiert werden.



(d) Aus folgendem Plot kann man folgern, dass $\lambda 1 > \lambda 2$ ist.



Aufgabe 4

Rennwagen **R1** und **R2** trainieren abwechselnd auf einer Strecke, d.h. die Autos fahren unabhängig voneinander. Ein Beobachter notiert die auf der Strecke erreichten Zeiten (in Minuten) x_i von R1 und y_i von R2 für $i = 1, 2, \ldots, 10$ Runden und berechnet folgende Größen: $\overline{x}_{10} = 25.23$, $\sum_{i=1}^{10} x_i^2 = 6399.79$, $\overline{y}_{10} = 25.49$ und $\sum_{i=1}^{10}(y_i - \overline{y})^2 = 9.269$.

1. Bestimmen Sie die Stichprobenvarianzen der erreichten Zeit von R1 und R2. (**Hinweis:** Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden $s^2_{X;10} = 4$ und $s^2_{Y;10} = 1$.)

Die Zufallsvariablen X: "Erreichte Zeit (in Min.) von R1" und Y: "Erreichte Zeit (in Min.) von R2" seien in der Grundgesamtheit normalverteilt: $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$.

2. Testen Sie mit einer Irrtumswahrscheinlichkeit von 5% die Vermutung, dass R1 im Durchschnitt besser abschneidet als R2, d.h. $\mu_X < \mu_Y$, wenn es gilt $\sigma_X^2 = \sigma_Y^2$.

3. Zusätzlich wird die Hypothese $H_0 : \mu_X = 25$ gegen $H_1 : \mu_X \neq 25$ mit $\alpha = 0.05$ getestet. Der p-Wert des Tests beträgt 0.71.

   (a) Treffen Sie eine Testentscheidung und begründen Sie diese kurz (keine Berechnung erforderlich).

   (b) Geben Sie den theoretischen kritischen Bereich des Tests an.

R1 trainiert anschließend allein auf einer Strecke der Länge 120 km. Auf dem ersten Streckenabschnitt fährt R1 mit einer Geschwindigkeit von 280 km/h. Die folgenden zwei gleich langen Abschnitte bewältigt R1 mit jeweils 300 km/h und 295 km/h.

4. Bestimmen Sie die durchschnittliche Geschwindigkeit des R1 auf der Strecke. Bekannt ist, dass der erste Abschnitt $\frac{2}{3}$ und die letzten beiden jeweils $\frac{1}{6}$ der Strecke ausmachen.

5. Geben Sie an, ob die folgenden Aussagen "wahr" oder "falsch" sind (ohne Begründung). Jede **richtige** Antwort wird mit **0.5**, jede **falsche** mit **-0.5** und keine Antwort wird mit **0** Punkten bewertet. **Insgesamt** können in dieser Teilaufgabe **mindestens 0** und **höchstens 2** Punkte erreicht werden.

   Seien $(X_1, \ldots, X_n)$ und $(Y_1, \ldots, Y_m)$ zwei einfache Stichproben aus normalverteilten

Grundgesamtheiten. Die Realisationen der zweidimensionalen Stichprobe seien als `x` und `y` in `R` gespeichert.

(a) Mit dem Befehl `cov(x,y)/sqrt(var(x)*var(y))` wird der Stichprobenkorrelationskoeffizient von `x` und `y` berechnet.

(b) Gegeben ist die folgende Funktion:

```
f1=funktion(x){
l=mean(x)-qt(0.975,(length(x)-1))*sqrt(var(x)/length(x)))
u=mean(x)+qt(0.975,(length(x)-1))*sqrt(var(x)/length(x)))
i=c(l,u)
return(i)}
```

i. Die Funktion `f1` berechnet ein Konfidenzintervall für den Mittelwert von X.

ii. Die Varianz der Grundgesamtheit ist bekannt.

iii. Das Konfidenzniveau beträgt $1 - \alpha = 0.95$.

Lösung Aufgabe 1

1.

$$n(X_2) = 200(1 - 0.21) = 158$$

2.

	Nutzungsdauer ↓	X_1 Smartphone besitzen	X_2 kein Smartphone besitzen	
Y_1	weniger als 20 Std.	0		
Y_2	20 bis 60 Std.	14/200		
Y_3	mehr als 60 Std.			0.39
		0.21	$1 - 0.21$	1

$$f(X_1|Y_3) = \frac{n(X_1 \cap Y_3)}{n(Y_3)} = \frac{200 \cdot 0.21 - 14}{200 \cdot 0.39} = 0.359$$

3.

$$f(Y_1 \cap X_2) = f(Y_1|X_2) \cdot f(X_2) = 0.1899(1 - 0.21) = 0.15$$

$$n(Y_1) = n(Y_1 \cap X_1) + n(Y_1 \cap X_2) = 0 + 200 \cdot 0.15 = 30$$

4. $F_{Y,Z}(36, 16) = 2/8 = 0.25$
   Der Anteil der Handy-Besitzer, deren Nutzungsdauer des Internets nicht länger als 36 Stunden und Ausgaben für Handy nicht höher als 16 Euro, beträgt 25%.

5. $\overline{yz} = 17203/8 = 2150.375$
   $s_{YZ} = \overline{yz} - \overline{y} \cdot \overline{z} = 2150.375 - 40.375 \cdot 42.625 = 429.3906$
   $r_{YZ} = \frac{s_{YZ}}{s_Y s_Z} = 0.9211$

6. (a) `t.test(dauer,alternative='greater',mu=20,conf.level=0.99)`
   (b) H_0 verwerfen, da $p < \alpha$.

7. wahr, wahr, wahr, wahr

Lösung Aufgabe 2

1. $Q_{90\%}^{\text{Bin}(3,0.4)} = 2$: Mit Wahrscheinlichkeit von 90% wird eine Fahrt von Paul zur Uni von höchstens zweiRotphasen verzögert.

2. $X_{mod} = 1$,
   $\text{E}[X] = np = 1.2$,
   $\text{Var}[X] = npq = 1.2 \cdot 0.6 = 0.72.$

3. Punktvergabe: An den Stellen 1 und 2; Wahrscheinlichkeiten berechnen in R ; Wahrscheinlichkeiten addieren

   Beispiele:

   (a) `0.432+0.288`

   (b) `sum(dbinom(1:2,3,0.4))`

   (c) `sum(choose(3,1:2)*0.4^(1:2)*(1-0.4)^(2:1))`

4. $Y_{mod} = 17 + 1.5 \cdot X_{mod} = 18.5$,
   $\text{E}[Y] = 17 + 1.5 \cdot \text{E}[X] = 18.8$,
   $\text{Var}[Y] = (1.5)^2 \cdot \text{Var}[X] = 1.62$.

5. 90% -Quantil aus Aufgabe 2: 2

   $$8:00 \text{ Uhr} \quad - \quad \left(17 + 1.5 \cdot Q_{90\%}^{\text{Bin}(3,0.4)}\right) \text{Minuten " = " } 7:40 \text{ Uhr}$$

6. (a) falsch

   (b) wahr

   (c) falsch

   (d) falsch

Lösung Aufgabe 3

1. $$E(X) = 1.25$$
   $$\lambda = \frac{1}{E(X)} = \frac{1}{1.25} = 0.8$$

2. $$P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = e^{-0.8 \cdot 3} = 0.0907$$

3. `pexp(2,0.8)`

4. $P(a \leq X < 4) = F(4) - F(a) = 1 - e^{-0.8 \cdot 4} - (1 - e^{-0.8 \cdot a}) = 0.6$
   $a = 0.5563$

5. $$\begin{aligned} KI &= \left[\bar{Y}_n - \lambda_{1-\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}}, \bar{Y}_n + \lambda_{1-\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}}\right] \\ &= \left[17.5 - 1.6448 \cdot \sqrt{\frac{2}{32}}, 17.5 + 1.6448 \cdot \sqrt{\frac{2}{32}}\right] \\ &= [17.0888, 17.9112] \end{aligned}$$

6. (a) $S \sim \mathcal{N}(51, 6)$
   (b) `pnorm(60,51,sqrt(6))`

7. wahr, wahr, falsch, falsch

Lösung Aufgabe 4

ALT

1. $s^2_{X;10} = \frac{n}{(n-1)}\overline{x^2} - \overline{x}^2 = \frac{(\sum_{i=1}^{10} x_i^2) - n*\overline{x}^2}{(n-1)} = (6399.79 - 10*(25.23)^2)/9 = 3.8068$
   $s^2_{Y;10} = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{9.269}{9} = 1.0299$

2. (a) Ein Mittelwertdifferenztest für unverbundene Stichproben, da x_i und y_i paarweise stochastisch unabhängig sind (verschiedene Autos).

   (b) $H_0: \ \mu_x - \mu_y \geq 0$ gegen $H_1: \mu_x - \mu_y < 0$. $\Rightarrow$
   Kritischer Bereich: $\frac{\overline{X}_n - \overline{Y}_m}{\overline{S}}\sqrt{\frac{mn}{n+m}} < -t_{1-\alpha;n+m-2}$.

   (c) kritische Schranke: $-t_{1-\alpha;n+m-2} = -t_{0.95;18} = -1.734$

   (d)

$$\overline{S}^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} = \frac{9*3.8068 + 9*1.0299}{18} = 2.41 \text{ bzw. } \left[\frac{9*4 + 9*1.0299}{18} = 2.515\right]$$

   (e) Teststatistik: $T = \frac{25.23-25.49}{1.5551}\sqrt{\frac{100}{20}} = -0.3739$ bzw. $\left[\frac{25.23-25.49}{1.5859}\sqrt{\frac{100}{20}} = -0.3666\right]$

   (f) Testentscheidung: $T = -0.3739[-0.3666] \nless -1.734 = -t_{0.95;18}$, d.h die Teststatistik liegt nicht im kritischen Bereich, somit wird H_0 nicht abgelehnt.

3. (a) Die H_0 wird nicht abgelehnt, da $p > \alpha$ ist, somit liegt die Teststatistik nicht im kritischen Bereich.

   (b) Der kritische Bereich des Tests lautet: $T_n = \left|\frac{\overline{X}_n - \mu_0}{S_{X;n}/\sqrt{n}}\right| > t_{1-\alpha/2;n-1}$.

4. <u>Berechnung der durchschnittlichen Geschwindigkeit:</u>
   $x_1 = \frac{2}{3} * 120 = 80$und $x_2 = x_3 = \frac{1}{6} * 120 = 20$.
   Harmonisches Mittels, da eine Verhältniszahl und benötigte Zeit (Nenner) unbekannt. (Alternativ arithmetisches Mittel möglich)

$$\overline{v}_H = \left(\frac{1}{n}\sum_{i=1}^{k}\frac{n_i}{x_i}\right)^{-1} = \left(\frac{1}{120}\left(\frac{80}{280} + \frac{20}{300} + \frac{20}{295}\right)\right)^{-1} = 285.5935 \text{ km/h}$$

5. wahr, wahr, falsch, wahr.