Musterlösung

1. **a)** **(2 Pt)** (1 Pt pro richtige Antwort).

$$\mathcal{P}(G\Big|W^c) = 1 - \mathcal{P}(G^c\Big|W^c) = 1 - 0.9 = 0.1 = 10\%$$

$$\mathcal{P}(G) = \mathcal{P}(G \cap W) + \mathcal{P}(G \cap W^c) = 0.75 \cdot \frac{1}{3} + 0.1 \cdot \frac{2}{3} = \frac{19}{60} = 31.67\%$$

b) **(2 Pt)** (1 Pt für die Formel $\mathcal{P}(W|G) = \mathcal{P}(G|W) \cdot \mathcal{P}(W)/\mathcal{P}(G)$

$$P\Big(W\Big|G\Big) = P\Big(G\Big|W\Big) \cdot \frac{\mathcal{P}(W)}{\mathcal{P}(G)} = 0.75 \cdot \frac{1}{3} : \frac{19}{60} = \frac{15}{19} \cong 78.95\%$$

c) **(2 Pt)** (1 Pt für $\mathcal{P}(W \cap G^c)$ oder $\mathcal{P}(W^c \cap G)$ richtig berechnet).

$$P\Big(W \cap G^c\Big) + P\Big(W^c \cap G\Big) = \mathcal{P}(W) \cdot P\Big(G^c\Big|W\Big) + \mathcal{P}(G) \cdot P\Big(W^c\Big|G\Big)$$

$$= \frac{1}{3} \cdot \frac{25}{100} + \frac{2}{3} \cdot \frac{10}{100} = \frac{1}{12} + \frac{1}{15} = \frac{3}{20} = 15\%$$

d) **(2 Pt)** Sei X die Zufallsvariable des Lohns.

$$\mathbf{E}[X] = 5000 \cdot \mathcal{P}(G \cap W) + 0 \cdot \mathcal{P}(G \cap W^c) + 1000 \cdot \mathcal{P}(G^c)$$

$$= 5000 \cdot \frac{1}{4} + 1000 \cdot \frac{41}{60} = 1933.3.$$

e) **(1 Pt)** $1 - 0.75^{10} = 0.944$.

2. **(10 Punkte)**
Die Krankenkasse easyHealth bietet einen sehr günstigen Studententarif für 160 Franken pro Jahr an. Sie hat pro Kunde einen jährlichen Administrationsaufwand von 100 Franken. Zusätzlich fallen die variierenden Gesundheitskosten (Arztrechnungen etc.) an. Zur Modellierung der Gesamtkosten X pro Kunde und Jahr benutzt die Krankenkasse die folgende Dichte

$$f(x) = \begin{cases} 0 & x \le c \\ \frac{\alpha c^\alpha}{x^{\alpha+1}} & x > c \end{cases}$$

wobei $c > 0, \alpha > 1$ Parameter sind.

Tipp: Beachten Sie bitte für alle untenstehenden Berechnungen, dass die Dichte 0 ist für $x \le c$.

a) (1 Punkt) Wofür steht in dem Modell der Parameter c und welchen Wert hat er? *Hinweis:* Nichts rechnen. **(1 Punkt)** c ist der Minimalwert der Verteilung und hat daher hier den Wert 100 (die Administrationskosten, die in jedem Fall anfallen).

b) (2 Punkte) Was sind die erwarteten Gesamtkosten pro Kunde und Jahr? **(2 Punkte)**

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx \text{ (0.5Pt)} = \int_c^{\infty} x\frac{\alpha c^\alpha}{x^{\alpha+1}}dx \text{ (0.5Pt)} = \frac{\alpha c^\alpha}{-\alpha+1}\left[x^{-\alpha+1}\right]_c^\infty = c \cdot \frac{\alpha}{\alpha-1} \text{ (1Pt)}.$$

Nehmen Sie von nun an $\alpha = 3$ und den Wert für c aus Teilaufgabe a). Falls Sie Aufgabenteil a) nicht lösen konnten, verwenden Sie $c = 120$.

c) (2 Punkte) Was ist die Wahrscheinlichkeit, dass die Gesamtkosten für einen Kunden pro Jahr weniger als 200 Franken betragen? *Ersatzresultat:* 0.8. **(2 Punkte)**

$$P(X < 200) = F(200)\ (0.5\text{Pt})\ = \int_c^{200} \frac{\alpha c^\alpha}{x^{\alpha+1}} dx\ (0.5\text{Pt})\ = \left[-\frac{c^\alpha}{x^\alpha}\right]_c^{200} = 1 - \left(\frac{c}{200}\right)^\alpha = \frac{7}{8} = 0.875\ (1\text{Pt})$$

Oder mit Ersatzresultat $c = 120$ aus a): $\frac{98}{125} = 0.784$

d) (2 Punkte) Ein Kunde ist 5 Jahre (für die Dauer des Studiums) in diesem Tarif versichert. Angenommen die Gesamtkosten in den einzelnen Jahren sind unabhängig, was ist die Wahrscheinlichkeit, dass die jährlichen Gesamtkosten für einen Kunden in mindestens 2 Jahren mehr als 200 Franken betragen? **(2 Punkte)** Wir definieren die Zufallsvariable Y als die Anzahl Jahre, in denen die Gesamtkosten mehr als 200 Franken betragen haben. Dann gilt $Y \sim B(5, p)$, wobei $p = 1 - P(X < 200)$ (aus Aufgabenteil c)). Somit:

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y < 2) = 1 - P(Y = 0) - P(Y = 1)\ (1\text{Pt}) \\ &= 1 - (1-p)^5 - 5p(1-p)^4 = 0.1207\ (1\text{Pt}) \end{aligned}$$

Oder mit Ersatzresultat $P(X < 200) = 0.8$: 0.2627
Oder mit Ersatzresultat $P(X < 200) = 0.784$: 0.2958

e) (3 Punkte) easyHealth hat zur Zeit 250 studentische Kunden. Was ist (approximativ) die Wahrscheinlichkeit, dass easyHealth mit diesem Tarif in einem Jahr Gewinn macht? Sie können die folgende Formel ohne Herleitung benutzen:

$$\text{Var}(X) = \frac{c^2\alpha}{(\alpha-1)^2(\alpha-2)}$$

Falls Sie b) nicht gelöst haben, verwenden Sie als Erwartungswert von X den Wert 155.
(3 Punkte) Wir definieren die Zufallsvariable G als Differenz aus Einnahmen und Kosten: $G = 250 \cdot 160 - \sum_{i=1}^{250} X_i$, wobei X_i iid Kopien von X sind (1Pt). Dann gilt (mit dem zentralen Grenzwertsatz):

$$\begin{aligned} P(G > 0) &= P\left(\sum_{i=1}^{250} X_i < 250 \cdot 160\right) = P\left(\frac{\sum_{i=1}^{250} X_i - 250\mu}{\sqrt{250\sigma^2}} < \frac{250(160-\mu)}{\sqrt{250\sigma^2}}\right) \\ &\approx \Phi^{-1}\left(\sqrt{\frac{250}{\sigma^2}}(160-\mu)\right)\ (1\text{Pt}) \end{aligned}$$

wobei $\mu = E[X]$ und $\sigma^2 = var(X)$. Somit erhalten wir

$$P(G > 0) \approx \Phi^{-1}(1.8257) = 0.96\ (1\text{Pt})$$

Oder mit Ersatzresultat $c = 120$:

$$P(G > 0) \approx \Phi^{-1}(-3.0429) = 1 - \Phi^{-1}(3.0429) = 1 - 0.9988 = 0.0012$$

Oder mit Ersatzresultat $E[X] = 155$: $P(G > 0) \approx \Phi^{-1}(0.9129) = 0.8186$.
Oder mit beiden Ersatzresultaten: $P(G > 0) \approx \Phi^{-1}(0.7607) = 0.7764$.

3. (**7 Punkte**)
Wir betrachten die Daten

x_1	x_2	x_3	x_4	x_5
1.52	2.12	2.28	1.72	1.86

die wir als Realisierungen von i.i.d. Zufallsvariablen mit folgender Dichte auffassen:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ c \cdot \frac{1}{\theta^4} x^3 e^{-\frac{x}{\theta}} & x > 0 \end{cases}$$

wobei c eine (bekannte) Konstante ist und $\theta > 0$ der Parameter ist. Wir wollen den unbekannten Parameter θ aus obiger Stichprobe schätzen.

Hinweis: Für eine Zufallsvariable X mit obiger Dichte gilt $\mathcal{E}[X] = 4 \cdot \theta$.

a) (2 Punkte) Bestimmen Sie die Likelihood- und die log-Likelihoodfunktion basierend auf n unabhängigen identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer Zufallsvariablen mit obiger Dichte. Schreiben Sie die allgemeine Form auf. Die konkreten Werte brauchen Sie nicht einzusetzen, auch die Maximierung müssen Sie nicht durchführen.

$$\ell(\theta; x_1, \ldots x_n) = \frac{c^n}{\theta^{4n}} \prod_{i=1}^{n} x_i^3 e^{-\frac{x_i}{\theta}}$$

$$\log \ell(\theta; x_1, \ldots x_n) = n \ln(c) - 4n \ln(\theta) + 3 \sum_{i=1}^{n} \ln(x_i) - \frac{1}{\theta} \sum_{i=1}^{n} x_i$$

(jeweils 1 Punkt, insgesamt: 2 Punkte)

b) (2 Punkte) Bestimmen Sie den Momentenschätzer für θ, sowohl basierend auf n unabhängigen Beobachtungen $x_1, \ldots, x_n$, wie auch den realisierten Wert für die gegebene Stichprobe.

Gleichsetzen der Momente, $\mathbf{E}[X] = \frac{1}{n} \sum_{i=1}^{n} x_i$, ergibt $\widehat{\theta}_{MoM} = \frac{\sum_{i=1}^{n} x_i}{4n} = 0.475$. In diesem Fall ist $\widehat{\theta}_{MLE} = \widehat{\theta}_{MoM}$.

(jeweils 1 Punkt für richtige Formel und korrekten Wert, insgesamt: 2 Punkte)

Der Maximum-Likelihood Schätzer ist hier gegeben durch

$$\widehat{\theta} = \frac{1}{4} \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right).$$

c) (2 Punkte) Wie gross ist der Erwartungswert und die Varianz von $\widehat{\theta}$, falls die Daten (mit $n = 5$) von obiger Verteilung mit $\theta = 1$ stammen?

Hinweis: $\mathrm{Var}(X) = 4 \cdot \theta^2$.

$$\mathbf{E}\left[\hat{\theta}\right] = \frac{1}{4n} \sum_i \mathbf{E}[X_i] = \theta = 1$$

(1 Punkt) und

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{16n^2} \sum_i \mathrm{Var}(X_i) = \frac{1}{4n} \theta^2 = 1/20$$

(1 Punkt).

d) (1 Punkt) Was ist die Wahrscheinlichkeit, dass der Maximum-Likelihood Schätzer bei einer Stichprobe von 5 Werten gerade dem wahren Parameter entspricht? $P(\hat{\theta} = \theta) = 0$ (1 Punkt).

4. **a)** t-Test für $U_i \sim \mathcal{N}(\mu, \sigma^2)$:

- $H_0 : \mu = 0, \quad H_A : \mu \neq 0$ (0.5 Pt)
- Teststatistik $T = \frac{\bar{U}_n}{S_n/\sqrt{n}} \sim t_{n-1}$ unter H_0 (0.5 Pt)
- Verwerfungsbereich $K = (-\infty, -t_{8,0.975}] \cup [t_{8,0.975}, \infty) = (-\infty, -2.31] \cup [2.31, \infty)$ (0.5 Pt)
- Testentscheid: $t = 2.53 \in K$, also wird H_0 verworfen. (0.5 Pt)

Notwendige Annahme: die Daten sollten normalverteilt sein (sonst stimmt die Verteilung der Teststatistik und damit Verwerfungsbereich und Niveau nicht). (1 Pt)

b) Das zweiseitige Vertrauensintervall für den t-Test ist gegeben durch $I = \bar{u}_9 \pm \frac{s_9}{\sqrt{9}} t_{8,0.975}$ (0.5 Pt). Also $I = [1.27, 26.95]$ (0.5 Pt).

c) Vorzeichentest für $X_i \sim F$:

- $H_0 : F^{-1}(0.5) = 0, \quad H_A : F^{-1}(0.5 \neq 0)$ (0.5 Pt)
- Teststatistik $T = \#\{X_i > 0\} \sim B(9, 0.5)$ unter H_0 (0.5 Pt)
- Verwerfungsbereich: Hierfür benötigen wir die folgenden Werte der Binomialverteilung:
  - $P(T = 0) = P(T = 9) = 0.002$
  - $P(T = 1) = P(T = 8) = 0.018$
  - $P(T = 2) = P(T = 7) = 0.07$

  Für den Verwerfungsbereich K gilt $P_{H_0}(T \in K) \leq 5\%$, somit ist $K = \{0; 1; 8; 9\}$ (0.5 Pt)
- Testentscheid: $T = 7$, also wird H_0 *nicht* verworfen (0.5 Pt)

d) Die Macht ist umso grösser je weiter die Alternativhypothese von der Nullhypothese entfernt ist, also berechnet Alain eine grössere Macht. (1 Pt) Für $H_A : \mu \to 0$ stimmen Null- und Alternativhypothese überein, also ist Macht gleich Niveau (in diesem Fall 5%). (1 Pt)

5. (**12 Punkte**)

a) Die Ereignisse A und B seien disjunkt und $\mathcal{P}(A) > 0$ sowie $\mathcal{P}(B) > 0$. Dann gilt in diesem Fall

	Richtig	Falsch
• $\mathcal{P}(A \mid A) = 1$	○	○
• $\mathcal{P}(A \cap B) = 0$	○	○
• $1 - \mathcal{P}(A \cup B) > \mathcal{P}(A^c) + \mathcal{P}(B^c) - 1$	○	○
• $\mathcal{P}(A^c \cup B) = 1 - \mathcal{P}(A)$	○	○

R,R,F,R

b) Betrachten Sie folgende Aussagen zu stetigen Zufallsvariablen.

	Richtig	Falsch
• Sei $X \sim \mathcal{N}(8, 4)$, dann ist $\mathcal{P}(X \geq 4) = 0.5$.	○	○
• Für jede beliebige stetige Zufallsvariable X mit $\mathcal{E}[X] = \mu$ gilt $$\mathcal{E}[X - \mu] = 0.$$	○	○
• Die gemeinsame Dichte von X und Y sei unbekannt. Sie können diese aber immer aus der Randdichte $f_X(x)$ und der bedingten Dichte $f_{Y\|X=x}(y)$ bestimmen, falls Sie diese kennen.	○	○
• Die Zufallsvariable U sei Uni(0,1)-verteilt und $\lambda > 0$ eine Konstante. Dann ist $$-\frac{\log(1-U)}{\lambda}$$ eine Exp(λ)-verteilte Zufallsvariable.	○	○

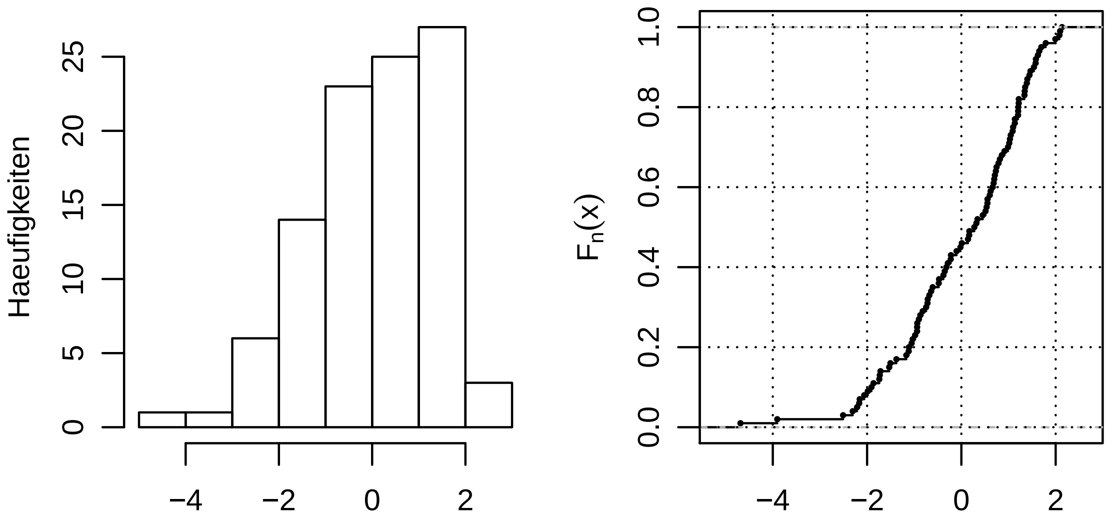
F,R,R,R

c) Wir betrachten den statistischen Test $H_0 : \theta = \theta_0$ und $H_A : \theta \neq \theta_0$, wobei θ ein Parameter in einem Modell und θ_0 ein vorgegebener Sollwert ist.

	Richtig	Falsch
• Wenn die Nullhypothese auf dem Signifikanzniveau 5% nicht verworfen wird, so überdeckt das entsprechende 95%-Vertrauensintervall für θ den Wert θ_0 nicht.	○	○
• Wenn wir H_0 verwerfen, obwohl H_0 in Tat und Wahrheit stimmt, dann machen wir einen Fehler 1. Art.	○	○
• Mit obigen Angaben ist es möglich, die Macht des Tests zu berechnen.	○	○
• Falls wir H_0 verwerfen, wenn immer der p-Wert kleiner gleich 10% ist (und sonst nicht), so haben wir das Signifikanzniveau kontrolliert auf dem Niveau 10%.	○	○

F, R, F, R

d) Betrachten Sie das Histogramm und die empirische kumulative Verteilungsfunktion einer Stichprobe mit 100 (verschiedenen) Beobachtungen.



	Richtig	Falsch
• Der Median ist (leicht) grösser als 0.	◯	◯
• Ca. 20% der Daten sind kleiner als -1.	◯	◯
• Nur 1 Beobachtung ist (strikt) kleiner als -4.	◯	◯
• Die Daten sind symmetrisch verteilt.	◯	◯

R, R, R, F

e) Betrachten Sie folgende Aussagen zu statistischen Tests.

	Richtig	Falsch
• Der Annahmebereich bei einem t-Test (für eine Stichprobe) entspricht gerade dem Vertrauensintervall für den Erwartungswert.	◯	◯
• Bei gepaarten Vergleichen arbeitet man stets mit den Differenzen der einzelnen Messwert-Paare und verwendet dann gewöhnliche Methoden für eine Stichprobe.	◯	◯
• Gepaarte Vergleiche kann man sicher nur dann durchführen, wenn die beiden Stichproben gleich gross sind.	◯	◯
• Wenn man bei einem Z-Test die Anzahl Beobachtungen verdoppelt, dann wird das 95%-Vertrauensintervall für den Erwartungswert auch doppelt so breit.	◯	◯

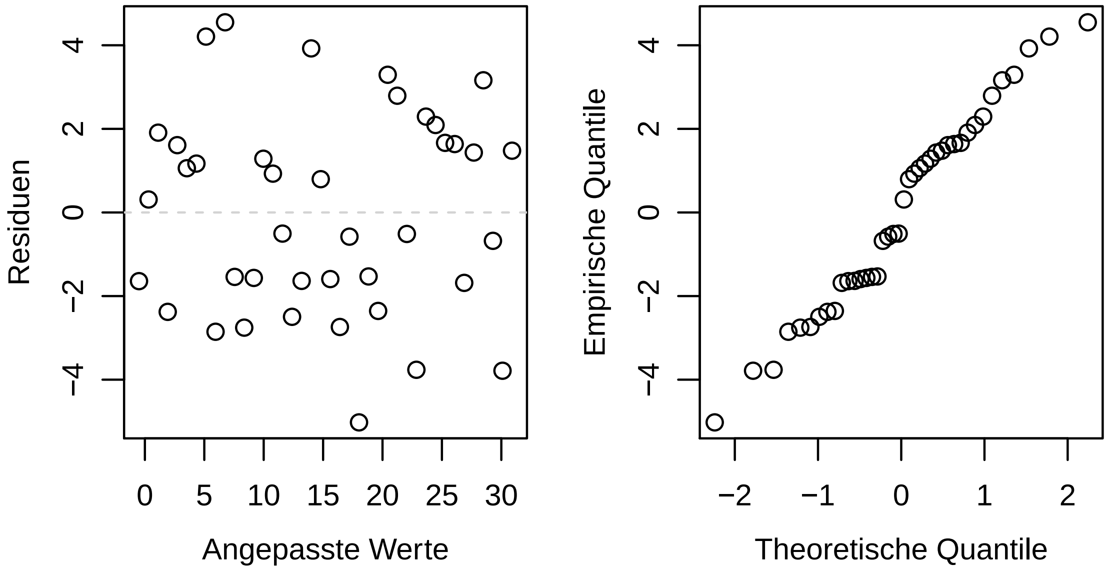
F, R, R, F

f) Sie haben für einen Datensatz $(x_i, y_i), i = 1, \ldots, n$ das einfache lineare Regressionsmodell

$$Y_i = \beta_0 + \beta_1 x_i + E_i,\ E_i \text{ i.i.d. } \sim \mathcal{N}(0, \sigma^2)$$

angepasst und folgenden Output bzw. Plots erhalten.

```
Estimated Coefficients:
                   Estimate     SE          tStat       pValue
    (Intercept)     3.3287       1.0918     3.0489      0.0093195
    x1             0.46741     0.12008      3.8925      0.0018521
```



	Richtig	Falsch
• Die Modellvoraussetzungen sind hier gut erfüllt.	○	○
• Es ist $\widehat{\beta}_1 = 0.46741$.	○	○
• Die Nullhypothese $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$ wird auf dem 5%-Niveau verworfen.	○	○
• Den Wert der Geraden an der Stelle $x = 10$ schätzen wir hier (gerundet) mit 8.	○	○

R, R, R, R