Musterlösung

1. **a)** Richtig, da $Var(X) = n \cdot p \cdot (1 - p) = 4.8$.
b) Richtig, da $E[X] = n \cdot p = 20 \times 0.4 = 8$.
c) Richtig.
d) Richtig.

2. **a)** Richtig, da $\alpha = 0.05$.
b) Richtig, da P (Fehler 2. Art) = P (Beibehalten von H_0 falls H_A stimmt)=P(X < 9)=0.6242
c) Richtig, da Macht = 1 - P (Fehler 2. Art).
d) Falsch.

3. **a)** Richtig, $T \sim Bin(n = 13, p = 0.4)$ und der Verwerfungsbereich ist $K = [0, c]$. Aus $P(X = 0) = 0.001306$, $P(X = 1) = 0.0113$ und $P(X = 2) = 0.0453$, daraus folgt $c = 1$.
b) Falsch, der Verwerfungsbereich ist in diesem Fall definiert als $K = \{0, ..., c\}$, wobei c die grösste ganze Zahl ist, so dass $P_{H_0}(X \in \{0, ..., c\}) \leq 5\%$
c) Richtig, der Verwerfungsbereich wird dann weniger Elemente enthalten.
d) Falsch, der Testentscheid beruht auf dem Widerspruchs-Prinzip: die Null-Hypothese kann nur falsifiziert und nicht verifiziert werden.

4. **a)** Falsch. Die Macht des einseitigen Tests ($H_A : p > 0.6$) wäre grösser als die des zweiseitigen Tests.
b) Richtig.
c) Falsch, der P-Wert ist $P_{p=0.5}(X \geq 53) = 0.31$
d) Richtig, der zweiseitige Test eignet sich dazu sehr kleine und sehr grosse Erfolgswahrschienlichkeiten zu erkennen.

5. **a)** Richtig, da es sich hier um die Anzahl Würfe mit Augenzahl 4 bei $n = 93$ unabhängigen Würfen handelt und P(Augenzahl = 4) = $1/5 = 0.2$.
b) Falsch. Als Faustregel gilt, dass die Normalapproximation gut ist, falls $n\pi > 5$ und $n(1 - \pi) > 5$. Das ist hier nicht der Fall, da $\pi = 0.1$ und somit $n\pi = 0.8$.
c) Falsch, $Pois(\lambda = 134)$
d) Falsch, die Dichtefunktion der Exponentialverteilung ist stetig und daher ist $P(X = x) = 0$ für alle $x \geq 0$.

6. **a)** Richtig, weil jeder Beobachtung vor Einnahme des Medikaments eine Beobachtung nach Einnahme des Medikaments zugeordnet werden kann.
b) Falsch.
c) Falsch, Unabhängigkeit wird vorausgesetzt.
d) Falsch, das spielt keine Rolle, da man nur an der Differenz der beiden Werte interessiert ist.

7. **a)** Richtig.
b) Falsch. Die richtige Lösung ist $\sqrt{93} \cdot 9.81/7.01 = 13.49$.
c) Richtig. Die Verteilung von T unter H_0 ist $T \sim t_{n-1} = t_{92}$.
d) Richtig, da μ_D unter H_0 gleich 0 ist und 0 nicht innerhalb des Vertrauensintervalls liegt, wird die Nullhypothese verworfen.

8. **a)** Falsch, diese Aussage stimmt zwar in den meisten Fällen, kann aber durch Datenpunkte, die sehr nahe beieinanderliegen, widerlegt werden. Für nahe beieinanderliegende Datenpunkte kann es sein, dass $\hat{\sigma}$(t-test) $<$ σ(z-test). Damit wäre dann das Vertrauensintervall des t-Tests kleiner.

b) Falsch. Der P-Wert könnte zwischen 0.5% und 1% sein.

c) Richtig. $[9.81 \pm 7.01 * 2.63/\sqrt{93}] = [7.90, 11.72]$

d) Richtig.

9. **a)** Richtig. Der Wert der t-Teststatistik ist $16.765/5.035 = 3.3$. Mit Hilfe der t-Tabelle sieht man, dass der Effekt von β_0 auf 1% signifikant ist.

b) Richtig. Es wurden 36 Waldstücke betrachtet und es gibt zwei Parameter im Modell (β_0 und β_1). Somit ergeben sich 36-2 = 34 Freiheitsgrade.

c) Falsch. Die Schätzung berechnet sich als $\hat{s.e.}(\hat{\beta}_1) = \frac{\hat{\beta}_1}{t} = 0.104$.

d) Falsch.

10. **a)** Richtig. Der Achsenabschnitt der geschätzten Regressionsgeraden ist $16.765 \approx 16.8$.

b) Richtig. Das geschätzte Modell ist Ertrag $= 16.765 + 0.441 \cdot$ Aufforstung. Somit ergibt sich für einem erwarteten Ertrag von $41 \frac{m^3}{ha}$ eine Aufforstung von $(41 - 16.765)/0.441 = 55$ Stunden.

c) Richtig.

d) Falsch. Das Vorhersageintervall beschreibt den Bereich, in welchem sich der **wahre Ertrag** für ein Waldstück mit $90h$ Aufforstung mit 99% Wahrscheinlichkeit befindet.

11. **a)** Falsch. Da das Modell linear in den Koeffizienten (und nicht in den erklärenden Variablen) ist, handelt es sich um eine lineare Regression.

b) Falsch. Der TA Plot zeigt eine systematische Verteilung der Fehler.

c) Richtig.

d) Falsch.

12. **a)** Richtig, da $E[4Y - 3X - 10] = 4E[Y] - 3E[X] - 10 = 2$.

b) Richtig, da $Var(4Y - 2X + 5) = 16Var(Y) + 4Var(X) = 64$.

c) Richtig. Es gilt $Var(X - 1/2) = Var(X) = 3^2 = 9$.

d) Falsch. Korrelation impliziert keinen kausalen Zusammenhang. So kann man zeigen, dass die Anzahl Piraten mit der globalen Erderwärmung korreliert, ein kausaler Zusammenhang aber wohl eher fragwürdig scheint.

13. **a)** Falsch, es gilt $P(A \cap B) = 0.06$.

b) Richtig. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.5 - 0.2 = 0.6$. Deshalb $P((A \cup B)^C) = 1 - P(A \cup B) = 1 - 0.6 = 0.4$.

c) Falsch. $P(A|B) = P(B|A) * \frac{P(A)}{P(B)} \Rightarrow \frac{P(A)}{P(B)} < 1 \Rightarrow P(A) < P(B)$.

d) Richtig. Definiere die folgenden Ereignisse:

$$\begin{aligned} B &= \text{Frau hat Brustkrebs} \\ B^c &= \text{Frau hat keinen Brustkrebs} \\ Pos &= \text{positives Testresultat} \\ Neg &= \text{negatives Testresultat.} \end{aligned}$$

Gesucht ist $P(B|Pos) = \frac{P(Pos|B) \cdot P(B)}{P(Pos)}$, wobei wir den Satz von Bayes angewendet haben. Die Wahrscheinlichkeit ein positives Testresultat zu erhalten berechnet sich als $P(Pos) = P(Pos|B^c)P(B^c) + P(Pos|B)P(B) = (1 - P(Neg|B^c))P(B^c) + P(Pos|B)P(B) = (1 - 0.8) * 0.98 + 0.9 * 0.02 = 0.214$. Deshalb ist $P(B|Pos) = \frac{0.9 \cdot 0.02}{0.214} \approx 0.084$.

14. **a)** Richtig.

b) Falsch, es wäre $\frac{Z-4}{2} \sim \mathcal{N}(0,1)$.

c) Richtig.

d) Falsch. Es befinden sich 68% aller Werte innerhalb von jeweils einer Standardabweichung vom Mittelpunkt entfernt.

15. **a)** Falsch. Damit das zutreffen würde, müssten die Daten symmetrisch verteilt sein - da diese Daten hier aber linksschief verteilt sind, ist der Mittelwert kleiner als der Median.

b) Falsch. Die IQR beträgt etwa 1.5.

c) Richtig. Die gestrichelten Whiskers bezeichnen den Bereich, welchen wir als üblich erachten. Punkte ausserhalb sind vermutlich besonders einflussreich oder Ausreisser.

d) Falsch, sie sind linksschief.

16. **a)** Falsch. Das Eintreten des Ereignisses F ist 2 mal so wahrscheinlich, wie das Eintreten des Komplements.

b) Richtig. Definiere das Gesamtgewicht als $Z = \sum_{i=1}^{30} X_i$ mit $X_i :=$ das Gewicht vom i-ten Lachs. Gemäss des Zentralen Grenzwertsatzes folgt das Gesamtgewicht einer Normalverteilung mit Erwartungswert $10 * 30$ und Standardabweichung $2\sqrt{30}$. Deshalb gilt

$$\begin{aligned} P(Z > 330) &= P(\frac{Z - 10 * 30}{2\sqrt{30}} > \frac{330 - 300}{2\sqrt{30}}) \\ &= P(\frac{Z - 10 * 30}{2\sqrt{30}} > 2.74) = 1 - P(Z^* \leq 2.74) \\ &= 1 - 0.9969 = 0.0031 < 0.01. \end{aligned}$$

Hier ist $Z^* := \frac{Z-10*30}{2\sqrt{30}} \sim \mathcal{N}(0,1)$ das standardisierte Gesamtgewicht.

c) Falsch. Das $\sqrt{n}$-Gesetz besagt, dass es für eine Verbesserung der Genauigkeit um den Faktor 5 das $5^2 = 25$-fache an Messungen braucht.

d) Richtig, nämlich $0.5 * 1 + 0.5 * 49/99 = \frac{1.494949...}{2} = 0.747474... \approx 0.747$.