Musterlösung

1. **a)** Richtig, da es sich hier um die Anzahl Würfe mit Augenzahl 3 bei $n = 126$ unabhängigen Würfen handelt und P(Augenzahl $= 3$) $= 1/8 = 0.125$.

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 $n\pi = 0.7$.

c) Falsch, $Pois(\lambda = 1324)$

d) Falsch, die Dichtefunktion der Exponentialverteilung ist stetig und daher ist $P(X = x) = 0$ für alle $x \geq 0$.

2. **a)** Richtig, da $Var(X) = n \cdot p \cdot (1 - p) = 2.4$.

b) Richtig, da $E[X] = n \cdot p = 15 \times 0.2 = 3$.

c) Richtig.

d) Falsch. $P[X = 15] = 3 * 10^{-11}$

3. **a)** Falsch. Die Macht des einseitigen Tests ($H_A : p > 0.5$) 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.

4. **a)** Richtig, $T \sim Bin(n = 11, p = 0.4)$ und der Verwerfungsbereich ist $K = [0, c]$. Aus $P(X = 0) = 0.0036$, $P(X = 1) = 0.027$ und $P(X = 2) = 0.089$, 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.

5. **a)** Richtig, da $\alpha = 0.05$.

b) Richtig, da P (Fehler 2. Art) = P (Beibehalten von H_0 falls H_A stimmt)=P(X > 1)=0.6242

c) Richtig, da Macht = 1 - P (Fehler 2. Art).

d) Falsch.

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 $\sqrt{93} \cdot 3.87/2.01 = 18.57$.

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) $< \sigma$(z-test). Damit wäre dann das Vertrauensintervall des t-Tests kleiner.

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

c) Falsch. Mit dem angegebenen Quantil berechnet sich das Vertrauensintervall als $[3.87 \pm 2.63 * 2.01/\sqrt{93}] = [3.32, 4.42]$.

d) Richtig.

9. **a)** Falsch, beim z-Test muss die Standardabweichung im Vergleich zum t-Test bereits bekannt sein.

b) Falsch. Das 99%-Vertrauensintervall ist grösser als das 95%-Vertrauensintervall und enthält dieses komplett. Da die 2 im 95%-Vertrauensintervall liegt, ist sie auch im 99%-Vertrauensintervall enthalten.

c) Falsch. Die t-Verteilung ist langschwänziger.

d) Falsch. Beim t-Test müssen die Daten normalverteilt sein.

10. **a)** Falsch. Es wurden 35 Waldstücke betrachtet und es gibt zwei Parameter im Modell (β_0 und β_1). Somit ergeben sich 35-2 = 33 Freiheitsgrade.

b) Richtig. Der Wert der t-Teststatistik ist $10.4653/3.1174 = 3.357$. Mit Hilfe der t-Tabelle sieht man, dass der Effekt von β_0 auf 1% signifikant ist.

c) Falsch. Die Schätzung berechnet sich als $\hat{s.e.}(\hat{\beta}_1) = \frac{\hat{\beta}_1}{t} = 0.06565$.

d) Richtig.

11. **a)** Richtig. Das geschätzte Modell ist Ertrag $= 10.46533 + 0.79637 \cdot$ Aufforstung. Somit ergibt sich für einem erwarteten Ertrag von $67\frac{m^3}{ha}$ eine Aufforstung von $(67 - 10.46533)/0.79637 \approx 71$ Stunden.

b) Richtig. Der Achsenabschnitt der geschätzten Regressionsgeraden ist $10.46533 \approx 10.5$.

c) Richtig.

d) Falsch. Das Vorhersageintervall beschreibt den Bereich, in welchem sich der **wahre Ertrag** für ein Waldstück mit $80h$ Aufforstung mit 95% Wahrscheinlichkeit befindet.

12. **a)** Falsch.

b) Falsch. Der TA Plot zeigt eine Trichterform.

c) Richtig.

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

13. **a)** Richtig, da $E[4Y - 2X + 5] = 4E[Y] - 2E[X] + 5 = 8 - 14 + 5 = -1$.

b) Richtig, da $Var(4Y - 2X + 5) = 16Var(Y) + 4Var(X) = 16 * 3 + 16 = 64$.

c) Richtig. Es gilt $Var((X - 5)/6) = 1/36 \cdot Var(X) = 1/36 \cdot 12 = 1/3$.

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.

14. **a)** Richtig. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.6 - 0.2 = 1.0 - 0.2 = 0.8$. Deshalb $P((A \cup B)^C) = 1 - P(A \cup B) = 1 - 0.8 = 0.2$.

b) Falsch, es gilt $P(A \cap B) = 0.08$.

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

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 $P(Pos) = P(Pos|B^c)P(B^c) + P(Pos|B)P(B) = (1 - P(Neg|B^c))P(B^c) + P(Pos|B)P(B) = (1 - 0.9) * 0.99 + 0.8 * 0.01 = 0.107$. Deshalb ist $P(B|Pos) = \frac{0.8 \cdot 0.01}{0.107} \approx 0.075$.

15. **a)** Richtig, da die Verteilung symmetrisch um $\mu = 2$ ist.

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

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

d) Richtig.

16. **a)** Falsch. Die IQR beträgt etwa 1.5.

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

c) Richtig.

d) Falsch. Damit das zutreffen würde, müssten die Daten symmetrisch verteilt sein - da diese Daten hier aber rechtsschief verteilt sind, ist der Mittelwert grösser als der Median.

17. **a)** Richtig.

b) Falsch, die Zufallsvariabeln müssen unabhängig sein. Falls z.B. $Y = X$, dann gilt nicht $X + Y = 2X \sim Pois(2\lambda_1)$.

c) Richtig. Wir haben unbeschränkte Zähldaten.

d) Richtig. Mit den Satz von Bayes haben wir $P(X_1 = x_1|X_1 + X_2 = k) = \frac{P(X_1+X_2=k|X_1=x_1) \cdot P(X_1=x_1)}{P(X_1+X_2=k)} = \frac{P(X_2=k-x_1) \cdot P(X_1=x_1)}{P(X_1+X_2=k)}$. Da X_1 und X_2 unabhängig ist $X_1 + X_2 \sim Pois(2)$. Deshalb $P(X_1 = x_1|X_1 + X_2 = k) = \frac{e^{-1}}{(k-x_1)!} \cdot \frac{e^{-1}}{x_1!} \cdot \frac{k!}{2^k e^{-2}} = \binom{k}{x_1} 2^{-k+x_1} 2^{-x_1}$.

18. **a)** Richtig. Da $n = 9$ und $\alpha = 0.4$, folgt $n \cdot \alpha = 3.6$ und somit ist $q_{0.4} = x_{(4)} = 7$.

b) Falsch. Es gilt $P(Y \geq 1) = 1 - P(Y \leq 0)$.

c) Richtig. $E(X) = 1 * 0 + 2 * 0.1 + 3 * 0.3 + 4 * 0.2 + 5 * 0.1 + 6 * 0.3 + 7 * 0 + 8 * 0 = 4.2$.

d) Falsch, der dritte Wert sollte 0.4 betragen und nicht 0.3.

19. **a)** Richtig.

b) Richtig. Die Studentsche t-Verteilung ist um den Nullpunkt herum symmetrisch.

c) Falsch. $P(X = x) = 0$ für stetige Zufallsvariablen.

d) Richtig. Die Wahrscheinlichkeit $P(X \leq 1)$ ist gerade die Fläche unter der Dichtefunktion bis $x = 1$. Deshalb gilt: $P(X \leq 1) = 3 * 0.1 + 1 * 0.3 = 0.6$.

20. **a)** Falsch. Das Eintreten des Ereignisses F ist 3 mal so wahrscheinlich, wie das Eintreten des Komplements.

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

c) Richtig. Gemäss ZGS folgt das Gesamtgewicht Z einer Normalverteilung mit Erwartungswert 4500 und Varianz 2025.

$$P(Z > 4600) = P(\frac{Z - 4500}{\sqrt{2025}} > \frac{4600 - 4500}{\sqrt{2025}}) = P(Z^* > 2.222) = 1 - P(Z^* \leq 2.222).$$

Mit $Z^* = \frac{Z-4500}{\sqrt{2025}} \sim \mathcal{N}(0,1)$ kann man die gesuchte Wahrscheinlichkeit in der Tabelle nachschauen. Deshalb ist die Wahrscheinlichkeit für einen Wert über 4600 kg sicher kleiner als 5%.

d) 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.