Musterlösung

1. **a)** Richtig, da $E[X] = n \cdot p = 15 \times 0.2 = 3$.
b) Falsch, da $Var(X) = n \cdot p \cdot (1-p) = 2.4$.
c) Falsch. Die Erfolgswahrscheinlichkeit ist $p = 0.2$ und nicht 0.8.
d) Richtig. $P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] = 0.398$

2. **a)** Richtig, da es sich hier um die Anzahl Erfolge (Kopf) bei $n = 25$ unabhängigen Münzwürfen handelt und P(Kopf) $= 0.5$.
b) Falsch, die Bernoulliverteilung realisiert nur ob es einen Erfolg gibt, nicht aber eine Anzahl.
c) Falsch. Es muss mit der hypergeometrischen Verteilung berechnet werden. Es handelt sich um die Wahrscheinlichkeit für einen Gewinn aus $n = 3$ Versuchen einer Population von $N = 5$ Losen und $K = 3$ Erfolgen ohne Zurücklegen.
d) Richtig. Als Faustregel gilt, dass die Normalapproximation gut ist, falls $n\pi > 5$ und $n(1-\pi) > 5$. Das ist hier der Fall.

3. **a)** Falsch, $T \sim Bin(n = 10, p = 0.5)$ und der Verwerfungsbereich ist $K = [c, 10]$. Aus $P(X = 10) = 0.001$, $P(X = 9) = 0.01$ und $P(X = 8) = 0.04$ folgt $c = 9$.
b) Richtig.
c) Falsch, die Macht berechnet sich als
$P_{H_A}(X \in K) = P_{p=0.7}(X = 9) + P_{p=0.7}(X = 10) = 0.12 + 0.03 = 0.15$.
d) Falsch, es ist genau der errechnete Wert oder einen <u>noch extremeren</u>!

4. **a)** Richtig.
b) Richtig, das Vertrauensintervall ist gegeben durch

$$I \approx \frac{28}{50} \pm 1.96\sqrt{\frac{28}{50}(1-\frac{28}{50})\frac{1}{50}} = [0.4224, 0.6976]$$

c) Falsch, $T \sim Bin(15, 0.2)$.
d) Falsch, der p-Wert gibt das kleinste Signifikanzniveau an, bei dem die Nullhypothese (gerade noch) verworfen wird. Wir haben $0.06 > 0.05$.

5. **a)** Richtig, weil jeder Beobachtung aus Basel genau eine Beobachtung aus Zürich zugewiesen werden kann.
b) Falsch.
c) Richtig.
d) Falsch.

6. **a)** Richtig
b) Falsch, es ist $\sqrt{n} = \sqrt{7}$ und nicht $\sqrt{n-1}$.
c) Richtig, die Verteilung von T unter H_0 ist $T \sim t_{n-1} = t_6$.
d) Falsch, da μ_D unter H_0 gleich 0 ist und 0 innerhalb des Vertrauensintervalls liegt, können wir die Nullhypothese eben nicht verwerfen.

7. **a)** Falsch, durch die Schätzung von σ im Falle des t-Tests wird dort die t-Verteilung verwendet (statt die Normalverteilung), diese ist aber weiter als die Normalverteilung und somit sind die Grenzen des Verwerfungsbereiches "weiter aussen" als beim z-Test. Der Verwerfungsbereich ist somit kleiner.

b) Falsch.

c) Falsch, sie unterscheiden sich auch in der verwendeten Standardabweichung.

d) Falsch, der beobachtete Wert der Teststatistik liegt nicht im Verwerfungsbereich $K = (-\infty; -t_{14,0.975}] \cup [t_{14,0.975}, \infty)$. Damit ist der P-Wert grösser als 5%.

8. a) Falsch

b) Falsch

c) Falsch

d) Falsch, der t-Test setzt ebenfalls normalverteilte Daten voraus.

9. a) Richtig. Die degrees of freedom sind 47 und es gibt zwei Parameter im Modell (β_0 und β_1). Also wurden $47 + 2 = 49$ Messungen gemacht.

b) Falsch. Der bobachtete Wert der Teststatistik liegt nicht im Verwerfungsbereich $K = (-\infty; -t_{47,0.995}] \cup [t_{47,0.995}; \infty)$.

c) Richtig. Die Schätzung berechnet sich als $\hat{\beta}_1 = se(\beta_1) \cdot t(\beta_1) = 0.185 \cdot 10.7$.

d) Falsch. Das Vertrauensintervall enthält die Null. Das exakte zweiseitige 95%-Vertrauensintervall für β_1 ist:
$[\hat{\beta}_1 - t_{47,0.975} \cdot se(\hat{\beta}_1), \hat{\beta}_1 + t_{47,0.975} \cdot se(\hat{\beta}_1)] = [1.98 - 2.01 \cdot 0.185, 1.98 + 2.01 \cdot 0.185]$.

10. a) Falsch: $-0.419 + 0.83 * 1.5 = 0.676$.

b) Falsch: Die Steigung ist 0.83.

c) Falsch. Es ist das Vertrauensintervall gegeben. Um die Frage beantworten zu können, bräuchten wir das Vorhersageintervall.

d) Richtig.

11. a) Richtig.

b) Falsch, da das Modell linear in den Koeffizienten (und nicht in den erklärenden Variablen) sein muss.

c) Falsch.

d) Falsch.

12. a) Richtig. Die Normalverteilung ist um den Erwartungswert herum symmetrisch.

b) Richtig. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.4 - 0.2 = 1 - 0.2 = 0.8$.

c) Richtig. Die odds für A sind $P(A)/(1 - P(A)) = 0.75/0.25 = 3$.

d) Richtig. $P(A|B) = P(B|A) * \frac{P(A)}{P(B)} = P(B|A) * 0.7 < P(B|A)$.

13. a) Richtig. Die dicke Linie im Rechteck zeigt den Median an. Diese ist etwa bei 6.

b) Falsch. Nur 25% der Daten befinden sich unterhalb des Rechtecks, also sind weniger als 25% der Daten kleiner als 4.

c) Falsch. Die Daten sind linksschief, somit ist der Mittelwert kleiner als der Median.

d) Falsch. Da nur ein Datenpunkt verändert wurde und auf der gleichen Seite des Medians geblieben ist, ändert sich der Wert des Medians nicht.

14. a) Richtig, da $E[2X + 4 - Y] = 2E[X] + 4 - E[Y] = 2 * 3 + 4 - 10 = 0$

b) Falsch, die richtige Antwort ist: $Var(2X + 4 - Y) = 4Var(X) + Var(Y) = 4 * 2 + 1 = 9$.

c) Falsch. Der erste Teil der Aussage bezieht sich auf $P(A|B)$, der zweite Teil aber auf $P(B|A)$. Im Allgemeinen sind diese beiden Grössen nicht gleich.

d) Falsch. Es könnte einen nichtlinearen Zusammenhang geben.

15. **a)** Richtig. Gemäss ZGS folgt das Gesamtgewicht einer Normalverteilung mit Erwartungswert 6300 und Varianz 5760. Das 99%-Quantil dieser Verteilung ist 6477. Also ist die Wahrscheinlichkeit für einen Wert über 6500 sicher kleiner als 1%.

b) Falsch. Das $\sqrt{n}$-Gesetz gilt für die Standardabweichung des Mittelwerts, nicht für die Standardabweichung einer Einzelbeobachtung. Die geschätzte Standardabweichung wird sich substantiell nicht ändern, wenn man mehr Beobachtungen macht.

c) Richtig. Die Wahrscheinlichkeit kann mit der hypergeometrischen Verteilung berechnet werden.

d) Falsch. Die Varianz von Y ist $\mathrm{Var}(Y) = (1/5)^2\,\mathrm{Var}(X) = \frac{1}{5}$ und nicht 1. Richtig wäre $\sqrt{5}$ im Nenner.

16. **a)** Falsch. $P(X \leq 2) = 4 * 0.2 + 1 * 0.1 = 0.9$.

b) Falsch. Zum Beispiel stimmt die Gleichung nicht für die kumulative Verteilungsfunktion der Wahrscheinlichkeitsdichte

$$f(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases}$$

c) Falsch. Die angegebene Formel gilt für die Wahrscheinlichkeitsdichte.

d) Richtig.