Musterlösung

1. **a)** Richtig. Als Faustregel gilt, dass die Normalapproximation gut ist, falls $n\pi > 5$ und $n(1-\pi) > 5$. Das ist hier knapp der Fall.

b) Richtig, da es sich hier um die Anzahl Erfolge (Kopf) bei $n = 25$ unabhängigen Münzwürfen handelt und P(Kopf) $= 0.5$.

c) Falsch, die Bernoulliverteilung realisiert nur ob es einen Erfolg gibt, nicht aber eine Anzahl - die Überlegung der Binomialverteilung basiert aber auf einer Wiederholung von n Bernoulliverteilten.

d) Richtig, es handelt sich um die Wahrscheinlichkeit für einen Gewinn aus $n = 3$ Versuchen einer Population von $N = 7$ und $K = 3$ Erfolgen ohne Zurücklegen.

2. **a)** Richtig, da $E[X] = n \cdot p = 12 \times 0.1 = 1.2$.

b) Falsch, da $Var(X) = n \cdot p \cdot (1-p) = 1.08$.

c) Falsch. Die Erfolgswahrscheinlichkeit ist $p = 0.1$ und nicht 0.9.

d) Richtig. $P[X <= 2] = P[X = 0] + P[X = 1] + P[X = 2] = 0.88913$

3. **a)** Richtig.

b) Falsch. Die Macht des einseitigen Tests ($H_A : p > 0.2$) wäre viel grösser als die des zweiseitigen Tests.

c) Falsch, $T \sim Bin(28, 0.2)$.

d) Richtig, der P-Wert gibt das kleinste Signifikanzniveau an, bei dem die Nullhypothese (gerade noch) verworfen wird und $0.03 < 0.05$.

4. **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, der Verwerfungsbereich wird dann weniger Elemente enthalten.

c) Richtig.

d) Falsch, der Test-Entscheid beruht auf dem Widerspruchs-Prinzip: die Null-Hypothese kann nur falsifiziert und nicht verifiziert werden.

5. **a)** Falsch, die Macht ist 1 - P(Fehler 2. Art).

b) Richtig, da Macht = 1 - P (Fehler 2. Art) = P (Verwerfen von H_0 falls H_A stimmt).

c) Falsch, die Macht berechnet sich als
$P_{H_A}(X \in K) = P_{p=0.2}(X = 0) + P_{p=0.2}(X = 1) = 0.11 + 0.27 = 0.38$.

d) Richtig, die Macht für die Alternative $H_A : p = 0.3$ ist 0.15, also kleiner.

6. **a)** Falsch, es ist genau der errechnete Wert oder einen noch extremeren!

b) Falsch, der P-Wert ist $P_{p=0.5}(X \geq 8) = P_{p=0.5}(X = 8) + P_{p=0.5}(X = 9) + P_{p=0.5}(X = 10) = 0.011 > 0.01$.

c) Richtig, das Vertrauensintervall ist gegeben durch

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

d) Richtig.

7. **a)** Falsch, weil jeder Beobachtung aus Basel genau eine Beobachtung aus Zürich zugewiesen werden kann.

b) Richtig.

c) Richtig.

d) Falsch.

8. **a)** Richtig, man möchte im Sinne der Stadt Zürich statistisch nachweisen, dass die Konzentration von MDMA im Abwasser statistisch nicht signifikant grösser ist als in Basel. Dabei handelt es sich um eine einseitige Alternativhypothese von der Form $H_A : \mu_D > 2$.

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.

9. **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) Richtig.

c) Falsch. Es wurde das Quantil der Normalverteilung und nicht das der t-Verteilung verwendet.

d) Falsch. Das t-Quantil mit 15 Freiheitsgraden und 97.5% ist 2.131 und das t-Quantil mit 15 Freiheitsgraden und 95% ist 1.75. Der Wert der Teststatistik liegt sehr viel näher an 2.131. Daher ist der P-Wert etwa 5% und nicht 10%.

10. **a)** Richtig

b) Falsch

c) Falsch

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

11. **a)** Falsch. Die degrees of freedom sind 49 und es gibt zwei Parameter im Modell (β_0 und β_1). Also wurden $49 + 2 = 51$ Messungen gemacht.

b) Falsch. Der Verwerfungsbereich ist etwa (kann aus unserer Tabelle mit nur einer Nachkommastelle abgelesen werden, aber das reicht für diese Aufgabe): $K = (-\infty, -t_{49,0.975}] \cup [t_{49,0.975}, \infty) = (-\infty, -2.0] \cup [2.0, \infty)$ und $t = 1.43$ ist nicht in K.

c) Falsch. Die Schätzung berechnet sich als se(b)$\cdot$ t-value(b) $= 8.37 \cdot (0.172) = 1.44$.

d) Falsch. Das exakte zweiseitige 95%-Vertrauensintervall für β_1 ist (unter der o.g. Annahme):
$[\hat{\beta}_1 - t_{49,0.975} \cdot se(\hat{\beta}_1), \hat{\beta}_1 + t_{49,0.975} \cdot se(\hat{\beta}_1)] = [1.35 - 2.01 \cdot 0.172, 1.35 + 2.01 \cdot 0.172] = [1.00, 1.70]$.

12. **a)** Richtig: $0.304 + 0.83 * 1.5 = 1.549 \approx 1.5$.

b) Richtig: Die Steigung ist (laut Annahme) etwa $0.83 \approx 0.8$.

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

d) Falsch. Das Vorhersageintervall ist in der Regel grösser.

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

b) Richtig.

c) Richtig.

d) Falsch.

14. **a)** Richtig. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7 + 0.5 - 0.3 = 1.2 - 0.3 = 0.9$.

- **b)** Richtig. Die odds für A sind $P(A)/(1-P(A)) = 0.8/0.2 = 4$.
- **c)** Richtig. Die Normalverteilung ist um den Erwartungswert herum symmetrisch.
- **d)** Richtig.

15. **a)** Richtig, da $E[3X - Y - 3] = 3E[X] - E[Y] - 2 = 15 - 2 - 3 = 10$
- **b)** Falsch, die richtige Antwort ist: $Var(3X - Y - 3) = 9Var(X) + Var(Y) = 45 + 2 = 47$.
- **c)** Richtig. $P(A|B) = P(B|A) * \frac{P(A)}{P(B)} = P(B|A) * 0.7 < P(B|A)$.
- **d)** 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.

16. **a)** Richtig. Die dicke Linie im Rechteck zeigt den Median an. Diese ist etwa bei 4.
- **b)** Richtig.
- **c)** Falsch. Die Daten sind linksschief, somit ist der Mittelwert kleiner als der Median.
- **d)** Richtig. Da nur ein Datenpunkt verändert wurde und auf der gleichen Seite des Medians geblieben ist, ändert sich der Wert des Medians nicht.

17. **a)** Falsch.
- **b)** Richtig. Gemäss ZGS folgt das Gesamtgewicht einer Normalverteilung mit Erwartungswert 3500 und Varianz 2450. Das 99%-Quantil dieser Verteilung ist 3615. Also ist die Wahrscheinlichkeit für einen Wert über 3650 sicher kleiner als 1%.
- **c)** Falsch. Das $\sqrt{n}$-Gesetz gilt für den Mittelwert, nicht die Einzelbeobachtungen. Die geschätzte Standardabweichung wird sich nicht substantiell ändern, wenn man mehr Beobachtungen macht.
- **d)** Falsch. Der P-Wert ist die Wahrscheinlichkeit für die Beobachtung (oder etwas extremeres), falls die Nullhypothese stimmt.

18. **a)** Falsch. Es wurde mit der Varianz normalisiert und nicht mit der Standardabweichung.
- **b)** Richtig. Die Wahrscheinlichkeit $P(X \leq 2)$ ist gerade die Fläche unter der Dichtefunktion bis $x = 2$. Deshalb gilt: $P(X \leq 2) = 1 * 0.4 + 2 * 0.2 = 0.8$.
- **c)** Falsch. Es könnte einen nichtlinearen Zusammenhang geben.
- **d)** Falsch. Die Wahrscheinlichkeit kann mit der hypergeometrischen Verteilung berechnet werden.

19. **a)** Richtig.
- **b)** Richtig.
- **c)** Falsch. Das Quantil erfüllt die Gleichung $q_{0.95} = F^{-1}(0.95)$.
- **d)** Falsch. Die angegebene Formel gilt für die Wa.dichte.

20. **a)** Richtig. Die Anzahl verworfener Nullhypothesen X folgt der Verteilung $X\ Bin(n = 100, p = 0.05)$.
- **b)** Falsch. Die Koeffizienten in der einfachen und multiplen Regression sind im Allgemeinen unterschiedlich.
- **c)** Richtig. Der Wert $p = 0.1$ liegt nicht im Vertrauensintervall. Per Definition wird die Nullhypothese für diesen Parameter also verworfen.
- **d)** Falsch.