Musterlösung

Es gibt verschiedene Version der Prüfung. Die Aufgaben sind jeweils in einer anderen Reihenfolge.

Gruppe A

1. a) Richtig. $P(X_i = 0) = 1 - \pi = 1 - 0.6 = 0.4$.
   b) Richtig. Der Erwartungswert einer Bernoulli Zufallsvariable ist π.
   c) Falsch. $\sum_{i=1}^{5} \mathrm{Var}(X_i) = 5 \cdot 0.6 \cdot 0.4$ da unabhängig.
   d) Falsch. $P(X_1 = 0 \cap X_2 = 0) = P(X_1 = 0)P(X_2 = 0) = (1 - 0.6) \cdot (1 - 0.6) = 0.16$.

2. a) Richtig. Per Definition der Binomialverteilung.
   b) Richtig. Da $P(Y \leq 4) = 1 - P(Y > 4) = 1 - P(Y = 5) = 0.92$.
   c) Falsch. Der Erwartungswert einer Binomialverteilung ist $n\pi = 5 \cdot 0.2 = 1$.
   d) Falsch. Der Erwartungswert hier ist $E\left[\sum_{i=1}^{5}(1 - X_i)\right] = \sum_{i=1}^{5} E[1 - X_i] = 5E\left[1 - X_1\right]$, wobei wir verwenden, dass $E[X_1] = ... = E[X_5]$.

3. a) Falsch. Die Macht ist definiert als $P_{H_A}(X \in K)$. Der Verwerfungsbereich für diesen Test ist $K = \{6, 7, 8\}$. Es gilt $P_{\pi=0.6}(X \geq 6) > P_{\pi=0.4}(X \geq 6)$.
   b) Richtig. Da $P(X \geq 5) < P(X \geq 2)$.
   c) Richtig. Ein grösserer Fehler 1.Art vergrössert den Verwerfungsbereich K, somit wird auch $P_{H_A}(X \in K)$ grösser.
   d) Falsch. Es gilt $P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) = \binom{8}{5}0.3^5 0.7^3 + \binom{8}{6}0.3^6 0.7^2 + \binom{8}{7}0.3^7 0.7^1 + \binom{8}{5}0.3^8 0.7^0 = 0.058 > 0.05$, deshalb kann 5 nicht im Verwerfungsbereich liegen. Somit ist die Aussage falsch.
   Achtung: $P(X \geq 5)$ nicht mit $P(X = 5) = 0.047 < 0.05$ verwechseln!

4. a) Falsch. Aus den Faustregeln $n(1 - \pi) > 5$ und $n\pi > 5$ folgt hier $0.3 \cdot 8 = 2.4 < 5$, somit kann man keine Normalapproximation durchführen.
   b) Richtig.
   c) Richtig.
   d) Falsch. Die Verteilung ist hypergeometrisch. Man kann sich das Ganze als ein Urnenmodell mit 10 weissen Kugeln und 5 schwarzen Kugeln vorstellen. Aus der Urne wird 4 Mal ohne Zurücklegen gezogen. Gesucht ist die Verteilung für die Anzahl weisser Kugeln unter den 4 gezogenen Kugeln.

5. a) Richtig. $P(X = 2) = \binom{30}{2}0.2^2 0.8^{28} = 0.03$
   b) Richtig. Aus $P_{H_0}(X \notin K) = 1 - P_{H_0}(X \in K)$ und $P_{H_0}(X \in K) \leq 0.05$ folgt die Aussage.
   c) Richtig. Die Macht ist definiert als $P_{H_A}(X \in K)$.
   d) Falsch. Der P-Wert ist definiert als die Wahrscheinlichkeit ein noch extremeres Ergebnis als $x = 2$ zu erhalten. Da wir zweiseitig testen, sind sowohl kleine wie auch grosse Anzahlen extrem. In der Aussage werden aber nur kleinere Anzahlen berücksichtigt.

6. a) Falsch. Es kann nicht jedem männlichen Proband auf eindeutige Weise eine weibliche Probandin zugeordnet werden.
   b) Richtig. Das ist zum Beispiel bei dieser Aufgabe der Fall.
   c) Richtig.
   d) Falsch. Der Welch-Test setzt nicht voraus, dass die Standardabweichungen in beiden Gruppen gleich sind.

7. a) Falsch. Der beidseitige Test hat die kleinere Macht als ein einseitiger Test.
   b) Richtig. Da $S_{pool}^2 = \frac{1}{38}(19 \cdot 3.1^2 + 19 \cdot 4.5^2) = 14.93$, ist $T = (20.4 - 18.9)/(S_{pool} \cdot \sqrt{0.1}) = 1.22$.
   c) Falsch. Die Verteilung von T unter H_0 ist $T \sim t_{n+m-2} = t_{38}$.
   d) Falsch. Die Teststatistik liegt nicht im Verwerfungsbereich. Somit kann die Nullhypothese auf dem 5%-Niveau nicht verworfen werden.

Gruppe A

8. **a)** Richtig.
b) Falsch. Richtig wäre $\mathrm{Bin}(40, 0.5)$.
c) Richtig. Die Grenzen des Intervalls sind gegeben durch $18.9 \pm t_{19,0.975} \cdot 4.5/\sqrt{20}$.
d) Richtig.

9. **a)** Falsch. Wir können die Standardabweichung aus den Daten schätzen. Nur beim z-Test ist es notwendig die wahre Standardabweichung zu kennen.
b) Richtig.
c) Falsch. Beim Wilcoxon-Test müssen die Daten symmetrisch verteilt sein.
d) Richtig. Sei $\rho = 1/2$ die Hälfte der Breite des Vertrauensintervalls. Wir möchten n so, dass $\rho = \frac{1}{2} \geq 2 \cdot \frac{\sigma}{\sqrt{n}}$. Somit $n \geq 4^2\sigma^2 = 16$.

10. **a)** Falsch. Es gibt zwei Parameter im linearen Modell. Somit ergibt sich die Anzahl Datenpunkte aus der Anzahl Freiheitsgrade 46 plus zwei.
b) Richtig. β_1 ist signifikant auf dem 0.1%-Niveau, da der P-Wert für β_1 kleiner als 0.01 ist.
c) Richtig. Das approximative, zweiseitige 95%-Vertrauensintervall für β_1 ist $[\hat{\beta}_1 - 2\hat{s.e.}(\hat{\beta}_1), \hat{\beta}_1 + 2\hat{s.e.}(\hat{\beta}_1)]$.
d) Richtig. Der t-Wert ergibt sich aus $\hat{\beta}_0/\hat{s.e.}(\hat{\beta}_0)$.

11. **a)** Richtig. Die Zunahme des erwarteten Ertrages ist $5 \cdot 0.8388 \approx 4.19\,kg$.
b) Falsch. Der Achsenabschnitt der geschätzten Regressionsgeraden ist $1.0914 \approx 1.09$.
c) Falsch. Das geschätzte Modell ist ertrag $= 1.0914 + 0.8388 \cdot$ pestizid. Somit ergibt sich für einem erwarteten Ertrag von $40\,kg$ der Einsatz von $(40 - 1.0914)/0.8388 \approx 46.39\,l$ Pestizid.
d) Richtig.

12. **a)** Richtig.
b) Richtig.
c) Falsch. Der QQ Plot weist keine Kurzschwänzigkeit auf. Die Normalitätsannahme ist nicht stark verletzt.
d) Richtig.

13. **a)** Richtig, da $E[2Y - 2X - 1] = 2E[Y] - 2E[X] - 1 = 10 - 4 - 1 = 5$.
b) Falsch, da $Var[2Y - 2X - 1] = 4Var[Y] + 4Var[X] = 16$.
c) Falsch. Es gilt $\frac{1}{2}(x_{(1)} + x_{(2)}) = 1$
d) Richtig. Der Erwartungswert ist $\frac{1}{5}(1 + 1 + 3 + 5 + 6) = 3.2$ und die Standardabweichung

$$\sqrt{\frac{1}{4}(2 * (1 - 3.2)^2 + (3 - 3.2)^2 + (5 - 3.2)^2 + (6 - 3.2)^2)} = 2.28$$

14. **a)** Richtig. Da mit dem Additionssatz $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ gilt, folgt $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.1 + 0.2 - 0.28 = 0.02 = P(A) \cdot P(B)$. Deshalb sind die Ereignisse unabhängig.
b) Falsch. Es ist sinnvoll einen Regenschirm mitzunehmen, da $odds(Regen) = \frac{P(Regen)}{P(keinRegen)} = 90$. Deshalb ist die Wahrscheinlichkeit, dass es regnet, 90 mal höher.
c) Richtig. Da $P(E|G) = \frac{P(E \cap G)}{P(G)} = \frac{1/4}{1/3} = 3/4$. Deshalb $odds(E|G) = \frac{P(E|G)}{1 - P(E|G)} = \frac{3/4}{1 - 3/4} = 3$.
d) Falsch. $odds(B) = \frac{2P(B)}{2(1 - P(B))} = \frac{P(A)}{2 - P(A)} \neq 2odds(A)$.

15. **a)** Richtig. $P(S \cap B) = P(B|S)P(S) = 0.9 \cdot 0.7 = 0.63$.

b) Richtig. Die Wahrscheinlichkeit $P(B)$ lässt sich mit dem Satz der totalen Wahrscheinlichkeit berechnen:

$$P(B) = P(B|S)P(S) + P(B|S^c)P(S^C)$$

wobei $P(B|S) = 0.9, P(S) = 0.7, P(S^C) = 1 - P(S) = 0.3$.

c) Falsch. $0.7 = P(B) \neq P(B|S) = 0.9$.

d) Richtig.

16. **a)** Falsch. Wenn man die Anzahl Bins in B grösser wählt, könnte man dasselbe Histogramm erhalten.

b) Falsch. Das Histogramm A zeigt eine rechtsschiefe Verteilung.

c) Falsch. Der Wertebereich der Daten geht von mindestens 0 bis 5 und ist somit grösser wie der Wertebreich der Daten bei A. Ausserdem ist der Median an einem anderen Ort und der Boxplot zeigt eine symmetrische Verteilung der Daten um den Median.

d) Richtig.

17. **a)** Richtig.

b) Falsch.

c) Falsch. Das arithmetische Mittel hat approximativ eine Normalverteilung nicht F.

d) Richtig. Definiere den Saft von Orange i in Liter als O_i. Der Ewartungswert ist $E[O_i] = 0.3l$ und die Varianz $Var[O_i] = 0.09^2$. Der Gesamtsaft kann geschrieben werden als

$$S := \sum_{i=1}^{38} O_i.$$

Mit dem Zentralen Grenzwertsatz ist $S \approx \mathcal{N}(38 * 0.3, 38 * 0.09^2) = \mathcal{N}(11.4, 0.3078)$ verteilt. Deshalb können wir die Wahrscheinlichkeit wie folgt berechnen:

$$\begin{aligned} P(S < 10l) &= P\left(\frac{S - 11.4}{0.555} < \frac{10 - 11.4}{0.555}\right) \\ &= P(Z < -2.52) = P(Z > 2.52) = 1 - P(Z \leq 2.52) \\ &= 1 - 0.9941 = 0.0059 < 0.01. \end{aligned}$$

18. **a)** Falsch. Integrieren ergibt 1 und kein Wert ist negativ.

b) Falsch. Es gilt $P(X = x) = 0$ für stetige Zufallsvariablen.

c) Richtig. $P(T > 2) = 1 - P(T \leq 2) = 1 - (1 - e^{-\lambda * 2}) = e^{-3*2} = e^{-6}$.

d) Falsch, $X + Y$ ist Poisson verteilt mit $\lambda = 1 + 1 = 2$.

19. **a)** Falsch. Im Allgemeinen ergibt sich aus neuen Daten ein anderer Wert für den Schätzer.

b) Richtig, der Momentenschätzer ist $\hat{\lambda}_{MoM} = x = 3$. Der Maximum Likelihood Schätzer $\hat{\lambda}_{MLE}$ berechnet sich wie folgt:

$$\log(e^{-\lambda}\lambda^x/x!) = -\lambda + x\log(\lambda) - \log(x!).$$

Ableiten ergibt

$$\frac{x}{\hat{\lambda}_{MLE}} = 1.$$

Somit ist $\hat{\lambda}_{MLE} = x = 3$, also identisch zum Momentenschätzer.

c) Richtig. $\hat{\pi}_{MoM} = \frac{x}{n} = 0.7$.

d) Falsch. Das $\sqrt{n}$- Gesetz sagt, dass man viermal soviele Daten braucht.

20. **a)** Falsch, über Kausalität kann keine Aussage gemacht werden.

b) Richtig.

c) Richtig. $P_{S_1}(w) = 0.5\frac{6}{10} + 0.5 \cdot 1 = 0.8$.

d) Richtig. $P_{S_2}(w) = \frac{16}{20} = 0.8$.

Gruppe B

1. **a)** Falsch. Es kann nicht jedem männlichen Proband auf eindeutige Weise eine weibliche Probandin zugeordnet werden.

b) Richtig. Das ist zum Beispiel bei dieser Aufgabe der Fall.

c) Richtig.

d) Falsch. Der Welch-Test setzt nicht voraus, dass die Standardabweichungen in beiden Gruppen gleich sind.

2. **a)** Falsch. Der beidseitige Test hat die kleinere Macht als ein einseitiger Test.

b) Richtig. Da $S^2_{pool} = \frac{1}{38}(19 \cdot 3.1^2 + 19 \cdot 4.5^2) = 14.93$, ist $T = (20.4 - 18.9)/(S_{pool} \cdot \sqrt{0.1}) = 1.22$.

c) Falsch. Die Verteilung von T unter H_0 ist $T \sim t_{n+m-2} = t_{38}$.

d) Falsch. Die Teststatistik liegt nicht im Verwerfungsbereich. Somit kann die Nullhypothese auf dem 5%-Niveau nicht verworfen werden.

3. **a)** Richtig.

b) Falsch. Richtig wäre $\text{Bin}(40, 0.5)$.

c) Richtig. Die Grenzen des Intervalls sind gegeben durch $18.9 \pm t_{19,0.975} \cdot 4.5/\sqrt{20}$.

d) Richtig.

4. **a)** Falsch. Wir können die Standardabweichung aus den Daten schätzen. Nur beim z-Test ist es notwendig die wahre Standardabweichung zu kennen.

b) Richtig.

c) Falsch. Beim Wilcoxon-Test müssen die Daten symmetrisch verteilt sein.

d) Richtig. Sei $\rho = 1/2$ die Hälfte der Breite des Vertrauensintervalls. Wir möchten n so, dass $\rho = \frac{1}{2} \geq 2 \cdot \frac{\sigma}{\sqrt{n}}$. Somit $n \geq 4^2\sigma^2 = 16$.

5. **a)** Falsch. Es gibt zwei Parameter im linearen Modell. Somit ergibt sich die Anzahl Datenpunkte aus der Anzahl Freiheitsgrade 46 plus zwei.

b) Richtig. β_1 ist signifikant auf dem 0.1%-Niveau, da der P-Wert für β_1 kleiner als 0.01 ist.

c) Richtig. Das approximative, zweiseitige 95%-Vertrauensintervall für β_1 ist $[\hat{\beta}_1 - 2\widehat{s.e.}(\hat{\beta}_1), \hat{\beta}_1 + 2\widehat{s.e.}(\hat{\beta}_1)]$.

d) Richtig. Der t-Wert ergibt sich aus $\hat{\beta}_0/\widehat{s.e.}(\hat{\beta}_0)$.

6. **a)** Richtig. Die Zunahme des erwarteten Ertrages ist $5 \cdot 0.8388 \approx 4.19\,kg$.

b) Falsch. Der Achsenabschnitt der geschätzten Regressionsgeraden ist $1.0914 \approx 1.09$.

c) Falsch. Das geschätzte Modell ist ertrag $= 1.0914 + 0.8388 \cdot$ pestizid. Somit ergibt sich für einem erwarteten Ertrag von $40\,kg$ der Einsatz von $(40 - 1.0914)/0.8388 \approx 46.39\,l$ Pestizid.

d) Richtig.

7. **a)** Richtig.

b) Richtig.

c) Falsch. Der QQ Plot weist keine Kurzschwänzigkeit auf. Die Normalitätsannahme ist nicht stark verletzt.

d) Richtig.

8. **a)** Richtig, da $E[2Y - 2X - 1] = 2E[Y] - 2E[X] - 1 = 10 - 4 - 1 = 5$.

b) Falsch, da $Var[2Y - 2X - 1] = 4Var[Y] + 4Var[X] = 16$.

c) Falsch. Es gilt $\frac{1}{2}(x_{(1)} + x_{(2)}) = 1$

d) Richtig. Der Erwartungswert ist $\frac{1}{5}(1+1+3+5+6) = 3.2$ und die Standardabweichung

$$\sqrt{\frac{1}{4}(2*(1-3.2)^2+(3-3.2)^2+(5-3.2)^2+(6-3.2)^2)} = 2.28$$

9. **a)** Richtig. Da mit dem Additionssatz $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ gilt, folgt $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.1 + 0.2 - 0.28 = 0.02 = P(A) \cdot P(B)$. Deshalb sind die Ereignisse unabhängig.

b) Falsch. Es ist sinnvoll einen Regenschirm mitzunehmen, da $odds(Regen) = \frac{P(Regen)}{P(keinRegen)} = 90$. Deshalb ist die Wahrscheinlichkeit, dass es regnet, 90 mal höher.

c) Richtig. Da $P(E|G) = \frac{P(E \cap G)}{P(G)} = \frac{1/4}{1/3} = 3/4$. Deshalb $odds(E|G) = \frac{P(E|G)}{1-P(E|G)} = \frac{3/4}{1-3/4} = 3$.

d) Falsch. $odds(B) = \frac{2P(B)}{2(1-P(B))} = \frac{P(A)}{2-P(A)} \neq 2odds(A)$.

10. **a)** Richtig. $P(S \cap B) = P(B|S)P(S) = 0.9 \cdot 0.7 = 0.63$.

b) Richtig. Die Wahrscheinlichkeit $P(B)$ lässt sich mit dem Satz der totalen Wahrscheinlichkeit berechnen:

$$P(B) = P(B|S)P(S) + P(B|S^c)P(S^C)$$

wobei $P(B|S) = 0.9, P(S) = 0.7, P(S^C) = 1 - P(S) = 0.3$.

c) Falsch. $0.7 = P(B) \neq P(B|S) = 0.9$.

d) Richtig.

11. **a)** Falsch. Wenn man die Anzahl Bins in B grösser wählt, könnte man dasselbe Histogramm erhalten.

b) Falsch. Das Histogramm A zeigt eine rechtsschiefe Verteilung.

c) Falsch. Der Wertebereich der Daten geht von mindestens 0 bis 5 und ist somit grösser wie der Wertebreich der Daten bei A. Ausserdem ist der Median an einem anderen Ort und der Boxplot zeigt eine symmetrische Verteilung der Daten um den Median.

d) Richtig.

12. **a)** Richtig.

b) Falsch.

c) Falsch. Das arithmetische Mittel hat approximativ eine Normalverteilung nicht F.

d) Richtig. Definiere den Saft von Orange i in Liter als O_i. Der Erwartungswert ist $E[O_i] = 0.3l$ und die Varianz $Var[O_i] = 0.09^2$. Der Gesamtsaft kann geschrieben werden als

$$S := \sum_{i=1}^{38} O_i.$$

Mit dem Zentralen Grenzwertsatz ist $S \approx \mathcal{N}(38 * 0.3, 38 * 0.09^2) = \mathcal{N}(11.4, 0.3078)$ verteilt. Deshalb können wir die Wahrscheinlichkeit wie folgt berechnen:

$$\begin{aligned}P(S < 10l) &= P\left(\frac{S-11.4}{0.555} < \frac{10-11.4}{0.555}\right)\\ &= P(Z < -2.52) = P(Z > 2.52) = 1 - P(Z \leq 2.52)\\ &= 1 - 0.9941 = 0.0059 < 0.01.\end{aligned}$$

13. **a)** Falsch. Integrieren ergibt 1 und kein Wert ist negativ.

b) Falsch. Es gilt $P(X = x) = 0$ für stetige Zufallsvariablen.

c) Richtig. $P(T > 2) = 1 - P(T \leq 2) = 1 - (1 - e^{-\lambda * 2}) = e^{-3*2} = e^{-6}$.

d) Falsch, $X + Y$ ist Poisson verteilt mit $\lambda = 1 + 1 = 2$.

Gruppe B

14. **a)** Falsch. Im Allgemeinen ergibt sich aus neuen Daten ein anderer Wert für den Schätzer.

b) Richtig, der Momentenschätzer ist $\hat{\lambda}_{MoM} = x = 3$. Der Maximum Likelihood Schätzer $\hat{\lambda}_{MLE}$ berechnet sich wie folgt:
$$\log(e^{-\lambda}\lambda^x/x!) = -\lambda + x\log(\lambda) - \log(x!).$$
Ableiten ergibt
$$\frac{x}{\hat{\lambda}_{MLE}} = 1.$$
Somit ist $\hat{\lambda}_{MLE} = x = 3$, also identisch zum Momentenschätzer.

c) Richtig. $\hat{\pi}_{MoM} = \frac{x}{n} = 0.7$.

d) Falsch. Das $\sqrt{n}$- Gesetz sagt, dass man viermal soviele Daten braucht.

15. **a)** Falsch, über Kausalität kann keine Aussage gemacht werden.

b) Richtig.

c) Richtig. $P_{S_1}(w) = 0.5\frac{6}{10} + 0.5 \cdot 1 = 0.8$.

d) Richtig. $P_{S_2}(w) = \frac{16}{20} = 0.8$.

16. **a)** Richtig. $P(X_i = 0) = 1 - \pi = 1 - 0.6 = 0.4$.

b) Richtig. Der Erwartungswert einer Bernoulli Zufallsvariable ist π.

c) Falsch. $\sum_{i=1}^{5} \mathrm{Var}(X_i) = 5 \cdot 0.6 \cdot 0.4$ da unabhängig.

d) Falsch. $P(X_1 = 0 \cap X_2 = 0) = P(X_1 = 0)P(X_2 = 0) = (1 - 0.6) \cdot (1 - 0.6) = 0.16$.

17. **a)** Richtig. Per Definition der Binomialverteilung.

b) Richtig. Da $P(Y \leq 4) = 1 - P(Y > 4) = 1 - P(Y = 5) = 0.92$.

c) Falsch. Der Erwartungswert einer Binomialverteilung ist $n\pi = 5 \cdot 0.2 = 1$.

d) Falsch. Der Erwartungswert hier ist $E\left[\sum_{i=1}^{5}(1 - X_i)\right] = \sum_{i=1}^{5} E[1 - X_i] = 5E\left[1 - X_1\right]$, wobei wir verwenden, dass $E[X_1] = ... = E[X_5]$.

18. **a)** Falsch. Die Macht ist definiert als $P_{H_A}(X \in K)$. Der Verwerfungsbereich für diesen Test ist $K = \{6, 7, 8\}$. Es gilt $P_{\pi=0.6}(X \geq 6) > P_{\pi=0.4}(X \geq 6)$.

b) Richtig. Da $P(X \geq 5) < P(X \geq 2)$.

c) Richtig. Ein grösserer Fehler 1.Art vergrössert den Verwerfungsbereich K, somit wird auch $P_{H_A}(X \in K)$ grösser.

d) Falsch. Es gilt $P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) = \binom{8}{5}0.3^5 0.7^3 + \binom{8}{6}0.3^6 0.7^2 + \binom{8}{7}0.3^7 0.7^1 + \binom{8}{5}0.3^8 0.7^0 = 0.058 > 0.05$, deshalb kann 5 nicht im Verwerfungsbereich liegen. Somit ist die Aussage falsch.
Achtung: $P(X \geq 5)$ nicht mit $P(X = 5) = 0.047 < 0.05$ verwechseln!

19. **a)** Falsch. Aus den Faustregeln $n(1 - \pi) > 5$ und $n\pi > 5$ folgt hier $0.3 \cdot 8 = 2.4 < 5$, somit kann man keine Normalapproximation durchführen.

b) Richtig.

c) Richtig.

d) Falsch. Die Verteilung ist hypergeometrisch. Man kann sich das Ganze als ein Urnenmodell mit 10 weissen Kugeln und 5 schwarzen Kugeln vorstellen. Aus der Urne wird 4 Mal ohne Zurücklegen gezogen. Gesucht ist die Verteilung für die Anzahl weisser Kugeln unter den 4 gezogenen Kugeln.

20. **a)** Richtig. $P(X = 2) = \binom{30}{2}0.2^2 0.8^{28} = 0.03$

b) Richtig. Aus $P_{H_0}(X \notin K) = 1 - P_{H_0}(X \in K)$ und $P_{H_0}(X \in K) \leq 0.05$ folgt die Aussage.

c) Richtig. Die Macht ist definiert als $P_{H_A}(X \in K)$.

d) Falsch. Der P-Wert ist definiert als die Wahrscheinlichkeit ein noch extremeres Ergebnis als $x = 2$ zu erhalten. Da wir zweiseitig testen, sind sowohl kleine wie auch grosse Anzahlen extrem. In der Aussage werden aber nur kleinere Anzahlen berücksichtigt.

Gruppe C

1. **a)** Richtig, da $E[2Y - 2X - 1] = 2E[Y] - 2E[X] - 1 = 10 - 4 - 1 = 5.$

b) Falsch, da $Var[2Y - 2X - 1] = 4Var[Y] + 4Var[X] = 16.$

c) Falsch. Es gilt $\frac{1}{2}(x_{(1)} + x_{(2)}) = 1$

d) Richtig. Der Erwartungswert ist $\frac{1}{5}(1 + 1 + 3 + 5 + 6) = 3.2$ und die Standardabweichung

$$\sqrt{\frac{1}{4}(2 * (1 - 3.2)^2 + (3 - 3.2)^2 + (5 - 3.2)^2 + (6 - 3.2)^2)} = 2.28$$

2. **a)** Richtig. Da mit dem Additionssatz $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ gilt, folgt $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.1 + 0.2 - 0.28 = 0.02 = P(A) \cdot P(B)$. Deshalb sind die Ereignisse unabhängig.

b) Falsch. Es ist sinnvoll einen Regenschirm mitzunehmen, da $odds(Regen) = \frac{P(Regen)}{P(keinRegen)} = 90.$ Deshalb ist die Wahrscheinlichkeit, dass es regnet, 90 mal höher.

c) Richtig. Da $P(E|G) = \frac{P(E \cap G)}{P(G)} = \frac{1/4}{1/3} = 3/4$. Deshalb $odds(E|G) = \frac{P(E|G)}{1-P(E|G)} = \frac{3/4}{1-3/4} = 3.$

d) Falsch. $odds(B) = \frac{2P(B)}{2(1-P(B))} = \frac{P(A)}{2-P(A)} \neq 2odds(A).$

3. **a)** Richtig. $P(S \cap B) = P(B|S)P(S) = 0.9 \cdot 0.7 = 0.63.$

b) Richtig. Die Wahrscheinlichkeit $P(B)$ lässt sich mit dem Satz der totalen Wahrscheinlichkeit berechnen:

$$P(B) = P(B|S)P(S) + P(B|S^c)P(S^C)$$

wobei $P(B|S) = 0.9, P(S) = 0.7, P(S^C) = 1 - P(S) = 0.3.$

c) Falsch. $0.7 = P(B) \neq P(B|S) = 0.9.$

d) Richtig.

4. **a)** Falsch. Wenn man die Anzahl Bins in B grösser wählt, könnte man dasselbe Histogramm erhalten.

b) Falsch. Das Histogramm A zeigt eine rechtsschiefe Verteilung.

c) Falsch. Der Wertebereich der Daten geht von mindestens 0 bis 5 und ist somit grösser wie der Wertebreich der Daten bei A. Ausserdem ist der Median an einem anderen Ort und der Boxplot zeigt eine symmetrische Verteilung der Daten um den Median.

d) Richtig.

5. **a)** Richtig.

b) Falsch.

c) Falsch. Das arithmetische Mittel hat approximativ eine Normalverteilung nicht F.

d) Richtig. Definiere den Saft von Orange i in Liter als O_i. Der Erwartungswert ist $E[O_i] = 0.3l$ und die Varianz $Var[O_i] = 0.09^2$. Der Gesamtsaft kann geschrieben werden als

$$S := \sum_{i=1}^{38} O_i.$$

Mit dem Zentralen Grenzwertsatz ist $S \approx \mathcal{N}(38 * 0.3, 38 * 0.09^2) = \mathcal{N}(11.4, 0.3078)$ verteilt. Deshalb können wir die Wahrscheinlichkeit wie folgt berechnen:

$$\begin{aligned} P(S < 10l) &= P\left(\frac{S - 11.4}{0.555} < \frac{10 - 11.4}{0.555}\right) \\ &= P(Z < -2.52) = P(Z > 2.52) = 1 - P(Z \leq 2.52) \\ &= 1 - 0.9941 = 0.0059 < 0.01. \end{aligned}$$

Gruppe C

6. **a)** Falsch. Integrieren ergibt 1 und kein Wert ist negativ.

b) Falsch. Es gilt $P(X = x) = 0$ für stetige Zufallsvariablen.

c) Richtig. $P(T > 2) = 1 - P(T \leq 2) = 1 - (1 - e^{-\lambda * 2}) = e^{-3*2} = e^{-6}$.

d) Falsch, $X + Y$ ist Poisson verteilt mit $\lambda = 1 + 1 = 2$.

7. **a)** Falsch. Im Allgemeinen ergibt sich aus neuen Daten ein anderer Wert für den Schätzer.

b) Richtig, der Momentenschätzer ist $\hat{\lambda}_{MoM} = x = 3$. Der Maximum Likelihood Schätzer $\hat{\lambda}_{MLE}$ berechnet sich wie folgt:

$$\log(e^{-\lambda}\lambda^x/x!) = -\lambda + x\log(\lambda) - \log(x!).$$

Ableiten ergibt

$$\frac{x}{\hat{\lambda}_{MLE}} = 1.$$

Somit ist $\hat{\lambda}_{MLE} = x = 3$, also identisch zum Momentenschätzer.

c) Richtig. $\hat{\pi}_{MoM} = \frac{x}{n} = 0.7$.

d) Falsch. Das $\sqrt{n}$- Gesetz sagt, dass man viermal soviele Daten braucht.

8. **a)** Falsch, über Kausalität kann keine Aussage gemacht werden.

b) Richtig.

c) Richtig. $P_{S_1}(w) = 0.5\frac{6}{10} + 0.5 \cdot 1 = 0.8$.

d) Richtig. $P_{S_2}(w) = \frac{16}{20} = 0.8$.

9. **a)** Richtig. $P(X_i = 0) = 1 - \pi = 1 - 0.6 = 0.4$.

b) Richtig. Der Erwartungswert einer Bernoulli Zufallsvariable ist π.

c) Falsch. $\sum_{i=1}^{5} \mathrm{Var}(X_i) = 5 \cdot 0.6 \cdot 0.4$ da unabhängig.

d) Falsch. $P(X_1 = 0 \cap X_2 = 0) = P(X_1 = 0)P(X_2 = 0) = (1 - 0.6) \cdot (1 - 0.6) = 0.16$.

10. **a)** Richtig. Per Definition der Binomialverteilung.

b) Richtig. Da $P(Y \leq 4) = 1 - P(Y > 4) = 1 - P(Y = 5) = 0.92$.

c) Falsch. Der Erwartungswert einer Binomialverteilung ist $n\pi = 5 \cdot 0.2 = 1$.

d) Falsch. Der Erwartungswert hier ist $E\left[\sum_{i=1}^{5}(1 - X_i)\right] = \sum_{i=1}^{5} E[1 - X_i] = 5E[1 - X_1]$, wobei wir verwenden, dass $E[X_1] = ... = E[X_5]$.

11. **a)** Falsch. Die Macht ist definiert als $P_{H_A}(X \in K)$. Der Verwerfungsbereich für diesen Test ist $K = \{6, 7, 8\}$. Es gilt $P_{\pi=0.6}(X \geq 6) > P_{\pi=0.4}(X \geq 6)$.

b) Richtig. Da $P(X \geq 5) < P(X \geq 2)$.

c) Richtig. Ein grösserer Fehler 1.Art vergrössert den Verwerfungsbereich K, somit wird auch $P_{H_A}(X \in K)$ grösser.

d) Falsch. Es gilt $P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) = \binom{8}{5}0.3^5 0.7^3 + \binom{8}{6}0.3^6 0.7^2 + \binom{8}{7}0.3^7 0.7^1 + \binom{8}{5}0.3^8 0.7^0 = 0.058 > 0.05$, deshalb kann 5 nicht im Verwerfungsbereich liegen. Somit ist die Aussage falsch.
Achtung: $P(X \geq 5)$ nicht mit $P(X = 5) = 0.047 < 0.05$ verwechseln!

12. **a)** Falsch. Aus den Faustregeln $n(1 - \pi) > 5$ und $n\pi > 5$ folgt hier $0.3 \cdot 8 = 2.4 < 5$, somit kann man keine Normalapproximation durchführen.

b) Richtig.

c) Richtig.

d) Falsch. Die Verteilung ist hypergeometrisch. Man kann sich das Ganze als ein Urnenmodell mit 10 weissen Kugeln und 5 schwarzen Kugeln vorstellen. Aus der Urne wird 4 Mal ohne Zurücklegen gezogen. Gesucht ist die Verteilung für die Anzahl weisser Kugeln unter den 4 gezogenen Kugeln.

13. **a)** Richtig. $P(X=2)=\binom{30}{2}0.2^2 0.8^{28}=0.03$

b) Richtig. Aus $P_{H_0}(X \notin K)=1-P_{H_0}(X \in K)$ und $P_{H_0}(X \in K) \leq 0.05$ folgt die Aussage.

c) Richtig. Die Macht ist definiert als $P_{H_A}(X \in K)$.

d) Falsch. Der P-Wert ist definiert als die Wahrscheinlichkeit ein noch extremeres Ergebnis als $x=2$ zu erhalten. Da wir zweiseitig testen, sind sowohl kleine wie auch grosse Anzahlen extrem. In der Aussage werden aber nur kleinere Anzahlen berücksichtigt.

14. **a)** Falsch. Es kann nicht jedem männlichen Proband auf eindeutige Weise eine weibliche Probandin zugeordnet werden.

b) Richtig. Das ist zum Beispiel bei dieser Aufgabe der Fall.

c) Richtig.

d) Falsch. Der Welch-Test setzt nicht voraus, dass die Standardabweichungen in beiden Gruppen gleich sind.

15. **a)** Falsch. Der beidseitige Test hat die kleinere Macht als ein einseitiger Test.

b) Richtig. Da $S_{pool}^2=\frac{1}{38}(19 \cdot 3.1^2+19 \cdot 4.5^2)=14.93$, ist $T=(20.4-18.9)/(S_{pool} \cdot \sqrt{0.1})=1.22$.

c) Falsch. Die Verteilung von T unter H_0 ist $T \sim t_{n+m-2}=t_{38}$.

d) Falsch. Die Teststatistik liegt nicht im Verwerfungsbereich. Somit kann die Nullhypothese auf dem 5%-Niveau nicht verworfen werden.

16. **a)** Richtig.

b) Falsch. Richtig wäre $\mathrm{Bin}(40, 0.5)$.

c) Richtig. Die Grenzen des Intervalls sind gegeben durch $18.9 \pm t_{19,0.975} \cdot 4.5/\sqrt{20}$.

d) Richtig.

17. **a)** Falsch. Wir können die Standardabweichung aus den Daten schätzen. Nur beim z-Test ist es notwendig die wahre Standardabweichung zu kennen.

b) Richtig.

c) Falsch. Beim Wilcoxon-Test müssen die Daten symmetrisch verteilt sein.

d) Richtig. Sei $\rho=1/2$ die Hälfte der Breite des Vertrauensintervalls. Wir möchten n so, dass $\rho=\frac{1}{2} \geq 2 \cdot \frac{\sigma}{\sqrt{n}}$. Somit $n \geq 4^2\sigma^2=16$.

18. **a)** Falsch. Es gibt zwei Parameter im linearen Modell. Somit ergibt sich die Anzahl Datenpunkte aus der Anzahl Freiheitsgrade 46 plus zwei.

b) Richtig. β_1 ist signifikant auf dem 0.1%-Niveau, da der P-Wert für β_1 kleiner als 0.01 ist.

c) Richtig. Das approximative, zweiseitige 95%-Vertrauensintervall für β_1 ist $[\hat{\beta}_1 - 2\hat{s.e.}(\hat{\beta}_1), \hat{\beta}_1 + 2\hat{s.e.}(\hat{\beta}_1)]$.

d) Richtig. Der t-Wert ergibt sich aus $\hat{\beta}_0/\hat{s.e.}(\hat{\beta}_0)$.

19. **a)** Richtig. Die Zunahme des erwarteten Ertrages ist $5 \cdot 0.8388 \approx 4.19\,kg$.

b) Falsch. Der Achsenabschnitt der geschätzten Regressionsgeraden ist $1.0914 \approx 1.09$.

c) Falsch. Das geschätzte Modell ist ertrag $=1.0914+0.8388 \cdot$ pestizid. Somit ergibt sich für einem erwarteten Ertrag von $40\,kg$ der Einsatz von $(40-1.0914)/0.8388 \approx 46.39\,l$ Pestizid.

d) Richtig.

20. **a)** Richtig.

b) Richtig.

c) Falsch. Der QQ Plot weist keine Kurzschwänzigkeit auf. Die Normalitätsannahme ist nicht stark verletzt.

d) Richtig.