Technische Universität Berlin
SoSe 2025

Prof. Dr. Carlos Améndola
Janike Oldekop

**Stochastik für Informatik (9 LP)**

# Musterlösung der Klausur vom 07.10.2025

**Aufgabe 1** $2 \times 10 = 20$ *Punkte*

Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind.

| | | Wahr | Falsch |
|---|---|---|---|
| a) | In einem Laplace-Raum ist jedes Ereignis gleich wahrscheinlich. | ○ | ⊗ |
| b) | Für alle Ereignisse $A, B$ aus einem Wahrscheinlichkeitsraum gilt $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$. | ⊗ | ○ |
| c) | Der Erwartungswert einer diskreten Zufallsvariable kann niemals negativ sein. | ○ | ⊗ |
| d) | Die Varianz einer stetigen Zufallsvariable kann niemals null sein. | ⊗ | ○ |
| e) | Bei einer stetigen Zufallsvariable ist die Wahrscheinlichkeit, dass sie genau einen bestimmen Wert annimmt, gleich null. | ⊗ | ○ |
| f) | Für $X \sim \mathrm{Ber}(p)$ gilt $\mathbb{E}[X] = \mathbb{E}[X^2]$. | ⊗ | ○ |
| g) | Der Maximum-Likelihood-Schätzer ist immer erwartungstreu. | ○ | ⊗ |
| h) | Die Varianz eines Schätzers sollte möglichst groß sein, um zuverlässige Schätzungen zu erhalten. | ○ | ⊗ |
| i) | Ein Fehler 1. Art bedeutet, dass man die Nullhypothese beibehält, obwohl sie falsch ist. | ○ | ⊗ |
| j) | Ist der $p$-Wert größer als das Signifikanzniveau, so wird die Nullhypothese verworfen. | ○ | ⊗ |

**Aufgabe 2** $3 + 3 + 3 + 6 = 15$ *Punkte*

Angenommen, 20 Prozent der Raucher*innen und ein Prozent der Nichtraucher*innen erkranken an Lungenkrebs. In einer Bevölkerung gibt es gleich viele Männer und Frauen, jedoch sind 30 Prozent der Männer und 20 Prozent der Frauen Raucher*innen. Der Einfachheit halber nehmen wir an, dass jede Person in der Bevölkerung entweder männlich oder weiblich ist. Eine Person wurde zufällig aus der Bevölkerung ausgewählt.

a) Wie groß ist die Wahrscheinlichkeit, dass die Person raucht und weiblich ist?

b) Wie groß ist die Wahrscheinlichkeit, dass die Person nicht raucht und Lungenkrebs hat?

c) Wie groß ist die Wahrscheinlichkeit, dass die ausgewählte Person Lungenkrebs hat?

d) Wenn die ausgewählte Person Lungenkrebs hat, wie groß ist die bedingte Wahrscheinlichkeit, dass es sich um einen Nichtraucher (männlich) handelt?

Bezeichne $K$ das Ereignis, dass eine zufällig ausgewählte Person Lungenkrebs hat, $R$ das Ereignis, dass eine zufällig ausgewählte Person Raucher*in ist, und $M$ das Ereignis, dass eine zufällig ausgewählte Person männlich ist.

**a)** Laut Aufgabenstellung gilt

$$\mathbb{P}(R \mid M^c) = \frac{\mathbb{P}(R \cap M^c)}{\mathbb{P}(M^c)} = 0.2.$$

Außerdem gilt $\mathbb{P}(M^c) = 0.5$ und damit $\mathbb{P}(R \cap M^c) = 0.2 \cdot 0.5 = 0.1$.

**Punktschema:** Ein Punkt für die richtige Formel, ein Punkt für das richtige Einsetzen der Werte und ein Punkt für das richtige Ergebnis.

**b)** Analog zu a) gilt $\mathbb{P}(R \cap M) = 0.3 \cdot 0.5 = 0.15$. Nach Aufgabenstellung folgt

$$\mathbb{P}(K \mid R^c) = \frac{\mathbb{P}(K \cap R^c)}{\mathbb{P}(R^c)} = 0.01.$$

Außerdem gilt $\mathbb{P}(R) = \mathbb{P}(R \cap M) + \mathbb{P}(R \cap M^c) = 0.15 + 0.1 = 0.25$ und $\mathbb{P}(R^c) = 0.75$. Damit ergibt sich $\mathbb{P}(K \cap R^c) = 0.01 \cdot 0.75 = 0.0075$.

**Punktschema:** Ein Punkt für die richtige Formel, ein Punkt für das richtige Einsetzen der Werte inklusive Berechnung der fehlenden Wahrscheinlichkeit und ein Punkt für das richtige Ergebnis.

**c)** Analog zu b) gilt $\mathbb{P}(K \cap R) = 0.2 \cdot 0.25 = 0.05$. Dann folgt

$$\mathbb{P}(K) = \mathbb{P}(K \cap R) + \mathbb{P}(K \cap R^c) = 0.05 + 0.0075 = 0.0575.$$

**Punktschema:** Ein Punkt für die richtige Formel, ein Punkt für das richtige Einsetzen der Werte inklusive Berechnung der fehlenden Wahrscheinlichkeit und ein Punkt für das richtige Ergebnis.

**d)** Gesucht ist die Wahrscheinlichkeit

$$\mathbb{P}(M \cap R^c \mid K) = \frac{\mathbb{P}(M \cap R^c \cap K)}{\mathbb{P}(K)}.$$

Nach Teilaufgabe c) ist $\mathbb{P}(K) = 0.0575$. Da ein*e Nichtraucher*in mit Wahrscheinlichkeit 0.01 an Lungenkrebs erkrankt, unabhängig davon, ob sie männlich oder weiblich ist, gilt weiter

$$\mathbb{P}(M \cap R^c \cap K) = \mathbb{P}(M \cap R^c) \cdot \mathbb{P}(K \mid R^c).$$

Da $\mathbb{P}(M \cap R^c) = \mathbb{P}(R^c \mid M) \cdot \mathbb{P}(M) = 0.7 \cdot 0.5 = 0.35$, erhalten wir

$$\mathbb{P}(M \cap R^c \cap K) = 0.35 \cdot 0.01 = 0.0035.$$

Insgesamt ergibt sich die gesuchte Wahrscheinlichkeit

$$\mathbb{P}(M \cap R^c \mid K) = \frac{\mathbb{P}(M \cap R^c \cap K)}{\mathbb{P}(K)} = \frac{0.0035}{0.0575} \approx 0.06.$$

**Punktschema:** Ein Punkt für den richtigen Ansatz inklusive Bayes, vier Punkte für die richtige Berechnung des Zählers (davon ein Punkt für die richtige Berechnung von $P(M \cap R^c)$), ein Punkt für das richtige Endergebnis.

**Aufgabe 3** $3 + 2 + 4 + 3 + 5 = 17$ *Punkte*

Die Flugzeit (in Minuten) eines Wildvogels zu seinem Nestplatz wird durch eine Zufallsvariable $X$ mit Erwartungswert $\mu = 45$ und Standardabweichung $\sigma = 8$ beschrieben. Es wird eine Stichprobe von 64 Vögeln beobachtet.

a) Bestimmen Sie eine obere Schranke für die Wahrscheinlichkeit, dass die Flugzeit eines einzelnen Vogels um mehr als 20 Minuten vom Erwartungswert abweicht.

b) Bestimmen Sie den Erwartungswert und die Standardabweichung des Stichprobenmittelwertes.

c) Bestimmen Sie unter Verwendung der Normalverteilung die Wahrscheinlichkeit, dass der Stichprobenmittelwert zwischen 43 und 47 Minuten liegt.

d) Bestimmen Sie ein 95%-Konfidenzintervall für den Mittelwert der Flugzeit, wobei nun angenommen wird, dass der Erwartungswert unbekannt ist und ein mittlerer Flugzeitwert von $\bar{\mu} = 45$ gemessen wurde.

e) Ein Vogelschutzgebiet möchte eine Ruhezone einrichten, in der die Flugzeit im Durchschnitt höchstens $d$ Minuten beträgt, sodass 90% der beobachteten Mittelwerte der 64 Vögel darunter liegen. Berechnen Sie $d$.

**a)** Unter Verwendung der Chebyshev-Ungleichung gilt

$$\mathbb{P}(|X - 45| \geq 20) \leq \frac{8^2}{20^2} = \frac{4}{25} \approx 0.16.$$

**Punktschema:** Ein Punkt für den richtigen Ansatz, ein Punkt für die richtige Anwendung der Chebyshev-Ungleichung und ein Punkt für die richtige Berechnung.

**b)** Es gilt

$$\mathbb{E}[\bar{X}_{64}] = \mu = 45 \quad \text{und} \quad \bar{\sigma}_{64} = \frac{8}{\sqrt{64}} = 1.$$

**Punktschema:** Ein Punkt für den richtigen Erwartungswert und ein Punkt für die richtige Standardabweichung.

**c)** Da $n = 64$ hinreichend groß ist, ist $\bar{X}_{64}$ annähernd normalverteilt mit $\mu_{\bar{X}} = 45$ und $\sigma_{\bar{X}} = 1$. Damit gilt

$$\begin{aligned}
\mathbb{P}(43 \leq X \leq 47) &= \mathbb{P}(X \leq 47) - \mathbb{P}(X \leq 43) \\
&\approx \Phi\left(\frac{47 - 45}{1}\right) - \Phi\left(\frac{43 - 45}{1}\right) \\
&= \Phi(2) - \Phi(-2) \\
&= \Phi(2) - (1 - \Phi(2)) \\
&\approx 0.9772 - 0.0228 \approx 0.9544.
\end{aligned}$$

**Punktschema:** Ein Punkt für den richtigen Ansatz mittels Differenz, ein Punkt für die richtige Standardisierung, ein Punkt für die Symmetrie der Verteilungsfunktion und ein Punkt für das richtige Ergebnis.

**d)** Es gilt $z_{0.975} = 1.96$. Damit ist ein 95%-Konfidenzintervall gegeben durch

$$\left[\bar{\mu} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{\mu} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = [45 - 1.96 \cdot 1,\ 45 + 1.96 \cdot 1] \approx [43.04, 46.96].$$

**Punktschema:** Ein Punkt für die richtige Intervallart/richtige Formel, ein Punkt für das richtige Einsetzen inklusive dem richtigen Quantil und ein Punkt für das richtige Ergebnis.

**e)** Wir bestimmen $d$, sodass $\mathbb{P}(\bar{X}_{64} \leq d) = 0.9$. Es gilt $z_{0.9} = 1.282$ (t-Tabelle) und damit

$$\mathbb{P}(\bar{X}_{64} \leq d) \approx \Phi\left(\frac{d - \mu}{\sigma_{\bar{X}_{64}}}\right) = 0.9 \quad \Leftrightarrow \quad \frac{d - \mu}{\sigma_{\bar{X}_{64}}} = z_{0.9} = 1.282.$$

Insgesamt

$$d = \mu + z_{0.9}\bar{\sigma}_{64} = 45 + 1.282 = 46.282.$$

**Punktschema:** Ein Punkt für den richtigen Ansatz, ein Punkt für die Normalapproximation inklusive richtiger Standardisierung, ein Punkt für das richtige Umstellen, ein Punkt für das Einsetzen der richtigen Werte inklusive dem richtigen Quantil und ein Punkt für das richtige Ergebnis.

**Aufgabe 4** $3 + 6 + 4 + 2 + 4 = 19$ *Punkte*

Ein Softwareunternehmen untersucht bei einem Entwicklerteam den Zusammenhang zwischen täglichem Kaffeekonsum (in Tassen) und der Codequalität (Score 0-100). Bei fünf Entwickler:innen wurden die folgenden Daten erhoben:

| **Entwickler** | **Kaffee (x)** | **Codequalität (y)** |
|---|---|---|
| A | 0 | 60 |
| B | 1 | 65 |
| C | 2 | 72 |
| D | 3 | 75 |
| E | 4 | 80 |

Es wird angenommen, dass ein linearer Zusammenhang vorliegt.

a) Visualisieren Sie die Daten in einem Punktplot. Wirkt ein linearer Zusammenhang plausibel?

b) Bestimmen Sie die empirische Varianz von $x$ und die empirische Kovarianz von $x$ und $y$.

c) Berechnen Sie die Regressionsgerade $y = ax + b$.

d) Schätzen Sie die Codequalität bei fünf Tassen Kaffee pro Tag.

e) Berechnen Sie den empirischen Korrelationskoeffizienten und interpretieren Sie dessen Wert hinsichtlich der Stärke und Richtung des Zusammenhangs.

**a)** Die Punkte liegen etwa auf einer aufsteigenden Linie. Daher scheint ein linearer Zusammenhang plausibel.

**Punkteschema:** Zwei Punkte für ein richtiges Diagramm und ein Punkt für die richtige Beantwortung der Frage.

**b)** Es gilt $\bar{\mu}_x = 2$ und $\bar{\mu}_y = 352/5 = 70.4$. Damit ergibt sich

$$\bar{\sigma}_x^2 = \frac{1}{4}\sum_{i=1}^{5}(x_i - \bar{\mu}_x)^2 = \frac{5}{2} = 2.5$$

und

$$\bar{c}_{xy} = \frac{1}{4}\sum_{i=1}^{5}(x_i - \bar{\mu_x})(y_i - \bar{\mu_y}) = \frac{25}{2} = 12.5.$$

**Punkteschema:** Jeweils ein Punkt für die richtigen empirischen Mittelwerte, jeweils ein Punkt für die richtige Formel der empirischen Varianz und der empirischen Kovarianz, jeweils ein Punkt für die richtige Formel der empirischen Kovarianz und das richtige Ergebnis.

**c)** Die lineare Regressionsgerade ist durch $y = ax + b$ gegeben, wobei

$$a = \frac{\bar{c}_{xy}}{\bar{\sigma}_x^2} = \frac{12.5}{2.5} = 5 \quad \text{und} \quad b = \bar{\mu}_y - a\bar{\mu}_x = 70.4 - 5 \cdot 2 = 60.4.$$

**Punkteschema:** Jeweils ein Punkt für die richtigen Formeln und jeweils ein Punkt für die richtigen Ergebnisse.

**d)** Bei fünf Tassen Kaffee pro Tag erreicht die Codequalität einen Wert von

$$y = 5x + 60.4 = 5 \cdot 5 + 60.4 = 85.4.$$

**Punkteschema:** Ein Punkt für das richtige Einsetzen des Wertes und ein Punkt für das richtige Ergebnis.

**e)** Für den empirischen Korrelationskoeffizienten gilt

$$\bar{\sigma}_y^2 = \frac{1}{4}\sum_{i=1}^{5}(y_i - \bar{\mu}_y)^2 = \frac{633}{10} = 63.3$$

und damit

$$\bar{r}_{xy} = \frac{\bar{c}_{xy}}{\bar{\sigma}_x\bar{\sigma}_y} = \frac{12.5}{1.581 \cdot 7.596} \approx 0.9937.$$

Der empirische Korrelationskoeffizient zeigt einen sehr starken, positiven Zusammenhang zwischen dem Kaffeekonsum und der Codequalität.

**Punkteschema:** Ein Punkt für die richtige empirische Varianz, ein Punkt für die richtige Formel des empirischen Korrelationskoeffizienten, ein Punkt für das richtige Ergebnis und ein Punkt für die richtige Interpretation.

**Aufgabe 5** $2 + 3 + 2 + 3 + 5 = 15$ *Punkte*

Ein Softwareunternehmen möchte überprüfen, ob die neue Version ihres Algorithmus im Durchschnitt schneller ist, als die alte Version. Für die alte Version ist bekannt, dass die durchschnittliche Laufzeit zehn Sekunden beträgt. Es werden zwölf Messungen mit der neuen Version durchgeführt, die folgende Laufzeiten (in Sekunden) ergeben:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9,1 | 8,7 | 9,5 | 9,3 | 8,9 | 9,8 | 9,0 | 8,6 | 9,2 | 8,8 | 9,4 | 9,1 |

Hinweis: Es gilt: $9,1 + 8,7 + \cdots + 9,1 = 109,4$ und $9,1^2 + 8,7^2 + \cdots + 9,1^2 = 998,7$.
Die Analyse soll auf dem Signifikanzniveau $\alpha = 0,05$ durchgeführt werden.

a) Berechnen Sie den empirischen Mittelwert und die empirische Standardabweichung der Laufzeiten.

b) Welcher statistische Test ist geeignet, um zu überprüfen, ob die neue Version im Durchschnitt schneller ist, als die alte Version? Nennen Sie die Vorraussetzungen, die für die Anwendung dieses Tests erfüllt sein müssen.

c) Formulieren Sie die Null- und Alternativhypothese für den Test, den Sie in Teilaufgabe b) vorgeschlagen haben.

d) Berechnen Sie den Wert der Teststatistik der Stichprobe.

e) Entscheiden Sie, ob die Nullhypothese verworfen werden kann. Interpretieren Sie das Ergebnis im Kontext des Problems.

**a)** Der empirische Mittelwert ist

$$\bar{\mu}_{12} = \frac{1}{12}\sum_{i=1}^{12} x_i = \frac{109.4}{12} \approx 9.1167.$$

Die empirische Standardabweichung ist

$$\bar{\sigma}_{12} = \sqrt{\frac{1}{11}\left(\sum_{i=1}^{12} x_i^2 - 12\bar{\mu}_{12}^2\right)} = \sqrt{\frac{1}{11}\left(998.7 - \frac{109.4^2}{12}\right)} \approx 0.3486.$$

**Punkteschema:** Ein Punkt für den richtigen Stichprobenmittelwert, und ein Punkt für die richtige empirische Standardabweichung.

**b)** Ein geeigneter Test ist der einseitige t-Test für eine Stichprobe. Hierzu müssen die Daten Realisierungen von unabhängig, identisch verteilten Zufallsvariablen sein, die annährend normalverteilt sind.

**Punkteschema:** Ein Punkt für die Wahl des richtigen Tests und zwei Punkte für die richtigen Voraussetzungen.

**c)** Das Testproblem lautet

$$H_0 : \mu \geq 10 \qquad \text{gegen} \qquad H_1 : \mu < 10.$$

**Punkteschema:** Zwei Punkte für die richtige Formulierung.

**d)** Für den Wert der Teststatistik gilt

$$t = \sqrt{12}\frac{\bar{\mu}_{12} - 10}{\bar{\sigma}_{12}} = \sqrt{12}\frac{9.1167 - 10}{0.3486} = -8.7775.$$

**Punkteschema:** Ein Punkt für die richtige Formel, ein Punkt für das richtige Einsetzen und ein Punkt für das richtige Ergebnis.

**e)** Der kritische Wert der Freiheitsgrade ist $f = 11$. Aus der angehängten Tabelle ergibt sich $t_{11,0.95} = 1.796$ und somit $t_{n-1,\alpha} = -1.796$. Da $t < -1.796$, wird die Nullhypothese verworfen. Mit einem 5%-Signifikanzniveau gibt es genügend statistische Evidenz, dass die neue Version des Algorithmus signifikant schneller ist als die alte Version.

**Punkteschema:** Ein Punkt für den richtigen Freiheitsgrad, zwei Punkte für den richtigen Vergleichswert, ein Punkt für die richtige Testentscheidung, ein Punkt für eine richtige Interpretation.

**Aufgabe 6** $4 + 2 + 3 + 3 + 2 = 14$ *Punkte*

In einer Universität steht ein Getränkeautomat mit drei möglichen Getränken: Kaffee, Kakao und Mangoshake (neu eingeführt vom tropikaffinen Institutsleiter). Im Rahmen eines Studierendenprojekte wird das Trinkverhalten einer großen Anzahl an Personen über mehrere Tage beobachtet. Dabei zeigt sich folgendes Muster:

- Wer heute Kaffee trinkt, entscheidet sich am nächsten Tag mit 70% Wahrscheinlichkeit für Kakao, mit 10% für Mangoshake und mit 20% wieder für Kaffee.
- Wer heute Kakao trinkt, wählt am nächsten Tag zu 60% erneut Kakao, zu 30% Mangoshake und zu 10% Kaffee.
- Wer heute Mangoshake trinkt, entscheidet sich am nächsten Tag mit 30% für Kaffee, mit 20% für Kakao und mit 50% wieder für einen Mangoshake.

Es wird angenommen, dass sich dieses Trinkverhalten durch eine Markov-Kette modellieren lässt.

a) Geben Sie die Übergangsmatrix $P$ der Markov-Kette an und stellen Sie ihren Übergangsgraphen dar.

b) Ist die Markov-Kette irreduzibel? Ist sie aperiodisch? Begründen Sie Ihre Entscheidung.

c) Berechnen Sie die Wahrscheinlichkeit, dass eine Person in genau zwei Tagen von Kaffee auf Mangoshake wechselt.

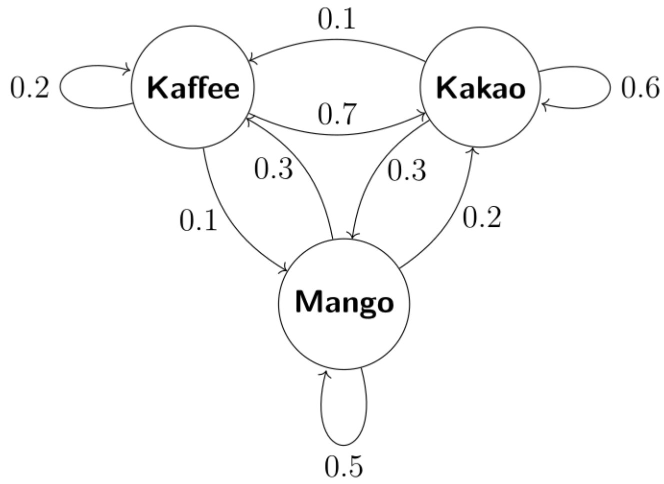
d) Zeigen Sie, dass $\pi^T = (\pi_{\text{Kaffee}}, \pi_{\text{Kakao}}, \pi_{\text{Mango}}) = \frac{1}{76}(14, 37, 25)$ eine invariante Verteilung der Markov-Kette ist. Ist diese eindeutig?

e) Wie groß ist der langfristige Anteil der Personen, die Mangoshake trinken?

**a)** Die Übergangsmatrix ist gegeben durch

$$P = \begin{pmatrix} 0.2 & 0.7 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}.$$

Die Markov-Kette lässt sich durch folgenden Übergangsgraphen darstellen:




**Punkteschema:** Zwei Punkte für die richtige Übergangsmatrix und zwei Punkte für den richtigen Übergangsgraphen.

**b)** Die Markov-Kette ist offensichtlich irreduzibel, da jeder Zustand in endlich vielen Schritten von jedem anderen Zustand erreichbar ist. Da alle Zustände Periode eins haben, ist die Markov-Kette aperiodisch.

**Punkteschema:** Jeweils ein Punkt für eine richtige Antwort inklusive richtiger Begründung.

**c)** Es gilt

$$P^2_{\text{Kaffee, Mango}} = 0.2 \cdot 0.1 + 0.1 \cdot 0.5 + 0.7 \cdot 0.3 = 7/25 = 0.28.$$

**Punkteschema:** Ein Punkt für den richtigen Ansatz, ein Punkt für die richtigen Pfade und ein Punkt für das richtige Ergebnis.

**d)** Wir betrachten das Gleichungssystem $(P - I)^T \pi \stackrel{!}{=} 0$, $\pi_1 + \pi_2 + \pi_3 = 1$, d.h.

$$\begin{aligned} -0.8\pi_1 + 0.1\pi_2 + 0.3\pi_3 &= 0, \\ 0.7\pi_1 - 0.4\pi_2 + 0.2\pi_3 &= 0, \\ 0.1\pi_1 + 0.3\pi_2 - 0.5\pi_3 &= 0, \\ \pi_1 + \pi_2 + \pi_3 &= 1. \end{aligned}$$

Einsetzen liefert, dass die gegebene Verteilung tatsächlich invariant ist. Da die Markov-Kette irreduzibel und aperiodisch ist, ist die invariante Verteilung eindeutig.

**Punkteschema:** Ein Punkt für das richtige Gleichungssystem, ein Punkt für die Substitution der richtigen Werte und ein Punkt für die Eindeutigkeit.

**e)** Dank des Ergodensatzes und nach Teilaufgabe d) trinken ca. 32.9% der Personen langfristig Mangoshake.

**Punkteschema:** Ein Punkt für das richtige Ergebnis und ein Punkt für die Begründung.