

Lösung der Probeklausur Statistik IHK  
WS 1920

**Aufgabe 1 (a)** *(insgesamt 10 Punkte: Für jedes korrekt markierte Feld 2 Punkte.)*

1. Welche Aussage ist richtig? (2 Punkte)

	Das arithmetische Mittel darf für metrisch skalierte, ordinal skalierte und nominal skalierte Merkmale berechnet werden.
x	Das arithmetische Mittel darf nur für metrisch skalierte Merkmale berechnet werden, jedoch nicht für ordinal oder nominal skalierte Merkmale.
	Das arithmetische Mittel darf nur für metrisch skalierte und für ordinal skalierte Merkmale berechnet werden, jedoch nicht für nominal skalierte Merkmale.

2. Welche Aussage ist richtig? (2 Punkte)

	Die Maßeinheit der Standardabweichung ist das Quadrat der Maßeinheit der Beobachtungen.
x	Die Maßeinheit der Standardabweichung ist die Maßeinheit der Beobachtungen.
	Die Maßeinheit der Standardabweichung ist die Wurzel der Maßeinheit der Beobachtungen.

3. Welche Aussage ist richtig? (2 Punkte)

	Das Merkmal "Anzahl von Auffahrunfällen aufgrund von Glatteis" ist qualitativ und stetig.
	Das Merkmal "Anzahl von Auffahrunfällen aufgrund von Glatteis" ist quantitativ und stetig.
x	Das Merkmal "Anzahl von Auffahrunfällen aufgrund von Glatteis" ist quantitativ und diskret.

4. Welche Aussage ist richtig? (2 Punkte)

	Modus, Median und MAD sind Lagemaße.
	Median, MAD und Varianz sind Streuungsmaße.
x	MAD, Varianz und Standardabweichung sind Streuungsmaße.

5. Welche Aussage ist richtig? (2 Punkte)

x	Die Größen von Modus, Median und arithmetischem Mittel hängen mit der Schiefe zusammen.
	Der Median ist immer kleiner als das arithmetische Mittel.
	Der Modus ist immer größer als der Median.

**Aufgabe 1 (b)** *(insgesamt 10 Punkte: Für jedes korrekt markierte Feld 1 Punkt.)*

richtig	falsch	
	x	Ältere Skifahrer (ab 61 Jahren) verursachten nur 19% der Skiunfälle.
	x	130 Unfälle gingen ohne Sachschaden aus.
x		Es gab insgesamt 97 Unfälle mit Personenschaden.
	x	Mehr als die Hälfte der Unfälle wurden von Jugendlichen (11-18 Jahre) verursacht.
	x	Es waren 227 Personen in Unfälle verwickelt.
x		Etwas mehr als 57% der Unfälle hinterließen nur einen Sachschaden.
x		Personen im Alter von 19-60 Jahren verursachten ungefähr doppelt so viele Unfälle wie Personen im Alter von über 60 Jahren.
	x	5% der über 60-Jährigen verursachten schwere Unfälle mit Sach- und Personenschaden.

**Aufgabe 2 (a):** *(insgesamt 6 Punkte)*

Randsummen zur Bestimmung der bedingten Verteilungen:

		Jahr			
		2000	2003	2006	
Nationalität	deutsch	1 287	1 743	2 095	5 125
	nicht deutsch	836	980	981	2 797
		2 123	2 723	3 076	7 922

*(je 0.5 Punkte für korrekten Wert = 1.5 Punkte)*

Bedingte Verteilungen des Merkmals “Nationalität”, gegeben das Merkmal “Jahr”:

		Jahr			
		2000	2003	2006	
Nationalität	deutsch	0.606	0.640	0.681	
	nicht deutsch	0.394	0.360	0.319	

*(muss nicht in Tabellenform angegeben sein; pro richtig angegebener relativer Häufigkeit 0.5 Punkte = 3 Punkte; falsche bedingte Verteilung: nur 1.5 Punkte)*

Vergleicht man die beiden bedingten Verteilungen, so stellt man keine wesentlichen Unterschiede fest *(1 Punkt)*. Es kann daher nicht auf einen Zusammenhang zwischen den beiden Merkmalen geschlossen werden *(0.5 Punkte)*.

**Aufgabe 2 (b):** *(insgesamt 4 Punkte)*

$\chi^2$ -Koeffizient:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(h_{ij} - e_{ij})^2}{e_{ij}} \quad (1 \text{ Punkt})$$

Konkret ist

$$\begin{aligned} \chi^2 &= \frac{(1\,287 - 1\,373.438)^2}{1\,373.438} + \frac{(1\,743 - 1\,761.598)^2}{1\,761.598} + \frac{(2\,095 - 1\,989.965)^2}{1\,989.965} \\ &\quad + \frac{(836 - 749.562)^2}{749.562} + \frac{(980 - 961.403)^2}{961.403} + \frac{(981 - 1\,086.035)^2}{1\,086.035} \\ &= 5.440 + 0.196 + 5.544 + 9.968 + 0.360 + 10.158 \\ &= 31.666 \end{aligned}$$

*(1 Punkt für Ansatz, 2 Punkte für Ausrechnen = insgesamt 3 Punkte)*

**Aufgabe 2 (c):** *(insgesamt 5 Punkte)*

Nutzung von  $\chi^2 = 35$ . Zur Einschätzung des Zusammenhangs zwischen den beiden Merkmalen wird der korrigierte Kontingenzkoeffizient  $K^*$  berechnet *(0.5 Punkte)*:

$$K^* = \frac{\sqrt{\frac{\chi^2}{\chi^2+n}}}{\sqrt{\frac{M-1}{M}}} \quad (0.5 \text{ Punkte})$$

mit  $M = \min\{k, m\}$  *(0.5 Punkte)*, hier  $M = 2$  *(0.5 Punkte)*

Damit :

$$K^* = \frac{\sqrt{\frac{35}{35+7922}}}{\sqrt{\frac{1}{2}}} = 0.094 \quad (1 \text{ Punkt})$$

Es besteht praktisch kein Zusammenhang *(1 Punkt)* zwischen den Merkmalen “Nationalität” und “Jahr” ( $K^*$  kleiner als 0.2 *(1 Punkt)*).

*Korrekturhinweis: Falls mit  $\chi^2 = 31.666$  oder irgendeinem anderen (falschen) Ergebnis aus (b) gerechnet wurde: 1 Punkt abziehen; korrektes Ergebnis ist  $K^* = 0.089$  für  $\chi^2 = 31.666$ .*

**Aufgabe 3:** (insgesamt 19 Punkte)

**Aufgabe 3 (a):** (insgesamt 12 Punkte)

Die Zuversicht (ausgedrückt auf einer Skala von “ganz pessimistisch” bis “sehr optimistisch”) ist ein ordinal skaliertes Merkmal (1 Punkt). Nutze daher den Rangkorrelationskoeffizienten nach Spearman. (1 Punkt)

Da in den Daten keine Bindungen vorliegen, kann die vereinfachte Formel zur Berechnung benutzt werden:

$$r_{SP} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n \cdot (n^2 - 1)} \quad (1 \text{ Punkt})$$

Arbeitstabelle:

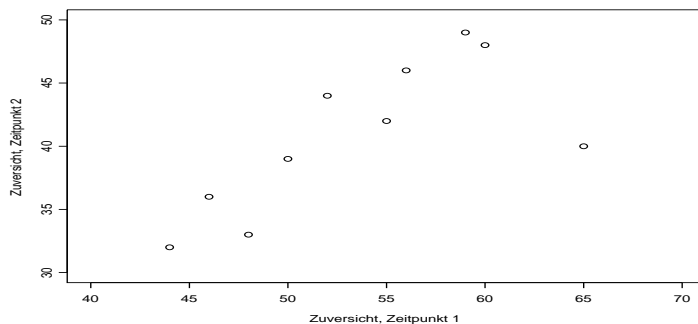
$i$	1	2	3	4	5	6	7	8	9	10	$\Sigma$
$x_i$	48	65	56	59	44	46	60	55	50	52	
$y_i$	33	40	46	49	32	36	48	42	39	44	
$rg(x_i)$	3	10	7	8	1	2	9	6	4	5	(1 Punkt)
$rg(y_i)$	2	5	8	10	1	3	9	6	4	7	(1 Punkt)
$d_i$	1	5	-1	-2	0	-1	0	0	0	-2	(1 Punkt)
$d_i^2$	1	25	1	4	0	1	0	0	0	4	36 (2 Punkte)

Also ist

$$r_{SP} = 1 - \frac{6 \cdot 36}{10 \cdot 99} = 0.78 \quad (1 \text{ Punkt}) \text{ Hinweis: } r_{XY} = 0.72; \text{ 5 Punkte geben, wenn richtig}$$

Mit  $r_{SP} = 0.78$  ist ein deutlicher bis starker monotoner positiver Zusammenhang (1.5 Punkte) zwischen den Zuversichtswerten zum ersten und zum zweiten Zeitpunkt festzustellen. Unternehmen, die zum ersten Zeitpunkt tendenziell niedrige Zuversichtswerte hatten, hatten auch zum zweiten Zeitpunkt eher niedrige Zuversichtswerte, Unternehmen mit höherer Zuversicht zum ersten Zeitpunkt sind auch zum zweiten Zeitpunkt eher zuversichtlich (1.5 Punkte).

**Aufgabe 3 (b):** (insgesamt 4 Punkte)



(Abbildung: 2 Punkte)

Beobachtung Nummer 2 (Koordinaten (65,40)) fällt aus dem beobachteten Zusammenhang der übrigen Punkte heraus (1 Punkt) → Ausreißer (1 Punkt)

**Aufgabe 3 (c):** (insgesamt 3 Punkte)

Beobachtung Nummer 2 (= Ausreißer) beeinflusst die Berechnung des Zusammenhangs (1 Punkt). Ohne diese Beobachtung ist der Zusammenhang zwischen den beiden Variablen deutlicher (1 Punkt). Es ist zu erwarten, dass sich dies in einem höheren Wert des Korrelationskoeffizienten niederschlägt (1 Punkt).

**Aufgabe 4:** (insgesamt 17 Punkte)

**Aufgabe 4 (a):** (insgesamt 11 Punkte)

Damit die angegebene Funktion eine Dichtefunktion ist, muss gelten, dass  $f(x) \geq 0$  f. alle  $x$  (1 Punkt), und es muss

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

gelten (1 Punkt).

Es ist

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_0^{0.5} c dx + \int_{0.5}^{1.5} (0.5 - x + c) dx \quad (1 \text{ Punkt}) \\ &= c \cdot x \Big|_0^{0.5} + (0.5 \cdot x - 0.5 \cdot x^2 + c \cdot x) \Big|_{0.5}^{1.5} \quad (\text{je } 1 \text{ P. f\"ur Stammfunktionen} = 2 \text{ Punkte}) \\ &= 0.5 \cdot c + 0.5 \cdot 1.5 - 0.5 \cdot 1.5^2 + c \cdot 1.5 - 0.5 \cdot 0.5 + 0.5 \cdot 0.5^2 - c \cdot 0.5 \quad (1 \text{ Punkt}) \\ &= 1.5 \cdot c + 0.75 - 1.125 - 0.25 + 0.125 \\ &= 1.5 \cdot c - 0.5 \quad (1 \text{ Punkt}) \\ &= 1 \end{aligned}$$

Damit ist  $1.5 \cdot c = 1.5$ , also  $c = 1$  (1 Punkt)

Insgesamt ist also

$$f(x) = \begin{cases} 1, & 0 \leq x < 0.5 \\ 1.5 - x, & 0.5 \leq x \leq 1.5 \\ 0 & \text{sonst} \end{cases}$$

und damit  $f(x) \geq 0$ , da

$f(x) = 1 \geq 0$  für  $0 \leq x < 0.5$  (1 Punkt)

$f(x) = 1$  für  $x = 0.5$ ,  $f(x) = 0$  für  $x = 1.5$  und  $f$  linear zwischen 0.5 und 1.5 (2 Punkte)

**Aufgabe 4 (b):** (insgesamt 6 Punkte)

Es ist  $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$  (1 Punkt)

Daher

$$\begin{aligned} E(X) &= \int_0^{0.5} x \cdot 1 dx + \int_{0.5}^{1.5} x \cdot (1.5 - x) dx \quad (1 \text{ Punkt}) \\ &= \int_0^{0.5} x dx + \int_{0.5}^{1.5} 1.5 \cdot x - x^2 dx \\ &= 0.5 \cdot x^2 \Big|_0^{0.5} + (1.5 \cdot 0.5 \cdot x^2 - \frac{1}{3} \cdot x^3) \Big|_{0.5}^{1.5} \quad (2 \text{ Punkte}) \\ &= 0.5 \cdot 0.25 + 0.5 \cdot 1.5 \cdot 1.5^2 - \frac{1}{3} \cdot 1.5^3 - 0.5 \cdot 1.5 \cdot 0.5^2 + \frac{1}{3} \cdot 0.5^3 \quad (1 \text{ Punkt}) \\ &= 0.125 + 1.6875 - 1.125 - 0.1875 + 0.0417 \\ &= 0.5417 \quad (1 \text{ Punkt}) \end{aligned}$$

**Aufgabe 5:** (insgesamt 20 Punkte)

**Aufgabe 5 (a):** (insgesamt 8 Punkte)

Vorbereitungsdauer  $X \sim N(44, 21.16)$  (1 Punkt)

Gesucht:  $P(X > 72)$  (1 Punkt)

Es ist

$$\begin{aligned} P(X > 72) &= 1 - P(X \leq 72) \quad (1 \text{ Punkt}) \\ &= 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{72 - \mu}{\sigma}\right) \\ &= 1 - P\left(Z \leq \frac{72 - 44}{4.6}\right) \quad (1 \text{ Punkt}) \\ &= 1 - \Phi(6.087) \quad (1 \text{ Punkt}) \\ &= 1 - 1 \approx 0 \quad (1 \text{ Punkt}) \end{aligned}$$

Der Wert für  $\Phi(6.087)$  kann der angegebenen Tabelle nicht entnommen werden. Man erkennt jedoch: mit wachsendem  $x$  nähern sich die Werte 1 an. (1 Punkt) Mit 0% ist es höchst unwahrscheinlich, dass ein zufällig ausgewählter Studierender sich länger als drei Tage auf die Prüfung vorbereitet hat. (1 Punkt).

**Aufgabe 5 (b):** (insgesamt 4 Punkte)

Die Aussage ist korrekt- es ist höchst unwahrscheinlich. (1 Punkt)

Begründung durch Rechnung oder Argumentation. (3 Punkte)

Auf Grund der gegebenen Werte für  $\mu$  und  $\sigma$  kann geschlussfolgert werden, dass es für einen zufällig ausgewählten Studierenden unwahrscheinlich ist, dass er sich weniger als 36 Stunden vorbereitet hat. Entsprechend ist es für drei Studierende noch unwahrscheinlicher.

$[P(X < 36)]^3 = 0.0411^3 = 0.0001$  , berechne zunächst

$$\begin{aligned} P(X < 36) &= P\left(Z \leq \frac{36 - \mu}{\sigma}\right) \\ &= \Phi(-1.7391) \\ &= 1 - \Phi(1.7391) \\ &= 0.0411 \end{aligned}$$

**Aufgabe 5 (c):** (insgesamt 8 Punkte)

Diese Aussage könnte mit einem statistischen Hypothesentest untermauert werden, Vorgehen:

- Testproblem
- Passender Test
- Teststatistik ermitteln
- Vergleich mit kritischem Wert
- Entscheidung

Fehlende Informationen (*2 Punkte*)

- repräsentative Stichprobe bzw. Vollerhebung (keine Aussage im Text) bzw.  $\bar{x}$  für Berechnung der Teststatistik
- alternativ: Annahme, dass gegebene Vorbereitungsdauer ( 44 Stunden) Erfahrungswert der letzten Jahre ist