

Statistik-Klausur vom 30. März 2005

Bearbeitungszeit: 90 Minuten

Aufgabe 1

An einer Klausur nahmen 179 Studierende teil. Es konnten die Punktzahlen 0; 0,5; 1; 1,5; 2; 2,5; ...; 80 erreicht werden. Eine Auswertung der Ergebnisse ergab folgendes Bild:

Note	Punkte	Anzahl
1	70 70,5 ... 80	28
2	60 60,5 ... 69,5	34
3	50 50,5 ... 59,5	53
4	40 40,5 ... 49,5	44
5	0 0,5 ... 39,5	20
Summe		179

- Die Tabelle basiert auf den statistischen Variablen X = „erreichte Punktzahl eines Prüflings“ und Y = „Note eines Prüflings“. Sind diese Variablen stetig oder diskret? Welche Skalierung haben diese statistischen Variablen?
- Beurteilen Sie das Punkte-Niveau der Klausur durch Berechnung eines geeigneten Lagemaßes.
- Beurteilen Sie das Noten-Niveau der Klausur durch Berechnung eines geeigneten Lagemaßes.
- Berechnen Sie für beide statistischen Variablen die Quartile $x_{0,25}$ und $x_{0,75}$ bzw. $y_{0,25}$ und $y_{0,75}$.

Aufgabe 2

VIVA und MTV zählen laut einer Umfrage zu den beliebtesten Sendern von jungen Leuten. Über die tägliche Sehdauer (in Minuten) machten 100 befragte Jugendliche folgende Angaben:

Sehdauer (in Minuten) für VIVA von ... bis unter ... (x_i)	Sehdauer (in Minuten) für MTV von ... bis unter ... (y_j)			
	0 – 30	30 – 60	60 – 120	120 – 240
0 – 30	5	18	2	0
30 – 60	4	4	7	0
60 – 120	0	8	16	11
120 – 240	1	0	9	15

- Wie viel Prozent der Jugendlichen schauen pro Tag sowohl mindestens eine Stunde VIVA als auch mindestens 30 Minuten MTV?

- b) Welchen der beiden Sender verfolgen die Jugendlichen täglich im Durchschnitt länger? Beantworten Sie diese Frage durch Berechnung einer geeigneten statistischen Maßzahl.
- c) Es wird vermutet, dass zwischen der Sehdauer für VIVA und der Sehdauer für MTV ein positiver Zusammenhang besteht. Bestätigen Sie anhand einer geeigneten statistischen Maßzahl diese Vermutung. Wie lässt sich Ihr Ergebnis interpretieren?

Aufgabe 3

Ein öffentlich-rechtlicher Fernsehsender hat aufgrund der zunehmenden Konkurrenz durch private Fernsehsender mit abnehmenden Zuschauerzahlen zu kämpfen. Um sein Programm stärker an die Wünsche der Zuschauer anpassen zu können, führt er eine Kundenbefragung durch. Die Befragung ergab folgende Ergebnisse (Mehrfachnennungen waren möglich):

- 70% der Zuschauer bevorzugen den öffentlich-rechtlichen Sender wegen der geringeren Werbeblöcke
 - 35% der Zuschauer bevorzugen den öffentlich-rechtlichen Sender wegen der solideren Nachrichtensendungen
 - 50% der Zuschauer bevorzugen den öffentlich-rechtlichen Sender wegen der interessanteren Spielfilme
 - 40% der Zuschauer bevorzugen den öffentlich-rechtlichen Sender wegen der geringeren Werbeblöcke und wegen der interessanteren Spielfilme
 - 35% der Zuschauer, die den öffentlich-rechtlichen wegen der interessanteren Spielfilme bevorzugen, bevorzugen ihn auch wegen der solideren Nachrichtensendungen
 - 10% der Zuschauer bevorzugen den öffentlich-rechtlichen Sender wegen der geringeren Werbeblöcke, wegen der solideren Nachrichtensendungen und wegen der interessanteren Spielfilme
- a) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Fernsehzuschauer den öffentlich-rechtlichen Sender wegen mindestens einer der beiden Gründe geringere Werbeblöcke und interessantere Spielfilme bevorzugt?
- b) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Fernsehzuschauer den öffentlich-rechtlichen Sender weder wegen der geringeren Werbeblöcke noch wegen der interessanteren Spielfilme bevorzugt?
- c) Ein zufällig ausgewählter Fernsehzuschauer bevorzugt den öffentlich-rechtlichen Sender wegen der solideren Nachrichtensendungen. Wie groß ist die Wahrscheinlichkeit, dass er ihn auch wegen der interessanteren Spielfilme bevorzugt?

- d) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Fernsehzuschauer den öffentlich-rechtlichen Sender nur wegen dessen interessanteren Spielfilmen, aber nicht wegen dessen solideren Nachrichtensendungen oder geringeren Werbeblöcke bevorzugt?

Aufgabe 4

Verspätet sich ein Zug über 60 Minuten, so zahlt eine Eisenbahngesellschaft einem Fahrgast eine Entschädigung in Höhe von 20 €. Ist der Zug sogar mehr als 90 Minuten verspätet, so beträgt die Entschädigung pro Fahrgast 30 €.

Pünktlich oder unter 60 Minuten verspätet sind 91,5% aller Züge. Zwischen 60 und 90 Minuten verspätet sind 8% aller Züge, über 90 Minuten verspätet sind 0,5% aller Züge.

Die Entschädigung wird nur aufgrund eines Antrages ausgezahlt. Ist ein Zug verspätet, so rechnet die Eisenbahngesellschaft mit 40 Anträgen pro Zug.

- a) Mit welcher Entschädigungshöhe pro Zugfahrt muss die Eisenbahngesellschaft rechnen?
- b) Berechnen Sie die Varianz der Variablen „Höhe der Entschädigungszahlungen (in €) der Eisenbahngesellschaft pro Zugfahrt“.
- c) Wie groß ist annähernd die Wahrscheinlichkeit, dass bei 100 Zugfahrten höchstens 12 000 € als Entschädigung zu zahlen sind (bzw. dass bei 100 Zugfahrten die durchschnittliche Entschädigungshöhe höchstens 120 € beträgt)?

Aufgabe 5

Bei der letzten Bundestagswahl am 22.9.2004 hat die SPD 38,5% aller Stimmen erhalten. Für die nächste Bundestagswahl im Jahr 2006 soll eine Wahlprognose für den Anteil der SPD-Wähler erstellt werden. Dazu wird vorgeschlagen, ein 0,95-Konfidenzintervall zu berechnen.

- a) Wie viele Wahlberechtigte sind zu befragen, damit die maximale Abweichung vom wahren Anteil höchstens $\pm 3\%$ -Punkte beträgt?
- b) Bei einer Umfrage unter 1 000 Wahlberechtigten gaben 34% an, für die SPD stimmen zu wollen. Berechnen Sie aus diesen Angaben das gewünschte Konfidenzintervall.
- c) Interpretieren Sie das unter b) berechnete Intervall.

Lösungen:

Lösung zu Aufgabe 1:

- a) Note ist eine diskrete Variable mit ordinaler Skalierung
Punkte ist eine diskrete Variable mit metrischer Skalierung
- b) X = erreichte Punktzahl eines Prüflings
Wir klassieren die Daten wie folgt:

Klasse von ... bis unter ...	n_j	kum. abs. H.
0 - 40	20	20
40 - 50	44	64
50 - 60	53	117
60 - 70	34	151
70 - 80 (einschl.)	28	179
Σ	179	•

$$x_{0,50} \approx 50 + \frac{0,5 - 64/179}{53/179} \cdot 10 = 54,8$$

d.h. 50% der Prüflinge haben mindestens 55 Punkte erzielt.

c) $Y =$ Note eines Prüflings

Note	n_i	kum. abs. H.
1	28	28
2	34	62
3	53	115
4	44	159
5	20	179
Σ	179	•

$$y_{0,50} \approx 3$$

d.h. mindestens 50% aller Prüflinge haben mindestens die Note 3 erzielt.

$$d) x_{0,25} \approx 40 + \frac{0,25 - 20/179}{44/179} \cdot 10 = 45,6$$

$$x_{0,75} \approx 60 + \frac{0,75 - 117/179}{34/179} \cdot 10 = 65,1$$

$$y_{0,25} \approx 2$$

$$y_{0,75} \approx 4$$

Lösung zu Aufgabe 2:

$$a) \frac{8 + 16 + 11 + 9 + 15}{100} = 0,59$$

d.h. 59% der Jugendlichen schauen täglich sowohl mindestens eine Stunde VIVA und als auch mindestens 30 Minuten MTV.

$$b) \bar{x} = \frac{1}{100} [25 \cdot 15 + 15 \cdot 45 + 35 \cdot 90 + 25 \cdot 180] = 87$$

$$\bar{y} = \frac{1}{100} [10 \cdot 15 + 30 \cdot 45 + 34 \cdot 90 + 26 \cdot 180] = 92,4$$

d.h. die befragten Jugendlichen schauen im Durchschnitt länger MTV.

$$c) s_{xy} = \frac{1}{100} [15 \cdot 15 \cdot 5 + 15 \cdot 45 \cdot 18 + \dots + 180 \cdot 180 \cdot 15] - 92,4 \cdot 87 = \frac{1\,029\,825}{100} - 92,4 \cdot 87 = 2\,259,45$$

d.h. es besteht ein positiver linearer Zusammenhang zwischen der Sehdauer von VIVA und MTV, da die empirische Kovarianz positiv ist.

Lösung zu Aufgabe 3:

W =geringere Werbeblöcke

N =solidere Nachrichtensendungen

S =interessantere Spielfilme

$$P(W) = 0,7 \quad P(W \cap S) = 0,4$$

$$P(N) = 0,35 \quad P(N | S) = 0,35$$

$$P(S) = 0,5 \quad P(W \cap N \cap S) = 0,1$$

$$\Rightarrow P(S \cap N) = P(N | S) \cdot P(S) = 0,35 \cdot 0,5 = 0,175$$

Arbeitstabellen:

	W	\bar{W}			N	\bar{N}	
S	0,4	0,1	0,5	S	0,175	0,325	0,5
\bar{S}	0,3	0,2	0,5	\bar{S}	0,175	0,325	0,5
	0,7	0,3	1		0,350	0,650	1

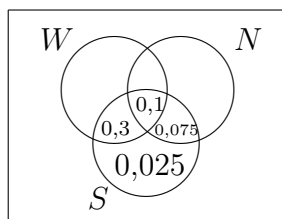
insb. sind N, S stoch. unabh.

a) $P(W \cup S) = P(W) + P(S) - P(W \cap S) = 0,7 + 0,5 - 0,4 = 0,8$
d.h. die Wahrscheinlichkeit beträgt 0,8.

b) $P(\bar{W} \cap \bar{S}) = 0,2$.
d.h. die Wahrscheinlichkeit beträgt 0,2.

c) $P(S | N) = \frac{P(S \cap N)}{P(N)} = \frac{0,175}{0,35} = 0,5$
d.h. die Wahrscheinlichkeit beträgt 0,5.

d) Um die Aufgabe zu lösen, tragen wir die Anzahlen ins Venn-Diagramm ein.



$$P(S \cap \bar{N} \cap \bar{W}) = P(S) - P(N \cap S) - P(W \cap S) + P(W \cap N \cap S) = 0,5 - 0,175 - 0,4 + 0,1 = 0,025$$

d.h. die Wahrscheinlichkeit beträgt 0,025.

Lösung zu Aufgabe 4:

X =Höhe der Entschädigungszahlungen (in €) der DB AG pro Zugfahrt

x	0	$40 \cdot 20 = 800$	$40 \cdot 30 = 1200$
$P(X = x)$	0,915	0,08	0,005

a) $E[X] = 0 \cdot 0,915 + 800 \cdot 0,08 + 1200 \cdot 0,005 = 70$

d.h. pro Zugfahrt muss die Eisenbahngesellschaft mit einer Entschädigungshöhe von 70 € rechnen.

$$\begin{aligned}
\text{b) } \text{Var}[X] &= (0 - 70)^2 \cdot 0,915 \\
&\quad + (800 - 70)^2 \cdot 0,08 \\
&\quad + (1\,200 - 70)^2 \cdot 0,005 \\
&= 53\,500
\end{aligned}$$

d.h. die Varianz beträgt 53 500.

c) ZGWS

$$P(X_1 + \dots + X_{100} \leq 12\,000) \approx F_U\left(\frac{12\,000 - 100 \cdot 70}{\sqrt{100 \cdot 53\,500}}\right) = F_U\left(\frac{12\,000 - 7\,000}{2\,313,0067}\right) = F_U(2,1617) = 0,985$$

d.h. die Wahrscheinlichkeit beträgt 0,985.

Lösung zu Aufgabe 5:

p =Anteil der SPD-Wähler in der BRD

$$\text{a) } n \geq \frac{1,96^2 \cdot 0,385 \cdot 0,615}{0,03^2} = 1\,010,6609$$

d.h. es sind mindestens 1 011 Wahlberechtigte zu befragen.

b) Faustregel $n = 1\,000 \geq 100$ o.k.

0,95-KI für p :

$$\left[0,34 - 1,96 \cdot \sqrt{\frac{0,34 \cdot 0,66}{1\,000}}; 0,34 + 1,96 \cdot \sqrt{\frac{0,34 \cdot 0,66}{1\,000}}\right]$$

$$= [0,34 - 0,0294; 0,34 + 0,0294] = [0,3106; 0,3694]$$

c) das Intervall [31%; 37%] ist ein geschätztes Intervall für den Bereich, in dem p mit der Wahrscheinlichkeit von 0,95 liegt.