

Klausur:

Explorative Datenanalyse

Prüfer: Vogt

Datum: 31.07.2009

Prüfungs-Nr.: 11015

Name: .....

Vorname: .....

Matr.-Nr.: .....

Fakultät: .....

Aufgabe	1	2	3	4	Gesamtpunkte	Note
Punkte						

Unterschrift des Prüfers: .....

- Als Hilfsmittel sind zugelassen:
- Nicht-programmierbare Taschenrechner ohne Kommunikations- oder Datenverarbeitungsfunktion (lt. Aushang des Prüfungsamtes)
  - Sechs nichtkopierte original handbeschriebene Blätter nach eigener Wahl; diese sind mit den Klausurheften abzugeben.

- Hinweise:**
1. Bitte tragen Sie oben auf diesem Deckblatt zuerst Ihre persönlichen Daten ein!
  2. Die Klausur besteht aus 4 Aufgaben.
  3. Bei den Ankreuzaufgaben gibt es immer mehrere Antwortmöglichkeiten. Von diesen ist genau eine richtig.
  4. Für eine korrekte Antwort erhalten Sie einen Punkt, für eine nicht beantwortete Frage gibt es keinen Punkt und für eine falsche Antwort wird Ihnen ein halber Punkt abgezogen. Die Punkte werden mit Gewichtungsfaktoren multipliziert, um zur Gesamtpunktzahl zu gelangen. Die jeweiligen Gewichte sind in der Aufgabenstellung angegeben.
  5. Die Klausur ist bei 50% der Gesamtpunktzahl auf jeden Fall bestanden.
  6. Nachstehend finden Sie die Aufgabensammlung mit integrierten Lösungsfeldern. Geben Sie Ihre Antworten bitte sorgfältig in den dafür vorgesehenen Bereichen! Wenn Sie zu einer Aufgabe mehr als eine Antwort markieren oder angeben, wird diese als falsch bewertet. Falls Sie eine Korrektur vornehmen müssen, kennzeichnen Sie diese bitte deutlich!
  7. Das Klausurheft besteht aus diesem Deckblatt (2 Seiten) plus 4 Aufgaben (8 Seiten); bitte zählen Sie nach! Die Heftung darf nicht gelöst werden!
  8. Zusätzlich erhalten Sie Papier für eventuelle Nebenrechnungen. Dieses ist nach Klausurende mit dem Aufgabenheft und den von Ihnen möglicherweise mitgebrachten handschriftlichen Blättern vollständig abzugeben!
  9. Alle numerischen Ergebnisse sind auf zwei Stellen genau zu runden.
  10. Viel Erfolg!!!!!!

## Aufgabe 1

20 Personen in einer Stadt wurden nach Familienstand und Religionszugehörigkeit befragt. Dabei wurden folgende Ergebnisse ermittelt:

k,v	e,l	e,g	e,g	k,l
e,g	k,v	e,v	k,g	s,v
e,v	e,l	e,v	k,v	e,l
k,g	e,v	k,l	e,g	e,v

(k: katholisch, e: evangelisch, s: sonstige, l: ledig, v: verheiratet, g: geschieden)

- a) Der Anteil der evangelischen Mitbürger unter den befragten Personen beträgt (auf zwei Dezimalstellen nach dem Komma gerundet): (Gewicht 4)

- 0,30  
 0,40  
 0,60  
 0,90  
 keine der obigen Antworten ist richtig

- b) Der Anteil der Verheirateten unter den katholischen Personen beträgt (auf zwei Dezimalstellen nach dem Komma gerundet): (Gewicht 5)

- 0,50  
 0,43  
 0,38  
 0,90  
 keine der obigen Antworten ist richtig

## Aufgabe 2

- a) Nach welcher Formel wird der Korrelationskoeffizienten von Bravais-Pearson für zwei Merkmale  $x$  und  $y$  berechnet? (Gewicht 4)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\left( \sum_{i=1}^n (x_i - \bar{x}) \right) * \left( \sum_{i=1}^n (y_i - \bar{y}) \right)}$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) * \left( \sum_{i=1}^n (y_i - \bar{y})^2 \right)}}$$

$$\square r = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{n * \sqrt{\left(\sum_{i=1}^n (x_i - \bar{x})\right) * \left(\sum_{i=1}^n (y_i - \bar{y})\right)}}$$

keine der obigen Antworten ist richtig

b) Gegeben seien  $n$  Punktepaare  $(x_i, y_i)$ ,  $i = 1, \dots, n$  und  $a$  und  $b$  seien reelle Zahlen. Für jedes  $i$  gilt

$$x_i = a + b^3 \cdot y_i$$

mit  $b < 0$

Was gilt für beliebige Werte von  $(x_i, y_i)$  für den Korrelationskoeffizienten von Bravais-Pearson? (Gewicht 5)

- $r = 1$
- $r = -1$
- $r = 0$
- $0 < r < 1$
- $-1 < r < 0$
- keine der obigen Antworten ist richtig

c) Gegeben seien  $n$  Punktepaare  $(x_i, y_i)$ ,  $i = 1, \dots, n$  und  $a$  und  $b$  seien reelle Zahlen. Für jedes  $i$  gilt

$$x_i = a - \frac{\sqrt{b}}{b} \cdot y_i$$

mit  $b > 0$

Was gilt für beliebige Werte von  $(x_i, y_i)$  für den Korrelationskoeffizienten von Bravais-Pearson? (Gewicht 5)

- $r = 1$
- $r = -1$
- $r = 0$
- $0 < r < 1$
- $-1 < r < 0$
- keine der obigen Antworten ist richtig

- d) In der folgenden Tabelle ist die Betriebszugehörigkeit  $x$  (in Jahren) und das monatliche Bruttoeinkommen  $y$  (in Euro) der Mitarbeiter einer Firma angegeben.

$i$	$x_i$	$y_i$
1	2	3.480
2	4	2.520
3	7	3.100
4	9	3.830
5	11	4.220
6	15	5.350

Bestimmen Sie den Wert des Korrelationskoeffizienten von Bravais-Pearson. Der Wert ist (auf zwei Dezimalstellen nach dem Komma gerundet)? (Gewicht 5)

- 0,67
- 0,86
- 0,79
- 0,79
- keine der obigen Antworten ist richtig

Die Betriebszugehörigkeit wird nun nicht in Jahren, sondern in Monaten und das Einkommen in Dollar und nicht in Euro angegeben wird, wobei ein Dollar damals 1,20 Euro entsprach. Der Koeffizient ist dann gleich dem gerade berechneten Koeffizienten multipliziert mit: (Gewicht 5)

- $12 * 12 * 1,20$
- 1
- 1
- $\frac{1}{12} * \frac{1}{1,20}$
- $12 * 1,20$
- keine der obigen Antworten ist richtig

### Aufgabe 3

Bei einer Untersuchung ergaben sich folgende Körpergewichte in kg bei 25 Kindern einer Schulklasse.

48, 55, 38, 63, 51, 49, 60, 44, 40, 47, 64, 39, 50, 55, 53, 46, 41, 54, 37, 49, 51, 54, 44, 57, 46

Es wurden folgende Klassen gebildet:

- 1. Klasse von 0 bis unter 40
- 2. Klasse von 40 bis unter 50
- 3. Klasse von 50 bis unter 60
- 4. Klasse von 60 bis unter 65
- 5. Klasse größer oder gleich 65

a) Die relative Häufigkeit in der 2. Klasse ist: (Gewicht 4)

- doppelt so hoch wie in der 4. Klasse.
- gleich hoch wie in der 3. Klasse
- doppelt so hoch wie in der 1. Klasse
- dreimal so hoch wie in der 4. Klasse.
- größer als in der 1. Klasse
- Keine der obigen Antworten ist richtig.

b) Die empirische Dichtefunktion  $\hat{f}(x)$  lautet: (Gewicht 4)

$$\hat{f}(x) = \begin{cases} \frac{0,12}{40} * x & \text{für } 0 \leq x < 40 \\ \frac{0,4}{10} * x & \text{für } 40 \leq x < 50 \\ \frac{0,36}{10} * x & \text{für } 50 \leq x < 60 \\ \frac{0,12}{5} * x & \text{für } 60 \leq x < 65 \\ 0 & \text{sonst} \end{cases}$$

$$\hat{f}(x) = \begin{cases} \frac{0,12}{40} & \text{für } 0 \leq x < 40 \\ \frac{0,4}{10} & \text{für } 40 \leq x < 50 \\ \frac{0,36}{10} & \text{für } 50 \leq x < 60 \\ \frac{0,12}{5} & \text{für } 60 \leq x < 65 \\ 0 & \text{sonst} \end{cases}$$

□

$$\hat{f}(x) = \begin{cases} \frac{3}{40} & \text{für } 0 \leq x < 40 \\ \frac{10}{10} & \text{für } 40 \leq x < 50 \\ \frac{9}{10} & \text{für } 50 \leq x < 60 \\ \frac{3}{5} & \text{für } 60 \leq x < 65 \\ 0 & \text{sonst} \end{cases}$$

□ keine der obigen Antworten ist richtig

c) Die empirische Verteilungsfunktion  $\hat{F}(x)$  lautet: (Gewicht 4)

☺

□

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0 + \frac{x-0}{40} * 0,12 & \text{für } 0 \leq x < 40 \\ 0,12 + \frac{x-40}{10} * 0,4 & \text{für } 40 \leq x < 50 \\ 0,52 + \frac{x-50}{10} * 0,36 & \text{für } 50 \leq x < 60 \\ 0,88 + \frac{x-60}{5} * 0,12 & \text{für } 60 \leq x < 65 \\ 1 & \text{für } x \geq 65 \end{cases}$$

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x-0}{40} * 0,12 & \text{für } 0 \leq x < 40 \\ \frac{x-40}{10} * 0,4 & \text{für } 40 \leq x < 50 \\ \frac{x-50}{10} * 0,36 & \text{für } 50 \leq x < 60 \\ \frac{x-60}{5} * 0,12 & \text{für } 60 \leq x < 65 \\ 1 & \text{für } x \geq 65 \end{cases}$$

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x-0}{40} & \text{für } 0 \leq x < 40 \\ \frac{x-40}{10} & \text{für } 40 \leq x < 50 \\ \frac{x-50}{10} & \text{für } 50 \leq x < 60 \\ \frac{x-60}{5} & \text{für } 60 \leq x < 65 \\ 1 & \text{für } x \geq 65 \end{cases}$$

keine der obigen Antworten ist richtig

d) Bestimmen Sie den Wert der empirischen Verteilungsfunktion an der Stelle  $x = 45$ ! Der Wert ist (auf zwei Dezimalstellen nach dem Komma gerundet):  
(Gewicht 4)

- 0,32
- 0,50
- 0,20
- 0,45
- Keine der obigen Antworten ist richtig.



- e) Bestimmen Sie aus den Rohdaten den Mittelwert und die Stichprobenvarianz (auf zwei Dezimalstellen nach dem Komma gerundet).

Der Mittelwert ist: (Gewicht 4)

- 48
- 49
- 49,4
- 50,5
- keine der obigen Antworten ist richtig

Die Stichprobenvarianz ist: (Gewicht 5)

- 45,2
- 55,5
- 15,28
- 7,45
- keine der obigen Antworten ist richtig

- f) Bestimmen Sie den Median, das untere Quartil und das obere Quartil mit Hilfe der empirischen Verteilungsfunktion (auf zwei Dezimalstellen nach dem Komma gerundet).

Der Median ist: (Gewicht 4)

- 49
- 50
- 49,5
- 50,5
- keine der obigen Antworten ist richtig

Das untere Quartil ist: (Gewicht 4)

- 44
- 43,25
- 52
- 41
- keine der obigen Antworten ist richtig

Das obere Quartil ist: (Gewicht 4)

- 57,2
- 55
- 54
- 56,39
- keine der obigen Antworten ist richtig

#### Aufgabe 4

- a) Ist der Korrelationskoeffizient von Bravais-Pearson für die Merkmale x und y Null, dann gilt: (Gewicht 6)

- Es besteht kein Zusammenhang zwischen den Merkmalen.
- Es besteht ein quadratischer Zusammenhang zwischen den Merkmalen.
- Der Rangkorrelationskoeffizient von Spearman ist dann in jedem Fall auch Null.
- Alle Beobachtungen sind identisch.
- keine der obigen Antworten ist richtig

b) Welche Aussage gilt für die Beziehung zwischen mittlerer quadratischer Abweichung  $d^2$  und Stichprobenvarianz  $s^2$ , wenn der Stichprobenumfang größer oder gleich 2 ist? (Gewicht 6)

- Die Stichprobenvarianz ist immer kleiner als die mittlere quadratische Abweichung.
- Die Stichprobenvarianz und die mittlere quadratische Abweichung sind immer gleich.
- Falls mindestens zwei Beobachtungen verschieden sind, ist die Stichprobenvarianz größer als die mittlere quadratische Abweichung.
- Die Stichprobenvarianz ist nur größer als die mittlere quadratische Abweichung, falls die mittlere quadratische Abweichung kleiner als Null ist.
- Die Stichprobenvarianz ist mindestens doppelt so groß wie die mittlere quadratische Abweichung.
- keine der obigen Antworten ist richtig

c) Es seien für  $a$  und  $b$  reelle Zahlen und  $y_i = a - b \cdot x_i, i = 1, \dots, n$ .

Dann gilt für die mittlere quadratische Abweichung  $d_y^2$ : (Gewicht 6)

- $d_y^2 = a - b \cdot d_x^2$
- $d_y^2 = -b^2 \cdot d_x^2$
- $d_y^2 = b^2 \cdot d_x^2$
- $d_y^2 = d_x^2$
- keine der obigen Antworten ist richtig

d) Wenn beide Korrelationskoeffizienten berechenbar sind, so ist der Rangkorrelationskoeffizient von Spearman in jedem Fall größer als der Korrelationskoeffizient von Bravais-Pearson. (Gewicht 6)

- ja
- nein
- keine der obigen Antworten ist richtig

e) Für die empirische Kovarianz zweier Merkmale gilt: (Gewicht 6)

- Die empirische Kovarianz ist immer größer oder gleich Null.
- Die empirische Kovarianz ist nur größer als Null, wenn der Korrelationskoeffizient von Bravais-Pearson kleiner als Null ist.
- Die empirische Kovarianz kann nur berechnet werden, falls alle Beobachtungen größer oder gleich Null sind.
- keine der obigen Antworten ist richtig