Klausur: Explorative Datenanalyse

Prüfer: Vogt

Datum: 25. Juli 2007

Prüfungs-Nr.: 11015

Name: .... **Vorname:** ....

Matr.-Nr.: .... **Fakultät:** ....

Aufgabe	1	2	3	4	Gesamtpunkte	Note
Punkte						

Unterschrift der Prüfer: ....

....

Als Hilfsmittel sind zugelassen:

- **Nicht-programmierbare Taschenrechner ohne Kommunikations- oder Datenverarbeitungsfunktion (lt. Aushang des Prüfungsamtes)**
- **Drei nichtkopierte 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 vier 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 zur Statistik besteht aus diesem Deckblatt (2 Seiten) plus vier Aufgaben; 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. Viel Erfolg!!!!!!

Aufgabe 1

In einer Untersuchung wurden folgende Gewichte in Kilogramm bei 25 Schulkindern einer Schulklasse registriert:

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 als 65

a) Die relative Häufigkeit in der 2. Klasse ist: *(Gewicht 5)*

- ☐ 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)*

☐

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

☐

$$\hat{f}(x) = \begin{cases} \frac{0{,}12}{40} & \text{für } 0 \le x < 40 \\ \frac{0{,}40}{50} & \text{für } 40 \le x < 50 \\ \frac{0{,}36}{60} & \text{für } 50 \le x < 60 \\ \frac{0{,}12}{70} & \text{für } 60 \le x < 65 \\ 0 & \textit{sonst} \end{cases}$$

☐

$$\hat{f}(x) = \begin{cases} \frac{0{,}03}{10} * x & \text{für } 0 \le x < 40 \\ \frac{0{,}04}{10} * x & \text{für } 40 \le x < 50 \\ \frac{0{,}036}{10} * x & \text{für } 50 \le x < 60 \\ \frac{0{,}12}{10} * x & \text{für } 60 \le x < 65 \\ 0 & \textit{sonst} \end{cases}$$

☐ keine der obigen Antworten ist richtig

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

☐

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

☐

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

☐

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

☐ 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 5)*

☐ 0,22
☐ 0,32
☐ 0,42
☐ 0,52
☐ Keine der obigen Antworten ist richtig.

e) Bestimmen Sie mit Hilfe der Verteilungsfunktion den Anteil der Kinder, die mindestens 40 kg und höchstens 55 kg wiegen. Der Wert ist (auf zwei Dezimalstellen nach dem Komma gerundet): *(Gewicht 5)*

☐ 0,38
☐ 0,48
☐ 0,58
☐ 0,68
☐ Keine der obigen Antworten ist richtig.

f) Für die Verteilungsfunktion gilt: *(Gewicht 5)*

☐ Die Höhe der Verteilungsfunktion entspricht der relativen Häufigkeit des Merkmals.
☐ Die Fläche unterhalb des Histogramms entspricht der Fläche unterhalb der Verteilungsfunktion.
☐ Die Fläche unterhalb der Verteilungsfunktion beträgt 1.
☐ Die Höhe der Verteilungsfunktion beträgt maximal 1.
☐ Die Verteilungsfunktion kann nur unterhalb der Höhe des Histogramms verlaufen.
☐ Keine der obigen Antworten ist richtig.

Aufgabe 2

20 Personen in der Stadt M wurden nach Geschlecht und Abitur (Ja, Nein) befragt. Dabei wurden folgende Ergebnisse ermittelt:

w, j	m, n	w, j	m, j	w, n
m, j	w, n	w, j	w, j	m, j
m, j	w, j	m, n	w, n	m, n
w, j	m, n	w, n	m, j	m, j

(m = männlich; w = weiblich; j = Abitur; n = kein Abitur)

a) Der Anteil der männlichen Mitbürger unter den befragten Personen beträgt (auf zwei Dezimalstellen nach dem Komma gerundet): *(Gewicht 3)*

- ☐ 0,30
- ☐ 0,50
- ☐ 0,75
- ☐ 0,80
- ☐ keine der obigen Antworten ist richtig

b) Der Anteil derjenigen ohne Abitur unter den männlichen Personen beträgt (auf zwei Dezimalstellen nach dem Komma gerundet): *(Gewicht 4)*

- ☐ 0,25
- ☐ 0,40
- ☐ 0,60
- ☐ 0,90
- ☐ keine der obigen Antworten ist richtig

c) Beim Kontingenzkoeffizienten $K = \sqrt{\frac{x^2}{x^2 + n}}$ mit $x^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$

gilt: *(Gewicht 5)*

- ☐ Je größer x^2, desto schwächer der Zusammenhang zwischen den Merkmalen.
- ☐ x^2 hat keinen Einfluss auf die Stärke des Zusammenhangs.
- ☐ x^2 kann nur Werte von 0 bis 1 annehmen.
- ☐ Kann nicht auf die Merkmale in Aufgabe 2 angewendet werden.
- ☐ keine der obigen Antworten ist richtig.

d) Für die bedingte relative Häufigkeit $h_{j|i} = \frac{n_{ij}}{n_{i.}}$ gilt: *(Gewicht 5)*

☐ Sie ist immer gleich der absoluten relativen Häufigkeit h_i.

☐ Sie ist nie gleich der absoluten relativen Häufigkeit h_i.

☐ Sie kann größer als die absolute relative Häufigkeit h_i sein.

☐ Sie ist nie größer als die absolute relative Häufigkeit h_i.

☐ Sie kann auch Werte von größer als 1 annehmen.

☐ Keine der obigen Antworten ist richtig.

Aufgabe 3

Die folgende Tabelle zeigt die Erlöse der Hopp & Reiter GmbH aus dem Verkauf von Satteln Y_i und den Stückpreis von kleinen Glückshufeisen X_i für letzten i Jahre in Euro:

i	X_i	Y_i
1	2	3.400
2	4	2.500
3	7	3.120
4	9	3.800
5	11	4.200
6	15	5.300

a) Der Korrelationskoeffizienten von Bravais Pearson wird nach der folgenden Formel berechnet: *(Gewicht 5)*

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

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

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

☐ keine der obigen Antworten ist richtig

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

- ☐ 0,67
- ☐ 0,77
- ☐ 0,87
- ☐ 1
- ☐ keine der obigen Antworten ist richtig

c) Die Erlöse und Stückpreise werden nun nicht in Euro, sondern in Dollar angegeben, wobei ein Dollar damals 0,75 Euro entsprach. Der Koeffizient ist dann gleich dem Koeffizienten aus Teil b) multipliziert mit: *(Gewicht 5)*

- ☐ 0,75
- ☐ 1
- ☐ -1
- ☐ $1/0{,}75$
- ☐ $0{,}75 * 0{,}75$
- ☐ keine der obigen Antworten ist richtig

d) Bei einem Korrelationskoeffizienten von Spearman von $r_s = 0$ zwischen den Merkmalen X und Y gilt: *(Gewicht 4)*

- ☐ Es besteht kein Zusammenhang zwischen den Merkmalen.
- ☐ Es kann ein Zusammenhang zwischen den Merkmalen bestehen.
- ☐ Es besteht ein vollständig positiver linearer Zusammenhang zwischen den Merkmalen.
- ☐ Es besteht ein vollständig negativer linearer Zusammenhang zwischen den Merkmalen.
- ☐ keine der obigen Antworten ist richtig.

e) Bei Multiplikation des Merkmals X mit dem Faktor 3 gilt für die mittlere quadratische Abweichung d^2_{3x} : *(Gewicht 4)*

- ☐ $d^2_{(3x)} = 3 * d^2_x$
- ☐ $d^2_{(3x)} = d^2_x + 3$
- ☐ $d^2_{(3x)} = 27 * d^2_x$
- ☐ $d^2_{(3x)} = 9 * d^2_x$
- ☐ $d^2_{(3x)} = d^2_x$
- ☐ keine der obigen Antworten ist richtig.

f) Für die mittlere quadratische Abweichung $d^2_{(X+Y)}$ der Summe der Merkmale $(X+Y)$ gilt bei einer quadratischen Abweichung $d^2_x > 0$ des Merkmals X und $d^2_Y > 0$ des Merkmals Y: *(Gewicht 4)*

- ☐ $d^2_{(X+Y)}$ ist immer größer als null.
- ☐ $d^2_{(X+Y)}$ kann auch null sein.
- ☐ $d^2_{(X+Y)}$ ist immer kleiner als null
- ☐ $d^2_{(X+Y)}$ ist immer ungleich null.
- ☐ $d^2_{(X+Y)} \leq d^2_x + d^2_Y$
- ☐ Keine der obigen Antworten ist richtig

g) Beim Korrelationskoeffizienten von Bravais-Peason gilt: *(Gewicht 4)*

- ☐ Ist nur 1, wenn die X und Y exakt gleich sind für jede Beobachtung.
- ☐ Kann auch 1 sein, wenn X und Y nicht exakt gleich sind für jede Beobachtung.
- ☐ Ist immer –1, wenn X und Y ein umgekehrtes Vorzeichen aufweisen.
- ☐ Ist 0, wenn X und Y niemals gleich sind.
- ☐ keine der obigen Antworten ist richtig.

h) Bei der Berechnung der Korrelationskoeffizienten von Bravais-Person und Spearman für die gleichen empirischen Beobachtungen gilt: *(Gewicht 4)*

- ☐ Sie führen stets zum exakt gleichen Ergebnis.
- ☐ Sie führen stets zu unterschiedlichen Ergebnissen.
- ☐ Sie können unterschiedliche Vorzeichen aufweisen.
- ☐ Sie sind beide ein Maß für den linearen Zusammenhang.
- ☐ Sie sind beide kein Maß für den linearen Zusammenhang.
- ☐ Keine der obigen Antworten ist richtig.

Aufgabe 4

a) Welche Aussagen gelten für die Beziehung zwischen dem Korrelationskoeffizienten von Bravais-Pearson und Spearman? Wenn der Korrelationskoeffizient von Bravais-Pearson berechnet werden kann, so kann *(Gewicht 4)*

- ☐ nicht der Rangkorrelationskoeffizient von Spearman für die Ränge der Beobachtungen bestimmt werden
- ☐ nicht in jedem Fall, aber in einigen Fällen der Rangkorrelationskoeffizient von Spearman für die Ränge der Beobachtungen bestimmt werden
- ☐ der Rangkorrelationskoeffizient von Spearman ist größer Null
- ☐ keine der obigen Antworten ist richtig

b) Wenn beide Korrelationskoeffizienten berechenbar sind, so ist der Korrelationskoeffizient von Bravais-Pearson in jedem Fall kleiner als der Rangkorrelationskoeffizient von Spearman *(Gewicht 4)*

- ☐ ja
- ☐ nein
- ☐ keine der obigen Antworten ist richtig

c) Sie wollen wissen, ob zwischen den Leistungen von Schülern einer Klasse in Mathematik und Deutsch ein Zusammenhang besteht. Daher besorgen Sie sich die letzten Noten aller Schüler in Mathematik und Deutsch. Welche Größe berechnen Sie als Maßzahl für den Zusammenhang? *(Gewicht 4)*

- ☐ Rangkorrelationskoeffizient von Spearman
- ☐ Korrelationskoeffizient von Bravais-Pearson
- ☐ die mittlere quadratische Abweichung der Notensumme
- ☐ keine der obigen Antworten ist richtig

d) Nachdem Sie alles berechnet haben, stellen der Deutschlehrer und der Mathelehrer fest, dass für einen Schüler, der in Deutsch und Mathe der beste war, eine falsche Note angegeben wurde. Statt einer 1, die sonst kein anderer Schüler bekommen hatte, war die Note dieses Schülers eine 1+. Wie ändert sich Ihr Ergebnis für den Korrelationskoeffizienten unter c) *(Gewicht 4)*

- ☐ der Korrelationskoeffizient wird größer
- ☐ der Korrelationskoeffizient wird kleiner.
- ☐ der Korrelationskoeffizient bleibt gleich
- ☐ der Korrelationskoeffizient ändert sich, das Vorzeichen der Änderung ist nicht eindeutig
- ☐ keine der obigen Antworten ist richtig

e) Was passiert mit Ihrem Ergebnis in c), wenn die Noten nicht auf einer Skala von 1-6, sondern auf einer Skala von 2-12 vergeben werden, d.h. alle Noten werden verdoppelt? *(Gewicht 4)*

- ☐ der Korrelationskoeffizient wird mit 2 multipliziert.
- ☐ der Korrelationskoeffizient bleibt gleich
- ☐ der Korrelationskoeffizient ändert sich, das Vorzeichen der Änderung ist nicht eindeutig
- ☐ keine der obigen Antworten ist richtig