Klausur: Explorative Datenanalyse

Prüfer: Vogt

Datum: 26.07.2011

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 (10 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

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

$$x_i = a - b^2 \cdot y_i$$

mit $$b > 0$$

Was gilt für beliebige Werte von (x_i, y_i) für den Korrelationskoeffizienten r von Bravais-Pearson? *(Gewicht 4)*

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

b) Nach welcher Formel wird der Korrelationskoeffizienten von Bravais-Pearson für zwei Merkmale *x* und *y* berechnet? *(Gewicht 3)*

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

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

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

☐ keine der obigen Antworten ist richtig

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

$$x_i = a - b^2 \cdot y_i$$

mit $b < 0$

Was gilt für beliebige Werte von (x_i, y_i) für den Korrelationskoeffizienten r von Bravais-Pearson? *(Gewicht 3)*

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

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

$$x_i = a + \sqrt{b} \cdot y_i$$

mit $b > 0$

Was gilt für beliebige Werte von (x_i, y_i) für den Korrelationskoeffizienten r von Bravais-Pearson? *(Gewicht 4)*

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

e) Berechnen Sie den Korrelationskoeffizienten von Bravais-Pearson für die folgenden 4 Tabellen und tragen sie ihn in das dafür vorgesehene Kästchen ein (auf zwei Dezimalstellen nach dem Komma gerundet).

i	1	2	3	4	5
x_i	9	4	0	1	9
y_i	3	2	0	-1	-3

Der Korrelationskoeffizient von Bravais-Pearson ist: *(Gewicht 3)* ☐

i	1	2	3	4	5
x_i	2	4	6	8	10
y_i	5	9	13	17	21

Der Korrelationskoeffizient von Bravais-Pearson ist: *(Gewicht 3)* ☐

Aufgabe 2

Nach der Vorlesung Explorative Datenanalyse kommt ein Kommilitone zu Ihnen und fragt, ob Frauen und Männer unterschiedliche Studiengänge studieren, also ob es einen Zusammenhang zwischen Studiengang und Geschlecht gibt. Ihr Kommilitone zeigt eine Tabelle, welche er in einer Datenbank des Statistischen Bundesamtes gefunden hat. Daraus können Sie die folgende Kontingenztabelle der Merkmale Geschlecht und Studienrichtung entnehmen.

Statistik der Studienanfänger in Deutschland		
Studienrichtung \ WS2006/ 07	männlich	weiblich
Betriebswirtschaftslehre	9804	9013
Intern. Betriebswirtschaft/ Management	1290	1881
Volkswirtschaftslehre	1754	912
(C)opyright Statistisches Bundesamt, Wiesbaden 2008		

a) Der Anteil der weiblichen Studenten unter den Studienanfängern dieser drei Studienrichtungen beträgt (auf zwei Dezimalstellen nach dem Komma gerundet): *(Gewicht 3)*

- ☐ 0,52
- ☐ 0,48
- ☐ 0,59
- ☐ 0,34
- ☐ keine der obigen Antworten ist richtig

b) Der Anteil der männlichen Studenten unter den Studienanfängern in Volkswirtschaftslehre beträgt (auf zwei Dezimalstellen nach dem Komma gerundet): *(Gewicht 3)*

- ☐ 0,07
- ☐ 0,70
- ☐ 0,34
- ☐ 0,66
- ☐ keine der obigen Antworten ist richtig

c) Mit Hilfe des Kontingenzkoeffizienten K lässt sich eine Maßzahl für den Zusammenhang zwischen 2 Merkmalen bestimmen. Nach welcher Formel wird er berechnet? (dabei sind n der Stichprobenumfang, n_{ij} die beobachteten Häufigkeiten und $\hat{n}_{ij}$ die erwarteten Häufigkeiten) *(Gewicht 4)*

☐ $K=\sqrt{\frac{X^2}{n}}$ mit $X^2=\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-\hat{n}_{ij})^2}{\hat{n}_{ij}}$

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

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

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

☐ keine der obigen Antworten ist richtig

Aufgabe 3

Bei einer Untersuchung ergaben sich folgende Körpergrößen in cm bei 25 Kindern einer Schulklasse.

148, 155, 138, 163, 151, 149, 160, 144, 140, 147, 164, 139, 150, 155, 153, 146, 141, 154, 137, 149, 151, 154, 144, 157, 146

Es wurden folgende Klassen gebildet:

1. Klasse	von 0 bis unter 140
2. Klasse	von 140 bis unter 150
3. Klasse	von 150 bis unter 160
4. Klasse	von 160 bis unter 165
5. Klasse	größer als 165

a) Die relative Häufigkeit in der 2. Klasse ist: *(Gewicht 3)*

☐ 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)*

□

$$\hat{f}(x) = \begin{cases} \frac{0{,}12}{140} & \text{für } 0 \leq x < 140 \\ \frac{0{,}4}{10} & \text{für } 140 \leq x < 150 \\ \frac{0{,}36}{10} & \text{für } 150 \leq x < 160 \\ \frac{0{,}12}{5} & \text{für } 160 \leq x < 165 \\ 0 & \text{sonst} \end{cases}$$

□

$$\hat{f}(x) = \begin{cases} \frac{0{,}12}{140} * x & \text{für } 0 \leq x < 140 \\ \frac{0{,}4}{10} * x & \text{für } 140 \leq x < 150 \\ \frac{0{,}36}{10} * x & \text{für } 150 \leq x < 160 \\ \frac{0{,}12}{5} * x & \text{für } 160 \leq x < 165 \\ 0 & \text{sonst} \end{cases}$$

□

$$\hat{f}(x) = \begin{cases} \frac{3}{140} & \text{für } 0 \leq x < 140 \\ \frac{10}{10} & \text{für } 140 \leq x < 150 \\ \frac{9}{10} & \text{für } 150 \leq x < 160 \\ \frac{3}{5} & \text{für } 160 \leq x < 165 \\ 0 & \text{sonst} \end{cases}$$

□ keine der obigen Antworten ist richtig

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

☐

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x-0}{140} * 0{,}12 & \text{für } 0 \leq x < 140 \\ \frac{x-140}{10} * 0{,}4 & \text{für } 140 \leq x < 150 \\ \frac{x-150}{10} * 0{,}36 & \text{für } 150 \leq x < 160 \\ \frac{x-160}{5} * 0{,}12 & \text{für } 160 \leq x < 165 \\ 1 & \text{für } x \geq 165 \end{cases}$$

☐

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x-0}{140} & \text{für } 0 \leq x < 140 \\ \frac{x-140}{10} & \text{für } 140 \leq x < 150 \\ \frac{x-150}{10} & \text{für } 150 \leq x < 160 \\ \frac{x-160}{5} & \text{für } 160 \leq x < 165 \\ 1 & \text{für } x \geq 165 \end{cases}$$

☐

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0 + \frac{x-0}{140} * 0{,}12 & \text{für } 0 \leq x < 140 \\ 0{,}12 + \frac{x-140}{10} * 0{,}4 & \text{für } 140 \leq x < 150 \\ 0{,}52 + \frac{x-150}{10} * 0{,}36 & \text{für } 150 \leq x < 160 \\ 0{,}88 + \frac{x-160}{5} * 0{,}12 & \text{für } 160 \leq x < 165 \\ 1 & \text{für } x \geq 165 \end{cases}$$

☐ keine der obigen Antworten ist richtig

d) Bestimmen Sie aus den Rohdaten den Median, das untere Quartil und das obere Quartil (auf zwei Dezimalstellen nach dem Komma gerundet).

Der Median ist: *(Gewicht 3)*

☐ 150
☐ 148
☐ 149
☐ 151
☐ keine der obigen Antworten ist richtig

Das untere Quartil ist: *(Gewicht 3)*

☐ 141
☐ 145
☐ 146
☐ 144
☐ keine der obigen Antworten ist richtig

Das obere Quartil ist: *(Gewicht 3)*

☐ 154
☐ 155
☐ 153
☐ 154,5
☐ keine der obigen Antworten ist richtig

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

- ☐ 150
- ☐ 149,5
- ☐ 150,5
- ☐ keine der obigen Antworten ist richtig

Das untere Quartil ist: *(Gewicht 3)*

- ☐ 144
- ☐ 143,25
- ☐ 152
- ☐ 141
- ☐ keine der obigen Antworten ist richtig

Das obere Quartil ist: *(Gewicht 3)*

- ☐ 157,2
- ☐ 155
- ☐ 154
- ☐ 156,39
- ☐ keine der obigen Antworten ist richtig

Aufgabe 4

a) Es seien für a und b reelle Zahlen und $y_i = a + b \cdot x_i$, $i = 1,\ldots,n$.
Dann gilt für die Stichprobenvarianz s_y^2: *(Gewicht 5)*

- ☐ $s_y^2 = a + b \cdot s_x^2$
- ☐ $s_y^2 = a^2 + b^2 \cdot s_x^2$
- ☐ $s_y^2 = b^2 \cdot s_x^2$
- ☐ $s_y^2 = s_x^2$
- ☐ 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) Ist der Rangkorrelationskoeffizient von Spearman 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 Korrelationskoeffizient von Bravais-Pearson ist dann in jedem Fall auch Null.
- ☐ Alle Beobachtungen sind identisch.
- ☐ 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.
- ☐ Falls die empirische Kovarianz Null ist, so ist auch der Korrelationskoeffizient von Bravais-Pearson Null.
- ☐ keine der obigen Antworten ist richtig

f) Für den Median und den Mittelwert einer Stichprobe gilt: *(Gewicht 6)*

- ☐ Der Mittelwert ist kleiner als der Median, falls die Stichprobe negative Beobachtungen beinhaltet.
- ☐ Ist der Mittelwert Null, so ist auch der Median Null.
- ☐ Falls alle Elemente der Stichprobe verschieden sind, sind Mittelwert und Median gleich.
- ☐ Mittelwert und Median sind immer verschieden.
- ☐ keine der obigen Antworten ist richtig

g) Bei einem Korrelationskoeffizienten von Bravais-Person von $r_{xy} = 0{,}5$ zwischen den Merkmalen X und Y gilt: *(Gewicht 6)*

- ☐ Wenn X um 0,5 Prozent steigt, dann steigt auch Y um 0,5 Prozent.
- ☐ Wenn X um ein Prozent steigt, dann steigt Y um 0,5 Prozent.
- ☐ 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.