

Klausur:

Explorative Datenanalyse

Prüfer: Prof. Dr. Dr. Bodo Vogt

Datum: 30.07.2013

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)
 - Drei nicht kopierte 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. Für eine korrekte Antwort erhalten Sie einen Punkt, für eine nicht beantwortete Frage und eine falsch beantwortete Frage gibt es keinen Punkt. Die Punkte werden mit Gewichtungsfaktoren multipliziert, um zur Gesamtpunktzahl zu gelangen. Die jeweiligen Gewichte sind in der Aufgabenstellung angegeben.
 4. Die Klausur ist bei 50% der Gesamtpunktzahl auf jeden Fall bestanden.
 5. Nachstehend finden Sie die Aufgabensammlung mit integrierten Lösungsfeldern. Geben Sie Ihre Antworten bitte sorgfältig in den dafür vorgesehenen Bereichen an! 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!
 6. Das Klausurheft besteht aus diesem Deckblatt (2 Seiten) plus 4 Aufgaben (7 Seiten); bitte zählen Sie nach! Die Heftung darf nicht gelöst werden!
 7. Zusätzlich erhalten Sie Papier für eventuelle Nebenrechnungen. Dieses ist nach Klausurende mit dem Aufgabenheft und den von Ihnen möglicherweise mitgebrachten handschriftlichen Blättern vollständig abzugeben!
 8. Alle numerischen Ergebnisse sind auf zwei Stellen genau zu runden.
 9. Viel Erfolg!!!!!!

Aufgabe 1

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

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

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

mit $b > 0$

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

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

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

mit $b < 0$

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

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

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

mit $b > 0$

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

- 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	3	3	3	4
y_i	3	2	3	4	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)

i	1	2	3	4	5
x_i	1	2	3	4	5
y_i	7	9	10	12	15

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)

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

- 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)

- d) Der Wert der Kontingenzkoeffizienten für die Merkmale Studienrichtung und Geschlecht lautet (auf zwei Dezimalstellen nach dem Komma gerundet): *(Gewicht 3)*

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)*

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

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

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)

Das untere Quartil ist: (Gewicht 3)

Das obere Quartil ist: (Gewicht 3)

- e) 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 3)

Das untere Quartil ist: (Gewicht 3)

Das obere Quartil ist: (Gewicht 3)

Aufgabe 4

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

- b) Welche der beiden Größen, mittlere quadratischer Abweichung d^2 und Stichprobenvarianz s^2 , ist größer, wenn der Stichprobenumfang größer oder gleich 2 ist? (Gewicht 5)

- c) Ist der Rangkorrelationskoeffizient von Spearman für die Merkmale x und y Null, dann gilt, dass ein Zusammenhang zwischen den Merkmalen „möglich“ ist oder „nicht möglich“ ist. Füllen Sie bitte die richtige Variante ein. (Gewicht 5)

- d) Ist der Rangkorrelations-koeffizient von Spearman in jedem Fall größer als der Korrelationskoeffizient von Bravais-Pearson, wenn beide Korrelationskoeffizienten berechenbar sind? (Gewicht 5)

- e) Was gilt für die empirische Kovarianz zweier Merkmale, falls die empirische Kovarianz Null ist: (Gewicht 5)

- f) Für den Median und den Mittelwert einer Stichprobe können folgende Definitionen gelten: Median kleiner Mittelwert, Median größer Mittelwert, Median gleich Mittelwert. Füllen Sie die richtige(n) ein. (Gewicht 5)