

Klausur: 11015 Explorative Datenanalyse

Prüfer: Vogt

Datum: 14. Februar 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 den 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

Bei der Auflistung der jährlichen Fahrleistung (in km) von 25 Außendienstmitarbeitern einer Firma ergibt sich folgende Urliste:

7000, 30000, 25000, 50000, 45000, 9000, 29000, 47000, 60000, 55000, 32000, 27000, 19000, 51000, 32000, 54000, 70000, 21000, 33000, 41000, 40000, 36000, 18000, 25000, 45000

Es wurden folgende Klassen gebildet:

1. Klasse von 0 bis unter 20000
2. Klasse von 20000 bis unter 40000
3. Klasse von 40000 bis unter 60000
4. Klasse von 60000 bis unter 80000
5. Klasse von 80000 bis unter 100000

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 hoch wie in der 4. Klasse.
- Keine der obigen Antworten ist richtig.

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

$$\hat{f}(x) = \begin{cases} \frac{0,16}{20000} & \text{für } 0 \leq x < 20000 \\ \frac{0,40}{20000} & \text{für } 20000 \leq x < 40000 \\ \frac{0,36}{20000} & \text{für } 40000 \leq x < 60000 \\ \frac{0,08}{20000} & \text{für } 60000 \leq x < 80000 \\ 0 & \text{sonst} \end{cases}$$

□

$$\hat{f}(x) = \begin{cases} \frac{0,16}{20000} & \text{für } 0 \leq x < 20000 \\ \frac{0,40}{40000} & \text{für } 20000 \leq x < 40000 \\ \frac{0,36}{60000} & \text{für } 40000 \leq x < 60000 \\ \frac{0,08}{80000} & \text{für } 60000 \leq x < 80000 \\ 0 & \text{sonst} \end{cases}$$

□

$$\hat{f}(x) = \begin{cases} \frac{0,16}{20000} * x & \text{für } 0 \leq x < 20000 \\ \frac{0,40}{20000} * x & \text{für } 20000 \leq x < 40000 \\ \frac{0,36}{20000} * x & \text{für } 40000 \leq x < 60000 \\ \frac{0,08}{20000} * x & \text{für } 60000 \leq x < 80000 \\ 0 & \text{sonst} \end{cases}$$

□ keine der obigen Antworten ist richtig

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

□

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x-0}{20000} * 0,16 & \text{für } 0 \leq x < 20000 \\ \frac{x-20000}{20000} * 0,40 & \text{für } 20000 \leq x < 40000 \\ \frac{x-40000}{20000} * 0,36 & \text{für } 40000 \leq x < 60000 \\ \frac{x-60000}{20000} * 0,08 & \text{für } 60000 \leq x < 80000 \\ 1 & \text{für } x \geq 80000 \end{cases}$$

□

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0 + \frac{x-0}{20000} * 0,16 & \text{für } 0 \leq x < 20000 \\ 0,16 + \frac{x-20000}{20000} * 0,40 & \text{für } 20000 \leq x < 40000 \\ 0,56 + \frac{x-40000}{20000} * 0,36 & \text{für } 40000 \leq x < 60000 \\ 0,92 + \frac{x-60000}{20000} * 0,08 & \text{für } 60000 \leq x < 80000 \\ 1 & \text{für } x \geq 80000 \end{cases}$$

$$\hat{F}(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{x-0}{20000} & \text{für } 0 \leq x < 20000 \\ \frac{x-20000}{20000} & \text{für } 20000 \leq x < 40000 \\ \frac{x-40000}{20000} & \text{für } 40000 \leq x < 60000 \\ \frac{x-60000}{20000} & \text{für } 60000 \leq x < 80000 \\ 1 & \text{für } x \geq 80000 \end{cases}$$

keine der obigen Antworten ist richtig

d) Bestimmen Sie den Wert der empirischen Verteilungsfunktion an der Stelle $x = 44000$! Der Wert ist (auf zwei Dezimalstellen nach dem Komma gerundet): (Gewicht 4)

- 0,33
- 0,43
- 0,53
- 0,63
- Keine der obigen Antworten ist richtig.

e) Bestimmen Sie den Anteil der Außendienstmitarbeiter, die eine jährliche Fahrleistung von mindestens 20000 und höchstens 60000 km haben. Der Wert ist (auf zwei Dezimalstellen nach dem Komma gerundet): (Gewicht 4)

- 0,46
- 0,56
- 0,66
- 0,76
- Keine der obigen Antworten ist richtig.

f) Beim Histogramm gilt: (Gewicht 5)

- Die Höhe der Balken beträgt maximal 1.
- Die Breite der Balken beträgt maximal 1.
- Die Gesamtfläche liegt zwischen 0 und 1.
- Die Fläche zwischen dem unteren und dem oberen Quartil beträgt 0,5.
- Die Fläche unter der Verteilungsfunktion entspricht der Fläche des Histogramms.
- Keine der obigen Antworten ist richtig.

Aufgabe 2

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

| | | | | |
|-----|-----|-----|-----|-----|
| k,v | e,l | e,g | s,g | k,l |
| e,g | k,v | k,v | k,g | s,v |
| e,v | k,l | e,v | k,v | e,l |
| k,g | e,v | k,l | e,g | e,v |

(k: kath., e: ev., s: sonst., l: led., v: verh., g: gesch.)

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

- 0,30
- 0,45
- 0,10
- 0,88
- keine der obigen Antworten ist richtig

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

- 0,46
- 0,44
- 0,38
- 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 stärker der Zusammenhang zwischen den Merkmalen.
- 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.

Aufgabe 3

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 |

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

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

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

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

- 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 4)

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

- c) 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 0,75 Euro entsprach. Der Koeffizient ist dann gleich dem Koeffizienten aus Teil b) multipliziert mit: (Gewicht 6)

- $12 * 12 * 0,75$
- 1
- 1
- $\frac{1}{12} * \frac{1}{0,75}$
- $12 * 0,75$
- keine der obigen Antworten ist richtig

d) Bei einem Korrelationskoeffizienten von Bravais-Pearson von $r_{xy} = 0$ zwischen den Merkmalen X und Y gilt: (Gewicht 5)

- 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_x^2 : (Gewicht 5)

- $d_{(-3x)}^2 = -3 * d_x^2$
- $d_{(-3x)}^2 = d_x^2 - 3$
- $d_{(-3x)}^2 = 3 * d_x^2$
- $d_{(-3x)}^2 = 9 * d_x^2$
- $d_{(-3x)}^2 = -d_x^2$
- keine der obigen Antworten ist richtig.

f) Für die Kovarianz zwischen den Merkmalen A und B und die mittlere quadratische Abweichung des Merkmals A gilt: (Gewicht 5)

- Empirische Kovarianz und mittlere quadratische Abweichung können gleich sein.
- Empirische Kovarianz und mittlere quadratische Abweichung sind keinesfalls identisch.
- Empirische Kovarianz und mittlere quadratische Abweichung sind gleich bei einem Korrelationskoeffizienten von -1.
- Empirische Kovarianz und mittlere quadratische Abweichung sind gleich bei einem Korrelationskoeffizienten von 0.
- 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 6*)
- auch der Rangkorrelationskoeffizient von Spearman für die Ränge der Beobachtungen bestimmt werden
 - 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 größer als der Rangkorrelationskoeffizient von Spearman (*Gewicht 6*)
- 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 6*)
- 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 schlechteste war, eine falsche Note angegeben wurde. Statt einer 5 minus, die sonst kein anderer Schüler bekommen hatte, war die Note dieses Schülers eine 6. Wie ändert sich Ihr Ergebnis für den Korrelationskoeffizienten unter c) (*Gewicht 6*)
- 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 6)

- 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