

Prüfungsklausur - Explorative Datenanalyse

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Σ

Name: _____, Vorname: _____

Matr.-Nr. _____ Studiengang: _____

Hinweise: (Bitte lesen Sie diese Hinweise genau durch.)

- Bitte tragen Sie als erstes Ihre persönliche Daten auf diesem Blatt ein.
- Zugelassene Hilfsmittel: Ein beidseitig mit der Hand beschriebenes DIN-A4-Blatt (Markierungen erlaubt), ein (vom WiWi-Prüfungsamt erlaubter) Taschenrechner und ein Geodreieck (oder Lineal und Winkelmesser).
- Die Klausur besteht aus fünfzehn (15) Aufgaben.
- Bei den Multiple-Choice-Aufgaben ist jeweils **genau eine** Antwort richtig.
- Für eine korrekte Antwort erhalten Sie einen Punkt. Für eine falsche Antwort oder eine nicht beantwortete Frage erhalten Sie weder einen Punkt noch wird Ihnen etwas abgezogen.
- Bitte markieren Sie bei jeder Frage die von Ihnen ausgewählte Antwort klar erkennbar. Mehrfache Antworten bei einer Frage werden als falsch beantwortet bewertet. Kennzeichnen Sie daher Korrekturen deutlich.
- Beschriften Sie Ihre Grafiken **sinnvoll**.
- Die Klausur ist bestanden, wenn mindestens **acht** (8) Punkte erzielt werden.

Aufgabe 1.

Um welchen Typ von Daten handelt es sich bei den folgenden Merkmalen:

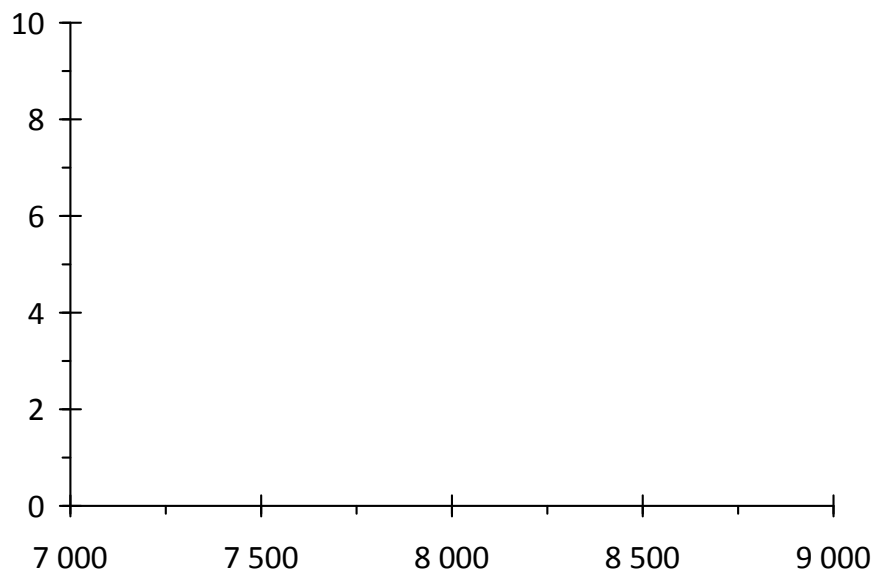
- Umsatz (in T Euro)
 - Schulnoten
 - Herkunftsland eines Produkts
 - Anzahl Kunden
- Umsatz und Schulnoten sind nominal
- Anzahl der Kunden und Umsatz sind metrisch diskret
- Herkunftsland und Anzahl der Kunden sind qualitativ
- Umsatz ist stetig und Schulnoten sind ordinal
- Herkunftsland ist ordinal und Umsatz ist nominal
-

Aufgabe 2.

In der folgenden Tabelle sind die Kupferpreise in US-\$ pro Tonne für die Werkzeuge aus dem Zeitraum vom 01.05.12 bis 07.06.12 angegeben.

7459,5	7480,5	7517,0	7545,0	7630,5
7660,5	7685,5	7708,0	7710,0	7740,0
7775,0	7800,5	7854,0	7980,0	8089,0
8195,5	8234,0	8318,0	8401,0	8528,0

Zeichnen Sie das zugehörige Histogramm mit Klassenbreiten 250, wobei die unterste Klasse bei 7250 beginnt.

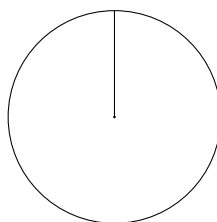


Aufgabe 3.

Bei einem Marketing-Experiment wurden Verkaufsmengen von Produkten A, B, C und D während eines festgesetzten Zeitraums betrachtet. Folgende Tabelle enthält die Verkaufsmengen der einzelnen Produkttypen.

Produkttyp	Verkaufsmenge
A	18
B	6
C	16
D	8

Vervollständigen Sie das nachfolgende Tortendiagramm unter Berücksichtigung des Pareto-Prinzips.

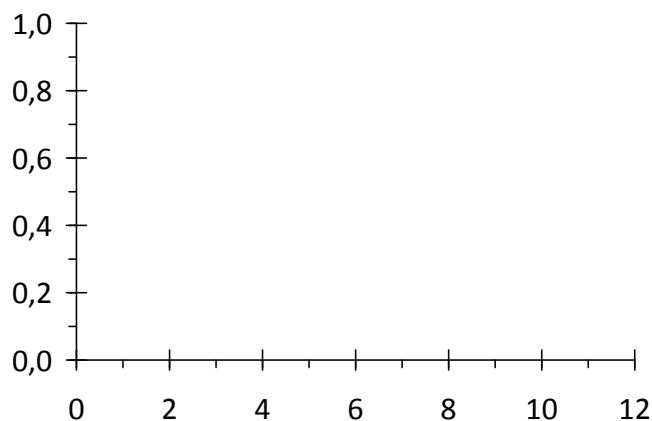


Aufgabe 4.

Für einige Autohäuser wurden die Mitarbeiterzahlen ermittelt. Die entsprechenden absoluten Häufigkeiten sind in folgender Tabelle enthalten.

Anzahl der Mitarbeiter	absolute Häufigkeit
2	2
3	3
5	10
9	4
10	1

Zeichnen Sie die zugehörige empirische Verteilungsfunktion.



Aufgabe 5.

An aufeinander folgenden Werktagen wurden folgende Haferpreise in US-\$ pro Bushel notiert:

4. Juni	5. Juni	6. Juni	7. Juni	11. Juni	12. Juni	13. Juni
2,88	2,89	2,94	3,00	2,95	2,96	3,06

Der mittlere Preis für diese Periode beträgt 2,95 US-\$ pro Bushel. Berechnen Sie den mittleren Preis in Euro pro kg Hafer. Verwenden Sie den Umrechnungskurs 1 Euro ist 1,23 US-\$ und 1 Bushel Hafer entspricht 14,515 kg.

- 2,40
 - 0,17
 - 34,81
 - 0,25
 - 52,68
-

Aufgabe 6.

Folgende Tabelle enthält die Kaffeepreise in US-\$ pro Pfund (lb.) an zufällig ausgewählten Tagen des Zeitraums vom 3.1.2012 bis 12.1.2012.

Tag	1	2	3	4	5	6
Kaffeepreis	2,27	2,20	2,22	2,26	2,35	2,34

Das arithmetische Mittel beträgt:

- 2,27
 - 2,73
 - 13,64
 - 2,21
 - 2,35
-

Aufgabe 7.

Bestimmen Sie die Stichprobenstandardabweichung zu den Daten aus Aufgabe 6.

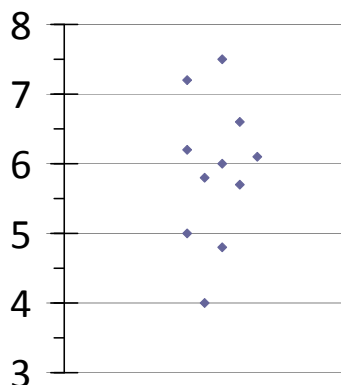
- 0,004
 - 0,061
 - 0,020
 - 0,140
 - 1,510
-

Aufgabe 8.

Für die US-Filiale Ihrer internationalen Spedition haben Sie von Ihren elf fest angestellten Fahrern den mittleren Kraftstoffverbrauch bei ihren Fahrten in miles per gallon notiert:

Fahrer	1	2	3	4	5	6	7	8	9	10	11
Verbrauch	6,1	7,2	5,7	5,0	6,6	4,0	6,0	7,5	4,8	5,8	6,2

Zeichnen Sie für diese Daten einen Boxplot in folgendes Diagramm:



Aufgabe 9.

Für Lage- und Steuungsmaßzahlen gilt:

- Das arithmetische Mittel kann aus dem Boxplot abgelesen werden.
 - Die Stichprobenstandardabweichung s_x ist immer größer als das arithmetische Mittel \bar{x} .
 - Das arithmetische Mittel \bar{x} ist immer größer als der Median x_{med} .
 - Der Median kann auch für ordinale Daten bestimmt werden.
 - Das harmonische Mittel kann auch negativ sein.
-

Aufgabe 10.

Für die Abhängigkeitsmaßzahlen χ^2 und K gilt:

- $\chi^2 > 0$ bedeutet, dass ein positiver monotoner Zusammenhang besteht.
- $\chi^2 < 0$ bedeutet, dass ein negativer linearer Zusammenhang besteht.
- χ^2 kann auch für stetige Variablen sinnvoll interpretiert werden.
- $K > 1$ bedeutet, dass ein eklatanter Zusammenhang besteht.
- K ist kleiner oder gleich $\sqrt{\chi^2}$.

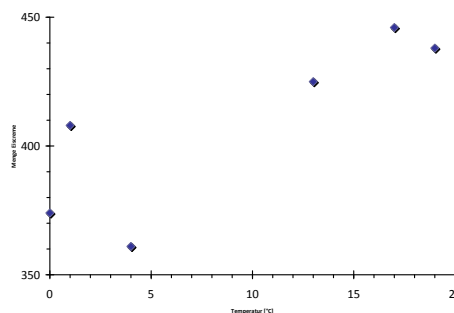
Aufgabe 11.

Es wird untersucht, ob ein Zusammenhang zwischen der verkauften Menge an Eiscreme und der mittleren Tagestemperatur besteht. In der folgenden Tabelle sind für 6 verschiedene Perioden von jeweils vier Wochen die verkaufte Menge Eiscreme y_i und die mittlere Tagestemperatur x_i angegeben.

i	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	13	425	16	266,78	65,33
2	17	446	64	1393,78	298,67
3	0	374	81	1201,78	312,00
4	4	361	25	2272,11	238,33
5	19	438	100	860,44	293,33
6	1	408	64	0,44	5,33
Summe	54	2452	350	5995,33	1213,00

Der Pearsonsche Korrelationskoeffizient ist:

- 0,02
- 0,71
- 0,55
- 1,02
- 0,84



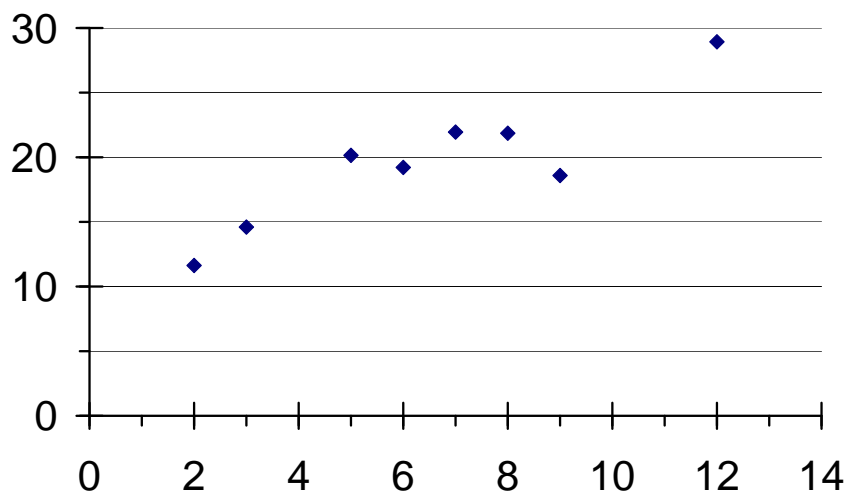
Aufgabe 12.

Bestimmen Sie die Parameter a und b der Regressionsgerade $y = a + bx$ zu den Daten aus Aufgabe 11.

- $a = 377,48, \quad b = 3,47$
 - $a = 406,85, \quad b = 0,20$
 - $a = 408,19, \quad b = -0,06$
 - $a = 401,13, \quad b = 0,84$
 - $a = 341,422, \quad b = 2,50$
-

Aufgabe 13.

Tragen Sie in das nachfolgende Streudiagramm die entsprechende Regressionsgerade ein. Die Regressionsparameter a und b betragen hier 10, 3 und 1, 43.



Aufgabe 14.

Für die Regressionsgerade $y = a + bx$ und den Pearsonschen Korrelationskoeffizienten r_{xy} gilt:

- Die Achsenabschnitt a und der Pearsonsche Korrelationskoeffizient r_{xy} haben immer umgekehrte Vorzeichen.
- Die Steigung b ist immer kleiner als der Pearsonschen Korrelationskoeffizient r_{xy} .
- Aus $b > 0$ folgt $r_{xy} > 0$.
- Der Nullpunkt $(0, 0)$ liegt immer auf der Regressionsgerade.
- Für die Steigung b gilt immer $-1 \leq b \leq 1$.

Aufgabe 15.

Zum Vergleich des stetigen Merkmals Einkommen [in Euro] über mehr als zwei Gruppen ist folgende grafische Darstellung sinnvoll:

- vergleichende Boxplots.
- vergleichendes Balkendiagramm.
- eine gemeinsame empirische Verteilungsfunktion.
- Mosaikplot.
- Tortendiagramm.