

**Prüfungsklausur - Explorative Datenanalyse**

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\Sigma$

Name: \_\_\_\_\_, Vorname: \_\_\_\_\_

Matr.-Nr. \_\_\_\_\_ Studiengang: \_\_\_\_\_

**Hinweise:**

- 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 15 Aufgaben.
- Bei den Multiple-Choice-Aufgaben ist jeweils nur eine Antwort richtig.
- Für eine korrekte Antwort erhalten Sie einen Punkt, für eine falsche Antwort wird Ihnen ein halber Punkt abgezogen. Für 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 6 Punkten erzielt werden.

## Aufgabe 1.

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

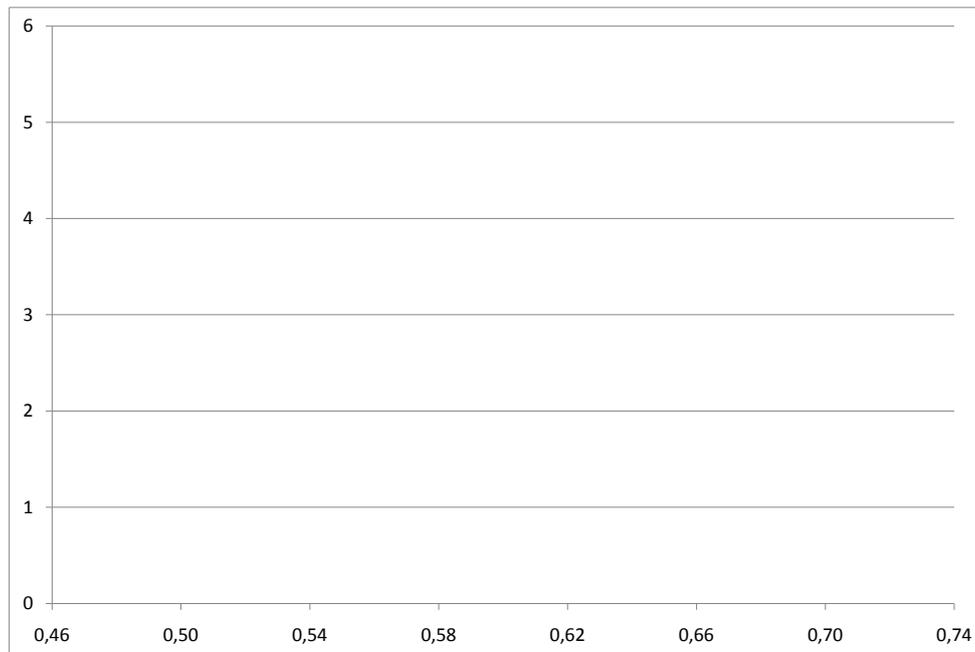
- Alter (in Jahren)
  - Anzahl der Zimmer pro Wohnung
  - Geschlecht
  - Lohnsteuerklasse
- Alter und Lohnsteuerklasse sind ordinal
- Anzahl der Zimmer und Lohnsteuerklasse sind metrisch diskret
- Lohnsteuerklasse und Geschlecht sind qualitativ
- Geschlecht und Anzahl der Zimmer sind ordinal
- Alter ist ordinal und Geschlecht ist nominal
- 

## Aufgabe 2.

In der folgenden Tabelle sind die jährlichen risikoneutralen Zinssätze (EUREPO) in % für die Werktage aus dem Zeitraum vom 15.6.10 bis 12.7.10 angegeben.

0,492	0,496	0,508	0,515	0,545
0,557	0,568	0,569	0,573	0,590
0,576	0,555	0,629	0,629	0,624
0,622	0,639	0,666	0,707	0,679

Zeichnen Sie das zugehörige Histogramm mit Klassenbreiten 0,04, wobei die unterste Klasse bei 0,48 beginnt.

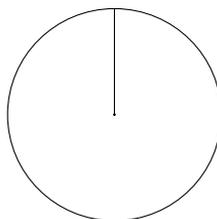


### Aufgabe 3.

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

Produkttyp	Verkaufsmenge
A	10
B	60
C	50

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

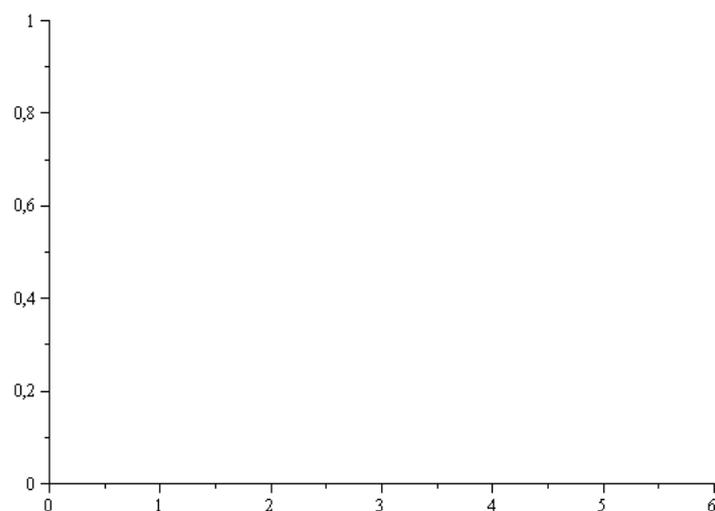


### Aufgabe 4.

In einem Ausschnitt aus dem Münchner Mietspiegel wurde die Anzahl der Zimmer pro Wohnung angegeben. Die entsprechenden absoluten Häufigkeiten sind in folgender Tabelle enthalten.

Anzahl der Zimmer	absolute Häufigkeit
1	2
2	4
3	10
4	3
5	1

Zeichnen Sie die zugehörige empirische Verteilungsfunktion.



### Aufgabe 5.

An aufeinander folgenden Werktagen wurden folgende Ölpreise in US-\$ pro Barrel notiert:

1. Juli	2. Juli	5. Juli	6. Juli	7. Juli	8. Juli	9. Juli
75,43	72,60	72,35	72,65	74,94	76,41	76,74

Der mittlere Preis für diese Periode beträgt 74,45 US-\$ pro Barrel. Berechnen Sie den mittleren Preis in Euro pro Liter Öl. Verwenden Sie den Umrechnungskurs 1 Euro ist 1,41 US-\$ und die Annäherung 1 Barrel entspricht 159 Liter.

- 0,66
  - 0,24
  - 0,33
  - 8394,94
  - 74,45
- 

### Aufgabe 6.

Folgende Tabelle enthält die Preise von Commerzbank-Aktien an zufällig ausgewählten Tagen des Zeitraums vom 1.4.2010 bis 9.7.2010.

Tag	1	2	3	4	5	6	7
Aktienpreis	6,42	6,34	6,27	5,93	5,55	5,62	6,29

Das arithmetische Mittel beträgt:

- 5,93
  - 7,07
  - 5,30
  - 42,42
  - 6,06
- 

### Aufgabe 7.

Bestimmen Sie die Stichprobenstandardabweichung zu den Daten aus Aufgabe 6.

- 0,13
  - 0,39
  - 0,02
  - 0,36
  - 0,15
-

### Aufgabe 8.

Zeichnen Sie einen Boxplot für die Daten der Aufgabe 6.



---

### Aufgabe 9.

Für die Maßzahl  $\chi^2$  und den Kontingenzkoeffizienten  $K$  gilt:

- $K$  kann auch negative Werte annehmen
- $K > 1$  bedeutet, dass ein starker Zusammenhang besteht
- $\chi^2 > 0$  bedeutet, dass ein positiver monotoner Zusammenhang besteht
- $\chi^2 < 0$  bedeutet, dass ein negativer linearer Zusammenhang besteht
- $\chi^2 \geq 0$

---

### Aufgabe 10.

Für die Regressionsgerade  $y = a + bx$  gilt:

- Alle Punkte  $(x_i, y_i)$  liegen immer auf der Regressionsgerade,  $i = 1, \dots, n$
- Der Achsenabschnitt  $a$  und die Steigung  $b$  haben immer dasselbe Vorzeichen
- Die Steigung  $b$  und der Pearsonsche Korrelationskoeffizient  $r_{xy}$  haben immer dasselbe Vorzeichen
- Die Steigung  $b$  ist immer gleich dem Pearsonschen Korrelationskoeffizienten  $r_{xy}$
- Aus  $r_{xy} > 0$  folgt  $b < 0$

### Aufgabe 11.

Gegeben sind folgende Daten:

	$x_i$	$y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	11	8,33	6,61	1,05	2,64
	14	9,96	31,04	7,05	14,80
	6	7,24	5,90	0,004	0,16
	4	4,26	19,61	9,27	13,48
	12	10,84	12,76	12,50	12,63
	7	4,82	2,04	6,17	3,55
	5	5,68	11,76	2,64	5,57
Summe	59	51,13	89,71	38,69	52,82

Der Pearsonsche Korrelationskoeffizient ist:

- 1,02
  - 0,90
  - 0,02
  - 0,90
  - 0,80
- 

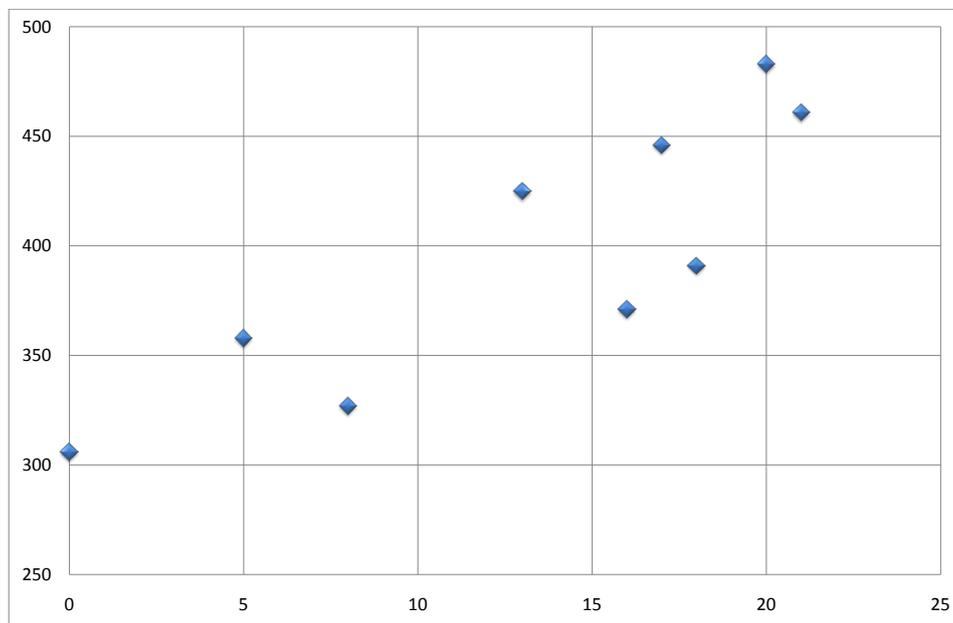
### Aufgabe 12.

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

- $a = -4,20, b = 1,37$
  - $a = 26,85, b = 2,32$
  - $a = 2,33, b = 0,59$
  - $a = -0,25, b = 0,90$
  - $a = 85,86, b = 0,59$
-

### Aufgabe 13.

Tragen Sie in das nachfolgende Streudiagramm die entsprechende Regressionsgerade ein. Die Regressionsparameter  $a$  und  $b$  betragen hier 301,5 und 7,2.



---

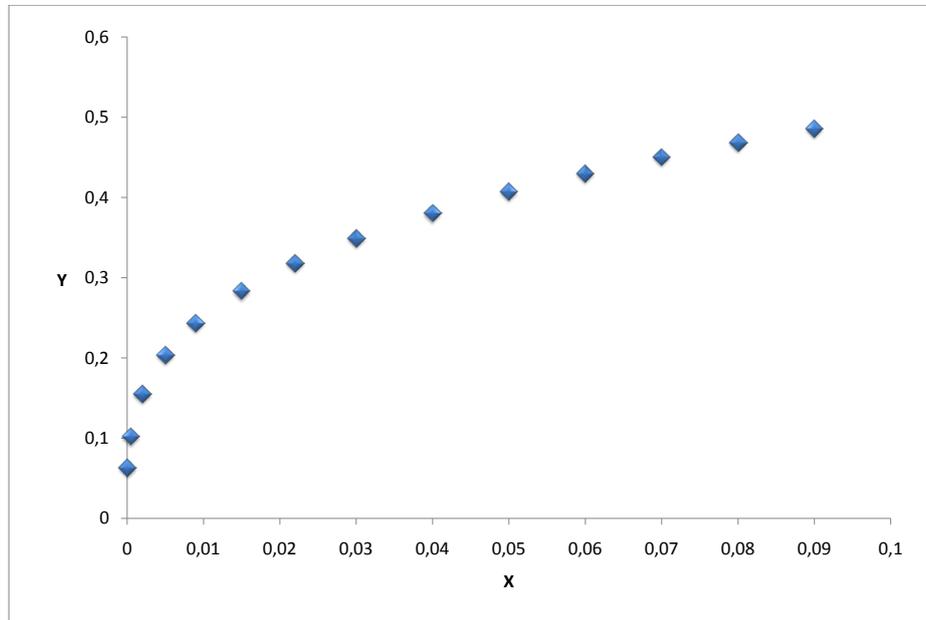
### Aufgabe 14.

Für den Spearmanschen Rangkorrelationskoeffizienten  $r_{S,xy}$  gilt:

- Keine Rangkorrelation ( $r_{S,xy} = 0$ ) bedeutet, dass kein Zusammenhang zwischen den zwei Merkmalen besteht
  - $|r_{S,xy}| = 1$  bedeutet, dass Daten auf einer Gerade liegen
  - Positive Rangkorrelation ( $r_{S,xy} > 0$ ) bedeutet, dass ein positiver linearer Zusammenhang vorhanden ist
  - $r_{S,xy}$  ist auch für ordinalskalierte Daten anwendbar
  - $0 \leq r_{S,xy} \leq 1$
-

## Aufgabe 15.

Für die im nachfolgenden Streudiagramm dargestellten Daten



- ist die Berechnung der Spearmanschen Rangkorrelationskoeffizienten **nicht** sinnvoll.
  - ist die Berechnung der Spearmanschen Rangkorrelationskoeffizienten genau so wenig sinnvoll wie die Berechnung der Pearsonschen Korrelationskoeffizienten.
  - liegt ein starker Zusammenhang der beiden Merkmale  $X$  und  $Y$  vor, den man am besten mit dem Spearmanschen Rangkorrelationskoeffizienten quantifiziert.
  - liegt ein starker Zusammenhang der beiden Merkmale  $X$  und  $Y$  vor, den man am besten mit dem Pearsonschen Korrelationskoeffizienten quantifiziert.
  - ist am besten eine lineare Regression durchzuführen.
-