

# Statistik – Übungen SS 2024

## Blatt 1: Beschreibende Statistik

### Daten und Merkmale

1. Die folgende Tabelle zeigt die Ergebnisse einer Befragung von 15 Mitarbeiterinnen und Mitarbeitern eines großen Universitätsinstituts mit insgesamt 127 Beschäftigten. Folgende Merkmale wurden erhoben: **Alter** (in vollendeten Lebensjahren), **Tätigkeitsbereich** (Forschung, Lehre, Administration), **Postleitzahl** (Wohnort), **Körpergröße** (in cm), **Familienstand** (0: ledig, 1: verheiratet, 2: geschieden), **Sport** (wie oft betreiben Sie Sport) (0: nie, 1: selten, 2: mäßig, 3: häufig).

ID	Alter	Tätigkeit	Postleitzahl	Körpergröße	Familienstand	Sport
1	42	Forschung	8302	178	1	2
2	30	Forschung	1190	185	0	0
3	39	Lehre	8075	179	1	1
4	58	Lehre	8041	175	0	3
5	45	Forschung	8700	176	1	1
6	31	Forschung	8044	161	0	3
7	25	Forschung	8010	173	0	2
8	45	Lehre	8020	174	1	1
9	28	Forschung	8045	171	0	0
10	43	Administration	8010	167	2	3
11	32	Administration	8112	184	1	1
12	36	Lehre	8020	172	0	3
13	59	Forschung	8043	168	2	3
14	45	Administration	8020	181	1	1
15	36	Lehre	8042	172	0	3

- a) Erklären Sie anhand der vorliegenden Datentabelle die Begriffe
- Grundgesamtheit
  - Stichprobe
  - Merkmalsträger (Objekt)
  - Merkmal (Variable)
  - Merkmalsausprägung
  - Realisation
- b) Geben Sie bei jedem Merkmal das dazugehörige Skalenniveau an und bestimmen Sie, ob das Merkmal stetig oder diskret ist.

2. **P 1** Geben Sie bei jedem Merkmal ein Beispiel und das dazugehörige Skalenniveau an und bestimmen Sie, ob das Merkmal stetig oder diskret ist:

<b>Merkmal</b>	<b>Ausprägung</b>	<b>Skalenniveau</b>	<b>diskret/stetig</b>
Arbeitslosenquote			
Anmeldezahlen einer Statistik VO-Prüfung			
Parteipräferenz			
BMI (Body-Mass-Index)			
Platzierung bei einem 100-m-Lauf			
IBAN			
Monatliches Nettoeinkommen			
Gehaltsgruppe für wissenschaftliches Personal einer österreichischen Universität			

## Verteilung eines Merkmals, Graphische Darstellungen

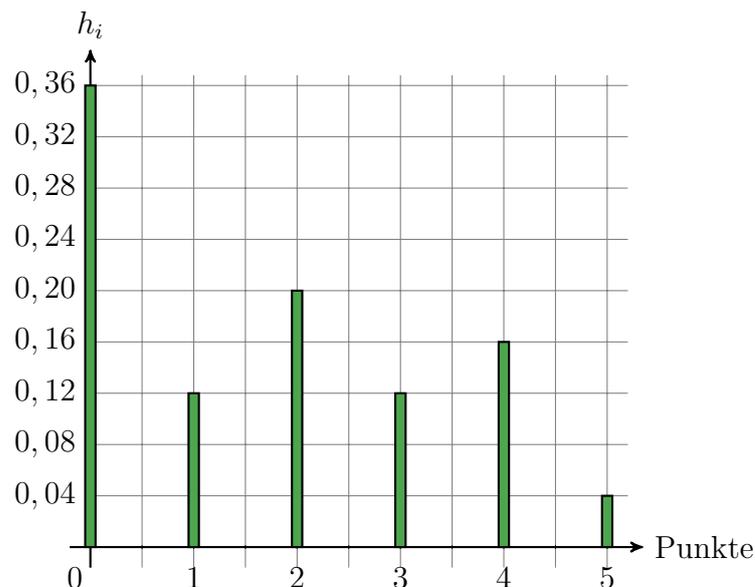
3. In einer Umfrage wurden 60 Studierende der TU-Graz nach der Anzahl Ihrer Prüfungsantritte in Mathematik I gefragt. Die Antworten finden Sie in nachfolgender Tabelle zusammengefasst:

1	1	1	5	2	4	2	5	3	2	3	2
1	5	2	3	4	4	2	3	4	5	5	2
2	4	3	2	4	2	1	1	1	1	3	2
1	4	3	1	2	5	2	2	3	4	1	1
2	2	3	1	2	3	2	3	3	1	4	1

Da viele gleiche Werte auftreten und das Merkmal diskret ist, bietet es sich an, die Werte zu zählen und die Ergebnisse in einer Häufigkeitstabelle zusammenzufassen:

Anzahl der Antritte	1	2	3	4	5
absolute Häufigkeit $H_i$	15	18	12	9	6

- a) Geben Sie die relativen und die kumulierten relativen Häufigkeiten an.
  - b) Stellen Sie die Daten durch ein Diagramm dar.
  - c) Zeichnen Sie eine geeignete Verteilungsfunktion.
  - d) Bestimmen und interpretieren Sie den Wert  $F(3)$ !
4. P 2 In der Statistik Übung können maximal 5 Mitarbeitspunkte erreicht werden. Die nachstehende Verteilung zeigt die relativen Häufigkeiten der erreichten Punkte, die die 50 Teilnehmenden einer Übungsgruppe im letzten Semester erreicht haben:



- a) Geben Sie die absoluten, relativen und die kumulierten relativen Häufigkeiten an.
- b) Zeichnen Sie eine geeignete Verteilungsfunktion.
- c) Bestimmen und interpretieren Sie den Wert  $F(4)$ !

5. In der letzten Statistikklausur mit 50 teilnehmenden Studierenden wurden folgende Punkte erreicht:

23	55,5	89	15	76	64	34	52,5	13	29,5
56,5	45	52	27	14	73	51,5	77	82	27,5
39	14	60	66	62	51	42	44,5	93	16,5
5	63	37	76,5	23	44	84	42	64,5	23
94	56	17	40	43,5	58	52	18	39	76

Da sehr viele Daten vorliegen, die fast alle voneinander abweichen, versucht man die Unübersichtlichkeit der Daten zu vermindern, indem man die Daten zu Klassen zusammenfasst. Eine mögliche Klasseneinteilung mit den absoluten Häufigkeiten für den vorliegenden Datensatz wäre:

Note	5	4	3	2	1
erreichte Punkte	[0; 50]	]50; 60]	]60; 70]	]70; 80]	]80; 100]
absolute Häufigkeit $H_i$	25	10	5	5	5

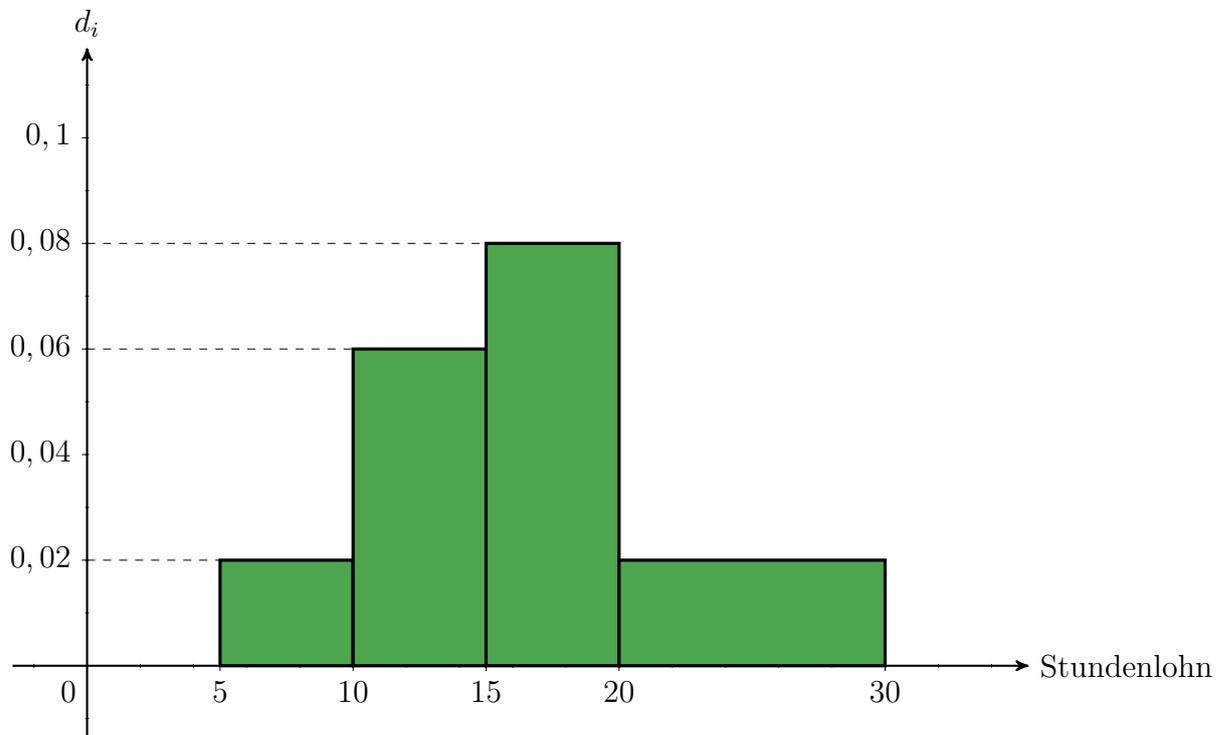
- Ergänzen Sie in der Häufigkeitstabelle die relativen Häufigkeiten sowie die absolute und die relative Besetzungsdichte.
  - Stellen Sie die Information auf geeignete Weise graphisch dar. Beachten Sie dabei insbesondere die unterschiedliche Klassenbreite des klassierten Merkmals "erreichte Punkte".
  - Zeichnen Sie die geeignete Verteilungsfunktion!
6. **P 3** Im Rahmen eines ambulanten Gewichtsreduktionsprogramms wurde bei der Aufnahme der sogenannte Body-Mass-Index (BMI) (Angaben in  $\text{kg}/\text{m}^2$ ), berechnet als Quotient aus dem Körpergewicht (Angaben in kg) und dem Quadrat der Körpergröße (Angaben in m), ermittelt. Gemäß einer medizinischen Adipositas-Klassifikation ist folgende Klasseneinteilung vorzunehmen:

Körpergewicht	BMI ( $\text{kg}/\text{m}^2$ )
Normalgewicht	20 $\text{kg}/\text{m}^2$ bis unter 25 $\text{kg}/\text{m}^2$
Adipositas Grad I	25 $\text{kg}/\text{m}^2$ bis unter 30 $\text{kg}/\text{m}^2$
Adipositas Grad II	30 $\text{kg}/\text{m}^2$ bis unter 40 $\text{kg}/\text{m}^2$
Adipositas Grad III	40 $\text{kg}/\text{m}^2$ bis unter 45 $\text{kg}/\text{m}^2$

Man erhielt die folgenden Daten:

BMI	[20; 25[	[25; 30[	[30; 40[	[40; 45[
Anzahl der Teilnehmenden	16	40	16	8

- a) Stellen Sie die Daten durch ein geeignetes Diagramm dar.
  - b) Zeichnen Sie eine geeignete Verteilungsfunktion und beantworten Sie mit deren Hilfe folgende Fragen:
    - i. wie groß ist näherungsweise der Anteil der Teilnehmenden mit einem BMI von mehr als 35?
    - ii. welcher BMI wird von 50 % der Teilnehmenden unterschritten?
7. P 4 Ein Betrieb hat insgesamt 300 Beschäftigte. In der nachstehenden Abbildung sind die Stundenlöhne dieser Beschäftigten dargestellt:



- a) Geben Sie tabellarisch die absoluten, relativen und kumulierten relativen Häufigkeiten an.
- b) Zeichnen Sie die zugehörige Verteilungsfunktion.
- c) Bestimmen und interpretieren Sie den Wert  $F(15)$ !

## Lagemaße, Streuungsmaße, Konzentrationsmaße

8. Die 12 Filialen eines Kaufhauskonzerns erzielten 2022 folgende Umsätze (in Mio. Euro):

Filiale $i$	1	2	3	4	5	6	7	8	9	10	11	12
Umsatz $x_i$	5	2	7	6	6	3	3	13	4	2	9	18

- Berechnen Sie Median, das 1. und das 3. Quartil, sowie Spannweite und den Quartilsabstand und zeichnen Sie einen Boxplot.
- Berechnen Sie das arithmetische Mittel und die Standardabweichung.
- Beurteilen Sie anhand des Boxplot, ob die Verteilung linksschief/rechtsschief/symmetrisch ist.

9. **P 5** Ein Unternehmen führt mit 8 Bewerbern einen Eignungstest durch. Es ergab sich folgende Punktebewertung:

Erreichte Punkte	5	3	8	31	2	15	22	14	10	8	7
------------------	---	---	---	----	---	----	----	----	----	---	---

- Wie ist das Merkmal skaliert? Ist es diskret oder stetig?
- Bestimmen Sie die für einen Boxplot relevanten Größen und zeichnen Sie den Boxplot!
- Bestimmen Sie anhand der Daten alle drei Werte der zentralen Tendenz (Modalwert, Median, Mittelwert)! Welcher dieser Werte beschreibt die Lage der Daten am besten, und warum?
- Beurteilen Sie anhand des Boxplot, ob die Verteilung linksschief/rechtsschief/symmetrisch ist.

10. **P 6** Um die Auswirkung der Studieneingangsphase zu demonstrieren, wurde die Studiendauer (in Semestern) von 100 Studierenden des Bachelorstudiums BWL erhoben, die in den vergangenen vier Semestern ihr Studium erfolgreich abgeschlossen haben. Es ergaben sich folgende Daten:

Studiendauer (in Semestern)	6	7	8	9	10	11	12
Anzahl Studierende	12	24	36	11	7	6	4

Berechnen Sie Median, arithmetisches Mittel und die Standardabweichung des Merkmals „Studiendauer“.

11. In der nachstehenden Tabelle ist die Anzahl der Kleinstunternehmen (Umsatz zwischen 0 und 10 Mio. Euro), Kleinunternehmen (Umsatz zwischen 10 und 20 Mio. Euro) und mittleren Unternehmen (Umsatz zwischen 20 und 60 Mio. Euro) in einer bestimmten Region, angegeben:

Umsatz (in Mio. €) [von; bis[	[0; 10[	[10; 20[	[20; 60]
Anzahl der Unternehmen	40	80	80

Berechnen sie näherungsweise das arithmetische Mittel sowie die Standardabweichung des Merkmals Umsatz.

12. **P 7** Im Rahmen einer Marktforschungsstudie für einen Wintersportartikelhersteller wurden Skifahrer an einer Liftstation nach verschiedenen Merkmalen befragt:

### Skiregion Dachstein-Tauern

Liebe Skifahrerinnen, liebe Skifahrer! Bitte teilen Sie mit uns Ihre Erfahrungen während Ihres Aufenthaltes in unserem Skigebiet.

<b>F1</b>	Ihr Geschlecht?	männlich ..... <input type="radio"/> (1) weiblich ..... <input type="radio"/> (2) divers ..... <input type="radio"/> (3)	
<b>F2</b>	Wie alt sind Sie?		(Jahre)
<b>F3</b>	Wie groß sind Sie?		(cm)
<b>F4</b>	Wie viel wiegen Sie?		(kg)
<b>F5</b>	Welches Sportgerät benutzen Sie?	Ski ..... <input type="radio"/> (1) Snowboard ..... <input type="radio"/> (2)	
<b>F6</b>	Welche Pisten haben Sie vorwiegend benutzt?	Planai ..... <input type="radio"/> (1) Hauser-Kaibling ..... <input type="radio"/> (2) Hochwurzen ..... <input type="radio"/> (3) Ramsau-Dachstein ... <input type="radio"/> (4)	
<b>F7</b>	Wie hat Ihnen unsere Skiregion gefallen?	sehr gut ..... <input type="radio"/> (1) gut ..... <input type="radio"/> (2) mittel ..... <input type="radio"/> (3) weniger gut ..... <input type="radio"/> (4) sehr schlecht ..... <input type="radio"/> (5)	

Zu diesem Fragebogen steht Ihnen folgende Stichprobe zur Verfügung:

Kunden ID	F1	F2	F3	F4	F5	F6	F7
1	1	42	178	78	1	1	1
2	2	28	170	51	1	2	2
3	2	36	156	46	1	3	1
4	3	47	151	47	2	1	3
5	1	51	175	63	1	3	2
6	1	61	178	75	1	1	1
7	2	22	171	62	2	2	4
8	1	48	185	95	1	2	3
9	2	19	163	55	2	1	3
10	2	21	157	48	1	3	1

- a) Berechnen Sie für die Merkmale, die Sie von Ihrer LV\_Leitung genannt bekommen, jeweils ein geeignetes Lagemaß und begründen Sie Ihre Wahl.
- b) Berechnen Sie für das Merkmal F2 ein geeignetes Streuungsmaß und begründen Sie Ihre Wahl.

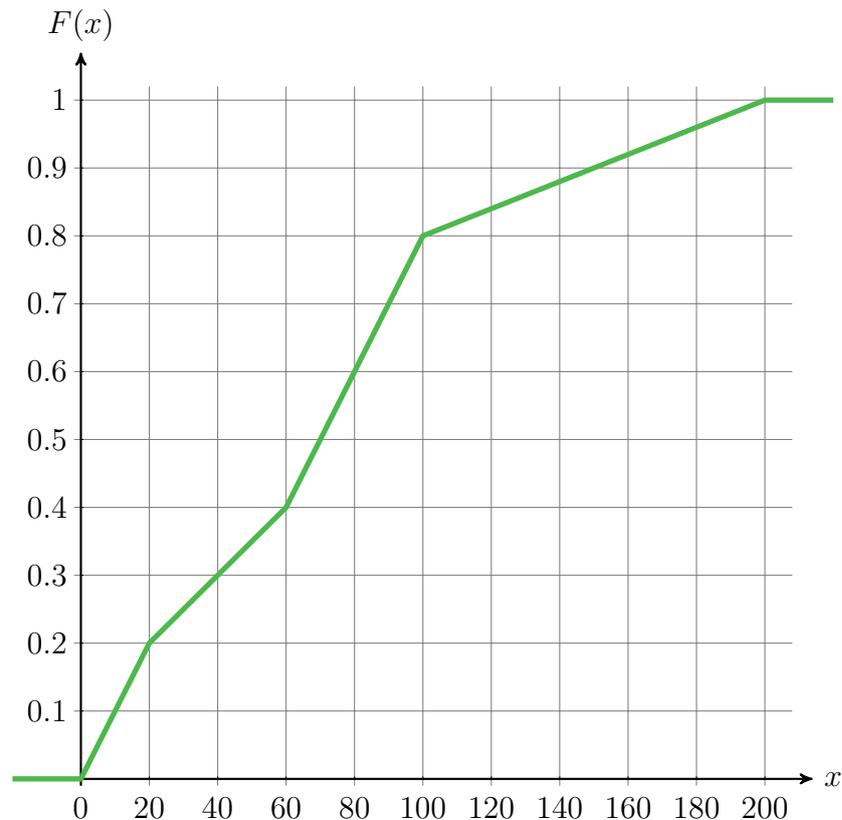
13. An zwei Standorten A und B eines Unternehmens werden Aggregate gleicher Sorte hergestellt. Zur Untersuchung der Qualität der Produktion wurden Lebensdaueranalysen mit je einer Stichprobe vom Umfang  $n_A = 5$  im Standort A und  $n_B = 6$  im Standort B gemacht. Es ergaben sich folgende Werte:

<b>A (Lebensdauer in Jahren)</b>	21	17	25	31	30	
----------------------------------	----	----	----	----	----	--

<b>B (Lebensdauer in Jahren)</b>	22	29	29	35	25	37
----------------------------------	----	----	----	----	----	----

- a) Bestimmen Sie Mittelwerte, Standardabweichungen und Variationskoeffizienten der Lebensdauer der Aggregate für die einzelnen Standorte. An welchem Standort war die Streuung der Lebensdauer höher?
- b) Bestimmen Sie unter Verwendung von a) Mittelwert und Standardabweichung der Lebensdauer für die Gesamtheit aller 11 Aggregate.

14. **P 8** Die folgende Grafik zeigt die Verteilung der Restlaufzeiten von 1000 Krediten einer Bank (in Monaten):



- a) Ergänzen Sie die nachstehende Tabelle

Restlaufzeit (in Monaten)	$[0; 20[$	$[20; 60[$	$[60; 100[$	$[100; 200]$
Anzahl der Kredite				

- b) Berechnen Sie näherungsweise die mittlere Restlaufzeit der Kredite.  
 c) Berechnen Sie näherungsweise die Standardabweichung der Restlaufzeit der Kredite.

15. **P 9** Ein Konzern ist weltweit tätig, er hat 14 Niederlassungen in Europa, 28 in Asien und sechs in den USA. Für die jährlichen Umsätze (in Mio. €) der Niederlassungen sind folgende Kennzahlen bekannt:

Region	Mittelwert der Umsätze (in Mio. €)	Standardabweichung der Umsätze
Europa	200	22
Asien	114	20
USA	240	25

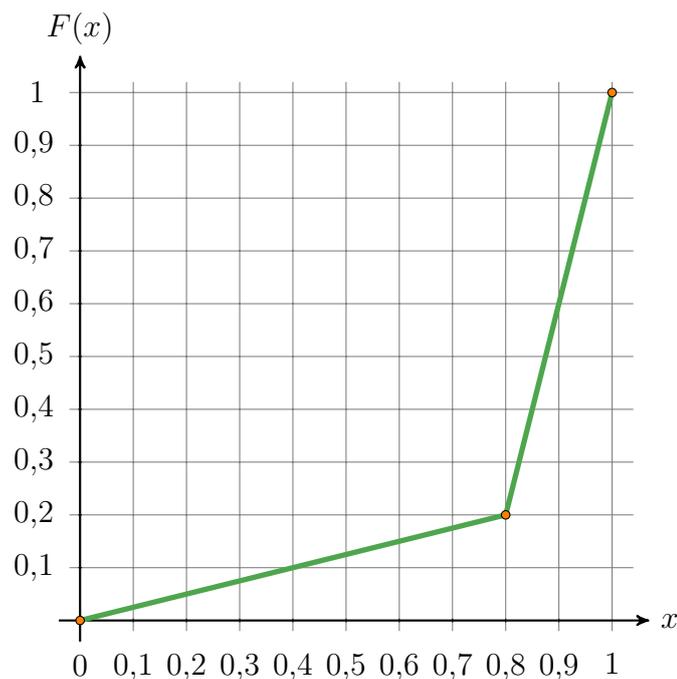
- a) An welchem Standort ist die Streuung der Umsätze höher?  
 b) Berechnen Sie den durchschnittlichen konzernweiten Umsatz sowie die Standardabweichung des Umsatzes für den gesamten Konzern.

16. In einem Land beträgt das gesamte Bruttovermögen (Sachvermögen und Finanzvermögen, ohne Schulden) 1.000 Mrd. Euro. Dieses verteilt sich auf vier Haushaltsgruppen („Untere Hälfte“, „Obere Mitte“, „Vermögende“ und „Top 5“). Die Gruppe „Untere Hälfte“ umfasst dabei 50%, „Obere Mitte“ 30%, „Vermögende“ 15% und „Top 5“ 5% aller Haushalte. Die Aufteilung des gesamten Bruttovermögens auf die vier Haushaltsgruppen ist in nachstehender Tabelle zusammengefasst:

Haushaltsgruppe	Untere Hälfte	Obere Mitte	Vermögende	Top 5
Bruttovermögen (in Mrd. Euro)	50	200	300	450

Es soll die Konzentration des Bruttovermögens untersucht werden.

- Zeichnen Sie die Lorenzkurve.
  - Bestimmen Sie den Gini-Koeffizienten für das Merkmal „Bruttovermögen“ und interpretieren Sie Ihr Ergebnis.
17. P 10 In einem Staat mit 10 Mio. Einwohnern gibt es nur zwei Haushaltsgruppen gemessen am Einkommen: arm und reich. Die Ungleichverteilung der Einkommen kann mit folgender Lorenz-Kurve dargestellt werden.



- Zeichnen Sie in das obige Diagramm ein, wie die Lorenz-Kurve bei perfekter Gleichverteilung der Einkommen aussehen müsste!
- Berechnen Sie den Gini-Koeffizienten für das Merkmal „Einkommen“ und interpretieren Sie ihr Ergebnis.
- Wie viele arme Personen (in absoluten Zahlen) leben in diesem Staat?
- Nehmen Sie an, dass es nun auch eine dritte Haushaltsgruppe gibt: eine Mittelschicht! Zeichnen Sie eine (beliebige) weitere Lorenz-Kurve in das Diagramm, in der die Einkommen weniger ungleich verteilt sind als jetzt, aber nicht gleichverteilt! Berücksichtigen Sie dabei die Existenz der Mittelschicht!

## Zusammenhangsmaße, Regressionsgerade

18. In einem Experiment zur Wirkung von Musik auf die Ausgabebereitschaft beim Einkauf in einem Supermarkt wurden 100 Kunden zufällig ausgewählt. Ein Teil der Kunden tätigte seine Einkäufe an Tagen, an denen im Supermarkt keine "Hintergrundmusik" eingespielt wurde. Der andere Teil an einem Tag, an dem der Einkauf durch Musik und Werbeanzeigen begleitet wurde. Jeder Kunde wurde hinsichtlich der Gesamtsumme der getätigten Einkäufe in eine der drei Gruppen der Ausgabebereitschaft (hoch, mittel und gering) eingeordnet. Es ergab sich folgende (bivariate) Verteilung:

	Ausgabebereitschaft		
Ambiente	gering	mittel	hoch
Mit Musik	8	16	21
Ohne Musik	24	20	11

Berechnen Sie eine geeignete Kennzahl, um eine Aussage über eine mögliche Abhängigkeit der beiden Merkmale treffen zu können!

19. **P 11** Die folgende Tabelle zeigt die Ergebnisse einer Befragung von 20 Studierenden einer Statistik-Übung zum Thema Umweltbewusstsein. Folgende Merkmale wurden erhoben: Selbst berichtetes Umweltbewusstsein (1: starkes Bewusstsein, 2: mittleres Bewusstsein, 3: geringes Bewusstsein), Studienrichtung (BWL, VWL, Soziologie) und Parteipräferenz (zuletzt gewählt bei der Nationalratswahl: ÖVP, SPÖ, FPÖ, Grüne):

Teilnahme-ID	Umweltbewusstsein	Studienrichtung	Parteipräferenz
1	1	Soziologie	Grüne
2	2	VWL	SPÖ
3	3	BWL	ÖVP
4	2	Soziologie	SPÖ
5	1	Soziologie	Grüne
6	3	BWL	ÖVP
7	3	VWL	FPÖ
8	1	BWL	Grüne
9	3	BWL	ÖVP
10	2	Soziologie	FPÖ
11	2	VWL	ÖVP
12	1	BWL	SPÖ
13	1	Soziologie	Grüne
14	2	BWL	Grüne
15	3	VWL	SPÖ
16	2	VWL	FPÖ
17	1	BWL	Grüne
18	2	Soziologie	SPÖ
19	1	VWL	ÖVP
20	1	BWL	FPÖ

- a) Welche Skalenniveaus haben hier die Variablen „Studienrichtung“ und „Umweltbewusstsein“?
- b) Erstellen Sie eine Kontingenztabelle für die beiden Merkmale „Studienrichtung“ und „Umweltbewusstsein“ und berechnen Sie eine geeignete Kennzahl, um eine Aussage über eine mögliche Abhängigkeit der beiden Merkmale treffen zu können. Interpretieren Sie Ihr Ergebnis!

20. Für acht Studierende einer Statistikübung wurde die Selbsteinschätzung ihrer Statistikkenntnisse (A: sehr gut, B: gut, C: teils gut/teils schlecht, D: schlecht, E: sehr schlecht) vor Beginn der Übung, sowie die erreichte Punktezahl am Ende der Übung erhoben. Gibt es einen Zusammenhang zwischen diesen beiden Merkmalen? Berechnen Sie eine geeignete Kennzahl.

Studierender	1	2	3	4	5	6	7	8
Statistikkenntnisse	D	B	C	D	E	A	C	D
erreichte Punkte	21	13	26	19	14	23	24	12

21. P 12 Ein Sportartikelhersteller führte einen Test an 10 Laufschuhen seiner aktuellen Kollektion hinsichtlich der Eigenschaften „Tragekomfort“ und „Wettkampfeignung“ durch. Dabei konnten die Noten 1: sehr gut, 2: gut, 3: befriedigend, 4: ausreichend, 5: mangelhaft vergeben werden. Folgende Daten wurden dabei erhoben:

Schuhmodell	A	B	C	D	E	F	G	H	I	J
Tragekomfort	1	3	4	5	2	1	2	4	2	2
Wettkampfeignung	3	3	1	1	4	5	2	2	3	4

Gibt es einen Zusammenhang zwischen diesen beiden Merkmalen? Berechnen Sie eine geeignete Kennzahl. Wie ist das Ergebnis zu interpretieren?

22. Die Marketingabteilung einer Handelskette möchte wissen, ob ihre Werbemaßnahmen wirken. Die Buchhaltung liefert Informationen über die monatlichen Umsätze. Die Umsätze von 6 aufeinanderfolgenden Monaten mit den entsprechenden Marketingausgaben liefern folgende Daten (Beträge in 1.000 Euro):

Monat	Jänner	Februar	März	April	Mai	Juni
Marketingausgaben (in Tausend €)	24	16	19	26	12	18
Umsatz (in Tausend €)	195	190	220	250	150	185

- Erstellen Sie ein Streudiagramm des Datensatzes.
- Bestimmen Sie den Pearson-Korrelationskoeffizienten.
- Geben Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit des Umsatzes von den Marketingausgaben an.
- Welchen Umsatz würde die Handelskette erzielen, wenn Marketingausgaben in Höhe von 10.000 € geplant werden?
- Um wie viel Prozent hat sich der Umsatz jeweils in den einzelnen Monaten verändert?
- Um wie viel Prozent hat sich der Umsatz in den Monaten Jänner bis Juni insgesamt verändert, um wie viel Prozent im Monatsdurchschnitt?

23. **P 13** Unter Leitzinsen versteht man die von der zuständigen Zentralbank festgelegten Zinssätze, zu denen sich Geschäftsbanken bei einer Zentral- oder Notenbank Geld beschaffen oder anlegen können. Die folgende Tabelle gibt die Entwicklung des Zinssatzes für das Hauptrefinanzierungsgeschäft (häufig als „der“ Leitzins bezeichnet) und die der Zinsen für einen privaten Konsumkredit (mit anfänglicher Zinsbindung, Laufzeit 1 bis 5 Jahre) in Österreich für 11 Zeitpunkte wieder (Werte in % p. a.):

Zeitpunkt	Jän. 22	Jul. 22	Sep. 22	Nov. 22	Dez. 22	Feb. 23	Mär. 23	Mai. 23	Jun. 23	Aug. 23	Sep. 23
Leitzins	0	0,50	1,25	2	2,50	3	3,50	3,75	4	4,25	4,5
Kreditzins	4,28	5,06	5,18	5,34	5,68	6,08	6,17	6,29	6,77	7,42	7,30

- Erstellen Sie ein Streudiagramm des Datensatzes.
- Bestimmen Sie die Koeffizienten der Regressionsgeraden zur Beschreibung der Abhängigkeit des Kreditzinses vom Leitzins.
- Beurteilen Sie mithilfe des Korrelationskoeffizienten, ob die Regressionsgerade ein geeignetes Modell darstellt, um diesen Zusammenhang zu beschreiben.
- Welchen Kreditzins prognostiziert das Modell für einen Leitzins von 2,25 % p. a.?

Die mit **P** gekennzeichneten Beispiele sind von den Studierenden vorzubereiten und nach Aufruf durch die Lehrveranstaltungsleitung zu präsentieren!