

Statistik – Übungen WS 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 132 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:

Merkmal	Ausprägung	Skalenniveau	diskret/stetig
Inflationsrate			
Reinheit von Diamanten			
Zeitdauer eines Fluges			
Automarke			
Betriebssystem eines Computers			
Hotelklassifikation			
Blutdruck (in mmHg)			
durchschnittliche tägliche Nutzungsdauer sozialer Medien			

Verteilung eines Merkmals, Graphische Darstellungen

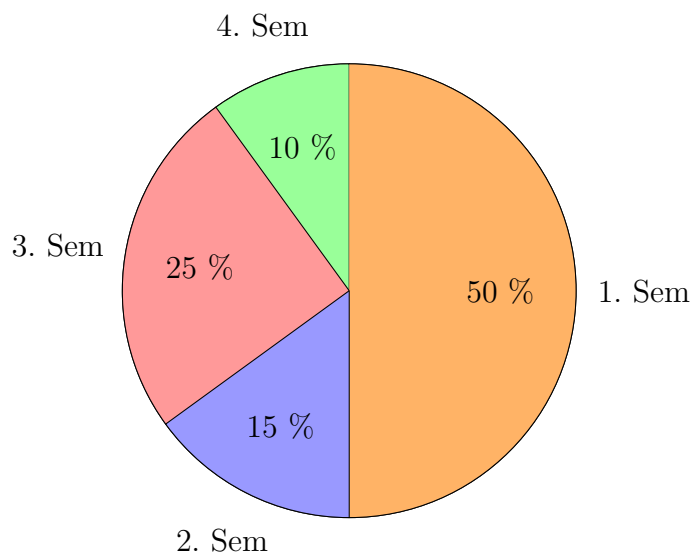
3. In einer Umfrage wurden 60 Studierende der TU-Graz nach der Anzahl Ihrer Prüfungsantritte in Mathematik I gefragt. Die Antworten finden Sie in nachfolgender Tabelle zusammengefasst:

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 einer Statistik Übung wurde erhoben, in welchem Semester sich die Studierenden zum Erhebungszeitpunkt befanden. Das nachstehende Kreisdiagramm zeigt die relativen Häufigkeiten für die 60 Teilnehmenden einer Übungsgruppe:



- 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 $F(3)$!

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:

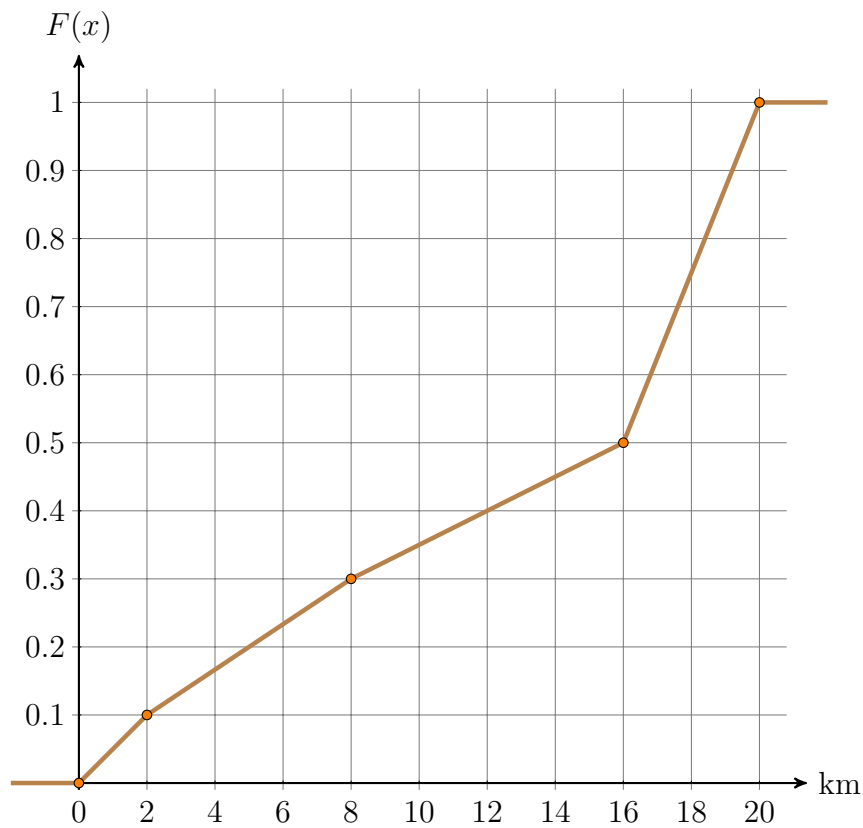
Note	5	4	3	2	1
erreichte Punkte	[0; 50]]50; 60]]60; 70]]70; 80]]80; 100]
absolute Häufigkeit H_i	25	10	5	5	5

- a) Ergänzen Sie in der Häufigkeitstabelle die relativen Häufigkeiten sowie die absolute und die relative Besetzungsdichte.
- b) Stellen Sie die Information auf geeignete Weise graphisch dar. Beachten Sie dabei insbesondere die unterschiedliche Klassenbreite des klassierten Merkmals „erreichte Punkte“.
- c) Zeichnen Sie die geeignete Verteilungsfunktion!
6. **P 3** 50 KFZ-Besitzer wurden nach der von ihrem privaten KFZ im Jahr 2023 zurückgelegten Kilometerzahl befragt. Man erhielt die folgenden Daten (in Tausend):

2,4	3,9	4,2	4,5	4,7	5,2	6,3	6,7	6,9	7,3
10,2	11,4	12,3	12,4	13,2	14,3	15,7	15,9	16,1	16,3
17,2	18,5	18,8	19,2	19,4	20,2	20,4	20,6	20,7	21,1
22,4	24,3	24,7	25,1	27,3	27,5	27,8	28,1	28,4	28,5
31,3	31,8	32,6	35,2	40,1	40,7	41,2	43,2	48,9	59,7

- a) Stellen Sie die Daten durch ein geeignetes Diagramm dar.
- b) Zeichnen Sie eine geeignete Verteilungsfunktion und beantworten Sie mit deren Hilfe folgende Fragen:
- wie groß ist näherungsweise der Anteil der KFZ-Besitzer mit einer zurückgelegten Kilometerzahl von weniger als 25.000?
 - welche zurückgelegte Kilometerzahl wird von 50 % der befragten KFZ-Besitzer überschritten?

7. **P 4** Im Zuge der Planung eines neuen Einkaufszentrums wurden für den in Frage kommenden Standort die Anreisewege potentieller Kunden ermittelt. Es ergab sich folgende Verteilung:



- Geben Sie tabellarisch die relativen und kumulierten relativen Häufigkeiten an.
- Stellen Sie die Daten durch ein Histogramm dar.
- Wie viel Prozent der potentiellen Kunden wohnen weniger als 8 km vom geplanten Standort entfernt?

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** Von der Verteilung der Bruttogehälter der 1.000 Beschäftigten eines Unternehmens sind die folgenden Daten bekannt:

$$\text{Minimalwert } x_{min} = \text{€ } 450.-$$

$$\text{unteres Quartil } x_{0,25} = \text{€ } 1500.-$$

$$\text{Median } x_{0,5} = \text{€ } 2.800.-$$

$$\text{oberes Quartil } x_{0,75} = \text{€ } 4500.-$$

Die beiden höchsten Bruttogehälter betragen: € 8.700.- bzw. € 13.000.-.

- Zeichnen Sie den Boxplot dieser Verteilung!
- Beurteilen Sie anhand des Boxplot, ob die Verteilung linksschief/rechtsschief/symmetrisch ist.

10. **P 6** In der folgenden Tabelle ist die Zahl der Kinder von 100 Haushalten zusammengefasst:

Zahl der Kinder	0	1	2	3	4	5
Haushalte	40	10	10	10	10	20

Berechnen Sie

- die durchschnittliche Kinderzahl pro Haushalt
- den Median sowie
- die Standardabweichung des Merkmals „Zahl der Kinder pro Haushalt“.

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 Führen Sie bitte eine Umfrage an 10 zufällig ausgewählten Personen durch. Verwenden Sie dabei unten stehenden Fragebogen und tragen Sie Ihre Ergebnisse in der nachfolgende Tabelle sowie in Excel ein:

ID	F1	F2	F3	F4	F5	F6	F7
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

- 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 F4 ein geeignetes Streuungsmaß und begründen Sie Ihre Wahl.

Omnivor (Ich konsumiere sowohl Pflanzen als auch Tiere.)

Vegetarisch (Ich meide Fleisch und Fisch, konsumiere jedoch Milchprodukte und Eier.)

Vegan (Ich meide alle tierischen Produkte.)

Pescetarisch (Meine Ernährung ist hauptsächlich pflanzlich, inkludiert aber Fisch.)

Flexitarisch (Ich esse meist pflanzliche Kost, gelegentlich jedoch auch Fleisch/Fisch.)

Fragebogen

Sehr geehrte Teilnehmerinnen und Teilnehmer, wir möchten Sie herzlich einladen, an unserer Umfrage teilzunehmen.

Zweck dieser Umfrage ist es, ein besseres Verständnis dafür zu entwickeln, wie sozioökonomische Faktoren, Lebensstile und individuelle Gesundheitspraktiken unser Wohlbefinden beeinflussen. Ihre Teilnahme liefert wertvolle Einblicke in Zusammenhänge zwischen dem sozialen Umfeld, der persönlichen Lebensweise und der Gesundheit.

Alle Ihre Angaben werden streng vertraulich behandelt werden. Die Ergebnisse dieser Befragung werden anonymisiert und dienen ausschließlich wissenschaftlichen Zwecken.

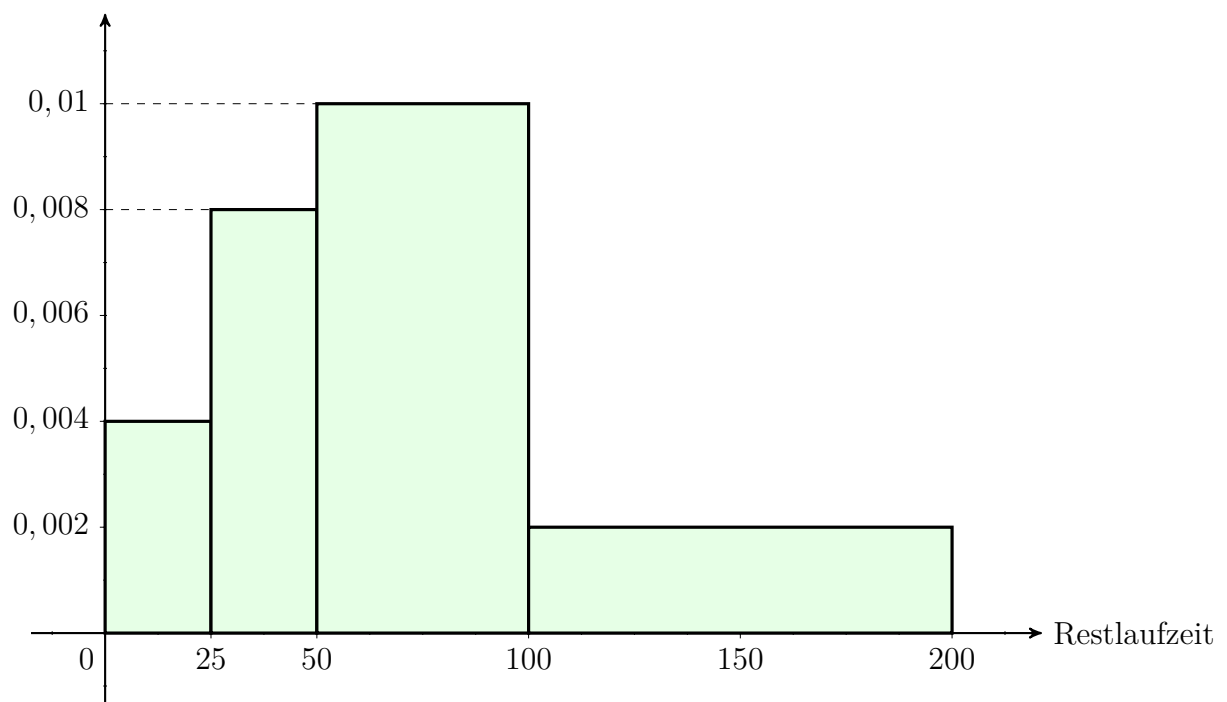
F1	Welches Geschlecht haben Sie?	männlich	<input type="radio"/>	(1)
		weiblich	<input type="radio"/>	(2)
		divers	<input type="radio"/>	(3)
F2	Wie alt sind Sie?			(Jahre)
F3	Wie groß sind Sie?			(cm)
F4	Wie viele Stunden nutzen Sie soziale Medien pro Tag?			(h)
F5	Wie würden Sie Ihre Ernährungsgewohnheiten beschreiben? Wählen Sie bitte die Option, die am besten zutrifft.	Omnivor	<input type="radio"/>	(1)
		Vegetarisch	<input type="radio"/>	(2)
		Vegan	<input type="radio"/>	(3)
		Pescetarisch	<input type="radio"/>	(4)
		Flexitarisch	<input type="radio"/>	(5)
F6	Wie oft betreiben Sie pro Woche Sport?	Weniger als 1 Mal	<input type="radio"/>	(1)
		1-2 Mal	<input type="radio"/>	(2)
		3-4 Mal	<input type="radio"/>	(3)
		5 Mal oder mehr	<input type="radio"/>	(4)
F7	Wie beurteilen Sie Ihren derzeitigen Gesundheitszustand?	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)

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:

A (Lebensdauer in Jahren)	21	17	25	31	30	
----------------------------------	----	----	----	----	----	--

B (Lebensdauer in Jahren)	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; 25[[25; 50[[50; 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** In einer Umfrage wurden die durchschnittlichen monatlichen Ausgaben für Nachhilfe in den Bundesländern Wien, Salzburg und Steiermark für das Jahr 2023 erhoben. Man erhielt folgende Daten:

Bundesland	Mittelwert der monatlichen Ausgaben	Standardabweichung der monatlichen Ausgaben	Anzahl der Befragten
Wien	120	30	500
Salzburg	100	25	600
Steiermark	90	15	450

- a) In welchem Bundesland ist die Streuung der Ausgaben am höchsten?
 b) Berechnen Sie den Gesamtdurchschnitt sowie die Gesamtstandardabweichung der monatlichen Ausgaben für Nachhilfe für diese drei Bundesländer.
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.

- a) Zeichnen Sie die Lorenzkurve.
 b) Bestimmen Sie den Gini-Koeffizienten für das Merkmal „Bruttovermögen“ und interpretieren Sie Ihr Ergebnis.
17. **P 10** Angenommen, in einer bestimmten Gesellschaft werden die Gesamtkonsumausgaben von Haushalten wie folgt dargestellt:
 Die unteren 20 % der Haushalte haben einen Anteil von 10 %,
 Die unteren 40 % der Haushalte haben einen Anteil von 25 %,
 Die unteren 60 % der Haushalte haben einen Anteil von 45 %,
 Die unteren 80 % der Haushalte haben einen Anteil von 70 % an den Gesamtkonsumausgaben.
- a) Zeichnen Sie die Lorenzkurve für die Verteilung der Konsumausgaben in dieser Gesellschaft.
 b) Berechnen Sie den Gini-Koeffizienten für die Verteilung der Konsumausgaben und interpretieren Sie Ihr Ergebnis.

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 kaufte an einem Tag, an dem der Einkauf durch Musik und Werbeansagen begleitet wurde, ein. 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 „Parteipräferenz“?
- b) Erstellen Sie eine Kontingenztabelle für die beiden Merkmale „Studienrichtung“ und „Parteipräferenz“ 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 Es soll untersucht werden, ob es einen Zusammenhang zwischen der Einschätzung des derzeitigen Gesundheitszustands und der wöchentlichen Sportaktivität gibt. Verwenden Sie die Daten Ihrer Umfrage aus Beispiel 12.

ID	1	2	3	4	5	6	7	8	9	10
Einschätzung Gesundheitszustand										
wöchentliche Sportaktivität										

- a) Berechnen Sie eine geeignete Kennzahl, um den Zusammenhang zwischen der Einschätzung des derzeitigen Gesundheitszustands und der wöchentlichen Sportaktivität zu beschreiben.
- b) Interpretieren Sie die Kennzahl im Kontext der Untersuchung. Gibt es Hinweise auf einen Zusammenhang zwischen der Einschätzung des derzeitigen Gesundheitszustands und der wöchentlichen Sportaktivität?
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

- a) Erstellen Sie ein Streudiagramm des Datensatzes.
- b) Bestimmen Sie den Pearson-Korrelationskoeffizienten.
- c) Geben Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit des Umsatzes von den Marketingausgaben an.
- d) Welchen Umsatz würde die Handelskette erzielen, wenn Marketingausgaben in Höhe von 10.000 € geplant werden?
- e) Um wie viel Prozent hat sich der Umsatz jeweils in den einzelnen Monaten verändert?
- f) 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** Die folgende Tabelle gibt die Steuereinnahmen sowie die Neuverschuldung der Republik Österreich für die Jahre 2015 bis 2023 (in Milliarden Euro) an:

Jahr	2015	2016	2017	2018	2019	2020	2021	2022	2023
Steuereinnahmen	98,9	97,7	100,9	106,4	110,4	101,8	114,0	127,2	132,9
Neuverschuldung	12,3	3,9	-6,2	-4,6	-4,9	35,4	18,4	16,4	20,4

Untersuchen Sie die Vermutung, dass höhere Steuereinnahmen tendenziell zu einer geringeren Neuverschuldung führen können (da mehr Einnahmen vorhanden sind, um Ausgaben zu decken), während niedrigere Steuereinnahmen oft eine höhere Neuverschuldung notwendig machen, damit der Staat seine finanziellen Verpflichtungen erfüllen kann.

- Erstellen Sie ein Streudiagramm des Datensatzes.
- Bestimmen Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit der Neuverschuldung von den Steuereinnahmen.
- Beurteilen Sie mithilfe des Korrelationskoeffizienten, ob die Regressionsgerade ein geeignetes Modell darstellt, um diesen Zusammenhang zu beschreiben.
- Welche Neuverschuldung prognostiziert das Modell für den Fall, dass die Steuereinnahmen 2024 auf 140 Mrd. € ansteigen?

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