Anlagen und Übungen zu Statistik:

Grundbegriffe und Definitionen

Man unterscheidet generell zwei Typen von Variablen:

a) diskrete Variablen:

Sie können nur endlich viele, streng voneinander getrennte Werte als Ausprägungen annehmen (Beispiel: Kinderzahl in Familien mit den Ausprägungen 0, 1, 2, usw.).

b) stetige Variablen:

Sie können innerhalb ihres Definitionsbereichs jeden beliebigen Wert (und Zwischenwert) annehmen (Beispiel: Körpergröße befragter Personen).

Beispiele:

Körpergröße (von Personen oder Tieren)

- Monatseinkommen (von Personen oder Haushalten)
- Kinderzahl (von Personen oder Familien)
- Reaktionszeit (von Personen oder Versuchsanordnungen)
- Bruchfestigkeit (von produzierten Werkstücken)
- Bevorzugte politische Partei (von Personen)
- Geschlecht (von Personen oder Tieren)
- Gattung (von Tieren oder Pflanzen)
- Sozialprodukt (von Regionen, Ländern oder Staaten)

Aufgabe 1:
Nennen Sie Variablen, die dem Merkmalsträger „Unternehmen" zugeordnet werden können!

Aufgabe 2:
Welche der folgenden Variablen sind vom diskreten, welche vom stetigen Typ?

- Geschlecht - Familienstand - Körpergewicht - gelaufene Zeiten im 100m-Lauf - Lebensdauer - Haarfarbe - Kinderzahl - durchschnittliche Kinderzahl - Einkommen - Kalorienverbrauch - Intelligenz

Aufgabe 3:
Welches sind die jeweiligen Merkmalsträger bei den in Aufgabe 2 genannten Variablen?

Aufgabe 4:

In allen deutschen Städten werden jeweils 100 Familien danach befragt, wie viele Kinder sie haben. Aus diesen Angaben wird dann in jeder der Städte eine durchschnittliche Kinderzahl berechnet.
Von welchem Typ ist diese Variable „durchschnittliche Kinderzahl in jeweils 100 Familien"?

Die Präsentation statistischer Daten

Aufgabe 1:

Gegeben sind die folgenden Angaben zur Kinderzahl in zufällig ausgewählten Haushalten:

1,2,0,0,1,1,2,5,0,3,0,0,2,1,1,2,3,0,0,1,2,4,0,1,0,1,

1,2,3,0,0,1,2,4,0,1,2,0,0,1,1,2,5,0,3,0,0,2,1,0,0,3

a) Von welchem Typ ist diese Untersuchungsvariable?
b) Welches ist das Skalenniveau dieser Variablen?
c) Erzeugen Sie eine Häufigkeitsverteilung!
d) Berechnen Sie die relativen Häufigkeiten in der Häufigkeitsverteilung!
e) Kumulieren Sie die relativen Häufigkeiten aufwärts!
f) Wie können aufwärts kumulierte Häufigkeiten interpretiert werden?

Aufgabe 2:

Gegeben sind die folgenden Angaben zum Körpergewicht zufällig ausgewählter Personen:

76,68,92,69,55,91,88,77,68,71,55,91,88,77,68,71

88,77,68,71,76,68,92,69,55,91,88,55,91,88,77,73

a) Von welchem Typ ist diese Untersuchungsvariable?
b) Welches ist das Skalenniveau?
c) Erzeugen Sie eine Häufigkeitsverteilung, indem Sie Klassen zu je 5 kg bilden!
d) Berechnen Sie die relativen Häufigkeiten!
e) Kumulieren Sie die relativen Häufigkeiten!
f) Erzeugen Sie eine Häufigkeitsverteilung, indem Sie Klassen zu je 10 kg bilden!
g) Berechnen Sie die relativen Häufigkeiten!
h) Kumulieren Sie die relativen Häufigkeiten!

Tabellarische Darstellung bivariater Daten

Aufgabe 1:

Gegeben sind die folgenden Angaben zu Körpergröße (in cm) und Körpergewicht (in kg) zufällig ausgewählter erwachsener Männer:

Größe	Gewicht	Größe	Gewicht	Größe	Gewicht
178	78	174	75	192	95
172	75	172	75	182	83
191	99	182	88	171	74
188	85	177	80	176	75
169	71	159	62	178	72
173	70	167	65	175	69
181	83	181	92	190	87
175	70	154	55	188	92
177	80	173	69	183	95
183	91	178	66	174	76

Erzeugen Sie aus diesen Angaben eine Kreuztabelle, indem Sie sowohl bei den Gewichts- als auch bei den Größenangaben Klassen der Breite 10 (kg bzw. cm) bilden!

Aufgabe 2:

Bestimmen Sie die drei Gruppen von Prozentwerten (relative Häufigkeiten), die oben genannt wurden!

Graphische Darstellung univariater Daten

Aufgabe 1:

Gegeben ist die folgende Häufigkeitsverteilung über die Fächerverteilung zufällig ausgewählter Schüler-LKs:

Fach	*Anzahl*
BWL	55
Mathematik	38
Deutsch	45
Englisch	22
IV	10

Stellen Sie diese Angaben in geeigneter Form graphisch dar!

Aufgabe 2:

Gegeben ist die folgende Häufigkeitsverteilung über die Lebensdauer untersuchter Glühbirnen (in Stunden):

Lebensdauer	Anzahl
5000 bis unter 6000	22
6000 bis unter 7000	37
7000 bis unter 8000	55
8000 bis unter 9000	25
9000 bis unter 10000	12
10000 bis unter 11000	8
11000 bis unter 12000	4

Stellen Sie diese Angaben in geeigneter Form graphisch dar!

Mittelwerte

Aufgabe 1: Warum ist $\sum (x_i + y_i) = \sum x_i + \sum y_i$?

Aufgabe 2: Warum ist $\sum (c \cdot x_i) = c \cdot \sum x_i$?

Aufgabe 3: Warum ist $\sum_{i=1}^{n} c = n \cdot c$?

Aufgabe 4:

Bilden Sie das arithmetische Mittel aus den folgenden Angaben zum Körpergewicht zufällig ausgewählter Personen (in kg):

67,5 88,2 73,4 76,0 75,1 82,7 88,3 101,7 55,0 66,2

Aufgabe 5:

Bei der Auszählung von Fehlern in Mathematikarbeiten der vergangenen Jahrgänge ergab sich die folgende Häufigkeitsverteilung:

Fehlerzahl	Häufigkeit	Fehlerzahl	Häufigkeit
0	23	4	22
1	45	5	13
2	67	6	2
3	38		

a) Stellen Sie diese Häufigkeitsverteilung graphisch dar!

b) Berechnen Sie das arithmetische Mittel dieser Verteilung!

c) Bestimmen Sie Modus und Median der Verteilung.

Aufgabe 6:

Bei der Befragung zufällig ausgewählter Schüler ergab sich die folgende Verteilung der monatlich (netto) zur Verfügung stehenden Geldbeträge.

Einkünfte	Häufigkeit
200 bis unter 400	7
400 bis unter 600	12
600 bis unter 800	19
800 bis unter 1000	14
1000 bis unter 1250	10
1250 bis unter 1500	8
1500 bis unter 2000	4

a) Wie hoch ist der durchschnittliche Monatsbetrag?

b) Stellen Sie diese Häufigkeitsverteilung graphisch dar!

c) Bestimmen Sie Modus und Median der Verteilung.

Aufgabe 7:

Beweisen Sie: Die Summe der Abweichungen aller Merkmalswerte vom arithmetischen Mittel ist immer null.

Aufgabe 8:

Gegeben sind die folgenden Angaben zu den Kinderzahlen von 40 zufällig befragten Familien:

1,1,3,2,0,0,0,1,1,0, 0,3,2,1,4,0,0,1,4,0, 1,3,2,1,0,0,1,1,3,2, 0,0,0,4,2,5,1,3,2,1

a) Erzeugen Sie eine Verteilung absoluter Häufigkeiten und stellen Sie diese Verteilung tabellarisch dar!
b) Bestimmen Sie auch die Verteilung der relativen Häufigkeiten!
c) Stellen Sie die gegebene Häufigkeitsverteilung in geeigneter Form graphisch dar!
d) Berechnen Sie das arithmetische Mittel der Verteilung!
e) Welches ist der häufigste Wert (Modus) der Verteilung?

Aufgabe 9:

Gegeben ist die folgende Verteilung über die zufällig ausgewählten Ludwigshafener Schüler am WG pro Monat zur Verfügung stehenden Geldbeträge:

Einkünfte	Häufigkeit
400 bis unter 600	8
600 bis unter 800	14
800 bis unter 1000	20
1000 bis unter 1200	15
1200 bis unter 1500	12
1500 bis unter 2000	6

a) Bestimmen Sie die relative Häufigkeitsverteilung!
b) Kumulieren Sie die relativen Häufigkeiten!
c) Berechnen Sie das arithmetische Mittel der Verteilung!
d) Stellen Sie die gegebene Häufigkeitsverteilung in geeigneter Form graphisch dar!
e) Stellen Sie in einem neuen Achsenkreuz auch die Verteilung kumulierter relativer Häufigkeiten graphisch dar!
f) Versuchen Sie, ausgehend von der zuletzt erzeugten Graphik, zu bestimmen, welches der Geldbetrag ist von dem gesagt werden kann, dass die Hälfte der Befragten weniger, die andere Hälfte aber mehr pro Monat zur Verfügung hat!

Aufgabe 10:

Gegeben sind die folgenden Angaben zur Verteilung der Nettoeinkommen in der Bundesrepublik Deutschland (2006 in %; die letzte Einkommensklasse wurde künstlich geschlossen):

Einkommen (€)	Anteil
unter 600	10,5
600 bis unter 1000	9,0
1000 bis unter 1200	6,1
1200 bis unter 1400	7,0
1400 bis unter 1800	19,4
1800 bis unter 2200	19,6
2200 bis unter 3000	15,0
3000 bis unter 4000	7,4
4000 bis unter 6000	5,8

a) Berechnen Sie das arithmetische Mittel der Verteilung!
b) Kumulieren Sie die relativen Häufigkeiten in einer Graphik!
c) Bestimmen Sie den Median der Verteilung!
d) Bestimmen Sie auch den Modus der Verteilung!
e) Warum fallen die drei Mittelwerte auseinander?
f) Stellen Sie die gegebene Häufigkeitsverteilung in geeigneter Form graphisch dar; beachten Sie dabei das Prinzip der Flächentreue!

Aufgabe 11:

Bestimmen Sie - ausgehend von den obigen Daten und den Ergebnissen - denjenigen Merkmalswert,
der die geordnete Reihe der Einkommenswerte im Verhältnis 25:75 teilt
und auch denjenigen, der im Verhältnis 75:25 teilt.
Versuchen Sie, die Differenz zwischen diesen beiden Werten inhaltlich zu interpretieren.

Streuung und Konzentration

Aufgabe 1:

Gegeben ist die folgende Häufigkeitsverteilung über die Anzahl von Kindern in befragten Familien:

Kinder	Häufigkeit	Kinder	Häufigkeit
0	27	3	12
1	53	4	7
2	38	5	3

a) Berechnen Sie die durchschnittliche Kinderzahl!

b) Berechnen Sie die Streuung dieser Verteilung unter Nutzung der Standardabweichung!

c) Dividieren Sie die Standardabweichung durch das arithmetische Mittel und interpretieren Sie das Ergebnis!

Aufgabe 2:

Gegeben ist die folgende Häufigkeitsverteilung über die Anzahl der Fernsehgeräte in zufällig ausgewählten Haushalten:

Anzahl der Geräte	Häufigkeit	Anzahl der Geräte	Häufigkeit
0	125	3	65
1	475	4	25
2	185		

a) Stellen Sie diese Verteilung in geeigneter Form graphisch dar!

b) Berechnen Sie das arithmetische Mittel der Verteilung!

c) Bestimmen Sie, ohne zu rechnen und ohne eine Graphik zu Hilfe zu nehmen, den Median der Verteilung!

d) Berechnen Sie die Standardabweichung und den Variationskoeffizienten dieser Verteilung!

Aufgabe 3:

Gegeben ist der folgende Datenbestand über Körpergrößen:

165 172 180 188 195

a) Berechnen Sie die mittlere absolute Abweichung der Merkmalswerte von ihrem eigenen arithmetischen Mittel (sog. mittlere lineare Abweichung) und die Standardabweichung.

b) Führen Sie die gleichen Berechnungen für den folgenden Datenbestand durch und vergleichen Sie diese Ergebnisse mit denen aus Aufgabe a)!

175 178 180 182 185

Datenbestand:

Bei der Befragung zufällig ausgewählter Schüler zu ihren monatlich zur Verfügung stehenden Einkommen, ergab sich die folgende Verteilung:

Einkommen	Häufigkeit
400 bis unter 600	25
600 bis unter 800	48
800 bis unter 1000	72
1000 bis unter 1200	45
1200 bis unter 1400	20
1400 bis unter 1600	10

Aufgabe 1:
Stellen Sie die Häufigkeitsverteilung in geeigneter Form graphisch dar!

Aufgabe 2:
Bestimmen Sie die aufwärtskumulierten relativen Häufigkeiten; stellen Sie auch diese graphisch dar!

Aufgabe 3:
Berechnen Sie das arithmetische Mittel und die Standardabweichung dieser Verteilung!

Aufgabe 4:
Welchen Wert hat der Median?

Aufgabe 5:
Zeichnen Sie die Lorenzkurve zu dieser Einkommensverteilung und bestimmen Sie als Maß der Konzentration den Lorenzkoeffizienten!

Aufgabe 6:

In einer Wirtschaftsbranche ergab sich die folgende Verteilung von Umsatzangaben:

Umsatz (Mio.)	Zahl der Unternehmen
unter 1	17
1 bis unter 1,5	25
1,5 bis unter 2	38
2 bis unter 3	29
3 bis unter 4	18
4 bis unter 5	15
5 bis unter 6	12
6 bis unter 8	8
8 bis unter 10	4

Werten Sie diesen Datenbestand so aus, wie es in den Aufgaben 1 bis 5 für den ersten Datenbestand verlangt wurde!

Aufgabe 7:

Gegeben ist die folgende Verteilung über die Einkommen zufällig ausgewählter Personen:

Einkommensklasse	Häufigkeit
unter 1000	12
1000 bis unter 1500	14
1500 bis unter 2000	18
2000 bis unter 3000	20
3000 bis unter 4000	16
4000 bis unter 5000	12
5000 bis unter 7500	8

a) Stellen Sie diese Verteilung als Histogramm dar!

b) Berechnen Sie das arithmetische Mittel und die Standardabweichung der Verteilung!

c) Zeichnen Sie die Lorenzkurve!

Regressionsrechnung

Die folgenden Aufgaben beziehen sich auf den nachstehenden Datenbestand:

Bei 10 zufällig ausgewählten Personen wurden die beiden Variablen

X: Körpergewicht und Y: Oberer Blutdruckwert erfasst, wobei sich ergab:

Gewicht	Blutdruck	Gewicht	Blutdruck
75	125	91	136
69	128	66	122
85	131	70	130
81	135	75	132
76	130	88	140

Aufgabe 1:
Stellen Sie diese Angaben in geeigneter Form graphisch dar!

Aufgabe 2:
Bestimmen Sie eine lineare Regressionsfunktion mit der Methode der kleinsten Quadrate und zeichnen Sie diese in die Graphik ein!

Aufgabe 3:
Berechnen Sie, welche Blutdruckwerte zu erwarten wären bei einem Körpergewicht von

a)	50 kg	b)	60 kg	c)	70 kg
d)	80 kg	e)	90 kg	f)	100 kg

Aufgabe 4:

Erledigen Sie für den folgenden Datenbestand die gleichen Aufgaben wie sie in den Aufgaben 1 und 2 beim ersten Datenbestand verlangt wurden:

Erfasst wurden bei zufällig ausgewählten Haushalten die monatlich zur Verfügung stehenden Nettoeinkommen einerseits und die monatlichen Ausgaben für Konsumgüter andererseits. Dabei ergab sich:

Einkommen	Ausgaben	Einkommen	Ausgaben
2650	1220	3900	2200
4000	2150	2800	1560
3280	1900		

Aufgabe 5:

Gegeben ist der folgende Datenbestand über schulische Leistungen:

Anzahl von Fehlern in zwei Klassenarbeiten (Mathematik und Deutsch):

Mathematik:	3	4	0	2	6	7	4	4	2	1
Deutsch	0	0	5	2	1	2	3	3	5	7

a) Stellen Sie dies bivariate Häufigkeitsverteilung in geeigneter Form graphisch dar, wobei die Mathematikfehler als X-Variable verwendet werden soll (welche Hypothese impliziert dies?)!

b) Bestimmen Sie mit der Methode der kleinsten Quadrate eine lineare Regressionsfunktion und zeichnen Sie diese in die Graphik ein!

c) Ermitteln Sie mit dieser Regressionsfunktion, welche Anzahl von Fehlern in einer Deutscharbeit bei einem Schüler zu erwarten ist, der in der Mathematikarbeit 3 Fehler macht!

d) Ermitteln Sie alle Erwartungswerte der Variablen Y!

Ordinaldaten (Rangkorrelationskoeffizient)

Aufgabe 1:

Bei einer Befragung von Schülern nach ihren Zensuren in den Fächern Deutsch und

Mathematik ergaben sich die folgenden Werte:

Mathematik	2	3	2	5	4	3	3	5	2
Deutsch	4	3	5	2	3	3	2	1	4

Bestimmen Sie die Stärke des Zusammenhangs zwischen beiden Untersuchungsvariablen mit einer geeigneten Maßzahl.

Metrische Daten (Korrelationskoeffizient)

Verwenden Sie für die folgenden Aufgaben 1 bis 10 den folgenden Datenbestand:

Gegeben seien Daten über die Leistungen von 4 zufällig ausgewählten Schülern:
X = Anzahl von Fehlern in einem Deutschdiktat
Y = Anzahl von Fehlern in einer Mathematikarbeit

Schüler	X	Y	Schüler	X	Y
Alfred	2	4	Claude	6	2
Bernd	0	5	Dagobert	5	3

Aufgabe 1:
Berechnen Sie den Korrelationskoeffizient von Bravais/Pearson!

Aufgabe 2:
Berechnen Sie die Steigung der Regressionsgeraden!

Aufgabe 3:
Bestimmen Sie auch den Ordinatenabschnitt der linearen Regressionsfunktion
(X = Leistung in Deutsch; Y = Leistung in Mathematik)!

Aufgabe 4:
Berechnen Sie die Varianz der Y-Werte!

Aufgabe 5:
Bestimmen Sie die bei den gegebenen X-Werten theoretisch zu erwartenden Y-Werte (y_i auf der Regressionsgeraden);

Aufgabe 6:
Berechnen Sie auch die Varianz dieser y_i -Werte!

Aufgabe 7:
Bestimmen Sie die „Reste" zwischen den beobachteten Y-Werten und den theoretisch zu erwartenden Y-Werten y_i!

Aufgabe 8:
Berechnen Sie schließlich auch die Varianz dieser „Reste"!

Aufgabe 9:
Berechnen Sie den Determinationskoeffizienten!

Aufgabe 10:
Vergleichen Sie die drei berechneten Varianzen mit diesem Determinationskoeffizienten!

Aufgabe 11:

Gegeben sei der folgende Datenbestand aus einer Untersuchung zum Geschlecht und zur bevorzugten politischen Partei zufällig ausgewählter Personen:

	männlich	weiblich
SPD	25	15
Nicht-SPD	20	22

Wie stark ist der statistische Zusammenhang?

Aufgabe 12:

Benutzen Sie die Daten der vorangegangenen Aufgabe, bezeichnen Sie die Variable „Geschlecht" mit X und die (dichotomisierte) Variable „Partei" mit Y und codieren Sie „m" (männlich) mit 0, „w" (weiblich) mit 1, „SPD" mit 0 und „Nicht SPD" mit 1.

Damit ergibt sich folgende Tabelle:

X	Y	Häufigkeit
0	0	25
0	1	20
1	0	15
1	1	22

Berechnen Sie aus diesem Datenbestand den Korrelationskoeffizienten r
und vergleichen Sie das Ergebnis mit dem Wert des oben berechneten PhiKoeffizienten!