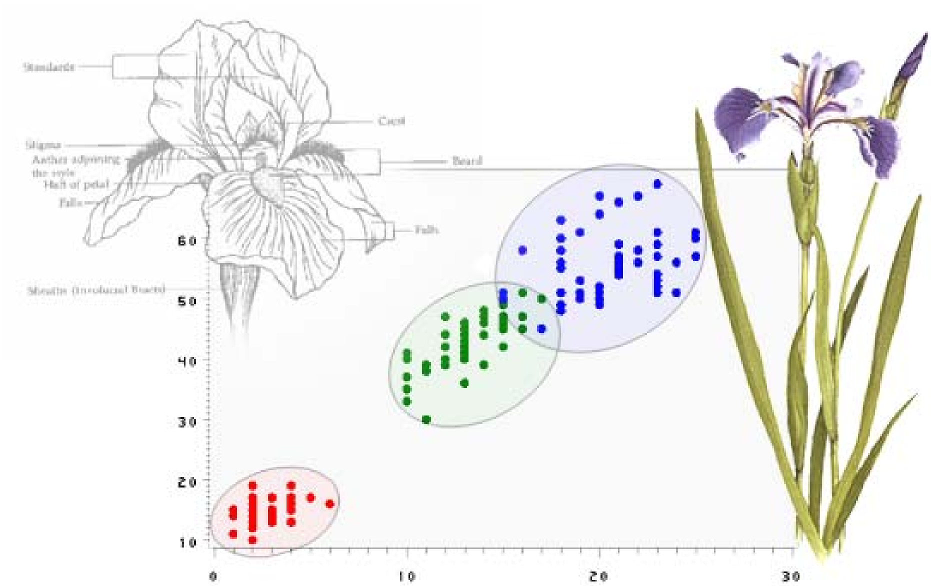
ZHA - Zusammenhänge



Lernziele und Überblick

In dieser Lerneinheit werden wir verschiedene Formen des Zusammenhanges zwischen Merkmalen darstellen.

Lernziele

Nach dem Durcharbeiten dieser Lerneinheit sollen Sie

- zwischen deterministischen und stochastischen Zusammenhängen unterscheiden können,
- zwischen positiven, negativen und keinen stochastischen Zusammenhängen unterscheiden können,
- zwischen den linearen und nicht-linearen Zusammenhängen unterscheiden können,
- mit Hilfe eines Streudiagramms die Stärke des Zusammenhanges erklären können.

Gliederung

In dieser Lerneinheit werden die folgenden Begriffe eingeführt:

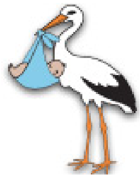
- der deterministische und stochastische Zusammenhang,
- der positive und negative Zusammenhang,
- der lineare und nicht-lineare Zusammenhang,
- das Streudiagramm.

Zeitbedarf und Umfang

Für die Durcharbeitung dieser Lerneinheit benötigen Sie ca. 120 Minuten und für die Übungen mit der Statistiksoftware **R** ca. 60 Minuten.

1 Einleitung

Bis jetzt haben wir stets eine Variable für sich betrachtet. Eigentlich ist das relativ langweilig. Erst wenn mindestens zwei Variable simultan betrachtet werden, kommen wir zu wirklich interessanten Fragestellungen.

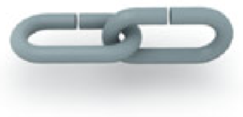


> *Was hat der Weltweizenertrag mit der Sonnenfleckenaktivität zu tun?*
>
> *Gibt es einen Zusammenhang zwischen Geburtenzahl und Anzahl der Storchennester?*

Funktionale Zusammenhänge sind uns vertraut: Heizkosten hängen von der durchschnittlichen Raumtemperatur und der Außentemperatur ab, Fahrtzeiten von der Entfernung etc. Im Folgenden werden wir Methoden kennenlernen, die geeignet sind, auch unbestimmtere Abhängigkeiten zwischen Variablen zu beschreiben.

Es gibt viele Situationen, in denen es wünschenswert ist, etwas über den Zusammenhang (oder die Abhängigkeit) zwischen zwei oder mehreren Eigenschaften eines Objektes, Produktes, Prozesses oder Individuums zu erfahren.

Die Analyse von Zusammenhängen ist ein Teilgebiet der multivariaten *(lat.: multus - vielfach, varia - Allerlei)* Statistik. Dabei erfassen statistische Untersuchungen mehrere Merkmale (sogenannte multivariate Datensätze) eines Merkmalsträgers gleichzeitig.



Erhebt man an der Untersuchungseinheit einer Erhebung mehrere Merkmale zugleich (z. B. Körpergröße, Gewicht, Haar- und Augenfarbe von Personen), so kann man die einzelnen Merkmale (eindimensionale Merkmale) auch zu einem mehrdimensionalen Merkmal zusammenfassen.

Fragen über Fragen

Wir wollen uns hier auf die Betrachtung von zwei Merkmalen (sogenannte bivariate Datensätze, Vorsilbe lat.: bi - zwei) beschränken, um die Darstellung zu vereinfachen.

Dabei können im Wesentlichen folgende Fragestellungen auftreten:

> *Besteht überhaupt ein Zusammenhang zwischen zwei Merkmalen (oder Variablen)?*

Wenn ein Zusammenhang besteht, dann fragen wir:

- *Wie stark ist dieser Zusammenhang?*
- *Wie lässt sich die Stärke (der Grad, die Intensität) des Zusammenhangs bzw. die Abhängigkeit zwischen zwei (oder mehr) als wechselseitig abhängig angenommenen Merkmalen bzw. Variablen messen?*
- *Lässt sich der Zusammenhang in einer bestimmten Form (Typ und Art) darstellen?*
- *Lassen sich die beobachteten Werte einer Variablen Y durch die Werte einer oder mehrerer anderer Variablen X (X_1, X_2, X_3, ...) näherungsweise bestimmen?*

2 Formen des Zusammenhanges

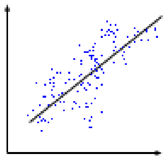
Je nach Art der vorliegenden Skalen von X und Y - ob die Merkmalausprägungen auf einer nominalen, ordinalen oder kardinalen Skala statistisch erhoben wurden - unterscheidet man in der statistischen Methodenlehre zwischen der Kontingenzanalyse (oder Assoziationsanalyse, lat.: contingentia - Zufälligkeit) und der Korrelationsanalyse (oder Maßkorrelationsanalyse) mit dem Spezialfall der Rangkorrelationsanalyse.

Die **Kontingenzanalyse** ist die Bezeichnung für eine statistische Zusammenhangsanalyse auf der Basis einer Kontingenztabelle. In der deskriptiven Statistik werden Kontingenztabellen in der Regel nur für nominale (manchmal für ordinale) Merkmale erstellt und analysiert.

	I	II	III	Summe
a	123	345	26	444
b	645	1234	6356	7777
Summe	768	1579	6382	8221

Die **Rangkorrelationsanalyse** ist eine Analyse eines Zusammenhanges zweier ordinalskalierter Merkmale mit Hilfe von Rangzahlen. Der Rangkorrelationskoeffizient nach SPEARMAN hat eine besondere praktische Bedeutung wegen seiner einfachen Berechnung.

Die **Korrelationsanalyse** ist eine Analyse von Zusammenhängen zwischen zwei kardinalen Merkmalen. Dabei erweist es sich als vorteilhaft, einer Korrelationsanalyse eine grafische Datenanalyse auf Basis eines Streudiagramms vorzulagern.



Im Rahmen der Korrelationsanalyse wird geprüft, ob zwei kontinuierliche Variablen X und Y **linear zusammenhängen** und wie stark dieser **Zusammenhang** ist.

Anmerkung

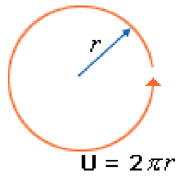
Typen des Zusammenhanges

Es gibt viele Beispiele, die dadurch gekennzeichnet sind, dass die jeweiligen Merkmale exakt durch eine Funktionsgleichung miteinander verbunden sind. Diese Funktionsgleichung gestattet im Rahmen der Messgenauigkeit genaue Vorhersagen der Ausprägung des einen Merkmals bei ausschließlicher Bekanntheit der Ausprägung des anderen Merkmals.

Da der Wert von Y bei Kenntnis von X exakt berechnet werden kann, also vorher bestimmt ist, spricht man in solchen Fällen von **deterministischen Zusammenhängen**.

Im Unterschied zu deterministischen Zusammenhängen lassen **stochastische (zufallsabhängige) Zusammenhänge** je nach Stärke des Zusammenhangs unterschiedlich präzise Vorhersagen zu.

2.1 Deterministische Zusammenhänge



Der Umfang U und der Radius r aller Kreise sind perfekt - funktional - korreliert, da $U = 2\pi r$. Mit zunehmendem Radius steigt der Umfang: gleiche Radien ergeben stets dieselben Umfänge und umgekehrt.

Der Umfang (U) ist eine Funktion des Radius (r). Zwischen U und r besteht ein funktionaler Zusammenhang. Hierbei ist es gleichgültig, welche Variable man fest vorgibt und welche man misst.

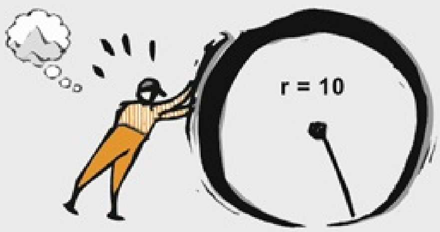
Bei funktionaler Abhängigkeit ist

X fest vorgegeben und
Y eine wohldefinierte Funktion von X: $Y = f(X)$.

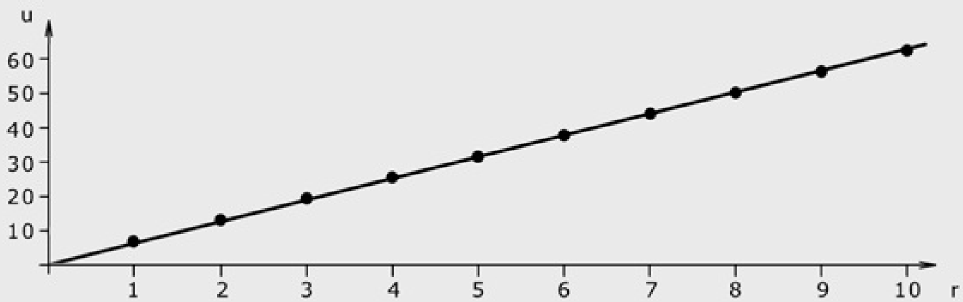
Die folgende Diashow visualisiert den Zusammenhang zwischen dem Radius r und dem Umfang U. Für einzelne Radien lassen sich entsprechende Umfänge berechnen. Die Wertepaare sind in der Tabelle verzeichnet. Mit dieser Wertetafel lassen sich zunächst einzelne Punkte der Funktionskurve im Bereich 1 bis 10 zeichnen. Berechnet man die Umfänge für weitere Zwischenwerte von r, so entsteht bald eine zusammenhängende Punktreihe, die sich als Abschnitt einer Geraden erweist.

Diashow

Der Zusammenhang zwischen Umfang und Radius



Radius r	1	2	3	4	5	6	7	8	9	10
Umfang u	6,28	12,57	18,85	25,13	31,42	37,70	43,98	50,26	56,55	62,83



2.2 Beispiele für deterministische Zusammenhänge

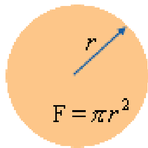
Dieser Absatz soll keinesfalls das Studium mathematischer Funktionen ersetzen. Natürlich gehen wir davon aus, dass Sie wissen, was eine Funktion vom Typ $y = f(x)$ ist.
Damit Sie aber zwischen näherungsweisen Abhängigkeiten und deterministischen Beziehungen unterscheiden können, haben wir hier zunächst noch einmal Beispiele aufgeführt.

Beispiel

Funktionale Zusammenhänge

Kreisfläche

Die Kreisfläche F ist gleich der Konstante π mal dem Radius des Kreises r zum Quadrat: $F = \pi r^2$ mit $\pi \approx 3{,}14159$. D. h. jedem Radius entspricht eine ganz bestimmte Fläche.



Die Kreisfläche F ist also eine Funktion des Radius r.

Umgekehrt geht es auch:
Der Radius des Kreises r ist eine Funktion der Kreisfläche F
(d. h. jeder Kreisfläche entspricht ein ganz bestimmter Radius): $r = \sqrt{\frac{F}{\pi}}$.

Kinetische Energie

Die kinetische Energie E eines bewegten Körpers ist durch seine Masse m und seiner Geschwindigkeit v vorherbestimmt – determiniert:
$E = \frac{mv^2}{2}$

2.3 Stochastische Zusammenhänge

Was bedeutet stochastischer Zusammenhang?

In der Praxis gibt es kaum Fälle (es sei denn triviale), in denen tatsächlich ein funktionaler linearer Zusammenhang zwischen zwei empirischen Variablen besteht. In der Praxis ist man mehr oder weniger weit von der Idealvorstellung – funktionaler linearer Zusammenhänge – entfernt.

Stochastischer Zusammenhang

Wenn anstelle eines deterministischen Zusammenhanges eine mehr oder weniger lose Verkettung zweier (oder mehrerer) Zufallsvariablen getreten ist, dann spricht man von einem stochastischen Zusammenhang.

Unterscheidung zwischen funktionalem und stochastischem Zusammenhang

Wo liegen die Unterschiede zwischen funktionalem und stochastischem Zusammenhang und wofür werden sie benötigt?

Deterministischer Zusammenhang:

- Y lässt sich genau aus X vorhersagen.
- Unterschiede in Y korrespondieren perfekt mit Unterschieden in X.

Stochastischer Zusammenhang:

- Y lässt sich zwar aus X vorhersagen, jedoch ist die Ausprägung von Y noch von anderen Variablen außer X abhängig.
  Gleiches gilt für Vorhersagen von X durch Y.
- Unterschiede in Y korrespondieren zwar mit Unterschieden in X, aber es treten im Einzelfall Ungenauigkeiten bei der Vorhersage auf.

Exkurs

Exkurs: Linearer Zusammenhang zwischen den Variablen X und Y (Siehe Anhang)

3 Typen und Arten von Zusammenhängen

Wir sprechen von einem positiven Zusammenhang, wenn die Werte der zwei Variablen gleich gerichtet sind. Es ergeben sich folgende Konstellationen:

- große Werte von X treten zusammen mit großen von Y auf und
- kleine Werte von X mit kleinen von Y.

Anders ausgedrückt:
die Punkte liegen hauptsächlich in den Quadranten I und III, d. h. das Streudiagramm zeigt einen ansteigenden Verlauf der Werte von X und Y (siehe Abbildung)

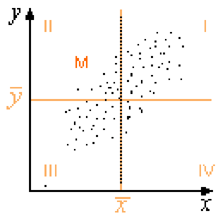


Abb.: Streudiagramm mit positivem Zusammenhang

Interpretation

Aus dem Streudiagramm ist erkennbar, dass ein positiver (Art) stochastischer (Typ) Zusammenhang zwischen X und Y besteht:
Je größer X ist, desto größer ist Y im Allgemeinen.

Noch einmal:

Einen positiven stochastischen Zusammenhang erkennt man daran, dass die Punkte in einer Wolke mit positiver Orientierung angeordnet sind.

Wird die zentrale Lage der Punktwolke gut durch eine Gerade beschrieben, spricht man von linearem Zusammenhang.

Hinweis

Lineare Funktionen werden oft als Näherung beliebiger Funktionen betrachtet.

Beispiel

Beispiele positiver Zusammenhänge

Die folgende Tabelle zeigt Merkmale, zwischen denen ein positiver Zusammenhang besteht.

Merkmal X	Merkmal Y
Körpergröße	Schrittlänge
Körpergröße	Schuhgröße
Mathematisches Verständnis	Klausurergebnis
Länge von Bohnen	Gewicht von Bohnen
Wortschatz	Alter von Kleinkindern
Fahrgeschwindigkeit	Bremsweg von Fahrzeugen

Tab.: Beispiele positiver Zusammenhänge

3.1 Negative stochastische Zusammenhänge

Ersetzen Sie bei Variablen, die in negativem Zusammenhang stehen, einfach Y durch -Y und Sie erhalten eine positive Relation. Inhaltlich wird ein negativer Zusammenhang aber oft anders aufgefasst, deshalb alles noch einmal. Negative Zusammenhänge treten auf, wenn kleine Werte von X eher mit großen von Y bzw. große von X mit kleinen von Y einhergehen.

Anders gesagt:
Die Punkte liegen hauptsächlich in den Quadranten II und IV, d. h. das Streudiagramm zeigt einen fallenden Verlauf zwischen den X- und Y-Werten (siehe Abbildung).

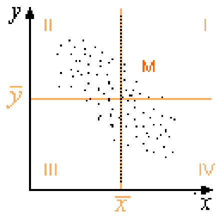


Abb.: Streudiagramm
mit negativem Zusammenhang

Aus diesem Streudiagramm lässt sich die **Art** und der **Typ** des Zusammenhanges erkennen:
Es besteht ein negativer **(Art)** linearer stochastischer **(Typ)** Zusammenhang zwischen X und Y: je größer die Werte von X, desto kleiner die von Y.

Beispiel

Beispiele negativer Zusammenhänge

Die folgende Tabelle zeigt Merkmale, zwischen denen ein negativer Zusammenhang besteht:

Merkmal X	Merkmal Y
Arbeitszeit im Nebenjob	Studienleistungen
Gefahrene Kilometer im Auto	Profilstärke des Reifens
Alter eines Menschen	Sehschärfe des Menschen
Anzahl zu lernender Vokabeln	Prozentsatz der behaltenen Wörter
...	...

Tab.: Beispiele negativer Zusammenhänge

Kennen Sie weitere negative Zusammenhänge?

3.2 Kein erkennbarer Zusammenhang

Zwischen zwei Variablen X und Y besteht kein erkennbarer Zusammenhang, wenn hohe Ausprägungen des Merkmals X keinen Hinweis darauf geben, ob die Ausprägung in Y eher hoch oder niedrig sein wird .

Anders ausgedrückt: die Punkte liegen etwa gleichmäßig und ohne Struktur in allen vier Quadranten (siehe Abbildung):

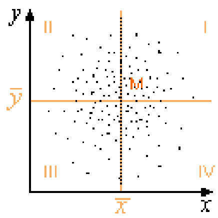


Abb.: Daten ohne erkennbaren Zusammenhang im Streudiagramm

Aus diesem Streudiagramm lassen sich Art und Typ des Zusammenhangs erkennen:

- Es besteht kein stochastischer Zusammenhang zwischen X und Y:
- Hohe (niedrige) Ausprägungen in X geben keinen Hinweis darauf, ob die Ausprägung in X bzw. Y eher hoch oder niedrig sein wird.

Beispiel

Merkmale ohne erkennbaren Zusammenhang

Die folgende Tabelle zeigt Merkmale, zwischen denen kein stochastischer Zusammenhang besteht:

Merkmal X	Merkmal Y
Körperbau	Persönlichkeitsmerkmale
Körpergröße	Haarlänge des Menschen
Länge eines Schiffes	Alter des Kapitäns
Körpergröße	Intelligenz
Punktzahl Würfel 1	Punktzahl Würfel 2

Tab.: Beispiele für Merkmale ohne erkennbaren Zusammenhang

Achtung

In der Wahrscheinlichkeitsrechnung ist die Aussage, dass „kein linearer Zusammenhang zwischen Zufallsvariablen besteht“, **nicht gleichbedeutend** mit der Aussage, dass die beiden Variablen **stochastisch unabhängig** voneinander sind.

Ein Zusammenhang kann dennoch existieren, muss aber nicht linear sein.

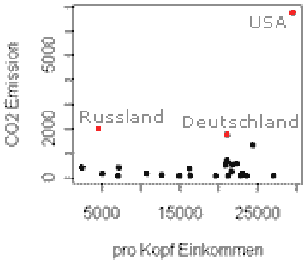
Beispiel

Bei der Funktion $Y = aX^2$ liegt ein quadratischer Zusammenhang vor, jedoch kein linearer.

4 Streudiagramm

Was jeder weiß: „Ein Bild sagt oft mehr als 1000 Worte!“
wird jetzt noch einmal festgehalten.

Das wichtigste Werkzeug in einer zweidimensionalen Darstellung von Merkmalen sind sog. Streudiagramme (Punktwolken).



Definition

Streudiagramm

Sind X und Y zwei kardinale Merkmale, dann heißt die grafische Darstellung von n beobachteten Wertepaaren $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ als Punktwolke in einem Koordinatensystem **Streudiagramm**.

Synonyme

Synonyme:
Scatterplot *(engl.: scatter - zerstreuen, plot - grafische Darstellung, Plan, Parzelle, Grundriss)*, Scattergram, Streuungsdiagramm.

Ein **Punktdiagramm** *(engl.: scattergram)* entspricht einem Streudiagramm, nur ist die Bezeichnung Punktdiagramm in der Präsentationsgrafik verbreiteter.

4.1 Eigenschaften und Interpretationen des Streudiagramms

Darstellung

Ein Streudiagramm ist die grafische Darstellung von Datenpaaren.

Punktwolke

Die n Wertepaare (x_1, y_1), (x_2, y_2), ... , (x_n, y_n) erscheinen als Punktewolke in einem zweidimensionalen Achsenkreuz (Merkmalsebene). Also, ein Streudiagramm erhält man, wenn man alle Punkte zweier Variablen X und Y in ein zweidimensionales Achsenkreuz einzeichnet. Pro (x, y)-Datenpaar wird ein Punkt dargestellt.

Die Werte des Merkmals X werden auf der Abszisse (horizontal skaliert) abgetragen, die des Merkmals Y auf der Ordinate (vertikal skaliert).

Typen, Art und Stärke

Aus dem Verlauf und der Form des Streudiagramms lassen sich Schlussfolgerungen hinsichtlich

- **Art** (positiv oder negativ) und
- **Stärke**

eines Zusammenhangs zwischen den zwei Merkmalen X und Y ziehen.

Schwerpunkt

Der Punkt M mit den Koordinaten $(\overline{x}, \overline{y})$ ist der Mittelpunkt (physikalisch der Schwerpunkt) der Punkte des Streudiagramms.

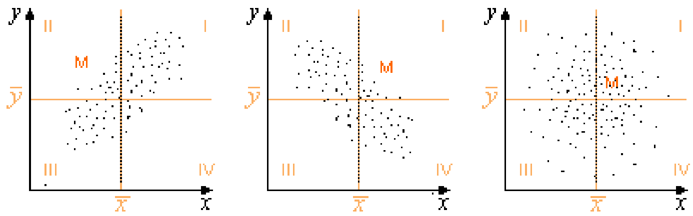


Abb.: Schwerpunkt im Streudiagramm

Interpretation

Positiver oder gleichläufiger Zusammenhang: gestreckter steigender Verlauf.
Negativer oder gegenläufiger Zusammenhang: gestreckter fallender Verlauf.
Kein Zusammenhang: Punkte gleichmäßig gestreut.

Der Korrelationskoeffizient ist eine Maßzahl für die **Stärke** und die **Art (Richtung)** eines linearen Zusammenhangs.

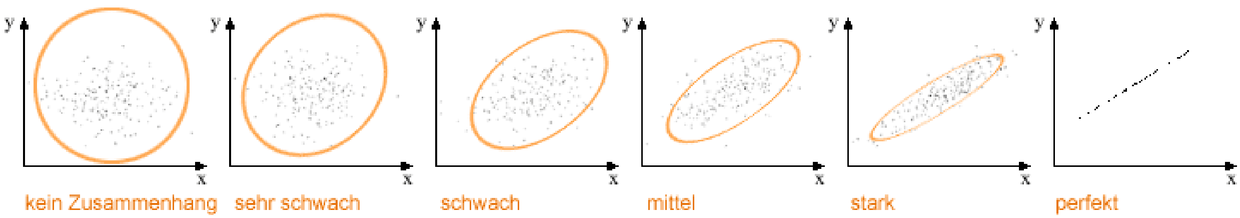


Abb.: Positive lineare Zusammenhänge

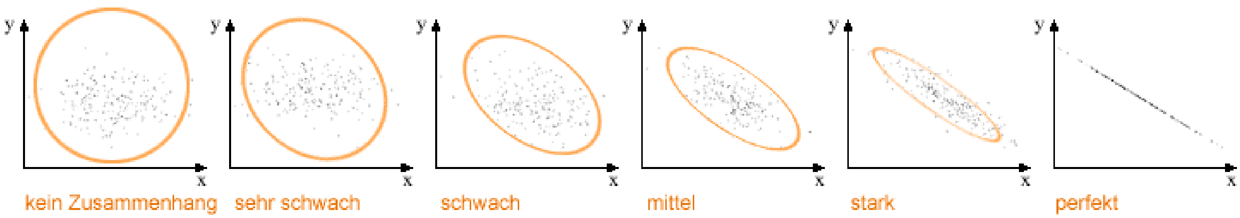


Abb.: Negative lineare Zusammenhänge

4.2 Stärke von Zusammenhängen

Aus einem Streuungsdiagramm lässt sich auch die Stärke des Zusammenhanges erkennen.

Die folgenden Abbildungen zeigen Streudiagramme mit positiven linearen sowie negativen linearen Zusammenhängen verschiedener Stärke.

Hinweis

Wir verwenden in den folgenden Abbildungen den Korrelationskoeffizienten **r**, der in Lerneinheit *„KOR - ‚Korrelation'“* systematisch behandelt wird.

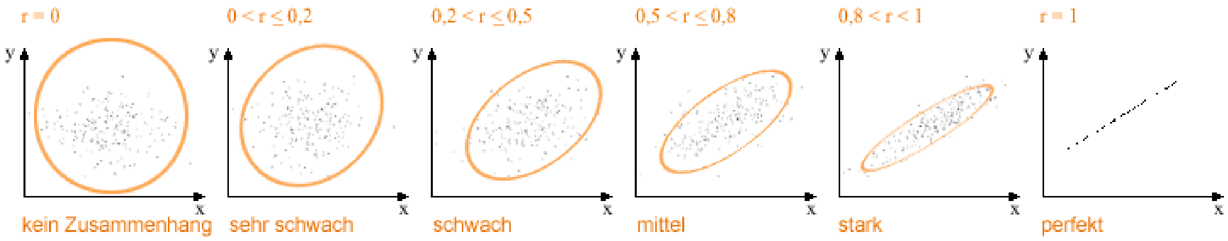


Abb.: Streudiagramme positiver linearer Zusammenhänge verschiedener Stärke

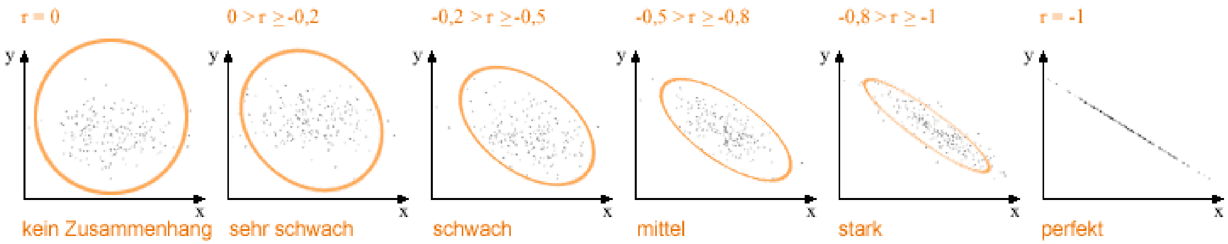


Abb.: Streudiagramme negativer linearer Zusammenhänge verschiedener Stärke

4.3 Beispiel für positive Zusammenhänge

Die zwei Variablen Körpergröße (X) und Gewicht (Y) von Individuen scheinen einen Zusammenhang (oder Korrelation) aufzuweisen.

Hierbei stellt sich die Frage, ob es einen Zusammenhang zwischen Körpergröße und Körpergewicht eines Menschen gibt.

Man kann sich leicht vorstellen, dass diese beiden Größen nicht unabhängig voneinander sind. Der Zusammenhang gilt aber nur im Mittel, denn man kann zwar erwarten, dass eine große Person schwerer als eine kleine ist, aber das Umgekehrte ist im Einzelfall immer möglich. Vermutlich liegt hier ein stochastischer Zusammenhang vor.

Beispiel

Körpergröße und Gewicht von 26 Studierenden

Es wurden von einer zufällig ausgewählten Gruppe von Studierenden (26 Personen) die Daten Körpergröße (Zufallsvariable X) und Gewicht (Zufallsvariable Y) ermittelt (siehe nachfolgende Tabelle).

i	Person	x_i (cm)	y_i (kg)	i	Person	x_i (cm)	y_i (kg)
1	Magda	158	48	14	Jörg	173	73
2	Anna	160	59	15	Volker	174	76
3	Roland	163	102	16	Stefan	174	63
4	Swetlana	165	57	17	Heike	174	83
5	Alexander	165	80	18	Karpo	177	65
6	Tamara	168	53	19	Vladimir	177	77
7	Iwan	168	58	20	Stanislaw	178	72
8	Eva	168	58	21	Felix	178	85
9	Karoline	169	66	22	Andrej	186	67
10	Nikolaj	170	87	23	Walerij	190	80
11	Alexandra	171	70	24	Jost	191	95
12	Ingrid	171	79	25	Stefan	192	90
13	Oksana	172	68	26	Witalij	194	72

Tab.: Beispiel für Körpergröße und Gewicht

4.4 Der Mittelpunkt im Streudiagramm

Zur Aufteilung der Wolke in Quadranten wird der Schwerpunkt der Daten verwendet. Den müssen wir zunächst bestimmen. Die Koordinaten ($\overline{x}, \overline{y}$) des Mittelpunktes M ergeben sich wie folgt:

$$\overline{x} = \frac{1}{26}\sum_{i=1}^{26} x_i = \frac{4526}{26} = 174,08\,cm$$

$$\overline{y} = \tfrac{1}{26}\sum_{i=1}^{26} y_i = \tfrac{1883}{26} = 72,42\,kg$$

Trägt man die Daten aus der Tabelle der Studierendengruppe auf, so ergibt sich folgendes Streudiagramm der Körpergröße und Gewichte im 4-Quadranten-Schema.

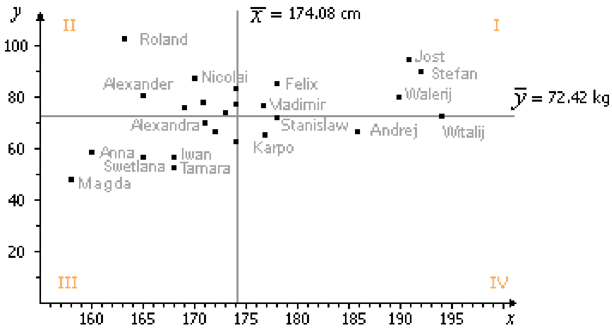


Abb.: Streudiagramm Körpergröße und Gewicht

Die Aufteilung verwendet die berechneten Mittelpunktskoordinaten.

4.5 Interpretationen zum Beispiel

Obwohl manchen Merkmalsausprägungen x_i auf der x-Achse zwei oder sogar drei verschiedene Merkmalsausprägungen y_k, y_i und y_m zugeordnet werden können,
resultiert eine Punktwolke mit erkennbarem Trend:

Das Körpergewicht steigt mit zunehmender Körpergröße an.

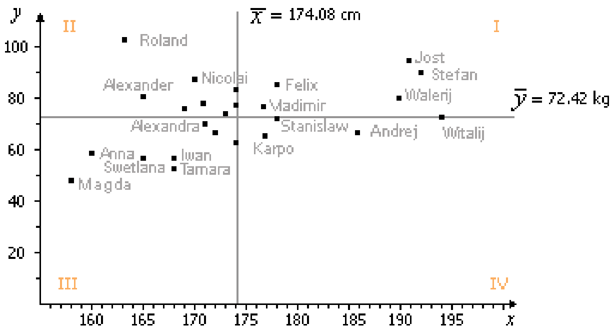


Abb.: Streudiagramm der Körpergröße und Gewicht

Na klar, das wissen wir doch, größere Personen haben tendenziell auch ein größeres Körpergewicht.

Durch die Betrachtung der Punktwolke im X,Y- Streudiagramm (Körpergröße-Körpergewicht) erhält man einen ersten Eindruck über

- einen stochastischen
- positiven
- mittleren (nach Augenmaß)

Zusammenhang zwischen zwei Variablen X und Y.

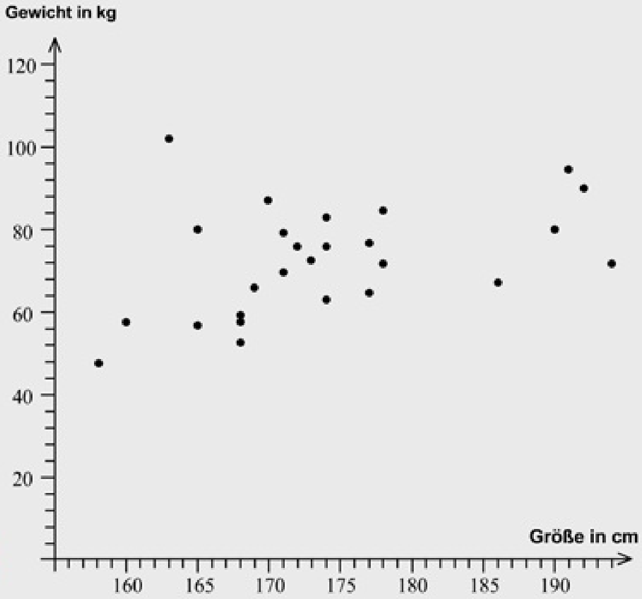
Im Unterschied zum funktionalen Zusammenhang ist zwischen den Zufallsvariablen Körpergröße und Gewicht eine mehr oder weniger lose Verkettung getreten, die man als positiven linearen stochastischen Zusammenhang bezeichnen kann.

Die folgende Diashow illustriert der Aufbau des Streudiagramms der Körpergrößen und Gewichte der 26 Studierenden.

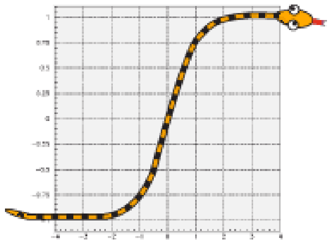
Diashow

Reihung

Person	x_i(cm)	y_i(kg)
1	158	48
2	160	59
3	163	102
4	165	57
5	165	80
6	168	53
7	168	58
8	168	59
9	169	66
10	170	87
11	171	70
12	171	79
13	172	68
14	173	73
15	174	76
16	174	63
17	174	83
18	177	65
19	177	77
20	178	72
21	178	85
22	186	67
23	190	80
24	191	95
26	192	90
26	194	72



5 Andere Arten von Zusammenhängen



Gelegentlich kann man beobachten, dass ein nicht-linearer Zusammenhang eine bessere Vorhersage gestattet als ein linearer. Einige charakteristische und gut erkennbare Typen wollen wir jetzt studieren.

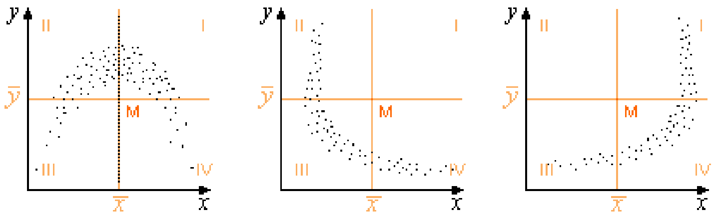
Ein nicht-linearer Zusammenhang liegt zum Beispiel dann vor, wenn zwischen den Werten von X und Y z. B eine der folgenden Relationen auftritt:

- exponentieller Zusammenhang
- logarithmischer Zusammenhang
- S-förmiger oder kubischer Zusammenhang
- umgekehrt U-förmiger oder parabolischer Zusammenhang

5.1 U-förmiger und exponentielle Zusammenhänge

Es gibt viele Beispiele für typische, nicht-lineare Zusammenhänge. Betrachten Sie den Verlauf der Punktwolken in den folgenden drei Diagrammen.
Wie würden Sie die Form charakterisieren?

Abb.: Streudiagramme im 4-Quadranten-Schema



Streudiagramme mit umgekehrt U-förmigem (parabolischem) und mit exponentiellen Zusammenhängen

Beispiel

Beispiele für Merkmale mit nicht-linearem umgekehrt U-förmigem stochastischen Zusammenhang

Tab.: Beispiele für typisch, nicht-lineare Zusammenhänge

Merkmal X	Merkmal Y
Informationsgehalt der Reize	Bewertung ästhetischer Reize
Körperliches Erregungsniveau	Konzentrationsfähigkeit
Anspannung bei Bewerbungsgespräch	Einstellungschance

5.2 Logarithmische und kubische Zusammenhänge

Sie erinnern sich an das Kapitel über Transformationen (vgl. Lerneinheit *„GST - Grundbegriffe der Statistik“*, Abschnitt 8.7 Einfach-logarithmische Transformation). Dort haben wir unter anderem die Logarithmus-Transformation untersucht.
Für Daten ergibt sich bei logarithmischen Zusammenhängen die folgende Form:

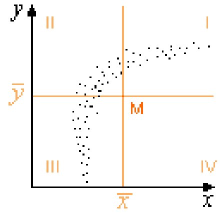


Abb.: Streudiagramm eines logarithmischen Zusammenhangs im 4-Quadranten-Schema

Häufig sind gewisse funktionale Formen eines Zusammenhanges physikalisch naheliegend.

Das Gewicht von Kartoffeln ist in der Regel nicht proportional zu ihrem Durchmesser, sondern zu ihrem Volumen. Das Volumen ist näherungsweise proportional zum Durchmesser hoch drei.

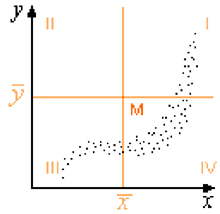


Abb.: Streudiagramm eines kubischen Zusammenhangs

Hinweis

Natürlich gibt es auch noch nicht-korrelierte Zusammenhänge. Mit diesen beschäftigen wir uns in diesem Kurs jedoch nicht.

Zusammenfassung

- Die **Kontingenzanalyse** ist die Bezeichnung für eine statistische Zusammenhangsanalyse auf der Basis einer Kontingenztabelle.
- Die **Rangkorrelationsanalyse** ist eine Analyse eines Zusammenhanges zweier ordinalskalierter Merkmale mit Hilfe von Rangzahlen.
- Die **Korrelationsanalyse** ist eine Analyse von Zusammenhängen zwischen zwei kardinalen Merkmalen.
- Wenn anstelle eines deterministischen Zusammenhanges eine mehr oder weniger lose Verkettung zweier (oder mehrerer) Zufallsvariablen getreten ist, dann spricht man von einem stochastischen Zusammenhang.
- Stochastische Zusammenhänge sind gekennzeichnet durch Form, Tendenz und Stärke.
- Ein Streudiagramm ist die grafische Darstellung von Datenpaaren.
- Streudiagramme, bei denen die Punktwolke nahe einer Geraden liegt, entsprechen einem linearen Zusammenhang der Merkmale.
- Wir sprechen von einem positiven Zusammenhang, wenn die Werte der zwei Variablen gleich gerichtet sind.
- Wird die zentrale Lage der Punktwolke gut durch eine Gerade beschrieben, spricht man von linearem Zusammenhang.
- Stochastische Zusammenhänge können auch nicht-linear sein.
- Das Streudiagramm ist das wichtigste Hilfsmittel zur Darstellung stochastischer Zusammenhänge.

Sie sind am Ende dieser Lerneinheit angelangt. Auf der folgenden Seite finden Sie noch die Übungen zur Wissensüberprüfung und weitere Übungen.

Wissensüberprüfung

Formulieren

Übung ZHA-01

Werbung und Umsatz

In der nachfolgenden Tabelle sind die Umsatzzahlen einer bestimmen Sorte Prosecco sowie die zugehörigen Werbungskosten in zehn Absatzregionen des Weinfachgeschäftes Maestro zusammengestellt.

Absatzregionen	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Werbung (in €)	90	120	50	90	60	120	140	50	30	110
Umsatz (in Tsd. €)	20	28	13	15	12	25	30	10	8	17

1. 1. Stellen Sie die Daten in einem Streudiagramm dar.
2. Ergänzen Sie dieses Streudiagramm durch die jeweiligen Mittelwertlinien.
3. Ziehen Sie daraus Schlussfolgerungen über den Zusammenhang zwischen Umsatz und Werbung.
4. Welche Art und welcher Typ lässt sich aus diesem Streudiagramm erkennen?

Lösung (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Multiple Choice

Übung ZHA-02

Welche Aussagen sind falsch und welche richtig?

	Richtig	Falsch	Auswertung
Bevor man versucht, die Stärke des Zusammenhangs zwischen zwei Merkmalen mit statistischen Methoden zu erfassen, sollte man prüfen, ob überhaupt ein sachlicher Zusammenhang angenommen werden kann.			
Ein statistischer Zusammenhang ist stets der Beweis eines ursächlichen Zusammenhangs.			
Stochastische Zusammenhänge können nicht anhand von Daten beobachtet werden.			
Ein positiver linearer stochastischer Zusammenhang zwischen X und Y besteht dann, wenn größere X - Werte eher gemeinsam mit größeren Y - Werten auftreten.			

Multiple Choice

Übung ZHA-03

Welche Aussagen sind falsch und welche richtig?

	Richtig	Falsch	Auswertung
Aus dem Verlauf und der Form des Streudiagramms lassen sich Schlussfolgerungen hinsichtlich des Typs, der Art und der Stärke eines Zusammenhangs zwischen Merkmalen X und Y ziehen.			
Die Stärke eines Zusammenhangs zwischen zwei Merkmalen X und Y ist die Enge (Grad, Intensität) der Relation zwischen den Beobachtungen der Merkmale X und Y.			
Aus dem gestreckten und steigenden Verlauf einer Punktewolke ist zu erkennen, dass zwischen zwei Merkmalen ein gegenläufiger statistischer Zusammenhang besteht.			
Die Korrelationsanalyse ist eine Analyse von Zusammenhängen zwischen zwei kardinalen Merkmalen.			

Übungen mit der Statistiksoftware R

Statistiksoftware R

Die in der Lerneinheit behandelten Themen können Sie anhand der folgenden Übungsaufgaben mit der Statistiksoftware **R** bearbeiten. Um die Übungen zu bearbeiten, muss die Software „**R**" auf Ihrem Rechner installiert sein.

Installationshinweise [Manuals | R Installation and Administration]

Für einige Übungen stehen auch Musterlösungen für das Programm Excel bereit.

Berechnen

Übung ZHA-04a

Hubble

1929 untersuchte Edwin Hubble die Beziehung zwischen dem Abstand der Galaxien von der Erde und der Geschwindigkeit, mit der sie sich scheinbar entfernt. Galaxien scheinen sich immer von uns zu entfernen, unabhängig davon in welche Richtung wir blicken. Man betrachtet dies als eine Folge des Urknalls. Hubble hoffte Kenntnisse darüber zu gewinnen, wie sich das Universum entwickelt hat und wie es sich in Zukunft entwickeln wird. Diese Untersuchungen bildeten die Grundlage zur Entwicklung der Hubble-Konstante H_o. Die Datei **`hubble.txt`** enthält den Abstand in Lichtjahren zwischen der Erde und 24 verschiedenen Galaxien und ihre Rückzugsgeschwindigkeit in km/s. Speichern Sie die Datei im gleichen Ordner wie Ihre R-Datei.

`hubble.txt` (2 KB)

Aufgabe

1. Lässt sich ein Zusammenhang zwischen Geschwindigkeit und Abstand feststellen? Benutzen Sie zur Überprüfung geeignete Maßzahlen und bestätigen Sie Ihr Ergebnis grafisch.
2. Interpretieren Sie Ihr Ergebnis.

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 20 Minuten

Berechnen

Übung ZHA-04b

Preise

Die Werte in der Tabelle geben die Preise für verschiedene Lebensmittel in Geschäft A und zum Vergleich im Geschäft B wieder.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
Preise A	1,5	0,25	9	4,75	3,5	8,75	6	2	1	3,25	1	5,5	7,75
Preise B	0,5	0,75	6	3,75	2,5	6,75	4	1	2	2,25	1	4,5	5,75
i	14	15	16	17	18	19	20	21	22	23	24	25	
Preise A	5	0,75	2,5	1,5	8,5	4,5	3	3,5	5	7	2	2,5	
Preise B	4	0,25	1,5	2,5	5,5	3,5	2	3,5	2	6	1	0	

1. Berechnen Sie jeweils das arithmetische Mittel sowie die Kovarianz und veranschaulichen Sie die Kovarianz grafisch. Zeichnen Sie dazu das 4-Quadrantenschema.

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 20 Minuten

Berechnen

Übung ZHA-04c

Umsatz

Es soll der Zusammenhang zwischen dem Umsatz verschiedener Abteilungen und deren Ausgaben für Werbung untersucht werden.

i	1	2	3	4	5	6	7	8	9	10
Umsatz	20	28	13	15	12	25	30	10	8	22
Werbung	90	120	50	90	60	120	140	50	30	110

1. Zeichnen Sie dazu ein Streudiagramm.
2. Was für ein Zusammenhang (Art,Typ) lässt sich erkennen?

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 20 Minuten

Zusätzliche Übungsaufgaben

Zum Vertiefen der vorgestellten Inhalte finden Sie hier zusätzliche Übungsaufgaben zur freiwilligen Bearbeitung.

Berechnen

Übung ZHA-05

Alltagsbeispiele

Geben Sie fünf Beispiele aus dem Alltag für den funktionalen Zusammenhang.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung ZHA-06

Beobachtungen

Stellen Sie für die folgende Menge von Beobachtungen das Streudiagramm auf:

(2;4), (3;5), (4;4), (5;2), (6;4), (8;4), (7;5), (2;7), (4;7), (6;7), (8;7), (5;9)

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung ZHA-07

Alter und monatliches Einkommen

Es wurden zehn zufällig ausgewählte Personen nach den Daten:
Alter (Zufallsvariable X) und monatliches Einkommen (Zufallsvariable Y) befragt (siehe nachfolgende Tabelle).

Stellen Sie die Daten in einem Streudiagramm dar und ergänzen Sie dieses Streudiagramm durch die jeweiligen Mittelwertlinien.

Person	Alter	monatliches Einkommen (in 100 €)
A	22	12
B	28	24
C	32	14
D	36	26
E	40	18
F	44	28
G	48	32
H	52	16
I	56	30
J	62	20

Lösung (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Appendix

Exkurs: Linearer Zusammenhang zwischen den Variablen X und Y

Die einfachste Beziehung zwischen zwei Variablen ist der lineare Zusammenhang.

Linearer Zusammenhang heißt, dass sich Y aus der linearen Transformation von X nach der Vorschrift Y = a + b X ergibt.
Trägt man X und Y in einem Koordinatensystem ab, so beschreibt dieser Zusammenhang eine Gerade (siehe Abbildung).

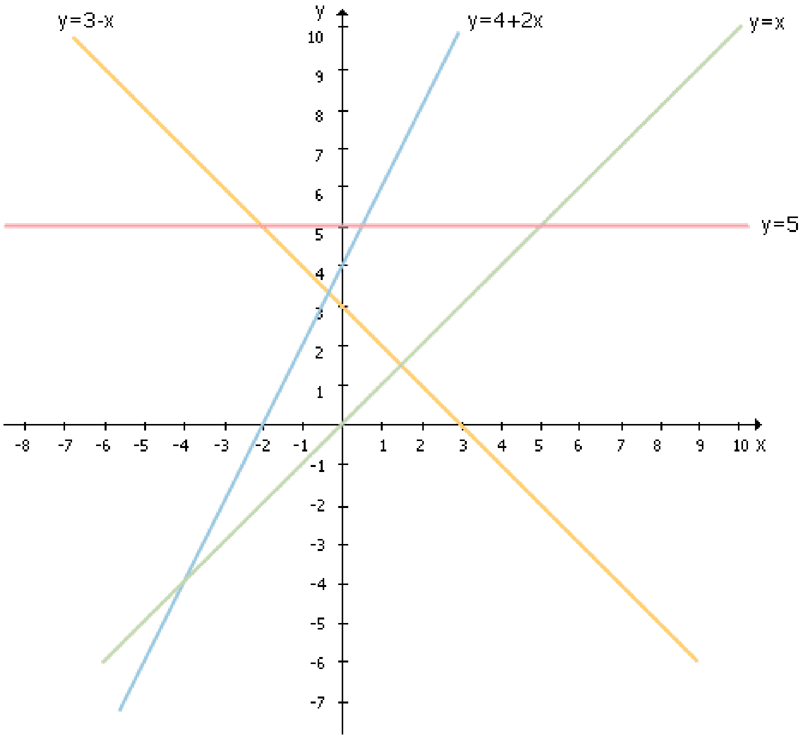


Abb.: Lineare Zusammenhänge

In der allgemeinen, linearen Funktionsgleichung kennzeichnet X die unabhängige Veränderliche, Y die abhängige Veränderliche.
Die Steigung der Geraden b ist der Tangens des Winkels zwischen der x-Achse und der Geraden:

$$b = tan\ \alpha = \frac{\Delta y}{\Delta x} = \frac{y_B - y_A}{x_B - x_A}.$$

Die Höhenlage a bezeichnet den Schnittpunkt der Geraden mit der y-Achse.
Im rechtwinkligen Dreieck ABC ist der Tangens des Winkels definiert als Quotient der Gegenkathete $\Delta y = y_B - y_A$ und der Ankathete $\Delta x = x_B - x_A$.
Somit ist die Steigung b einer Gerade gleich dem Tangens des Steigungswinkels α: b = tan α.

Die folgende Abbildung stellt die grafische Interpretation zur Steigung b und Höhenlage a dar.

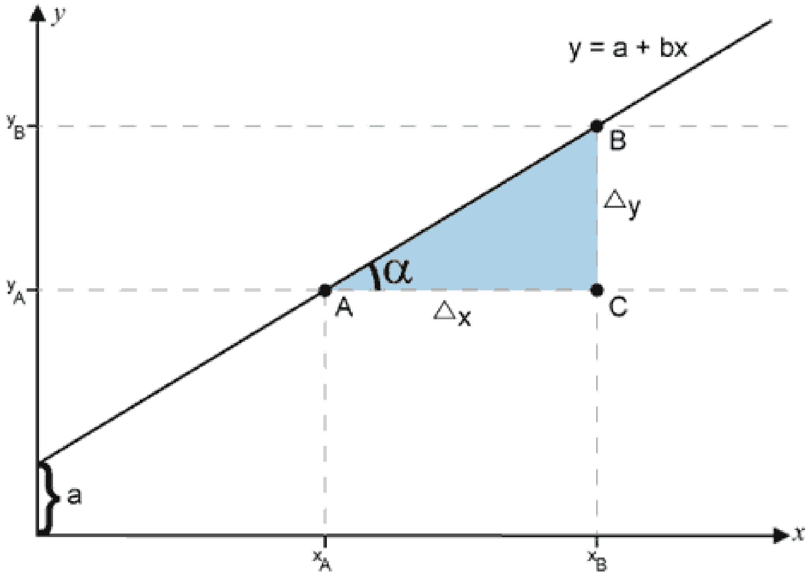


Abb.: Lineare Transformation: Steigung b und Höhenlage a.

Die Steigung b einer Gerade kann positiv, negativ oder gleich Null (b = 0) sein
(die Höhenlage soll hier kurz vernachlässigt werden, d. h. a = 0):

- ist die Steigung b positiv (b > 0), wachsen die y-Werte mit steigenden x-Werten.
- ist die Steigung b negativ (b < 0), wallen die y-Werte mit steigenden x-Werten.
- ist die Steigung b = 0, so verläuft die Gerade parallel zur x-Achse

---

Lösung zur Übung ZHA-01

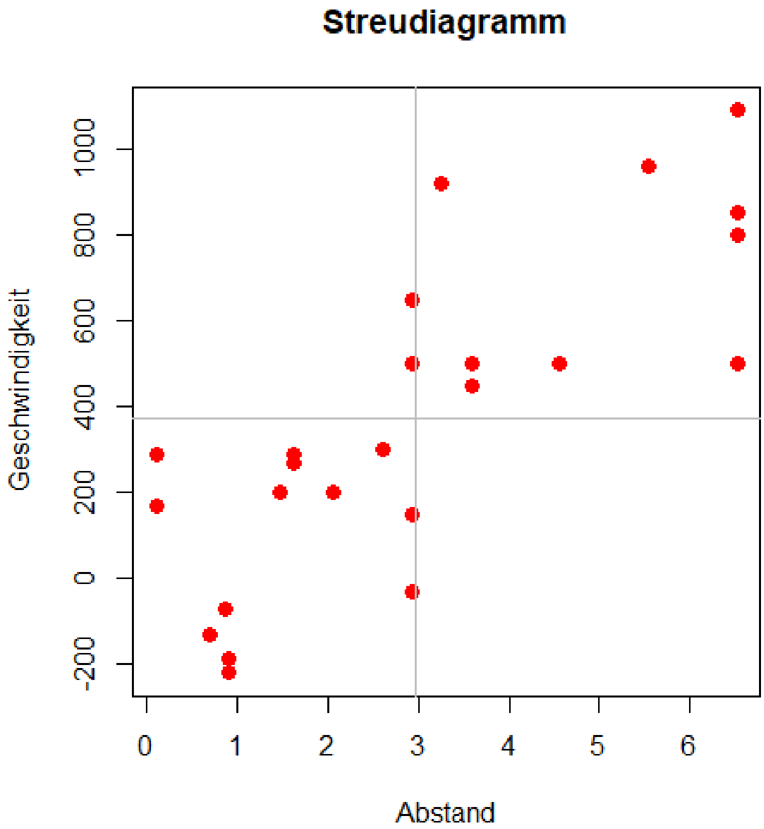
Werbung und Umsatz

Für diese Aufgabe finden Sie die Lösung sicher selbst, wenn Sie die Daten grafisch darstellen.

Lösung Übung ZHA-04a

Hubble

Die Kovarianz beträgt 616,87. Die Korrelation beträgt 0,79



Interpretation
Die positive Kovarianz lässt auf einen positiven Zusammenhang schließen. Das 4-Quadranten-Schema verdeutlicht das Ergebnis der Kovarianz: Fast alle Werte liegen im I. und III. Quadranten. Die Annahme eines Zusammenhangs wird durch den Korrelationskoeffizient bestätigt, dessen Wert von 0,8 eine starke Korrelation der Daten bedeutet. Das Streudiagram zeigt deutlich die gemeinsame Tendenz der x- und y-werte. Das heißt zusammengefasst, dass die Daten stark positiv korreliert sind.

Lösung mit R

hubble_loesung.R

```
# Einlesen der Daten aus der Datei hubble.txt
hubble<-read.table("hubble.txt",sep="\t",header=TRUE)

attach(hubble)

#  Bestimmung der Kovarianz und der Korrelation
round(cov(Abstand,Geschw),2)
round(cor(Abstand,Geschw),2)

#  Zeichnen eines Streudiagramms
plot(
  Abstand,
  Geschw,
  pch = 16,
  cex = 1.2,
  col = "red",
  xlab = "Abstand",
  ylab = "Geschwindigkeit",
  main = "Streudiagramm"
)
abline(h=mean(Geschw),col="aliceblue")
abline(v=mean(Abstand),col="aliceblue")
```

Lösung mit Excel

WMS_ZHA_04_Hubble.xlsx (12 KB)

Aufgabe 1: *Lässt sich ein Zusammenhang zwischen Geschwindigkeit und Abstand feststellen? Benutzen Sie zur Überprüfung geeignete Maßzahlen und bestätigen Sie Ihr Ergebnis grafisch. Interpretieren Sie Ihr Ergebnis.*

Um herauszufinden, ob es einen Zusammenhang zwischen Geschwindigkeit und Abstand gibt, berechnen wir die Kovarianz zwischen den beiden Variablen und den Korrelationskoeffizienten. Die Kovarianz wird mit der Funktion `KOVARIANZ.S` berechnet und der Korrelationskoeffizient mit der Funktion `KORREL`.

A31 | fx | =KOVARIANZ.S(B2:B25;C2:C25)

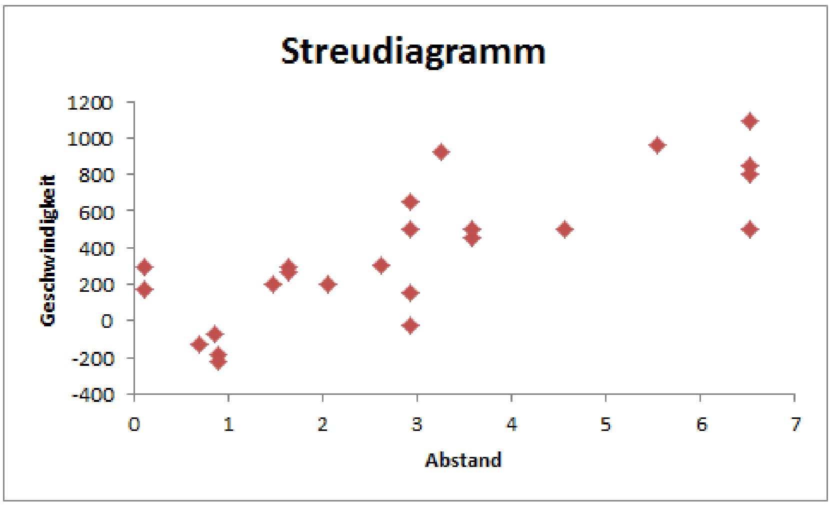
	A	B	C	D	E	F	G
29							
30	Berechnen der Kovarianz:						
31	616,8709						
32							
33	Berechnen des Korrelationskoeffizienten:						
34	0,789622						

A34 | fx | =KORREL(B2:B25;C2:C25)

	A	B	C	D	E	F
29						
30	Berechnen der Kovarianz:					
31	616,8709					
32						
33	Berechnen des Korrelationskoeffizienten:					
34	0,789622					

Dabei stehen in den Zellen B2 bis B25 die Werte für den Abstand und in den Zellen C2 bis C25 die Werte für die Geschwindigkeit. Das Ergebnis und die grafische Darstellung sehen so aus:

Berechnen der Kovarianz:				
616,87				
Berechnen des Korrelationskoeffizienten:				
0,79				



Interpretation: Positive Kovarianz und großer Korrelationskoeffizient besagen: Je größer der Abstand desto größer die Geschwindigkeit.

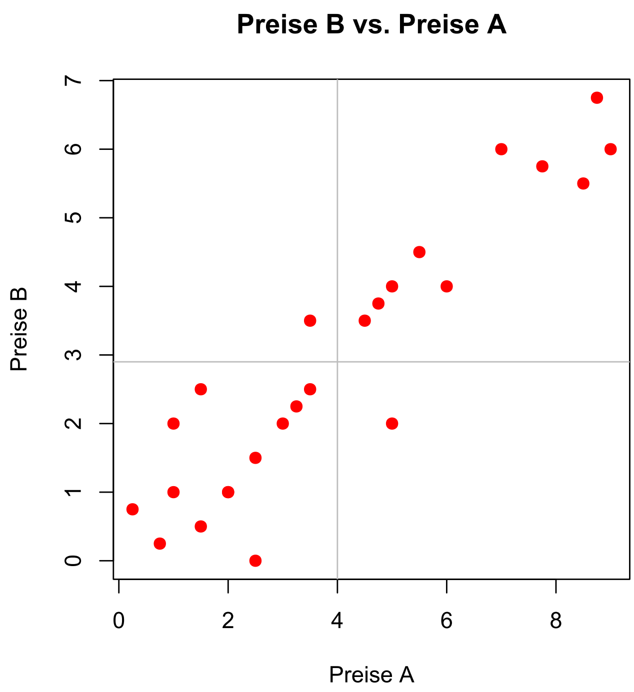
Die Grafik bestätigt das, denn je weiter rechts die Punkte liegen, desto weiter oben liegen sie auch. Erstellt wird sie über die Schaltfläche **`Einfügen`**, wo wir ein Punkt (XY)-Diagramm auswählen, für das wir die beiden Datenreihen Abstand und Geschwindigkeit (also die Zellen B2 bis B25 und C2 bis C25) markieren.

Lösung Übung ZHA-04b

Preise

Das arithmetische Mittel der Preise **A = 4**, der Preise **B = 2,9**

Die Kovarianz beträgt: 4,942708



Lösung mit R

preise_loesung.R

```
#  Preise - Aufgabe

#  Einlesen der Preise in Geschäft A und Geschäft B.
preise.a<-c(1.5,0.25,9,4.75,3.5,8.75,6,2,1,3.25,1,5.5,
            7.75,5,0.75,2.5,1.5,8.5,4.5,3,3.5,5,7,2,2.5)

preise.b<-c(0.5,0.75,6,3.75,2.5,6.75,4,1,2,2.25,1,4.5,
            5.75,4,0.25,1.5,2.5,5.5,3.5,2,3.5,2,6,1,0)

#  Berechnung der Mittelwerte
mean(preise.a)
mean(preise.b)

#  Berechnung der Kovarianz
cov(preise.a,preise.b)

#  4-Quadranten-Schema
plot(
  preise.a,
  preise.b,
  pch = 16,
  cex = 1.2,
  col = "red",
  main = "Preise B vs. Preise A",
  xlab = "Preise A",
  ylab = "Preise B"
)
abline(h=mean(preise.b),col="grey")
abline(v=mean(preise.a),col="grey")
```

Lösung mit Excel

`WMS_ZHA_04_Preise.xlsx` (13 KB)

***Aufgabe 1**: Berechnen Sie jeweils das arithmetische Mittel sowie die Kovarianz und veranschaulichen Sie die Kovarianz grafisch. Zeichnen Sie dazu das 4-Quadranten-Schema.*

Wir schreiben den Vektor der Preise von Geschäft A in die Zellen B2 bis B26 und den Vektor der Preise von Geschäft B in die Zellen C2 bis C26. Dann berechnen wir mit der Funktion `MITTELWERT` jeweils das arithmetische Mittel der Preise der beiden Geschäfte.

31	Arithmetisches Mittel:		
32	A:	=MITTELWERT(B2:B26)	
33	B:	=MITTELWERT(C2:C26)	
34			

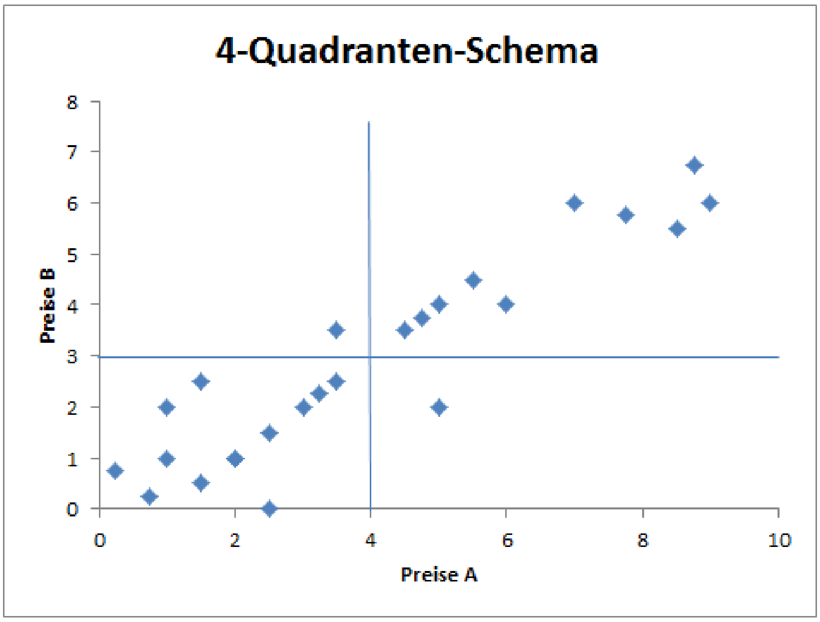
Arithmetisches Mittel:	
A:	4
B:	2,9

Die Kovarianz bestimmen wir wieder mit der Funktion `KOVARIANZ.S`.

35	Kovarianz:	
36	=KOVARIANZ.S(B2:B26;C2:C26)	

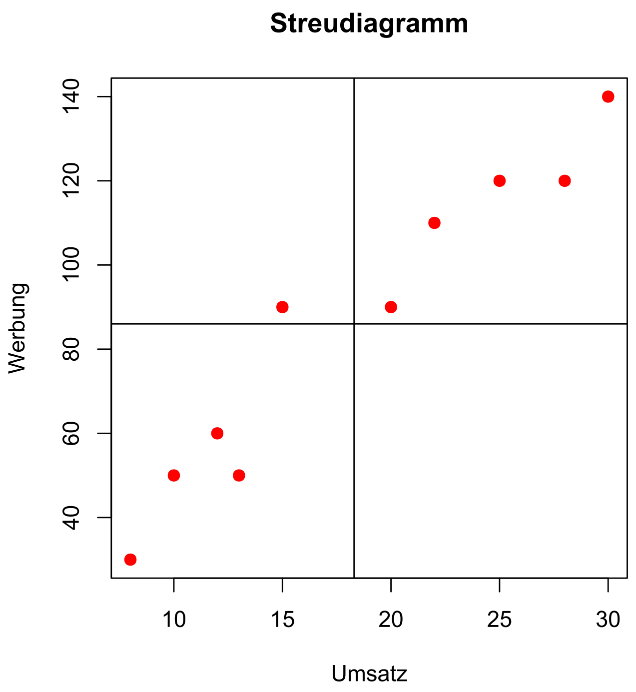
Kovarianz:
4,94

Die grafische Darstellung erhalten wir mittels eines Punkt (XY)-Diagramms. Es werden dafür die Zellen B2 bis B26 und C2 bis C26 markiert. Die Einteilung in die vier Quadranten erhalten wir, indem wir eine vertikale Linie beim Mittelwert der Preise von Geschäft A (also bei 4 auf der x-Achse) einzeichnen und eine horizontale Linie beim Mittelwert der Preise von Geschäft B (also bei 2.9 auf der y-Achse).



Lösung Übung ZHA-04c

Umsatz



Aufgabe 2: Die Art des Zusammenhangs ist positiv: Denn je größer der Umsatz, desto höher die Ausgaben für Werbung. Der Typ ist stochastisch, denn Unterschiede der Ausgaben für Werbung korrespondieren nicht perfekt mit den Unterschieden des Umsatzes. Die Ausgaben für Werbung hängen auch noch von anderen Variablen ab.

Lösung mit R

umsatz_loesung.R

```
#  Umsatz - Aufgabe

#  Einlesen der Werte für Umsatz und Werbung in einen Data Frame
daten <-
  data.frame(
    Umsatz = c(20, 28, 13, 15, 12, 25, 30, 10, 8, 22),
    Werbung = c(90, 120, 50, 90, 60, 120, 140, 50, 30, 110)
  )

#  Streudiagramm
plot(
  daten$Umsatz,
  daten$Werbung,
  pch = 16,
  cex = 1.2,
  col = "red",
  main = "Streudiagramm",
  xlab = "Umsatz",
  ylab = "Werbung"
)
abline(h=mean(daten$Werbung))
abline(v=mean(daten$Umsatz))
```

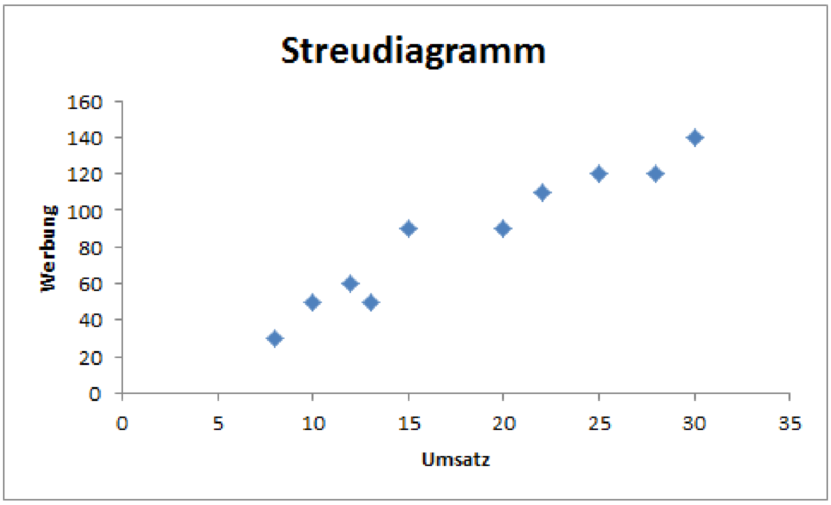
Lösung mit Excel

WMS_ZHA_04_Umsatz.xlsx (14 KB)

Aufgabe 1: *Zeichnen Sie ein Streudiagramm und geben Sie die Art sowie den Typ des Zusammenhangs an.*

Die grafische Darstellung erhalten wir wieder mittels eines Punkt (XY)-Diagramms. In der ersten Abbildung ist zu sehen, wie die Daten markiert werden müssen, darunter finden Sie das dazugehörige Streudiagramm.

	A	B	C	D
1		Umsatz	Werbung	
2	1	20	90	
3	2	28	120	
4	3	13	50	
5	4	15	90	
6	5	12	60	
7	6	25	120	
8	7	30	140	
9	8	10	50	
10	9	8	30	
11	10	22	110	
12				



Die Art des Zusammenhangs ist positiv: Denn je größer der Umsatz, desto höher die Ausgaben für Werbung. Der Typ ist stochastisch, denn Unterschiede der Ausgaben für Werbung korrespondieren nicht perfekt mit den Unterschieden des Umsatzes. Die Ausgaben für Werbung hängen auch noch von anderen Variablen ab.

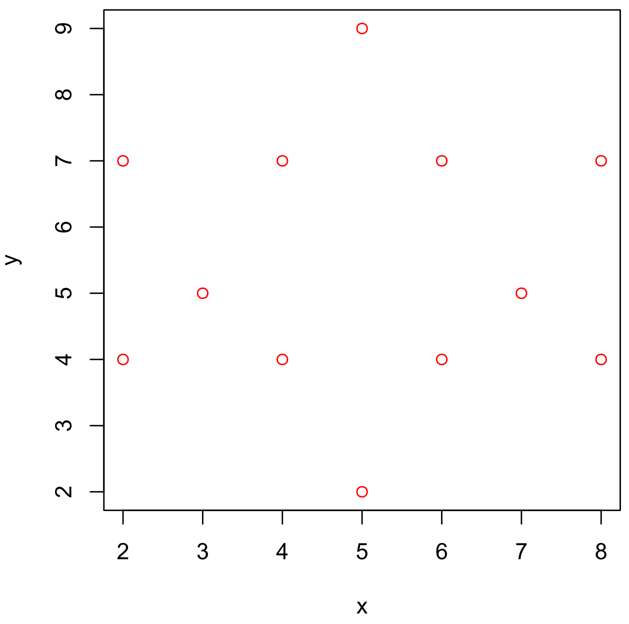
Lösung für Übung ZHA-05

Alltagsbeispiele für den funktionalen Zusammenhang

- Der Preis einer Ware hängt von der verkauften Menge ab.
- Die gemessene Außentemperatur hängt von der Tageszeit ab.
- Der zurückgelegte Weg eines Radfahrer/in hängt bei gleichbleibender Geschwindigkeit von der Fahrzeit ab.
- Der Bremsweg eines Fahrzeugs hängt im wesentlichen von seiner Geschwindigkeit ab.
- Das Wachstum eines Kindes hängt von seinem jeweiligen Alter ab.
- Die Note einer Mathematikarbeit hängt von der erreichten Punktzahl ab.
- Der Zinsertrag eines Kapitals hängt bei festem Zinssatz von der Laufzeit ab.

Lösung für Übung ZHA-06

Beobachtungen



Lösung mit R

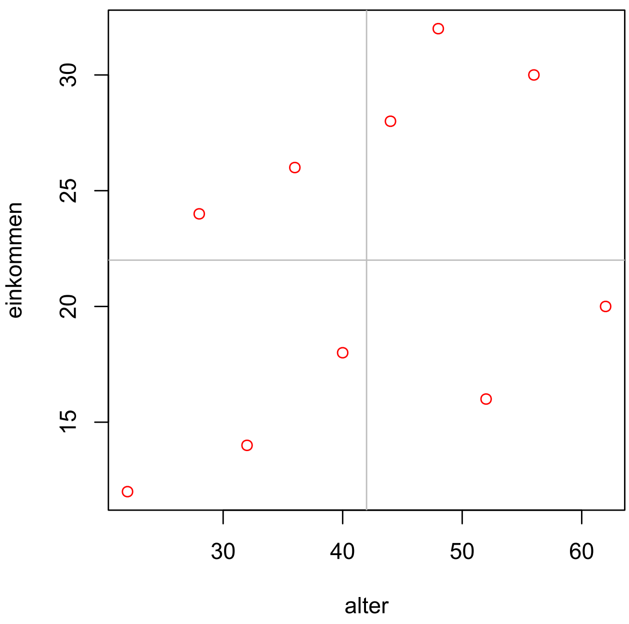
ZHA-Beobachtungen.R

```
# Einlesen der Daten
x<-c(2,3,4,5,6,8,7,2,4,6,8,5)
y<-c(4,5,4,2,4,4,5,7,7,7,7,9)

# Streudiagramm
plot(x, y, col = "red")
```

Lösung für Übung ZHA-07

Alter und monatliches Einkommen



Lösung mit R

ZHA-Einkommen.R

```
# Einlesen der Werte von Alter und Einkommen
alter<-c(22,28,32,36,40,44,48,52,56,62)
einkommen<-c(12,24,14,26,18,28,32,16,30,20)

(mean(alter))

(mean(einkommen))

plot(alter,einkommen, col="red")
abline(22,0,col="grey")
abline(0,0,v=42, col="grey")
```