11. Korrelation

Monovariate Daten

In Kapitel 4 und 5 haben wir uns mit Schätzern für Mittelwerte und Variationsmaße beschäftigt. Beispiele waren:

Objekt	Variable X
Ratte	Kreatinin-Clearance
Hühnerei	Masse
Mensch	systolischer Blutdruck

Den Beispielen ist gemeinsam, dass jeweils an einem Objekt (Hühnerei) eine Variable (Masse) ermittelt wurde. Derartige Daten (Variable X) nennen wir monovariate oder univariate Daten.

Bivariate Daten

In vielen Bereichen in denen statistische Methoden angewendet werden, finden wir Situationen, in denen an einem Objekt mehrere Variable ermittelt werden müssen. Beispiele sind Fragen nach einem Zusammenhang zwischen zwei Variablen.

Objekt	Variable X	Variable Y
Mensch	diastolischer Blutdruck	systolischer Blutdruck
Mensch	Alter	Körpergröße
Patient	Dosis eines Arzneistoffs	Stärke der therapeutischen Wirkung
Ratte	Masse des Körpers	Masse des Gehirns
Insekt	Masse des Körpers	Sauerstoffverbrauch
Bakterienkultur	Zellzahl/µL	Kulturtemperatur
DNA	„Molekülgröße“ in kbp	Laufstrecke bei der Elektrophorese

Datensätze, für die an einem Objekt zwei Variablen ermittelt wurden, nennen wir **bivariate Datensätze.**

Korrelation
Wenn wir für zwei Datensätze einen numerischen Zusammenhang finden, dann sprechen wir von einer **Korrelation** zwischen den Daten, die beiden Variablen sind korrelliert. Ein Maß für die Stärke der Korrelation ist z.B. der Korrelationskoeffizient r, ($-1 \le r \le +1$).
Je nach Datenstruktur gibt es verschiedene Korrelationskoeffizienten.

Von **multivariaten** Datensätzen sprechen wir, wenn mehr als zwei Variablen bei einem Objekt ermittelt werden, z.B.

Objekt	Variable U	Variable V	Variable W	Variable X	Variable Y	Variable Z
Mensch	Körpergröße	Körpermasse	$RR_{syst.}$	$RR_{diast.}$	Atemfrequenz	Pulsfrequenz

In diesem Kapitel werden wir uns mit **bivariaten Daten** beschäftigen.

Vorgehen bei der Prüfung auf einen Zusammenhang zweier Variablen.

Im Vorfeld einer solchen Untersuchung steht in der Regel ein Sachproblem. Beispielsweise wollen wir wissen, ob bei Hühnereiern ein numerischer Zusammenhang besteht zwischen dem Durchmesser und der Masse. Das grundsätzliche Vorgehen zur Klärung solcher Fragen besteht in folgenden Schritten:

1. Es wird die Hypothese aufgestellt, dass der Korrelationskoeffizient r = 0 ist, dass also keine Korrelation besteht.
2. Es wird experimentiert oder wir greifen retrospektiv auf Daten zurück.
3. Die Daten werden in einer Graphik auf Hinweise für eine Korrelation untersucht.
4. Der Korrelationskoeffizient wird berechnet.
5. Es wird geprüft, ob die Hypothese falsifizierbar ist.

Da eine Einführung in die Hypothesenprüfung erst in einem späteren Kapitel folgt, untersuchen wir hier nur Punkt 3 wobei folgende Fragen zu beantworten sind:

1. Besteht eine Korrelation zwischen der Masse (Variable X) und dem Durchmesser (Variable Y)?
2. Welcher Art ist eine eventuelle Korrelation? (Zu „Art“ siehe unten.)
3. Wie stark ist eine eventuelle Korrelation? (Zu „stark“ siehe unten.)

Beispiel 1 soll zeigen, wie wir zur Beantwortung dieser Fragen vorgehen. (Die Berechnungen zu den Daten folgen in Kapitel 12.) Wir stellen zunächst die Daten in Tabellenform und dann als Streudiagramm dar.

Beispiel 1
Die Tabelle
In Tabelle 1 liegen die Massen und die Durchmesser von 38 Hühnereiern in rangierter Folge vor. Masse und Durchmesser sind, da empirisch ermittelt, Zufallsvariable. Bei genauer Betrachtung der Tabelle mit den rangierten Daten fällt auf, dass mit steigenden X-Werten meist auch die Y-Werte größer werden. In der Urtabelle, die die Daten in zufälliger Folge enthält, zeigt sich dieser Zusammenhang auf den ersten Blick nicht so schnell. (Zur Frage, warum der Masse die Variable X und dem Durchmesser Y zugewiesen wurde, siehe bei Beispiel 4.)

Nr.	X Masse (g)	Y Durchmesser (mm)	Nr.	X Masse (g)	Y Durchmesser (mm)	Nr.	X Masse (g)	Y Durchmesser (mm)	Nr.	X Masse (g)	Y Durchmesser (mm)
1	53.3	41.2	11	57.9	43.4	21	60.6	43.2	31	61.9	43.9
2	56.0	42.0	12	58.3	43.2	22	60.6	43.3	32	62.0	43.5
3	56.1	42.7	13	58.4	42.6	23	60.7	43.1	33	62.1	43.1
4	56.3	42.2	14	59.1	43.5	24	61.2	43.9	34	62.2	43.9
5	56.6	43.0	15	59.4	42.5	25	61.2	44.1	35	62.3	42.6
6	57.0	42.9	16	60.0	43.8	26	61.3	44.2	36	62.3	43.6
7	57.2	42.0	17	60.0	42.8	27	61.5	43.5	37	62.4	44.3
8	57.4	42.3	18	60.2	43.2	28	61.6	42.9	38	64.0	45.0
9	57.7	42.1	19	60.4	43.3	29	61.6	43.8			
10	57.7	42.8	20	60.5	43.5	30	61.8	43.2			

Tabelle 1

Das Streudiagramm (x/y-Diagramm)
Besser als in der Tabelle erkennen wir einen Zusammenhang durch Visualisierung der Daten in Form eines Streudiagramms. In Abb.1 sehen wir sofort, dass mit steigendem X tendenziös auch Y steigt. Das bedeutet, dass ein korrelativer Zusammenhang zwischen den beiden Variablen besteht. Oder vorsichtiger gesagt, dass ein Hinweis auf einen solchen Zusammenhang zu erkennen ist.

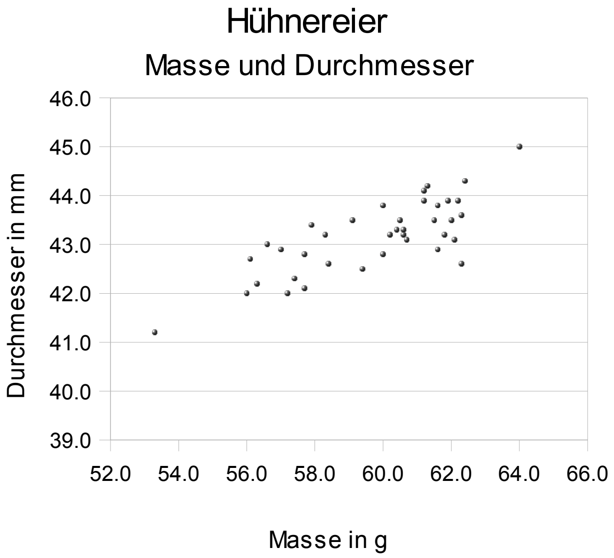


Abb.1

Ein <u>Vorteil der Graphik</u> besteht auch darin, dass <u>Extremwerte</u> wie $P_{x=53,3;y=41,2}$ und $P_{x=64;y=45}$ in Abb.1 schnell erkannt werden. Auch wird schnell deutlich, wenn <u>unplausible Werte</u> auftauchen. Läge für den Wert x = 53,3 der y-Wert nicht bei 41,2 sondern bei 45,5, so würde das in der Tabelle wahrscheinlich zunächst kaum, in der Graphik aber sofort als unplausibel auffallen.

Die Punktwolke zeigt eine Tendenz zur Linearität. Wir passen ihr daher eine Gerade an, die wie in Abb.2 liegen könnte. Die Lage dieser Geraden wurde über ein Tabellenkalkulationssystem berechnet. Wie diese Anpassung durch eine „von Hand“-Berechnung durchgeführt wird, sehen wir im Zusammenhang mit der Regression in Kapitel 13.

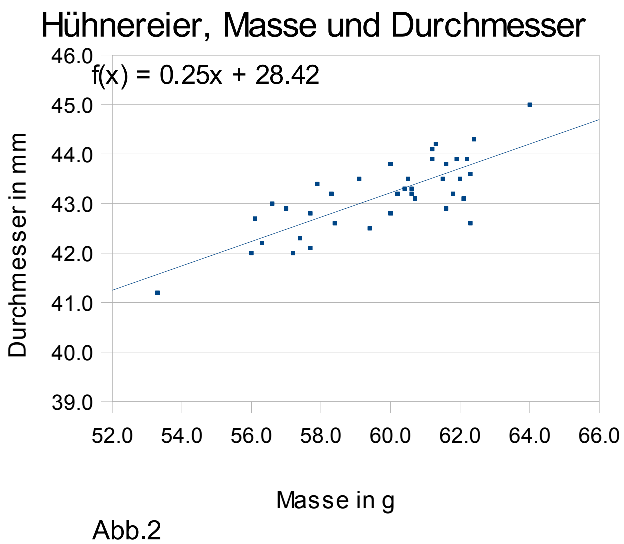


Abb.2

Aus Beispiel 1 führen zwei Fragestellungen zur Korrelation und Regression.

Fragen zur Korrelation

* **Gibt es einen numerischen Zusammenhang zwischen der Masse und dem Durchmesser?**
In Abb.1 deutet die Punktwolke an, dass mit steigender Masse die Durchmesser steigen. Die Gerade in Abb.2 verdeutlicht dies und weist auf eine lineare Korrelation hin.
* **Wie stark ist dieser Zusammenhang?**
Diese Frage wird quantitativ durch eine Rechnung beantwortet, deren Ergebnis der Korrelationskoeffizient r ist. Die Korrelation ist um so stärker, je näher die Punkte an der Geraden liegen und je näher der Koeffizient r bei |1| liegt.

Die Korrelation beschäftigt sich mit der Stärke und Richtung des Zusammenhangs zweier Zufallsvariablen.

Fragen zur Regression

* **Falls eine Korrelation besteht, um wie viel mm nimmt der Durchmesser zu, wenn die Masse um 1 g steigt?**
* **Um wie viel g nimmt die Masse zu, wenn der Durchmesser um einen mm steigt?**
Diese Fragen werden quantitativ durch eine Rechnung beantwortet, deren Ergebnis der Regressionskoeffizient b ist. Mit b können wir den Wert der einen Variablen aus der Kenntnis eines Wertes der anderen Variablen berechnen.

Die Regression beschäftigt sich mit der Schätzung von Y-Werten durch X-Werte.

11.1 Korrelationen

Bei der Untersuchung bivariater Daten können verschiedene Korrelationsformen auftreten. In der Literatur finden wir verschiedene Möglichkeiten diese zu strukturieren. Wir unterscheiden hier:

Nichtlineare Korrelationen - Lineare Korrelationen
Positive Korrelationen - Negative Korrelationen
Unterschiedlich starke Korrelationen
Kausale Korrelationen - Formale Korrelationen
Funktionale Korrelationen - Stochastische Korrelationen

11.1.1 Lineare und nichtlineare Korrelationen

Dem Verlauf der Punkte in Abb.3,4,5 kann keine Gerade angepasst werden, weil keine Tendenz zur Linearität zu erkennen ist. Es handelt sich um nichtlineare Korrelationen. Um solche Daten für die Rechentechniken der linearen Korrelationen nutzbar zu machen, können sie u.U. durch Transformation linearisiert werden. So korrelieren z.B. bei einem Lochplattentest in der Mikrobiologie die Konzentrationen eines Antibiotikums mit den resultierenden Hemmhofdurchmessern nicht linear. Dagegen sind die Logarithmen der Konzentrationen und die Hemmhofdurchmesser linear korreliert.

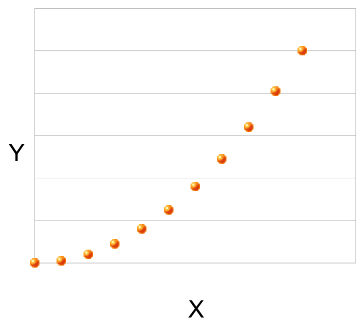


Abb.3

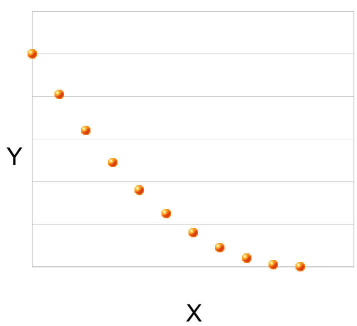


Abb.4

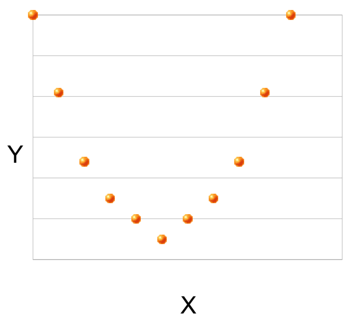


Abb.5

Bei Abb.6 und 7 handelt es sich um lineare Korrelationen. Abb.6 ist monoton steigend, das bedeutet, jedes Ansteigen eines X-Wertes hat auch ein Ansteigen des entsprechenden Y-Wertes zur Folge. Abb.7 steigt nicht monoton. Die angepasste Gerade ist zwar steigend aber an einigen Stellen folgt einem steigenden X-Wert ein fallender Y-Wert. Abb.8 zeigt eine partiell lineare Funktion in deren oberem Teil die Punktfolge von der Linearität abweicht.

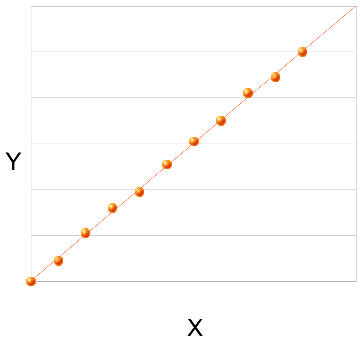


Abb.6

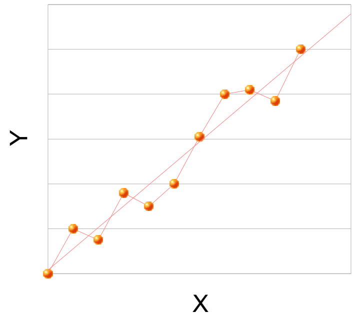


Abb.7

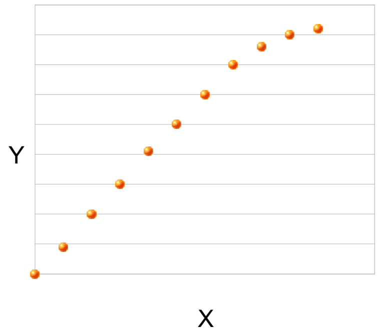


Abb.8

In der Folge beschränken wir uns auf lineare Korrelationen.

11.1.2 Positive und negative lineare Korrelationen

Korrelationen, bei denen Y steigt wenn X steigt – ob monoton oder nicht monoton – nennen wir positive Korrelationen.

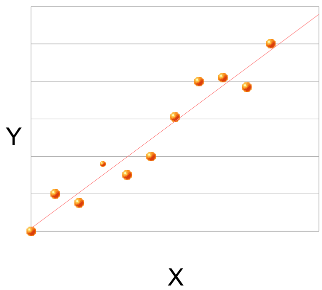


Abb.9

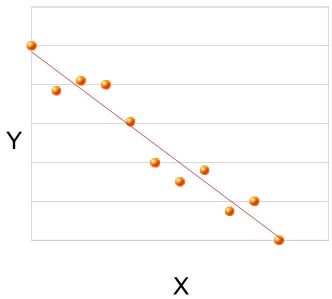


Abb.10

Abb. 6 bis 9 zeigen solche. Abb.10 stellt eine negative Korrelation dar. Hier fallen tendenziös die Y-Werte, wenn die X-Werte steigen.

11.1.3 Stärke linearer Korrelationen

Je nachdem, wie nahe die Punkte an der Geraden liegen, unterscheiden wir schwächere und stärkere Korrelationen. Ein Maß für die Stärke wird durch den Korrelationskoeffizienten r ($-1 \leq r \leq +1$) angegeben. Je näher der Wert bei |1| liegt, um so stärker ist die Korrelation. Bei positiven Korrelationen ist r positiv, bei negativen Korrelationen negativ. Wenn $r = 0$, dann sind die Werte nicht korreliert. Die r-Werte in Abb. 11 bis 15 wurden über eine Tabellenkalkulation berechnet.

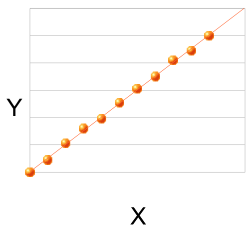


Abb.11
$r = +1$

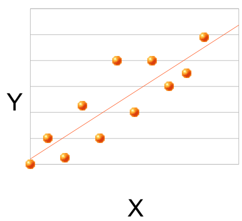


Abb.12
$r = +0{,}86$

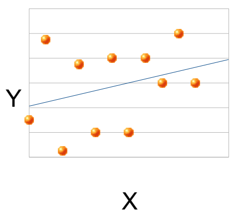


Abb.13
$r = +0{,}32$

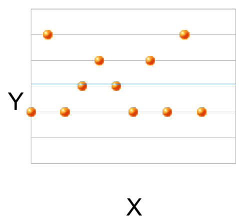


Abb.14
$r = 0$

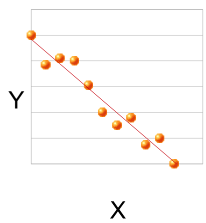


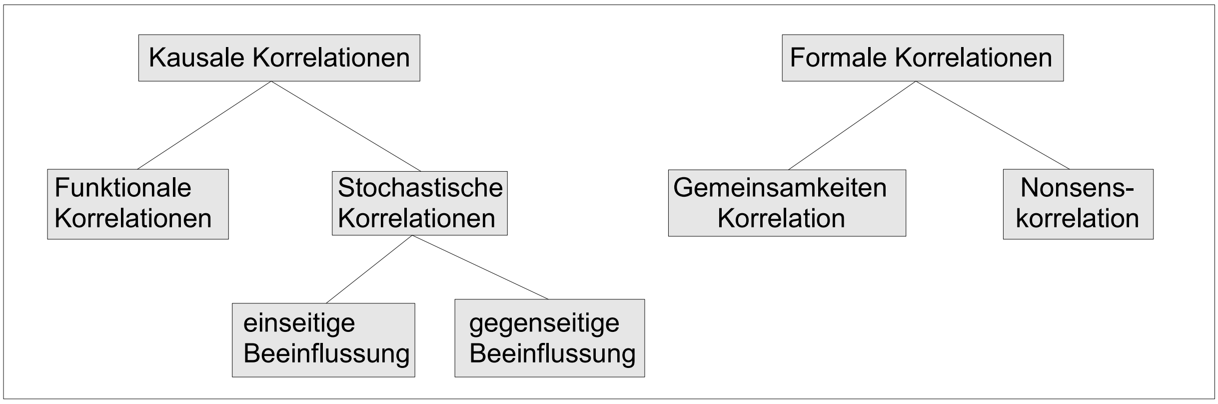
Abb.15
$r = -0{,}97$

11.1.4 Kausale und formale Korrelationen

Bei korrelativen Zusammenhängen unterscheiden wir kausale und formale Korrelationen.

Kausale Korrelationen
Wir finden einen numerischen Zusammenhang zwischen der Dosis eines pharmakologischen Wirkstoffs und der Stärke des therapeutischen Effekts. Der Zusammenhang erscheint uns plausibel, da wir die Pharmakodynamik des Wirkstoffs kennen. Wir erkennen einen ursächlichen Zusammenhang zwischen beiden Variablen und sprechen dann von einer **kausalen Korrelation**.

Formale Korrelationen
Wir finden für eine bestimmte Zeitspanne einen numerischen Zusammenhang zwischen den Verkaufszahlen von Computern und dem Import einer bestimmten Automarke. Das gemeinsame „Objekt“ an dem die Variablen ermittelt wurden, ist dann der Wirtschaftsraum. Einen ursächlichen Zusammenhang zwischen den beiden Variablen („PC“ und „PKW“) können wir uns nicht vorstellen. Möglicherweise gibt es einen Zusammenhang, aber keinen kausalen. Wir kämen sicher nicht auf die Idee zu folgern, dass die Autoimporte den PC-Absatz beeinflussen oder umgekehrt. Solche Korrelationen werden **formale Korrelationen** genannt.



Korrelation und Kausalität
Ein bei der Interpretation von Korrelationen häufig begangener Fehler – sei es unbewusst oder bewusst – ist die Annahme, eine numerische Korrelation begründe eine Kausalität. Eine Kausalität mag zwar in Einzelfalle vorliegen, kann aber nicht daraus abgeleitet werden, dass die Daten korreliert sind. Ob bei einer Korrelation ein kausaler Zusammenhang besteht, kann mit den Methoden der Statistik nie geklärt werden. Hier ist sachlogisch die Plausibilität zu prüfen. Der Korrelationskoeffizient ist nur eine statistische Maßzahl, die nichts mit Ursache und Wirkung zu tun hat.

Mit statistischen Verfahren können Kausalzusammenhänge nicht nachgewiesen werden.

11.1.4.1 Kausale Korrelationen

Hier unterscheiden wir zwischen funktionalen Korrelationen und stochastischen Korrelationen

11.1.4.1.1 Funktionale Korrelationen

Beispiel 2
Von dem Farbstoff X liegt eine **berechnete** arithmetische Konzentrationsreihe (c in mol L^{-1}) mit dem Summanden 0,01 vor. Der molare Extinktionskoeffizient ist ε = 140 L mol^{-1} cm^{-1}. Die Schichtdicke der Photometerküvette ist d =1 cm. Nach dem Lambert-Beer-Gesetz ($\varepsilon = E\,c^{-1}\,d^{-1}$) wurden für jede Konzentration im Gültigkeitsbereich des Gesetzes die Extinktion **berechnet**. Vor dem physikalischen Hintergrund, auf den wir hier nicht eingehen, liegt eine kausale Korrelation vor. Zur Berechnung der Extinktion für c = 0,0034 mol L^{-1} stellen wir die Gleichung von $\boldsymbol{\varepsilon = E\,c^{-1}\,d^{-1}}$ nach $\boldsymbol{E = \varepsilon\,c\,d}$ um, dann ist

$E = \varepsilon\, c\, d$
$E = 140\ L\ mol^{-1}\ cm^{-1} * 0{,}0034\ mol\ L^{-1} * 1\ cm$
$\underline{E = 0{,}4760}$

Tabelle 2 zeigt die **berechneten** Extinktionen für alle Konzentrationen und in Abb.16 sehen wir, dass die Punkte alle exakt auf der Geraden liegen. Hier sprechen wir von einer funktionalen Korrelation. Sie basiert auf einem Zusammenhang gemäß der Funktion f(c) = E = ε c d. Solche funktionsgebundenen Zusammenhänge sind eindeutig umkehrbar, d.h. wir können zu jedem X-Wert den Y-Wert (und umgekehrt) exakt berechnen. Für den Korrelationskoeffizienten gilt in diesem Beispiel r = +1. Allgemein gilt für funktionale Korrelationen r = |1|.

c mol L^{-1}	E Extinktion **berechnet**
0,0034	0,4760
0,0044	0,6160
0,0054	0,7560
0,0064	0,8960
0,0074	1,0360
0,0084	1,1760
0,0094	1,3160
0,0104	1,4560
0,0114	1,6960
0,0124	1,7360

Tabelle 2

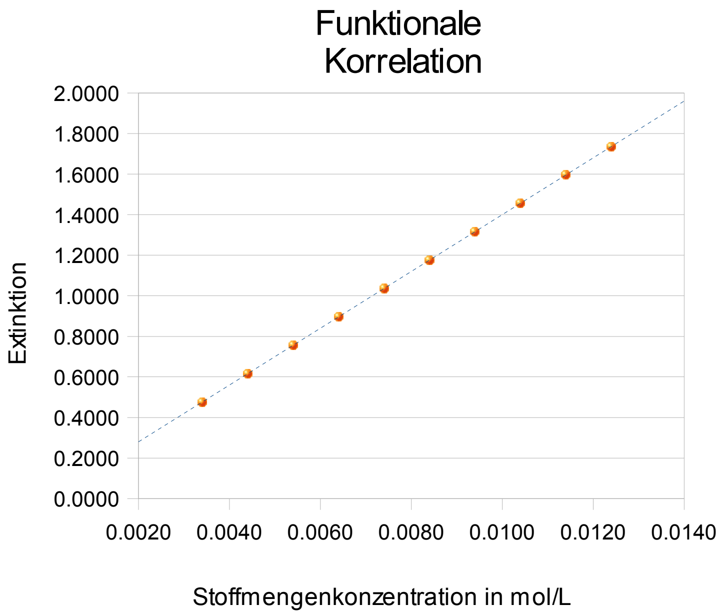


Abb.16

11.1.4.1.2 Stochastische Korrelationen

Wir nehmen die gleiche Konzentrationsreihe wie vorhin, nur stellen wir jetzt die Lösungen praktisch her anstatt sie nur theoretisch vorliegen zu haben. Und wir messen die Extinktionen an einem Spektralphotometer bei $\lambda_{max.}$ statt sie zu berechnen. Die Ergebnisse liegen in Tabelle 3 vor.

c[mol $^{L-1}$	0.0034	0.0040	0.0054	0.0064	0.0074	0.0084	0.0094	0.0104	0.0114	0.0124
E Extinktion berechnet	0.4650	0.6000	0.8000	0.8990	1.0000	1.1790	1.3900	1.4600	1.5900	1.7400

Tabelle 3

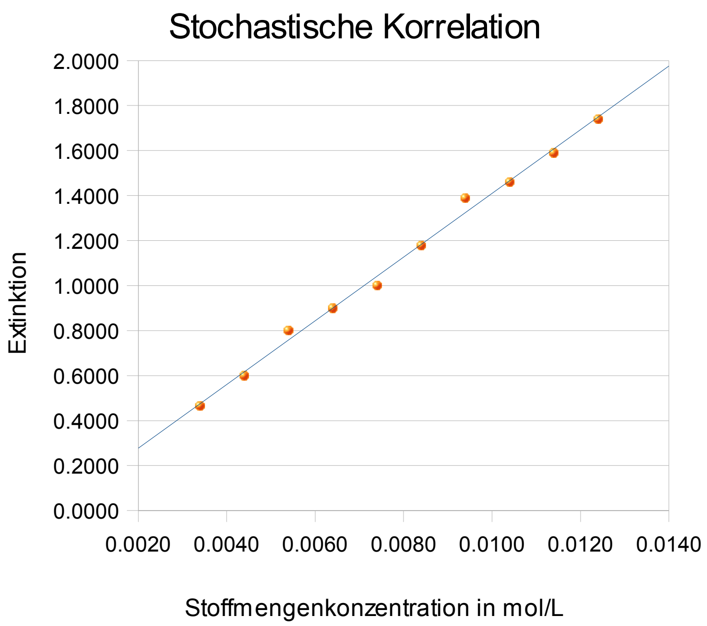


Abb.17

Konzentrationen und Extinktionen sind hier, da experimentell gewonnen, durch Messunsicherheit behaftete Zufallsvariablen. Dies zeigt sich in Abb.17, in der die Punkte zwar eine deutliche Tendenz zur Linearität zeigen aber nicht exakt auf der Ausgleichsgeraden liegen. Sie streuen um diese. E und c sind kausal linear korreliert. Wenn mindestens eine der Variablen durch Messunsicherheit behaftet ist, sprechen wir von stochastischen Korrelationen. Eine Vorhersage von Y-Werten aus X-Werten bzw. umgekehrt ist nur mit der messtechnisch bedingten Unsicherheit möglich. Für den Korrelationskoeffizienten gilt $-1 < r < 1$.

Regressor und Regressand

Bei der Analyse korrelierter Daten tritt die Frage auf, welche Variable die andere steuert (beeinflusst). Untersuchen wir die Stärke der Wirkung eines Medikamentes im Zusammenhang mit der Höhe der Dosis, so steht die Dosis als steuernde Variable eindeutig fest, denn sie beeinflusst die Wirkung, nicht umgekehrt. So eindeutig ist das aber nicht immer. Wenn wir in der klinischen Chemie die Stoffmengenkonzentrationen des Kaliumions und die des Calciumions im Seren bestimmen, dann ist die Zuordnung einer der beiden Variablen als steuernde Variable nicht so eindeutig möglich. Es ist üblich, der steuernden Variablen (Regressor) ein X zuzuordnen. Die Werte dieser Variablen werden in der Graphik auf der Abszisse aufgetragen. Die gesteuerte Variable (Regressand) heißt Y und wird auf der Ordinate dargestellt. Wie wir X und Y im Falle von Unsicherheit (K^+-Ca^{++}) zuordnen, wird im Zusammenhang mit Beispiel 4 erklärt.

Korrelationen mit <u>einseitiger</u> Beeinflussung

Beispiel 3
Regressor ist bekannt.
Tabelle 4 enthält die Massen von sechs Mäusen verschiedener Altersgruppen. Abb.18 zeigt die Punkteschar mit der Geraden. Da die Masse und das Alter der Tiere als Zufallsvariable unsicher sind, liegen die Punkte also nicht exakt auf der Geraden (Stochastische Korrelation). Das **Alter ist der Regressor** und die **Masse der Regressand**. Wenn eine Maus älter wird, nimmt ihre Masse in der Wachstumsphase in der Regel zu. Allein eine Gewichtszunahme lässt die Maus dagegen nicht älter werden.

X	Alter in d	10	11	18	2	6	17
Y	Masse in g	4.7	5.2	6.6	1.8	3.5	6.4

Tabelle 4

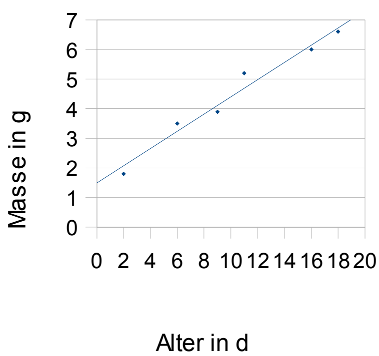
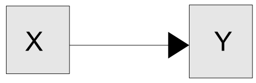
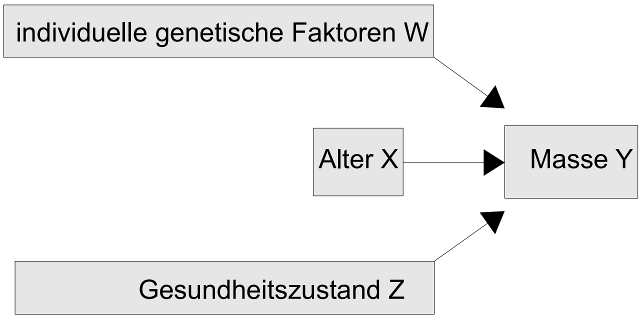


Abb.18

In vielen Fällen der empirischen Datengewinnung sind Regressor und Regressand eindeutig erkennbar. Wenn X die einzige Einflussgröße auf Y ist, können wir das so formulieren.



Bei näherer Betrachtung fällt aber auf, dass auch andere Variable einen Einfluss auf die Masse haben können. So können etwa genetische Faktoren, der Gesundheitszustand oder die Futterqualität die Gewichtszunahme beeinflussen (Komplexe Beeinflussung, multivariate Datensätze).



Beispiel 4
Regressor ist nicht eindeutig festzulegen.
Bei einem Patienten wurden über mehrere Wochen die „Kaliumwerte" und die „Calciumwerte" im Serum bestimmt. Tabelle 5 zeigt die Daten.

Kaliumionen mval/L	Calciumionen mval/L	Kaliumionen mval/L	Calciumionen mval/L
4.3	4.37	4.9	4.02
4.6	4.30	4.1	4.40
5.9	4.22	5.9	4.70
4.9	4.02	5.9	5.64
5.4	5.42	5.9	4.70
5.9	4.96	6.2	4.98
5.4	5.42	4.1	4.40

Tab.5

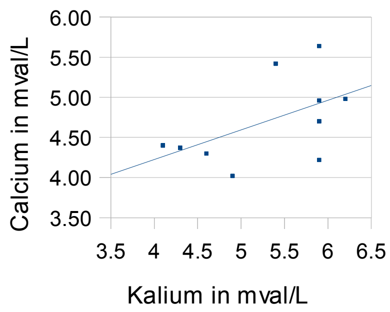


Abb.19

Hier ist zunächst nicht zu sagen, ob die Kaliumwerte die Calciumwerte beeinflussen oder umgekehrt, wer also der Regressor und wer der Regressand ist. Um solche Wertepaare in eine Graphik eintragen zu können, weisen wir einer der beiden Variablen arbiträr, d.h. willkürlich, das Zeichen X (oder X_1) und der anderen Y (oder X_2) zu und erhalten so die Abb.19, die auf eine positive Korrelation hinweist.

Korrelationen mit gegenseitiger Beeinflussung

Neben der in Beispiel 3 gezeigten Korrelation, bei der Y von X beeinflusst wird ($X \longrightarrow Y$), sind auch Situationen der Form $X \longleftrightarrow Y$ denkbar, bei denen sich die beiden Variablen gegenseitig beeinflussen.

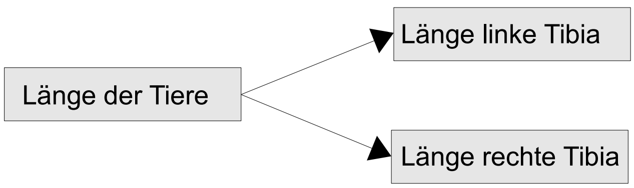
Beispiel 5
Bei einem Lernversuch in der Skinner-Box muss eine Ratte lernen, durch Druck auf die richtige von mehreren Tasten einen Strom abzuschalten, der an ihren Füßen ein Kribbeln [wir nehmen an, dass sie das so empfindet] verursacht. Dies erreicht sie nur, wenn sie auf die richtige Taste drückt. Irgendwann drückt sie zufällig auf eine Taste aber es ist die Falsche. Es kribbelt weiter. Dann trifft sie zufällig die richtige Taste und das Kribbeln in den Füssen hört auf. Bei Wiederholungen nach neuerlichem Einsetzen des Kribbelns macht sie immer weniger Fehler bei der Tastenwahl, sie lernt also, dass ein Zusammenhang zwischen dem Ende des Kribbelns und dem Druck auf die richtige Taste besteht. Sie ist nun motiviert, die als richtig erkannte Taste zu drücken. Auf die Motivation (X_1) „durch Tastendruck ein Beenden des Kribbelns zu erreichen“ folgt immer häufiger der Erfolg (X_2) nämlich „der Druck auf die richtige Taste“. Hier führt die Motivation zunächst zum Erfolg und der Erfolg führt dann zur Steigerung der Motivation. Wenn es dem Experimentator gelingt, Daten zur Motivation und zum Erfolg in geeigneter Form zu ermitteln, kann er eine Korrelationsanalyse durchführen.

Gerade im Bereich des Lernens sind Korrelationen mit gegenseitiger Beeinflussung häufig.

11.1.4.2 Formale Korrelationen

Im Gegensatz zu kausalen Korrelationen stehen formale Korrelationen. Wenn wir bei der Datenanalyse einen korrelativen Zusammenhang finden, für den offensichtlich keine sachliche Erklärung zu erkennen ist, dann kann das zwei Gründe haben.

Gemeinsamkeitenkorrelationen
Es kann sich um eine Korrelation handeln, bei der kein direkter Zusammenhang zwischen den beiden Variablen besteht. Wenn wir die Zunahme des Gewichts von Grundschulkindern und die Zunahme deren Rechenfähigkeit auf einen numerischen Zusammenhang hin untersuchen, werden wir eine positive Korrelation finden. Mit der Zunahme des Gewichtes nimmt auch die Rechenfähigkeit zu. Ein kausaler Zusammenhang zwischen Gewicht und Rechenfertigkeit besteht aber sicher nicht. Keine der beiden Variablen hat einen Einfluss auf die andere. Es wäre unsinnig zu sagen, eine Verbesserung der Rechenfähigkeit hinge ursächlich mit einer Gewichtszunahme zusammen oder umgekehrt. Das Alter der Kinder ist eine gemeinsame Einflussgröße. In der Regel werden die Kinder schwerer <u>und</u> lernen durch Übung besseres Rechen, wenn sie älter werden. Es gibt viele Situationen in denen zwei Variable sich nicht gegenseitig beeinflussen sondern von einer dritten Variablen beeinflusst werden. Wenn wir die Längen der Tibiae des mittleren rechten und linken Beines von Käfern (Zophobas morio) messen und prüfen, ob die Längen der beiden Tibiae korrelieren, dann werden wir eine positive Korrelation finden. Auch hier können wir nicht sagen, dass die Länge der linken Tibia die Länge der rechten beeinflusst oder umgekehrt. Möglicherweise steuert hier die Körperlänge der Tiere die Länge beider Tibiae.



In manchen Fällen ist es nicht so einfach, die gemeinsame Beeinflussungsvariable zu erkennen. Wie bei dem vielzitierten Fall des numerischen Zusammenhangs zwischen der Abnahme der Zahl der Storchennester und der Abnahme der Geburtenzahlen in verschiedenen Gegenden Europas in einen bestimmten Zeitraum. (Zahlenangaben über Google „Störche Geburten“.) Je nach den verwendeten Daten resultieren Korrelationskoeffizienten zwischen 0,6 und 0,9. Keiner würde annehmen, dass die Anzahl der Störche in einem kausalen Zusammenhang mit der Geburtenzahl steht. Die scherzhafte Schlussfolgerung, die Störche brächten die Kinder, beruht auf der falschen Annahme, dass ein nachgewiesener numerischer Zusammenhang ein Indiz für einen Kausalzusammenhang wäre. Wir wissen, dass das falsch ist. Gründe für die Korrelation liegen möglicherweise darin, dass andere Faktoren sowohl die Nistgewohnheiten der Störche wie auch den Kinderwunsch junger Paare beeinflussen.

Nonsenskorrelationen
Bei dem numerische Zusammenhang zwischen PC-Absatz und PKW-Import könnte ein Ökonom möglicherweise eine gemeinsame steuernde Variable entdecken. Irgendwie, so lesen wir hin und wieder, soll ja alles, auf welchen Wegen auch immer, miteinander zusammenhängen. Aber es könnte sich auch um ein rein zufälliges Zusammenkommen der Zahlen handeln. In solchen Fällen sprechen wir dann von einer sinnlosen Korrelation.

In Kapitel 12 werden wir uns mit der Berechnung von Korrelationskoeffizienten beschäftigen.