

Mathe 2

Corona - Notversion

M. Oettinger

14.05.2020

Bisher wurden stets Beispiele von Funktionen betrachtet, die von einer einzelnen Variablen x abhängen. Sie wurden als Modelle benutzt, um Abhängigkeiten (Zusammenhänge) zwischen (meist) technischen Größen darzustellen. Oft treten aber in der Anwendung auch Größen auf, die von zwei oder mehreren Variablen abhängen, was sich recht einfach behandeln lässt, indem man den Begriff der Funktion erweitert und mehrere unabhängige Variable zulässt.

Beispiel: Ohmsches Gesetz

Nach dem ohmschen Gesetz ist die an einem (ohmschen) Widerstand abfallende Spannung U der eingepprägten Stromstärke I proportional, wir bezeichnen die Proportionalitätskonstante mit dem Formelzeichen R

Beispiel: Ohmsches Gesetz

$$U = R \cdot I \text{ mit } R = \text{const.}$$

Man kann die Spannung U als eine Funktion der Stromstärke I betrachten:

$$U = f(I) = U(I).$$

Auch bei nicht ohmschem Verhalten eines Systems kann man sich natürlich eine Größe 'Widerstand' als das Verhältnis U/I definieren - der Unterschied ist schlicht, dass diese Größe R nicht mehr konstant ist, sie kann von der Stromstärke I abhängen, muss dies aber nicht. Damit hängt die Spannung im allgemeinen Fall von der Stromstärke I und dem (nicht mehr konstanten) Widerstand ab:

$$U = f(R, I) = U(R, I)$$

Wir können den Begriff der Funktion etwas erweitern, indem wir mehrere voneinander unabhängige variable Größen als Argument einführen:

Definition:

Unter einer Funktion von zwei unabhängigen Variablen versteht man eine Abbildungsvorschrift, die jedem geordneten Zahlenpaar $(x; y)$ aus einer Menge D genau ein Element z aus einer Menge W zuordnet. Man schreibt dafür symbolisch

$$z = f(x, y) \quad (1)$$

Wir bezeichnen die Mengen D bzw. W wieder als Definitions- bzw. Wertebereich der Funktion f . Der Definitionsbereich der Funktion (geordnete, voneinander unabhängige Zahlenpaare) bildet eine flächenhafte Punktmenge der x, y -Ebene.

Beispiel:

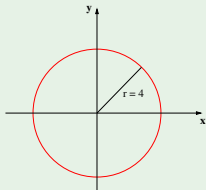
Die reellwertige Funktion zweier Variablen

$$z = z(x, y) = \sqrt{16 - x^2 - y^2} \in \mathbb{R}$$

besitzt als Definitionsbereich einen Kreis mit Radius $r = 4$ in der x, y -Ebene, denn der Ausdruck unter der Wurzel muss ≥ 0 sein.

$$16 - x^2 - y^2 \geq 0 \iff x^2 + y^2 \leq 16,$$

Damit kann der Wertebereich W bestimmt werden: $0 \leq z \leq 4$
(wegen $0 \leq 16 - x^2 - y^2 \leq 16$).



Beschränkt man sich auf Funktionen von zwei unabhängigen Variablen (ein in der Anwendung häufiger Fall), so kann eine Funktion grafisch recht einfach dargestellt werden.

Durch die Funktionsgleichung $z = f(x, y)$ wird jedem Zahlenpaar aus dem Definitionsbereich der Funktion genau ein Funktionswert $z_0 = f(x_0, y_0)$ zugeordnet. Wir können die drei Zahlen x_0, y_0 und z_0 als kartesische Koordinaten eines Punktes im dreidimensionalen Raum deuten, dem ein rechtwinkliges Koordinatensystem zugrunde liegt. Der Funktionswert z_0 kann dann geometrisch als Höhenkoordinate oberhalb der x, y -Ebene interpretiert werden.

Grafische Darstellung

Ordnet man jedem Zahlenpaar $(x; y) \in D$ einen Punkt $P = (x; y; z = f(x, y))$ als Höhenkoordinate zu, erhält man in der Regel eine über dem Definitionsbereich D liegende Fläche.

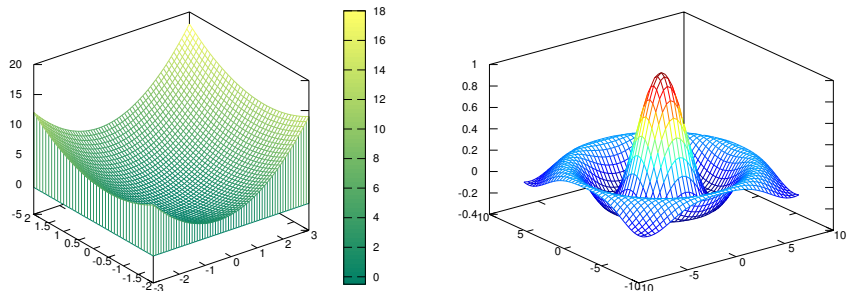


Abbildung: zwei Beispiele für die grafische Darstellung von Funktionen als Flächen im Raum. Links die Funktion $f(x, y) = z = x + x^2 + y + y^2$, rechts die 'Sombbrero-Funktion' $f(r) = \sin(r)/r$ mit $r = \sqrt{x^2 + y^2}$.

Ableitung von Funktionen mehrerer Variabler

Skalarwertige Funktionen einer Variablen lassen sich ohne Probleme ableiten:

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Das Konzept lässt sich einfach auf vektorwertigen Funktionen einer Variablen anwenden - es werden hier einfach die einzelnen Komponenten der vektorwertigen Funktion, die alle von derselben Variablen x abhängen, differenziert:

$$\vec{f}'(x) = \begin{pmatrix} f'_1(x) \\ \vdots \\ f'_n(x) \end{pmatrix}$$

Wie aber differenziert man eine skalarwertige Funktion mehrerer Variablen, beispielsweise eine Funktion wie $f(x_1, x_2) = x_1^3 \cos x_2$?

Ableitung von Funktionen mehrerer Variabler

Da die Funktion von mehreren Variablen abhängt, gibt es in diesem Fall auch mehrere Möglichkeiten, diese Variablen zu verändern (im obigen Beispiel könnte ja nach der Variablen x_1 oder nach der davon unabhängigen Variablen x_2 differenziert werden). Zunächst muss also festgelegt werden, welche der Variablen betrachtet werden sollen.

Differenziert man einfach nach einer Variablen und hält alle anderen fest (man betrachtet sie als Parameter), so gelangt man zu partiellen Ableitungen einer Funktion mehrerer Variabler:

Definition: partielle Ableitung

Die Funktion $f(\vec{x})$ ist in \vec{x}_0 nach der Variablen x_i partiell differenzierbar, falls der Grenzwert

$$\frac{\partial}{\partial x_i} f(\vec{x}) = \lim_{h \rightarrow 0} \frac{f(\vec{x}_0) + h\vec{e}_i - f(\vec{x}_0)}{h} \quad (2)$$

existiert. Dabei bezeichnet \vec{e}_i den i -ten kanonischen Einheitsvektor. Er besteht aus einem Nullvektor, in dem die i -te Komponente durch den Wert 1 ersetzt wird:

$$\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

Die Größe $\frac{\partial}{\partial x_i} f(\vec{x})$ heißt *partielle Ableitung* von $f(\vec{x})$ nach x_i im Punkt \vec{x}_0 .

Ableitung von Funktionen mehrerer Variabler

- Die partielle Ableitung einer Funktion mehrerer Variabler $(x_1, x_2, \dots) \mapsto f(x_1, x_2, \dots)$ ist also die Ableitung der Funktion nach einer dieser Variablen. Man schreibt dafür auch:

$$f_{x_i}(x_1; x_2; \dots) = \frac{\partial}{\partial x_i} f \quad (3)$$

- Für den Fall der beiden unabhängigen Variablen können wir die Definition etwas einfacher schreiben:
Partielle Ableitung 1. Ordnung nach x :

$$f_x(x; y) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x; y) - f(x; y)}{\Delta x} \quad (4)$$

Partielle Ableitung 1. Ordnung nach y :

$$f_y(x; y) = \lim_{\Delta y \rightarrow 0} \frac{f(x; y + \Delta y) - f(x; y)}{\Delta y} \quad (5)$$

Beispiel:

$f(x_1, x_2) = x_1^3 \cos x_2$. Dann sind die beiden partiellen Ableitungen

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 3x_1^2 \cos x_2$$

und

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = -x_1^3 \sin x_2.$$

Die bekannten Ableitungsregeln für Funktionen einer Variabler übertragen sich direkt auf Funktionen mehrerer Variabler.

Ableitung von Funktionen mehrerer Variabler

Für eine Funktion $(x, y) \mapsto f(x, y)$ von zwei unabhängigen Variablen versteht man unter dem totalen Differential den Ausdruck

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \quad (6)$$

Der Ausdruck wird als *totaler* Differential bezeichnet, weil er die gesamte Information über die Ableitung enthält, während die partiellen Ableitungen nur Information über die Ableitung in Richtung der jeweiligen Koordinatenachsen enthalten.

Ähnlich wie die Tangente die Gerade beschreibt, die mit der Steigung der Funktion in diesem Punkt an der Kurve anliegt, beschreibt das totale Differential eine Ebene. Diese Tangentialebene wird in einem Punkt $P = f(x_0; y_0)$ durch die beiden senkrecht zueinander an die (flächenartige) Kurve abgelegten Tangenten in Richtung der beiden Variablen x und y aufgespannt.

Das totale Differential beschreibt also näherungsweise das Verhalten der Funktion in der Umgebung des Punktes.

Statistische Einheit, Merkmalsträger: Personen, Gegenstände aber auch Ereignisse, die (üblicherweise) in einer Stichprobe erfasst werden.

Merkmale: die bei einer statistischen Einheit interessierenden Eigenschaften, z.B. die Größe oder das Alter. Merkmale genannt.

Merkmalsausprägungen: Alternativen, die von einem Merkmal angenommen werden können. Beispiele für Merkmalsausprägungen sind 'blond', '20a'.

Grundgesamtheit / statistische Masse: ist die Menge aller relevanten statistischen Einheiten mit übereinstimmenden sachlichen, räumlichen und zeitlichen Identifikationskriterien.

Stichprobe: ist eine Menge von Merkmalsausprägungen $\{x_i\}$ mit $i \in \mathbb{N}$.
Die Mächtigkeit der Menge heißt Stichprobenumfang n .

Der Zweck der Statistik ist der Umgang mit großen Mengen von Daten, die meist mit zufälligen Abweichungen behaftet sind. Im einfachsten Fall werden umfangreiche Datenmengen dabei vereinfacht dargestellt (kondensiert).

Oft ist der erste Schritt die Bildung von Häufigkeitsverteilungen - dabei treten einzelne Merkmalswerte mehrfach auf, so dass nur $m \leq n$ unterschiedliche Werte mit der jeweiligen Häufigkeit h_i bzw. $f_i = h_i/n$ in einer Stichprobe enthalten sind.

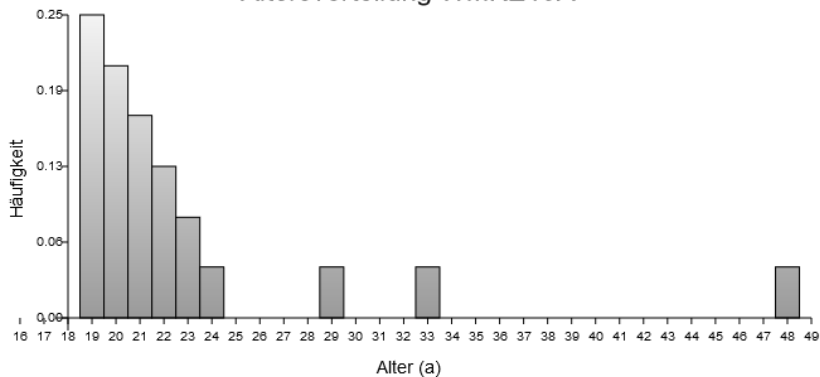
Die Urliste für eine Altersverteilung (aus WMKE19A in Ravensburg) ist

$$\{x_i\} = (19, 19, 29, 20, 20, 20, 20, 33, 21, 23, 22, 22, 19, 23, 24, 19, 20, 21, 21, 21, 19, 22, 19, 48)$$

Der Umfang der Stichprobe ist $n = 24$. Die Werte für das Alter lassen sich in eine Häufigkeitsverteilung umschreiben

Alter x_i	Häufigkeit h_i	rel. Häufigkeit f_i	kumuliert F_i
19	6	0.250	0.250
20	5	0.208	0.458
21	4	0.167	0.625
22	3	0.125	0.750
23	2	0.083	0.833
24	1	0.042	0.875
29	1	0.042	0.917
33	1	0.042	0.959
48	1	0.042	1

Grafik der Häufigkeitsverteilung:
Altersverteilung WMKE19A



Gauß-Verteilung

Die unten gezeigte Funktion heißt Gauß-Verteilung (oft Gauß-Glocke oder Glockenkurve). Sie beschreibt die Wahrscheinlichkeitsdichte für eine normalverteilte Zufallsgröße.

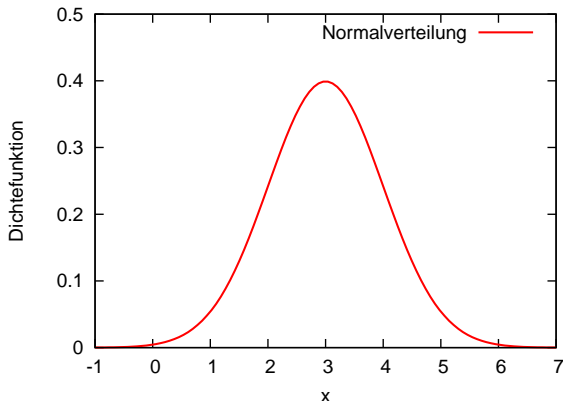


Abbildung: Normalverteilung mit $\sigma = 1, \mu = 3$.

Falls μ den Erwartungswert und σ die Standardabweichung von X bezeichnet, so hat die Dichtefunktion von X die Form

$$p(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (7)$$

die Variable X wird dann als *normalverteilt* bezeichnet.

Statistische Parameter (auch statistische Maßzahlen) sind charakteristische Werte, die eine Menge von Beobachtungen *einfach* beschreiben. Zweck ist die Verdichtung von Daten einer Stichprobe in einzelne, möglichst einfache Parameter. Dabei wird Information vernichtet, aber Übersichtlichkeit gewonnen! Für eine Menge von Beobachtungen lassen sich viele solcher Maßzahlen angeben.

Das arithmetische Mittel ist der am weitesten verbreitete Mittelwert. Es kann nur für kardinale Merkmale berechnet werden.

Definition: arithmetisches Mittel (AM)

Das AM ist die Summe der Merkmalswerte geteilt durch ihre Anzahl:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Das arithmetische Mittel ist ein *Lageparameter*, es ist ein Schätzwert für das Maximum der Verteilung (den Erwartungswert μ einer Gaußverteilung). Verschiedene Mittelwerte unterscheiden sich durch das Kriterium, wann ein Wert als typisch gesehen wird.

Die Summe aller Merkmalswerte ist

$$S = x_1 + x_2 + x_3 + \cdots + x_n = \sum_{i=1}^n x_i$$

wählt man als typischen Wert \bar{x} für die Daten den Wert, der n -mal summiert dieselbe Summe S ergibt, folgt das arithmetische Mittel:

$$\underbrace{\bar{x} + \bar{x} + \bar{x} + \bar{x} + \cdots + \bar{x}}_{n \cdot \bar{x}} = \sum_{i=1}^n \bar{x} = S = \sum_{i=1}^n x_i$$
$$\Leftrightarrow n \cdot \bar{x} = \sum_{i=1}^n x_i \quad \Leftrightarrow \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Eine weitere Möglichkeit ist eine Kennzahl, die die Breite der Verteilung beschreibt (und einer Schätzung der Varianz σ^2 einer Gaußverteilung entspricht). Streumaße beschreiben die typische Abweichung eines Merkmalswerts vom typischen Wert: berechnet man die Abweichungen

$$\Delta_i = \bar{x} - x_i,$$

erhält man eine neue Stichprobe, die die Information über die Abweichungen enthält. Mitteln der Abweichungen führt zu

$$\begin{aligned}\bar{\Delta} &= \frac{1}{n} \sum_{i=1}^n \bar{x} - x_i = \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{n} \sum_{i=1}^n \bar{x} \\ &= \bar{x} - \frac{1}{n} n \bar{x} = 0\end{aligned}$$

Die Abweichungen nach oben bzw. unten heben sich genau auf.

Für die mittlere Abweichung spielt das Vorzeichen keine Rolle, man benutzt deshalb das Quadrat der Abweichungen: Die *durchschnittliche quadratische Abweichung* der Einzelwerte einer (kardinalen) Stichprobe (x_1, x_2, \dots, x_n) vom arithmetischen Mittel \bar{x} wird als die empirische Varianz s^2 bezeichnet:

Definition: empirische Varianz

$$s^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (8)$$

Die Wurzel der Varianz wird als Standardabweichung bezeichnet:

$$s = \sqrt{s^2}$$

Mit Hilfe der 2. binomischen Formel kann die empirische Varianz umgeschrieben werden:

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - 2\bar{x} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) + \frac{1}{n} \cdot n\bar{x}^2, \end{aligned}$$

wobei der mittlere Term wegen $1/n \sum x_i = \bar{x}$ genau $-2\bar{x}^2$ ergibt und der letzte Term sich auf \bar{x}^2 reduziert. Damit ergibt sich

$$s^2 := \frac{1}{n} \left(\sum_{i=1}^n x_i^2 \right) - \bar{x}^2, \quad (9)$$

was sich einfacher berechnen lässt.

bivariate Verteilungen

Liegen Stichproben vor, die mehrere Merkmalswerte erfassen (die prinzipiell voneinander abhängig sein können), so spricht man von multivariaten Verteilungen. Der Einfachheit halber werden wir uns hier auf Verteilungen zweier Merkmale, sog. bivariate Verteilungen, beschränken. Die interessante Frage ist hier natürlich die nach der Beziehung der einzelnen Merkmale untereinander.

Die Varianz berechnet sich aus den quadrierten Abweichungen $(x_i - \bar{x})^2$ einzelner Werte vom Schwerpunkt des Datensatzes \bar{x} . Die Kovarianz berechnet sich analog aus den Abständen von Beobachtungspaaren (x_i, y_i) von den jeweiligen arithmetischen Mitteln. An die Stelle der Abweichungsquadrate treten die Produkte der Abweichungen der Merkmale, die wieder gemittelt werden:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (10)$$

Definition: empirische Kovarianz

Die Größe s_{xy} bezeichnet man als *Kovarianz*, sie lässt sich natürlich nur berechnen, wenn beide Merkmale kardinal skaliert sind.

Auch für die Kovarianz findet man einen alternativen Ausdruck:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (11)$$

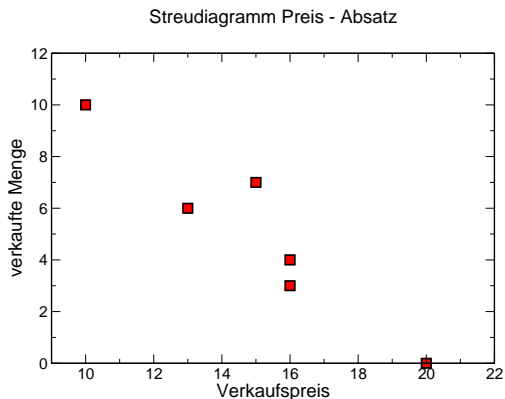
Beispiel: lineare Regression.

Ein Hersteller von Dingen produziert mehrere Modelle (nummeriert mit dem Index i) zum Preis von jeweils x_i . Um den Absatz zu optimieren, soll eine Preis-Absatz-Funktion ermittelt werden.

	Preis	Menge			
i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
1	20	0	5	-5	-25
2	16	3	1	-2	-2
3	15	7	0	2	0
4	16	4	1	-1	-1
5	13	6	-2	1	-2
6	10	10	-5	5	-25
Summe	90	30	0	0	-55
Mittelwert	15	5			

Tabelle: Preise verschiedener Modelle von Dingen

Beispiel: lineare Regression.



Im Streudiagramm kann man einen linearen Zusammenhang (eine Gerade) erahnen.

Die Merkmale sind miteinander über eine lineare Beziehung verknüpft.

Abbildung: Merkmalswerte zum Beispiel des Herstellers von Dingen.

Im Folgenden gehen wir davon aus, dass eine Datenreihe zweier Merkmale vorliegt, bei der ein Zusammenhang zwischen den Merkmalen existiert. Der Einfachheit halber wird ein linearer Zusammenhang angenommen:

$$y = a \cdot x + b$$

Es liegen also viele (n) unabhängige Angaben zum selben Sachverhalt vor (für jedes der n Paare):

$$y_i = a_i \cdot x + b_i$$

Da jedes einzelne Datenpaar mit einer (zufälligen) Abweichung vom realen Wert behaftet ist, kann für den Zusammenhang lediglich ein Schätzwert bestimmt werden. Gedanklich könnte jedes einzelne Paar von Punkten im Streudiagramm über eine Funktion (im einfachsten Fall eine Gerade) verbunden werden.

lineare Regression

Man erhält eine Anzahl von Funktionen, die sich wahrscheinlich ähnlich, aber nicht identisch sein werden - es existiert eine große Zahl von denkbaren Zusammenhängen (bei n Wertepaaren gibt es insgesamt $(n - 1)!$ mögliche Geraden).

Welche der Geraden ist die Richtige?

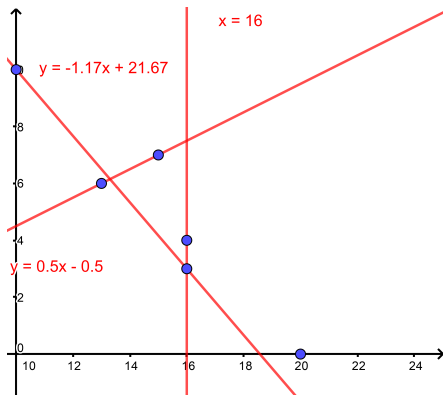


Abbildung: Einige mögliche Geraden $y = a_i x + b_i$

Das Problem bei der Regression besteht darin, ein Kriterium für eine gut angepasste Gerade zu finden. Prinzipiell sind immer mehrere Möglichkeiten des 'Ausgleichens' solcher redundanten Messungen denkbar, beispielsweise könnte über die Steigungen a_i und Achsenabschnitte b_i einfach gemittelt werden.

Aus historischen Gründen (1801: Entdeckung von Ceres durch Giuseppe Piazzi) hat sich aber die *Methode der kleinsten Quadrate* durchgesetzt: in Abb. 3 sind Punkte in einem System der Variablen x, y gegeben. Die 'Punktwolke' soll durch eine Gerade möglichst gut repräsentiert werden. Gesucht sind nun die Parameter jener Geraden, die für diese Approximation 'am besten' geeignet ist. Wir nennen diese Gerade auch *Ausgleichsgerade*.

Für die Regression nimmt man an, dass die wahren Werte der gemessenen Größen auf einer Geraden liegen, d.h. zwischen der Variablen x und den wahren Werten der beobachteten Größe y wird ein linearer Zusammenhang angenommen:

$$y = f(x) = a + b \cdot x$$

Mathematisch kann zur Bestimmung der beiden Geradenparameter Steigung a und Ordinatenabschnitt b ein Gleichungssystem aufgestellt werden, wobei jeder Punkt eine Gleichung beisteuert (dabei wird vorausgesetzt, dass mehr Datenpunkte vorliegen, als zur Bestimmung der Unbekannten notwendig sind) \Rightarrow es handelt sich bei unserem Beispiel um ein Gleichungssystem mit zwei Unbekannten und 6 Gleichungen. Ein solches System wird als überbestimmt bezeichnet, wegen der leichten Abweichungen realer Messwerte kann es nicht eindeutig gelöst werden.

Eine Gerade ist in der Ebene durch zwei Punkte definiert. Gibt es mehr als zwei Punkte, die auf einer Geraden liegen sollen, kann im Allgemeinen keine eindeutige Lösung angegeben werden. Es muss also ein Kriterium dafür gefunden werden, welche Gerade 'möglichst gut' an die Punktwolke angepasst ist. Wir fordern dafür, dass die Abweichung der Funktionswerte einer unbekanntes Funktion $y = f(x)$ von den gemessenen Werten möglichst klein ist.

Die Abweichung kann unterschiedlich definiert werden - plausibel erscheint eine Definition, in der ein größerer Fehler überproportional mehr wiegt als ein kleiner. Zudem darf das Fehlermaß nicht vorzeichenbehaftet sein (sonst würde ein negativer Fehler einen betragsgleichen positiven Fehler ausgleichen). Das Quadrat der Abweichung zwischen Mess- und Funktionswert erfüllt beide Forderungen.

$$\begin{aligned} f(a, b) &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - (ax_i + b))^2 \end{aligned}$$

Zunächst kann die Summe noch über die binomische Formel umgeformt werden:

$$\begin{aligned} \sum_{i=1}^n (y_i - (ax_i + b))^2 &= \sum_{i=1}^n (y_i^2 - 2y_i(ax_i + b) + (ax_i + b)^2) \\ &= \sum_{i=1}^n (y_i^2 - 2ax_iy_i - 2by_i + a^2x_i^2 + 2abx_i + b^2) \end{aligned}$$

Die partiellen Ableitungen der Funktion $f(a, b)$ sind

$$\frac{\partial f(a, b)}{\partial a} = \sum_{i=1}^n -2x_i y_i + 2ax_i^2 + 2bx_i$$

$$= 2 \sum_{i=1}^n x_i (ax_i + b - y_i)$$

$$\frac{\partial f(a, b)}{\partial b} = \sum_{i=1}^n -2y_i + 2ax_i + 2b$$

$$= 2 \sum_{i=1}^n (ax_i + b - y_i)$$

Die partielle Ableitung nach b kann jetzt gleich Null gesetzt und aufgelöst werden:

$$\frac{\partial f(a, b)}{\partial b} = 2 \sum_{i=1}^n (ax_i + b - y_i) = 0$$

$$2 \sum_{i=1}^n ax_i + 2 \sum_{i=1}^n b - 2 \sum_{i=1}^n y_i = 0$$

$$2nb = 2 \sum_{i=1}^n y_i - 2 \sum_{i=1}^n ax_i$$

$$b = \frac{2}{2n} \sum_{i=1}^n y_i - \frac{2}{2n} \sum_{i=1}^n ax_i$$

oder mit der Definition des arithmetischen Mittels $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

$$b = \frac{1}{n} \sum_{i=1}^n y_i - a \frac{1}{n} \sum_{i=1}^n x_i$$

$$b = \bar{y} - a \cdot \bar{x} \tag{12}$$

Daraus folgt sofort, dass

$$\bar{y} = a\bar{x} + b,$$

der Punkt $(\bar{x}; \bar{y})$ liegt immer auf der optimal angepassten Geraden.

Setzt man Gleichung (12) in die partielle Ableitung nach a ein, erhält man für die Minimierung

$$\begin{aligned}\frac{\partial f(a, b)}{\partial a} &= 2 \sum_{i=1}^n x_i (ax_i + b - y_i) \\ &= 2 \sum_{i=1}^n x_i (ax_i + \bar{y} - a \cdot \bar{x} - y_i) = 0\end{aligned}$$

Ausmultiplizieren liefert

$$\begin{aligned}0 &= \sum_{i=1}^n ax_i^2 + x_i(\bar{y} - a \cdot \bar{x}) - x_i y_i \\ &= \sum_{i=1}^n ax_i^2 + x_i \bar{y} - ax_i \bar{x} - x_i y_i\end{aligned}$$

$$0 = \sum_{i=1}^n (ax_i^2 - a\bar{x}x_i) + \sum_{i=1}^n x_i\bar{y} - \sum_{i=1}^n x_iy_i$$

$$a \sum_{i=1}^n (x_i^2 - \bar{x}x_i) = \sum_{i=1}^n x_iy_i - \sum_{i=1}^n x_i\bar{y}$$

$$a = \frac{\sum_{i=1}^n x_iy_i - \sum_{i=1}^n x_i\bar{y}}{\sum_{i=1}^n x_i^2 - \bar{x}x_i}$$

Erweitern mit $\frac{1}{n}$ liefert

$$a = \frac{\frac{1}{n} \sum_{i=1}^n x_iy_i - \frac{1}{n} n\bar{x}\bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n} n\bar{x}x_i} = \frac{\frac{1}{n} \sum_{i=1}^n x_iy_i - \bar{x}\bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}$$

Jetzt entsprechen Zähler und Nenner den alternativen Darstellungen der Kovarianz s_{xy} bzw. der Varianz s_{xx}

Regression

Die gesuchten Regressionskoeffizienten sind die Lösungen

$$a = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}}$$

und (die Regressionsgerade geht immer durch den Punkt (\bar{x}, \bar{y}))

$$\begin{aligned}\bar{y} &= a \cdot \bar{x} + b \\ \Rightarrow b &= \bar{y} - b\bar{x}\end{aligned}$$

Beispiel: lineare Regression

Für die Daten des Herstellers von Dingen

i	Preis x_i	Menge y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	20	0	5	-5	-25	25
2	16	3	1	-2	-2	1
3	15	7	0	2	0	0
4	16	4	1	-1	-1	1
5	13	6	-2	1	-2	4
6	10	10	-5	5	-25	25
Summe	90	30	0	0	-55	56
Mittelwert	15	5				

Table: lineare Regression am Beispiel von Dingen

Beispiel: lineare Regression

ergeben sich für die beiden gesuchten Regressionsparameter die Werte

$$a = \frac{s_{xy}}{s_{xx}} = \frac{-55}{56} = -0,98$$

und

$$b = \bar{y} - a \cdot \bar{x} = 5 + 0,98 \cdot 15 = 19,73$$

Die jeweilige verkaufte Menge an Produkten y hängt also mit dem Preis x angenähert wie

$$y = a \cdot x + b = -0,98x + 19,73$$

zusammen. Mit einer Preiserhöhung um eine Einheit sinkt der Absatz um etwa eins.

Beispiel: lineare Regression

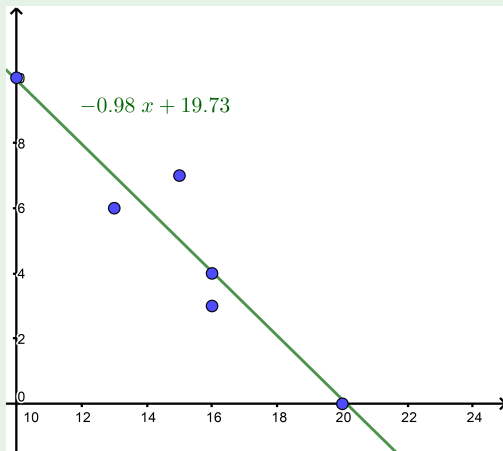


Abbildung: Daten zum Beispiel des Herstellers von Dingen mit Ausgleichsgerade.