

Einführung in statistische Analysen

Andreas Thams

Econ Boot Camp 2008

Wozu braucht man Statistik?

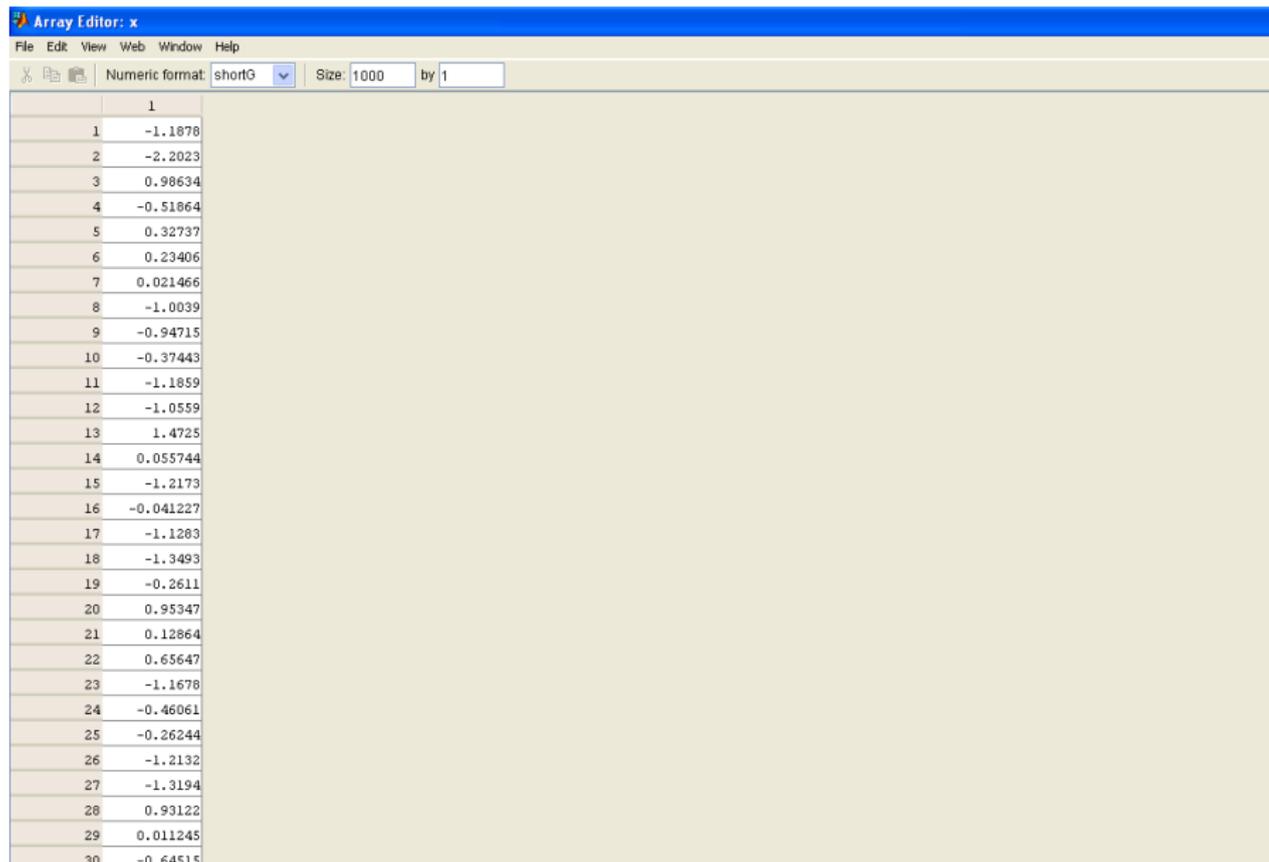
- ▶ Statistik begegnet uns jeden Tag...
 - ▶ **"Weihnachten macht Deutschen Einkaufslaune.** Im Advent überkommt die Deutschen wieder Einkaufslust: Die sogenannte Anschaffungsneigung der Verbraucher ist der Gesellschaft für Konsumforschung zufolge wieder gestiegen. Hinsichtlich ihrer Einkommenssituation und der Konjunktur sind die Bundesbürger aber pessimistisch."
 - ▶ **"Unternehmer zweifeln am Aufschwung und schaffen Jobs.** Die Stimmung in der deutschen Wirtschaft hat sich im Dezember stärker als erwartet eingetrübt. Der Ifo-Geschäftsklima-Index sank von 104,2 Punkten im Vormonat auf 103,0 Punkte. Ökonomen hatten einen moderateren Rückgang prognostiziert."

Aus Spiegel Online vom 20.12.2007

Was macht Statistik?

Ziel der Statistik ist es, Massendaten zu reduzieren und zu komprimieren, um Gesetzmäßigkeiten und Strukturen in den Daten sichtbar zu machen.

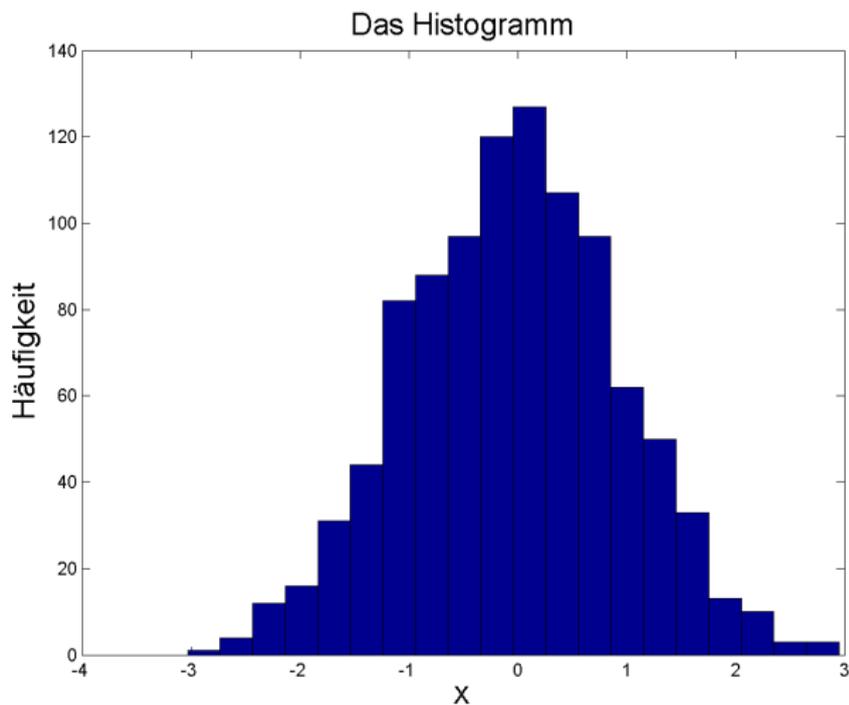
Was macht Statistik? Ein Beispiel... (1)



The screenshot shows the 'Array Editor' application window. The title bar reads 'Array Editor: x'. The menu bar includes 'File', 'Edit', 'View', 'Web', 'Window', and 'Help'. Below the menu bar, there are icons for a file, a printer, and a trash can. To the right of these icons, the 'Numeric format' is set to 'shortG' with a dropdown arrow, and the 'Size' is set to '1000' and 'by 1'. The main area of the window contains a table with 30 rows and 2 columns. The first column contains integers from 1 to 30, and the second column contains floating-point numbers.

	1
1	-1.1878
2	-2.2023
3	0.98634
4	-0.51864
5	0.32737
6	0.23406
7	0.021466
8	-1.0039
9	-0.94715
10	-0.37443
11	-1.1859
12	-1.0559
13	1.4725
14	0.055744
15	-1.2173
16	-0.041227
17	-1.1283
18	-1.3493
19	-0.2611
20	0.95347
21	0.12864
22	0.65647
23	-1.1678
24	-0.46061
25	-0.26244
26	-1.2132
27	-1.3194
28	0.93122
29	0.011245
30	-0.64515

Was macht Statistik? Ein Beispiel... (2)



Statistik in der Wirtschaftswelt

- ▶ Die Lage der Unternehmen heute ist geprägt von Globalisierung, Konkurrenz und Kostendruck.
- ▶ Ein Unternehmen trifft Entscheidungen unter Unsicherheit.
 - ▶ Wie entwickelt sich mein Umsatz / Gewinn in der Zukunft?
 - ▶ Wie entwickelt sich das Marktumfeld?
 - ▶ Werden die Zinsen steigen oder sinken?
- ▶ Ein Hilfsmittel für die Entscheidung unter Unsicherheit ist die Statistik.
 - ▶ Statistik kann Gesetzmäßigkeiten aufdecken.
 - ▶ Mit Statistik können wahrscheinliche Szenarien ermittelt werden.
 - ▶ Mit statistischen Methoden können zukünftige Entwicklungen prognostiziert werden.

Untergliederung der Statistik

1. **Deskriptive (beschreibende, empirische) Statistik:** Man untersucht ein Phänomen und fasst die Daten zusammen, ordnet sie, stellt sie grafisch dar.
2. **Induktive (schließende, folgernde, mathematische, analytische) Statistik:** Grundlage ist die Wahrscheinlichkeitstheorie. Ergebnisse der deskriptiven Statistik dienen häufig als Ausgangspunkt für verallgemeinernde Aussagen.

Einige Begriffe der deskriptiven Statistik

1. Lagemaße

- ▶ Mittelwert
- ▶ Median

2. Streuungsmaße

- ▶ Varianz
- ▶ Standardabweichung

Mittelwert / Arithmetisches Mittel

Der Mittelwert einer Variable x ist definiert als

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$$

Mittelwert / Arithmetisches Mittel am Beispiel Bevölkerungsentwicklung in Deutschland (1)

Jahr	1999	2000	2001	2002
Gesamtbevölkerung in 1000	82163475	82259540	82440309	82536680
Jahr	2003	2004	2005	2006
Gesamtbevölkerung in 1000	82531671	82500849	82437995	82314906

Quelle: Statistisches Bundesamt

- ▶ Damit ergibt sich eine durchschnittliche Bevölkerungsgröße von

$$\frac{1}{8}(82163475+82259540+82440309+\dots+82314906) = 82,398 \text{ Mill.}$$

Mittelwert / Arithmetisches Mittel am Beispiel Bevölkerungsentwicklung in Deutschland (2)

Wie hoch ist das durchschnittliche Bevölkerungswachstum?

1. Wachstumsraten berechnen
2. Mittelwert der Wachstumsraten berechnen

Mittelwert / Arithmetisches Mittel am Beispiel: Bevölkerungsentwicklung in Deutschland (3)

Um wieviel Prozent ist die Bevölkerung 1999-2000 gewachsen?

$$\frac{82259540 - 82163475}{82163475} \cdot 100 = 0.1169\%$$

Allgemein: Die prozentuale Wachstumsrate zur vorherigen Periode ist gegeben als

$$\frac{x_t - x_{t-1}}{x_{t-1}} \cdot 100$$

Das durchschnittliche Bevölkerungswachstum für den Zeitraum 2000-2006 beläuft sich damit auf

$$\frac{1}{7}(0,1169\% + 0,2198\% + \dots - 0,0373\% - 0,0762\% - 0,1493\%) = 0,0264\%$$

Median

In der Statistik halbiert der Median eine Stichprobe. Gegenüber dem arithmetischen Mittel hat der Median den Vorteil, robuster gegenüber Ausreißern zu sein.

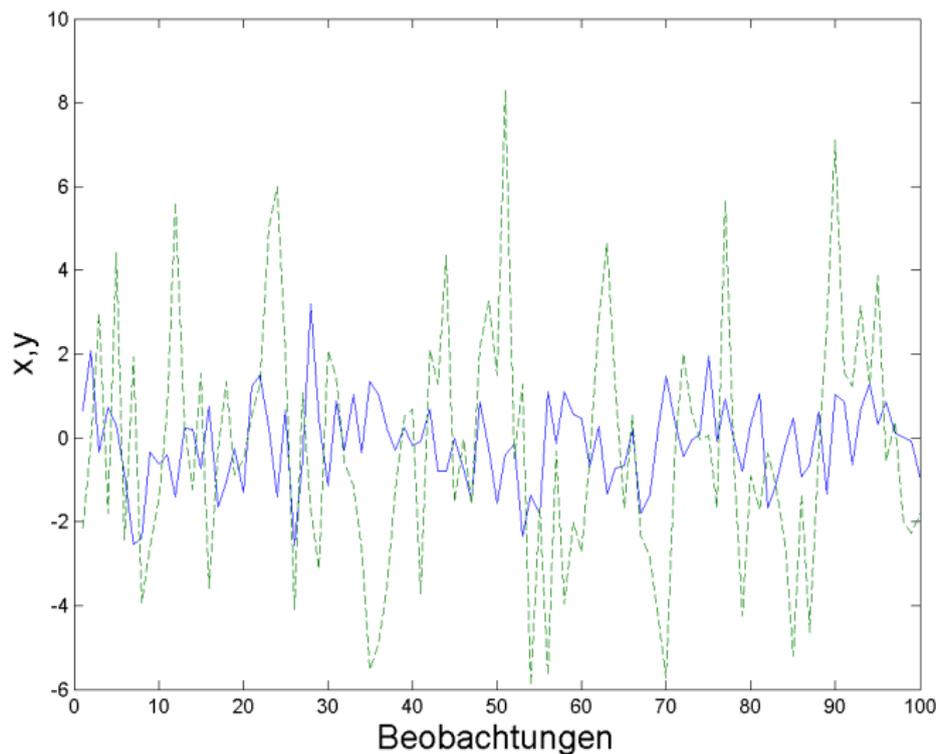
- ▶ Die Stichprobe wird dafür aufsteigend gereiht $\rightarrow 1\ 3\ 3\ 4\ 5\ 6\ 7$
- ▶ Ist die Anzahl an Beobachtungen gerade, ist der Median der Durchschnitt der beiden mittleren Werte der geordneten Reihe:
 $1\ 3\ 3\ 4\ 5\ 6\ 7\ 7 \rightarrow \text{Median} = 4,5$

Varianz

Die Varianz ist ein Streuungsmaß, d.h. ein Maß für die Abweichung einer Variable x von ihrem mittleren Wert. Sie ist definiert als

$$\text{Var}(x) = \sigma^2 = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})^2$$

Welche Variable hat die höhere Varianz?



Standardabweichung

Die Standardabweichung einer Variable x ist definiert als

$$\sigma = \sqrt{\sigma^2} = \sqrt{\text{Var}(x)}$$

Die Standardabweichung hat gegenüber der Varianz den Vorteil, dass sie die gleiche Einheit hat wie die ursprünglichen Messwerte.

Kovarianz

Die Kovarianz ist eine Maßzahl für den Zusammenhang zweier Variablen x und y .

- ▶ Die Kovarianz ist positiv, wenn x und y einen gleichsinnigen linearen Zusammenhang besitzen.
- ▶ Die Kovarianz ist negativ, wenn x und y einen gegensinnigen linearen Zusammenhang besitzen.

Kovarianz

Die Kovarianz zweier Variablen x und y ist definiert als

$$\text{Cov}(x, y) = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y})$$

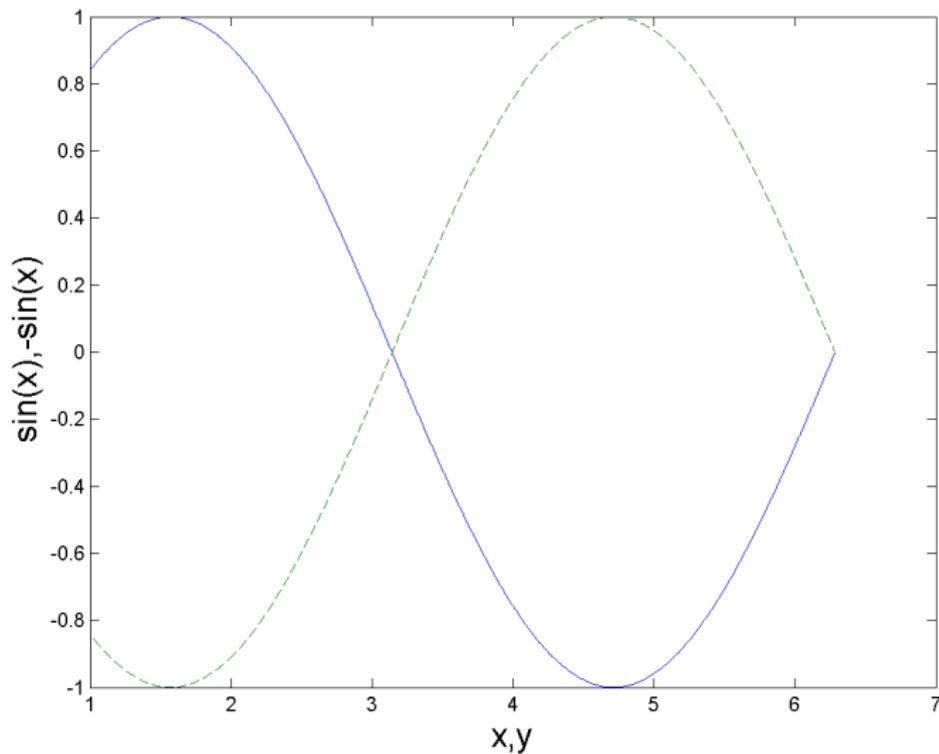
Korrelation

Wird die Kovarianz auf das Produkt der Standardabweichungen von x und y bezogen, so spricht man vom Korrelationskoeffizienten.

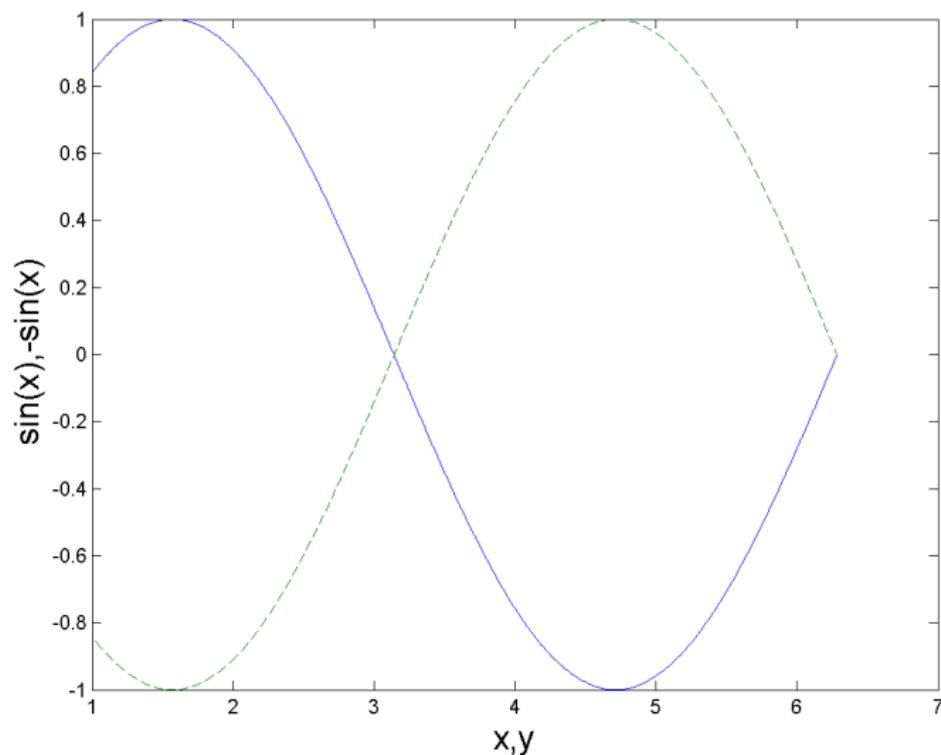
$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

- ▶ Die Korrelation kann lediglich Werte zwischen -1 und 1 annehmen.
- ▶ Bei einem Wert von +1 (bzw. -1) besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Variablen.
- ▶ Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Variablen überhaupt nicht linear voneinander ab.
- ▶ **ABER VORSICHT:** Korrelation bedeutet nicht automatisch Kausalität.

Korrelation am Beispiel: Wie groß ist die Korrelation? (1)

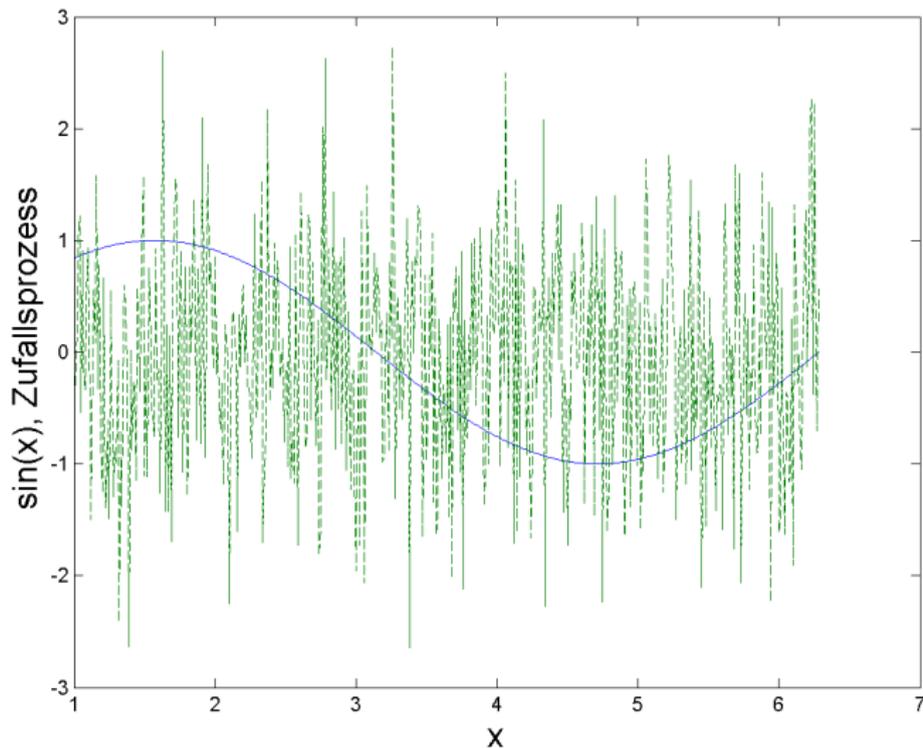


Korrelation am Beispiel: Wie groß ist die Korrelation? (1)

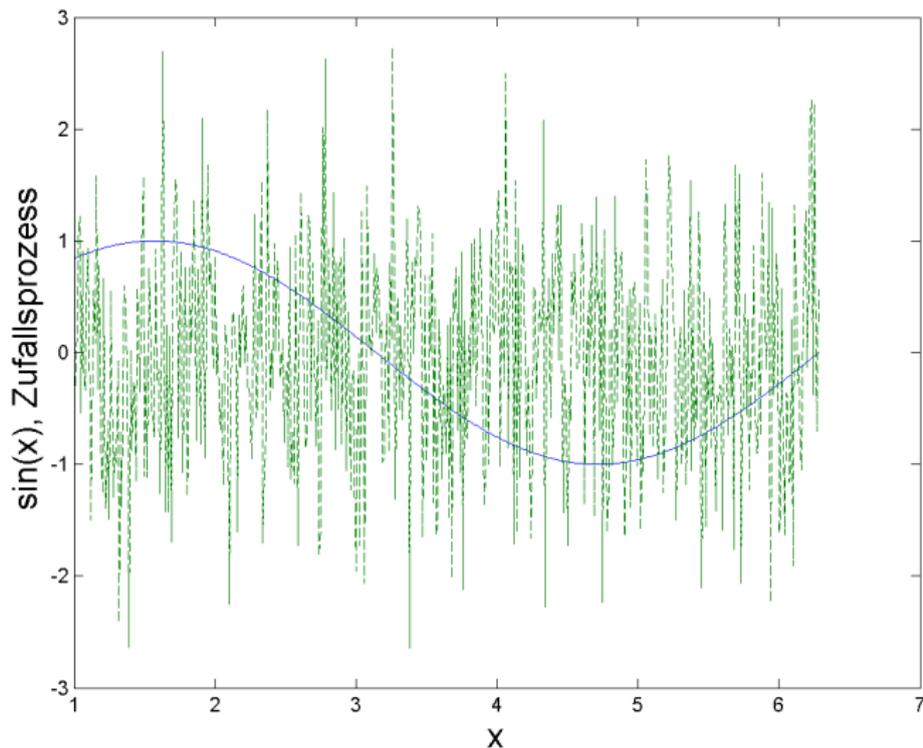


$$\text{Cov}(x, y) = -0,54, \text{Corr}(x, y) = -1$$

Korrelation am Beispiel: Wie groß ist die Korrelation? (2)

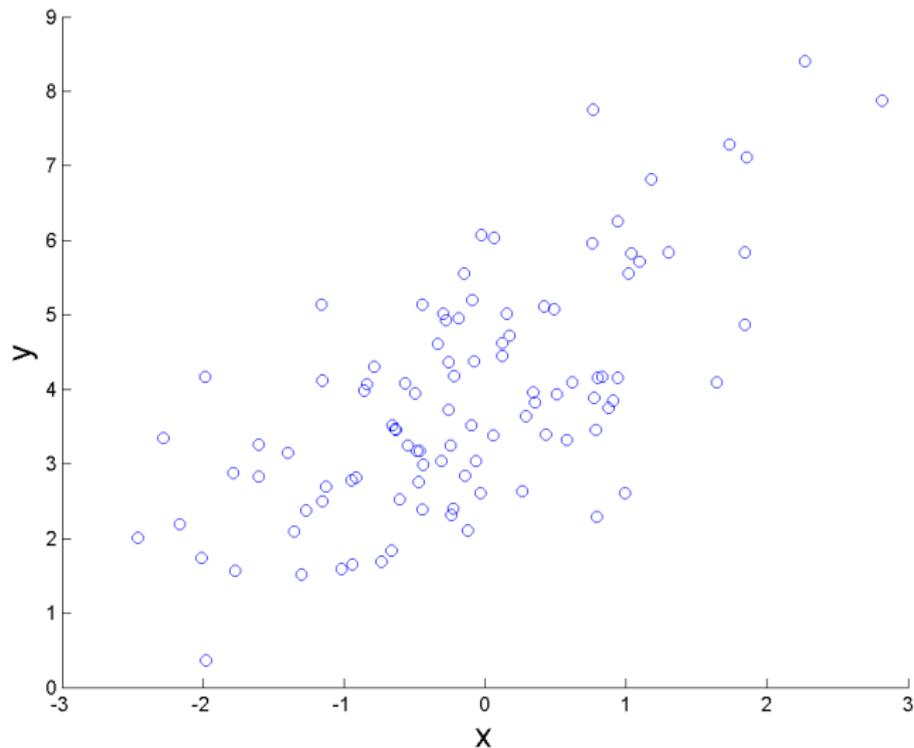


Korrelation am Beispiel: Wie groß ist die Korrelation? (2)



$$\text{Cov}(x, y) = -0,07, \text{Corr}(x, y) = -0,01$$

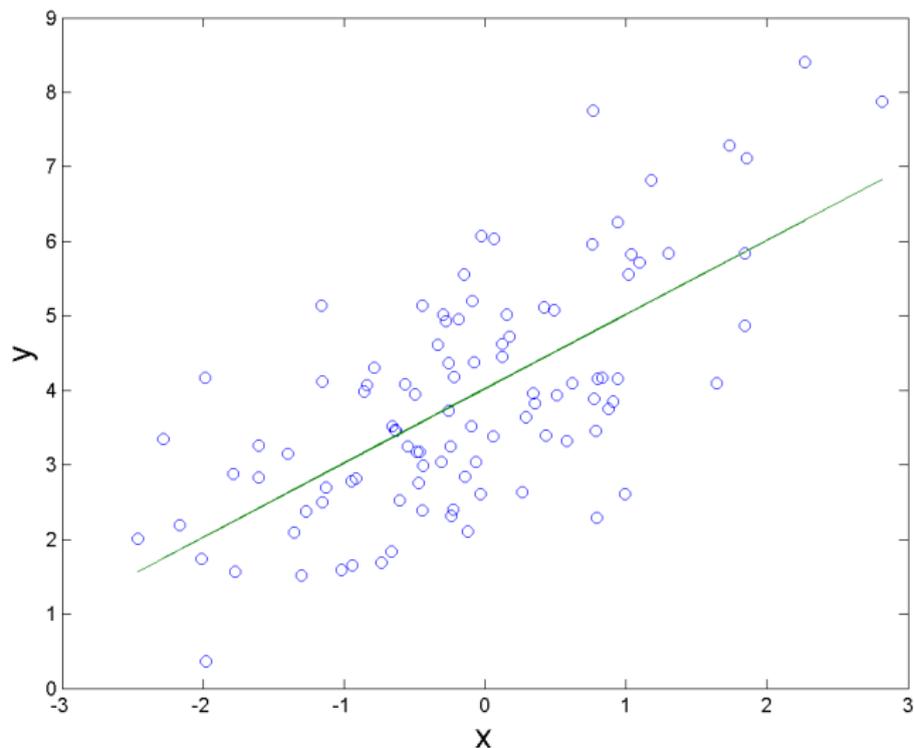
Streudiagramm (Scatter Plot)



Lineare Regressionsanalyse nach der Kleinste-Quadrate-Methode (1)

Die Regressionsanalyse ist ein statistisches Analyseverfahren. Ziel ist es, Beziehungen zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen festzustellen.

Lineare Regressionsanalyse nach der Kleinste-Quadrate-Methode (2)



Lineare Regressionsanalyse nach der Kleinste-Quadrate-Methode (2)

Geschätzter Zusammenhang: $\hat{y}_t = 4 + 1 \cdot x_t$

