

# Wie kann man Kurvenverläufe auswerten?

## Allgemein

Ich möchte hierzu zwei verschiedene Verfahren vorstellen, die auf anderen Wegen jeweils zu einem ähnlichen Ziel – der Trennung der Briefmarken anhand der Spektralkurven – erreichen. Das erste Verfahren von Pearson ist ausführlich in der Wikipedia beschrieben. Es wird normalerweise nicht auf Spektralkurven von Briefmarken, sondern elektrische Signalverläufe und ähnliches angewendet. Das zweite Verfahren hat als Grundlage die Differenzialrechnung und deren Kurvenbeschreibungsmöglichkeiten.

## Korrelation

Die Korrelation wird in Mathematik /Statistik dazu benutzt, den Grad der Übereinstimmung von Kurven zu ermitteln. Zur genauen Erklärung siehe <http://de.wikipedia.org/wiki/Korrelationskoeffizient> (wurde im Anhang aus Wikipedia kopiert). Bei der Bestimmung der Kurvenverläufe bei Briefmarken wurde der Korrelationskoeffizient nach Pearson eingesetzt. Er liefert schon ohne Anpassung recht gute Werte.

Die normale Korrelation wird (siehe auch wikipedia)

$$\rho(x,y) = (Cov(x,y))/(\sqrt{Var(x) * Var(y)})$$

Wobei  $Cov(x,y) = 1/(n-1) \sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))$  = Kovarianz von x,y

Mit  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  und  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  = Mittelwert

Und  $Var^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  und  $Var^2(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$  = Varianz

Bei näherer Betrachtung vereinfacht sich die Wurzel im Nenner von  $\rho$ , da sie sich mit dem Quadrat der Varianz aufhebt.

Verbessern lässt sich die Korrelation, wenn man die spektralen Eigenschaften des Auges mit berücksichtigt. Dadurch werden die Werte am Rand weniger berücksichtigt als die Werte in der Mitte. Es wurde dazu eine Kurve aus der Summation der spektralen Empfindlichkeit (bei D65, 10°, siehe dazu auch die Abbildung oben) gebildet. Mit diesem Empfindlichkeitsfaktor wird die Kovarianz und die Varianz multipliziert. Daraus ergibt sich die Formel für den gewichteten Korrelationskoeffizient nach der gleichen Formel für  $\rho(x,y)$ . Lediglich die Bestimmung der Kovarianz

$$Cov_{gew}(x,y) = 1/(n-1) \sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}) * gew_i)$$

Dabei ist  $gew_i$  die spektrale Empfindlichkeit des Auges (als Summation der 3 Komponenten) an der betreffenden Wellenlänge.

Die Varianz ändert sich dementsprechend auch auf

$$Var_{gew}^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) * gew_i$$

Und für y

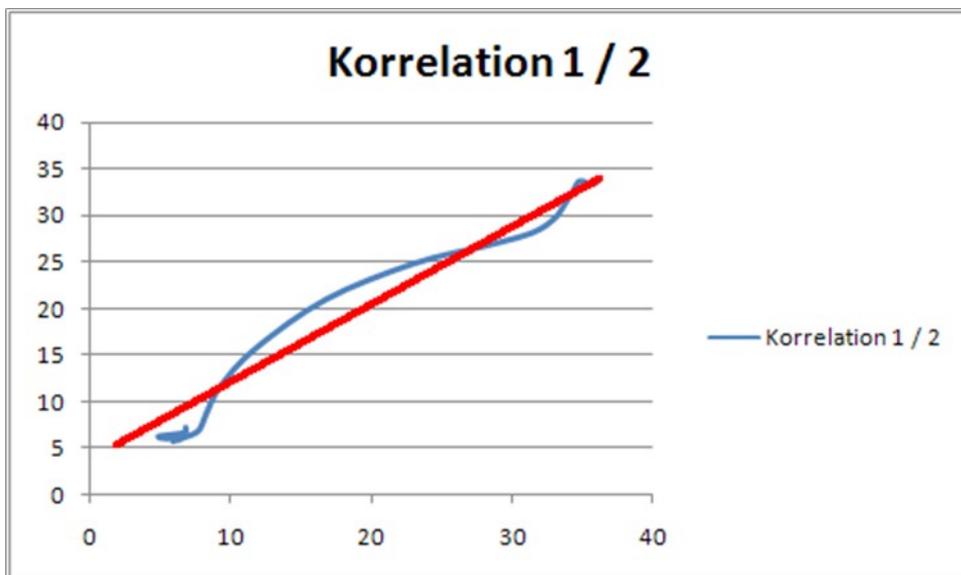
$$Var_{gew}^2(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}) * gew_i$$

Dieser Wert hat sich als geringfügig besser herausgestellt. Interessant ist die Umschaltung zwischen beiden Werten für jede Marke. Ist die Zuordnung einer Marke für beide Korrelationskoeffizienten nicht gleich, ist sie meist auch für das menschliche Auge nicht trennbar.

Man kann nun für jede Briefmarke diesen Korrelationskoeffizienten zu Referenzmarken berechnen. Je näher der Korrelationskoeffizient zwischen der neuen Marke und Referenzmarke an 1 liegt, desto ähnlicher ist die neue Marke der Referenzmarke. Hat man alle Koeffizienten zu den Referenzmarken berechnet, ordnet man die neue Marke der Referenzfarbe zu, deren Koeffizient am nächsten zu 1 ist.

Was wird hierbei getan?

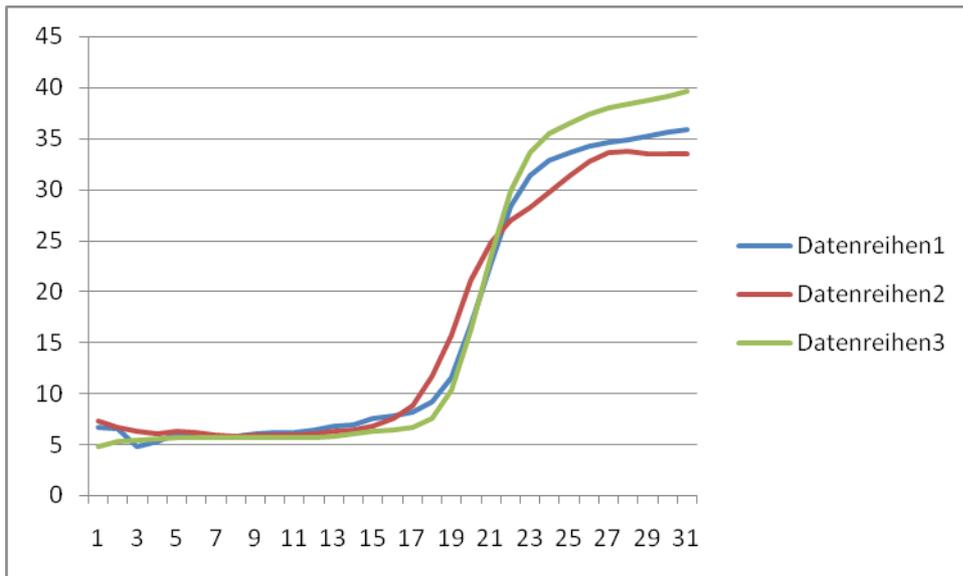
Es werden für die erste Marke die Spektralwerte als X und für die zweite Marke deren Spektralwerte als Y aufgetragen. Für die beiden 8-Pfg.- Marken des obigen Beispiels ist dies folgende blaue Kurve:



Mit Hilfe der Korrelation wird nun die Abweichung dieser Kurve einer gedachten Graden -im Bild rot dargestellt - ermittelt. Je geringer die Abweichung beider Kurven ist, desto ähnlicher sind die Kurven. Beim Korrelationskoeffizienten erkennt man dieses daran, wie nahe er an dem Wert 1 liegt.

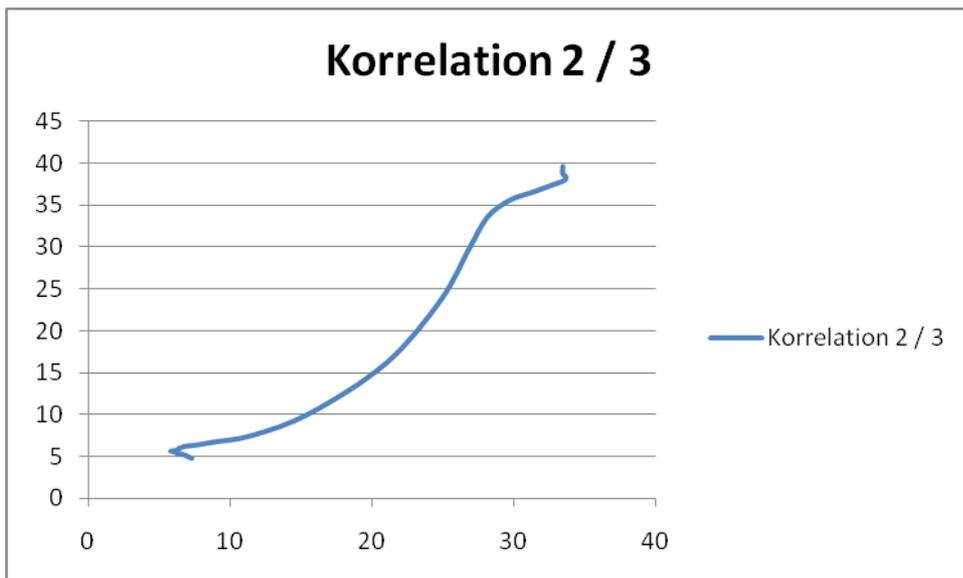
Deutlicher wird es wenn man eine 3. Marke hinzunimmt und die Korrelation zu den beiden anderen Marken berechnet:

Ausgangslage:

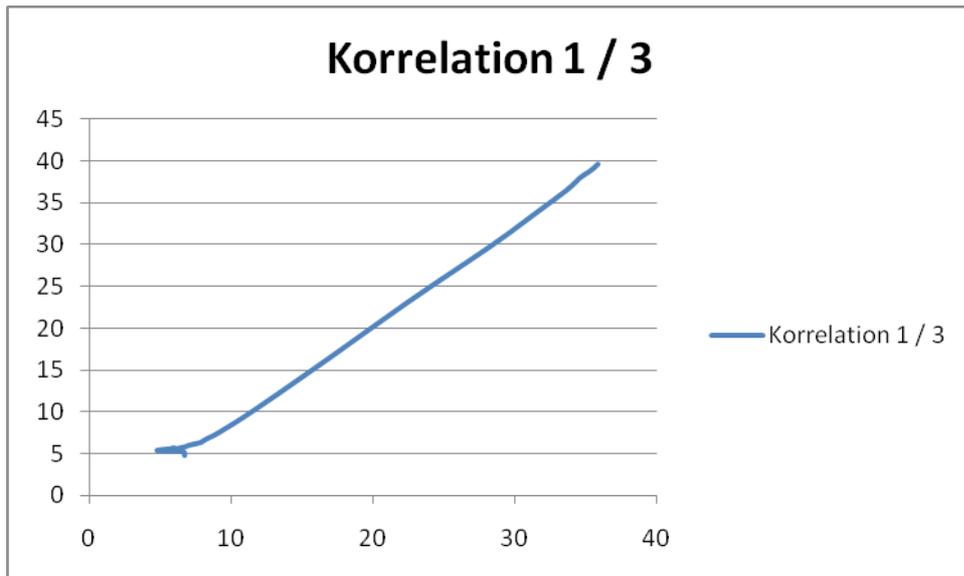


Der Kurvenverlauf der grünen Marke wurde hinzugefügt.

Es ergeben sich demnach neu die Korrelationen



Und



Man erkennt deutlich, dass sich die Marken 1 und 3 sehr ähnlich sind, während zu der Marke 2 doch deutliche Abweichungen erkennbar sind.

Dieses drückt sich in der Korrelation wie folgt aus:

1 zu 2 = 0,9842

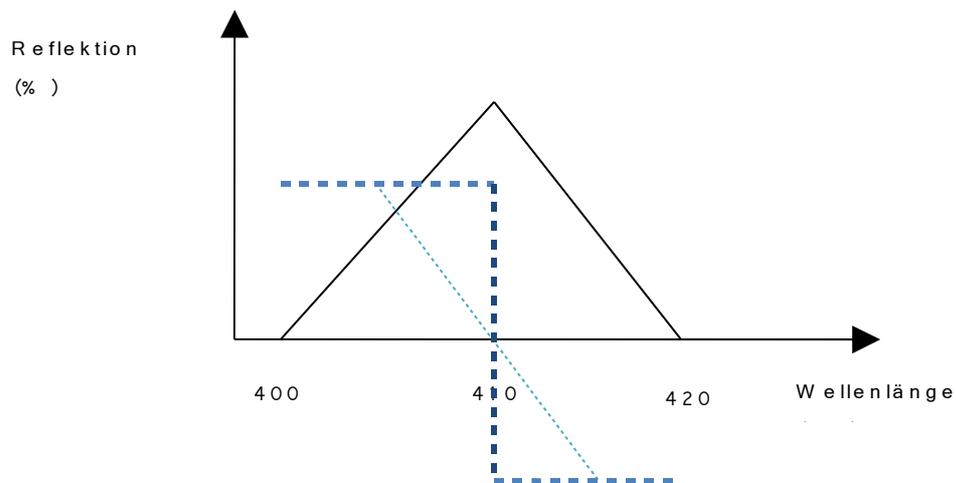
1 zu 3 = 0,9978

2 zu 3 = 0,9762

Die Korrelation 1 zu 3 hat die höchsten Werte und wen wundert es, wenn auch die Kurvenverläufe am ähnlichsten sind.

## Differenzialrechnung (Kurvendiskussion)

Die Differenzialrechnung bietet ein weiteres Mittel zu Beschreibung von Kurvenverläufen. Mit Hilfe von Ableitungen kann man Extremwerte und Wendepunkte bestimmen. Dieses erfolgt durch Ableitungen. Im digitalen Bereich hat man das Problem, dass sich zu 3 Werten von Kurven nur 2 Werte der Ableitung gehören. Dieses kommt dadurch, dass es für 3 Punkte einer Kurve nur 2 Steigungen gibt, also immer eine weniger als Kurvenpunkte. Damit die differenzierten Werte noch mit der Originalkurve übereinstimmen, werden die differenzierten Werte einfach der Mitte zwischen den beiden Originalpunkten zugeordnet. Die Steigung der Kurve zwischen den Punkten 400nm und 410 nm wird einfach durch einen Punkt bei 405 nm dargestellt.



In diesem Bild wird schematisch dargestellt, wie der Verlauf der Ableitung dargestellt wird. Die schwarze Kurve ist der gemessene Kurvenverlauf. Die Ableitung davon ist eigentlich die blau gepunktete Linie. Würde man diese Linie zeichnen, wäre der Kurvenverlauf ziemlich eckig. Da man auch die Originalkurve mit geraden Verbindungslinien gezeichnet hat, habe die Verbindung auch als Kurve dargestellt, die blaue durchgezogene Kurve. Da der Verlauf der Kurve zwischen den Werten nicht definiert ist, dürfte dieses auch zulässig sein.

Der Nullstelle wird durch die Bestimmung des Nulldurchgangs der durch die beiden Begrenzungspunkte beschriebenen Geraden errechnet.

Durch diese Darstellung erhält man für jede Spektralkurve eine Beschreibung ihres Verlaufes durch die Extremwerte und Wendepunkte.

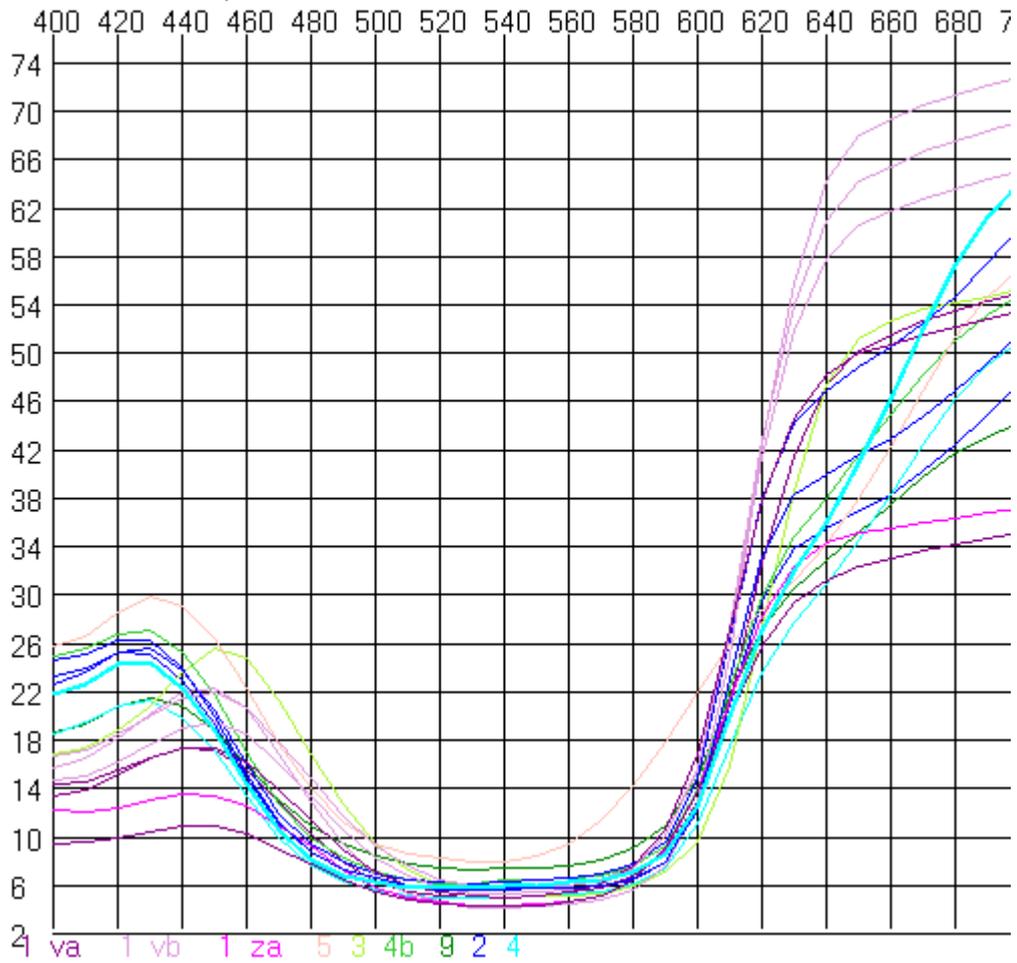
Unterschiedliche Kurvenverläufe erkennt man durch die unterschiedliche Lage der Extremwerte /Wendepunkte in dem Kurvenverlauf der 1. und 2. Ableitung.

Um eine Übersicht über die verschiedenen Kurvenverläufe zu erhalten, habe ich auf der X Achse die Marken und auf der y Achse die Extremwerte dargestellt. Dadurch erhält man einen guten Überblick über die Extremwerte aller vermessenen Marken.

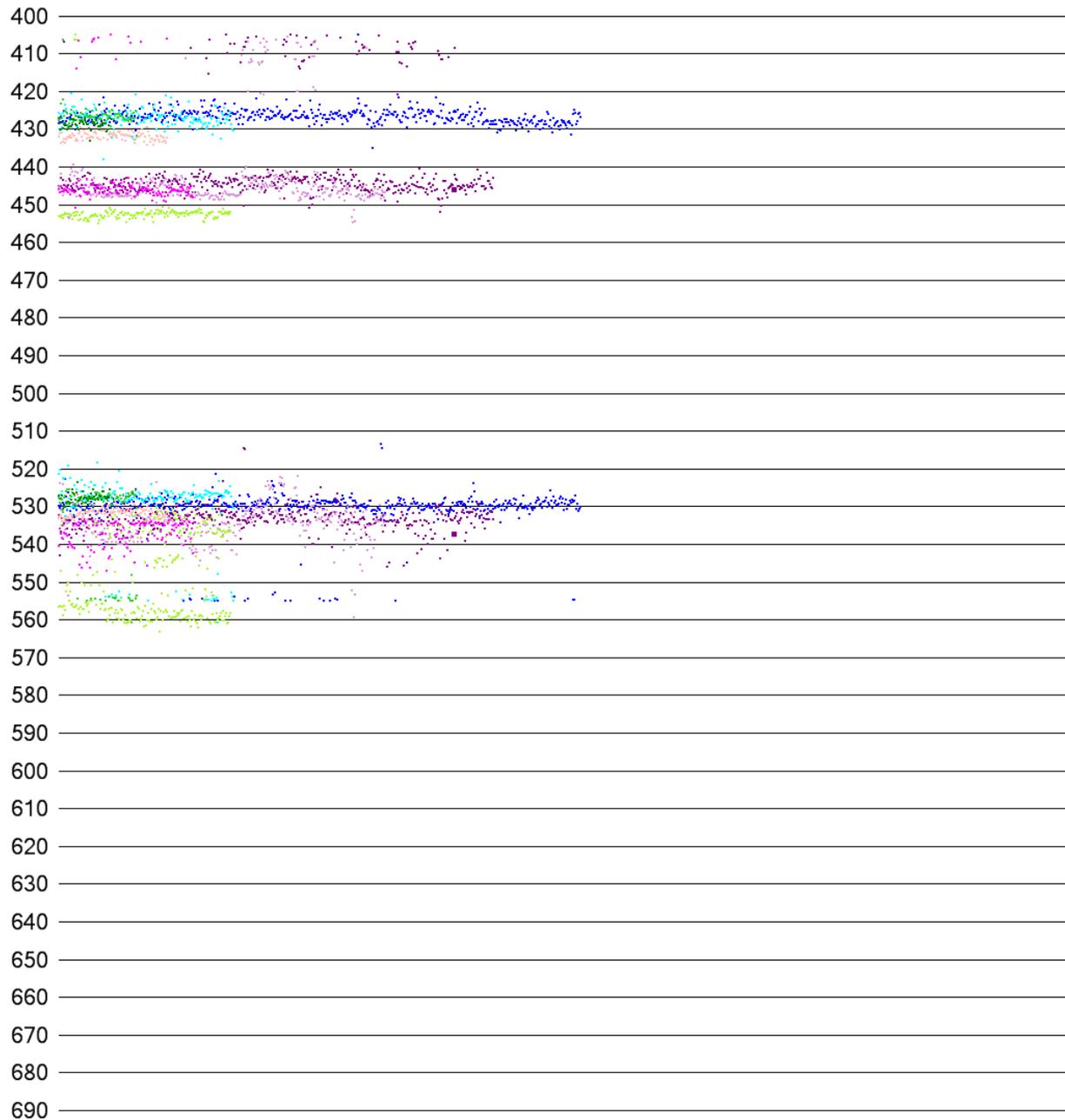
Dieses möchte ich am Beispiel der 40 Pf Marke der Köpfe 1 Serie betrachten.

Die Spektren der verschiedenen Marken sehen wie folgt aus:

### 40 Pf Referenzspektren



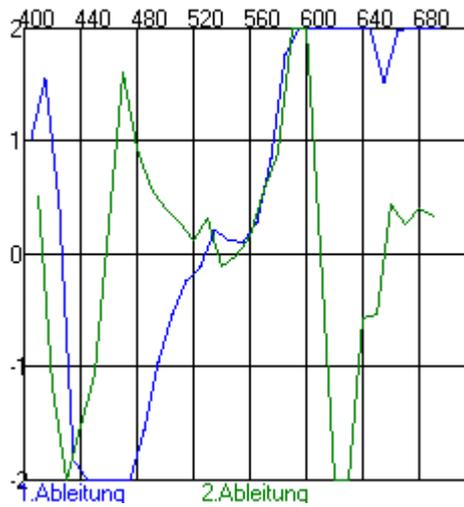
Man kann deutlich die verschiedenen Maxima im Bereich 420 nm bis 455 nm erkennen. Bei der Betrachtung aller Marken sieht dies wie folgt aus:



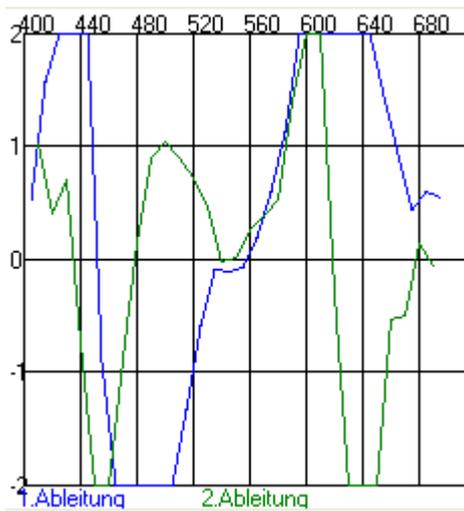
Die farbigen Punkte um die 410 nm sind Fehlmessungen. Die blauen Punkte (hellblau und dunkelblau) im Bereich um 430 nm stellen Maxima von Marken dar, die sich wohl nicht trennen lassen werden. Die hellroten Marken darunter haben wiederum einen kleinen Abstand, eine Trennung könnte möglich sein. Die violetten und dunkelroten Marken darunter sind gut trennbar, ebenso die grünen Marken.

Der Vorteil dieser Darstellung ist, dass man auf einen Blick die Trennbarkeit von Kurven erkennen kann. Überlappen sich Gebiete, dürfte eine Trennung auf Grund des Spektralverlaufes schwer erreichbar sein. Dabei reicht es aus, wenn die Marken in einem Gebiet, in diesem Falle dem Bereich von 420 nm bis 460 nm trennbar sind. Dass sie im Bereich 520 nm bis 560 nm nicht trennbar sind, spielt keine Rolle.

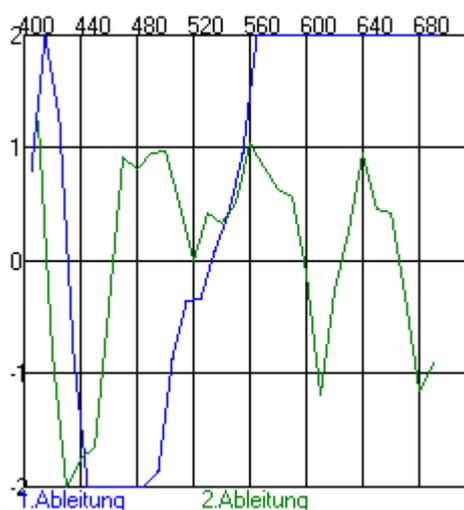
Die Ableitungen von 3 Referenzmarken sehen Als Beispiel so aus:



Referenz Marke 40 Pf blau (Forschungsfarbe 2)



Referenz Marke 40 Pf grün (Forschungsfarbe 3)



Referenz Marke 40 Pf hellrot (Forschungsfarbe 5)

Die 2. Ableitung sieht wie folgt aus:



Hier erkennt man, dass sich die Farbe Blau und hellblau im Bereich von 638 nm gut trennen lassen. Übrigens lassen sich auch Messfehler aus der Kurve erkennen. Bei den blauen Punkten fällt im Bereich von 410 nm auf, dass sie wohl mit unterschiedlichen Messgeräten gemessen wurden.

Zur Zuordnung der Marken zu einer Farbe können Kriterien (z.B. 1. Ableitung Nullstelle im Bereich 460 nm bis 470 nm) herangezogen werden und hiermit eine Zuordnung der Marken zu bestimmten Forschungsfarben gemacht werden.

## Anhang Korrelation (aus Wikipedia)

# Korrelationskoeffizient

aus Wikipedia, der freien Enzyklopädie

Der **Korrelationskoeffizient** oder die **Produkt-Moment-Korrelation** (von [Bravais](#) und [Pearson](#), daher auch *Pearson-Korrelation* genannt) ist ein dimensionsloses Maß für den Grad des *linearen* Zusammenhangs ([Zusammenhangsmaße](#)) zwischen zwei mindestens [intervallskalierten Merkmalen](#). Er kann Werte zwischen  $-1$  und  $1$  annehmen. Bei einem Wert von  $+1$  (bzw.  $-1$ ) besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen. Wenn der Korrelationskoeffizient den Wert  $0$  aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab. Allerdings können diese ungeachtet dessen in *nicht-linearer* Weise voneinander abhängen. Damit ist der Korrelationskoeffizient kein geeignetes Maß für die (reine) [stochastische Abhängigkeit](#) von Merkmalen.

Je nachdem, ob der lineare Zusammenhang zwischen zeitgleichen Messwerten zweier verschiedener Merkmale oder derjenige zwischen zeitlich verschiedenen Messwerten eines einzigen Merkmals betrachtet wird, spricht man entweder von der [Kreuzkorrelation](#) oder von der [Autokorrelation](#) (siehe [Zeitreihenanalyse](#)).

## Inhaltsverzeichnis

[\[Verbergen\]](#)

- [1 Definitionen](#)
  - [1.1 Korrelationskoeffizient \(nach Pearson\)](#)
  - [1.2 Empirischer Korrelationskoeffizient](#)
- [2 Eigenschaften](#)
- [3 Voraussetzungen für die Pearson-Korrelation](#)
  - [3.1 Skalierung](#)
  - [3.2 Normalverteilung](#)
  - [3.3 Linearitätsbedingung](#)
  - [3.4 Signifikanzbedingung](#)
- [4 Bildliche Darstellung und Interpretation](#)
- [5 Verteilung des Korrelationskoeffizienten](#)
- [6 Partieller Korrelationskoeffizient](#)
- [7 Fisher-Transformation](#)
- [8 Schätzung der Korrelation zwischen nicht-metrischen Variablen](#)
- [9 Siehe auch](#)
- [10 Literatur](#)
- [11 Weblinks](#)

## Definitionen [\[Bearbeiten\]](#)

### Korrelationskoeffizient (nach Pearson) [\[Bearbeiten\]](#)

Für zwei quadratisch integrierbare [Zufallsvariablen](#)  $X$  und  $Y$  mit jeweils positiver [Varianz](#)  $\text{Var}$  ist der Korrelationskoeffizient (Pearsonscher Maßkorrelationskoeffizient) durch

$$\begin{aligned}\varrho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} \\ &= \frac{E((X - EX)(Y - EY))}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}\end{aligned}$$

definiert, wobei  $E$  den [Erwartungswert](#)-Operator und  $\sqrt{\text{Var}(X)}$  die [Standardabweichung](#) von  $X$  bezeichnet. Weitere übliche Schreibweisen sind  $\text{Kor}(X, Y) := r_{XY} := \varrho(X, Y)$

Ferner heißen  $X, Y$  *unkorreliert*, falls  $\text{Cov}(X, Y) = 0$ .

### Empirischer Korrelationskoeffizient [\[Bearbeiten\]](#)

Sind für die beiden Zufallsvariablen lediglich zwei Messreihen  $x_1, x_2, \dots, x_n$  und  $y_1, y_2, \dots, y_n$  bekannt, so wird der *empirische Korrelationskoeffizient* berechnet durch

$$\text{Kor}_e(X, Y) := \varrho_e(X, Y) := r_{xy} := \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Dabei sind

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

die empirischen Erwartungswerte  $X$  und  $Y$  anhand der Messreihe.

Sind diese Messreihenwerte [z-transformiert](#) mit  $z_i := \frac{x_i - \bar{x}}{s_x}$ , wobei  $s_x = \sqrt{\text{Var}(X)}$  die [Streuung](#) bezeichnet, gilt auch:

$$\text{Kor}(X, Y) := \varrho(X, Y) := r_{xy} := \frac{\sum z_x z_y}{df},$$

wobei  $df$  für die Freiheitsgrade bei der Berechnung der zugrundegelegten Streuung zu ersetzen ist (also  $n$  oder  $(n - 1)$ ).

### Eigenschaften [\[Bearbeiten\]](#)

Mit der Definition des Korrelationskoeffizienten gilt unmittelbar

- $\text{Kor}(X, Y) = \text{Kor}(Y, X)$  bzw.  $r_{xy} = r_{yx}$
- $\text{Kor}(X, X) = 1$
- $\text{Kor}(X, -X) = -1$

Mit der [Cauchy-Schwarz-Ungleichung](#) sieht man, dass

- $\text{Kor}(X, Y) \in [-1, 1]$ .

Durch Optimieren ergibt sich, dass  $Y = aX + b$  [fast sicher](#) genau dann, wenn  $|\text{Kor}(X, Y)| = 1$ .

Sind die Zufallsgrößen  $X$  und  $Y$  voneinander unabhängig, dann gilt:

- $\text{Kor}(X, Y) = 0$ . Die Umkehrung dieses Satzes gilt jedoch im allgemeinen nicht (Voraussetzungen für die Umkehrung des Satzes sind z.B. in [en: normally distributed and uncorrelated does not imply independent](#) skizziert).

## Voraussetzungen für die Pearson-Korrelation [\[Bearbeiten\]](#)

Der Korrelationskoeffizient nach Pearson erlaubt Aussagen über statistische Zusammenhänge unter folgenden Bedingungen:

### Skalierung [\[Bearbeiten\]](#)

Der Pearsonsche Korrelationskoeffizient liefert korrekte Ergebnisse bei [intervallskalierten](#) und bei [dichotomen](#) Daten. Für niedrigere Skalierungen existieren andere Korrelationskonzepte (z. B. [Rangkorrelationskoeffizienten](#)).

### Normalverteilung [\[Bearbeiten\]](#)

Beide Variablen müssen annähernd [normalverteilt](#) sein. Bei zu starken Abweichungen von der Normalverteilung muss auf [Rangkorrelationen](#) zurückgegriffen werden.

### Linearitätsbedingung [\[Bearbeiten\]](#)

Zwischen den Variablen  $x$  und  $y$  wird ein linearer Zusammenhang vorausgesetzt. Diese Bedingung wird in der Praxis nicht selten ignoriert; daraus erklären sich mitunter enttäuschend niedrige Korrelationen, obwohl der Zusammenhang zwischen  $x$  und  $y$  bisweilen trotzdem hoch ist. Ein einfaches Beispiel für einen *hohen Zusammenhang trotz niedrigem Korrelationskoeffizienten* ist die [Fibonacci-Folge](#). Alle Zahlen der Fibonacci-Folge sind durch ihre Position in der Reihe durch eine mathematische Formel exakt determiniert (siehe die Formel von Jacques-Philippe-Marie Binet in [Fibonacci-Folge](#)). Der Zusammenhang zwischen der Positionsnummer einer Fibonacci-Zahl und der Größe der Zahl ist vollkommen determiniert. Dennoch beträgt der Korrelationskoeffizient zwischen den Ordnungsnummern der ersten 360 Fibonacci-Zahlen und den betreffenden Zahlen nur 0,20; das bedeutet, dass in erster Näherung nicht viel mehr als 4 % (= 0,2<sup>2</sup>) der Varianz durch den Korrelationskoeffizienten erklärt werden und 96 % der Varianz „*unerklärt*“ bleiben. Der Grund ist die Vernachlässigung der Linearitätsbedingung, denn die Fibonacci-Zahlen wachsen progressiv an: In solchen Fällen ist der Korrelationskoeffizient nicht korrekt interpretierbar. Eine mögliche Alternative, welche ohne die Voraussetzung der [Linearität](#) des Zusammenhangs auskommt, ist die [Transformation](#).

## Signifikanzbedingung [\[Bearbeiten\]](#)

Ein Korrelationskoeffizient  $> 0$  bei positiver Korrelation bzw.  $< 0$  bei negativer Korrelation zwischen  $x$  und  $y$  berechtigt *nicht a priori* zur Aussage, es bestehe ein statistischer Zusammenhang zwischen  $x$  und  $y$ . Eine solche Aussage ist nur gültig, wenn der ermittelte Korrelationskoeffizient *signifikant* ist. Der Begriff 'signifikant' bedeutet hier 'signifikant von Null verschieden'. Je höher die Anzahl der Wertepaare  $(x, y)$  und das Signifikanzniveau sind, desto niedriger darf der *Absolutbetrag* eines Korrelationskoeffizienten sein, um zur Aussage zu berechtigen, zwischen  $x$  und  $y$  gebe es einen linearen Zusammenhang. Ein [t-Test](#) zeigt, ob die Abweichung des ermittelten Korrelationskoeffizienten von Null auch signifikant ist.

## Bildliche Darstellung und Interpretation [\[Bearbeiten\]](#)

Sind zwei Merkmale vollständig miteinander korreliert (d. h.  $|r| = 1$ ), so liegen alle Messwerte in einem 2-dimensionalen Koordinatensystem auf einer Geraden. Bei einer perfekten positiven Korrelation ( $r = +1$ ) steigt die Gerade; wenn die Merkmale perfekt negativ miteinander korreliert sind ( $r = -1$ ), sinkt die Gerade. Besteht zwischen zwei Merkmalen eine sehr hohe Korrelation, sagt man oft auch, sie erklären dasselbe.

Je kleiner der Betrag von  $r$ , desto kleiner der lineare Zusammenhang. Für  $r = 0$  kann der statistische Zusammenhang zwischen den Messwerten nicht mehr durch eine eindeutig steigende oder sinkende Gerade dargestellt werden. Dies ist z. B. der Fall, wenn die Messwerte [rotationssymmetrisch](#) um den Mittelpunkt verteilt sind. Dennoch kann dann ein nicht-linearer statistischer Zusammenhang zwischen den Merkmalen gegeben sein. Umgekehrt gilt jedoch: Wenn die Merkmale statistisch unabhängig sind, nimmt der Korrelationskoeffizient stets den Wert 0 an.

Der Korrelationskoeffizient ist kein Indiz eines [ursächlichen Zusammenhangs](#) zwischen den beiden Merkmalen: Die Besiedlung durch Störche im Süd-[Burgenland](#) korreliert zwar positiv mit der dortigen Geburtenzahl, doch das bedeutet noch lange keinen kausalen Zusammenhang. Trotzdem ist ein statistischer Zusammenhang gegeben. Dieser leitet sich aber aus einem dritten, vierten etc. Faktor ab, wie in unserem Beispiel der Industrialisierung, der Wohlstandssteigerung, die einerseits den Lebensraum der Störche einschränkten und andererseits zu einer Verringerung der Geburtenzahlen führten.

Der Korrelationskoeffizient kann schon gar kein Indiz über die *Richtung eines Zusammenhanges* sein: Steigen die Niederschläge durch die höhere Verdunstung oder steigt die Verdunstung an, weil die Niederschläge mehr Wasser liefern? Bedingt das eine das andere möglicherweise in beiderlei Richtung?

Ob ein gemessener Korrelationskoeffizient *groß* oder *klein* ist, hängt stark von der Art der untersuchten Daten ab. Bei psychologischen Fragebogendaten werden z. B. Werte bis ca. 0,3 häufig als klein angesehen, während man ab ca. 0,8 von einer sehr hohen Korrelation spricht. Das Quadrat des Korrelationskoeffizienten  $r^2$  nennt man [Bestimmtheitsmaß](#). Es gibt in erster Näherung an, wieviel Prozent der Varianz durch die untersuchte Beziehung erklärt werden. *Beispiel: Bei  $r = 0,3$  bzw.  $0,8$  werden 9 % bzw. 64 % der gesamten auftretenden Varianz im Hinblick auf einen statistischen Zusammenhang erklärt.*

## Verteilung des Korrelationskoeffizienten [\[Bearbeiten\]](#)

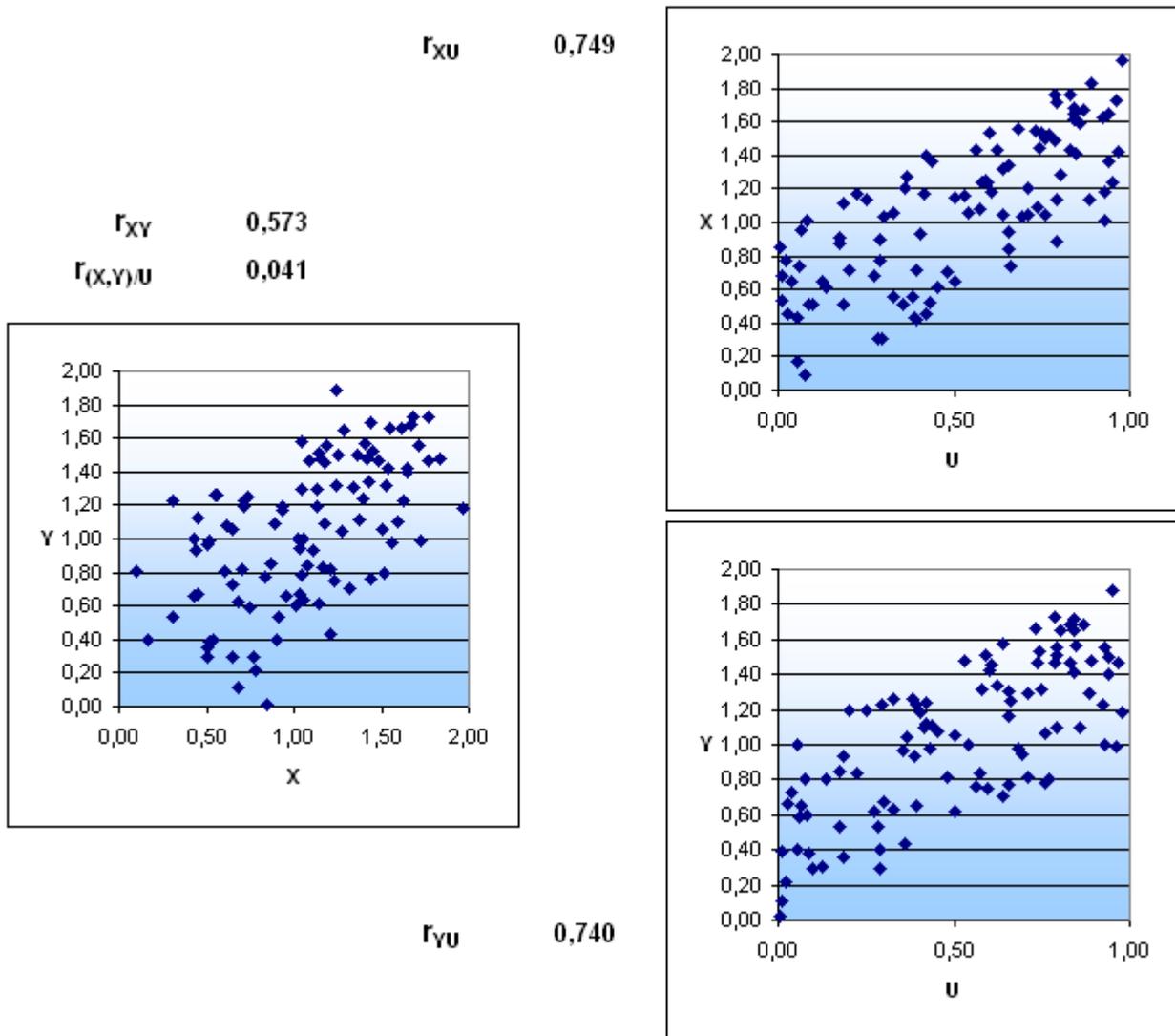
Korrelationskoeffizienten sind nicht normalverteilt. Sie verteilen sich eingipfelig nach rechts verzerrt ([rechtssteil](#) oder auch [linksschief](#)). Vor der Berechnung von Vertrauensbereichen ([Konfidenzintervallen](#)) muss daher erst eine Korrektur der Verteilung mit Hilfe der Fisher-Transformation (s.u.) vorgenommen werden. Die transformierten Korrelationen sind dann annähernd normalverteilt. Das so errechnete Konfidenzintervall wird danach wieder auf die ursprünglichen Korrelationen zurückgeführt. Konfidenzintervalle von Korrelationen liegen in aller Regel *unsymmetrisch* bezüglich ihres Mittelwerts.

## Partieller Korrelationskoeffizient [\[Bearbeiten\]](#)

Eine Korrelation zwischen zwei Zufallsvariablen X und Y kann unter Umständen auf einen gemeinsamen Einfluss einer dritten Zufallsvariablen U zurück geführt werden. Um solch einen Effekt zu messen, gibt es das Konzept der partiellen Korrelation. Die „partielle Korrelation von X und Y unter U“ ist gegeben durch

$$r_{(X,Y)/U} = \frac{r_{XY} - r_{XU} \cdot r_{YU}}{\sqrt{(1 - r_{XU}^2)(1 - r_{YU}^2)}}.$$

Das folgende Bild zeigt ein Beispiel:



Zwischen X und Y besteht eine merkliche Korrelation. Betrachtet man die beiden rechten Plots, so erkennt man, dass X und Y jeweils stark mit U korrelieren. Die beobachtete Korrelation zwischen X und Y basiert nun fast ausschließlich auf diesem Effekt.

## Fisher-Transformation [\[Bearbeiten\]](#)

Der Korrelationskoeffizient  $r$  ist rechtssteil [unimodal](#) verteilt. Durch die Transformation

$$f(r) = 0,5(\ln(1 + r) - \ln(1 - r))$$

ist  $f(r)$  annähernd normalverteilt. Das auf Basis dieser Normalverteilung errechnete Konfidenzintervall der Form

$$z_1 \leq \mu \leq z_2$$

wird sodann retransformiert zu

$$r_1 = (e^{2z_1} - 1)/(e^{2z_1} + 1)$$

$$r_2 = (e^{2z_2} - 1)/(e^{2z_2} + 1)$$

$$\mu(r) = (e^{2\mu} - 1)/(e^{2\mu} + 1).$$

Das Konfidenzintervall für die Korrelation lautet sodann

$$r_1 \leq \mu(r) \leq r_2.$$

## Schätzung der Korrelation zwischen nicht-metrischen Variablen [\[Bearbeiten\]](#)

Die Schätzung der Korrelation mit dem Korrelationskoeffizient nach Pearson setzt voraus, dass beide Variablen [intervallskaliert](#) und [normalverteilt](#) sind. Dagegen können die [Rangkorrelationskoeffizienten](#) immer dann zur Schätzung der Korrelation verwendet werden, wenn beide Variablen mindestens [ordinalskaliert](#) sind. Die Korrelation zwischen einer [dichotomen](#) und einer intervallskalierten und normalverteilten Variablen kann mit der [punktbiserialen Korrelation](#) geschätzt werden. Die Korrelation zwischen zwei dichotomen Variablen kann mit dem [Vierfelderkorrelationskoeffizienten](#) geschätzt werden.

### Siehe auch [\[Bearbeiten\]](#)

- [Zusammenhangsmaß](#)
- [Bestimmtheitsmaß](#)
- [Transinformation](#)
- [Kontingenztafel](#)
- [Streudiagramm](#)
- [Kovarianz](#)
- [Faktorenanalyse](#)

### Literatur [\[Bearbeiten\]](#)

- Hartung, Joachim: *Statistik*, 12. Auflage, Oldenbourg Verlag 1999, S. 561 f, [ISBN 3-486-24984-3](#)
- Zöfel, Peter: *Statistik für Psychologen*, Pearson Studium 2003, München, S. 154.

### Weblinks [\[Bearbeiten\]](#)

 [Wikibooks: Einfache Erläuterung des Korrelationskoeffizienten](#) – Lern- und Lehrmaterialien

- <http://mathworld.wolfram.com/CorrelationCoefficient.html> – Darstellung des Korrelationskoeffizienten als [Kleinste-Quadrate-Schätzer](#)
- <http://www.zotteljedi.de/software/kkf/index.xhtml> – Demonstration der Kreuzkorrelation, wie sie in der digitalen Signalverarbeitung Anwendung findet, um verrauschte Signale zu finden.

Von „<http://de.wikipedia.org/wiki/Korrelationskoeffizient>“