

# Korrelation, Kausalität und Regression

Zusammenhänge werden in den verschiedenen Bereichen der Mathematik adressiert, beispielsweise bei funktionalen Zusammenhängen aber auch in der Statistik. Außerhalb der Mathematik wird der Begriff ebenso vielfältig verwendet, wie die folgenden Zitate verdeutlichen:

„Je stärker Unternehmen sich wandeln, desto nachhaltiger werden sie“ (Bertelsmann Stiftung, 2024),

„Klimawandel führt zu weltweiten Einkommensverlusten“ (Tagesschau, 2024),

„Einer Analyse zufolge verschlimmerte die Erderwärmung den Starkregen um bis zu zehn Prozent“ (n-tv, 2024).

Alle drei Zitate adressieren die allgemeine Frage: „Inwieweit hängen die zwei Größen  $X$  und  $Y$  zusammen?“. Die differenzierte und umfassende Beantwortung dieser Frage führt unweigerlich zu den (mathematischen) Begriffen **Korrelation**, **Kausalität** und **Regression**, die nicht nur mit dem Begriff des Zusammenhangs in Verbindung stehen, sondern auch untereinander eng verflochten sind.

## 1. Korrelation



### Aus dem Lehrplan PLUS:

„Sie berechnen die Werte von Korrelationskoeffizienten und interpretieren diese als Maß für den Zusammenhang zweier Größen. Dass bei der Betrachtung dieses Zusammenhangs Fehlinterpretationen auftreten können (z. B. Simpson-Paradoxon), weisen sie anhand von geeigneten Beispielen graphisch und rechnerisch nach.“

Der Begriff der Korrelation beschreibt einen **Zusammenhang** zwischen zwei oder mehr Variablen<sup>1</sup>. Die Korrelation gibt somit an, inwieweit eine Änderung in einer Variablen mit einer Änderung in einer anderen Variablen einhergeht<sup>2</sup>. Während im Allgemeinen Variablen auch nicht-linear zusammenhängen können, beschränkt sich der Begriff der Korrelation im allgemeinen Sprachgebrauch oftmals auf lineare Zusammenhänge – also solche, welche sich näherungsweise durch eine lineare Funktion beschreiben lassen. Im Falle eines nicht-linearen Zusammenhangs kann man abgrenzend von *kurvi-linearen* Zusammenhängen bzw. nicht-linearen Korrelationen sprechen<sup>3</sup>. Kurvi-lineare Zusammenhänge können dabei beliebige funktionale Zusammenhänge aufweisen, die einer Kurve folgen. Bei Verwendung des Begriffs Korrelation ist daher aus dem Kontext zu schließen, ob damit nur lineare oder allgemeinere Zusammenhänge gemeint sind. Im Folgenden wird der Begriff Korrelation in seiner allgemeinen Form genutzt.

### 1.1. Graphische Überprüfung einer Korrelation

Hat man Daten gegeben, so ist es für die Analyse der Korrelation wichtig, zunächst erste Vermutungen über die Art (linear bzw. kurvi-linear) und Stärke des Zusammenhangs zwischen

<sup>1</sup> Kuckartz et al. (2013)

<sup>2</sup> Schäfer (2010)

<sup>3</sup> Schäfer (2016)

den Variablen auf Basis des **visuellen Eindrucks** der Datenpunkte in einem Streudiagramm aufzustellen<sup>4</sup>. Die Stärke beschreibt dabei, wie gut die Datenpunkte auf einer (gegebenenfalls nicht-linearen) Kurve liegen.

Im **linearen Fall** liegt eine positive Korrelation vor, wenn die Datenpunkte um eine Gerade mit positiver Steigung herum liegen, und eine negative Korrelation, wenn die Datenpunkte um eine Gerade mit negativer Steigung herum liegen. Je weniger die Datenpunkte um die Gerade herum „streuen“, desto stärker ist die Korrelation. Neben der graphischen Betrachtung kann die Stärke der Korrelation im linearen Fall auch rechnerisch ermittelt werden – nämlich mithilfe des Korrelationskoeffizienten.

## 1.2. Berechnung des Korrelationskoeffizienten

Im Falle eines linearen Zusammenhangs lässt sich die Stärke der Korrelation mathematisch mithilfe des sogenannten **Korrelationskoeffizienten  $r$**  (nach Bravais-Pearson) ausdrücken.

Um den Korrelationskoeffizienten zu bestimmen, wird das Konzept der **Kovarianz  $s_{xy}$**  benötigt. Diese beschreibt den empirischen Zusammenhang der beiden Variablen  $x$  und  $y$ . Für die Berechnung werden die Datenpunkte  $(x_1, y_1)$  bis  $(x_N, y_N)$  und die jeweiligen arithmetischen Mittel  $\bar{x}, \bar{y}$  betrachtet. Es gilt:

$$s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y}).$$

Betrachtet man die Definition der Kovarianz lässt sich erkennen, dass diese dimensionsbehaftet ist, also die Größe der Kovarianz abhängig von der Maßeinheit ist.

Um dieses Problem zu umgehen, wird die Kovarianz normiert durch Division durch die Standardabweichungen  $s_x := \sqrt{s_{xx}}$  und  $s_y := \sqrt{s_{yy}}$

$$s_x^2 := s_{xx} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Dieses normierte Maß heißt Korrelation. Es gilt:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

Dabei sind  $s_x$  und  $s_y$  die jeweiligen **Standardabweichungen**. Der Wert  $r_{xy}$  heißt **Korrelationskoeffizient**. Dieser ist dimensionslos und es gilt  $|r_{xy}| \leq 1$ . Dabei kann der Korrelationskoeffizient positiv, negativ, oder gleich null sein. Bei einem positiven Korrelationskoeffizient liegen die Datenpunkte um eine Gerade mit positiver Steigung. Bei einem negativen Korrelationskoeffizient liegen die Datenpunkte um eine Gerade mit negativer Steigung. Je größer  $|r_{xy}|$ , desto kleiner ist die Abweichung der Datenpunkte von einer gedachten Geraden (sog. Trendgerade).<sup>5</sup>

<sup>4</sup> Fox (2000)

<sup>5</sup> Vgl. Scheid und Vogl (2021)

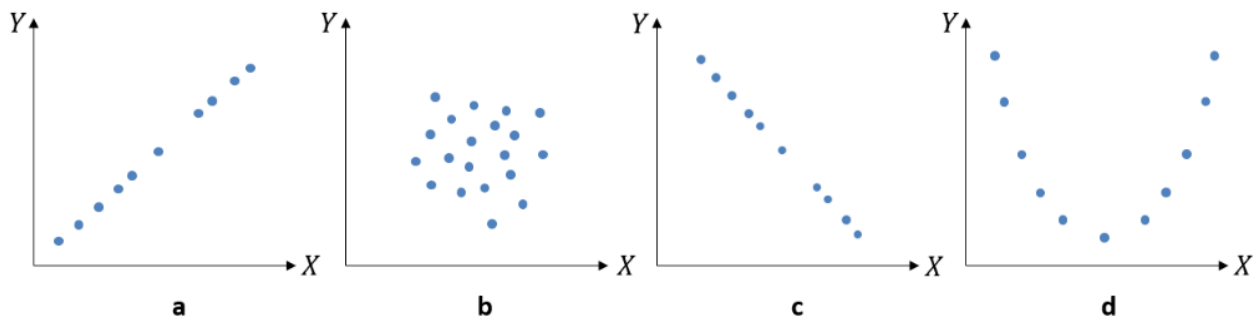


Abbildung 1: Datenpunkte mit unterschiedlichen Korrelationskoeffizienten **a**  $r = 1$ ,  
**b**  $r \approx 0$ , **c**  $r = -1$ , **d**  $r = 0$

### 1.3. Interpretation des Korrelationskoeffizienten

Der Korrelationskoeffizient gibt nur die Stärke des *linearen* Zusammenhangs an. Ist  $|r_{xy}| = 0$ , lässt sich daraus nicht schließen, dass es keinen statistischen Zusammenhang zwischen den Daten gibt, es gibt lediglich keinen linearen Zusammenhang; es können aber andere funktionale Zusammenhänge bestehen, beispielsweise quadratische (Abbildung 1d). Zudem impliziert ein hoher Korrelationskoeffizient nicht, dass ein linearer Zusammenhang vorliegt. Deshalb sollte immer zunächst graphisch überprüft werden, ob ein linearer Zusammenhang naheliegt.

Die Interpretation des Werts des Korrelationskoeffizienten hängt sehr stark vom jeweiligen Forschungsgebiet ab<sup>6</sup>. Eine mögliche Faustregel bietet Cohen<sup>7</sup>, der die Werte folgendermaßen interpretiert: Für  $|r| < 0,1$  liegt kein linearer Zusammenhang vor, für  $0,1 \leq |r| < 0,3$  ein geringer, für  $0,3 \leq |r| < 0,5$  ein mittlerer und für  $|r| \geq 0,5$  besteht ein starker linearer Zusammenhang. Je größer der Betrag des Korrelationskoeffizient  $r$  ist, desto kleiner ist die Abweichung der Datenpunkte von einer gedachten Geraden (sog. Trendgerade).

Die Stärke eines nicht-linearen Zusammenhangs lässt sich nicht direkt durch einen solchen Koeffizienten angeben.



#### Auf den Punkt:

- **Korrelation** ist ein statistisches Maß zur Beschreibung der Beziehung zweier Variablen. Im allgemeinen Sprachgebrauch bezeichnet Korrelation meist einen linearen Zusammenhang.
- Der **Korrelationskoeffizient** gibt die Stärke des linearen Zusammenhangs zweier Variablen an.

## 2. Kausalität



#### Aus dem Lehrplan PLUS:

„Sie berechnen die Werte von Korrelationskoeffizienten und interpretieren diese als Maß für den Zusammenhang zweier Größen. Dass bei der Betrachtung dieses

<sup>6</sup> Rasch et al. (2021)

<sup>7</sup> Cohen (1988)

Zusammenhangs Fehlinterpretationen auftreten können (z. B. Simpson-Paradoxon), weisen sie anhand von geeigneten Beispielen graphisch und rechnerisch nach.“

Besteht eine Korrelation zwischen zwei Variablen, so stellt sich die Frage, ob diese einander auch beeinflussen, d.h. in einem kausalen Zusammenhang stehen. Ein kausaler Zusammenhang bzw. eine *Kausalität* ist eine **Ursache-Wirkungs-Beziehung**<sup>8</sup>. Sie steht in einem Abhängigkeitsverhältnis zu Korrelation: Aus Kausalität folgt Korrelation, aber (im Allgemeinen) nicht umgekehrt. Mathematisch ausgedrückt heißt das, dass Korrelation eine notwendige, aber keine hinreichende Bedingung für Kausalität ist<sup>9</sup>. Denn Drittvariablen können Einflüsse auf die Korrelation besitzen<sup>10</sup>, das heißt anstelle eines kausalen Zusammenhangs sind für eine Korrelation zwischen zwei Variablen auch **alternative Erklärungen** denkbar. Abbildung 2 zeigt dabei zwei wesentliche, jedoch nicht alle möglichen Fälle von alternativen Erklärungen für eine Korrelation ohne eine Kausalität zwischen den Variablen  $x$  und  $y$ . Zum einen könnte  $x$  lediglich kausal auf eine dritte Variable  $z$  wirken (ggf. auch auf mehrere), welche wiederum auf  $y$  wirkt, sodass keine direkte Ursache-Wirkungs-Beziehung zwischen  $x$  und  $y$  besteht. Man spricht in diesem Fall von einer **intervenierenden Variablen**  $z$ . Zum anderen könnte eine Drittvariable  $z$  die beiden Variablen  $x$  und  $y$  beeinflussen, wodurch eine Korrelation zwischen letzteren zustande kommt (**Scheinkorrelation**)<sup>11</sup>.

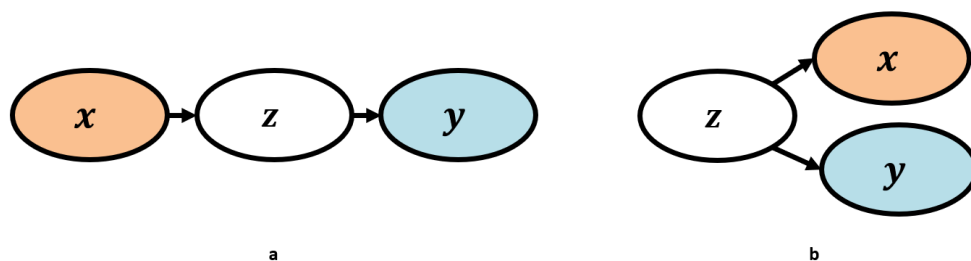


Abbildung 2: Mögliche Wirkungszusammenhänge, bei der die Variablen  $x$  und  $y$  korrelieren, aber nicht kausal zusammenhängen, sondern durch eine Drittvariable  $z$  beeinflusst werden **a** Intervenierende Variable **b** Scheinkorrelation

Solche alternativen Erklärungen müssen ausgeschlossen werden, um von einer Korrelation auf eine Kausalität schließen zu können<sup>12</sup>. Konkret sind für Kausalität neben einer Korrelation zwei weitere Bedingungen zu überprüfen<sup>13</sup>.

Die **erste Bedingung** drückt aus, dass die verursachende Variable zeitlich *vor* der beeinflussten Variable auftreten muss. Damit wird insbesondere auch eine umgekehrte Wirkrichtung zwischen den beiden Variablen ausgeschlossen. Diese naheliegende Eigenschaft verleiht dem Begriff der Kausalität eine zeitliche Dimension.

Die **zweite Bedingung** bedeutet (vereinfacht), dass es keine Drittvariable geben darf, sodass die Beziehung zwischen den beiden ursprünglichen Variablen verschwindet, wenn die Drittvariable

<sup>8</sup> Weiber & Mülhhaus (2014)

<sup>9</sup> Kuckartz et al. (2013)

<sup>10</sup> Schäfer (2016)

<sup>11</sup> Rasch et al. (2021)

<sup>12</sup> Kuckartz et al. (2013)

<sup>13</sup> Kenny (1979)

kontrolliert bzw. entfernt wird. Die Beziehung darf sich also nicht nur durch die Drittvariable „erklären“ oder verändern lassen.

**i Auf den Punkt:**

- Eine **Kausalität** bezeichnet eine Ursache-Wirkungs-Beziehung zwischen zwei Variablen.
- Um eine Kausalität nachzuprüfen, müssen zusätzlich zu einer Korrelation zwei Bedingungen erfüllt sein. Eine Korrelation allein, liefert keine Aussage über Kausalität (z.B. Scheinkorrelation).

## 3. Regressionsanalyse



**Aus dem Lehrplan PLUS:**

„Die Schülerinnen und Schüler [...] erläutern anhand von Streudiagrammen die Grundidee der linearen Regression und stellen Gleichungen von Regressionsgeraden auf, um Werte von Variablen im Sinne einer Vorhersage zu schätzen.“

Nun stellt sich abschließend die Frage, mit welchen statistischen Verfahren der Zusammenhang zwischen Variablen bzw. der Einfluss einer Variable auf die andere quantifiziert werden kann. Ein zentrales Instrument zur Untersuchung solcher Zusammenhänge ist die **Regressionsanalyse**. Sie ist eng verknüpft mit der Korrelation, aber erlaubt nicht nur die Stärke und Richtung des Zusammenhangs zwischen zwei Variablen zu bestimmen, sondern auch die Werte einer Variablen zu nutzen, um die Werte der zweiten Variablen (auf Grundlage von Daten) zu ermitteln<sup>14</sup>. Die Durchführung einer Regression kann somit verschiedene Ziele besitzen. Daten können *retrospektiv* analysiert werden, um Trends und Muster zu identifizieren. *Prognostische* Analysen ermöglichen die Vorhersage zukünftiger Trends oder Ereignisse basierend auf Daten, wenn auch mit Unsicherheit.

Als Voraussetzung für eine Regressionsanalyse muss ein Abhängigkeitsverhältnis zwischen den beiden Variablen vorliegen, d.h. eine (unabhängige) Variable hat einen Einfluss auf die (abhängige) Variable. Zudem wird angenommen, dass sich die Datenpunkte näherungsweise – das heißt bis auf einen kleinen zufälligen Fehler – durch eine Funktion  $f$  mit  $y = f(x)$  darstellen lassen. Dabei bezeichnet  $x$  die unabhängige Variable, welche die abhängige Variable  $y$  beeinflusst. Um diese Abweichung darzustellen, wird die Regressionsgleichung durch den Fehlerterm  $e$  ergänzt. Grundsätzlich können verschiedene Funktionstypen gewählt werden.

### 3.1. Einfache lineare Regression

Bei einer einfachen linearen Regression geht man zum Beispiel davon aus, dass sich die Datenpunkte näherungsweise durch eine Gerade beschreiben lassen. Diese soll nun so bestimmt werden, so dass sie „gut“ zu den tatsächlichen Datenpunkten „passt“. Wir erhalten die folgende **Regressionsgleichung**:

$$y = mx + t$$

$y$  ist der abhängige Variablenwert (der zu schätzende Wert oder die Zielvariable),  $x$  ist der unabhängige Variablenwert (der Zeitpunkt oder die Beobachtungsperiode),  $m$  ist die Steigung

<sup>14</sup> Schäfer (2016)

der Geraden, die den Anstieg des Trends darstellt, und  $t$  ist der y-Achsenabschnitt, der den Schnittpunkt der Geraden mit der y-Achse repräsentiert<sup>15</sup>.

Mit dem Fehlerterm erhalten wir die folgende Gleichung:

$$y = mx + t + e$$

mit  $m, t \in \mathbb{R}$ .

Das Ziel der linearen Trendschätzung besteht darin, die Werte von  $m$  und  $t$  zu finden, die am besten die Beziehung zwischen den unabhängigen und abhängigen Variablen in den Daten beschreiben.

Ein Verfahren hierfür ist die Methode der kleinsten Quadrate.

### 3.1.1. Methode der kleinsten Quadrate

Eine Methode der linearen Regression ist die **Methode der kleinsten Quadrate**, bei der die Summe der quadrierten Abweichungen zwischen den beobachteten Datenpunkten, genauer den Werten der abhängigen Variablen, und den geschätzten Werten minimiert wird.

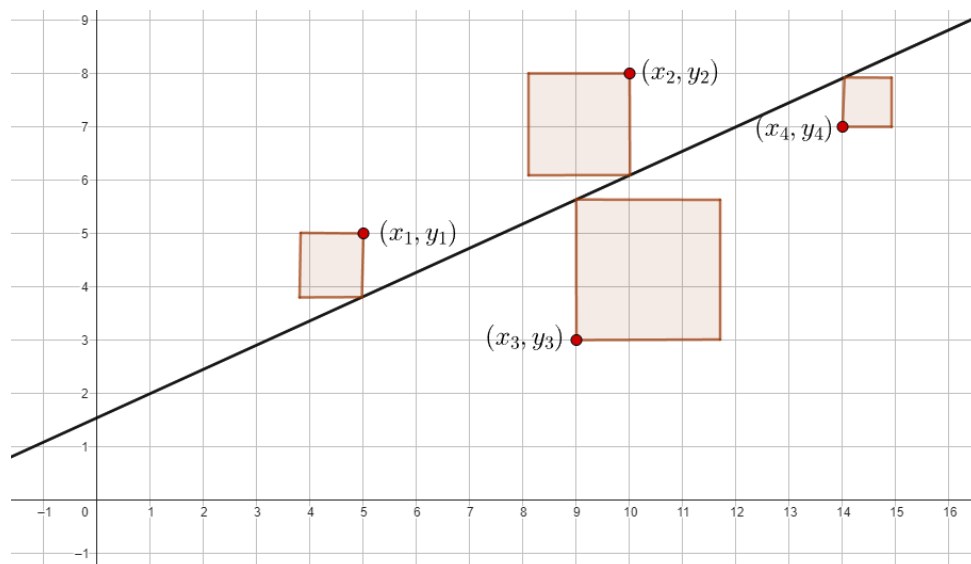


Abbildung 3 Applet Methode der kleinsten Quadrate; Verfügbar unter <https://www.dimu.mathematik.uni-wuerzburg.de/> (Bivariate Daten – Zusammenhang und Regression, Abb. 25)

Für  $N$  Datenpunkte  $(x_i, y_i)$  soll also

$$S(m, t) = \sum_{i=1}^N (mx_i + t - y_i)^2$$

minimiert werden. Dabei sind  $mx_i + t$  die durch die Regression geschätzten Werte.

Rechnerisch wird das Minimum des obigen Terms mithilfe der Ableitung bestimmt. Man leitet separat nach  $m$  und  $t$  ab, setzt beide Terme gleich 0 und löst nach  $m$  und  $t$  auf.

<sup>15</sup> Scheid und Vogl (2021)



### Vertiefung:

Ableitung nach  $m$ :

$$S'_t(m) = 2 \sum_{i=1}^N (mx_i + t - y_i)x_i \quad (I)$$

Ableitung nach  $t$ :

$$S'_m(t) = 2 \sum_{i=1}^N (mx_i + t - y_i) \quad (II)$$

Wir setzen beide Ableitungen gleich 0 und lösen (II) nach  $t$  auf:

$$I) \quad 2 \sum_{i=1}^N (mx_i + t - y_i)x_i = 0 \quad | : 2$$

$$\Leftrightarrow \sum_{i=1}^N mx_i^2 + tx_i - y_ix_i = 0$$

$$\Leftrightarrow \sum_{i=1}^N mx_i^2 + \sum_{i=1}^N tx_i - \sum_{i=1}^N y_ix_i = 0$$

$$\Leftrightarrow m \sum_{i=1}^N x_i^2 + t \sum_{i=1}^N x_i - \sum_{i=1}^N y_ix_i = 0$$

$$II) \quad 2 \sum_{i=1}^N (mx_i + t - y_i) = 0 \quad | : 2$$

$$\Leftrightarrow \sum_{i=1}^N mx_i + \sum_{i=1}^N t - \sum_{i=1}^N y_i = 0$$

$$\Leftrightarrow m \sum_{i=1}^N x_i + Nt - \sum_{i=1}^N y_i = 0 \quad | + \sum_{i=1}^N y_i - m \sum_{i=1}^N x_i$$

$$\Leftrightarrow Nt = \sum_{i=1}^N y_i - m \sum_{i=1}^N x_i \quad | : N$$

$$\Leftrightarrow t = \frac{\sum_{i=1}^N y_i - m \sum_{i=1}^N x_i}{N}$$

Wir setzen den letzten Ausdruck in (I) ein und lösen nach  $m$  auf:

$$m \sum_{i=1}^N x_i^2 + \left( \frac{\sum_{i=1}^N y_i - m \sum_{i=1}^N x_i}{N} \right) \sum_{i=1}^N x_i - \sum_{i=1}^N y_ix_i = 0$$

$$\Leftrightarrow m \sum_{i=1}^N x_i^2 + \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i}{N} - m \frac{(\sum_{i=1}^N x_i)^2}{N} - \sum_{i=1}^N y_ix_i = 0$$



$$\begin{aligned} &\Leftrightarrow m \left( \sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{N} \right) + \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i}{N} - \sum_{i=1}^N y_i x_i = 0 \\ &\Leftrightarrow m \left( \sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{N} \right) = \sum_{i=1}^N y_i x_i - \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i}{N} \\ &\Leftrightarrow m = \frac{\sum_{i=1}^N y_i x_i - \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i}{N}}{\sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{N}} \end{aligned}$$

Wir ersetzen  $\sum_{i=1}^N x_i$  durch  $N\bar{x}$  und  $\sum_{i=1}^N y_i$  durch  $N\bar{y}$ , wobei  $\bar{x}$  und  $\bar{y}$  die arithmetischen Mittel der Datenpunkte sind. Damit ist

$$m = \frac{\sum_{i=1}^N y_i x_i - \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i}{N}}{\sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{N}} = \frac{N \sum_{i=1}^N y_i x_i - N^2 \bar{x} \bar{y}}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2}$$

und

$$t = \frac{\sum_{i=1}^N y_i - m \sum_{i=1}^N x_i}{N} = \frac{N\bar{y} - mN\bar{x}}{N} = \bar{y} - m\bar{x}$$

### 3.1.2. Güte einer linearen Regression

Das **Bestimmtheitsmaß** oder **Determinationskoeffizient**  $R^2$  gibt die Genauigkeit einer Regressionsbeziehung an. Er ist der Anteil der sogenannten erklärten Varianz an der Gesamtvarianz:

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

Dabei ist  $\hat{y}$  die Variable, welche die (durch die Regression) ermittelten Werte der abhängigen Variablen angibt. Die erklärte Varianz  $s_{\hat{y}}$  die Varianz von  $\hat{y}$ .

Im einfachen linearen Modell gilt  $R^2 = r_{xy}^2$ .

Gilt  $R^2 = 1$ , so lässt sich die gesamte Variation in einer Variablen durch die Veränderung der anderen Variablen erklären. Liegt  $R^2$  hingegen nahe 0, so stellt die Trendgerade keine gute Anpassung an die Daten dar.



#### Vertiefung: Nicht lineare Modelle

Für **nicht-lineare Modelle** stellt das Bestimmtheitsmaß im Allgemeinen kein adäquates Gütekriterium dar, weshalb die graphische Veranschaulichung hier von großer Bedeutung ist. Insbesondere sollte ein spezifisches nicht-lineares Modell nach Möglichkeit theorie- bzw. forschungsbasiert ausgewählt und begründet werden. Ist dies aufgrund fehlender oder widersprüchlicher Vorarbeiten nicht möglich, bietet die (einfache) nichtparametrische



Regression eine induktive Möglichkeit zur Modellgenerierung<sup>16</sup>. Rechnerisch lassen sich nicht-lineare Modelle mit Hilfe von allgemeineren Kenngrößen wie dem *AIC* oder *BIC* vergleichen<sup>17</sup>, die aufgrund ihrer Komplexität hier aber nicht näher erläutert werden. In manchen nicht-linearen Modellen besteht allerdings auch die Möglichkeit, die Variablen (im zulässigen Wertebereich) so zu transformieren, dass zwischen den beiden Variablen eine lineare Regression durchgeführt wird und deren Vorteile genutzt werden können<sup>18</sup>.



#### **Vertiefung: Konfidenzintervalle**

Die Korrelations- und Regressionsanalyse sind Techniken, die dazu verwendet werden, Beziehungen zwischen Variablen zu analysieren und Schätzungen für Parameter abzuleiten. Die Parameter werden jedoch auf der Grundlage der jeweiligen Stichprobe bestimmt und entsprechen nicht unbedingt dem exakten wahren Wert der Gesamtpopulation. Ein **Konfidenzintervall** ergänzt daher die Schätzungen der Parameter, indem es Informationen über die Unsicherheit oder Genauigkeit dieser Schätzungen liefert.



#### **Auf den Punkt:**

- Die **lineare Regression** ist ein statistisches Verfahren, mithilfe dessen eine lineare Funktion ermittelt wird, die die Daten möglichst gut erklärt.
- Die **Methode der kleinsten Quadrate** ist ein Verfahren, das die optimalen Werte der Parameter in der Geradengleichung ermittelt.
- Die Güte der Anpassung an die Datenpunkte kann mithilfe des **Bestimmtheitsmaßes** überprüft werden.

<sup>16</sup> Fox (2000)

<sup>17</sup> Schneider (2020)

<sup>18</sup> Backhaus et al. (2021)

## 4. Literatur

- Backhaus, K., Weiber, R., Weiber, T., Erichson, B. & Gensler, S. (2021). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Onleihe. E-Book. Springer Fachmedien Wiesbaden; divibib GmbH.
- Buchberger, B. (1990). Should Students Learn Integration Rules? *ACM SIGSAM Bulletin*, 24(1), 10–17. <https://doi.org/10.1145/382276.1095228>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Lawrence Erlbaum Associates.
- Fox, J. (2000). *Nonparametric simple regression: smoothing scatterplots*. *Sage university papers. Series: Quantitative applications in the social sciences: Bd. 130*. Sage Publ.
- Kenny, D. A. (1979). *Correlation and Causality*.
- Kuckartz, U., Rädiker, S., Ebert, T. & Schehl, J. (2013). *Statistik*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-19890-3>
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2021). *Quantitative Methoden 1*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-63282-6>
- Schäfer, T. (2010). *Statistik I*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92446-5>
- Schäfer, T. (2016). *Methodenlehre und Statistik*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-11936-2>
- Scheid, S. & Vogl, S. (2021). *Data Science: Grundlagen, Methode und Modelle der Statistik*. Hanser.
- Schneider, M. (2020). *Datenanalyse für Naturwissenschaftler, Mediziner und Ingenieure* (1. Aufl. 2020). Springer Spektrum. Springer Spektrum. <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflucht-1808802> <https://doi.org/10.1007/978-3-662-61866-0>
- Straesser, R. (2007). Didactics of mathematics: more than mathematics and school! *ZDM*, 39(1-2), 165–171. <https://doi.org/10.1007/s11858-006-0016-x>
- Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-35012-2>